Towards safe and stable *Clostridium botulinum* strains for research and industrial applications

Daphne Esmé Groothuis, MSc

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

July 2021

Abstract

The pathogen *Clostridium botulinum* has gained significant infamy as the causative agent of botulism, a rare but potentially fatal paralytic condition. Unfortunately, research on these dangerous pathogens is limited due to the necessary high containment facilities and their associated costs. Therefore, the initial aim of this project was to develop a strategy, applicable to all members of the highly diverse *C. botulinum* family, to efficiently generate non-toxigenic strains, which would allow the expansion of the *C. botulinum* research scope. These highly desired surrogate strains are anticipated to find wide application in both industrial and research settings.

In this context, 'safe' *C. botulinum* ATCC 3502 strains were created by deleting the neurotoxin gene using various technologies. However, the project took an unexpected turn when the unstable nature of this widely studied *C. botulinum* model strain was discovered. The cause of the ATCC 3502 mutator phenotype was pinned down via whole genome comparisons and was experimentally confirmed to be due to a partial, in-frame deletion of the *mutL* gene. Interestingly, several other group I strains bearing a similar genetic background were identified, which are, probably, also mutators. It is quite surprising that this is the first study revealing the rapid degeneration of strains of this organism, despite them being extensively investigated by different research groups. Evidently, these findings might have a significant impact on our current understanding of *C. botulinum*, affecting both the industrial and scientific community. In view of this finding, previously published results with this organism should henceforth be considered with caution.

To pursue the initial aim of this thesis, similar *C. botulinum* family members were acquired and characterised. Various strategies to circumvent their restriction barriers, which hindered reliable and efficient DNA transfer, were successfully employed. Additionally, the effectiveness of a comprehensive donor customisation pipeline for improving DNA transfer efficiencies in previously 'untouched' strains is exemplified. Subsequently, empowered by CRISPR/Cas9 technology, different non-toxigenic strains were created and phenotypically assessed. Overall, the advancements presented here constitute a solid basis for the future generation of any desired safe *C. botulinum* strain.

Acknowledgements

I would like to thank everyone who has been part of this journey; for your support, patience and wisdom. I would like to thank my supervisor, Nigel Minton, for his invaluable advice, support and for giving me the opportunity to work on this project. Also, special thanks to my second supervisor, Alan Cockayne, for his guidance and for always being there, even if it meant dressing up in moon suits to clean the botlab. I would like to thank Alexandra Bailey for teaching me how to work with anaerobes and Christopher Humphreys for showing me all the tips and tricks on how to work with *C. botulinum*. I am grateful to everyone in the office and lab for making the long hours at the SBRC more enjoyable, laughing about all the good and the bad. I would like to specifically thank Andrew Dempster for reading and commenting on my thesis. Also, I am especially thankful to Rajan Patel for being my writing buddy and having endless chats. I am very grateful to Evangelos Karvonis, the best lab technician, who prepared crazy amounts of media and who had to deal with all our lab antics.

I am forever grateful to the best team one could have ever wished for; Raquel, Maria and Ines. To Raquel, for starting the PhD and finding our way together but mostly for always being there. To Maria, for her inspiring perseverance, limitless generosity, sharing all your many diverse talents and for making Nottingham feel like home. And to Ines, for her endless energy, brilliant mind and her positivity. Whenever we girls were down you were always there to find the silver lining. Spending many hours working alone in the botlab, having you peak around the door to update me on literally everything that was going on in the outside world, was always the highlight of my day. It has been an incredible adventure together and our friendship is what I cherish most.

Thanks to my mum and dad, for being understanding and supportive, whenever I needed it the most. My brother, for his jokes and pokes to push me to the finish line. My grandma for her believe and lovely encouraging cards. And last, but not least, Laurens, for his infinite support, love and patience. I am very grateful to have had you by my side during these intense years; I couldn't have done it without you.

Finally, I am grateful to the Marie Skłodowska-Curie Actions programme (European union) for funding my research. To everyone involved in the CLOSPORE network; thank you.

To my parents

Table of Contents

Abstract ..... **i**
Acknowledgements ..... **ii**
List of Figures ..... **viii**
List of Tables ..... **x**
List of Abbreviations ..... **xi**
CHAPTER 1: General introduction **1**
1.1 *Clostridium botulinum* 2
1.2 The toxin: expression and mode of action 4
1.3 Botulinum neurotoxin: the good and the bad 7
1.3.1 The bad: botulism 7
1.3.2 Treatment and prevention 11
1.3.3 The good: cosmetic and therapeutic applications 13
1.4 The current status of *C. botulinum* research 15
1.4.1 *C. botulinum* in the food industry 15
1.4.2 *C. botulinum* in the pharmaceutical industry 17
1.4.3 *C. botulinum* in academia 19
1.5 Methods to genetically engineer and control gene expression in clostridia 20
1.5.1 ClosTron-mediated mutagenesis 21
1.5.2 Allele-coupled exchange 22
1.5.3 CRISPR/Cas9-mediated genome editing 24
1.5.4 Inducible systems to control gene expression 26
1.6 Project aim and thesis overview 28
CHAPTER 2: Materials and methods **30**
2.1 Bacterial strains 31
2.2 Growth media, buffers and supplements 31
2.3 Growth and storage conditions 33
2.4 Preparation of electrocompetent *E. coli* cells 34
2.5 Transformation of *E. coli* cells by electroporation 35
2.6 Conjugation of R702 from *E. coli* donor into *E. coli* recipient 35
2.7 Conjugation of plasmid DNA from *E. coli* donor into *C. botulinum* 36
2.8 Plasmid loss 37
2.9 Phenotypic analysis 37

2.9.1 Optical density measurements....37
2.9.2 Growth curves....38
2.9.3 Conjugation efficiency assay....38
2.9.4 Sporulation assay....39
2.9.5 Spore survival curves....39
2.9.6 Fluctuation assay....40
2.9.7 Competitive fitness assay....41
2.9.8 SDS-PAGE and Western blotting....42
2.10 General molecular biology techniques....44
2.10.1 Plasmid DNA extraction....44
2.10.2 Genomic DNA extraction....44
2.10.3 DNA amplification....45
2.10.4 Agarose gel electrophoresis....46
2.10.5 DNA purification....47
2.10.6 DNA quantification....47
2.10.7 DNA restriction by endonuclease digestion....47
2.10.8 DNA ligation....47
2.10.9 Sanger sequencing....48
2.10.10 Illumina genome sequencing....48
2.11 Plasmid construction....48
2.11.1 Construction of plasmids used in Chapter 3....48
2.11.2 Construction of plasmids used in Chapter 4....51
2.11.3 Construction of plasmids used in Chapter 5....52
2.12 Targeted genome mutagenesis....53
2.12.1 ClosTron mutagenesis....53
2.12.2 Allele-Coupled Exchange....53
2.12.3 CRISPR/Cas9-mediated mutagenesis....54
2.13 List of bioinformatics tools....55
CHAPTER 3: Development of a donor customisation pipeline to improve DNA transfer efficiencies....56
3.1 Introduction....57
3.1.1 Horizontal gene transfer....57
3.1.2 Transformation....58
3.1.3 Transduction....60
3.1.4 Conjugation....61

3.1.5 Bacterial defence mechanisms against parasites: barriers to horizontal gene transfer 63
3.1.6 Bacterial immune systems that target nucleic acids 68
3.1.7 Strategies to overcome restriction barriers 74
3.2 Objectives 77
3.3 Results and discussion 78
3.3.1 Identification of restriction modification systems 78
3.3.2 Engineering the recipient: inactivation of type I nuclease in ATCC 3502 86
3.3.3 Engineering the donor 1: customised methylation of donor DNA 89
3.3.4 Engineering the donor 2: creating an all in one package by recombineering the methyltransferases into conjugal plasmid R702 ..94
3.3.5 Engineering the donor 3: evading type IV restriction systems 104
3.4 Key outcomes 106
CHAPTER 4: Hypermutability and the role of *mutL* 108
4.1 Introduction 109
4.1.1 Bacterial genome evolution 109
4.1.2 The balance between maintaining genome integrity and evolution ... 113
4.1.3 Bacterial systems to maintain genome integrity 116
4.1.4 ATCC 3502, a strain with high population heterogeneity 122
4.2 Objectives 125
4.3 Results and discussion 126
4.3.1 A fresh start: engineering a 'new' ATCC 3502 strain using ACE and ClosTron mutagenesis 126
4.3.2 Deletion of the neurotoxin in the genetically unstable ATCC 3502 strain using CRISPR/Cas9-based technology 135
4.3.3 First attempt to establish the mutation frequency of ATCC 3502 140
4.3.4 Investigating the role of DNA mismatch repair protein MutL 147
4.3.5 Competitive fitness: ATCC 3502 mutator *vs* ATCC 3502^{+} 152
4.3.6 Going back in time: the origin of mutator ATCC 3502 Δ*mutL*231 ... 164
4.3.7 Published ATCC 3502 research 171
4.3.8 What is a wild type genome? 173
4.3.9 Are there other mutators out there? 177
4.4 Key outcomes 178
CHAPTER 5: Generation and characterisation of safe *C. botulinum* strains .. 180
5.1 Introduction 181

5.1.1 The potential applications for non-toxigenic *C. botulinum* strains 181
5.1.2 Currently available safe strains 183
5.1.3 What is the perfect safe strain? 184
5.2 Objectives 186
5.3 Results and discussion 187
5.3.1 Selection of various group I *C. botulinum* type A strains 187
5.3.2 Basic characterisation of *C. botulinum* type A strains 192
5.3.3 Whole genome sequencing of *C. botulinum* type A strains 196
5.3.4 Genetic stability of *C. botulinum* type A strains 201
5.3.5 Spore heat resistance and investigating the role of the *spoVA*^{2} gene cluster 205
5.3.6 Generation and characterisation of non-toxigenic ATCC 19397 strains 211
5.3.7 Generation and characterisation of non-toxigenic NCTC 2916 strains ... .221
5.4 Key outcomes 226
CHAPTER 6: General discussion 227
6.1 Summary of findings 228
6.2 Future perspectives 230
6.3 Concluding remarks 235
Bibliography 236
Appendix 281
List of Appendix Figures 282
List of Appendix Tables 282
I. Bacterial strains 284
II. Plasmids 287
III. Representative plasmid maps 289
IV. Primers 291
V. Phenotypic characterisation of ATCC 19397 Δ*boNT/A1* mutants 299
VI. Whole genome sequencing and variant detection Chapter 4 300
VII. Whole genome sequencing and variant detection Chapter 5 329
VIII. SNPs/SNVs comparison table for ATCC 3502 350

List of Figures

Figure 1.1 Overview of normal neurotransmitter release and blockage of neurotransmitter release by botulinum neurotoxin. 7

Figure 1.2 Schematic overview of allele-coupled exchange (ACE). 23

Figure 3.1 Overview of horizontal gene transfer mechanisms..... 58

Figure 3.2 Overview of bacterial immune systems. 67

Figure 3.3 Overview of bacterial defence mechanisms that target foreign DNA. 74

Figure 3.4 Conjugation efficiency into ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916. 79

Figure 3.5 Restriction modification systems of ATCC 3502, ATCC 19397, 62A, Loch Maree and NCTC 2916. 81

Figure 3.6 Conjugation efficiency of various plasmids containing a type I RM system recognition site into ATCC 3502^{+}. 85

Figure 3.7 ClosTron-mediated inactivation of *hsdR* in strain ATCC 3502. 87

Figure 3.8 Conjugation efficiency of various plasmids containing a type I RM system recognition site into *hsdR*1600s::CT. 89

Figure 3.9 Comparison of conjugation efficiency into ATCC 3502^{+} using *E. coli* donors expressing different methyltransferases. 91

Figure 3.10 Comparison of conjugation efficiency into ATCC 19397 using *E. coli* donors expressing different methyltransferases. 93

Figure 3.11 Comparison of conjugal donors InterStellar and CA434 with and without methyltransferases in combination with ATCC 3502^{+}. 97

Figure 3.12 Comparison of conjugal donors InterStellar and CA434 with and without methyltransferases in combination with ATCC 19397. 98

Figure 3.13 Comparison of the conjugation efficiency achieved with different conjugal donors carrying the recombinant R702 plasmid with both MTases. 101

Figure 3.14 Acquired mutations in the recombineering cassette. 103

Figure 3.15 Comparison of conjugative plasmid transfer from donor strains CA434 and sExpress into 62A, Loch Maree and NCTC 2916. 105

Figure 4.1 Sporulation capacity of ATCC 3502 (IFR) *pyrE*$^{-}$, Δ*boNT/A1 pyrE*$^{-}$, Δ*boNT/A1* and ΔClA *pyrE*$^{-}$ mutants. 124

Figure 4.2 Truncation of *pyrE* gene in strain ATCC 3502..... 128

Figure 4.3 Phenotypic characterisation of ATCC 3502 Unilever *hsdR*1600s::CT and *pyrE*$^{-}$ mutants. 130

Figure 4.4 Deletion of *boNT/A1* gene in strain ATCC 3502..... 137

Figure 4.5 Phenotypic characterisation of ATCC 3502 Unilever Δ*boNT/A1* mutants. 138

Figure 4.6 Preliminary fluctuation assay with ATCC 3502.....143

Figure 4.7 Visualisation of the Δ*mutL*231 deletion identified in ATCC 3502. 145

Figure 4.8 Deletion and complementation of *mutL*231 and *mutL* in strain ATCC 19397.....149

Figure 4.9 Growth profiles of ATCC 19397 *mutL* mutants. 151

Figure 4.10 Mutation frequency of ATCC 3502 and ATCC 19397 *mutL* mutants.152

Figure 4.11 Competitive fitness assay of ATCC 3502^{+} *vs* ATCC 3502 Δ*mutL*231. 157

Figure 4.12 PCR screening of ATCC 3502 stocks for Δ*mutL*231. 166

Figure 4.13 PCR screening of ATCC 3502 SBRC stocks for Δ*mutL*231. 167

Figure 4.14 ATCC 3502 strain provenance. 170

Figure 4.15 PCR screening of published ATCC 3502 mutants for Δ*mutL*231. 173

Figure 4.16 Mutations in mutator genes *mutS* and *mutL* in published type A genome sequences.....178

Figure 5.1 Schematic diagram of botulinum neurotoxin gene cluster and possible non-toxic variants.....185

Figure 5.2 Basic phenotypic characterisation of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916.....194

Figure 5.3 Mutation frequency of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916. 201

Figure 5.4 Frequency distribution of the number of Rif^{R} mutants per culture for ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916. 203

Figure 5.5 Mutations identified in the *rpoB* gene of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916.....204

Figure 5.6 Survival of spores of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916 after 99°C thermal treatment. 206

Figure 5.7 Deletion of the *spoVA*^{2} operon in strain NCTC 2916.....209

Figure 5.8 Phenotypic characterisation of NCTC 2916 Δ*spoVA*^{2}.....211

Figure 5.9 Deletion of *boNT/A1*, *boNT/A1* gene cluster (ClA) and generation of i*boNT/A1*^{RY} in strain ATCC 19397.....214

Figure 5.10 Phenotypic characterisation of ATCC 19397 Δ*boNT/A1*, ΔClA and i*boNT/A1*^{RY}. 216

Figure 5.11 Western blot analysis of ATCC 19397 Δ*boNT/A1*, ΔClA and WT supernatants.....219

Figure 5.12 Deletion of *boNT/A1*, *boNT/B, boNT/A1* gene cluster (ClA) and *boNT/B* gene cluster (ClA) in strain NCTC 2916. 223

Figure 5.13 Phenotypic characterisation of NCTC 2916 Δ*boNT/A1*, Δ*boNT/B*, ΔClA, ΔClB and double mutants. 225

List of Tables

Table 1.1 Characteristics of the four distinct *C. botulinum* groups.....3

Table 2.1 Growth media and buffers used in this study.....31

Table 2.2 Supplements used in this study 33

Table 2.3 Ligation mixtures to create pM plasmids.....50

Table 3.1 Type I and type II restriction modification systems and associated methylation patterns detected by REBASE 83

Table 4.1 List of mutations identified in all non-mutator genome sequences.....175

Table 4.2 List of mutations identified in at least two, but not all non-mutator genome sequences.....176

Table 5.1 *C. botulinum* strains available from the NCTC (operated by Public Health England). 190

Table 5.2 Summary of the maximum achieved conjugation efficiencies of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916. 195

Table 5.3 Overview of the theoretically optimal customized donor strains for ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916 to achieve maximum conjugation efficiency. 195

Table 5.4 Comparison of SNPs detected between 62A NCBI (CFSAN064329) and 62A NCTC (NCTC 3805) to the published genomes of ATCC 3502 and ATCC 19397. 200

List of Abbreviations

5-FOA	5-fluoroorotic acid
Abi	Abortive infection
ACE	allele - coupled exchange
Apr	Apramycin
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
ATPase	ATP synthase
BER	base excision repair
BoNT	botulinum neurotoxin
bp	base pair
BREX	bacteriophage exclusion
Cas	CRISPR-associated
CDC	Centres of Disease Control and Prevention
CFU	colony forming unit
Cm	chloramphenicol
CRISPR	clustered regularly interspaced palindromic repeats
crRNA	CRISPR RNA
CSA	Campdem sporulating agar
CT	ClosTron
dH2O	distilled water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ECDC	European Centre of Disease Prevention and Control
eDNA	extracellular DNA
EDTA	ethylenediaminetetraacetic acid
Erm	Erythromycin
FDA	Food & Drug Administration
HA	haemagglutinin
hfr	high frequency of recombination
HSE	The Health and Safety Executive
IEP	intron encoded protein
IFR	Institute for Food Research
iGEM	International Genetically Engineered Machine
InDel	large insertion/deletion
IPTG	isopropyl-β-d-thiogalactoside
Kan	kanamycin

LB	Luria Bertani
LHA	left homology arm
LTEE	Long-Term Evolution Experiment
M	MTase subunits
MCS	multiple cloning site
MDS	Maximum-Depth Sequencing
Met	Methionine
MIC	Minimum Inhibitory Concentration
MMR	DNA mismatch repair
MOPS	3-(*N*-morpholino)propanesulfonic acid
MTase	methyltransferase
NCBI	National Center for Biotechnology Information
NCTC	National Collection of Type Cultures
NEB	New England Biolabs
NER	Nucleotide Excision Repair
NGS	next generation sequencing
NHEJ	Non-Homologous End Joining
NTNH	non-toxic-non-haemagglutinin
OD	optical density
ORF	open reading frame
oriT	origin of transfer
oriV	vegetative replication
pAgo	Prokaryotic Argonaute proteins
PAM	protospacer adjacent motif
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline-tween
PCR	polymerase chain reaction
PHE	Public Health England
PICIs	phage-inducible chromosomal islands
Pol	polymerase
Pro	Proline
R	REase subunits
R factor	resistance transfer factor
RAM	Retrotransposition-Activated Marker
RBS	ribosome binding site
REase	restriction endonucleases
RHA	right homology arm
Rif	Rifampicin
RM	restriction-modification
RNA	ribonucleic acid
RNase	ribonuclease

S	specificity subunit
SBRC	Synthetic Biology Research Centre
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
sgRNA	single-guide RNA
Sie	superinfection exclusion
SNP	single-nucleotide polymorphism
SNV	single-nucleotide variant
SOE	Splicing by overlap extension
ssDNA	single-stranded DNA
strp	streptomycin
TAE	Tris-acetate EDTA
TCA	trichloroacetic acid
TE	Tris-EDTA
tet	anhydrotetracycline
Tm	thiamphenicol
TMB	3,3′,5,5′-Tetramethylbenzidine
tra	conjugal transfer
tracrRNA	trans-activating CRISPR RNA
tRNA	transfer-RNA
Trp	Tryptophan
TYG	tryptone-yeast extract-sodium thioglycolate
USDA	US Department of Agriculture
UTR	untranslated region
WGS	whole genome sequecing
WHO	World Health Organisation
WT	wildtype

CHAPTER 1

General Introduction

1.1 *Clostridium botulinum*

At the beginning of the 19th century a medical officer, named Kerner, was the first to publish an accurate description of the symptoms of a severe paralytic condition, referred to as botulism. He based his clinical description on some well-documented outbreaks of "sausage poisoning" that occurred southwest of the German region Württemberg. However, he could not identify the origin of this so-called "sausage poison".[1] It took another 80 years before the responsible pathogen, initially named *Bacillus botulinus*, was discovered, following a large outbreak of botulism caused by pickled ham served at a funeral in a small Belgian village.[2] Since then, many more neurotoxin producing bacterial strains and neurotoxin variants have been identified all over the world.[3-6]

The *Clostridium botulinum* family comprises a wide variety of anaerobic bacteria with the shared ability to produce botulinum neurotoxins (BoNTs).[7] The various *C. botulinum* species have distinct metabolic and physiological characteristics which led to the classification of *C. botulinum* into four different groups (I, II, III and IV; Table 1.1).[8-12] Group I strains are mesophilic, proteolytic and form highly heat resistant spores. They primarily cause human botulism and are occasionally associated with botulism cases in animals.[13] Group II strains are psychrotrophic, non-proteolytic, produce moderately thermoresistant spores and cause human botulism. Group III strains are generally non-proteolytic mesophiles, form spores of moderate heat resistance and are mainly responsible for animal botulism. Group IV strains are also mesophilic and produce moderately heat resistant spores. It is a genetically distinct group and although Group IV strains have been isolated from human and animal specimens that displayed botulism symptoms, there is currently no definitive evidence

to confirm their association. It should be noted that within each group there are various levels of genetic diversity and hence phenotypic variability.

Table 1.1 Characteristics of the four distinct *C. botulinum* groups.[14-15]

Neurotoxigenic Clostridia	*C. botulinum* Group I (proteolytic)	*C. botulinum* Group II (non-proteolytic)	*C. botulinum* Group III	*C. botulinum* Group IV (*C. argentinense*)
Neurotoxins formed	A, B, F	B, E, F	C, D	G
Non-neurotoxigenic equivalent clostridia	*C. sporogenes*	No species name given	*C. novyi*	*C. subterminale*
Optimum growth temperature	37°C	30°C	40 °C	37 °C
Minimum growth temperature	10-12 °C	3.0 °C	15 °C	
Minimum pH for growth	4.6	5.0	5.1	
NaCL concentration preventing growth	10%	5%		6.5%
Minimum water activity for growth with NaCl/glycerol	0.94/0.93	0.97/0.94		
Spore heat resistance	$D_{121\ °C}$ = 0.21 min	$D_{82.2\ °C}$ = 2.4 min	$D_{104\ °C}$ = 0.1-0.9 min	$D_{104\ °C}$ = 0.8-1.1 min

Seven distinct neurotoxin serotypes (A to G) of this deadly toxin are currently recognised, with some toxin serotypes being identified in various bacterial strains that belong to different species groups (Table 1.1).[14, 16] The toxins can cause botulism in humans, animals or both and differ in their toxicity level, depending on the organism affected. The classification has been based on functionality studies and sequencing data, with amino acid differences between the serotypes ranging from 37.2% to 69.6%.[17-18] Additionally, sequencing has revealed the existence of chimeric/hybrid toxins and that there is considerable genetic variability within each serotype.[19] This explains the observed differences in catalytic activity among serotype family members

and why in some cases antibodies did not neutralise all toxin variants of a specific serotype.[18, 20-26] Recently, it has been proposed that a toxin variant should be considered a subtype when its amino acid sequence differs from all designated subtypes by at least 2.6%.[5] This to provide a level of clarity in the currently very complex *C. botulinum* nomenclature maze. However, whether the distinct bacterial species, now represented by the common name *C. botulinum*, should all get their own species name or if the current group classification should be maintained remains a topic of debate.[16, 27]

1.2 The toxin: expression and mode of action

To date, the molecular mechanisms behind botulism are well understood, particularly for type A toxins.[6, 28] It's a highly sophisticated process, targeting the neuroexocytosis machinery, which results in the powerful blockage of acetylcholine transmission between neurons and muscle cells, inevitably leading to muscle paralysis. It all starts with the expression of the neurotoxin and its associated, non-toxic, proteins, in the bacterial cell. Together they form an exceptionally stable protein complex that is resistant to, for example, proteolytic degradation and acidic pH.[10, 29-30] The non-toxic proteins play an important role in the protection and possibly absorption of the toxin, allowing the toxin to arrive at its destination safely and effectively.[31-32]

The genes that encode the neurotoxins and accessory proteins are always situated together, and are referred to as the botulinum locus or neurotoxin gene cluster (Cl/A-G). The cluster consists of two transcriptional units; the first encodes the universal non-toxic-non-haemagglutinin (NTNH) protein, conferring stability to BoNT when associated, and the neurotoxin; the second, located upstream of the *boNT* and *ntnh* genes, can either be the *ha* or the *orf-X* operon, which are associated with different

toxin serotypes.[17, 29, 33-34] The three proteins (haemagglutinin components) encoded by the *ha* operon form a complex directly with NTNH-BoNT. The alternative sigma factor BotR, encoded by the *botR* gene located within the toxin gene cluster, positively regulates the expression of BoNT, NTNH and HA proteins.[35-36] Currently, little is known about the *orf-X* operon, which encodes three proteins of unknown function. Several studies have found that these proteins form slightly fragile neurotoxin complexes, but further research is required to elucidate their purpose.[27, 37] Additionally, neurotoxin gene clusters with the *orf-X* operon contain gene *p47*, whose function is yet unknown.[38] Interestingly, some *C. botulinum* strains (only from group I) with the *orf-X* operon contain both the *p47* gene and the *botR* gene, normally associated with the *ha* operon.[10] Probably these 'mixed' clusters result from past DNA recombination events. Furthermore, some strains contain multiple toxin gene clusters of different serotypes.[39-42] The high diversity of bacterial strains expressing BoNTs has been attributed to the toxin gene clusters being flanked by insertion sequences and, although in some strains the botulinum locus is located on the chromosome, it is often found on plasmids, transposable elements and in phage genomes which aid horizontal gene transfer.[19, 43-50]

The neurotoxin and associated proteins are released by the cell via an unknown mechanism and subsequently the 150 kDa toxin is cleaved by proteases into a light chain (~50 kDa) and a heavy chain (~100 kDa), which remain connected via a disulphide bridge until they enter the neuronal cytosol, where they encounter reducing conditions.[51-53] This 'nicking' of the neurotoxin into its di-chain form increases toxicity substantially.[54] Proteolytic Group I *C. botulinum* strains express proteases which cleave the single 150-kDa polypeptide to fully activate the toxin. The proteolytic enzyme(s) responsible for the post-expression modification of the toxin

have not yet been identified. The non-proteolytic *C. botulinum* Group II, III and IV strains lack these enzymes, and therefore neurotoxin activation depends on host proteases.

Next, the neurotoxin enters the host's system by crossing the intestinal barrier or endothelial barrier and subsequently disseminates through the body via the bloodstream and lymph system, to reach its final destination; neurons.[55-56] The action of BoNT at the neuron can be roughly divided into four steps (Figure 1.1): (i) the heavy chain C-terminal domain binds to surface receptors on the presynaptic membrane of nerve cells; (ii) neurotoxin uptake occurs via receptor-mediated trapping of BoNT into recycling vesicles (transcytosis); (iii) conformational changes in the neurotoxin are triggered by acidification of the vesicle lumen and the disulphide bridge linking the light chain to the heavy chain is reduced, allowing the translocation of the unfolded light chain into the neuron's cytosol, and; (iv) in the pH neutral cytosol the neurotoxin light chain refolds and cleaves components of the synaptic vesicle docking/fusion complex (soluble N-ethylmaleimide sensitive factor attachment protein receptors; SNAREs), preventing the fusion of acetylcholine-containing vesicles with the synaptic membrane (exocytosis).[14, 57-62] The light chain of the various toxin serotypes (A-G) recognize different binding sites and cleave at specific SNARE protein sites.[18, 63] One of the reasons why BoNTs are so incredibly lethal is their long duration of activity.[6, 64] They can persist inside a neuron for several weeks up to months, with type A being the longest-acting one.[65-67] Hence, the inhibition of exocytosis can persist for months, despite the regeneration of SNAREs.

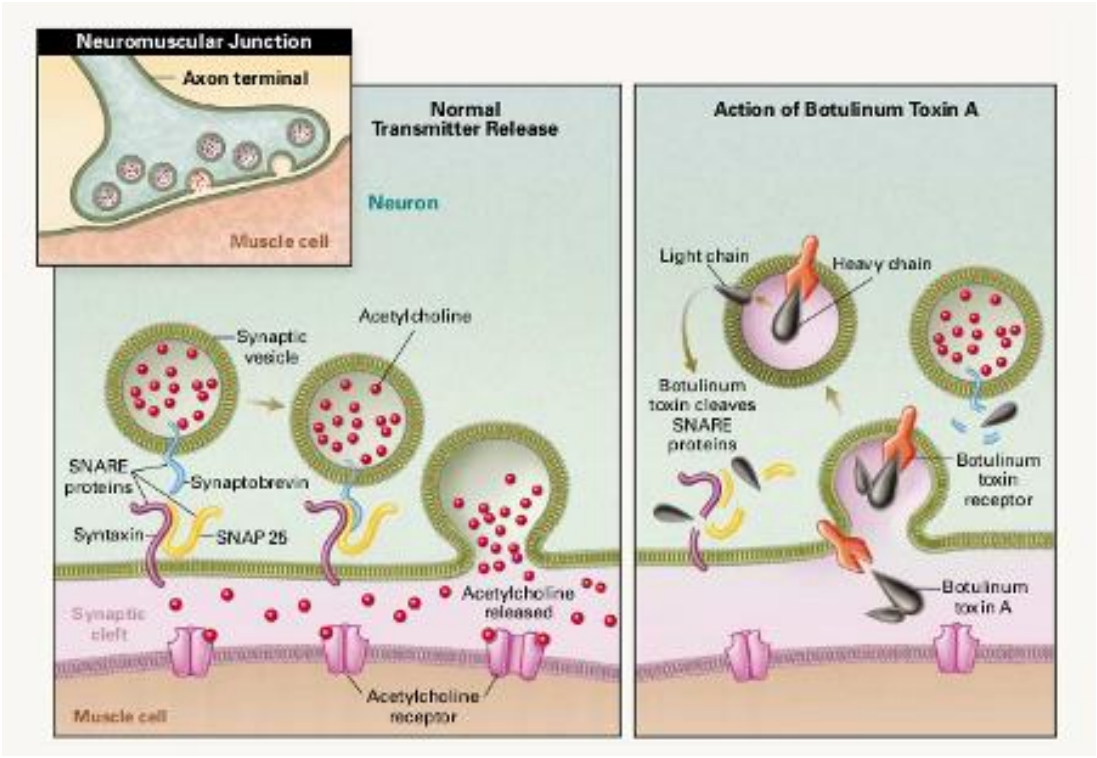


Figure 1.1 Overview of normal neurotransmitter release and blockage of neurotransmitter release by botulinum neurotoxin. In healthy neurons acetylcholine release from vesicles into the synaptic cleft is facilitated by SNARE proteins. The released neurotransmitter subsequently bind to ion-channel receptors embedded in the muscle cell membrane, initiating a sequence of steps that ultimately leads to muscle contraction. *C. botulinum* neurotoxin inhibits the release of acetylcholine, thereby causing muscle paralysis. First, the toxin's heavy chain binds to presynaptic membrane receptors of the neuron. During the reformation of synaptic vesicles from the cellular membrane receptor bound toxin gets internalised. Subsequently, the toxin light chain exits the vesicle into the cytosol, where it cleaves the SNARE proteins, which results in the inhibition of acetylcholine release into the synaptic cleft. Figure adapted from Rowland (2002).[62]

1.3 Botulinum neurotoxin: the good and the bad

1.3.1 The bad: botulism

Botulism refers to the rare paralytic condition caused by botulinum neurotoxin. The symptoms that characterize botulism are descending, symmetrical paralysis which usually starts in the head and neck muscles, blurred vision, difficulty speaking, dry mouth and generalized weakness. If the condition progresses into a severe form, the muscle paralysis leads to respiratory problems and mechanical ventilation becomes

vital.[68-69] Depending on the route of entry and toxin dose, the incubation period is usually 12 to 72 hours for symptoms to present themselves.[14] BoNTs are the most potent natural toxins currently known; it has been estimated that approximately 0.2-0.3 ng/kg in mice and 1 ng/kg in human of BoNT/A is a lethal parental dose, with 30 ng being sufficient to cause botulism in an adult human when ingested.[70-74] The World Health Organisation (WHO) has estimated that 15% of the severe botulism cases is fatal.[75]

C. botulinum strains are ubiquitous in the environment and are prevalent in soil, sediments, waters, cadavers and are occasionally found in the intestinal tract of healthy animals, such as pigs, birds and fish.[33, 76-87] They can grow and produce neurotoxin in anaerobic environments, such as canned foods and animal cadavers. Fortunately, botulism is rare, but due to the high strain diversity and them being widespread in the environment, there is always a chance *C. botulinum* will grow when favourable anaerobic conditions arise. There are several clinical forms of botulism, where classification is based on the toxin's route of entry into the body, and can be caused by an infection (colonisation of *C. botulinum*) or an intoxication (e.g., ingestion of toxin). The Centres of Disease Control and Prevention (CDC) lists five types of botulism: (i) foodborne botulism; (ii) iatrogenic botulism; (iii) wound botulism; (iv) infant botulism, and; v) adult intestinal botulism (extremely rare).[88] For all types the neurological symptoms and disease progression are similar, however, the presence of potential secondary symptoms, such as constipation, nausea and vomiting, depends on the botulism type.

C. botulinum infections are quite rare and occur when *C. botulinum* spores are able to colonize anaerobic wound tissue (necrotic abscess) or the human gastrointestinal tract. In necrotizing wounds occasionally favourable anaerobic conditions arise for spores

to germinate and outgrow into toxin-producing cells. Wound infections are primarily associated with intravenous drug abusers using needles or drugs contaminated with *C. botulinum* spores but can also occur after a traumatic injury.[89] Usually symptoms appear several days after injection but it can even take a couple of weeks.

Adult and infant intestinal botulism might occur after *C. botulinum* spores have been ingested by humans (or other animals), usually with highly disrupted gut microbiota where competitive bacterial flora is absent.[90] *C. botulinum* is unable to colonise the intestines of healthy adults, even though the intestinal content is anaerobic, and it is therefore referred to as an opportunistic pathogen. Medical treatments, such as extensive antibiotic usage, gastrointestinal surgery and bowel anomalies (e.g., Crohn's disease), can lead to a severely disrupted gut microbiota, and thus create an opportunity for *C. botulinum* to colonise the intestines and produce toxin *in situ*. Infants are more susceptible to *C. botulinum* colonisation of the intestine because new-borns still have to develop their gut microflora.[91-92] In addition to the common symptoms described above, signs of intestinal botulism in infants might include constipation, poor feeding and a weak cry. Luckily, infants usually make a full recovery. Adult intestinal botulism is extremely rare, is generally difficult to diagnose and the clinical presentation is similar to other types of botulism, but may also include gastrointestinal symptoms, distended abdomen, abdominal pain and constipation.[90]

In contrast to *C. botulinum* infections, where spores are ingested or contaminate a wound, foodborne botulism is caused by the uptake of preformed toxin. The issue of food spoilage by *C. botulinum* arises when food or feed, contaminated with spores, is sealed into anaerobic packaging.[72, 86, 93] Subsequently, spores can germinate into toxin-secreting vegetative cells. If there is no need to heat the food product to boiling point before consumption, the toxin will remain in its active form. After toxin ingestion, it

passes through the digestive track and epithelial wall, to eventually enter the bloodstream. Outbreaks of foodborne botulism in humans are typically related to home canned foods and linked to group I *C. botulinum* strains.[72] In addition to the general symptoms listed above, signs of foodborne botulism might also include vomiting, nausea, stomach pain and diarrhoea. Unfortunately, due to the condition being so rare, it is often misdiagnosed; the most commonly listed differential diagnosis were myasthenia gravis, stroke, psychiatric disorders, Guillain-Barré syndrome, poliomyelitis and other poisonings/intoxications.[88] The severity of the symptoms depends on the toxin type and dose.[71]

Iatrogenic botulism is a relatively new form of intoxication caused by inappropriate administration of the neurotoxin for cosmetic and therapeutic purposes.[88, 94-97] BoNT injections gained popularity in the second half of the previous century, and although it is generally safe, it can occasionally lead to serious local or systemic complications.[98] Usually the paralytic effects outside of the target muscle are limited to adjacent muscles and the adverse effects are generally mild to moderate. The more severe systemic complications are referred to as iatrogenic botulism and it usually presents itself with the common symptoms, such as general muscle weakness, slurred speech, blurred vision and in severe cases difficulty breathing.

The extreme lethal potency of BoNTs unfortunately also caught the eye of people intending to do harm. Several cases have been reported where BoNTs have been trialled as a biological weapon for warfare and bioterrorism.[99-101] A deliberate release via, for example, aerosols or the contamination of food and beverages (water systems), could lead to a large number of fatalities and flooding of the health care system.[69, 71, 102] Therefore, BoNTs are considered a great biosecurity threat and are classified as a Category A threat agent by the Centers for Disease Control and Prevention Select

Agent Program.[99, 103] To prevent strains from falling into the wrong hands, records are kept of the laboratories which have *C. botulinum* in their strain collection and the transfer of strains between laboratories is heavily restricted. However, with *C. botulinum* strains being ubiquitous in nature, there is always the possibility to isolate BoNT-producing strains from the environment.

1.3.2 Treatment and prevention

The treatment of botulism is mainly supportive (e.g., mechanical ventilation, fluid and nutritional support) and may include the immediate administration of antitoxins to prevent further progression of the disease.[88] Ideally, antitoxin is given within 24 hours after the onset of symptoms, because it can only neutralise circulating toxin; after neurotoxins are internalised by the neuron they can no longer be neutralised.[104] However, the intravenous infusion of the antitoxin should be done with caution due to the risk of an immediate hypersensitivity reaction (usually in less than 2 hours) and potentially anaphylaxis.[105] In severe botulism cases mechanically-assisted respiration and other supportive care is generally required for 2 to 3 months but can last for extended periods of time.[69] Furthermore, for patients with botulism resulting from an infection, additional surgery and treatment with antibacterial agents might be required. The use of antibiotics is not recommended, as lysis of *C. botulinum* cells could increase the amount of free toxin, thereby aggravating the symptoms. Theoretically a complete recovery is possible because BoNTs are not cytotoxic and do not cause axonal degeneration. The BoNT concentration in neurons will gradually decrease over time, while SNAREs are cautiously regenerated which will eventually lead to the restoration of neuromuscular connections. Nonetheless, recovery is usually a slow process; paralysis and muscle weakness can last for several months or even years,

depending on the type of BoNT, dose, route of entry, type of neuromuscular junction (skeletal muscles or autonomic nerves) and animal affected.[106]

In the past, people at high risk of being exposed to botulinum neurotoxin, such as laboratory workers and military personnel, were vaccinated. The vaccination program was ended however, as the vaccine production was discontinued due to declining potency and efficiency.[107-108] Nowadays, actual use of vaccinations is scarce since its effectiveness is disputed. Widespread vaccination of the population is not proportional to the threat; not only because the disease is extremely rare, but also because it could interfere with the expanding therapeutic applications of BoNTs.[109-110] Even the value of vaccinating high risk workers, such as laboratory staff, is debatable, as laboratory accidents are rare due to high safety standards.[111] Additionally, the large diversity in toxin serotypes complicates vaccine development; preferably a vaccine provides high levels of efficacy against all types. Nevertheless, there have been ongoing efforts to develop next generation vaccines against botulism.[108] After all, it might not be such a bad idea to have an effective vaccine stored on the shelves, just in case its use as a bioweapon suddenly takes off.

Despite new vaccine variants showing promising efficacy against botulism, the most obvious way to reduce the risk of getting botulism is by taking preventative measures and creating awareness. The CDC states that most naturally occurring cases of botulism are preventable. Wound botulism, and wound infections in general, can be prevented by keeping lesions clean. The risk of getting botulism from intravenous drug abuse can be minimised by using clean needles and by avoiding drug injection into muscles or necrotic tissue. The cause of infant botulism is often unknown and therefore difficult to prevent.[92] For now, honey remains the only identified avoidable risk factor; honey is known to contain *C. botulinum* spores and has been linked to

various infant botulism cases. It is therefore recommended to not feed honey to babies (less than 12 months old). Many cases of foodborne botulism, however, are preventable by practicing good food hygiene, applying adequate storage conditions (refrigerating) and sufficiently heating high risk food products. Most cases result from contaminated home-canned, preserved or fermented foods. Foodborne botulism can be prevented by inactivating the spores before food is anaerobically packed, inhibiting spore germination or by destroying the toxin during cooking through sufficient heating at 80°C. In the food industry a heat treatment of 121°C for 3 minutes has become the minimum standard to ensure the inactivation of *C. botulinum* spores in canned foods and is referred to as the 'botulinum cook'.[72]

1.3.3 The good: cosmetic and therapeutic applications

The interesting observation that BoNTs mainly inhibit neurotransmitter release in peripheral cholinergic nerve terminals of the skeletal and autonomic nervous system, while mental functions and sensory signals are unaffected, inspired medical professionals to investigate its potential to treat human syndromes characterized by hyper function of such nerve terminals.[1, 73] Initially, it was trialled as a drug to relieve hypercontraction of small eyelid muscles (blepharospasm), other facial spams and cervical dystonia (repetitive involuntary contractions of neck muscles).[106] BoNT injections for the treatment of blepharospasm was first approved by the Food and Drug Administration (FDA) in 1989 and its usage for therapeutic applications has been growing ever since. Today, BoNTs are used as a therapeutic agent to treat a variety of medical indications, including limb spasticity, essential tremor, cerebral palsy, dystonia, migraines, urologic and gastrointestinal disorders, excessive sweating and chronic pain.[112-113] Additionally, the use of BoNTs for the treatment of various conditions in animals is being explored.[114-115] All currently licensed BoNT

preparations for clinical use by the FDA use BoNT/A1 or BoNT/B1. The three biggest brands on the market today are Botox/Vistabel (Allergan Inc.), Dysport/Azzalure (Ipsen) and Xeomin/Bocouture (Merz Pharmaceutical Gmbh).

The great success of BoNTs for therapeutic purposes can be attributed to a unique combination of favourable pharmacological properties; their remarkable selectivity, extreme potency, limited diffusion when injected locally and the complete reversibility of the neuroparalysis caused by the BoNTs.[106] Recovery of function after toxin-induced paralysis usually takes 3 to 5 months and can be repeated countless times extended over many years, without sustaining permanent neuronal damage.[116-118] The toxin has to be injected carefully though, to avoid iatrogenic botulism, especially because there is currently no drug available that can reverse or neutralise its effects once it has reached the interior of neurons. Patients who have accidentally been injected with a BoNT 'overdose' will have to wait until its effects naturally cease.[107] Despite their high level of toxicity, over the years it has become evident that the use of BoNT based therapeutics is very safe and efficacious when administered by trained personnel.[112, 119-120]

The many therapeutic applications of BoNTs, however, is not what they are most famous for; botulinum neurotoxin is better known by the general public under the brand name BOTOX®: 'this stuff that can magically make wrinkles disappear'. Interestingly, its remarkable effectiveness in smoothening facial lines was a rather serendipitous discovery that was noticed when used therapeutically.[121-122] The use of BoNTs for aesthetic medicine was first approved by the FDA in 2002 and is now a billion dollar global industry.[106] It is estimated that approximately half of the produced BoNT/A1 doses are used for cosmetic applications.

Given that when administered locally effects only last for approximately 4 to 5 months, it has to be injected frequently. Despite BoNT therapy often being continued over many years, they very rarely elicit an immunogenic response, leading to secondary treatment failure.[123] Various treatment and product-related factors can influence the development of neutralising antibodies, including toxin dose, the presence of toxoids (inactive or denatured toxin), the injection site (lymph node rich regions are more likely to produce immunogenic response) and injection frequency.[106] To minimise the chance of developing an adaptive immune response it is recommended to administer the lowest effective dose and maximise injection intervals.[124]

1.4 The current status of *C. botulinum* research

The complex mechanism of action of BoNTs inside the host has been studied quite extensively, however, relatively little is known about the organisms that produce them.[6] Research efforts have been limited because of the need for expensive specialised lab facilities and *C. botulinum* being listed as a schedule 5 agent further hampers *C. botulinum* research across disciplines. Nonetheless, there is quite some interest in this intriguing organism, evident from the large number (324) of genomes sequenced to date, which is more than for any other *Clostridium* spp., besides *Clostridium difficile*. The current status of *C. botulinum* research in the food industry, pharmaceutical industry and academia will be discussed briefly in the next three sections.

1.4.1 *C. botulinum* in the food industry

Food products contaminated with botulinum neurotoxin due to *C. botulinum* spores germinating inside the packaging is one of food companies' worst nightmares. An

outbreak can have major medical and economic consequences due to the potential severity of the condition and the high fatality risk.[72, 125] Improving food safety is thus very important for both customers and industry. The food industry has remained extremely aware of possible *C. botulinum* food contamination and both novel and established intervention strategies are routinely tested for efficacy and efficiency.[126]

C. botulinum spores are ubiquitous in the environment and can thus easily enter the food chain.[127-130] Fortunately, outbreaks involving commercially produced food products are extremely rare due to significant efforts by the food industry to prevent spore outgrowth.[72, 131-132] *C. botulinum* spores, especially the extremely heat resistant ones produced by group I strains, are particularly challenging to eliminate from food products, as they can survive harsh conditions, such as extreme temperatures and radiation.[133] Therefore, low acid canned foods are always treated with the botulinum cook (121°C for 3 min) just to ensure that any potential *C. botulinum* spores are inactivated.[14]

A second challenge for the food industry are the psychrotrophic group II strains, which can grow in low oxygen packages at refrigerating temperatures.[9, 134] Group II strains are particularly a concern in the context of minimally heated, chilled foods like smoked or dried meat and fish products. With the recent increase in popularity of such foods and most spores being able to survive the mild heat treatment, the food industry is exploring other possibilities to control the outgrowth of spores, such as artificial food additives, biopreservation, active packaging and hurdle technology.[135-140]

To determine the efficacy and efficiency of new preservation technologies food is inoculated with the pathogenic organism and subsequently its growth is monitored under the new preservation regime. However, to avoid the hazards and restrictions

associated with *C. botulinum*, surrogate organisms or non-toxic variants are used instead.[141] For example, non-toxic *Clostridium sporogenes* strains are biochemically and morphologically very similar to group I strains and are commonly used as a surrogate for *C. botulinum* in food challenging tests.[142] Currently, few surrogate organisms are available for Group II *C. botulinum* strains, hence food challenging studies have to be performed with the toxigenic strains.[131, 143] Additionally, the broad diversity in botulinum toxin producing strains, even within groups, further complicates the assessment of preservation techniques. Therefore, challenge tests are usually carried out using a mixture of strains to account for this variability.

Modelling of microbiological processes is becoming more popular and could be used to minimize the need for challenge tests with *C. botulinum*. Very few predictive models addressing the effects of certain preservatives on growth, thermal death and neurotoxin production have been published to date.[14, 144-147] By definition, a model is a simplified representation of reality but nonetheless they can be very informative. Potentially, the large scale analysis of the genetic diversity amongst *C. botulinum* strains combined with systems biology modelling and omics approaches will lead to a better understanding of what determines the survival and growth of these bacteria in food and may aid the identification of novel intervention strategies.[148]

1.4.2 *C. botulinum* in the pharmaceutical industry

Contrary to the food industry, where *C. botulinum* is associated with costs, the pharmaceutical industry makes billions of dollars each year by selling the purified neurotoxin for therapeutic and cosmetic applications. To drive market growth, BoNT producing companies are always looking for ways to improve their products and discover new applications. A growing field of research is focussing on changing toxin

specificity, toxicity and duration of action by modifying the BoNT protein in order to expand their current range of use and create optimised therapeutic agents.[149-155] The modulation of BoNT activity and specificity can be achieved by mutating protein domains at strategic positions. These recombinant proteins are usually expressed in *Escherichia coli* because of their easy genetic manipulation, short culturing times and the availability of optimised expressions strains, capable of producing a wide variety of different types of proteins.[156-157] Nonetheless, obtaining high yields of active BoNT derivatives in *E. coli* can be challenging due to solubility issues, protein associated toxicity, the formation of inclusion bodies and incorrect folding. Moreover, they do not express the yet to be identified proteases that 'activate' the toxin. Ideally, recombinant BoNTs would be expressed in their native host. Unfortunately, the development of non-sporulating *C. botulinum* expression host strains has been hindered by the limited number of tools available to manipulate their DNA.[157] It is however anticipated that this will change rapidly over the next few years with the development of efficient CRISPR/Cas9-based editing technologies for clostridial species (see section 1.5.3).[158]

Product quality and safety are paramount to pharmaceutical BoNT manufacturers, and therefore drug potency and stability is monitored continuously. For the past decades the mouse bioassay has been considered the best assay to detect BoNTs, determine their potency and assess their stability in both medical settings and in the cosmetics industry.[159] Despite the costs and ethical concerns of using live animals, it is still the most widely used detection method, and for good reasons; it is sensitive, can be quantitative, reliable (detecting all known types of BoNT) and detects only biologically active BoNTs. However, driven by the ethical concerns, there has been a

large push towards the development and improvement of alternative assays, such as the use of cultured neurons, to reduce and ultimately replace the animal-based assay.[160]

It should be noted that the current neuronal cell-based assays do not replicate the complexity of the *in vivo* intoxication pathway; there is no dilution by circulating fluids and various components of the neuromuscular junction are missing, such as muscle and gland cells.[106] Yet, they can detect only biologically active botulinum neurotoxin and are easily standardised. Although there can be striking differences between BoNT toxicity *in vivo* and cultured neurons, neuronal cell-based assays could potentially replace the mouse bioassay for potency and stability determinations performed on pharmaceutical BoNT products.[159, 161-163] In 2011, one of the major BoNT manufacturers (Allergan) obtained FDA approval to substitute the animal-based assay with a cell-based potency assay for quality control testing and according to their website this is now the only assay they use to perform BOTOX® drug product release and stability testing.[106]

1.4.3 *C. botulinum* in academia

C. botulinum research at universities has been quite diverse, albeit limited due to the small number of laboratories with the appropriate facilities. Hence, a large portion of the studies published during the past decade are primarily based on bioinformatics.[5, 7, 27, 33, 42, 164] It has been one of the forces driving the large on-going genome sequencing effort; as mentioned previously, *C. botulinum* is the second most sequenced *Clostridium* spp. to date. Another major factor contributing to this vastly growing dataset, besides the reduced costs of sequencing, is the species' diversity. The fascinating question 'why such a diverse group of bacteria evolved to produce this incredibly potent toxin' is still unanswered.[6] Possibly, evolution theories derived from

genomics based studies combined with experimental data will one day provide an answer to this intriguing question.[44-45, 165]

A majority of the current 'wet-lab' research focusses on the toxin rather than the organism and can be roughly divided into three groups: (i) the study of recombinant protein (discussed in section 1.4.2); (ii) the functional analysis of BoNTs and their non-toxic accessory proteins; and (iii) the characterisation of BoNT neutralising antibodies and vaccines.[108, 110, 149, 157, 166] The handful of studies published characterising the organism in a 'wet-lab' setting pales in comparison to the number of toxin-oriented studies published. Hence, many fundamental questions are yet unanswered, including how and what regulates toxin production, which native protease(s) cleave the toxin and via what route is the toxin secreted.[167-171] Most of the organism based research has focussed on cell processes relevant to the food industry, such as sporulation and resilience under specific growth conditions.[172-176] Progress however has been frustrated by the lack of efficient tools to manipulate its DNA. Fortunately, the highly anticipated tools have recently become available (discussed in more detail in section 1.5) and should aid future research projects which aim to elucidated more fundamental cell processes that define this unique group of bacteria and its toxin.[158, 177]

1.5 Methods to genetically engineer and control gene expression in clostridia

Synthetic biology is a relatively new field of research that aims to harness the power of living things for the benefit of the planet by engineering cell processes. Central to synthetic biology is the development of tools to modify the genetic code and alter protein expression levels in order to create cell systems with useful applications.[178]

Such tools first appeared for the model organisms *E. coli* and *Bacillus subtilis*, however for non-model organisms the available methods were fairly limited until the discovery of the revolutionising CRISPR/Cas9 (Clustered Regularly Interspaced Palindromic Repeats/ CRISPR-associated) based technologies, which will be discussed in more detail in section 1.5.3. The more fundamental molecular research fields have significantly benefitted from the synthetic biology advancements. For example, the availability of tools to manipulate bacterial DNA *in vivo* and control protein expression has accelerated the study of gene and protein function.

1.5.1 ClosTron-mediated mutagenesis

Engineering *Clostridium* spp., including *C. botulinum*, has been notoriously difficult for many years and only in 2007 the first easy to use method for knocking out genes in clostridia was published: the ClosTron.[126, 179-180] In current literature, it is still the most widely used tool to study gene functionality in *C. botulinum*.[170-171, 176, 181-185] The basic principle behind ClosTron mediated mutagenesis is the insertion of a DNA fragment into a specific DNA target to disrupt the coding sequence, thereby preventing the translation of a functional protein. It is based on a mobile group II intron, whose insertion is mediated by base-pairing between the excised intron RNA and the target site DNA. The intron excision is facilitated by an Intron Encoded Protein (IEP) which also possesses reverse transcriptase activity. This IEP is provided on the ClosTron expression plasmid and can be removed from the strain by plasmid loss to ensure that the insertional mutant is stable. To facilitate the identification and isolation of successful insertional mutants, the group II intron contains a Retrotransposition-Activated Marker (RAM), which comprises an erythromycin resistance gene (*ermB*) interrupted by a self-splicing group I intron. During transposition the group I intron loops out of the *ermB* gene, 'activating' erythromycin resistance upon successful

genomic integration of the group II intron. Hence, successful mutants can be selected based on the acquisition of erythromycin resistance.

The specificity of the group II intron is easily re-programmed by changing the DNA sequence of the intron that recognizes the target DNA. Despite the intron design being restricted by certain 'rules', using a computer algorithm it is possible to design an intron for almost every gene.[186] However, integration efficiency and target specificity vary depending on the target. Additionally, the intron occasionally inserts into a secondary location rather than or alongside the intended target, and therefore, single copy insertion should always be confirmed via western blotting or by sequencing the entire genome. Moreover, the ClosTron insertion can affect the expression of downstream genes or operons, which is referred to as polar effects.

1.5.2 Allele-coupled exchange

Shortly after the ClosTron insertional mutagenesis tool was published, a second, more versatile, homologous recombination-based DNA manipulation method was developed.[187-189] It is a multistep strategy which relies on counter selection markers (i.e., *pyrE* or *codA*) and splitting homologous recombination events (Figure 1.2).[190] First a counter-selectable gene is inactivated via two recombination events, referred to as allele-coupled exchange (ACE). The order of the two events is directed by homology arms of asymmetric length, with recombination occurring first at the longer homology arm. This results in the integration of the entire plasmid, which carries an antibiotic resistance gene, and is referred to as a single-crossover mutant. The single-crossover mutants will grow faster on media supplemented with antibiotics compared to cells where the plasmid is still an independent entity that is readily lost during the cell division process. In the second step, the plasmid is excised, mediated by the

smaller homology arm, and results in the truncation of the selection marker, yielding a selectable phenotype (double-crossover mutant). For example, truncation of the *pyrE* gene yields cells that can no longer metabolise 5-fluoroorotic acid (5-FOA) into the toxic metabolite 5-fluorouracil. Hence, acquisition of 5-FOA 'resistance' can be used to phenotypically select successful mutants. Simultaneously, a desired insert can be integrated into the genome when placed between the two asymmetric homology arms.

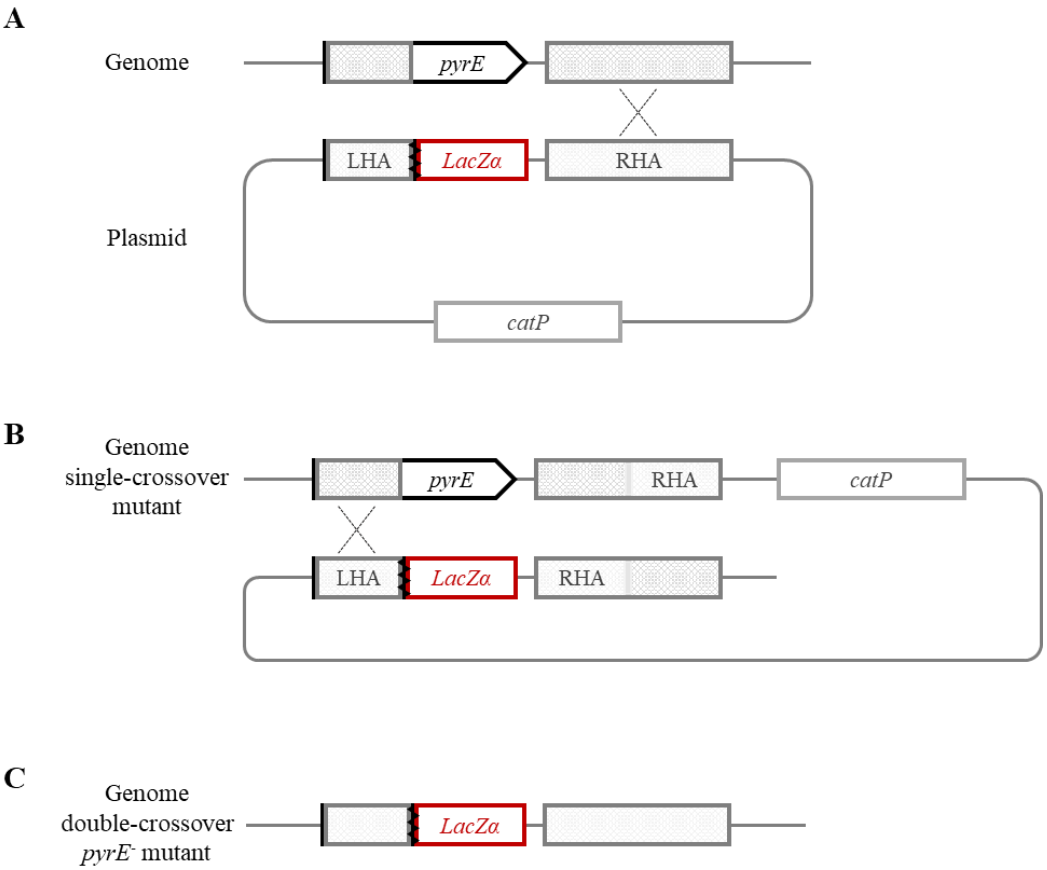


Figure 1.2 Schematic overview of allele-coupled exchange (ACE). (**A**) The insert of interest, here *lacZα*, if applicable, is placed between the two 'arms' homologous to the genomic DNA. The first recombination event is facilitated by the long right homology arm (RHA) downstream of the *pyrE* gene and results in the integration of the entire plasmid into the chromosome (**B**). If the second crossover event occurs at the short left homology arm (LHA) it leads to the truncation of the *pyrE* gene, while the plasmid is excised, and yields cells that are 5-FOA resistant (**C**). After a successful recombination event the sequence of interest is stably integrated into the chromosome.

Next, genetic material can be deleted or inserted at any desired target site via classical allelic exchange by placing a non-native, functional copy of the inactivated selection marker and an antibiotic resistance cassette on an unstable plasmid that is easily lost by the cell.[187] Subsequently, double-crossover mutants, which have lost the plasmid, can be selected based on the acquisition of the selectable phenotype due to loss of the counter-selectable gene. Although this method allows precise genome modifications, it is reliant on the availability of a counter selection marker, which is not equally effective in all *Clostridium* spp., and involves several time-consuming steps (personal communications; N. Minton).

1.5.3 CRISPR/Cas9-mediated genome editing

In bacteria CRISPRs and their associated Cas proteins provide adaptive immunity against invading exogenous DNA or RNA; they recognise and degrade these molecules. Over the past decade this system has been repurposed by scientists for genome editing applications, regulating gene expression and much more, with unprecedented success in both eukaryotes and prokaryotes.[191-194] In particular, the type II CRISPR/Cas9 system from *Streptococcus pyogenes* has found widespread use due to its simplicity.[195-196] It only requires one RNA-guided endonuclease (Cas9), which generates double-stranded breaks in the desired DNA targets via base-pairing with a corresponding guide RNA.[197] In most bacteria, if no repair template is provided, double-stranded DNA breaks are lethal, as they lack the capacity to repair such breaks via Non-Homologous End Joining (NHEJ).[196] Currently, few bacterial strains are known that can perform NHEJ to repair double-stranded breaks and can thus, theoretically, survive Cas9 DNA cleavage events.[198] However, if in bacterial strains deficient in NHEJ machinery a repair/editing template, homologous to sequences flanking the target site, is provided, cells can eliminate the Cas9 target site by

exchanging the site for the 'new' DNA sequence, thereby avoiding cleavage.[194, 199] Thus only cells, who have successfully swapped the target site for donor DNA via homologous recombination will survive in the presence of Cas9. Basically, Cas9 is just acting as a negative selection marker, enabling the identification of mutants, without requiring antibiotic resistance markers. Importantly, concomitantly with Cas9 target site removal, entire genes or operons can be removed, as well as cargo inserted.[200]

As mentioned, Cas9 requires a dual guide, consisting of a precursor CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA), for target recognition. To simplify the overall system the two RNAs have been 'fused' into a synthetic single-guide RNA (sgRNA).[201] To direct Cas9 to the desired locus, a specific 20 bp sequence within the sgRNA should be complementary to the genomic site of interest.[202] This region can be modified to match almost any genomic location of interest, providing that the sequence is upstream of the short sequence motif 5'-NGG-3', known as the protospacer adjacent motif (PAM).[203] If the target sequence is not adjacent to a PAM motif, Cas9 will not cleave the DNA sequence.

Initially, CRISPR/Cas based technologies were mainly employed in eukaryotes, while use in prokaryotes was hampered due to its inherent lethality, resulting in low transformation efficiencies for *cas9*-carrying plasmids.[195, 204-205] This is especially problematic when attempting to employ the system to engineer clostridial genomes, as *Clostridium* spp. usually already suffer from poor transformation efficiencies in the absence of a lethal cargo.[206-207] Nevertheless, several CRISPR/Cas9-based systems have been successfully used in a few *Clostridium* spp. that have reasonable DNA transfer efficiencies, albeit with mixed effectiveness.[177, 199, 204, 208-213] To overcome the above limitations, a very efficient inducible CRISPR/Cas9 system, named RiboCas,

was developed.[158] Controlling the expression of Cas9 provides time for homologous recombination to occur before Cas9 is expressed, thus enabling uncompromised plasmid transfer, while there is an adequate level of nuclease expression after induction, allowing strong selection. It seems however, that although there are ongoing effort to optimise the *S. pyogenes* CRISPR/Cas9 system for other bacterial species, interest has been shifting towards reprogramming endogenous CRISPR/Cas systems to edit a strain's genome.[199, 214-216]

1.5.4 Inducible systems to control gene expression

Several inducible systems have been developed for clostridia in order to control gene expression, including the well-known isopropyl-β-d-thiogalactoside (IPTG)/lactose-inducible system (IPTG system) and anhydrotetracycline-inducible system (tet-system).[217-218] In both the IPTG and tet-inducible system, a repressor binds to an operator sequence (lacOI/tetO) which is placed in between or next to the -10 and -35 boxes of a promoter, thereby inhibiting the binding of RNA polymerases to the promoter sequence and thus prevents the start of transcription. Addition of the inducer (Lactose/anhydrotetracycline) will release the repressor from the DNA and consequently expression is activated.

In addition to these 'famous' inducible systems, a lactose-inducible system, discovered in *Clostridium perfringens*, has been tested and characterised.[219-220] The system consists of a transcriptional regulator, *bgaR* and the divergent promoter P_{bgaL}. Although the exact mechanism of the system is unknown, experimental results suggest that *bgaR* acts as an activator that binds to the promoter region of P_{bgaL} in the presence of lactose.[221] In general, high expression levels can be obtained after induction,

however background expression levels are relatively high compared to, for example, the tet-system.

In search for systems which tightly control expression in order to express very toxic proteins, such as cas9, the interest in riboswitches has recently increased.[222] Riboswitches are RNA aptamers formed upstream of the 5' untranslated region (UTR) of mRNA molecules which undergo conformational changes upon interaction with a ligand, like theophylline, neomycin and tetracycline.[223] They can function as a switch to control translation when the ribosome binding site (RBS) upstream of a gene is included in the aptamer structure. In the absence of the ligand the RBS is included in the aptamer structure, preventing the ribosome to bind and initiate translation. In the presence of the ligand the aptamer structure will change, exposing the RBS and thus allowing expression of the protein. The theophylline responsive riboswitches are the best available system published so far and have been successfully implemented in *C. sporogenes* recently.[221] They tightly controlled gene expression, albeit the level of induction is limited. Coupling the riboswitches to the lactose-inducible system of *C. perfringens* (named RiboLac) significantly increased the systems dynamic range. Depending on the application different inducible systems might be appropriate.

1.6 Project aim and thesis overview

The overall aim of this thesis was to develop a strategy, applicable to all the members of the highly diverse *C. botulinum* family, to efficiently generate genetically stable non-toxigenic strains, eliminating biosafety requirements and reducing associated laboratory costs. The road towards achieving this goal is presented herein:

Chapter 3 explores various tactics for outwitting host restriction barriers in order to establish reliable and efficient DNA transfer in previously genetically inaccessible or poorly accessible strains. Additionally, the effectiveness of a comprehensive roadmap for improving DNA transfer efficiency in previously 'untouched' strains is exemplified using type A strains.

Chapter 4 embarks on a campaign to elucidate and circumvent the genomic instability of model strain ATCC 3502 which has to date hampered *C. botulinum* research at the Synthetic Biology Research Centre (SBRC). In the process, widespread occurrence of mutator genotypes amongst type A strains was discovered and evolutionary concepts are being discussed in an attempt to explicate this interesting phenomenon.

Chapter 5 describes the characterisation of various *C. botulinum* type A strains and engineered non-toxigenic versions thereof. Concomitantly, a highly efficient one-step CRSIPR/Cas9-based DNA editing system was developed and tested, thus improving upon the previously available engineering toolkit for *C. botulinum*.

Chapter 2 provides detailed information regarding the materials and methods used that enabled the research presented in this thesis.

Chapter 6 outlines the strategy developed to create *C. botulinum* safe strains and discusses various exciting future research opportunities.

Short statement concerning strain nomenclature. The work presented in this thesis revolves around five different type A *C. botulinum* strains referred to as: ATCC 3502, ATCC 19397 (NCTC 7272), 62A (NCTC 3805), Loch Maree (NCTC 2012) and NCTC 2916. The rationale behind selecting these specific strains can be found in Chapter 5. For more general information regarding the strains please see Table 5.1 and Appendix I. Specifically, ATCC 3502 refers to the mutator/non-mutator mixture, ATCC 3502^{+} refers to the non-mutator strain and ATCC 3502^{-} refers to its mutator variant. Chapter 4 discusses the ATCC 3502 variants in detail. It should be noted that various stocks of these strains can be found in laboratories across the world, occasionally under a different name, and their origin is not always clear.

CHAPTER 2

Materials and Methods

2.1 Bacterial strains

All bacterial strains used and generated in this study are listed in Appendix I.

2.2 Growth media, buffers and supplements

All media and buffers were prepared using dH_2O and autoclaved for 20 minutes at 121°C. To prepare solid media 1.5% (w/v) No. 1 Bacteriological Agar (Oxoid) was added to the media prior to autoclaving. All media and buffers used in this study are listed in Table 2.1.

Supplements were autoclaved or filter sterilised using a 0.2 µm pore size filter (Minisart®) before use and added to the autoclaved media. All supplements used in this study are listed in Table 2.2.

Table 2.1 Growth media and buffers used in this study

Medium/Buffer	Composition	Amount	pH
Luria-Bertani (LB)	Tryptone	10 g/L	7.2
	Yeast Extract	5 g/L	
	NaCl	10 g/L	
Defined media	Glucose 50% (w/v)	8 ml/L	7.2
	9M Salt solution	100 ml/L	
	Biotin	1 mg/L	
	$CaCl_2$ (1 M)*	0.3 ml/L	
	$MgSO_4$ (1 M)*	1 ml/L	
	Thiamine*	1 mg/L	
	Glycine	100 mg/L	
	Alanine	100 mg/L	
	Serine	100 mg/L	
	Threonine	100 mg/L	
	Cysteine	100 mg/L	
	Valine	100 mg/L	
	Isoleucine	100 mg/L	
	Phenylalanine	100 mg/L	
	Aspartic acid	100 mg/L	
	Glutamic acid	100 mg/L	
	Asparagine	100 mg/L	
	Histidine	100 mg/L	
	Lysine	100 mg/L	
	Arginine	100 mg/L	
	Leucine	100 mg/L	

Defined media 1.1	Defined media +		
	Methionine	100 mg/L	
	Tryptophan	100 mg/L	
Defined media 1.2	Defined media +		
	Proline	100 mg/L	
Tryptone-Yeast Extract-Sodium Thioglycolate medium (TYG)	Tryptone	30 g/L	7.2
	Yeast Extract	20 g/L	
	Sodium thioglycolate	1 g/L	
Campdem Sporulating Agar (CSA)	Tryptone	5 g/L	
	Bacterial peptone	5 g/L	
	Lab Lemco meat extract	1 g/L	
	Yeast extract	2 g/L	
	$CaCl_2$	56 mg/L	
	$MnSO_4 \cdot H_2O$	62 mg/L	
	Glucose	1 g/L	
	$CaCl_2$ (1 M)*	1 ml/L	
	$MgSO_4$ (1 M)*	1 m/L	
Phosphate-Buffered Saline (PBS)	NaCl	8 g/L	7.4
	KCl	0.2 g/L	
	Na_2HPO_4	1.44 g/L	
	KH_2PO_4	0.24 g/L	
9M Salt solution	$Na_2HPO_4 \cdot 7H_2O$	70 g/L	
	KH_2PO_4	30 g/L	
	NaCl	5 g/L	
	NH_4Cl	10 g/L	
MOPS solution	MOPS	209 mg/L	
	Glycerol	100 ml/L	

* Filter sterilised and added to the media after the media was autoclaved.

Table 2.2 Supplements used in this study

Supplement	Stock concentration	Solvent	Working concentration: *E. coli*	Working concentration: *Clostridium*
D-Cycloserine	50 mg/ml	dH_2O	-	250 μg/ml
Chloramphenicol (Cm)	25 mg/ml	100% ethanol	12.5 μg/ml (broth) 25 μg/ml (agar)	-
Thiamphenicol (Tm)	15 mg/ml	50% ethanol	-	15 μg/ml
Kanamycin (Kan)	50 mg/ml	dH_2O	50 μg/ml	-
Apramycin (Apr)	50 mg/ml	dH_2O	100 μg/ml	-
Erythromycin (Erm)	50 mg/ml	100% ethanol	300 μg/ml (broth) 500 μg/ml (agar)	30 μg/ml (broth) 50 μg/ml (agar)
Rifampicin (Rif)	50 mg/ml	DMSO	-	25 μg/ml
Glucose	50% (w/v)	dH_2O	-	0.05% (w/v)
Lactose	100 mM	dH_2O	-	10 mM
5-Fluorootic acid (5-FOA)	100 mg/ml	DMSO	-	5-5.5 mM
L-arabinose	1M	dH_2O	10 mM	-
Theophylline	266 mM	DMSO	-	2 mM (broth) 5 mM (agar)

2.3 Growth and storage conditions

E. coli strains were cultured aerobically at 30°C or 37°C (depending on cargo toxicity) on LB agar plates or in LB liquid media while shaking at 200 rpm. Supplements were added to the media as required (i.e., plasmid selection or induction). *E. coli* strains were stored at -80°C using the Microbank™ Long Term Bacterial and Fungal Storage System (Pro-lab Diagnostics). Cells for storage were harvested from LB agar plates and subsequently re-suspended in Microbank™ vials. Strains were stored in triplicates (green, yellow and red stock).

C. botulinum strains were cultured anaerobically at 37°C on TYG agar plates or in TYG liquid media without shaking. Supplements were added to the media as required. An anaerobic environment was created within a MG1000 Mark II Anaerobic Work Station (Don Whitley Scientific Ltd) with an atmosphere consisting of 80% N_2, 10% CO_2 and 10% H_2. Before use, all media was pre-reduced for approximately 4 hours in the anaerobic work station. *C. botulinum* strains were stored at -80°C in TYG supplemented with 20% v/v glycerol. Cells for storage were harvested from TYG agar plates and subsequently re-suspended in TYG-20% glycerol or x µl of glycerol was added to a liquid overnight culture to get a final concentration of 20%. Strains were stored in triplicates (green, blue and red stock).

2.4 Preparation of electrocompetent *E. coli* cells

To prepare competent *E. coli* cells 5 ml of LB broth was inoculated with cells from a -80°C stock and incubated overnight at 37°C while shaking (200 rpm). The following day, 1 ml of overnight cell culture was used to inoculate 200 ml of sterile, pre-warmed LB broth in a 500 ml conical flask and incubated at 37°C while shaking (200 rpm) until an OD_{600} of 0.5-0.7 was reached. Subsequently, cells were harvested by centrifugation at 6000 ×*g* for 10 minutes at 4°C in 40 ml Falcon tubes. Next, the cell pellet were washed three times by gently re-suspending the cells in 20 ml sterile, ice-cold distilled H_2O (dH_2O) followed by centrifugation at the same conditions. After each wash step two re-suspended cell pellets were combined into one Falcon tube. After washing the cells thoroughly but gently, the final cell pellet was re-suspended in 1 ml of 1 mM MOPS solution with 10% (v/v) glycerol, split into 50 µl aliquots and stored at -80°C until future use.

2.5 Transformation of *E. coli* cells by electroporation

For each *E. coli* transformation by electroporation a 50 µl aliquot of electrocompetent cells was taken from the -80°C freezer and thawed on ice for 20 minutes. Next, purified plasmid DNA (2 µl or 200-500 ng) or 10-20 µl of dialysed ligation mixture (0.025-µm MF-Millipore™ dialysis discs over dH_2O for ± 20 minutes) was added to the competent cells before transferring the cells to a chilled electroporation cuvette (Bio-Rad 0.2 cm electrode). Electroporation was performed using the Bio-Rad MicroPulser™, as per manufacturer's instructions. The applied conditions for the pulse were set to 2250 V (22.5 kV/cm, 200 Ω, 25 µF). Immediately after pulsation the cells were re-suspended in 400 µl SOC medium (Invitrogen™), transferred to a clean 1.5 ml tube and incubated for 1 hour at 37°C (or for 2 hours at 30°C) while shaking (200 rpm). Cell were subsequently plated onto selective LB agar plates (20 µl, 100 µl and rest) and incubated as required until growth was visible.

2.6 Conjugation of R702 from *E. coli* donor into *E. coli* recipient

Conjugative DNA transfer was used to move the engineered R702 plasmids from InterStellar, via LE392, to other *E. coli* strains (i.e., HB101 and CA434). Selection was based on amino acids deficiencies; InterStellar, HB101 and CA434 require proline supplementation for growth, while for LE392 methionine and tryptophan supplementation is essential for growth. Overnight cultures were diluted to a starting OD_{595} of 0.02 in fresh LB broth with the necessary supplements and the OD_{595} was monitored for ± 4 hours at 37°C while shaking. Next, the cells were mixed in a 1:1 OD_{595} ratio in triplicates and incubated for 2 hours at 30°C without shaking. The cells were harvested by centrifugation for 1 minute at 5000 ×*g* and washed twice by re-suspending the pellets in 1 ml PBS followed by centrifugation. The final pellet was re-

suspended in 200 µl PBS and plated on defined media 1.1 agar ($Met^+/Trp^+/Pro^-$) supplemented with Km and Apr. Successful transfer of the recombinant R702 plasmid was confirmed by PCR in three biological replicates from three independent conjugations per recipient strain using primers RE_F and RE_R, annealing to the recombineering cassette, and re-streaked to purity on defined media 1.1 agar. Similarly, plasmid R702 was transferred from LE392 to HB101 and CA434, however, after washing the conjugal mixture, cells were serially diluted in PBS as required and subsequently plated on defined media 1.2 agar ($Met^-/Trp^-/Pro^+$) supplemented with Km and Apr. Successful plasmid transfer was confirmed by PCR in three biological replicates from three independent conjugations per recipient strain using primers RE_F and RE_F and restreaked to purity on defined media 1.2 agar.

2.7 Conjugation of plasmid DNA from *E. coli* donor into *C. botulinum*

Conjugative transfer of plasmid DNA was used to introduce new DNA into *C. botulinum*. For each independent conjugation, 5 ml of LB broth with appropriate supplements (e.g., antibiotics and/or inducer) was inoculated with *E. coli* donor cells (harbouring a shuttle plasmid) from a -80°C stock and incubated overnight at 37°C while shaking (200 rpm). Similarly, 5 ml of TYG broth was inoculated with *C. botulinum* recipient cells and incubated overnight in the anaerobic work station. To harvest the donor cells, 1 ml of overnight culture was centrifuged for 1 minute at 5000 ×*g*. Subsequently, the cell pellet was washed carefully with 1 ml of PBS, harvested as before (1 minute at 5000 ×*g*) and transferred to the anaerobic work station. The resulting cell pellet was gently re-suspended in 200 µl of the *C. botulinum* recipient overnight culture and the mixture was carefully spotted (5 spots) onto a TYG-0.5% glucose plate. The plate was incubated for at least 6 hours (and at least 7 hours for

ATCC 19397) in the anaerobic work station. After incubation cells were harvested from the mating plate by re-suspending the cells in 500 µl of reduced PBS. Finally, the conjugal mixture was serial diluted in PBS (up to 10^{-2}) and 50 µl of the different dilutions was plated onto selective TYG/D-cycloserine plates to counter-select for *E. coli*. The plates were incubated at 37°C in the anaerobic work station until growth was visible (24-48 hours). Presence of the plasmid of interest in the recipient cells was confirmed by PCR (see Appendix IV).

2.8 Plasmid loss

To achieve plasmid loss in successful *C. botulinum* mutants that no longer required the conjugated plasmid, cells were cultured in non-selective TYG broth for 24-48 hours. Subsequently the cell cultures were serially diluted in PBS as required and the appropriate dilutions plated on TYG agar to obtain single colonies. After 24-48 hours of incubation, single colonies were replica-plated on selective and non-selective TYG agar. Colonies that exclusively grew on non-selective TYG agar were assumed plasmid free and plasmid loss was confirmed via PCR with plasmid-specific primers (see Appendix IV).

2.9 Phenotypic analysis

2.9.1 Optical density measurements

The optical density (OD_{595}) was measured at 595 nm using a Bio-Rad iMark microplate reader and flat bottomed 96-well plates (Costar, Corning®). Samples were diluted with TYG broth as required. Measurements were taken against a blank of sterile medium.

2.9.2 Growth curves

To establish the growth profiles of *C. botulinum* strains, overnight cultures were diluted to a stating OD_{595} of 0.01 in fresh TYG broth with the necessary supplements. Samples were taken every 1-2 hours and growth was monitored for 24 hours.

2.9.3 Conjugation efficiency assay

Conjugation efficiency assays were performed in a similar fashion as standard conjugations described in section 2.7. Conjugations were performed in triplicates, using three independent *E. coli* and *C. botulinum* overnight cultures. For each replicate an additional 1 ml of the *E. coli* donor overnight culture was washed with BPS as described in section 2.7., and the resulting cell pellet was re-suspended in 1 ml PBS. Subsequently, the sample was serially diluted in PBS as required and 20-µl aliquots from the appropriate dilutions were spotted in triplicates on LB agar with the necessary supplements. Furthermore, to determine the total *C. botulinum* CFU present in the conjugal mixture, the cells harvested from the mating plates were diluted as required and subsequently 20-µl aliquots from the appropriate dilutions were spotted in triplicates on both selective and non-selective TYG/D-cycloserine agar. To ensure that the total CFU at the end of each conjugation was similar between strains, mating plates with ATCC 3502 were incubated for ±6 hours; mating plates with 62A, NCTC 2916 and Loch Maree were incubated for ±7 hours; and mating plates with ATCC 19397 were incubated for ±8 hours. Both *E. coli* and *C. botulinum* colonies were enumerated after 24-48 hours of incubation. The conjugation efficiency was calculated as the total number of transconjugants per *E. coli* donor.

2.9.4 Sporulation assay

To quantify heat resistant spores, overnight cultures were diluted to a stating OD_{595} of 0.01 in fresh TYG broth with the necessary supplements and incubated for 120 hours. Samples were taken at 0, 24, 48, 72, 96 and 120 hours and split into two. One of the two samples was centrifuged for 30 seconds at 5000 ×*g* and subsequently heat treated for 15 minutes at 80°C. The heat-treated samples were serially diluted in reduced PBS as required and 20-µl aliquots from the appropriate dilutions were spotted in triplicates on TYG agar. To determine the total viable cell count the non-heat treated samples were serially diluted and spotted on agar in a similar fashion as the heat-treated samples. Colonies were counted after incubating the plates for 24-48 hours in the anaerobic work station (spots with CFUs between $5 \leq x \geq 50$ were enumerated).

2.9.5 Spore survival curves

To establish spore heat resistance, 100 µl overnight culture (TYG broth) was plated on CSA agar and subsequently incubated for 2 weeks in the anaerobic work station at 37°C.[224] Spores were harvested by adding 1-2 ml cold dH_2O to the plates, scraping the plate with a plate spreader, transferring the spore suspension to a tube (this was repeated twice) and centrifugation at 6000 ×*g* for 2 minutes. Next, the spores were washed three times by re-suspending the spores in 1 ml sterile, cold dH_2O followed by centrifugation at the same conditions. Finally, the spores were re-suspended in 1 ml dH_2O. To confirm that the spore concentration was sufficiently high to determine the spores heat resistance (10^6), a standard sporulation assay was performed, as described in section 2.9.4. The remainder of the spore suspension was stored at 4°C for ± 1 week.

The spore suspension was split into 50 µl aliquots in 0.5 ml PCR tubes and subjected to heat treatment at 99°C in a PCR machine. Samples were removed from the PCR machine after 0, 1, 2, 3, 4, 5, 10, 13, 15, 30, 45, 60 or 90 minutes of heat treatment as required. The heat-treated samples were serially diluted in reduced PBS as required and 20-µl aliquots from the appropriate dilutions were spotted in triplicates on TYG agar. Colonies were counted after incubating the plates for 24-48 hours in the anaerobic work station (spots with CFUs between $5 \leq x \geq 50$ were enumerated).

The *D*-values were calculated by dividing -1 by the slope of N/N_0. Where N stands for the number of viable spores at a specific time point and N_0 stands for the number of viable spores before any heat treatment was applied.

2.9.6 Fluctuation assay

To determine the mutation frequency, three independent overnight cultures were serially diluted in PBS as required and 100 µl from the appropriate dilution was plated onto TYG agar. After 24 hours of incubation random single colonies were harvested, re-suspended in TYG-20% glycerol and stored at -80°C. A total of 10 colonies were stored; three colonies from replicate culture 1, three colonies from replicate culture 2 and four colonies from replicate culture 3.

One overnight culture per single colony was prepared in TYG broth. The following day the cultures were serially diluted up to 10^{-7} in PBS and 20-µl aliquots from the appropriate dilutions were spotted in triplicates on TYG agar to confirm that the assay starting inoculum was ± 100 cells. Fresh 3 ml TYG broth was inoculated with 30 µl of the 10^{-5} overnight culture dilution. The 'assay' cultures were split into three 1 ml cultures and incubated in the anaerobic work station at 37°C for 24 hours. The cultures were subsequently serially diluted up to 10^{-7} in PBS and 20-µl aliquots from the

appropriate dilutions were spotted in triplicates on TYG agar without antibiotics to determine the total number of viable cells. Colonies were enumerated after 24 hours of incubation. The remainder of the 1 ml cultures was plated on TYG/Rif agar (4× 200 µl and 1× 50 µl) and colonies were enumerated after 48 hours of incubation. The mutation frequency was calculated by dividing the total number of Rif resistant colonies by the total CFU.

2.9.7 Competitive fitness assay

To assess the relative fitness of the ATCC 3502^{+} and ATCC 3502^{-} strain the mixed Unilever meat stock (*C. botulinum* ATCC $3502^{+/-}$ Unilever, 2011; by C. Meaney) was serially diluted in PBS as required in triplicate and 100-µl aliquots from the appropriate dilution were used to inoculate fresh TYG broth; two cultures per independent serial dilution series. The six fresh cultures (T0) were mixed thoroughly and 20-µl aliquots were serially diluted in PBS as required. Subsequently, 100-µl aliquots from the appropriate dilution were plated on TYG agar. The following day 32 colonies per replica were PCR screened using primers mutL_F_scr and mutL_R_scr to determine the starting ratio of non-mutator *vs* mutator.

After the first 24 hours of incubation three cultures from independent serial dilution series were used to inoculate fresh TYG broth (1:100; S1). The cultures were sub-cultured for 6 days in a row (S6) and the non-mutator:mutator ratio was determined for each sub-culture after 24 hours of growth, as described previously. Cells in the three static cultures were cultured for 168 hours in TYG broth and the non-mutator:mutator ratio was determined after 24, 48, 72 and 168 hours of incubation, as described previously.

2.9.8 SDS-PAGE and Western blotting

2.9.8.1 Preparation of cell-free culture supernatants and protein precipitation

C. botulinum overnight cultures were diluted to a starting OD_{595} of 0.01 in fresh TYG broth and incubated for 24-48 hours. Supernatants were collected by centrifugation at full speed for 10 minutes, filtered through a through a 0.22-µm membrane (Minisart®) filter and subsequently transferred to a clean tube. The supernatants were stored at -80°C until further processing.

To precipitate protein from the cell-free supernatants sodium deoxycholate (DOC) (0.03% final w/v) was added to the samples and incubated for 10 minutes at room temperature. Next, trichloroacetic acid (TCA) was added to a final concentration of 10% w/v and the samples were stored at -20°C overnight. The following day, the samples were thawed on ice, subsequently centrifuged at full speed for 20 minutes and the supernatants discarded. The pellets were washed by re-suspending them in 1 ml ice-cold ethanol and acetone consecutively. In between wash steps, the samples were centrifuged at full speed for 10 minutes and the ethanol/acetone discarded. After the final wash step, the pellets were air-dried for 1 hour in the microbiology safety cabinet.

2.9.8.2 SDS-PAGE

The protein content of the supernatants was analysed by SDS-PAGE using the XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Fisher Scientific) and 1.0 mm precast Bolt™ 4-12% Bis-Tris Protein Gels (Invitrogen). The protein pellets (from section 2.9.8.1) were combined in a 3:1 volume ratio with 4× NuPAGE® LDS Sample Buffer (Novex™), premixed with 400 mM dithiothreitol (DTT) and 0.1 M NaOH (resulting in 10× concentrated supernatants). The proteins were denatured by incubating the samples for 5 minutes at 100°C. Next, the samples were briefly vortexed and loaded in equal volumes onto the gels. The Precision Plus Protein™

Kaleidoscope™ Prestained Protein Standard (Bio-Rad) was used as a reference for the molecular weight. Electrophoresis was carried out at 180 V for 60 minutes in 1× NuPAGE® MES SDS running buffer (Novex™) and subsequently the gels were rinsed with dH_2O.

2.9.8.3 Western blotting

For protein transfer the BioRad Trans-Blot Turbo Transfer System and Trans-Blot Turbo Transfer Packs (PVDF membrane; cat # 170-4156) were used according to the manufacturer's instructions. To block the membranes, they were briefly washed in PBS-T buffer (PBS-0.05% Tween 20), moved to a plastic container with 30 ml PBS-T-M buffer (5% w/v Marvel milk powder in PBS-T) and incubated for 1 hour at room temperature while gently agitating. Next, the PBS-T-M buffer was removed and primary antibody (polyclonal sheep IgG BoNT-A primary antibody; R&D systems; AF4489), dissolved in 30 mL cold PBS-T-M buffer, was added to the membrane. The membranes were incubated overnight at 4°C while gently agitating. Subsequently, the membranes were washed in PBS-T buffer 6× for 10 minutes while agitating at room temperature. The secondary antibody (sheep IgG HRP-conjugated secondary antibody; R&D systems; HAF016), dissolved in 30 mL cold PBS-T-M buffer, was added to the membranes and incubated at room temperature for 1 hour while gently agitating. The labelled membranes were washed as described previously and subsequently 1 ml 3,3′,5,5′-Tetramethylbenzidine (TMB; Sigma-Aldrich) was added. The membranes were incubated for 1-2 minutes (until bands appeared) at room temperate while slowly moving the container.

2.10 General molecular biology techniques

2.10.1 Plasmid DNA extraction

Plasmid DNA was extracted from 1-2 ml overnight *E. coli* cultures using the Monarch® Plasmid Miniprep Kit (New England Biolabs; NEB) according to the manufacturer's instructions, eluted in dH_2O and stored at -20°C.

2.10.2 Genomic DNA extraction

Genomic DNA from *C. botulinum* was extracted according to S. Cartman's 'Phenol:Chloroform:Isoamyl Alcohol (25:24:1) DNA extraction' protocol.[225] 1 ml of overnight cell cultures were centrifuged at full speed for 3 minutes, the supernatants were removed using a pipette and the bacterial pellets were re-suspended in 180 µl lysis buffer (PBS with 10 mg/ml lysozyme). The samples were incubated for 30 minutes at 37°C while occasionally gently agitating the samples. Subsequently, 25 µl proteinase K solution (QIAGEN), 85 µl dH_2O and 110 µl 10% w/v SDS solution were added and mixed by inversion. Samples were incubated for 30 minutes at 65°C while occasionally gently agitating the samples, followed by the addition of 4 µl RNase solution (Sigma-Aldrich). Samples were mixed by inversion and incubated for an additional 15 minutes at room temperature. 2 ml Phase-Lock Gel Tubes Heavy (5 PRIME) were centrifuged at full speed for 1 minute, followed by pre-filling the Phase-Lock Tubes with 400 µl Phenol:Chloroform:Isoamyl Alcohol (25:24:1) (Sigma-Aldrich). Samples were added to the pre-filled Phase-Lock Tubes and the solutions were mixed thoroughly by inversion, followed by centrifugation at full speed for 3 minutes. After centrifugation, the top layer was transferred into a fresh Phase-Lock Tube. This process of mixing the sample solution with Phenol:Chloroform:Isoamyl Alcohol by inversion and subsequent separation of the two layers by centrifugation

was repeated a further two times. The final top layer was transferred into a clean 1.5 ml tube containing 40 µl 3 M NaAc and 800 µl ice cold 100% ethanol. Samples were mixed gently but thoroughly by inversion and incubated at -80°C for 30 minutes. Next, the samples were centrifuged at full speed for 15 minutes and the supernatant was removed quickly using a pipette to prevent re-suspension. The DNA pellets were washed by adding 1 ml of 70% ethanol and centrifugation at full speed for 3 minutes. After washing the DNA, the supernatant was removed carefully using a pipette and centrifuged again at full speed for 3 minutes to remove any remaining supernatant with a pipette. Finally, the DNA pellets were air dried for 45 minutes, followed by re-dissolving the DNA in 50 µl dH_2O. The samples were stored at -20°C.

2.10.3 DNA amplification

2.10.3.1 Polymerase chain reactions with purified DNA

Purified DNA was amplified using KOD Hot Start Master Mix (Novagen®). Polymerase chain reactions (PCRs) contained 0.8 µl forward primer (10 µM), 0.8 µl reverse primer (10 µM), 12.5 µl KOD Hot Start Master Mix, 9.9 µl H_2O and 1 µl of template DNA (1-50 ng). All PCRs were performed according to the manufacturer's instructions using a thermal cycler block (Applied Biosystems Gene Amp PCR system 2400). Difficult PCRs were performed using a touchdown cycling program; during the first 10 cycles, the annealing temperature decreased 1°C per cycle, and amplification is continued using the predicted primer annealing temperature.

2.10.3.2 Colony PCR

To rapidly screen for transformants and mutants, 1 µl of a bacterial cell suspension in dH_2O was used as template in PCRs using the DreamTaq Green PCR Master Mix kit.

Colony PCRs were assembled and performed according to the manufacturer's instructions.

2.10.3.3 Quick C. botulinum 'colony' PCR

Before colony PCR screening became available in the *C. botulinum* laboratory, cells were pre-treated with TE buffer (10 mM Tris pH 7.5, 1 mM EDTA pH 8.0) prior to the PCR. Colony material was re-suspended in 1 ml TE buffer, incubated at 37°C for 1 hour and subsequently centrifuged for 1 minute at full speed. The pellet was re-suspended in 500 µl dH_2O and 3 µl of this cell suspension was added to a PCR mix with KOD Hot Start Master Mix.

2.10.3.4 Splicing by overlap extension PCR

Knockout cassettes were generated by splicing by overlap extension (SOE) PCR. This approach relies on the design of oligonucleotides with overlapping sections, and is described by Horton, R.M.[226]

2.10.3.5 Primers

Primers were designed using SnapGene® version 3.2.1 (GSL Biotech LLC), synthesized by Eurofins Genomics or Sigma-Aldrich (salt-free, lyophilised) and were reconstituted in dH_2O as per the manufacturer's instruction. The concentrated stocks were diluted to a working stock of 100 pmol/µl. Stocks were stored at -20°C. All primers used in this study are listed in Appendix IV.

2.10.4 Agarose gel electrophoresis

DNA fragments were separated and visualised by performing gel electrophoresis. The agarose gels consisted of 1% (w/v) agarose (MP Biomedicals) in 1× TAE buffer (40 mM Tris, 1 mM EDTA, 0.1% Glacial Acetic Acid; VWR International Ltd) and SYBR® Safe DNA Gel Stain (Invitrogen). Prior to loading DNA on the gel, NEB

loading buffer (6×) was added to the sample. Gels were run at 90-120 V for 40-60 minutes in 1× TAE buffer and subsequently the DNA was visualised in a Gel Doc™ XR+ imaging system (Bio-Rad). When required, agarose gel slices containing the DNA fragment of interest were cut using a clean scalpel.

2.10.5 DNA purification

Amplified DNA fragments were purified using the Monarch® PCR & DNA Cleanup kit, according to the manufacturer's instructions. DNA was extracted from agarose gel slices using the Monarch® DNA Gel Extraction kit, following the manufacturer's instruction. DNA fragments were eluted from the columns in dH_2O.

2.10.6 DNA quantification

DNA concentrations were measured using a NanoDrop Lite Spectrophotometer (Thermo Fisher Scientific), using dH_2O as a blank. For NGS purposes the concentration of genomic DNA was determined using a Qubit 4 Fluorometer (Invitrogen), as per manufacturer's instruction.

2.10.7 DNA restriction by endonuclease digestion

DNA digestion reactions were performed using enzymes and buffers from NEB. The reaction mixtures consisted of ± 1 µg of short DNA fragments or plasmid DNA, 1 µl of the required restriction enzyme(s), 5 µl of the appropriate buffer (NEB) and x µl distilled water, making up the total reaction volume to 50 µl. Digestion reactions were incubated for at least 1 hour at 37°C.

2.10.8 DNA ligation

Plasmids were constructed and DNA fragments assembled using T4 DNA ligase and buffer (Promega), as per manufacturer's instruction. Digested DNA fragments were

mixed together at a molar ratio of 1:1, 1:3 or 1:5 (backbone:insert). Ligation reaction mixtures were incubated for 3 hours at room temperature or overnight in melting ice. Prior to electroporation, the 10 or 20 µl ligation reaction mixtures were dialysed using a 0.025-µm MF-Millipore™ dialysis discs floating on dH_2O in a petri dish for ± 20 minutes.

2.10.9 Sanger sequencing

DNA constructs and mutations in bacterial strains were verified by Sanger sequencing using the Eurofins Genomics' plasmid DNA and purified PCR products TubeSeq services. Samples were provided in accordance with the companies requirements.

2.10.10 Illumina genome sequencing

Genome sequencing was provided by DeepSeq (NGS platform of the University of Nottingham) using their Illumina-based MiSeq Standard and MiSeq v3 300PE (600cycle kit) Whole-Genome Service.

2.11 Plasmid construction

All the plasmids used and constructed in this study are listed in Appendix II.

2.11.1 Construction of plasmids used in Chapter 3

2.11.1.1 Conjugative plasmids with type I recognition sites

To create the plasmids with the type I recognition sites, recognition site one (R1) was created by mixing primers TI-R1_Cas9_F/TI-R1_Cas9_R (no PCR-step required), recognition site two (R2) was created by mixing primers TI-R2 cas9_F/TI-R2_Cas9_R (no PCR-step required), and the two recognition sites together were amplified using primers TI-R1+2_Cas9_NdeI_F/TI-R1+2_Cas9_NotI_R using pRECas1-MSC as a template. The primers used to create DNA fragments R1 and R2 have overhangs on

each 5’ and 3’ end that match a NotI and NdeI restriction site, respectively. The three inserts were cloned into the pMTL83151 plasmid via NotI/NdeI digestion, generating plasmids pMTL83151-R1, pMTL83151-R2 and pMTL83151-R1+R2. Insertion of the recognition sites was confirmed by Sanger sequencing with primer M13F.

2.11.1.2 Expression plasmids for MTases

To create the pM plasmids carrying the ATCC 19397 MTases eight fragments were amplified: terminator B0014 was amplified with primers T-B0014_F/T-B0014_R (no template required; primer dimer reaction), terminator B0010 was amplified with primers T-B0010_F/T-B0010_R (no template required; primer dimer reaction), low copy replicon pBBR1 was amplified with primers pBBR1_F_AscI/pBBR1_R_FseI using a pMTL80000 series plasmid as a template, the apramycin resistance gene was amplified with primers Apr_F_FseI/Apr_R_SbfI using RBV1 as a template, the type I genes *hsdM* and *hsdS* with their native RBS (TI) were amplified with primers CLB-RS03505_F_BsaI2/CLB-RS03510_R_NotI using genomic DNA as a template, the type II MTase with its native RBS (TII) was amplified using primers CLB-RS13900_F_BsaI/CLB-RS13900_R_NotI using genomic DNA as a template, insert TI+TII was amplified using primers CLB-RS13900_F_BsaI/ CLB-RS03510_R_NotI using RBV1-BTI+TII as a template, the inducible promoter P_{BAD}-*araC* was amplified with primers araC_R_XbaI/PBAD_R_BsaI using RBV1 as a template and weak promoter P_{J23105} was amplified with primers PJ23105_F_XbaI/PJ23105_R_BsaI (no template required; primer dimer reaction).

Subsequently, all fragments were digested with the appropriate restriction enzyme: B0014 was digested with NotI/AscI, B0010 was digested with SbfI/XbaI, pBBR1 was digested with AscI/FseI, the *apr* gene was digested with FseI/SbfI, insert TI was digested with BsaI/NotI, insert TII was digested with BsaI/NotI, P_{BAD}-*araC* was

digested with XbaI/BsaI and P_{J23105} was digested with XbaI/BsaI. Subsequently, the plasmids were assembled by mixing the appropriate fragments in one ligation mixture, generating plasmids pM-JTI, pM-JTII, pMJTI+TII, pM-BTI and pM-BTII (see Table 2.3). Two representative plasmid maps are depicted in Appendix III. The presence of all fragments in each plasmid was verified by Sanger sequencing using primers CLB_RS03505_S1_R, CLB_RS03510_S1_R, CLB_RS03510_S2_R and the primers used for cloning.

Table 2.3 Ligation mixtures to create pM plasmids

Ligation mix	B0014	B0010	pBBR1	*apr*	TI	TII	TI+TII	P_{BAD}-*araC*	P_{J23105}
1	X	X	X	X	X				X
2	X	X	X	X		X			X
3	X	X	X	X			X		X
4	X	X	X	X	X			X	
5	X	X	X	X		X		X	

2.11.1.3 Insertion of recombineering cassette into R702 plasmid in E. coli

The R702 plasmid carrying the ATCC 19397 MTases was created in two steps. First plasmid RBV1-BTI+TII was created and in the second step the recombineering cassette, amplified from RBV1-BTI+TII, was integrated into R701 *in vivo*.

For plasmid RBV1-BTI+TII, the inserts TI (type I genes *hsdM* and *hsdS* with native RBS) and TII (type I MTase gene CLB_RS13900 with native RBS) were amplified using primers CLB-RS03505_F_BsaI/CLB-RS03510_R_BsaI and CLB-RS13900_F_BsaI/CLB-RS13900_R_BsaI, respectively, using genomic DNA as a template. The inserts were cloned into RBV1 (Appendix III) via BsaI digestion and successful plasmid assembly was confirmed by PCR with primers p15a+tra-F1/M13R. To confirm that all three genes were free of mutations the RBV1-BTI+TII plasmid

was sequenced using primers CLB_RS03505_S1_R, CLB_RS03510_S1_R, CLB_RS03510_S2_R, RE_F and the primers used to amplify the inserts.

Once the sequences were confirmed, the entire recombineering cassette, comprised of LHA, P_{BAD}-*araC*, TI+TII, *apr* and RHA, was amplified with primers RE_LHA_F and RE_RHA_R. The cassette was electroporated into *E. coli* strains RE-CA434 and RE-InterStellar and the cells were plated onto LB/Km/Apr agar. Correct recombinants were confirmed by PCR in three biological replicates using primers RE_F and RE_R, annealing to the cassette and the R702 plasmid, respectively. The presence of the cassette in the final InterStellar R702TI+TII mutants was subsequently verified by Sanger sequencing.

2.11.2 Construction of plasmids used in Chapter 4

2.11.2.1 Plasmid to generate ATCC 3502 boNT/A1 deletion strains

CRISPR/Cas9-mediated mutagenesis was carried out using the pMTL431511 plasmid backbone with a *cas9* gene containing a proximal frameshift mutation (fs*cas9*).[227] The sgRNA SEED sequence was selected based on the on- and off- target scores calculated by the Benchling platform and generated via a primer dimer PCR with primers sgRNA1_boNT/A1_F_SalI and sgRNA_R_AsiSI. The *boNT/A1* deletion cassette left homology arm (LHA) was amplified with primers boNT/A1_LHA_AsiSI_F and boNT/A1_LHA_R; and the right homology arm (RHA) was amplified with primers boNT/A1_RHA_F and boNT/A1_RHA_R_AscI, using genomic DNA as a template. The two homology arms were fused via SOEing PCR. The sgRNA was digested with SalI/AsiSI and the boNT/A1 deletion cassette with AsiSI/AscI, and all fragments were ligated into a SalI/AscI digested pMTL431511 plasmid backbone in one step, generating plasmid pMTL431511_boNT/A1-G1.

2.11.2.2 Plasmids to generate ATCC 19397 mutL variants

CRISPR/Cas9-mediated mutagenesis was carried out using the pRECas2-MSC plasmid backbone.[158] The sgRNA SEED sequences were selected based on the on- and off- target scores calculated by the Benchling platform, generated via primer dimer PCRs and digested with SpeI/AatII (Appendix IV). The LHAs and RHAs of the deletion and insertion cassettes were amplified using primers listed in Appendix IV using genomic DNA as a template and digested with AatII/BsaI and BsaI/AscI, respectively. The digested sgRNA, LHA and RHA fragments were ligated into a SpeI/AscI digested pRECas2-MSC plasmid backbone in one step, generating plasmids pRECas2-mutL-G1d, pRECas2-mutL231-G1d, pRECas2-mutL_comp-G2c and pRECas2-mutL_comp-G1c.

2.11.3 Construction of plasmids used in Chapter 5

CRISPR/Cas9-mediated mutagenesis was carried out using the pRECas1-MSC plasmid backbone.[158] The sgRNA SEED sequences were selected based on the on- and off- target scores calculated by the Benchling platform, generated via primer dimer PCRs and digested with SalI/AatII (Appendix IV). The LHAs and RHAs of the deletion and toxoid cassettes were amplified using primers listed in Appendix IV using genomic DNA or plasmids DNA as a template and digested with AatII/BsaI and BsaI/AscI, respectively. The toxoid cassettes were assembled via SOEing PCR. The digested sgRNA, LHA and RHA fragments were ligated into a SalI/AscI digested pRECas1-MSC plasmid backbone in one step, generating the plasmids listed in Appendix II.

2.12 Targeted genome mutagenesis

2.12.1 ClosTron mutagenesis

The ClosTron system was used to inactivate target genes via intron insertion, as described by Heap *et al*.[179, 228] Re-targeted group II introns were designed using a freely available tool based on the Perutka algorithm (www.clostron.com). Standard ClosTron plasmid pMTL007C-E2 carrying the newly designed introns were supplied by ATUM and eluted from the membrane using dH_2O, according to the manufacturer's instruction. Plasmids were electroporated into CA434 and subsequently conjugated into *C. botulinum*, as described in section 2.5 and 2.7, respectively. Conjugal mixtures were plated on TYG/Tm/D-cycloserine agar and incubated for 24-48 hours. Successful transconjugants (Tm resistant) were streaked to purity on TYG/Tm/D-cycloserine agar. Next, all transconjugants were harvested, pooled by scraping, serially diluted in BPS as required and plated on TYG/Erm/D-cycloserine agar to select for intron insertion. Group II intron insertion was confirmed by PCR using primers flanking the target site and subsequently the plasmid was lost as described in section 2.8.

2.12.2 Allele-Coupled Exchange

The allele-coupled exchange (ACE) method was used to generate *pyrE* negative (*pyrE*$^-$) *C. botulinum* strains.[188] Plasmid pMTL-CH12 was transferred from CA434 to *C. botulinum* via conjugation and the conjugal mixture was plated onto TYG/Tm/D-cycloserine agar, as described in section 2.7. After 24-48 hours of incubation morphologically larger colonies were screened via PCR with a plasmid-specific and chromosomal-specific primer to identify single-crossover mutants. The single-crossover mutants were plated on TYG agar containing 5.5 mg/ml 5-FOA to isolate

double-crossover *pyrE*$^{-}$ mutants, resistant to 5-FOA. Plates were incubated under anaerobic conditions for 48-72 hours and the colonies resistant to 5-FOA were screened via PCR using a primer annealing to the chromosome and a primer annealing to the *lacZα* insert. Double-crossover mutans were streaked to purity on TYG agar containing 5.5 mg/ml 5-FOA and the plasmid was lost as described in section 2.8.

2.12.3 CRISPR/Cas9-mediated mutagenesis

CRISPR/Cas9-mediated mutagenesis was carried out using two different systems: (i) with constitutive Cas9 expression (pMTL40000 series); and (ii) with inducible Cas9 expression (RiboCas).[158, 227] The CRISPR/Cas9 plasmids were transferred from *E. coli* to *C. botulinum* by conjugation as described in section 2.7. Conjugal mixtures were plated onto TYG/Tm/D-cycloserine agar and incubated for 24-48 hours. Transconjugants harbouring pMTL431511 were directly screened via PCR, using the appropriate screening primers, to confirm the desired genomic modification, and subsequently the plasmid was lost as described in section 2.8. Transconjugants harbouring the RiboCas plasmid were restreaked onto TYG/Tm/ D-cycloserine agar containing 5 mM theophylline and incubated for 24-48 hours. Colonies were screened via PCR, using the appropriate primers, to confirm the desired genomic modification and subsequently the plasmid was lost as described in section 2.8.

2.13 List of bioinformatics tools

Tool	Description
SnapGene® version 3.2.1 (GSL Biotech LLC)	Plasmid map visualisation, primer design, cloning simulations and Sanger sequencing alignments
Blastn & Blastp algorithms (NCBI)	Sequence alignments and database searches
Benchling	Design of sgRNAs
NCBI conserved domain database	Identification of protein domains
CLC Genomics Workbench v10 (Qiagen)	Whole genome assembly, R-factor assembly, resequencing analysis and variant detection
Mauve Multiple Genome Alignment	Alignment of *C. botulinum* genomes
RAST server (Rapid Annotation using Subsystems Technology)[229]	Annotation of assembled sequences
GraphPad Prism 9.1.2 for windows	Data visualisation and statistical analysis

CHAPTER 3

Development of a donor customisation pipeline to improve DNA transfer efficiencies

3.1 Introduction

3.1.1 Horizontal gene transfer

Horizontal gene transfer refers to the movement of genetic material between organisms other than by the passing of genetic material from parent to offspring.[230] The ability of bacteria to acquire genetic material from their environment (e.g., directly or indirectly from other organisms) plays a major role in bacterial evolution and enables bacterial populations to overcome challenging environmental changes. The uptake of new (foreign) genetic material can serve three purposes: (i) generate genetic diversity; (ii) function as a template for DNA repair, and; (iii) provision of nutrients.[231-232] A good example of the impact of horizontal gene transfer is the fast increase of multidrug resistant bacteria worldwide, which is due to the transfer of resistance genes between bacterial species.[233-234] Another common phenomenon is the adaptation of bacteria to a pathogenic lifestyle, which usually involves the acquisition of new genetic material, e.g., a large cluster of virulence genes, through horizontal gene transfer.[230] The *boNT* gene cluster, responsible for the pathogenic phenotype of *C. botulinum*, is a good example of toxin genes that have moved between different species, although its exact evolutionary benefit is still unknown.[12, 44, 46-47, 165, 235]

In order to gain new abilities through horizontal gene transfer, donor DNA has to find its way into the recipient cell and subsequently must be maintained within the cell, which can either be via integration into the cell's genome or via autonomous replication.[230] Additionally, for long term stable maintenance, the DNA should be beneficial to the recipient cell. In bacteria, this process of horizontal gene transfer may occur via three different mechanisms: transformation, transduction and conjugation (Figure 3.1).[236-239] Each mechanism will be discussed in more detail below.

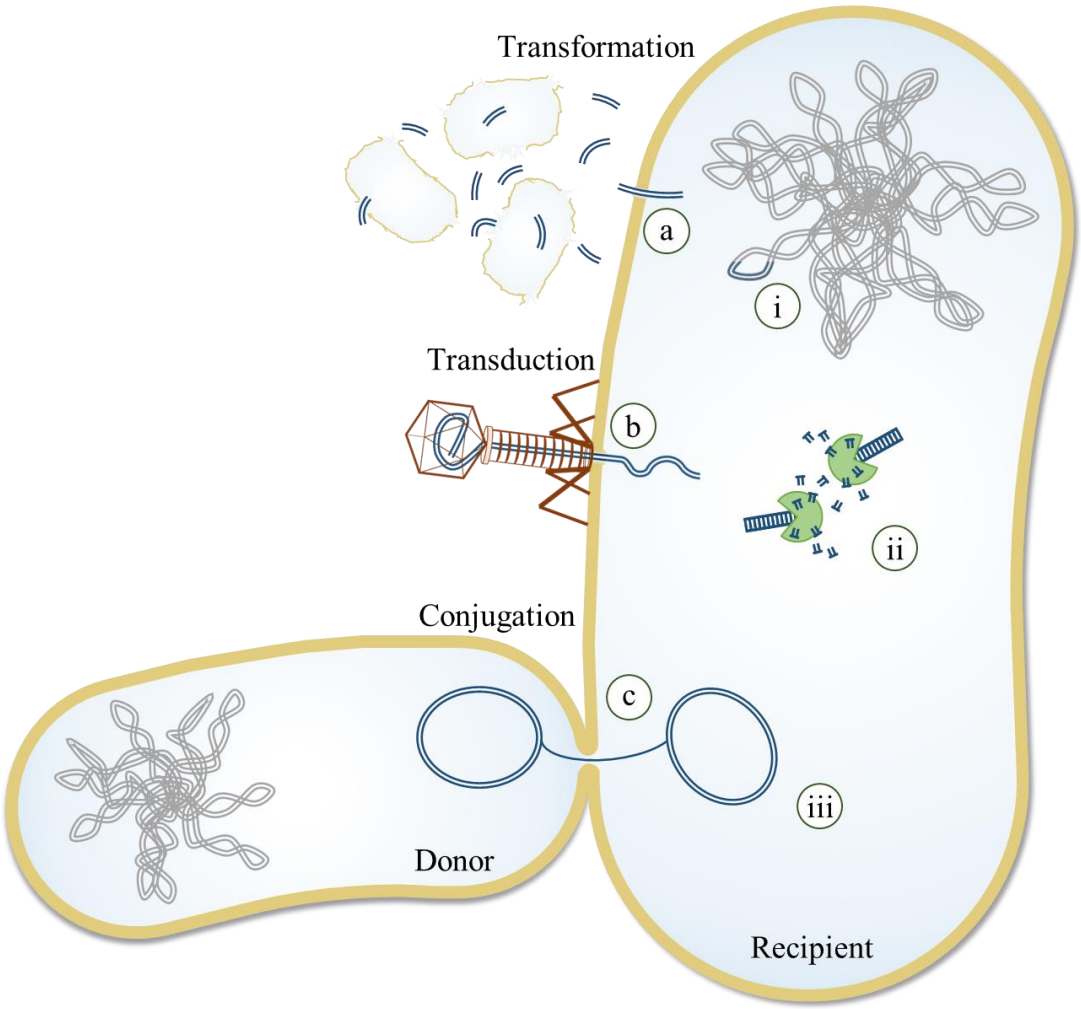


Figure 3.1 Overview of horizontal gene transfer mechanisms. a) Transformation: the uptake of free genetic material from the environment; b) Transduction: the accidental transfer of DNA from one bacterium to another by a bacteriophage; c) Conjugation: the transfer of genetic material between bacteria by physical contact. Genetic material that has entered the cell is subsequently (i) integrated into the host chromosome; (ii) degraded by nucleases, or; (iii) recircularised to become a self-replicating entity.

3.1.2 Transformation

Transformation was the first mechanism of horizontal gene transfer to be discovered and involves the uptake of free exogenous genetic material by the cell from its environment through the cell membrane(s). In 1928 Griffith was the first to describe the ability of bacteria to transform by acquiring a transforming principle from the

environment and a few years later, in 1944, Avery *et al*. demonstrated that the transforming factor was exogenous genetic material.[240-241] This phenomenon, that they termed bacterial transformation, can only occur if the recipient cells enter a specialised physiological state, known as 'competence'. Various bacteria can naturally develop competence whereas in other strains an artificial state of competence can be induced using laboratory procedures.

Natural transformation involves a physiological state transition of the recipient cell, triggered by environmental stimuli (e.g., population dynamics, DNA damaging conditions and starvation) and depends on the tightly regulated expression of genes connected to the DNA uptake process.[237] Naturally competent Gram-positive and Gram-negative bacteria have a related competence system, which involves the assembly of a structure that partially spans the cell membrane and includes components that share homology with proteins from the type IV pili and type II secretion systems.[242] The formed structure is coupled to a DNA translocation complex and together they transport single-stranded DNA into the cytoplasm. Once in the cytoplasm, single-stranded DNA gets a) recircularised (plasmids), b) integrated into the genome via homologues recombination or c) degraded by nucleases (Figure 3.1).[243]

Artificial competence is based on making cells passively permeable to DNA using laboratory procedures. The cells are exposed to conditions that cause the formation of small holes in the cell membrane and/or the opening of membrane pores through which DNA can enter the cell. The two most common methods to induce competence for transformation are chemical induction and electroshock treatment (electroporation). To chemically induce competence, cells are treated with ice-cold solutions containing divalent cations (commonly Ca^{2+}) and are subsequently subjected to a heat-shock

treatment to make the membrane permeable.[244] Electroshock treatment is currently the most efficient artificial transformation method available and involves thoroughly washing bacterial cells to remove excess salts (ions) before subjecting the cells to a high-voltage pulse, which elicits the uptake of macromolecules.[245-246] Since the discovery of these methods to induce an artificial state of competence in bacteria that are not naturally competent, they have become commonly used tools to genetically modify bacteria, rapidly advancing the fields of molecular biology and biotechnology.

3.1.3 Transduction

The second mechanism of horizontal gene transfer, termed transduction, involves the accidental transfer of DNA from one cell to another by a bacteriophage. When a bacteriophage infects a cell it can enter either the lytic or the lysogenic lifecycle. During the lytic cycle phages harness the host machinery to assemble new phage particles, which are released to the environment upon lysis of the host. The second route for bacteriophage reproduction, lysogeny, is characterised by coexistence of the phage with the host, as a so called prophage.[247-248] In most cases, the genetic material of a prophage is integrated into the host-cell chromosome, where it can remain dormant for extended periods of time and passively replicate from bacterial parent to offspring for many generations. Nonetheless, upon induction, the prophage can exit the bacterial chromosome, switching from the lysogenic to the lytic cycle. Prophage induction initiates the excision of its DNA and the subsequent hijacking of the host machinery to assemble new phage particles, which eventually results in cell lysis.

Normally, exact copies of the original genetic material is packaged into the new bacteriophage particles. However, occasionally erroneous packaging of the genetic material occurs and host DNA is included. When such a bacteriophage infects another

bacterium, this genetic material is released, resulting in horizontal gene transfer. The infecting bacteriophage could have packaged random DNA fragments from the previous host (generalised transduction) or packaged DNA flanking the prophage integration site due to imprecise excision of the prophage DNA (specialised transduction).[230] Generalised transduction can occur through the lytic and the lysogenic lifecycle whereas specialised transduction can only occur via the lysogenic lifecycle.

Like transformation, transduction does not require direct contact between the donor and recipient cell, the only attribute they must share is a common sensitivity for the transducing bacteriophage. Although host range of a bacteriophage is usually quite limited, making horizontal gene transfer via transduction highly specific, it is an important vehicle for DNA transfer, owing to their abundance in the environment and their ability to promote DNA integration into the host's chromosome, protecting the genetic material from degradation by restriction endonucleases (REases).

3.1.4 Conjugation

The third mechanism of horizontal gene transfer, known as bacterial conjugation, is the transfer of genetic material, usually plasmids or transposons, by cell-to-cell contact. When a donor and recipient cell come into contact a mating junction is formed through which DNA is exchanged. This junction can be formed between cells of different species, phyla and even major domains.[249] Despite the availability of quite a substantial amount of molecular data regarding the conjugative apparatus, the exact mechanism of conjugal DNA transfer remains to be elucidated. Currently, the available data suggests that a conjugative pilus produced by the donor cell extends from the cell surface, which facilitates attachment to the recipient and induces the

formation of a mating junction. However, these pili probably do not play a direct role in the transfer of the genetic material.[250] The formation of the actual junction requires the assembly of a cell membrane spanning translocation channel, which involves extensive remodelling of the donor and recipient cell membranes. Both distributed and polar patterns of localisation have been observed for the pili and channel components of the conjugation apparatus.[239]

The genes that encode the required components for the conjugation apparatus are usually located on a plasmid, but occasionally get integrated into the chromosome of a donor cell. Such strains, with the plasmid integrated, transfer chromosomal genes with a high efficiency and are therefore referred to as High frequency of recombination (Hfr) strains. The most common functional components constituting these plasmids, known as fertility factors (F factors), are the origin of transfer (*oriT*), origin of vegetative replication (*oriV*) and conjugal transfer (*tra*) genes.[251] Upon initiation of conjugation a relaxase (or relaxosome) binds at the *oriT* locus and cleaves one of the DNA strands.[252] Subsequently, the nicked strand is translocated into the recipient cell, while the remaining strand is replicated inside the donor cell.[239] As a result, after a successful conjugation event, both the donor and recipient have a copy of the F factor. Some conjugal plasmids have a very limited host range, whereas others are less host specific.

The host range of these conjugal plasmids is not only determined by the successful formation of a mating junction, but also depends on how well the transferred plasmid is maintained within the recipient cell. Moreover, the frequency of plasmid transfer is indirectly influenced by environmental factors, due to their effect on bacterial activity and population density. For example, the high bacterial density and metabolic activity in biofilms seems to increase the total number of DNA-transfer events.[249]

Furthermore, transfer efficiencies are affected by restriction barriers within the recipient, which will be discussed in more detail in section 3.1.6.

Just like transformation and transduction, conjugation is often utilised by researchers to transfer DNA from one organism to another for engineering purposes; its main advantage being that relatively large DNA molecules can be transferred. Conjugative plasmid transfer is a common method to genetically modify various clostridia, including the toxicogenic species *C. difficile* and *C. botulinum*.[206] For these two species the majority of methods rely on an *E. coli* donor harbouring a shuttle plasmid and the R702 factor, which is a resistance transfer factor (R factor). R factors encode all the required components for conjugation apparatus formation, allowing them to readily spread through populations. They are characterised by the multiple resistance genes they carry and are closely related to F factors. The shuttle plasmid carries an *oriT*, a selectable marker, cargo, a Gram-negative *E. coli* replicon and a Gram-positive *Clostridium* replicon to allow its successful transfer.[180] The exact composition of the shuttle vector can influence its final transfer efficiency; for example, plasmids with a stable Gram-positive replicon transfer with a higher frequency compared to plasmids with an unstable Gram-positive replicon.

3.1.5 Bacterial defence mechanisms against parasites: barriers to horizontal gene transfer

Engineering bacterial strains with the aforementioned DNA transfer methods is not always straightforward because, as so often in biology, it is not an easy 'one size fits all' case. It is not uncommon that with a specific transformation protocol high numbers of transformants (cells into which foreign DNA has been introduced) can be achieved for one strain whereas for another strain, of the same species, significantly less transformants are obtained.[244] This is not only due to the many physiological

differences between bacterial species, like cell membrane composition, ability to become naturally competent or form a conjugation apparatus, but is also influenced by the various defence mechanisms a cell might have developed and/or acquired to protect itself from potentially harmful foreign genetic material, such as viral DNA.

For bacteria there is a fine balance between acquiring beneficial genes and protection against bacteriophages and invasive plasmids. Especially between phages and bacteria an evolutionary arms race has been going on for ages, with both sides continuously evolving to overcome the defence mechanisms of the other.[253] This has led bacteria to develop multiple lines of defence, some very specific (e.g., CRISPR-Cas systems) and others more broad-range (e.g., restriction-modification (RM) systems), that are collectively referred to as the prokaryotic immune system.[254] The acquisition and evolution of these different immune strategies is driven by several key ecological factors, such as the frequency of phage attacks, parasite diversity, the genetic diversity in a phage population and the bacterial population density. Although the benefits of sustaining a diverse defence arsenal are evident, investing in and maintaining a multitude of defence mechanisms comes at a cost. The systems can impose a metabolic burden but more importantly can lead to autoimmunity; for example, defence mechanism that should attack foreign DNA can accidentally cleave the chromosome, resulting in cell death.[255] Moreover, permitting a low level of horizontal gene transfer can benefit the fight against harmful genetic material, since it can be employed to access defence systems encoded by other bacteria in the environment.[256]

Challenged by the abundance and diversity of bacteriophages and given the fitness costs associated with maintaining various defence mechanisms, bacteria have evolved in such a way that they can adapt quickly to the ever changing environment by gaining and losing these mechanisms at a relatively high rate.[254] To do so, the various defence

systems are usually clustered together at specific chromosomal regions, referred to as defence islands and oftentimes these islands also encode transposases, recombinases, conjugation genes or other typical mobile genetic elements.[257] It has been hypothesized that a bacterial immune system is not just the defence mechanisms within a single cell, but rather should be viewed as the sum of all immune systems available within a population (pan-genome), shared via horizontal gene transfer.[255] Additionally, diversifying the defence arsenal amongst cells within a population could be seen as a bet-hedging survival technique to combat the continuously evolving parasites.

In recent years, it has become evident that the bacterial immune system is much more diverse and complex than previously appreciated. Here, the focus is on how these systems can interfere with various horizontal gene transfer processes, limiting the successful uptake of foreign DNA.

In nature, bacterial transformation relies on the availability of DNA in the environment and hence, the major barriers to horizontal gene transfer via this route are related to the extracellular DNA (eDNA) stability, eDNA degradation by nucleases and other environmental factors that affect eDNA concentrations.[258] In addition, the specificity of the host DNA uptake system, limited bacterial competence and the lack of homology between foreign DNA and the host chromosome influence the transformation efficiency.

Horizontal gene transfer via transduction is hindered by the vast number of anti-phage defence systems bacteria have at their disposal.[259] These systems can roughly be divided into three categories: (i) systems that target viral nucleic acids; (ii) Abortive infection (Abi) systems which result in bacterial suicide upon infection, and; (iii) other

(Figure 3.2).[260] The various groups interfere with different stages of the phage lifecycle and most of the known systems have been studied extensively. Briefly, a very effective and widespread method employed by bacteria to fight phages is the degradation of the genetic material they inject into the cell. These defence mechanisms (e.g., RM and CRISPR-Cas systems) also form a major barrier to the uptake of foreign DNA via the other two horizontal gene transfer routes and hence, cells that harbour many of such systems are often genetically inaccessible to researchers. Both RM-systems and the CRISPR-Cas system will be discussed in more detail in section 3.1.6. Abi is quite a drastic stress response by the cell that leads to a general metabolism arrest to prevent the completion of the phage life cycle and can result in cell death, thus preventing the release of new phage particles. Essentially, Abi is an altruistic suicide response that limits the phage spread to protect the rest of the population. Furthermore, bacteria can modify their surface to prevent phage adsorption by masking, mutating or losing receptor proteins. Lastly, phages that are already inside a bacterial host, such as a prophages, can block the injection of foreign genetic material or the replication process of newly infecting phage.[261] This phenomenon is referred to as superinfection exclusion (Sie) and could be viewed as a collaboration between the prophage and its bacterial host. The almost infinite diversity of bacteriophages has resulted in a remarkable diversity of anti-phage defence mechanisms and new systems, such as phage-inducible chromosomal islands (PICIs), bacteriophage exclusion (BREX) and DISARM, are still being discovered. It could well be that the currently known defence mechanisms are just be the tip of the iceberg.

Conjugal immunity systems, encoded by conjugative elements, such as plasmids and transposons, can prevent the successful transfer of DNA, which is known as exclusion. Various distinct exclusion mechanisms encoded by conjugative elements have been

found in both Gram-positive and Gram-negative bacteria.[262-265] For example, the *tra* operon located on F factors encodes two exclusion genes, *traS* and *traT*. TraS is an inner membrane bound protein and inhibits DNA entry. TraT is an outer membrane lipoprotein and is thought to either reduce receptiveness of the recipient cell to conjugative pili or destabilize mating junctions (Figure 3.2). Initially the consensus view was that these exclusion mechanisms prevent recipient cells from acquiring other identical or closely related conjugative elements. However, more recently it has been hypothesized that the surface exclusion systems, such as *traT*, promote the breakage of a mating junction after a successful transfer event.[243]

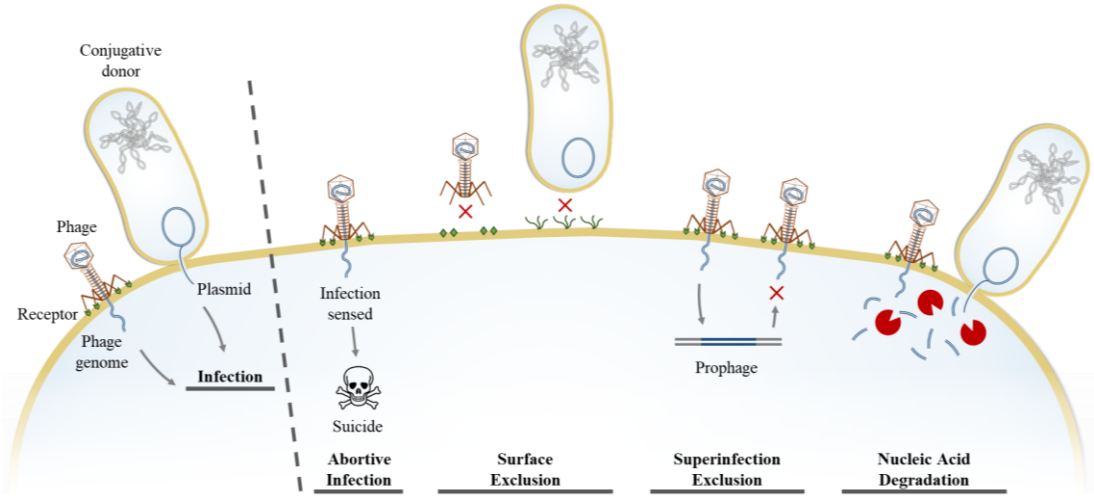


Figure 3.2 Overview of bacterial immune systems. From left to right: transfer and maintenance of invading DNA halted by Abortive infection, Surface Exclusion, Superinfection Exclusion and Nucleic acid degradation. Invading DNA (phage or plasmid): dark blue; Bacterial chromosome: grey; Phage receptors: dark green; Enzyme (complex) with nuclease activity: red.[254]

3.1.6 Bacterial immune systems that target nucleic acids

Bacteria have developed various ingenious mechanisms to attack invading genetic material. Prokaryotic Argonaute (pAgo) proteins, CRISPR-Cas systems and restriction-modification (RM) systems can all interfere with foreign genetic elements that are potentially harmful to the cell (Figure 3.3). Usually, these systems do not inhibit one specific method of horizontal gene transfer, but rather interfere with all genetic material that fit their requirements.

The Argonaute protein family is a diverse group of guide-dependent enzymes that target nucleic acids and members of this family can be found in all three domains of life.[266] pAgo proteins can be divided into long and short pAgos; Long pAgos consist of the amino-terminal (N), middle (MID) and two PIWI-Ago-Zwille (PAZ) domains. The MID and PAZ domains usually bind to the guide oligonucleotide and upon target binding the PIWI domain facilitates target cleavage. Short pAgo are basically incomplete pAgo proteins that lack nuclease activity. They consist of only a MID and a PIWI domain that is catalytically inactive. The primary function of short pAgos might be target binding, which on its own could interfere with RNA/DNA expression (silencing), or recruit additional nucleases to mediate target cleavage. The genes encoding pAgos often co-locate with other host defence mechanisms and short pAgo genes, lacking the cleavage domain, are typically associated with genes that encode putative nucleases. Most of the studied pAgos use small DNA guides to cleave DNA, however there are also examples of RNA-guided Argonaute nucleases that cleave RNA and DNA.[266-268] Both DNA- and RNA-guided pAgos seem to protect the cell against invading DNA by interfering with multicopy genetic elements, such as plasmids and phage DNA.[269] However, the function of pAgos might be more versatile; Jolly *et al.* demonstrated that the Ago of *Thermus thermophilus* is involved in the

unlinking of replicating chromosomes and Fu *et al*. report that an Ago of *Natronobacterium gregoryi* promotes homologous recombination.[270-271] Although in recent years various studies have led to a better understanding of the pAgos function in bacteria, their exact role, how they discriminate self from non-self and other mechanistic details remain to be elucidated.

The discovery of prokaryotic CRISPR-Cas immune systems has probably been the most significant revelation of the past decades. Not only because it is a very sophisticated adaptive immune system that memorizes previous infections but also because these programmable CRISPR-Cas systems have been repurposed for editing eukaryotic and prokaryotic genomes, DNA imaging, base editing and much more, revolutionizing many research fields.[272-273]

The CRISPR loci together with the Cas genes constitute the ubiquitous immune system found in approximately half of the sequenced bacteria and in a majority of archaea.[191-192, 273-274] The *cas* genes encode the protein machinery that drives the immune response, which can be divided into three phases: (i) adaptation; (ii) crRNA biogenesis, and; (iii) interference.[193, 275-276] During adaptation, an immunological memory is formed through the acquisition of invader-derived sequences (spacers) that are incorporated into the CRISPR loci on the host chromosome. This memory is retrieved when the CRISPR loci is transcribed and the resulting transcripts are processed to yield small crRNA guides. Subsequently, the crRNAs direct the interference machinery, consisting of one or more Cas proteins, to cleave complementary invading genetic material, which results in host immunity.[277-278] To prevent autoimmune targeting of the CRISPR loci, the interference machinery only targets sequences next to a specific short recognition sequence, called PAM, which is not included in the CRISPR array by the adaptation machinery.

CRISPR-Cas immune systems are highly diverse and are currently divided into two classes based on the composition of *cas* genes.[191] Class I systems accomplish interference using a multi-Cas protein complex, whereas class II systems employ a single effector protein. Although it has been well established that the various systems can provide protection against invading mobile DNA or RNA elements, such as phages and plasmids, the effect of CRISPR-Cas systems on horizontal gene transfer is debated.[279-280] On the one hand, lineages have been identified with strong horizontal gene transfer signatures that are depleted of CRISPR-Cas systems. On the other hand, studies suggest that the presence of CRISPR-Cas systems can enhance generalised transduction because they target the genetic material of a phage, but do not recognize bacterial DNA that was encapsulated accidentally.

RM systems provide innate immunity against invading DNA by recognising and cleaving foreign DNA that contains a specific sequence motif or DNA modification. They are present in over three quarters of the sequenced prokaryotic genomes and a bacterial genome harbours on average two RM systems.[281] There are roughly two kinds of RM systems: those that cleave DNA lacking appropriate methylation of adenosyl or cytosyl residues within the recognition motif and those that target modified DNA.[282] The RM systems belonging to the first group are typically comprised of enzymes with opposing activities: a restriction endonuclease (REase) that recognizes and cleaves unmethylated motifs and a methyltransferase (MTase) that modifies the REase recognition sequence in the host chromosome to prevent self-targeting.[283] Foreign DNA usually lacks the characteristic methylation of the REase recognition site and will therefore be degraded. To further minimize autoimmunity in bacterial cells expressing functional RM systems the cognate restriction site is

generally underrepresented in the genome, which is referred to as recognition site avoidance.[284-285]

In addition, emerging studies indicate that RM-systems might play a role in various other processes, beyond host defence.[286] RM-systems can restrict the acquisition of foreign (new) DNA by horizontal gene transfer, thus maintaining the genetic integrity of the population.[287] To date, many studies have shown that RM systems affect horizontal gene transfer and form a barrier to both natural DNA transfer as well as artificial transfer methods.[288-290] However, it has also been hypothesised that the REase generated fragments could promote homologous recombination with the host's genome.[286] Furthermore, the site specific methylation by RM systems might act on gene expression.

RM systems are classified into four different types based on their subunit composition and biochemical characteristics, such as substrate specificity, cleavage position and cofactor requirements.[291-292]

Type I systems are comprised of two REase subunits (R) for cleavage, two MTase subunits (M) that catalyse the DNA modification reaction and one specificity subunit (S), which recognises specific bipartite target sequences (e.g., AACNNNNNNGTGC of EcoKI[293]) and are encoded by the *hsdR*, *hsdM* and *hsdS* genes, respectively.[294] The DNA methylation status determines if the complex functions as a REase or as an MTase; fully methylated DNA is protected against restriction, the non-methylated strand of hemimethylated DNA is targeted for methylation and unmethylated molecules are susceptible to restriction. When the complex binds an unmethylated recognition site, the R subunit starts the ATP-dependent DNA translocation process, reeling in the DNA base pairs while the REases remain bound to the recognition

sequence, which results in the formation of DNA loops.[295-296] Stalling of the translocation (e.g., due to DNA topology or the collision of two translocating enzyme complexes) initiates cleavage of both strands within the formed DNA loop, and thus restriction occurs at a random distant site from the recognition site.[297]

Type II systems are a heterologous collection of hundreds of different REases with unique biochemical properties and their associated MTases.[298] Generally, the MTase and REase are separate and independently acting enzymes that recognise with extreme accuracy the same DNA sequence. This sequence is typically a 4-8 base pair palindrome and the type II REase cleaves at a fixed position within or adjacent to this specific recognition site. Type II REases can be active as homodimers, monomers or tetramers whereas MTases usually act as monomers to modify both DNA strands at a specific recognition site. In modern molecular biology type II enzymes are much more than just a prokaryotic immune system; in laboratories they have become daily used 'work horses' for the analysis of DNA and gene cloning, playing a key role in many developments within the life sciences.[299]

Type III systems share many features with type I systems but are less complex.[300] They are comprised of a REase ('res' subunit), which is only active when associated to the complex and of a MTase ('mod' subunit), which is responsible for site recognition and modification. The ratio of the different subunits within the final complex is debated, studies suggest that the complex either consisting of one or two 'res' and two 'mod' subunits.[301] Type III complexes bind to two inversely oriented copies of its non-symmetrical 5-6 base pair recognition site and as with type I systems, translocate DNA while the complex remains bound to the recognition site and loop formation precedes cleavage.[302] However, unlike type I systems, stalled DNA translocation initiates restriction at a defined point, which is usually 25–27 bases away from the recognition

site. In contrast to the res subunits, the mod subunits can act independently, methylating just one strand of a DNA duplex, which results in hemimethylated DNA. This DNA is protected against cleavage because type III systems only cleave completely unmethylated DNA molecules. In addition, there is competition between the res and mod activities of the complex. Hence, foreign DNA occasionally gets methylated before restriction could occur, making type III systems a less efficient defence mechanism.

Type IV systems are a relatively new class of restriction enzymes that is modification-dependent (e.g., bases that have been methylated, hydroxymethylated, or glucosyl-hydroxymethylated), in contrast to the conventional modification-blocked type I-III systems.[303-304] Many of the known enzymes described have been discovered during the past three decades and if the hypothesis is true that for each type of protective modification there are enzymes that only cleave DNA with this specific modification, there are still many undiscovered type IV systems. The currently described type IV enzymes are highly diverse and sequence analysis suggest that they have emerged independently at least six times during evolution. The currently known systems are basically a nuclease domain fused (covalently or via non-covalent polypeptide interactions) to domains that exhibit DNA binding and regulatory functions. Typically, they recognize modified DNA substrates with little sequence specificity. To date, the consensus view is that these modification-dependent systems have evolved to protect the host against attacking parasites with foreign methylation patterns and as suicidal defence strategy against the expression of foreign MTases.[305] In addition, self-cleavage of the chromosome has been observed when cells are exposed to stressful conditions, which suggests that these systems also play a role in the bacterial stress physiology.[306]

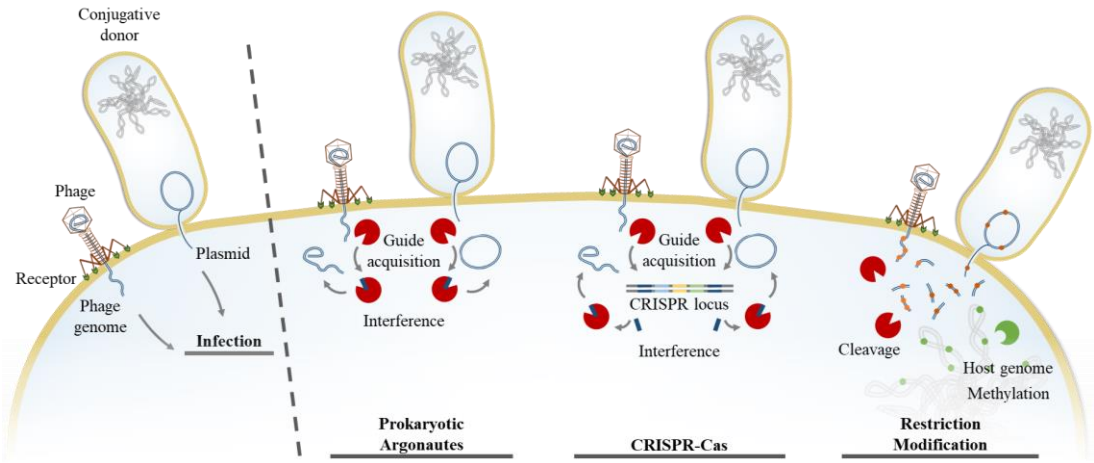


Figure 3.3 Overview of bacterial defence mechanisms that target foreign DNA. From left to right: transfer and maintenance of invading DNA halted by prokaryotic Argonautes, CRISPR-Cas systems and RM-systems. Invading DNA (phage or plasmid): dark blue; Bacterial chromosome: grey; Phage receptors: dark green; Enzyme (complex) with nuclease activity: red; Enzyme (complex) with methyltransferase activity: green; "Non-self" DNA modification pattern: orange and red dots; "Self" modification pattern: green dots.[254]

3.1.7 Strategies to overcome restriction barriers

The genetic manipulation of bacteria is a powerful approach for elucidating fundamental aspects of bacterial behaviour and for harnessing bacterial capabilities for biotechnology applications. However, the full power of the quickly expanding genetic toolkit can only be applied to a few 'domesticated' bacterial strains that readily accept exogenous genetic material.[307] At present, there are still many organisms with potentially interesting characteristics that are genetically inaccessible, and hence cannot be genetically engineered, due to the many barriers to DNA/RNA transfer described previously. Overcoming these barriers is challenging, mostly because of the large biological diversity and strain-level variations in defence mechanisms, even within species.[291, 308] As a result, projects with a new bacterial isolate usually start with the time consuming task of establishing efficient and reliable DNA transfer. This step can be a true bottleneck, sometimes delaying projects by several years.

Initial attempts to establish and increase DNA uptake in an 'untouched' bacterial strain are normally based on previously developed methods for closely related strains. Optimization of these artificial (or natural) transformation and conjugation protocols generally involves testing different media compositions, growth phases, incubation times, electroporation settings and other, less conventional, alterations[307, 309]; for example, a heat shock treatment prior to conjugation can significantly increase DNA transfer efficiency in specific species or strains, including *C. difficile*.[310-312] Nevertheless, the cumulative effect of tweaking all these different parameters is often insufficient to establish successful DNA uptake in strains with a variety of efficient nucleic acid targeting systems. From the three systems described in section 3.1.6, especially RM systems hinder DNA transfer and are the most common underlying cause of genetic inaccessibility.

Most of the currently available strategies to overcome RM systems have been inspired by the anti-restriction mechanisms employed by bacteriophages to avoid degradation. The three most common strategies are: (i) subversion of RM systems; (ii) recognition site avoidance, and; (iii) methylation mimicry.[206, 291, 313-314]

Phages have evolved several mechanisms to subvert RM systems, such as blocking recognition sequences by encoding DNA binding proteins, destroying cofactors essential for REase activity and directly interacting with RM systems to prevent for example the assembly of type I systems.[291] All these strategies essentially aim to inactivate the host REases. Researchers have followed a similar approach by deleting or mutating cell REases to increase DNA transfer efficiencies.[315-321] Although the inactivation of host RM systems via directed engineering can be very effective in some cases, this approach is unsuitable for a wide range of bacteria because of the need for an initial, albeit low, level of DNA uptake.

Recognition site avoidance is very common amongst phage genomes and conjugative elements as it is a simple but effective strategy to prevent cleavage by host RM systems: no recognition sites, no DNA restriction.[322-323] Additionally, it has been demonstrated that even a reduction in the number of recognition sites can increase the probability of overcoming this type of host defence.[289, 324] However, a recent study by Rusinov *et al*. suggests that although the avoidance strategy is widespread, it is not universal; this phenomenon is predominantly associated with type II systems and non-temperate phages with a DNA genome.[322] Nevertheless, it has been shown that removing RM system recognition sequences of any type from DNA templates is an effective approach to circumvent genetic inaccessibility.[289, 313] Unfortunately, it is not always possible to remove a recognition site; for example, a site within an origin of replication cannot easily be altered without affecting its function.

Methylation mimicry by bacteriophages involves matching the phage genome modifications to the methylation pattern of the host, so that upon infection the host cannot distinguish the exogenous DNA from endogenous DNA to prevent hijacking of the cell. This is also a very effective and common approach taken by researchers who aim to make the inaccessible accessible.[325] There are two frequently used methylation mimicry methods: (i) the *in vitro* methylation of DNA fragments using commercially available enzymes that target the site of interest, or; (ii) *in vivo* methylation of plasmids by passing the DNA through a surrogate strain which has been engineered to match the methylation pattern of the strain of interest.[326-329] However, these two strategies also have their limitations. For many REases identified in genetically inaccessible strains there is no commercial enzyme available that modifies the specific recognition sequence due to the high diversity of RM systems. If this is the case, *in vivo* methylation could offer a solution, although expression of

the host's cognate MTases can be challenging due to toxicity effects to the surrogate strain. Placing the MTases downstream of an inducible or weak promoter can mitigate toxicity effects without compromising plasmid methylation, though finding the right balance between MTase toxicity and conferred protection can be tricky.

Generally, when a strain of interest is genetically inaccessible, the full genome is sequenced using SMRT PacBio sequencing technology and RM systems with their recognition sequences are identified.[292] Next, several of the above strategies are attempted to establish efficient DNA transfer, which is time consuming and often results in a method that is specific for just one strain; the whole cycle needs to be repeated for the next strain. Hence, there is a desire for more widely applicable and systematic approaches to circumvent RM systems. For example, a roadmap which dictates subsequent steps depending on previous findings could speed up the whole process and increase the success rate.

3.2 Objectives

Establish and increase the efficiency of conjugative plasmid transfer from *E. coli* into *C. botulinum* type A strains via three different methods:

i. Inactivation of the recipient's REases,

ii. Expression of the recipient's MTases in the *E. coli* donor,

iii. Testing various conjugal *E. coli* donor strains with different native methylation profiles.

3.3 Results and discussion

3.3.1 Identification of restriction modification systems

Reliable and efficient DNA transfer is crucial for achieving the overall aim of this work, generating a library of *C. botulinum* safe strains. Based upon literature research the 62A strain should accept foreign DNA, however for the ATCC 19397, Loch Maree and NCTC 2916 strain this information could not be found.[126] Therefore, as a starting point the amenability of each strain to accepting foreign DNA was assessed following the standard conjugation protocol developed for ATCC 3502, as described in section 2.9.3. To calculate the conjugation efficiency the total number of transconjugants per mating was divided by the total number of *E. coli* donor added to the conjugal mixture. Additionally, the recipient CFU was enumerated at the end of each mating, in order to rule out that any potential differences observed in DNA transfer efficiencies resulted from variations in the total number of recipients per conjugal mixture.

It should be noted that the above method to determine the plasmid-transfer efficiency is imperfect because it does not discriminate between an increased number of transfer events and post-transfer division of the recipient cell. Ideally, the transfer efficiency is calculated based on the number of conjugation events per donor-recipient interaction.[249] To obtain data in this form single-cell analysis would be required, however the necessary equipment to study single cells was not available in the *C. botulinum* laboratory. Additionally, single-cell analysis is much more laborious. Moreover, the different donors and plasmids used are unlikely to have significantly different impacts on the post-transfer growth of recipient cells. Therefore, it was concluded that discrimination between the number of transfer events and growth of

the recipient population was not essential for this study and the simple, commonly used, transconjugant to donor ratio would be sufficient.

Surprisingly, although the strains are quite closely related, the conjugation efficiencies achieved under these conditions varied greatly (Figure 3.4). Strains ATCC 3502 and 62A were most susceptible to plasmid pMTL83151, yielding approximately 70000 and 30000 transconjugants per conjugal mixture, respectively. The NCTC 2916 strain was the next most efficient in terms of transformation, with on average 13000 transconjugants per mating plate, followed by Loch Maree with a consistent but low number of approximately 500 transconjugants per conjugal mixture. Despite the genomic similarity of strain ATCC 19397 to both ATCC 3502 and 62A, very few colonies appeared on the selection plates (on average 1.5 colony per conjugal mixture). Several of the putative transconjugants were subjected to a PCR screen using primers specific to the transferred plasmid (M13R and 83XXX_LR). Not a single colony screened came back positive, suggesting that probably the majority, if not all, of the CFUs were false positives.

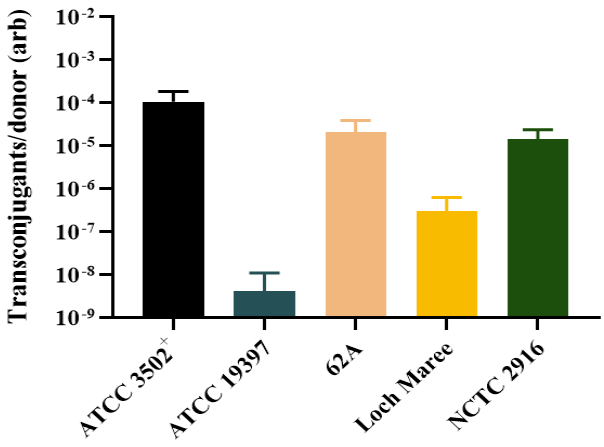


Figure 3.4 Conjugation efficiency into ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916. Transfer efficiency of plasmid pMTL83151 from conjugal donor CA434 into different WT *C. botulinum* recipient strains. Error bars represent the standard deviation of three or more independent experiments.

For the ATCC 3502, 62A and NCTC 2916 strains, the frequency of transfer obtained under the current conditions should be sufficient to introduce the requisite gene tools to engineer their genome. Only for suicide plasmid based methods or high-throughput studies like TraDIS higher transfer efficiencies would be required.[330] However, the relatively low conjugation efficiency of the Loch Maree strain might prove to be problematic when conjugating unstable plasmids (for ACE) or plasmids with toxic cargo (e.g., Cas9, due to its lethality caused by accidental off-target cuts).

Host specific RM systems are a prominent barrier to DNA transfer, as described in section 3.1.6, and could provide an explanation for the observed variation in conjugation efficiency amongst the strains. To identify putative RM systems in each strain the REBASE database, a restriction enzyme database, was consulted and the predicted methylation, restriction and identification genes are highlighted in Figure 3.5.[308] Only the RM systems with a restriction component were considered relevant as potential barriers to successful conjugation. Strains ATCC 3502, ATCC 19397, 62A and Loch Maree are predicted to encode one type I system. The REBASE profile of ATCC 19397 also includes one complete type II system. In strains 62A, Loch Maree and NCTC 2916 a type IV system has been detected. Specific to Loch Maree is a gene predicted to encode a methyltransferase restriction-endonuclease fusion.

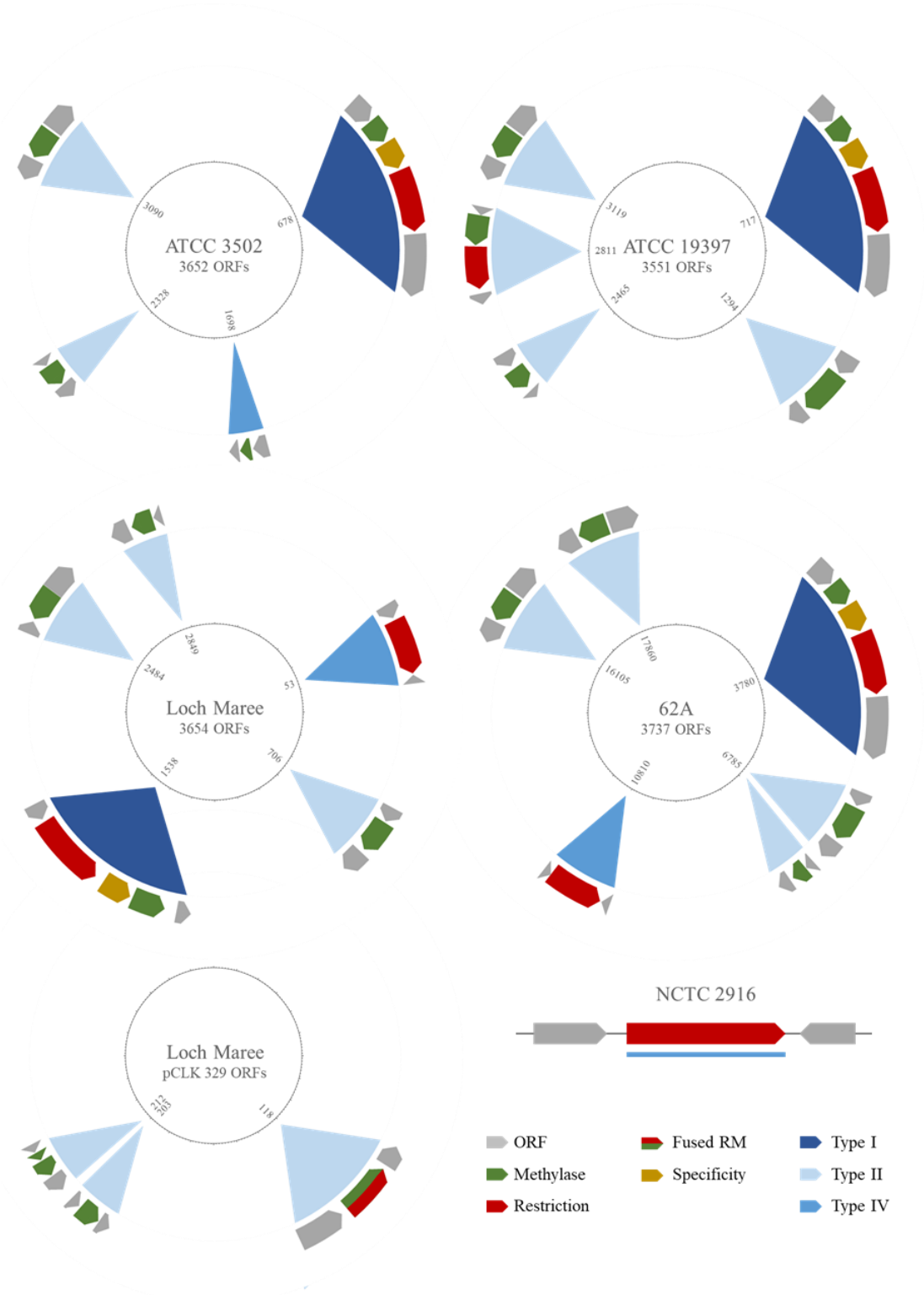


Figure 3.5 Restriction modification systems of ATCC 3502, ATCC 19397, 62A, Loch Maree and NCTC 2916. Schematic representation of *C. botulinum* genomes with predicted RM systems highlighted. Due to the assembly level of the NCTC 2916 genome (49 contigs) the schematic representation of its genome is linear instead of circular. Data was obtained from the New England Biolabs REBASE database.[308]

To further compare the RM system profiles of the strains, sequences were subjected to BLASTn/p searches and alignments using SnapGene. The type I systems in ATCC 3502, ATCC 19397 and 62A were 100% homologous, with identical recognition sequences predicted for ATCC 1937 and 62A (Table 3.1). It is therefore safe to assume that CTANNNNNNRTGAA is also the recognition sequence in ATCC 3502. This motif is not present in plasmid pMTL83151, which was used to establish the DNA transfer efficiency into all strains, and hence should not have been a barrier to plasmid transfer.

In the genome sequence of NCTC 2916 two type I RM systems were identified, which were not predicted by REBASE. The first system, comprised of genes CBN_0725, CBN_0726 and CBN_0728, is most similar to the type I genes found in ATCC 3502, ATCC 19397 and 62A. The other system, consisting of CBN_2138, CBN_2137 and CNB_2139, is more homologues to the type I system in Loch Maree. In both systems the specificity subunit has a low % of identity to sequences of the other strains, 35% and 38%, respectively, which could indicate that they recognise different sequence motifs. Currently the methylome of NCTC 2916 and Loch Maree is unknown. To determine the type I recognition sequences the methylation patterns of both strains will have to be identified using PacBio SMRT sequencing.

The type II system was unique to ATCC 19397 and its putative recognition sequence GCWGC, predicted by the REBASE database, is present nine times in pMTL83151 (Table 3.1). Therefore, it was assumed that this system is the strain's main barrier to DNA transfer, resulting in (almost) no transconjugants.

In both ATCC 3502 and ATCC 19397 a protein 100% homologous to the predicted type IV in 62A was identified. This eliminated the possibility that the reduced conjugation efficiency of 62A compared to ATCC 3502 was due to its type IV system.

It should be noted that the genome of 62A was published in the final year of this study.[331] Therefore, this system was not taken into consideration when designing and performing the experiments in sections 3.3.2, 3.3.3 and 3.3.4. At the time, the genome sequence of 62A (NCTC 3805) was determined using Illumina MiSeq sequencing. The short Illumina reads were assembled into 99 contigs using the CLC genomic workbench and annotated automatically with the RAST server.[332] Subsequently, the putative RM system genes detected by REBASE in ATCC 3502 and ATCC 19397 were aligned to the 62A contigs. As a result of this approach only the type I system was identified but no unknown systems were detected.

Table 3.1 Type I and type II restriction modification systems and associated methylation patterns detected by REBASE

Strain	Gene(s)	Type	Recognition sequence
ATCC 3502	hsdM, hsdS, hsdR	I	unknown
ATCC 19397	CLB_RS03505, CLB_RS03510, CLB_RS03525	I	CTANNNNNNRTGAA
	CLB_RS13900, CLB_RS13895	II	GCWGC
62A CFSAN064329	DBN47_03780, DBN47_03785, DBN47_03800	I	CTANNNNNNRTGAA
Loch Maree	CLK_1539, CLK_1538, CLK_1537	I	Unknown
	CLK_A0118	II	Unknown

Although the ATCC 19397 type I recognition sequence was absent from pMTL83151, the type I system is a restriction threat to the pRECas1 plasmid, carrying the most efficient clostridial engineering tool available to date, with two restriction sites present in the *Cas9* sequence. Additionally, motif CTANNNNNNRTGAA is present in the left homology arm of the editing template to delete the *boNT/A1* gene in ATCC 3502, ATCC 19397 and 62A.

To establish if the predicted type I recognition sequence is correct, the sequence motifs CTAAATACGATGAA (R1) and CTAGTCATTATGAA (R2), present in the *Cas9* gene, were cloned into plasmid pMTL83151, as described in section 2.11.1.1. The resulting plasmids pMTL83151-R1, pMTL83151-R2 and pMTL83151-R1+R2 were conjugated into ATCC 3502 via donor strain CA434, as detailed in section 2.9.3. Utilisation of ATCC 3502 for this assay was preferred over ATCC 19397 or 62A because of the relatively high transfer efficiencies obtained with pMTL83151 (at the time the genome sequence of 62A was yet to be determined). It was hypothesised that if the type I system is active and targets the CTANNNNNNRTGAA motif, the presence of R1/R2 would have a negative impact on the conjugation efficiency.

Insertion of R1 and/or R2 into plasmid pMTL83151 resulted in a significant reduction of the DNA transfer efficiency, relative to pMTL83151 (Figure 3.6). The total number of transconjugants was reduced by approximately 100-fold in the presence of either R1 or R2 and both together. Still, on average 280 transconjugants were obtained per conjugal mixture; in these cells the plasmid was able to escape the type I system. One explanation for these 'escapers' could be that mutations occurred either in the recognition sequence or in the recipient's type I system, but given the number of transconjugants, which is higher than one would normally expect based on normal mutation rates (see Chapter 5), this seems very unlikely. The survival of these plasmids

in some cells is more likely due to variations in the type I subunit concentrations between cells; in cells with a low concentration of the restriction subunit relative to the concentration of methyltransferase, accidental methylation of the recognition site could occur before digestion.

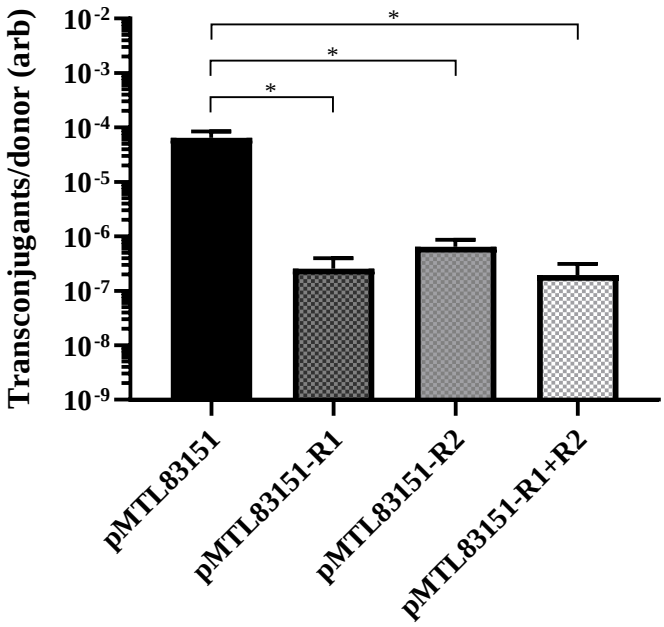


Figure 3.6 Conjugation efficiency of various plasmids containing a type I RM system recognition site into ATCC 3502^{+}. Transfer efficiency of plasmids pMTL83151, pMTL83151-R1, pMTL83151-R2 and pMTL83151-R1+R2 from conjugal donor CA434 into recipient strain ATCC 3502^{+}. Error bars represent the standard deviation of three independent experiments. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

Surprisingly, the addition of a second recognition site did not increase the effectiveness of the system, contrary to what has been observed for type II systems; additional restriction recognition sites on plasmids reduced the frequency of transfer.[289, 323] Possibly, this unexpected difference can be attributed to a difference in how the two systems interact with and cleave DNA (see 3.1.6). Type I systems reel in plasmid DNA at a rate of 1 kb per second and when it gets to the "end" it stalls, effectively destroying the plasmid or it just carries on looping the DNA. This looping

could make the DNA inaccessible to replicative machinery, but could also prevent binding of another type I complex to the second recognition site. Hence, the addition of a second recognition site did not result in a further decrease of the conjugation efficiency.

To confirm that the plasmids are cleaved at the CTANNNNNNRTGAA site, pure plasmid extract from a standard *E. coli* donor (using the plasmid with two recognition sequences) could be incubated with *C. botulinum* cell lysate. Subsequently, the reaction mixtures can be analysed by agarose gel electrophoresis to visualise plasmid digestion. If REases in the lysate cleave at the CTANNNNNNRTGAA site a specific band pattern is expected to show.

The drop in conjugation efficiency in the presence of recognition sequence CTANNNNNNRTGAA could prove troublesome when trying to engineer the strain with a *Cas9* based tool, especially when its expression is constitutive, due to off-target effects.

3.3.2 Engineering the recipient: inactivation of type I nuclease in ATCC 3502

As previously mentioned, the type I complex consists of a restriction, modification and specificity subunits, encoded by the *hsdR*, *hsdM* and *hsdS* genes, respectively. If the type I system targets the predicted recognition sequence inactivation of *hsdR* should abolish cleavage at the CTANNNNNNRTGAA site and thus result in a strain with improved properties for genome editing purposes.

To inactivate the HsdR subunit in ATCC 3502 ClosTron mutagenesis was carried out as described in section 2.12.1, using plasmid pMTL007C-E2-*hsdR*-1600s. The intron insertion was confirmed by PCR in three biological replicates from three independent conjugations using primers annealing to the chromosome, flanking the insertion site

(Figure 3.7). The plasmid was lost from the selected ATCC 3502 *hsdR*1600s::CT mutants, as described in section 2.8, and presence of the intron in the final mutants was subsequently verified by Sanger sequencing. To confirm the intron inserted only at the desired target gene the entire genome was sequenced. No additional intron insertions were detected.

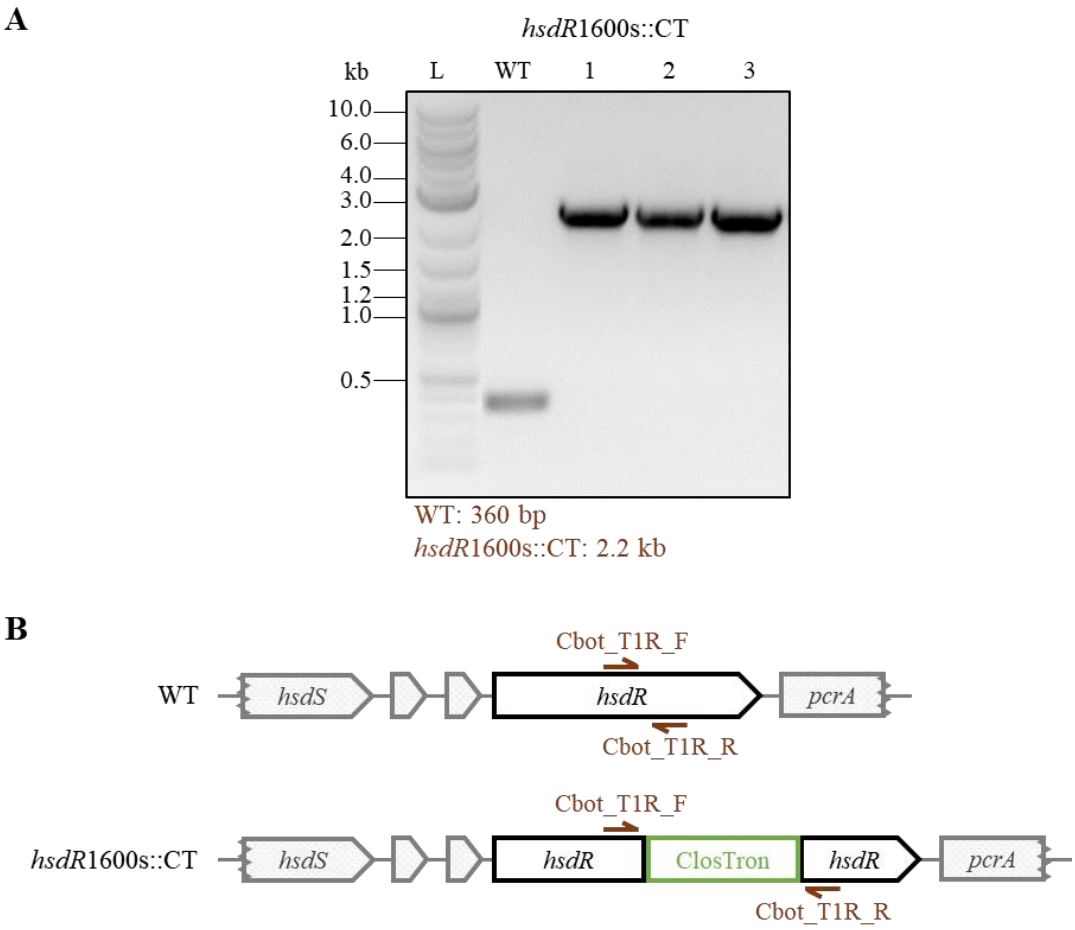


Figure 3.7 ClosTron-mediated inactivation of *hsdR* in strain ATCC 3502. (**A**) PCR for three biological replicates of ATCC 3502 *hsdR*1600s::CT mutants alongside WT. Products were amplified using flanking primers Cbot_T1R_F and Cbot_T1R_R. Lane L: NEB 1 kb Plus DNA Ladder. (**B**) Schematic diagram of the targeted chromosomal region showing the primer binding sites in the WT and mutant strains.

The conjugation efficiency of plasmid pMTL83151 with and without the type one recognition site(s) into the *hsdR*1600s::CT mutants was determined as described in section 2.9.3 (Figure 3.8). As expected, the transfer efficiency of pMTL83151 into wildtype (WT) and *hsdR*1600s::CT was similar, whereas there was a clear difference in efficiency using the plasmids containing R1 and/or R2. Inactivation of the restriction subunit resulted in a strain readily accepting plasmid DNA with the recognition sequence, implying that the putative type I system is responsible for cleavage at the CTANNNNNNRTGAA site. Inactivation of the HsdR subunit increased the conjugation efficiency of plasmids with the recognition sequence by approximately 100-fold compared to WT. Nevertheless, although the type I system in *C. botulinum* ATCC 3502 compromised DNA uptake, it is not a major barrier to DNA transfer, which is in line with outcomes of former studies.[293] This work was not repeated in strains ATCC 19397 and 62A because the type I system is identical across all three strains and no differences are anticipated.

Despite the fact that inactivating *hsdR* in ATCC 3502 was a success, this approach was not pursued in the other strains because of its limited potential. An initial, albeit low, level of DNA transfer is required, therefore this strategy is unsuitable for strains like ATCC 19397. In addition, the conjugation efficiency is improved to facilitate future projects and ideally the strain of interest is modified as little as possible. Methods resulting in a clean deletion of *hsdR* which require a minimal number of single colony stages to isolate mutants would be preferred over ClosTron mutagenesis to avoid polar effects and minimise the accumulation of random mutations.

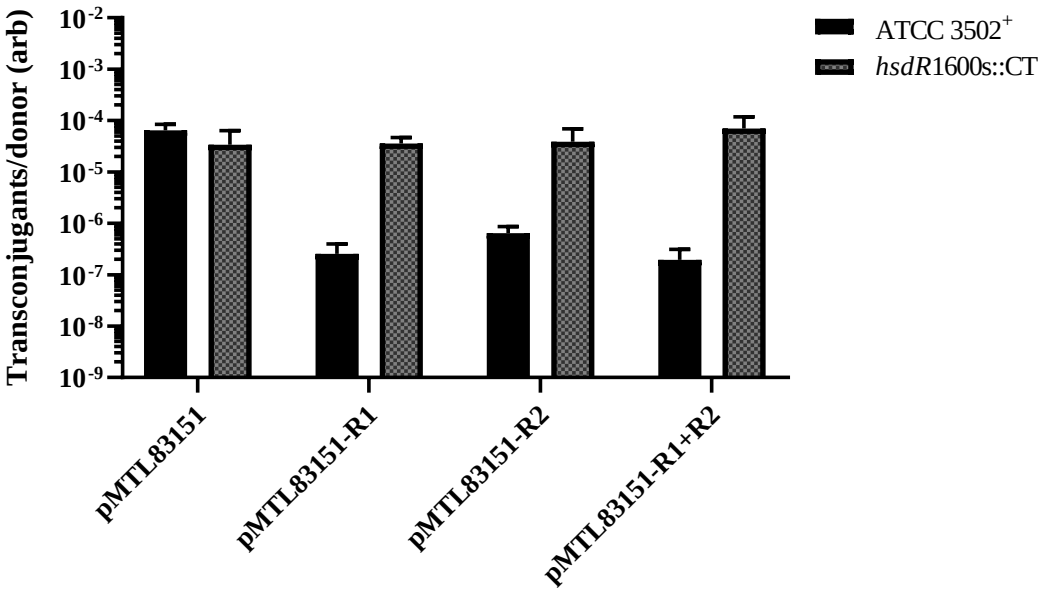


Figure 3.8 Conjugation efficiency of various plasmids containing a type I RM system recognition site into *hsdR*1600s::CT. Transfer efficiency of plasmids pMTL83151, pMTL83151-R1, pMTL83151-R2 and pMTL83151-R1+R2 from conjugal donor CA434 into recipient strains *hsdR*1600s::CT and ATCC 3502^{+} (data from Figure 3.6). Error bars represent the standard deviation of three biological replicates for *hsdR*1600s::CT from two independent experiments.

3.3.3 Engineering the donor 1: customised methylation of donor DNA

The second strategy to protect transferred plasmid DNA from restriction is based on the principle that if 'new' DNA has the same methylation pattern as the host's genome, it will not be recognised as foreign by its RM systems. To achieve this, the donor strain was engineered to express the MTases associated with the type I and type II RM systems in ATCC 19397 (ATCC 3502 and 62A).

Initially, the effect of expressing the MTases in *E. coli* on the conjugation efficiency was assed using five plasmid-based configurations with: a) the weak constitutive promoter P_{J23105} upstream of *hsdM-hsdS* (type I), RS13900 (type II) and both together and b) the inducible P_{BAD} promoter upstream of *hsdM-hsdS* and RS13900. The

MTases were placed under the control of a weak constitutive or inducible promoter to minimize the potential toxic effects to *E. coli*, often associated with the expression of non-native MTases. Additionally, adequate methylation can be attained with low levels of expression (R. Mansfield, personal communication).

The plasmids, whose construction is outlined in section 2.11.1.2, were transformed into the *E. coli* donor CA434. Conjugation efficiency assays were performed as described in section 2.9.3, using pMTL83151 with and without the type I recognition site(s). The *E. coli* cells harbouring plasmid pM-BTI or pM-BTII were grown overnight in the presence of 100 mM arabinose to induce expression of the MTase.

The transfer efficiencies of pMTL83151 into ATCC 3502 from the various donors did not differ significantly (Figure 3.9A). However, although not statistically significant, conjugation efficiencies were consistently slightly lower in the presence of the type II MTase. This suggests that methylation at the putative GCWGC recognition site, which is not part of its methylome, occasionally triggers plasmid degradation.

Conjugation efficiencies of plasmids with the type I recognition site significantly improved if *hsdM* and *hsdS* were expressed in the donor strain (Figure 3.9B). The correctly methylated plasmids transferred with the same efficiency as pMTL83151, suggesting that expression of these two proteins provides full protection against degradation by the type I RM system. Presence of both the type I and type II MTases in the donor strain again resulted in slightly lower plasmid transfer efficiencies compared to donor cells harbouring only the type I genes (Figure 3.9C).

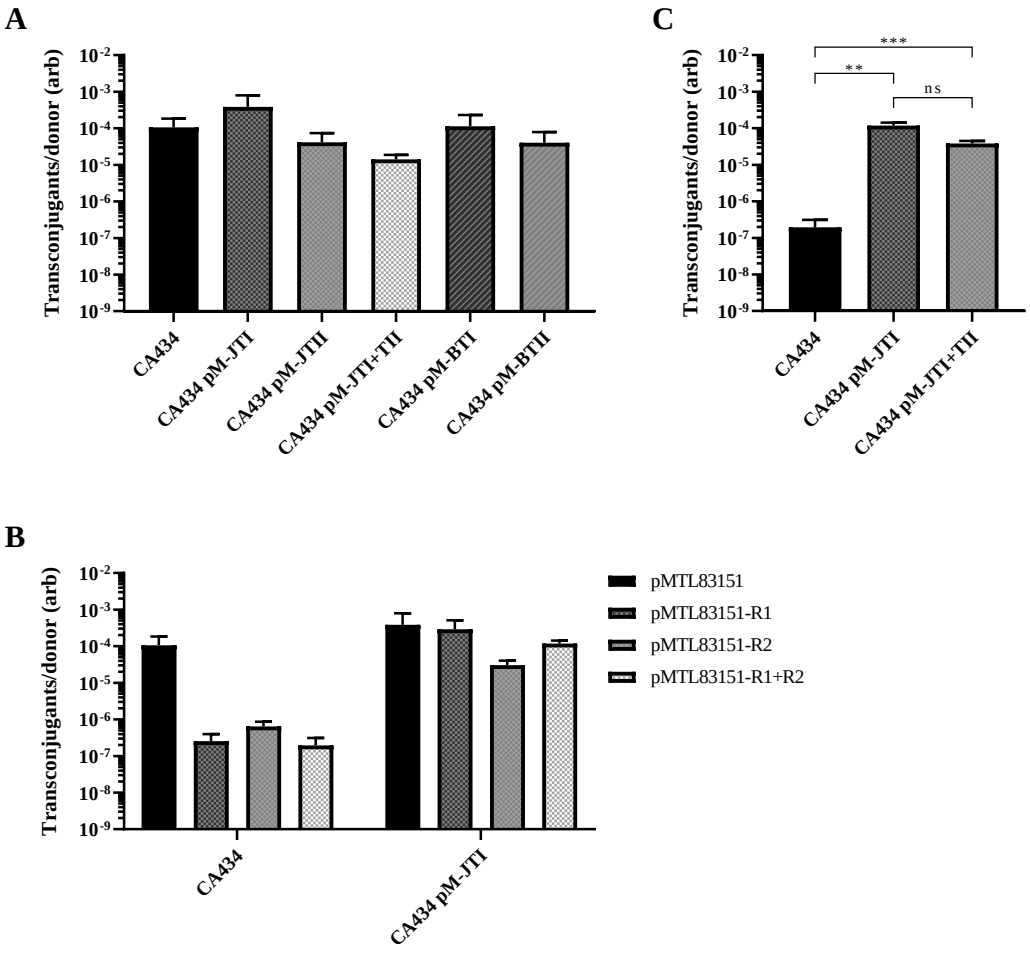


Figure 3.9 Comparison of conjugation efficiency into ATCC 3502^{+} using *E. coli* donors expressing different methyltransferases. (**A**) Transfer efficiency of plasmid pMTL83151 from CA434 conjugal donors carrying plasmids with different methyltransferase expression systems. (**B**) Transfer efficiency of plasmids pMTL83151, pMTL83151-R1, pMTL83151-R2 and pMTL83151-R1+R2 from conjugal donor CA434 carrying pM-JTI in comparison to transfer efficiencies obtained using CA434. (**C**) Transfer efficiency of plasmid pMTL83151_R1+R2 from conjugal donors CA434, CA434 pM-JT1 and CA434 pM-JTI+TII. Error bars represent the standard deviation of three independent conjugations. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

For ATCC 19397, the transfer efficiency of pMTL83151 increased by 10000-fold from almost zero to more than 15000 transconjugants per conjugal mixture in the presence of the type II MTase (Figure 3.10A). The achieved conjugation efficiency now equals the maximum level reached with ATCC 3502, which lacks the type II system, indicating that expression of the RS13900 MTase is a very effective way to circumvent the host's type II RM system. No differences were observed in the conjugation efficiency of pMTL83151 with RS13900 under the control of P_{J23105} *vs* P_{BAD}. As expected, expression of *hsdM* and *hsdS* had no positive effect on the conjugation efficiency because pMTL83151 is not recognised by this type I system.

Conjugation efficiency assays with pMTL83151-R1+R2 from the various donors into ATCC 19397 clearly showed the effect of both RM systems on plasmid transfer. (Figure 3.10B). Expression of the type II MTase in the donor is essential for successful DNA transfer and additional expression of the type I methylation and specificity subunits further increases the transfer efficiency.

To confirm that the MTases prevent cleavage by methylating the plasmid DNA, pure plasmid extract from a standard *E. coli* donor and pure plasmid extract from the engineered donors could be incubated with cell lysate. Subsequently, the reaction mixtures can be analysed by agarose gel electrophoresis to visualise plasmid digestion. If the MTases protect against cleavage at the identified recognition sequences there should be a clear difference in gel profiles.

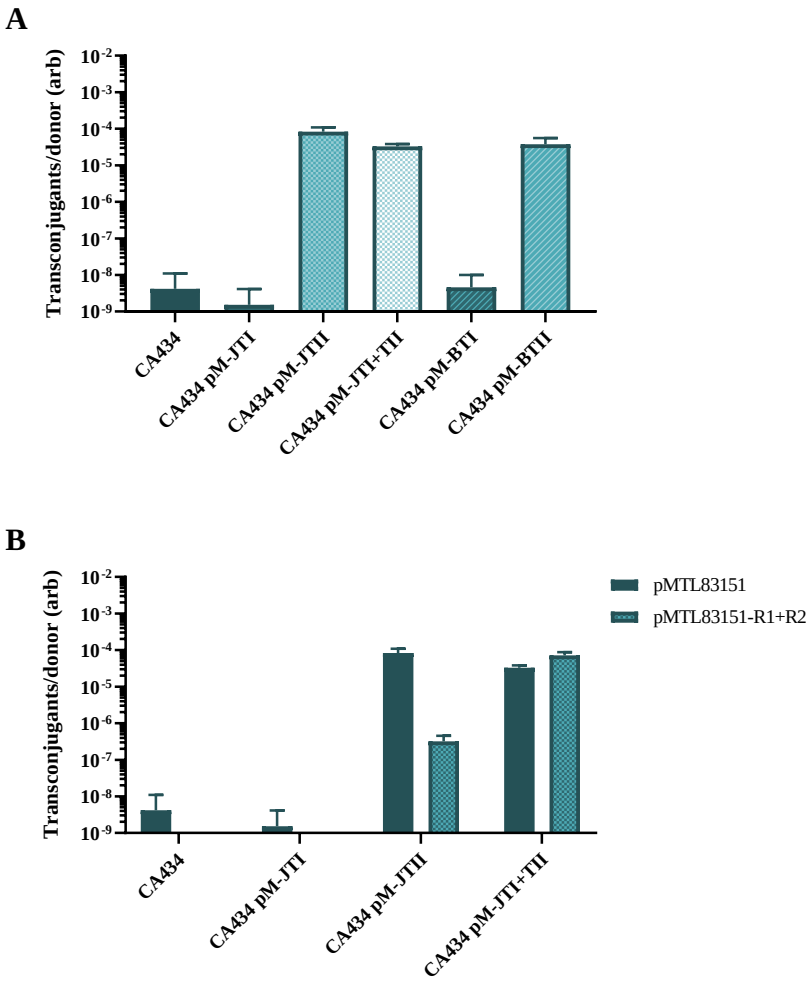


Figure 3.10 Comparison of conjugation efficiency into ATCC 19397 using *E. coli* donors expressing different methyltransferases. (**A**) Transfer efficiency of plasmid pMTL83151 from CA434 conjugal donors carrying plasmids with different methyltransferase expression systems. (**B**) Transfer efficiency of plasmid pMTL83151-R1+R2 from conjugal donor CA434 carrying no plasmid, pM-JTI, pM-JTII or pM-JTI+TII in comparison to transfer efficiencies obtained with pMTL83151. Error bars represent the standard deviation of three independent conjugations.

3.3.4 Engineering the donor 2: creating an all in one package by recombineering the methyltransferases into conjugal plasmid R702

In the conjugal donors optimised for DNA transfer into ATCC 3502 and ATCC 19397 the MTases are expressed from a plasmid, which is suboptimal; to maintain this plasmid, the plasmid to be transferred and the conjugal R factor R702 three different antibiotics have to be added to the growth medium. To reduce the number of plasmids in the donor strain a strategy was developed in the Clostridial Research Group whereby the MTases are expressed from R702 (R. Mansfield, unpublished data), resulting in a stable 'personalised' donor strain. Briefly, the relevant genes to circumvent to host's RM system(s) are identified and inserted into the *aadA1* spectinomycin resistance locus, located on R702, under the control of the arabinose inducible promoter P_{BAD}, using the lambda red recombineering system.[333]

To create the recombinant donor strains a recombineering cassette was constructed, as described in section 2.11.1.3. This cassette carried the RS13900, *hsdM* and *hsdS* genes combined in one operon, so that the resulting donor strain could be used for ATCC 3502, ATCC 19397 and 62A. Previous experiments showed that expression of RS13900 had no significant adverse effect on the transfer efficiency of plasmids into ATCC 3502 (section 3.3.3). The P_{BAD} promoter was placed upstream of the operon for consistency reasons within the group. Initially it was attempted to transform the recombineering cassette into RE-CA434 (harbouring the pKD46 lambda red recombineering plasmid), as detailed in section 2.11.1.3. Unfortunately, no recombinants were obtained in CA434 for unknown reasons.

Therefore, the RE-InterStellar strain was selected as an alternative based on previous work by other group members which showed successful insertion of recombineering cassettes into the R702 plasmid using RE-InterStellar as a host. This *E. coli* strain was

created by transforming the conjugal R factor R702 and plasmid pKD46 into Stellar Competent Cells (Clontech; J. Millard, unpublished data). There are numerous genetic differences between InterStellar and CA434, including their methylation profiles; CA434 contains *dam* and *dcm* genes (dam^+/dcm^+), encoding adenosine and cytosine MTases, respectively, whereas InterStellar is dam^-/dcm^-.[334] It has been shown that the methylation status of the donor can affect DNA transfer efficiencies into various bacterial species, including clostridia.[207, 220, 335-338] The importance of Dam/Dcm methylation depends on the recipient's type IV restriction systems, its Dam/Dcm methylation status and its ability to replicate hemimethylated DNA. At this stage the existence of the type IV REase in ATCC 3502 and ATCC 19397 was still unknown but no homologues to the Dam and Dcm MTases in *E. coli* were detected. Hence, no or a positive effect on the conjugation efficiency was anticipated.

It is worth nothing that one downside of using dam^- strains is their increased mutation rate due to the disruption of post-replicative mismatch repair caused by the absence of adenine-methylation.[339] This could be a problem, especially when the plasmid's cargo has a metabolic burden or is (slightly) toxic to the *E. coli* donor (e.g., Cas9 and MTases). Cells harbouring plasmids with accidental random detrimental mutations in the cargo potentially have a growth advantage and could become a substantial part of the population used for conjugations.

The recombineering cassette was transformed into RE-InterStellar and correct recombinants were confirmed by PCR in three biological replicates using primers RE_F and RE_R annealing to the cassette and the R702 plasmid, respectively. The plasmid was lost from the selected recombinants, as described in section 2.8, and presence of the cassette in the final InterStellar R702TI+TII mutants was subsequently verified by Sanger sequencing.

Conjugation efficiency assays were performed as described in section 2.9.3, alongside donor strains Stellar, InterStellar and InterStellar R702apr as controls. In InterStellar R702apr the empty recombineering cassette is inserted into the *aadA1* gene of the R factor. For ATCC 3502 the DNA transfer efficiencies achieved with InterStellar and InterStellar R702apr did not differ significantly compared to CA434, suggesting that Dam/Dcm methylation of the shuttle plasmid is not a major barrier to transfer. In line with previous results, the plasmid with the type I recognition sequences was transferred less efficiently (Figure 3.11A). As expected, no ATCC 19397 transconjugants were obtained using InterStellar and InterStellar R702apr as a donor (Figure 3.12A). Additionally, conjugal mixtures with Stellar yielded zero transconjugants for both strains; Stellar lacks the required functional elements for conjugation.

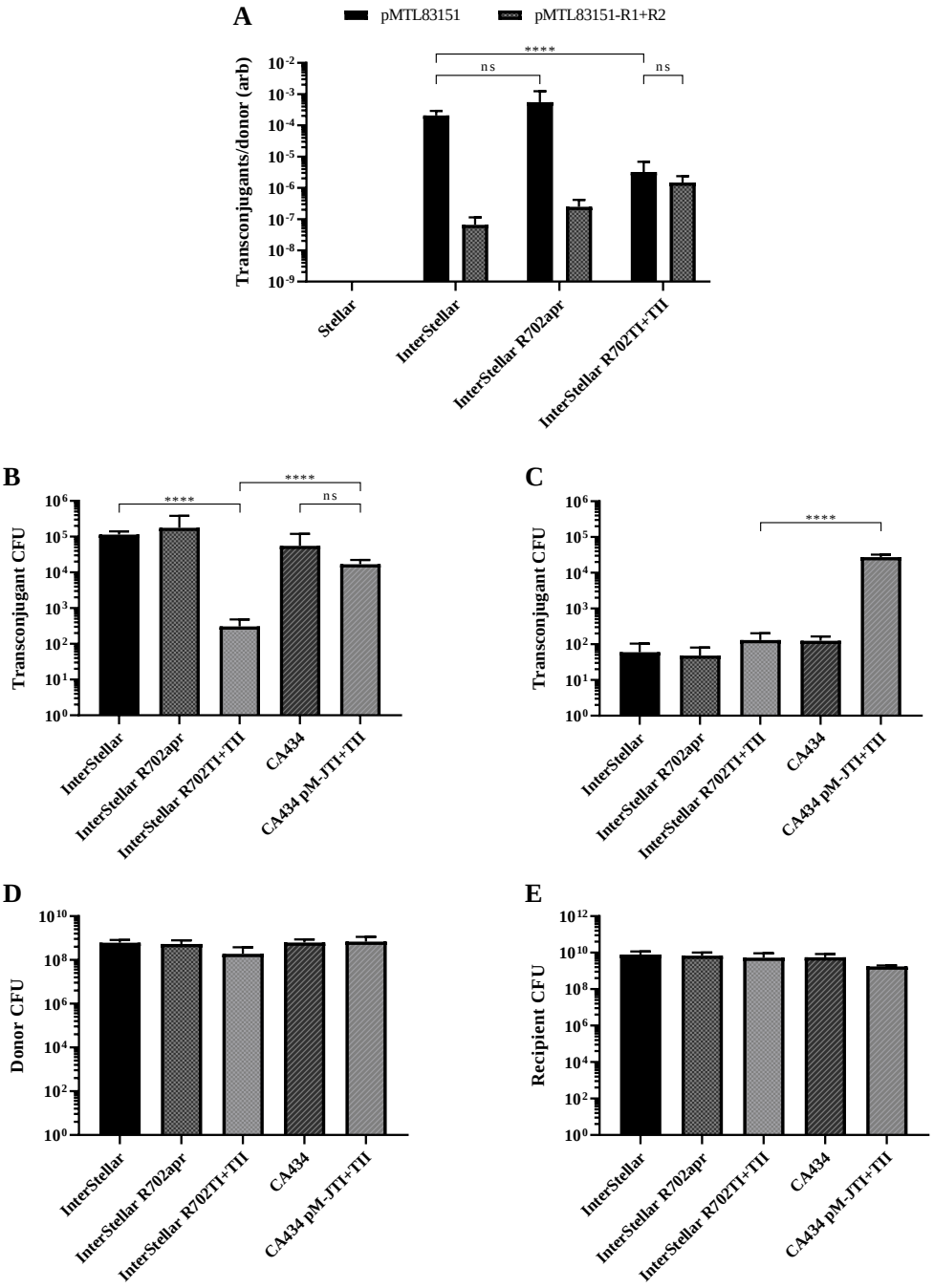


Figure 3.11 Comparison of conjugal donors InterStellar and CA434 with and without methyltransferases in combination with ATCC 3502^{+}. (**A**) Conjugation efficiency of plasmids pMTL83151 and pMTL83151-R1+R2 from conjugal donors Stellar, InterStellar, InterStellar R702apr and InterStellar R702TI+TII into ATCC 3502^{+}. (**B**) Total number of transconjugants carrying pMTL83151 per conjugal mixture. (**C**) Total number of transconjugants carrying pMTL83151-R1+R2 per conjugal mixture. (**D**) Conjugal donor CFUs carrying pMTL83151 at the start of the conjugation. (**E**) Recipient CFUs after harvesting the conjugal mixture (with the donor strains carrying pMTL83151). Error bars represent the standard deviation of three independent conjugations. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

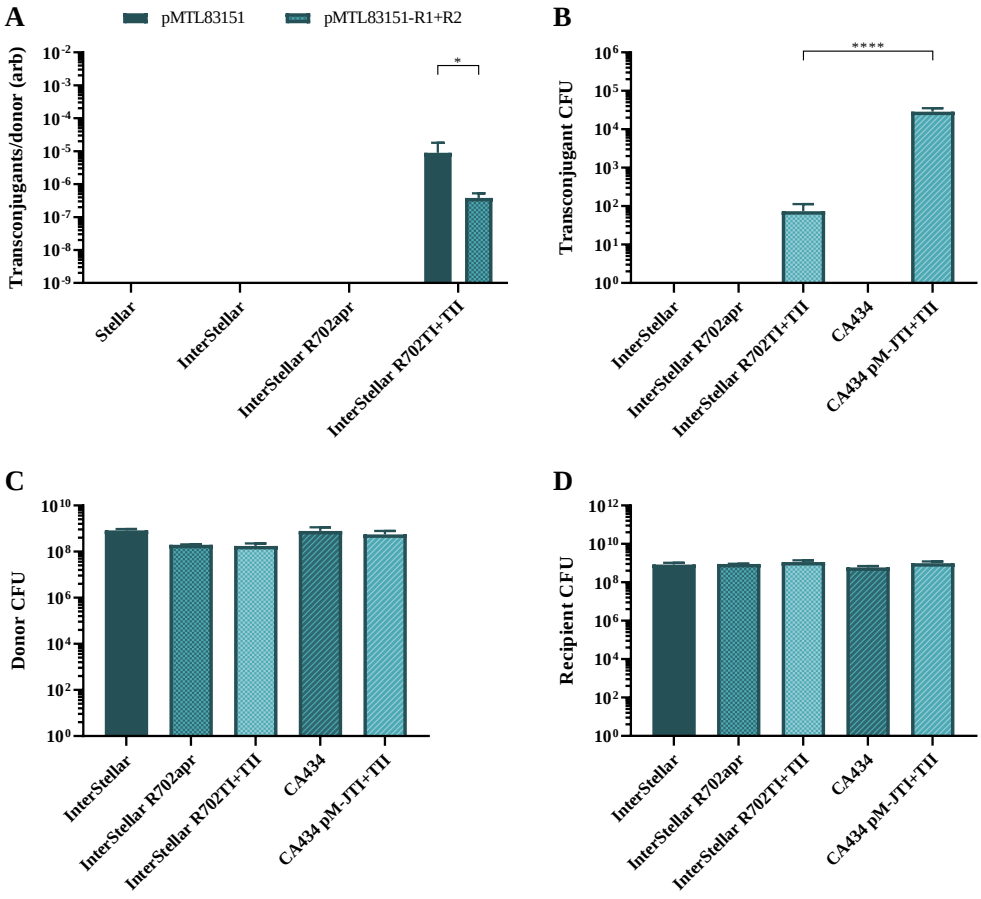


Figure 3.12 Comparison of conjugal donors InterStellar and CA434 with and without methyltransferases in combination with ATCC 19397. (**A**) Conjugation efficiency of plasmids pMTL83151 and pMTL83151-R1+R2 from conjugal donors Stellar, InterStellar, InterStellar R702apr and InterStellar R702TI+TII into ATCC 19397. (**B**) Total number of transconjugants carrying pMTL83151-R1+R2 per conjugal mixture. (**C**) Conjugal donor CFUs carrying pMTL83151-R1+R2 at the start of the conjugation. (**D**) Recipient CFUs after harvesting the conjugal mixture (with the donor strains carrying pMTL83151-R1+R2). Error bars represent the standard deviation of three independent conjugations. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

Surprisingly, ATCC 3502 mating mixtures with InterStellar R702TI+TII yielded significantly less transconjugants compared to InterStellar (Figure 3.11A). Furthermore, expression of the type I *hsdM* and *hsdS* genes did not result in the anticipated transfer efficiency increase for pMTL83151-R1+R2. Previous conjugation assays with pMTL83151/pMTL83151-R1+R2 and CA434 pJTI+TII yielded

approximately 17000 transconjugants, whereas with InterStellar R702TI+TII on average only 270 transconjugants were counted per conjugal mixture (Figure 3.11B/C). The total number of donor and recipient cells per mating mixture was similar for each experiment, eliminating the possibility that the lower yields were caused by unequal numbers of either donor or recipient (Figure 3.11D/E). It should be noted that the OD_{600} of InterStellar R702TI+TII overnight cultures was always slightly lower compared to the other donors (data not shown). To ensure that the donor/recipient ratio was equal across all assays cells were harvested from a larger volume.

Similar results were obtained for ATCC 19397, where conjugations using CA434 pJTI+TII yielded on average 16000 transconjugants compared to 75 using InterStellar R702TI+TII (Figure 3.12B). Additionally, InterStellar R702TI+TII transferred pMTL83151-R1+R2 with a significantly lower efficiency relative to pMTL83151 (Figure 3.12A). Again, there was no significant difference in the total number of donor and recipient present in each conjugal mixture (Figure 3.12C/D).

This could partially be due to lower expression levels of the MTases in the *E. coli* donor, now located on a single copy plasmid instead of the multicopy plasmids tested in Figure 3.9 and Figure 3.10, resulting in incomplete methylation of the shuttle plasmids. Nevertheless, this does not offer an explanation for the low number of ATCC 3502 transconjugants obtained from conjugal mixtures with InterStellar R702TI+TII harbouring pMTL83151. Expression of the MTases should be irrelevant for the success of this transfer. While in conjugation assays with CA434 expression of the type II MTase consistently lowered the efficiency of plasmid transfer into ATCC 3502, differences were always statistically insignificant unlike what has been observed for InterStellar R702TI+TII. Taken together, the incomplete and/or incorrect

methylation of the plasmids by InterStellar R702TI+TII are probably just minor contributors to the poor transfer efficiencies and other, yet unknown, factors also compromise its donating capacity.

Based on the data presented here and the increased rate of mutation exhibited by *dam*$^{-}$ *E. coli* strains, the following hypotheses were formulated: (i) the lack of self-methylation combined with the expression of foreign MTases is toxic to the cells, causing the strain to struggle in general, which indirectly effects the strain's transfer abilities, or; (ii) one/several detrimental random mutations occurred in genes located on the R702TI+TII plasmid important for successful formation of the conjugation apparatus and subsequent DNA transfer.

To further explore both options, conjugal plasmid R702TI+TII was moved from InterStellar into HB101 and CA434 (essentially HB101 harbouring R702) via LE392, as described in section 2.6. In addition, strain HB101 R702 was created by transforming the conjugal R factor R702 from CA434 into HB101 via electroporation, as outlined in section 2.5, and was assayed alongside the other strains as a positive control.

The DNA transfer efficiency was established using pMTL83151, as detailed in section 2.9.3; using pMTL83151 instead of pMTL83151-R1+R2 eliminates the type I RM-system from the equation, simplifying the analysis. The frequency of plasmid transfer into ATCC 3502 with HB101 and CA434 harbouring R702TI+TII was similar to the DNA transfer frequency determined for InterStellar R702TI+TII, while conjugation assays with HB101 or CA434 carrying the standard R702 yielded significantly more transconjugants (Figure 3.13A). This strongly suggest that the poor conjugation

efficiencies are linked to R702TI+TII (theory two), rather than being due to donor specific characteristics.

Interestingly, moving the R702TI+TII plasmid from InterStellar to HB101 and CA434 had a positive impact on the conjugation efficiency into ATCC 19397, albeit yielding half the number of transconjugants obtained with CA434 pTI+TII. Moreover, it should be noted that plasmid transfer from InterStellar R702TI+TII into ATCC 19397 was particularly poor (10-fold less *vs* ATCC 3502), while there was no pronounced difference between ATCC 3502 and ATCC 19397 for donors HB101 and CA434 carrying R702TI+TII. Overall, although not as clear-cut as for ATCC 3502, it appears that R702TI+TII is a worse conjugal R factor than the original R702 plasmid.

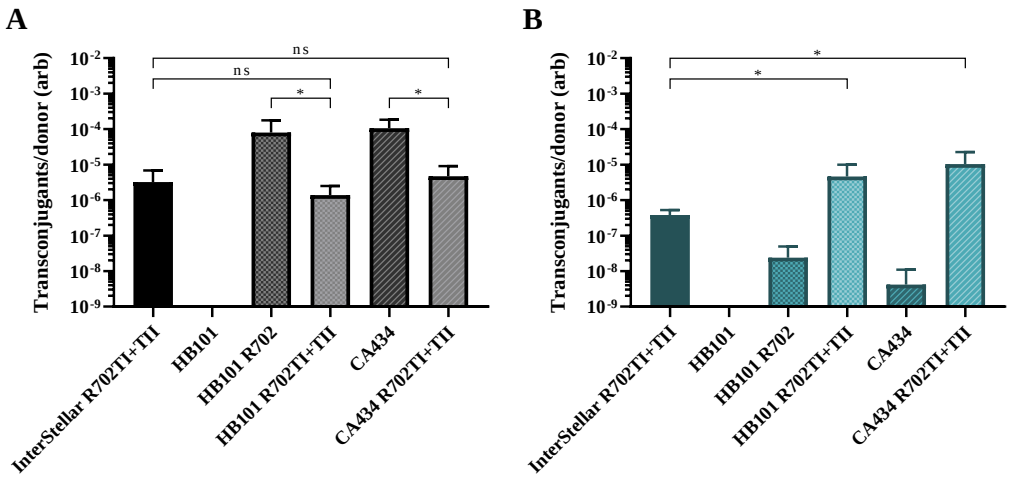


Figure 3.13 Comparison of the conjugation efficiency achieved with different conjugal donors carrying the recombinant R702 plasmid with both MTases. Transfer efficiency of plasmid pMTL83151 from various conjugal donors into recipient strains (**A**) ATCC 3502^{+} and (**B**) ATCC 19397. Error bars represent the standard deviation of three independent experiments. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

The full genome of Stellar, InterStellar, InterStellar R702, HB101, CA434 and the three InterStellar R702TI+TII replicates was determined using Illumina MiSeq sequencing. Genomes were assembled by mapping MiSeq paired end reads to reference genomes *E. coli* HST104 and RR1 using the CLC Workbench. Subsequently, the unmapped reads were used for the *de-novo* assembly of R702. The resulting contigs were annotated using the RAST pipeline. Although the quality of the final genomes was quite poor, for all strains harbouring an R factor one or several contigs could be identified, constituting the 76588 bp long R702 plasmid or the 80762 bp long R702TI+TII plasmid (data not shown). The P_{BAD} and *araC* sequences were not included in the *de-novo* assembly of all three plasmids due to incorrect mapping of these reads to another region by the software, therefore its presence was verified by PCR using primers M13R and CLB-RS13900_R_NotI (data not shown).

Alignment of the different R702 sequences using SnapGene revealed that no additional mutations had been acquired outside the inserted cassette, except for a base substation in the LHA, which overlaps with the integron integrase IntIPac coding sequence (Figure 3.14). Moreover, the assembly and annotation of R702 revealed that the LHA was homologues to a region in the middle of the IntIPac coding sequence. This design error results in the truncation and thus inactivation of the IntIPac gene when the recombineering cassette is inserted into R702. However, there is no literature to support a link between this intergron intergrase and conjugative plasmid transfer. In addition, the transfer efficiencies of InterStellar R702 and InterStellar R702apr were equal, indicating that truncation of the IntIPac gene does not affect conjugative plasmid transfer (Figure 3.11A).

The three other mutations detected were all inside the recombineering cassette; two in non-coding regions (Figure 3.14B; 2 and 4) and one in the *apr* gene (Figure 3.14B; 4).

It is highly unlikely that these mutations altered the DNA donating capacity. Hence, it was concluded that the poor conjugation efficiencies observed for strains harbouring R702TI+TII could not be attributed to mutations in the conjugal R factor.

With the above data suggesting that the poor conjugation efficiencies obtained with donor strains carrying R702TI+TII cannot be assigned to a specific donor nor mutations in the R factor, it remains a mystery what causes this peculiar defect.

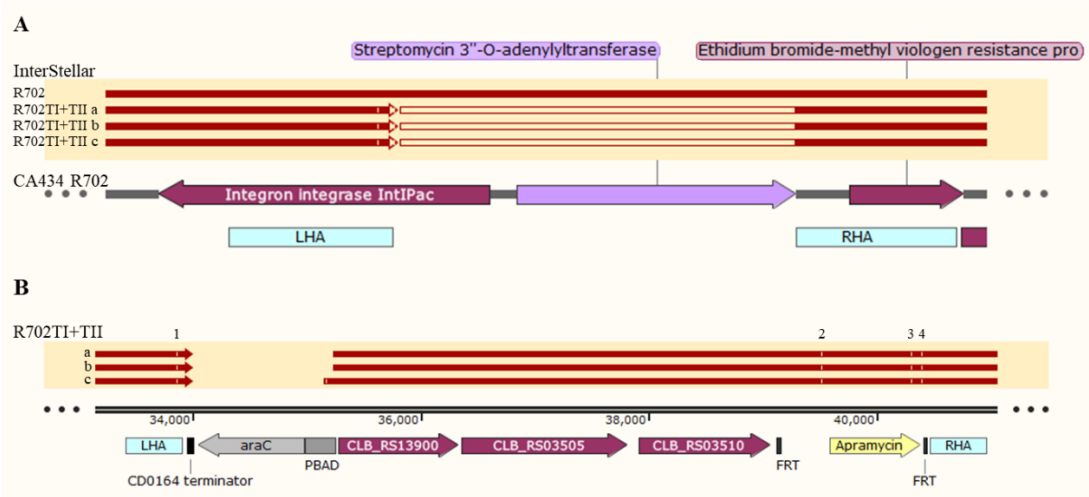


Figure 3.14 Acquired mutations in the recombineering cassette. Schematic representation showing sequence alignments of conjugal R factors from CA434, InterStellar R702 and three InterStellar R702TI+TII replicates using SnapGene. R factor sequences were determined by *de-novo* assembly using Illumina MiSeq sequencing reads. (**A**) Alignment of various InterStellar strains to reference R702 sequence from CA434. Gap in reads a, b and c indicates that this region could not be aligned; in these strains the *aadA1* gene (lavender) is replaced by the recombineering cassette. (**B**) Alignment of three replicates of R702TI+TII to reference R702TI+TII sequence assembled *in silico*. 1: C>A mutation in LHA, 2: G>A mutation non-coding region, 3: C>A (L241M) mutation in apramycin gene, 4: deletion (G) in non-coding region.

3.3.5 Engineering the donor 3: evading type IV restriction systems

Recently a novel conjugal *E. coli* donor strain was published, named sExpress, lacking Dcm methylation.[338] This strain offered a significant improvement in plasmid transfer efficiency compared to the standard CA434 donor for various clostridial strains, including *C. sporogenes*. This effect was attributed to the absence of methylated cytosine nucleotides from incoming DNA, which results in the circumvention of the recipient's type IV system(s).

As previously mentioned, initially the type IV REase in ATCC 3502 and ATCC 19397 went undetected and only came to light when the full genome of the 62A strain (CFSAN064329) was published. Unintentionally, the *dcm*$^-$ donor strain InterStellar had already been tested in combination with ATCC 3502 and ATCC 19397, which did not offer a significant increase in conjugation efficiency over CA434 (Figure 3.11 and Figure 3.12). Therefore, to further explore the potential benefit of using the *dcm*$^-$ conjugal donor sExpress was tested in combination with the three other *C. botulinum* strains.

The standard pMTL83151 plasmid was transformed into sExpress and conjugation assays were performed as described in section 2.9.3, using strains 62A, Loch Maree and NCTC 2916. Plasmid transfer efficiencies significantly increased for all three *C. botulinum* strains using sExpress compared to CA434 (Figure 3.15). A 100-fold change was observed for the Loch Maree strain and a 5-fold increase for 62A and NCTC 2916.

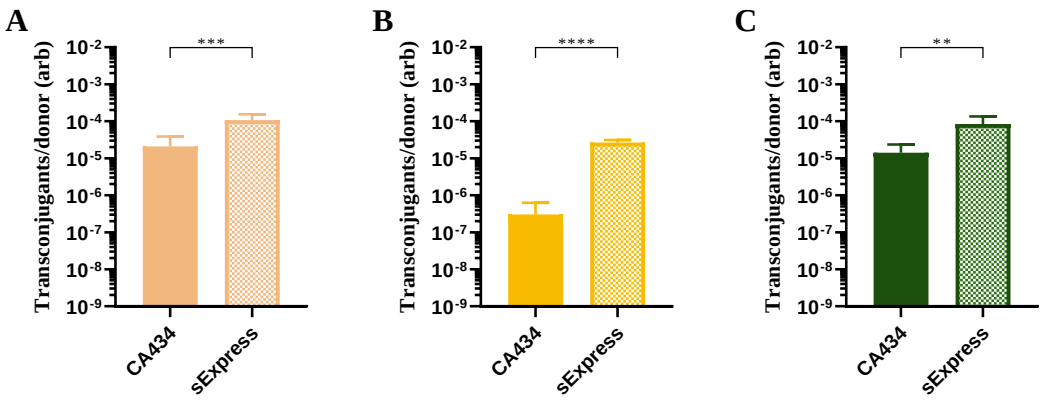


Figure 3.15 Comparison of conjugative plasmid transfer from donor strains CA434 and sExpress into 62A, Loch Maree and NCTC 2916. Transfer efficiency of plasmid pMTL83151 from conjugal donors CA434 and sExpress into recipient strains (**A**) 62A, (**B**) Loch Maree and (**C**) NCTC 2916. Error bars represent the standard deviation of three independent conjugations. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

Interestingly, the total number of transconjugants per conjugal mixture with sExpress was within a range of 15000-50000, which is similar to what has been previously observed for strains ATCC 3502 and ATCC 19397 combined with their best donor. This suggests that some sort of plateau has been reached, where restriction systems are no longer the limiting factor, which could also explain why InterStellar did not offer a significant increase in efficiency over CA434 in the assays with NCTC 3502 and ATCC 19397. Possibly, the procedure itself (media composition, incubation time, growth rate etc.) has now become the limiting factor and to further increase efficiencies changes to the experimental set-up should be considered.[340-342]

However, the above theory does not offer an explanation for the difference in conjugation efficiency observed between ATCC 3502 and 62A. Their restriction profiles are identical and therefore similar conjugation efficiencies were anticipated. Alignment of the type IV CBO2086 gene from ATCC 3502, including a 1 kb region upstream and downstream of the gene, to the ATCC 19397 and 62A genome showed

that these regions are 100% identical, excluding the possibility that the observed dissimilarity can be attributed to differences between the type IV genes. To elucidate why 62A less readily accepts foreign DNA a more in-depth genomic analysis will be required combined with additional experiments, which is beyond the scope of this project.

Furthermore, InterStellar and sExpress are two distinct *E. coli* strains, with different methylation patterns; *dam*$^{-}$/*dcm*$^{-}$ and *dam*$^{+}$/*dcm*$^{-}$, respectively. To further investigate the differences observed between the various strains and donors, conjugation assays using sExpress with ATCC 3502 and ATCC 19397 will be required. Additionally, the transfer efficiency of pMTL83151 from InterStellar into strains 62A, Loch Maree and NCTC 2916 should be determined. This will provide more insight into whether the differences are donor and/or recipient related or if the assay conditions are the limiting factor.

3.4 Key outcomes

- Inactivation of the type I REase in *C. botulinum* ATCC 3502 improves the conjugation efficiency for plasmids harbouring the type I specific recognition site CTANNNNNNRTGAA.
- The efficiency of conjugative plasmid transfer into *C. botulinum* type A strains is significantly increased through circumvention of the restriction barriers by expressing the associated MTases in the conjugal *E. coli* donor.
- Utilisation of the *dcm*$^{-}$ conjugal donor strain sExpress improves conjugation efficiencies for strains 62A, Loch Maree and NCTC 2916.

- Optimised conjugal donor strains for DNA transfer into *C. botulinum* strains can be created using the following customisation pipeline: (i) whole genome sequencing of the recipient (using PacBio SMRT); (ii) identification of its restriction systems (using REBASE); (iii) selection of an appropriate conjugal donor strain (Dam/Dcm methylation profile), and; (iv) expression of relevant MTases in donor.

CHAPTER 4

Hypermutability and the role of *mutL*

4.1 Introduction

4.1.1 Bacterial genome evolution

“Survival of the fittest” is one of the most famous quotes used to describe evolution. The phrase aims to describe the mechanism of natural selection, a concept that originated from the Darwinian evolution theory, in a simple, yet effective, way.[343] In modern evolutionary biology, fitness is still viewed as the most important property of any organism and is defined as "survival of the form that will leave the most copies of itself in successive generations".[344] Reproductive success is generally linked to being better adapted to the current local environment.[345] Although this theory in essence seems straightforward, the actual dynamics of this evolutionary process are far more complex; who is the fittest, and possibly even more important, how does one become the fittest?

Evolution is happening continuously, not only in fast changing environments but also under stable conditions, with the aim to optimise fitness. According to Darwinian evolution theory, there is a constant selection on many mutations of small effect and the resulting marginal gains lead to gradual adaptation.[343] However, experimental research has shown that in some lineages there are alternating periods of rapid phenotypic evolution and stasis.[346-347] Occasionally, rare genetic ‘revolutions’ occur that result in a fast sudden change in fitness; a single small change initiates a cascade of other changes in a population that was previously in a long near stasis state. For example, during the early phase of evolution for large populations in a new environment, where there are plenty of adaptive mutations available, the most beneficial mutations should dominate. Gradually, the adaptation rate should decrease with less adaptive mutations of large effect being available. Based on the empirical

evidence, an alternative model was proposed, describing the irregular fitness landscape of adaptive evolution, though controversy remains regarding the probably diverse underlying causes of adaptation rate changes.[348]

Typically, microorganisms, such as bacteria, live in a large community of varying degrees of complexity; with populations that originated from a single cell via clonal expansion and groups that include organisms from all three domains of life on opposite ends of the spectrum.[349] The characteristics of these communities, such as reproductive success and resilience to external disturbance, depend on internal genetic diversity, which is shaped by selective pressure.[350] Additionally, diversification and adaptive evolution in the community is determined by: (i) environmental changes (and bottlenecks); (ii) random genetic drift (stochasticity in mutation frequency), and; (iii) chance events, also referred to as historical accidents.[351-352] For example, although mutations arise spontaneously by chance, analysis of large populations in identical environments has revealed that the order at which these mutations occur is important.[352] Certain mutations, despite being beneficial, can lead to evolutionary dead-ends. Furthermore, adaptive mutations might have negative side effects, which creates opportunities for new beneficial mutations to compensate them.[348] Disentangling the intricate roles of "chance and necessity" in diversification and adaptive evolution is complicated. Nevertheless, thanks to technical advancements (e.g., sequencing technology), analysis of similarities between adaptive solutions achieved by replicate populations is starting to increase the general understanding of the different factors and dynamics contributing to the evolution of successful phenotypes.[353]

Genetic diversity plays a central role in adaptive evolution, however, the availability of beneficial mutations is not the only determinant of fitness landscapes. Phenotypic

heterogeneity (functional phenotypic variation between genetically identical individuals in the same environment) also influences the reproductive success and overall resilience of a population.[354-355] The modifying effect of phenotypic heterogeneity on evolutionary dynamics, together with genetic diversity, shape the adaptive landscape. A potential function of phenotypic heterogeneity is to ensure the production of offspring in variable conditions; some individuals suffer from reduced fitness in their environment in exchange for increased fitness in temporally fluctuating conditions.[356] In bacterial populations this risk-spreading (bet-hedging) strategy to diversify phenotypes, allows genotypes to persist in variable conditions, maximising survival. Variations in phenotype between genetically identical individuals is the result of random fluctuations in the biochemical reactions within cells. Stochastic variations in gene expression, so called 'noise', is a key determinant of phenotypic heterogeneity.[357] The effect of non-genetic phenotypic variation on adaptive evolution might be stronger than currently appreciated. Nevertheless, its potential contribution will not be included in further discussions. The primary focus of this chapter will be on genetic variation driving adaptive evolution.

In bacterial populations various mechanisms are important for generating the genetic variation among cells essential for adaptive evolution. Bacterial genomes can evolve via different routes, either by endogenous DNA changes or through the acquisition of exogenous DNA, such as plasmids and other mobile genetic elements (see Chapter 3 for more information about horizontal gene transfer).[358] Endogenous DNA changes, generally referred to as mutations, are considered the ultimate source of heritable variation and can emerge from coordinated cell processes or external stresses.[359-360] Mutations come in a variety of shapes and sizes and encompass sequence deletions, duplications, insertions, inversions and translocations. One base pair changes in the

DNA sequence, also known as point mutations or single nucleotide polymorphisms (SNPs), can be single nucleotide variants (SNVs) or single nucleotide insertions/deletions and they arise due to a variety of reasons, such as, erroneous DNA replication and a decreased efficacy of DNA repair. Genetic recombination events can lead to larger genomic insertions, deletions (also known as InDels) and rearrangements, and can occur due to DNA sequence repeats or mobile genetic elements (transposons).[361] Transposable elements are small mobile genetic sequences that encode enzymes which facilitate the movement of DNA within or between genomes.[362] The rate at which these mutations occur can vary widely between organisms and even within a clonal bacterial population, affecting genome stability.[360-361]

During the last two decades, next generation sequencing (NGS) technologies have revolutionised genetic studies and accelerated the analysis of genome evolution by making large scale whole genome sequencing (WGS) widely accessible.[363-364] The most significant factor fuelling this revolution was the vast cost reduction of sequencing, resulting from technical advancements. Large scale sequencing of many different bacteria revealed that the genetic diversity among bacterial genomes, even from the same species, is far greater than what was previously suspected. Two closely related bacterial genomes may contain quite a few different insertions and deletions due to recombination events, possibly as a result of constantly combating viral infections. Furthermore, WGS aided the understanding of genome evolution dynamics; a good example of this is the Long-Term Evolution Experiment (LTEE) with *E. coli*.[348, 352-353, 365] One of their surprising findings was that in their populations genomic evolution follows a near-linear trajectory whereas the fitness trajectory (adaptation) was strongly nonlinear. Although several theories have been proposed,

the reason for this divergence is yet unknown. Moreover, WGS has become an essential tool for the surveillance and diagnostics of infectious diseases and is now part of most infectious disease programs at the European Centre of Disease Prevention and Control (ECDC) and the US CDC.[366-367]

4.1.2 The balance between maintaining genome integrity and evolution

With mutations being the raw material on which natural selection can act, the rate at which they are being supplied affects evolution. Although it might seem logical that a higher mutation rate always benefits adaptive evolution, there is also a cost associated with an increased supply of beneficial mutations.[359, 368-369] Most of the mutations a cell acquires are harmful or neutral rather than beneficial.[370-372] Therefore, cells have to balance the need for mutations to adapt and the costs associated with an increase in deleterious mutations (genetic load). High mutation rates and the concomitant increase in genetic load are presumed to be a potent selective disadvantage, which has resulted in the low mutation rates typically observed for bacterial genomes.[373-374] Usually, the genomic mutation rate of bacteria is around 0.001 per generation, however, based on the known mechanisms responsible for maintaining genetic integrity, cells with a 10 to >10,000-fold higher mutation rate could, in theory, occur.[375-376]

Varying mutation rates within a species have been observed in laboratory and natural bacterial populations.[369, 377-388] This suggests that bacteria can 'tune' the rate at which they acquire new mutations by, for example, changing the fidelity of DNA replication.[359, 389-390] This tuning of mutation rates is probably a response to changes in selective pressure and is one of the many factors affecting adaptive evolution.[391]

Cells exploit various mechanisms to control their mutation rate, which will be discussed in more detail in section 4.1.3.[391-393] Mutations in these systems often alter

the type or rate of future mutations and are, therefore, also known as mutation rate modifiers. When the inactivation of a gene elevates the mutation rate, it is referred to as a mutator gene. Opposite to this are genes which often contain mutations that decrease the mutation rate and are so called anti-mutator genes.[359, 394-395]

Several different factors, such as environmental constancy and complexity, genome size and population size, could, theoretically, determine the optimal mutation rate.[374] A well-adapted population growing under constant conditions typically exhibits a low mutation rate, as the cost of deleterious mutations is relatively high compared to the potential fitness gain from the few beneficial mutations of low effect still available. Genomic mutation rates, however, cannot decrease indefinitely because the rate reduction is believed to be restricted by the costs associated with increasing and maintaining DNA integrity.[396-399]

Bacterial populations living in constantly changing conditions need to rapidly adapt and can, therefore, benefit from greater genome flexibility.[400-401] Furthermore, in bacterial populations that experience a sudden environmental shock (e.g., a natural isolate now grown under laboratory conditions) higher mutation rates could confer a selectable advantage. Here, greater mutation rates can speed up the adaptation process, as mutators (cells with an elevated mutation rate relative to the original WT) are more likely to acquire beneficial alleles compared to the non-mutator WT. In addition, new environments provide a fresh pool of adaptive mutations of large effect. This can shift the balance between the potential fitness increase and the concomitant genetic load.

Unexpected flexibility in bacterial mutation rates has been observed when populations were exposed to a near-lethal stress (e.g., alcohol or antibiotics) to avoid complete extinction.[389, 401-404] Moreover, viral pressure or frequent periods of strong selection

caused by other stresses, might result in long-term selection for elevated mutation rates, and thus lead to fixation of mutator alleles in bacterial populations.[405-407]

The dynamics of mutator evolution in spatially growing populations (cells attached to fixed surface) is slightly different, as the growth abilities of mutator strains decrease significantly over time compared to non-mutator strains.[389] This fitness decrease is less pronounced in bacterial populations without spatial fixation. During the range expansion, cells probably accumulate deleterious mutations affecting the growth rate but continue to persist because there is no direct selection (competition) against the decline in growth rate.[360, 408] Theory predicts that the fitness of individuals on the front of the expansion will continue to decline over time and space, which, especially in mutators, can ultimately lead to extinction depending on the population size.[409-411] Nevertheless, mutator alleles and other deleterious load can still persist when linked to beneficial mutations (also known as allele surfing).[412-413]

Although increased mutation rates indirectly confer a fitness advantage in certain situations, extremely high mutations rates (hypermutation) can also hinder adaptation.[390, 414] In the long term, an hypermutators fitness can rapidly decline due to the increasing genetic load; the majority of mutations a cell can acquire are deleterious, and the resulting fitness decrease cannot be compensated by the relatively few available beneficial mutations.[371, 415]

Mutator genotypes, which are normally a selective burden, can ‘hitchhike’ to a high frequency together with beneficial alleles and over short time scales even rise to fixation in bacterial populations.[382, 394, 416-417] The mutation rate of these mutators, which reached fixation during short-term selection, is probably higher than what is optimal for long-term evolution.[396] Mutation-evolution models and experimental data

show that mutation rates should rapidly decline once a population is well adapted to the environment.[359, 389-390, 394-395] There is no evidence that fixed mutator alleles confer a fitness advantage in the long-term.[418] Yet, mutator phenotypes can persist for thousands of generations in bacterial populations, despite the selective penalty incurred.[419] Shaver *et al*. proposed four possible reasons that could account for this phenomenon: (i) the mutator genotype is not revertible (e.g., a large deletion in a coding region); (ii) accumulated additional mutations prevent the mutator population to revert to WT; (iii) the rate of adaptation remains high enough for the mutator phenotype to still have a selective advantage over the WT, or; (iv) the fitness gain resulting from reverting to WT is marginal and, therefore, has no effect over the studied time scale. Although this question was raised quite a while ago (2003) and lots of new data has become available since then, the answer still remains unknown. The continuous efforts of many different researchers, studying evolution across the tree of life, will keep adding little pieces of information to this complex puzzle with the aim to further elucidate what drives mutation rate changes and ultimately adaptive evolution.

4.1.3 Bacterial systems to maintain genome integrity

All organisms, across the tree of life, exploit a series of mechanisms to control their genomic mutation rate in order to maintain genome integrity, while allowing enough genetic variation to occur for adaptive evolution. Cells have evolved a number of stress response and DNA repair pathways to recognize and fix the multitude of DNA mutations and modifications that can occur throughout a cell's life.[339, 420-422] The damage and errors can be caused by exogenous sources or by normal cellular metabolism.[423-424] The genome maintenance toolbox available to a cell can roughly be divided into two: systems for faithful genome duplication and systems that repair DNA

damage. Most systems are highly conserved across organisms, ranging from bacteria to mammals.[421, 425-426] Below a brief overview of the best known pathways in bacteria to control genome stability, specifically in Gram-positive bacteria, is provided.

The SOS response has been studied extensively over the past decades, which has revealed that its function varies between bacterial species and is not exclusively a DNA damage response pathway but is also involved in a variety of other stress response mechanisms.[427] The DNA damage response coordinates various cellular processes, such as DNA repair and translesion DNA synthesis (the bypassing or repairing of DNA lesions at stalled DNA replication sites using specialised, often low-fidelity, polymerases), to protect the cell against all sorts of stresses that affect DNA integrity.[427-429]

Upon DNA damage, the multifunctional protein RecA (also known as RecE) is activated. RecA is required for homologous recombination but also facilitates the self-cleavage of the transcriptional repressor LexA, inactivating LexA and preventing it from repressing the transcription of genes.[430-432] LexA represses the expression of multiple SOS genes by binding to their promoters, thus negatively regulating the SOS response.[433-434] To date, there is some evidence that LexA is not solely a repressor (as in *E. coli*), but can also positively regulate gene expression.[427]

Thus far, LexA is consistently found to regulate the SOS response. The genes under its control, however, vary considerably from organism to organism.[430, 435-440] Often, in addition to the various DNA maintenance systems, LexA controls the expression of genes with a different function, including toxin expression, antibiotic resistance or sporulation.[441] In *B. subtilis* activation of the SOS response upregulated a variety of genes involved in different DNA maintenance processes, including *uvrA* and *uvrB*

(nucleotide excision repair, discussed in more detail below).[420] Furthermore, it seems that in *B. subtilis* the SOS response is not necessarily the main DNA damage response pathway, but in addition to this is able to repair damaged DNA and tolerate replication fork disturbance using SOS-Independent DNA Damage Responses.[427, 442]

High-fidelity DNA replication is essential for accurate genome replication and thus genome stability.[443-444] In low-GC Gram-positive bacteria DNA polymerase III PolC and DnaE (MutI) are both present in most species. In *B. subtilis* both PolC and DnaE are essential, however, if PolC and DnaE are generally essential across low-GC Gram-positive bacteria is currently unknown.[420, 445-446] PolC consist of two domains, an exonuclease domain (for proofreading) and a polymerase domain, whereas DnaE only contains a polymerase domain (in *E. coli* the proofreading functionality is fulfilled by DnaQ).[447] Accordingly, characterisation of DnaE confirmed that it is an error-prone polymerase.[448-449] Moreover, it has been hypothesized that DnaE is not only essential for DNA replication but also has a function in mutagenesis and translesion synthesis.[420]

DNA mismatch repair (MMR) pathways work closely together with the DNA replication machinery to further increase replication fidelity.[375, 422, 450 451] MutS and MutL are the core MMR proteins, and depending on the organisms a variety of other proteins are involved.[452-453] Although the current understanding of MMR in Gram-positive bacteria *in vivo* is limited compared to *E. coli* and eukaryotes, a potential mechanism has been proposed by Lenhart *et al.*[454] DNA replication and MMR are coupled through DnaN (the β-clamp of the DNA replication complex), which recruits MutS. Following DNA mismatch identification by MutS, the endonuclease MutL is recruited to mismatch-proximal DNA and subsequently creates a DNA incision.[455-459] Next, the WalJ (YycJ) exonuclease loads to the nicked site and degrades the mismatch-

containing DNA strand.[460] It is thought that in the final steps, strand resynthesis and backbone ligation, is completed by PolC or DnaE and a DNA ligase.[454]

Interestingly, some bacteria possess a MutS paralog, named MutS2. Despite their similarity there is no clear evidence confirming MutS2 is part of the MMR pathway in Gram-positives.[420, 461] Analysis of MutS2 in *B. subtilis* and the Gram-negative bacteria *Helicobacter pylori* and *Thermus thermophilus* has revealed that MutS2 is an endonuclease, linked to homologous recombination.[462-464]

Furthermore, the mutator genes MutY and MutM are part of the base excision repair (BER) pathway, dedicated to the removal of damaged bases.[465-466] It specifically targets oxidative damage, such as the most common 8-oxo-G–A mismatch. MutM selectively removes 8-oxo-G, while MutY removes the mismatched adenine.[467-469] The site is subsequently closed by a repair polymerase (e.g., Pol I) and ligase. Inactivation of each protein significantly increases the mutation rate, albeit the impact of a mutM deletion on the mutation rate (approx. 5-fold) is not nearly as strong as the effect of a mutY-deficiency (approx. 100-fold).[394, 470] In addition, two anti-mutator proteins have been identified, YtkD and MutT, protecting growing *B. subtilis* cells against oxidizing agents.[471-473]

The more bulky DNA lesions caused by reactive oxygen species, drugs or UV are repaired via the Nucleotide Excision Repair (NER) pathway.[474] The conserved UvrABC nuclease complex recognises and excises the damaged bases of such lesions. In *C. difficile* the NER pathway seems to be controlled by sigB whereas in *B. subtilis* data suggests it is part of the SOS response.[430, 475] Furthermore, activation of the NER pathway has been linked to mutation frequency decline protein Mfd, which is involved in transcription coupled DNA repair.[476-478] Mfd removes stalled RNA polymerases

from the DNA template when it encounters damaged DNA and its disruption affects cells' capacity to perform homologous recombination and decreases sporulation efficiency in *B. subtilis.*[479-480]

Moreover, in *B. subtilis* the sporulation process is delayed when damaged DNA is detected by DNA integrity scanning protein DisA.[481] DisA scans the chromosome for DNA repair intermediates, such as Holliday junctions, to prevent cells with damaged genomes to progress into the spore formation process, which would inevitably lead to non-viable spores.

Homologous recombination is another key system for both DNA repair as well as adaptive evolution. This pathway is critical for cell survival; via this pathway bacteria can repair double-strand DNA breaks, which, if left unpaired, are lethal.[231, 482-483] The restoration of double-strand breaks via homologous recombination is a complex multistep process that requires the participation of a multitude of different proteins. Briefly, upon recognition of a double-strand break, a recombinase, usually RecA, is loaded onto the single-stranded DNA. Subsequently, the single-strand DNA is paired with intact homologous DNA, forming a crossover junction (also known as Holliday junction), followed by DNA synthesis on the invading strand to restore the DNA strand that was displaced during the invasion.[483]

RecA, which acts as an activator for the SOS response, as mentioned previously, also catalyses strand exchange during the recombinational process, hence the name, and when bound to single-stranded DNA searches for similar sequences along the double-stranded DNA.[484-485] In conjunction with RecA, RuvA and RuvB bind to Holliday junctions and subsequently direct branch migration.[486-488] To resolve the Holliday junction RecU is recruited, which catalyses DNA strand cleavage and regulates RecA

activities.[489-490] The inactivation of any of these three proteins substantially increases the cell's sensitivity to DNA damage and reduces transformation efficiency.

Other key components of this pathway are RecN, involved in the early stages of homologous recombination; AddA and AddB, required for double-stranded end processing; SsbA, involved in strand exchange and RecO, required for the repair of single-strand gaps.[482, 491-495] In addition, PriA, DnaD, DnaB, and DnaI are important for replication restarts after DNA damage repairs in *B. subtilis*.[496-497] Furthermore, many other proteins have been linked to the homologous recombination pathway and DNA break repair but their exact role is currently unknown (e.g., RecQ, RecS, RecJ, RecF, RecM, RecV, RecG, RecD2, SbcC, SbcD, ScpA, ScpB, RadA, RadB and YhaN (now known as SbcE).[420, 498]

Although most DNA modifying pathways in the cell serve to minimise genetic alterations, there are also cellular pathways that increase the mutation rate upon activation.[427] This temporary increase in mutation rate is part of the coordinated defence responses bacteria have evolved to survive environmental shocks.[499] To adapt cells can upregulate the expression of error-prone DNA polymerases IV (also known as DinB or PolY1 and PolY2) and V by approximately 10-fold.[429, 500] These low-fidelity DNA polymerases lack the intrinsic proofreading activity and they can both act on DNA damage by performing translesion synthesis at stalled replication fork sites as well as undamaged DNA. These more promiscuous polymerases allow cells to survive sudden increases in DNA damage.

Furthermore, repeat sequences seem to be overrepresented in stress response genes compared to the rest of the genome.[419, 501] It has been hypothesized these repeats evolved to promote survival under stressful conditions. Spontaneous duplications or

deletions occur more frequently at repetitive DNA sites, increasing the chance that, for example, a mutator phenotype emerges within the population. Under normal conditions this mutator, which suffer from a mutational burden, will disappear from the population quickly, however, as described previously, can aid adaptation in changing environments and survival during stress.[389, 406, 502]

4.1.4 ATCC 3502, a strain with high population heterogeneity

Genetic variability and mutation rate changes in bacterial populations might be very interesting from an evolutionary standpoint, but when bacteria are being repurposed for the production of useful compounds or engineered to study fundamental cell processes, this variation can be highly problematic. Strain degeneration is frequently observed during the industrial and lab-scale fermentation of bacterial strains, especially in continues culture.[503] The degenerates are usually compromised in their capacity to produce the compound or enzyme of interest (e.g., decreased growth rate or loss of the production trait). Additionally, SNVs might occur in the gene encoding the final enzyme product, which might result in (subtle) activity changes. Degeneration is usually due to mutation rate increases caused by mutations in the DNA replication, DNA mismatch repair or DNA damage repair systems, but can also be the result of a single recombineering event, the transposition of a mobile genetic element or non-chromosomal DNA loss.[503]

Furthermore, population heterogeneity and genome instability complicate fundamental research, especially when experiments require single colony stages. Discrete colonies are derived from a single cell and thus the genetic variation within the population is lost. Additionally, single cell bottlenecks often lead to the fixation of arbitrary mutations, which might cause phenotypic changes.[504] Therefore, minimising

single cell passages or other population bottlenecks is generally preferred. However, to study protein function the most powerful strategies require DNA modifications (e.g., gene knock-outs or protein fusions). To isolate individuals with the desired mutation single colony stages are necessary. Thus, a very homogeneous starting population and a stable genome are desirable when engineering bacterial strains.

Since the WGS revolution the awareness of secondary mutations and their potential impact has significantly increased. The study of the *C. botulinum* type A model strain ATCC 3502 is a perfect example where WGS data revealed that quite a significant issue had been hiding beneath the surface. In the SBRC *C. botulinum* laboratory several researchers struggled with generating phenotypically identical gene knock-out strains. For example, the sporulation capacity was often compromised in some, but not all, mutants. Maintaining a high yield of viable spores proved to be particularly difficult when generating non-toxicogenic strains via ACE (Figure 4.1).[225]

Sequencing the whole genome of a few mutants revealed that the first mutant (*pyrE*$^{-}$) carried 59 SNPs compared to the WT ATCC 3502 (IFR). During the second step, deletion of the *boNT/A1* gene in the *pyrE*$^{-}$ strain, and the third step, restoring the *pyrE* allele in the Δ*boNT/A1 pyrE*$^{-}$ strain, an additional 13 SNPs and 15 SNPs were acquired, respectively. At the time, many clostridial genomes were sequenced by the *Clostridium* research group (e.g., *Clostridium beijerinckii*, *C. sporogenes* and *C. difficile*) and usually only one or two secondary mutations were identified in the genetically modified strains (personal communication; N. Minton). The unexpected high number of secondary mutations in the *C. botulinum* Δ*boNT/A1* strains and the concomitant phenotypic changes render it unsuitable for future studies.

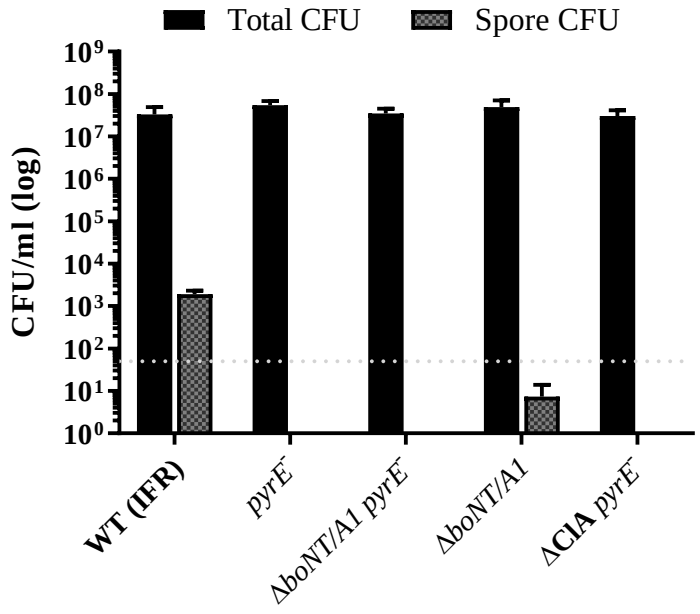


Figure 4.1 Sporulation capacity of ATCC 3502 (IFR) *pyrE*⁻, Δ*boNT/A1 pyrE*⁻, Δ*boNT/A1* and ΔClA *pyrE*⁻ mutants. Sporulation/germination capacity of *pyrE*⁻, Δ*boNT/A1 pyrE*⁻, Δ*boNT/A1* and ΔClA *pyrE*⁻ mutants compared to WT created by Christopher Humphreys. Strains were cultured in TYG broth and samples were taken after 72 hours. Total CFU corresponds to the total cell count (non-heat treated) and the spore CFU is the number of viable spores withstanding a 15 min, 80°C heat treatment. Error bars represent the standard deviation of three independent cultures for WT and of three biological replicates for the mutants. The dotted line at 50 CFU/ml represents the detection limit.

The sequencing data suggests that the ATCC 3502 (IFR) population kept at the SBRC has an unstable genome. This finding raised many questions, such as what causes the genome instability (e.g., stress, disruption of the DNA mismatch repair system or is it just an inherent strain characteristic that evolved to survive in the wild) and is only the ATCC 3502 (IFR) stock in Nottingham affected or are other ATCC 3502 populations also prone to accumulating mutations? If the latter is true, this might have far-reaching consequences, as ATCC 3502 is currently the most studied type A strain.[168-169, 181, 505-511] It is possible that these studies have employed degenerated strains whose associated phenotypes may differ compared to those of naturally occurring WT strains. Moreover, the majority of the more recently published work involved the study of

genetically modified strains; this leaves one contemplating if some of the observed phenotypic changes linked to a specific genotypic modification is genuine or if they should be attributed to, as yet unidentified, secondary mutations.[170-173, 176, 181-184, 512-515]

4.2 Objectives

i. Find a genetically stable *C. botulinum* ATCC 3502 population or isolate in order to create safe strains,

ii. Identify the cause of the observed genome instability in *C. botulinum* ATCC 3502.

4.3 Results and discussion

4.3.1 A fresh start: engineering a 'new' ATCC 3502 strain using ACE and ClosTron mutagenesis

At the beginning of this project all available information suggested that the observed genomic instability of strain ATCC 3502 was merely affecting research performed at the SBRC in Nottingham (personal communication; N. Minton). Initially all Nottingham based students were working with the same *C. botulinum* WT stock, which was acquired from the Institute for Food Research (IFR; nowadays known as the Quadram Institute Bioscience) in 2006. Interestingly, researchers at IFR stated they had no indications that supported a similar genetic instability in their stock. This suggested that either the strain had undergone some sort of unfortunate change when cultured for storage in Nottingham or that standard culturing methods in Nottingham somehow induced higher mutation frequencies. The latter was considered unlikely because in 2011 a different ATCC 3502 spore stock (provided by Unilever) came to Nottingham and since then both strains have been cultured under the exact same conditions. In the Unilever strain, however, no unaccountable discrepancies between biological replicates were ever observed (personal communication; C. Meaney). Based on the above information, in combination with no other cases of instability being described in literature, it was assumed that only the IFR SBRC WT stock suffered from an unusually high mutation rate. Although the behaviour of this particular ATCC 3502 WT stock raised plenty of interesting questions, they all strongly deviated from the initial aim of this project, the creation of safe *C. botulinum* strains. Therefore, it was decided to abandon the IFR WT strain without any further investigation and continue all work with the ATCC 3502 Unilever stock.

Initially, two genes were knocked out individually in the ATCC 3502 Unilever strain, using two different genome editing tools. In the first set of mutants the *hsdR* gene was inactivated via ClosTron mutagenesis with the aim to improve DNA transfer efficiencies, as described in section 2.12.1 and section 3.3.2. The three biological replicates of strain *hsdR*1600s::CT were isolated from independent conjugal mixtures. In the second set of mutants the *pyrE* gene was truncated using ACE as described in section 2.12.2. The deletion was confirmed by PCR using a primer annealing to the inserted *lacZα* fragment and a primer annealing to the chromosome in three biological replicates isolated from a single conjugal mixture (Figure 4.2). The plasmid was lost from the selected *pyrE*$^{-}$ mutants, as described in section 2.8, and the truncation of the *pyrE*$^{-}$ gene was subsequently verified by Sanger sequencing. Creation of these *pyrE* negative (*pyrE*$^{-}$) strains was part of a multi-step process to ultimately generate non-toxigenic *C. botulinum*. As previously described (section 1.5.2), the inactivation of the *pyrE* gene results in uracil auxotrophic cells, which allows the in-frame deletion of the neurotoxin gene(s) via standard allelic exchange using a plasmid based non-native *pyrE* gene for counter selection.[188]

Given the high mutation rate of the ATCC 3502 IFR strain and that the reason for this was still unknown, the Unilever WT stock as well as the *hsdR*1600s::CT and *pyrE*$^{-}$ mutants were handled with extra care to minimize their exposure to potential stresses that could induce mutagenesis, such as reactive oxygen species and temperature shocks.[423, 516-517] Additionally, during the generation/isolation of the mutants subculturing steps and single colony stages were kept to an absolute minimum to further decrease the risk of picking up undesired genetic changes.

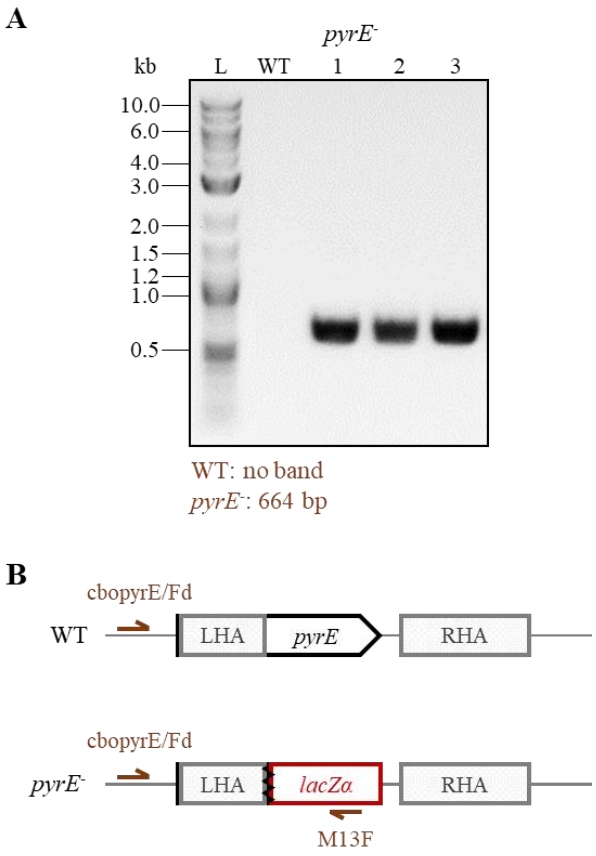


Figure 4.2 Truncation of *pyrE* gene in strain ATCC 3502. (**A**) PCR for three biological replicates of ATCC 3502 *pyrE*- mutants alongside WT. Products were amplified using flanking primers cbopyrE/Fd and M13F. Lane L: NEB 1 kb Plus DNA Ladder. (**B**) Schematic diagram of the targeted chromosomal region showing the primer binding sites in the WT and mutant strain.

Before continuing the different projects for which the mutants were generated, they were subjected to a basic phenotypic characterisation to confirm the presumed stability of the ATCC 3502 Unilever strain. The rational was that no significant differences between the original WT phenotype and its engineered variants beyond the expected effects of the intended genetic modifications should be observed if the Unilever strain is not susceptible to the accumulation of random SNPs and InDels. The ability to form viable spores was considered the most appropriate phenotype to verify stability because: (i) spore formation is a non-essential functionality in laboratory conditions and can thus be readily lost; (ii) sporulation efficiencies can be determined with an

easy phenotypic screen, and; (iii) the high number of genes involved increases the probability of the process being affected by mutations, hence there is a higher chance of observing a phenotypic change. Furthermore, neither the *hsdR* nor the *pyrE* gene have been linked to sporulation and therefore their inactivation should normally not impact viable spore counts.[206] The growth and sporulation/germination capacity of the mutants was assessed as described in sections 2.9.2 and 2.9.4, respectively.

Five out of the six mutants were able to attain similar total cell density levels as WT and grew at a comparable rate (Figure 4.3A). However, one mutant (*hsdR*1600s::CT 3) had a considerably longer lag phase and was not capable of reaching the same total cell density as WT. Furthermore, during the 'death' phase the growth profiles of all mutants started to deviate slightly from the WT profile, with strains *pyrE*$^{-}$ 2 and 3 showing a clear drop in cell density after a relatively short stationary phase not observed in WT.

Surprisingly, all six mutants were compromised in their ability to form viable spores compared to WT (Figure 4.3B). For two out of the three *hsdR*1600s::CT mutants a 10-fold reduction in sporulation/germination efficiency was observed. The viable spore count of *pyrE*$^{-}$ 2 and *pyrE*$^{-}$ 3 was approximately 560000- and 130-fold lower than WT, respectively. The third biological replicate of each mutant (*hsdR*1600s::CT 3 and *pyrE*$^{-}$ 1) did not generate any viable spores. These inexplicable viable spore count reductions strongly resemble what has been previously observed for the genetically unstable ATCC 3502 IFR strain. Possibly the ATCC 3502 Unilever strain also suffers from an unusually high mutation rate, contrary to what was anticipated based on previous work from other researchers using this strain.

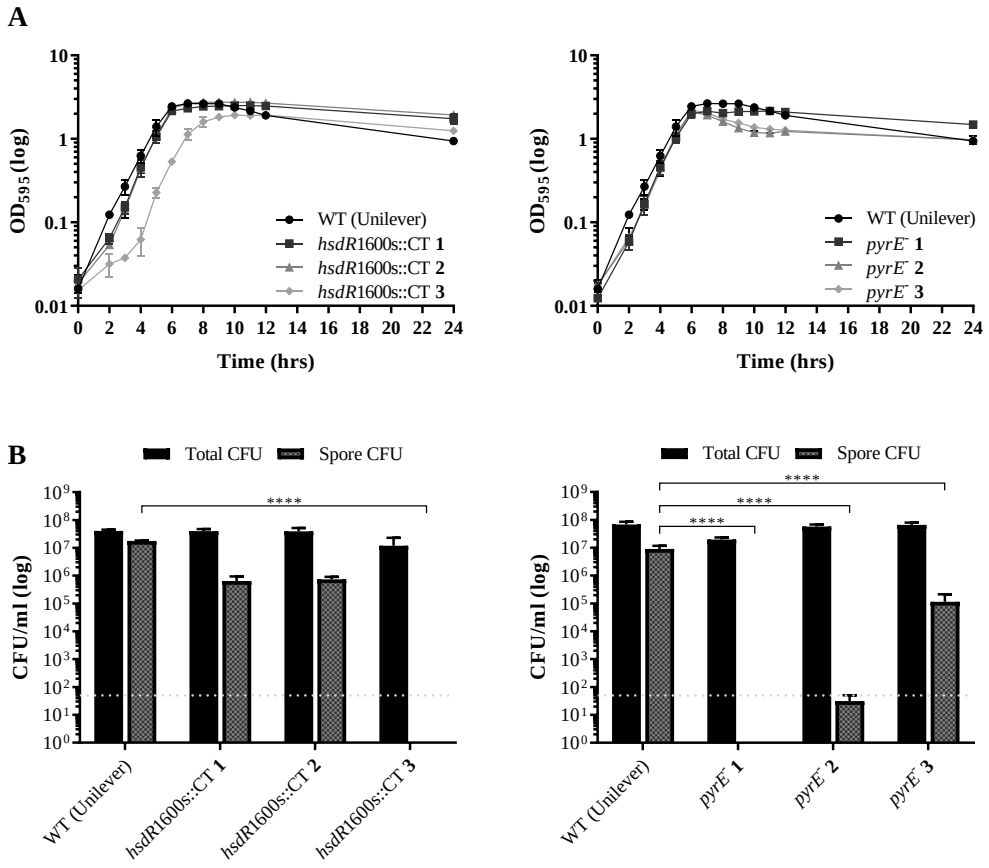


Figure 4.3 Phenotypic characterisation of ATCC 3502 Unilever *hsdR*1600s::CT and *pyrE*⁻ mutants. (**A**) Growth profiles of mutants compared to WT. Strains were cultured in TYG broth and the OD_{595} was monitored for 24 hours. Error bars represent the standard deviation of three independent cultures from one representative experiment out of two repeats. (**B**) Sporulation/germination capacity of mutants compared to WT after 120 hours of incubation in TYG broth. Total CFU corresponds to the total cell count (non-heat treated) and the spore CFU is the number of viable spores withstanding a 15 min, 80°C heat treatment. Error bars represent the standard deviation of three independent cultures from three separate assays. The dotted line at 50 CFU/ml represents the detection limit. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

To confirm this suspicion, the genome sequence of all six mutants was determined using Illumina MiSeq sequencing. The MiSeq paired end reads were mapped to the WT ATCC 3502 genome sequence deposited at the NCBI database (accession number: NC_009495) and variations identified using the CLC genomics workbench. In addition, the full genome of the ATCC 3502 WT Unilever stock (green), from

which the mutants were generated, was sequenced. It is common practice within the Clostridial research group that all strains are stored in triplicates (in a green, yellow and red tube), which is referred to as the 'traffic light' system. The green stock is the working stock, used for all ordinary laboratory experiments such as measuring growth or creating mutants. The yellow stock is only used to grow a bacterial strain to create a fresh green stock whenever a green stock runs out. The red stock serves as the ultimate back-up in case something happens with the other two stocks. The SNPs and InDels detected in the WT and mutant strains are listed in Appendix VI. SNPs and InDels detected in a region with low coverage (>50 reads) or with a low frequency (>75%) were manually checked to confirm the SNP or InDel is not an artefact resulting from alignment errors by the software. 'False' SNPs and InDels were removed from the mutation lists in Appendix VI.

In the WT Unilever stock nine variations were detected compared to the published NCBI genome (accession number: NC_009495) and all but one of the mutant strains had a significant number of additional mutations compared to WT. Surprisingly, not all single nucleotide variations between the NCBI ATCC 3502 genome and the WT Unilever genome were detected in every mutant strain. Only two point mutations, at positions 130004 (a silent SNP in the isoleucyl-tRNA synthetase) and 3724684 (a silent SNP in elongation factor Tu), were present in all strains. This suggests that there is considerable genetic cell-to-cell variation in the WT stock. When such a heterologous population is sequenced with a coverage ranging from 50 to 100 reads per base, only the most prevalent SNPs will show. Additionally, the relatively low frequency values (% of reads with the SNPs out of the total number of reads) for most of the mutations detected in the WT stock further supports this theory. The frequency of a point mutation in WT is often below 80%, whereas in mutant strains, which have

been through several single colony stages and are thus less likely to be a mixture, this value is generally above 95%. Overall, the sequencing data effectively confirmed that the genome instability previously observed in the ATCC 3502 IFR strain is also present in the ATCC 3502 Unilever strain.

A total of 124 unique mutations were identified, with 56 of these occurring in two or more strains. However, 49 of the 56 non-unique mutations were present in strains *pyrE*$^{-}$ 2 and *pyrE*$^{-}$ 3, with 38 SNPs only shared between these two strains. Given that the three *pyrE*$^{-}$ mutants were isolated from a single conjugal mixture and the *pyrE*$^{-}$ 2 and *pyrE*$^{-}$ 3 strains have many mutations in common, it is plausible that both mutants originated from the same transconjugant. Despite the high level of similarity between their SNP profiles, both strains have a distinctly different sporulation/germination capacity (a 3700-fold reduction for *pyrE*$^{-}$ 2 strain and 'only' a 130-fold reduction for the *pyrE*$^{-}$ 3 strain compared to the WT) (Figure 4.3B). The *pyrE*$^{-}$ 2 strain has four unique point mutations; one of them is a frameshift mutation in the *gerAB1* gene, which results in a premature stop codon in the mRNA of the encoded protein, a germinant receptor that responds to L-alanine.[518-521] The truncated receptor is probably inactive and is most likely causing the observed reduction in sporulation/germination capacity. In the *pyrE*$^{-}$ 3 strain only two unique mutations were identified. There is, however, no obvious connection between the affected genes and spore formation or germination processes that could explain the observed reduction in viable spore count compared to WT. Furthermore, none of the other 46 mutations were in genes that encode proteins which are known to be directly involved in spore formation or germination. Perhaps the reduction in viable spores is an indirect effect, resulting from the disruption of other cell processes by one or several point mutations.

The *pyrE*$^{-}$ 1 strain did not produce any detectable number of spores. This is probably due to a SNV in a gene annotated as *spoIV*, which creates a premature stop codon at position 21 of the normally 386 amino acid long polypeptide. The *spoIV* gene encodes a stage IV sporulation protein controlled by σ^{E}, known as YqfD, and disruption of this protein in *B. subtilis* results in a strain only producing phase-dark spores that fail to turn into fully mature phase-bright spores.[522]

Although the 10-fold reduction in viable spore count for strains *hsdR*1600s::CT 1 and 2 was not statistically significant, spore counts were consistently lower compared to WT. Analysis of the mutations detected in both strains did not provide any clues as to what could be responsible for this decrease. Moreover, only 4 additional SNPs were identified in *hsdR*1600s::CT 2, which is very little compared to the other mutants. It was hypothesized that if the WT stock is truly a heterogeneous mixture of cells due to elevated mutation rates, there might be a subpopulation within the stock with a higher sporulation or germination frequency. This would result in an overestimation of the sporulation/germination capacity of the WT strain and consequently leads to the incorrect assumption that strains *hsdR*1600s::CT 1 and 2 have a reduced viable spore count due to genetic alterations. To establish if such a subpopulation exists within the WT stock further analysis, for example, by determining the sporulation and germination capacity of single cells, would be required.

Finally, in cultures of strain *hsdR*1600s::CT 3 no viable spores were detected and analysis of its genome sequence revealed that two sporulation associated genes contain a mutation: *lonB* (σ^{F} regulation) and *spoIIP* (peptidoglycan degradation). The mutation within the *lonB* gene is silent, changing leucine codon TTA into TTG. Usually a single silent mutation in a gene has no consequences, though there are occasionally cases where a silent SNV causes an alteration in protein expression and

activity, especially when there is a strong codon usage bias (the difference in frequency of occurrence of codons that encode the same amino acid in a genome).[523] It is thought that codon usage influences protein expression and structure through its effect on translation speeds, mRNA stability, secondary DNA structures, translation accuracy and co-translational protein folding.[524] In this case the difference in codon frequency is quite large (TTA 53.3 *vs* TTG 8.3), but given that in a previous study, by Serrano *et al*., the inactivation and overexpression of *lonB* did not affect spore formation in *B. subtilis*, it seems very unlikely that this mutation has contributed to the asporogeneous phenotype.[525-526]

The second affected sporulation gene, *spoIIP*, contains a 1-bp insertion which shifts the open reading frame (ORF) and results in a premature stop codon, halving the normally 363 amino acid polypeptide. This truncated version of SpoIIP is probably inactive. SpoIIP is part of a peptidoglycan degradation complex, together with SpoIID and SpoIIM, which function is critical for successful engulfment of the forespore.[526-527] Loss of function in any of the above proteins prevents engulfment completion and hence, no heat resistant spores are formed.

For three out of the six genetically modified strains random mutations have disrupted genes that encode proteins involved in the process of either sporulation or germination. Nevertheless, a potential contribution of the other SNPs to the reduced viable spore counts or asporogeneous phenotypes cannot be excluded, despite the lack of an apparent connection between the mutations and sporulation or germination. Future examination of the spores under a microscope could provide more information concerning the stage at which spore formation or outgrowth has been disrupted and if this resembles the phenotype of strains with an inactivated *gerAB1*, *yqfD* or *spoIIP* gene.

These above findings led to the conclusion that creating a safe ATCC 3502 strain using ACE would not yield a worthy WT surrogate. By the time that the three sequential genetic modification steps required to generate a safe strain have been completed (*pyrE*$^{-}$, *pyrE*$^{-}$ Δ*boNT/A1* and finally *pyrE*$^{+}$ Δ*boNT/A1*), the strain will have accumulated many SNPs, which will undoubtedly affect its phenotype to some extent.

4.3.2 Deletion of the neurotoxin in the genetically unstable ATCC 3502 strain using CRISPR/Cas9-based technology

Based on the above findings, it appears that it is virtually impossible to create a non-toxicogenic version of ATCC 3502 without compromising other properties, making them less appropriate WT surrogates. However, the severity of its genetic instability is possibly amplified by the used engineering tools. ClosTron mutagenesis involves three single colony stages and several additional culturing steps. In other clostridial species this lengthy procedure does not results in a high number of additional random mutations, nevertheless in a strain like ATCC 3502, which appears to have a relatively high basal mutation rate, the high number of steps required to isolate a mutant can be detrimental to maintaining genome integrity (personal communication; N. Minton). The same applies to ACE, where the multitude of steps and single colony stages required increases the likelihood of picking up unwanted random mutations. Additionally, to isolate mutants the cells are grown on solid media supplemented with 5-FOA (see section 1.5.2), which is slightly toxic to the cells, depending on its concentration. The exposure of cells to toxic compounds can induce a general stress response that leads to elevated spontaneous mutation rates.[423] An unusually high number of additional random mutations has not been observed in other clostridial species after 5-FOA exposure, however, the concentration of 5-FOA required to select for *pyrE* negative cells varies between clostridial species; for example, for *C.*

beijerinckii a concentration of 800 µg/ml is sufficient, whereas for *C. sporogones* a concentration of 3.5 mg/ml is required.[206] To select for *pyrE* negative ATCC 3502 cells an exceptionally high concentrations of 5-FOA is required (5.5 mg/ml), but concentrations above 6 mg/ml are lethal to ATCC 3502 *pyrE*$^{-}$ mutants. WT ATCC 3502 tolerates 5-FOA concentrations as high as 5 mg/ml but colony morphology and growth rate are clearly affected, suggesting that the cells are experiencing some level of stress.

At the start of this project, ClosTron mutagenesis and marker-free engineering via ACE were the only available genetic tools for clostridia. Halfway through this project, however, a CRISPR/Cas9-based genome editing tool became available, which involves only 2 single colony stages, less culturing steps and no exposure to toxic compounds that are potentially mutagenic.[204, 227, 528] Additionally, with the CRISPR/Cas9 tool the *boNT/A1* gene can be deleted directly (one step), contrary to the indirect, three step-based ACE strategy. These advantages together with the knowledge that one of the *hsdR*1600s::CT mutants had just four additional SNPs led to the idea that adequate non-toxicogenic strains could be created using the CRISPR/Cas9 tool.

To test this hypothesis the *boNT/A1* gene was deleted, starting from the 'red' culture collection stock using CRISPR/Cas9, as described in section 2.12.3, and plasmid pMTL431511_boNT/A1-G1, whose construction is outlined in section 2.11.2.1. The deletion was confirmed by PCR in three biological replicates from three independent conjugations for each strain using primers annealing to the chromosome, flanking the editing template (Figure 4.4). The plasmid was lost from the selected mutants, as described in section 2.8 and the deletion in the final mutants was subsequently verified by sequencing.

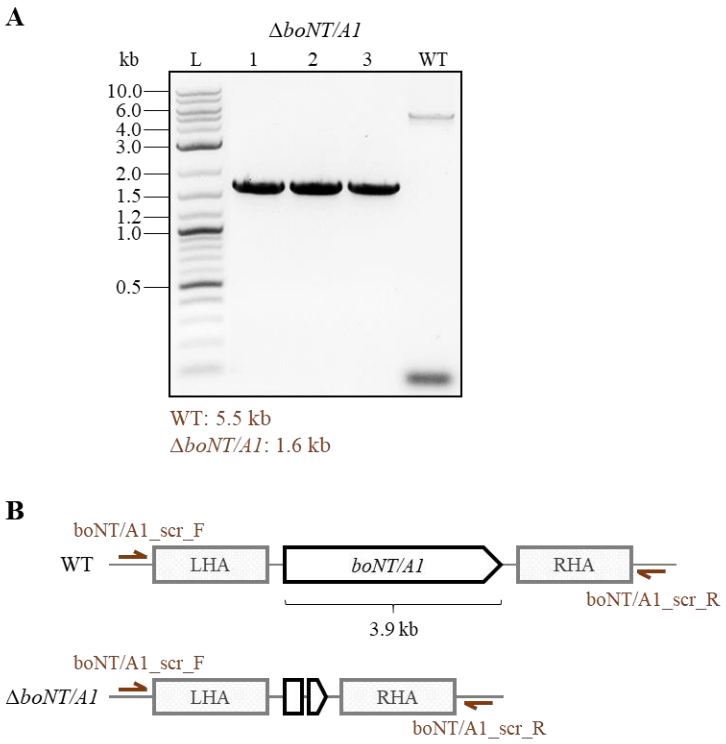


Figure 4.4 Deletion of *boNT/A1* gene in strain ATCC 3502. (**A**) PCR for three biological replicates of ATCC 3502 *pyrE*$^{-}$ mutants alongside WT. Products were amplified using flanking primers boNT/A1_scr_F and boNT/A1_scr_R. Lane L: NEB 1 kb Plus DNA Ladder. (**B**) Schematic diagram of the targeted chromosomal region showing the primer binding sites in the WT and mutant strain.

To establish if phenotypic characteristics of the WT strain had not been compromised by additional random mutations, the growth and sporulation/germination capacity of the Δ*boNT/A1* strains was assessed as described in sections 2.9.2 and 2.9.4, respectively. The growth profiles of all three mutants were similar to WT (Figure 4.5A). The sporulation/germination capacity however, varied among the safe strains (Figure 4.5B); mutants Δ*boNT/A1* 1 and 2 yielded approximately 130- and 36,000-fold less viable spores than WT, while the third mutant was not compromised in its ability to form heat resistant spores. Deletion of the neurotoxin in other *C. botulinum* strains did not have a negative impact on the total number of viable spores; presumably the reduction in sporulation/germination capacity of the mutants is again caused by

additional random mutations, even though culturing steps and single colony stages were kept to an absolute minimum.[126, 529]

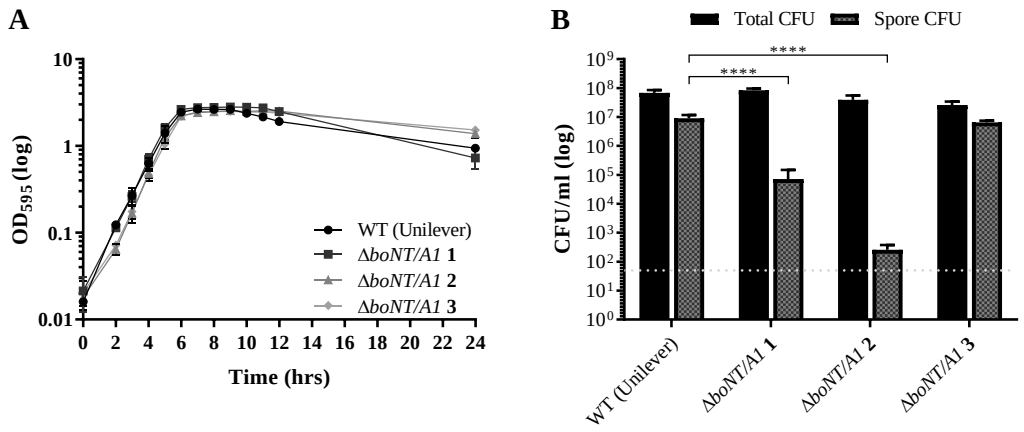


Figure 4.5 Phenotypic characterisation of ATCC 3502 Unilever Δ*boNT/A1* mutants. (**A**) Growth profiles of mutants compared to WT. Strains were cultured in TYG broth and the OD_{595} was monitored for 24 hours. Error bars represent the standard deviation of three independent cultures from one representative experiment out of two repeats. (**B**) Sporulation/germination capacity of mutants compared to WT after 120 hours of incubation in TYG broth. Total CFU corresponds to the total cell count (non-heat treated) and the spore CFU is the number of viable spores withstanding a 15 min, 80°C heat treatment. Error bars represent the standard deviation of three independent cultures from three separate assays. The dotted line at 50 CFU/ml represents the detection limit. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

To confirm that the reduction in viable spores was caused by point mutations, the full genome of each Δ*boNT/A1* replicate was sequenced, as described in section 2.10.10. In addition, the full genome of the ATCC 3502 WT Unilever 'red' stock, from which the mutants were generated, was sequenced. Genetic variations were identified using the CLC genomics workbench and are listed in Appendix VI Table A 6.

Nine additional point mutations were detected in the WT genome compared to the published reference genome (accession number: NC_009495); the red and green stock have eight mutations in common and both contain one unique mutation, albeit their

frequency was rather low. Not all the SNPs detected in the WT genome were present in each mutant, as previously observed for the *hsdR*1600s::CT and *pyrE*$^{-}$ mutants.

The two mutants with a reduced viable spore count both contained 33 additional SNPs, yet none of the mutations were in ORFs annotated as sporulation or germination genes. Nevertheless, several candidates were identified based on their (functional) homology to genes found in genome-wide studies of the sporulation and germination pathway in *C. difficile* and *C. sporogenes* (Unpublished data; R. Rodriques).[530-531] Homologues of the mutated genes *serS*, CBO1361 (manganese/zinc/iron chelate uptake ABC transporter permease) and *ppdK* in strain Δ*boNT/A1* 1 appeared to be essential for sporulation or germination in the *C. sporogenes* TraDIS library. Additionally, functional homologues of genes *serS*, CBO0719 (methyl-accepting chemotaxis protein), *chiA* and CBO1414 (ATP-dependent DNA helicase) occurred in the TraDIS and RNA-Seq libraries of *C. difficile*. For strain Δ*boNT/A1* 2 only gene CBO1580 (Pyridine nucleotide-disulfide oxidoreductase) was homologous to a *C. sporogenes* TraDIS library germination gene and genes CBO0465 (NUDIX family: hydrolase), *opuD*, CBO1991 (PTS system fructose family transporter subunit IIBC) had a similar description to genes identified in the various *C. difficile* studies. Furthermore, the mutated kinases CBO0906 (Δ*boNT/A1* 1), CBO0780 and CBO1967 (Δ*boNT/A1* 2) were listed as potential candidates. Most of the above have not been described in literature in relation to sporulation or germination, though there are some indications that genes such as the methyl-accepting chemotaxis protein and manganese/zinc/iron chelate uptake ABC transporter permease might affect sporulation and germination.[532-533] To establish which mutations actually contribute to the reduction in viable spores, they will have to be re-created in the WT strain; first independently and then, depending on the results, in various combinations.

The third replicate (Δ*boNT/A1* 3), which was not compromised in its ability to form viable spores, only contained 5 additional SNPs. Interestingly, the SNP profile of Δ*boNT/A1* 3 was identical to *hsdR*1600s::CT 2, except for mutation 3641879. Given that the chance both strains acquired four identical mutations independently is low, it was assumed that they have a common ancestor. At the time, the odd lack of additional mutations was not given much thought and was attributed to mere coincidence, however, at a later stage this turned out to be far more interesting (see section 4.3.3).

The above results show that despite the advantages the CRISPR/Cas9 based method offers over the other two tools, using this technology does not reduce the number of random SNPs accumulated. In conclusion, the observed genomic instability seems to be inherent to the strain and its predisposition for the acquisition of mutations is not aggravated by the used tool, as evidenced by the similar number of SNPs identified in the mutant strains generated using ClosTron, ACE and CRISPIR/Cas9 mediated mutagenesis. It might be worth contemplating if continuing research with a safe strain such as Δ*boNT/A1* 3 or even the WT strain is justifiable, knowing that few culturing steps already generates a hodgepodge of genetic variation within the population and any further downstream processing might result in significant phenotypic changes. This is however not a question that can be answered in a single PhD thesis, and will be a separate, more resource intensive investigation.

4.3.3 First attempt to establish the mutation frequency of ATCC 3502

Whole genome sequencing methods are rarely used to quantify a bacterial strain's genetic stability and establish its mutation rate. Despite recent advancements within the field, such as Maximum-Depth Sequencing (MDS) and the visualisation of emerging mutations using single cell analysis coupled to high-throughput sequencing,

the detection of random low frequency mutations in bacteria remains notoriously difficult and data might be skewed by mutational 'hot' or 'cold' spots.[534-536] In addition, the sequencing based methods are expensive and especially the single-cell based strategies require state-of-the-art technologies and are very laborious.

Instead, the 'old fashioned' relatively rapid and inexpensive fluctuation assay, which is based on the loss of sensitivity to a selective marker (e.g., antibiotic resistance), is still the most common method to estimate spontaneous mutation rates.[537-538] A large number of parallel cultures are grown in the absence of the selective marker and after many generations the cultures are plated onto non-selective and selective agar. The CFUs from the non-selective agar is used to estimate the total number of cells that could potentially acquire a mutation (N_t). The distribution of the observed number of resistant mutants per culture (r) is used to estimate the mean number of mutational events per culture (m) by applying the appropriate mathematical model. Subsequently, the mutation rate (μ) is usually calculated by dividing m (the mean number of mutations per culture) by some measure of N_t (the final number of cells per culture).

However, calculating the mutation rate can be problematic and the accuracy of the estimated rate is highly dependent on the experimental set-up (e.g., starting CFU, number of mutants per culture (r) and number of parallel cultures) as well as the mathematical models used to calculate m and μ. Therefore, many researchers opt for reporting the average mutation frequency (f), which is simply the number of resistant mutants (r) divided by the final number of cells (N_t) per parallel culture. Based on this value strains can be classified as hypomutators ($f \leq 8 \times 10{-}9$), normomutators ($8 \times 10^9 < f < 4 \times 10^8$), weak mutators ($4 \times 10^8 \leq f < 4 \times 10^7$), and strong mutators ($f \geq 4 \times 10^7$).[384-386] In literature these ranges can vary, depending on the organism studied and references used. Additionally, sometimes the term hypermutator is used, which can

refer to weak mutators and strong mutators together or only refer to strong mutators. Here the term hypermutator is used to refer to strong mutators only.

The fluctuation assay, as described in section 2.9.6, using rifampicin as a selection marker (rifampicin resistance occurs due to mutations in the RNA polymerase β subunit gene *rpoB*), was designed based on results of preliminary experiments to establish the required number of parallel cultures, culture volume and inoculum.[539] These initial experiments were conducted with both ATCC 3502 and ATCC 19397. While the number of observed resistant mutants per culture (*r*) was consistent amongst the ATCC 19397 colonies (data not shown) – independent of the experimental set-up – for ATCC 3502 *r* varied strongly between the different colonies (Figure 4.6A). It should be noted that not all colonies were treated equally and thus a direct comparison between the three different groups (1-5, 6-10, 11-13) should be undertaken with caution. Nonetheless, there is a clear-cut difference between the six colonies which are within the hypomutator range and the remainder of the colonies, which can be classified as weak or even strong mutators. This strong variation suggests that the ATCC 3502 population is mixed and possibly contaminated with a non-mutator strain, such as ATCC 19397. To establish if all thirteen colonies are ATCC 3502 or if the stock is contaminated with ATCC 19397, 62A, Loch Maree or NCTC 2916 (all strains that had been cultured in the same anaerobic cabinet), strain specific PCRs were performed using the primers listed in Appendix IV. The strain specific PCRs confirmed that all colonies were strain ATCC 3502 (data not shown).

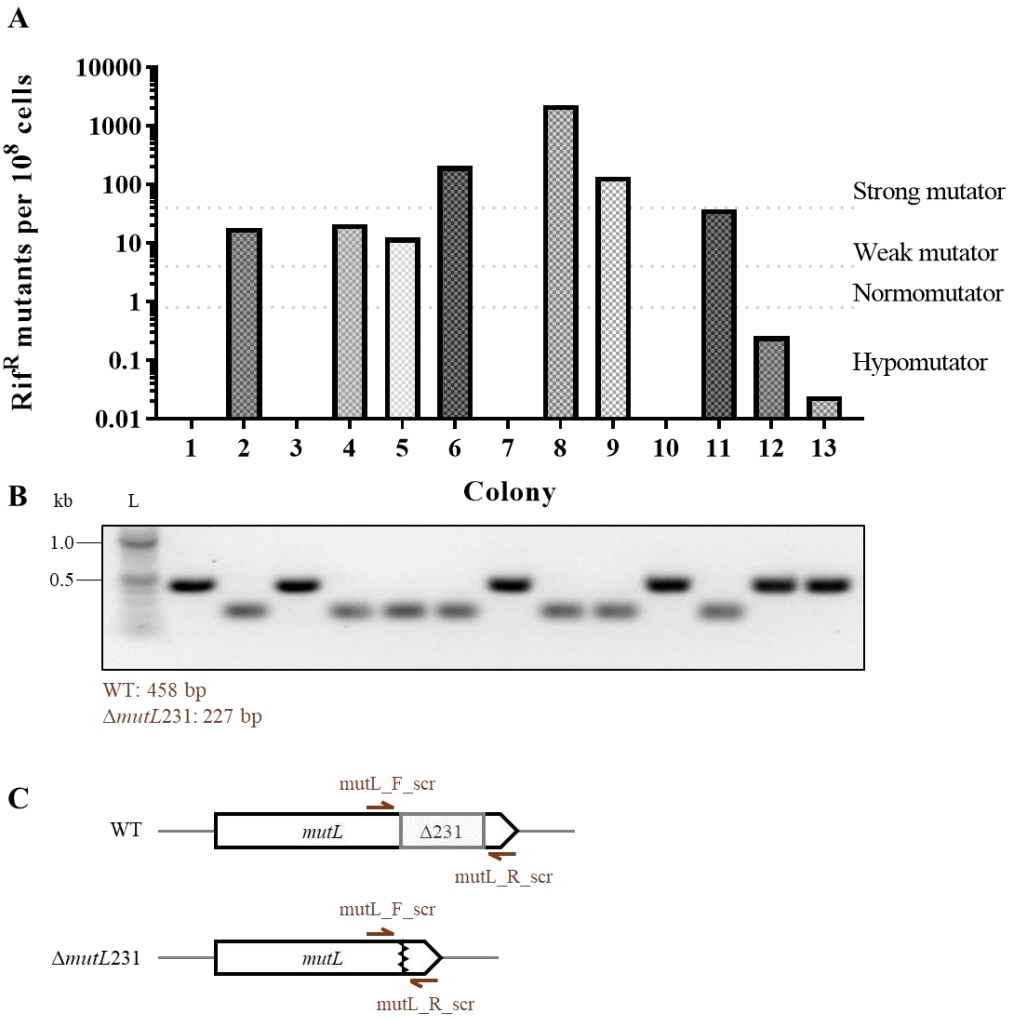


Figure 4.6 Preliminary fluctuation assay with ATCC 3502. (**A**) Results from preliminary fluctuation assay experiments using thirteen different 'WT' colonies to inoculate the parallel cultures. Note: experimental set-up varied between colonies 1-5, 6-10 and 11-13. (**B**) PCR for the thirteen ATCC 3502 'WT' colonies. Products were amplified using flanking primers mutL_F_scr and mutL_R_scr. Lane L: NEB 100 bp DNA Ladder. (**C**) Schematic diagram of the targeted chromosomal region showing the primer binding sites in the WT and the Δ*mutL*321 mutant strain.

Intrigued by the variation in genomic stability of ATCC 3502 isolates, its published genome sequence (accession number: NC_009495) was compared to the genome sequence of the genetically stable and highly similar ATCC 19397 strain (accession number: NC_009697), with the aim to identify possible clues towards elucidating the underlying cause of this strange behaviour (see Chapter 5 for a more detailed

comparison of the two strains). Usually strong mutator phenotypes are a consequence of defects in DNA repair or replication systems.[375, 540] Prior to the alignment, genes potentially important for maintaining the genome integrity in ATCC 3502 were identified based on the available literature. The full genome comparison was carried out manually using Mauve software (data not shown).

Two interesting variations were detected; the first being a SNP in *recO*, a gene that encodes a recombination mediator protein.[541] The SNP in *recO* results in an amino acid change (E200A). Among the type A strains in the Nottingham culture collection this glutamic acid at position 200 is conserved, however, a quick BLASTp analysis revealed that this amino acid is not conserved across different species. Predicting the effect of such a single amino acid change on protein functionality is difficult and therefore to establish its impact the mutation would have to be recreated in a genetically stable type A strain. Further analysis of the *recO* mutation was not pursued, mainly because the second mutation, discussed in more detail below, was considered to be the more likely cause of the genetic instability of strain ATCC 3502.

The second mutation is a 231 bp deletion in *mutL*, which encodes a DNA mismatch repair protein.[542] The mutation results in an in-frame deletion of 77 amino acids in the normally 666 amino acid long protein. This large deletion most likely has a negative effect the protein's functionality. Previous studies have shown that reduced MutL activity levels lead to elevated mutation rates.[419, 422, 452] Interestingly, alignment of the ATCC 3502 Unilever WT stock Illumina whole genome sequencing reads against the ATCC 19397 WT sequence (NCBI accession number: NC_009697) revealed that some reads map to this 231 bp region whereas other reads are cut-off (Figure 4.7). This suggest that the stock is a mixed population, consisting of cells carrying the full *mutL* gene and cells carrying the mutated *mutL* gene.

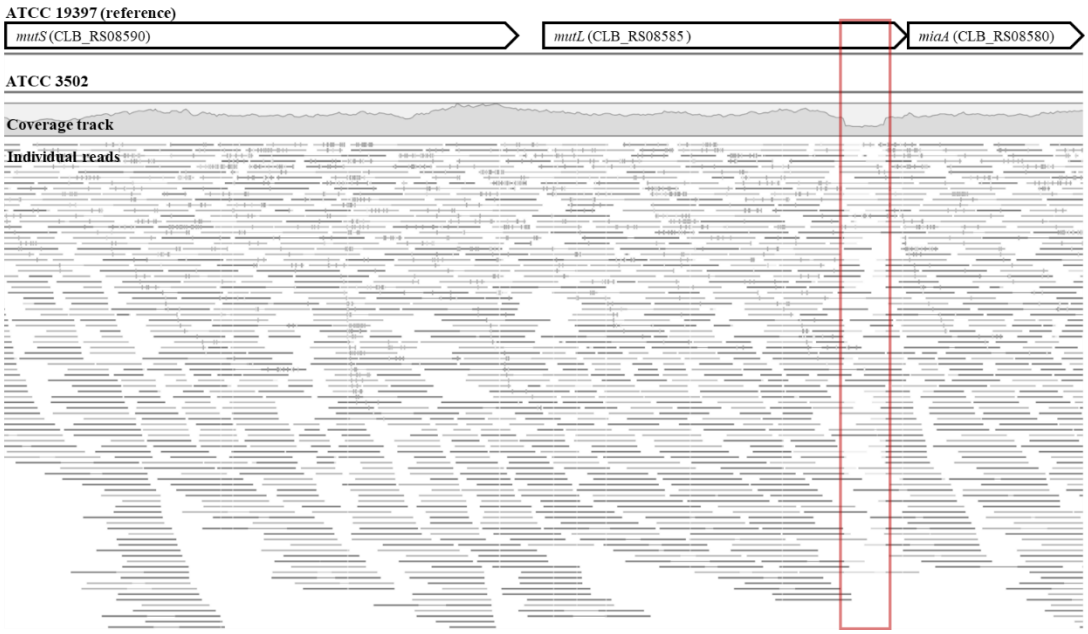


Figure 4.7 Visualisation of the Δ*mutL*231 deletion identified in ATCC 3502. Visualisation of individual ATCC 3502$^{+/-}$ NGS reads mapped to the ATCC 19397 *mutL* genome region using CLC Genomics Workbench. The deletion was identified by comparing the published ATCC 3502 genome sequence to the published ATCC 19397 genome sequence using Mauve. Coverage track: total number of reads for a specific nucleotide position. Individual reads: individual paired and non-paired sequencing reads aligning to the depicted genomic region of ATCC 19397. Within the red box is the 231 bp deletion region overlapping with *mutL*; some aligned reads reach into the deletion region whereas other reads abruptly stop (dark reads that turn transparent – the lighter coloured section of the read does not align to the reference).

To assess if there is a correlation between the mutated *mutL* gene and hypermutability in strain ATCC 3502, all thirteen colonies included in the initial fluctuation assays described previously were screened for the presence of the 231 bp deletion using primers flanking the deletion site (Figure 4.6B/C). There was a perfect match between the presence of the *mutL* deletion (Δ*mutL*231) and the mutator phenotype. All colonies with a full WT *mutL* gene were non-mutators.

Finally, the genome sequences of the ATCC 3502 mutants described in sections 4.3.1 and 4.3.2 were reanalysed using the ATCC 19397 genome as a reference. This revealed that the mutant strains where viable spore counts were strongly compromised

(*hsdR*1600s::CT 3, *pyrE*$^{-}$ 1, *pyrE*$^{-}$ 2, *pyrE*$^{-}$ 3, Δ*boNT/A1* 1 and Δ*boNT/A1* 2) all carried the *mutL* gene with the 231 bp deletion, whereas all other strains (*hsdR*1600s::CT 2 and Δ*boNT/A1* 3), except strain *hsdR*1600s::CT 2, carried the full *mutL* gene. This variation in *mutL* went unnoticed because the used reference genome of ATCC 3502 (accession number: NC_009495) contains the 231 bp deletion and therefore no clear disruption is visible in the read mapping track displayed by CLC Genomics Workbench. Up until this point the re-sequenced genomes were never checked for large insertions or deletions using the InDels variant detection functionality available on CLC Genomics Workbench, which in hindsight was a mistake.

Furthermore, the mutant strains with the disrupted *mutL* gene carried 29 or more random SNPs compared to only 12 and 13 mutations in *hsdR*1600s::CT 2 and Δ*boNT/A1* 3, respectively (see Appendix VI). Moreover, the SNP profiles of *hsdR*1600s::CT 2 and Δ*boNT/A1* 3 were almost identical; strain Δ*boNT/A1* 3 carried one additional mutation compared to *hsdR*1600s::CT 2, while the other 12 SNPs detected were the same in both strains. The two mutants were isolated from independent experiments and it is highly unlikely that both strains acquired exactly the same mutations independently. Probably both mutants originated from the same ancestor. This genetically stable ancestor with a full *mutL* gene could be considered the 'true' WT genome, instead of the published ATCC 3502 genome (accession number: NC_009495), which is likely comprised of the most common nucleotide sequences for each locus in the mixed population at the time of sequencing. In section 4.3.8 the mutator and non-mutator WT genomes will be discussed in more detail.

The data presented here were the first results that provided a logical explanation for the mysterious reliability issues and unusual genetic instability that had been haunting the lab for many years.

To distinguish between the different ATCC 3502 WT genotypes and stock compositions the following names were given to the strains:

- ATCC 3502^{+} – The genomically stable portion of the population or isolates that carry a full *mutL* gene.
- ATCC 3502 Δ*mutL*231 (or ATCC 3502^{-}) – The mutator portion of the population or isolates that carry the mutated *mutL* gene with the 231 bp deletion.
- ATCC 3502 (or ATCC $3502^{+/-}$) – The mixed ATCC 3502 population consisting of both mutator and non-mutator ATCC 3502 cells.

The above naming system has been implemented consistently across all chapters in this thesis.

4.3.4 Investigating the role of DNA mismatch repair protein MutL

The *mutL* deletion was recreated in ATCC 19397 as a means of establishing that the observed hypermutability in the ATCC 3502 Δ*mutL*231 strains is solely due to this mutation and not due to a set of different additional mutations. Additionally, the full *mutL* gene was deleted in ATCC 19397 and the resulting mutation frequency compared to the Δ*mutL*231 ATCC 19397 strain to assess if the truncated MutL protein in ATCC 3502 might have any residual activity.

To generate the Δ*mutL* and Δ*mutL*231 ATCC 19397 strains CRISPR/Cas9-mediated mutagenesis was carried out as described in section 2.12.3, using plasmids pRECas2-mutL-G1d and pRECas2-mutL231-G1d, respectively.[158] Construction of the two plasmids used to create these strains is outlined in section 2.11.2.2. The deletions were confirmed by PCR in three biological replicates from three independent conjugations

using primers annealing to the chromosome, flanking the insertion site (Figure 4.8). The plasmid was lost from the selected ATCC 19397 mutants, as described in section 2.8, and presence of the deletion in the final mutants was subsequently verified by Sanger sequencing.

In order to confirm that the phenotype observed in the mutant strain is directly attributable to the different *mutL* deletions, *mutL* complemented strains Δ*mutL*^{comp} and Δ*mutL*231^{comp} were generated via CRISPR/Cas9-mediated genome editing, incorporating *mutL* at its original chromosomal position. Plasmids pRECas2-mutL_comp-G2c and pRECas2-mutL_comp-G1c were used to create the complemented strains, whose construction is outlined in section 2.11.2.2. The chromosomal Cas9 target site, just downstream of *mutL* and inside a coding region (*miaA* gene), is also present in one of the homology arms that facilitates the insertion of *mutL*. To prevent self-cleavage of the editing template silent mutations were introduced in the PAM and SEED sequence present in the homology arm. The codon changes might affect the rate of translation, RNA stability and protein folding, which could cause changes in the protein's expression and/or activity levels.[523] Therefore, only nucleotide changes resulting in a codon with a similar frequency of occurrence were incorporated.

To screen for successful insertion events PCRs were performed using primers annealing to the chromosome, flanking the insertion site (Figure 4.8). The three biological replicates for each mutation were isolated from three independent conjugations. The plasmid was lost from the selected mutants, as described in section 2.8, and the genomic modifications in the final mutants were subsequently verified by Sanger sequencing.

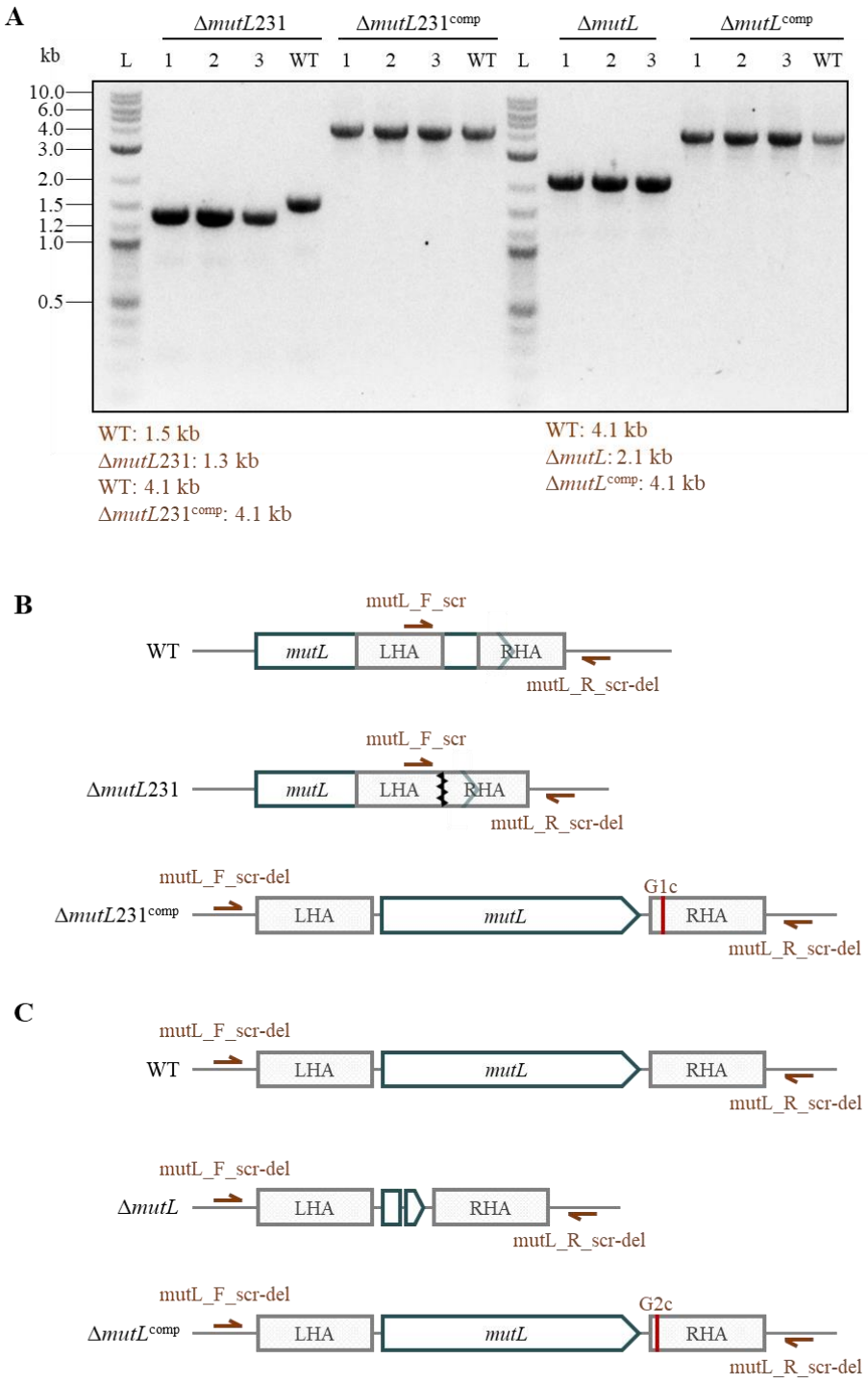


Figure 4.8 Deletion and complementation of *mutL*231 and *mutL* in strain ATCC 19397. (**A**) PCR for three biological replicates of ATCC 19397 Δ*mutL*231, Δ*mutL*231^{comp}, Δ*mutL* and Δ*mutL*^{comp} alongside WT. Products were amplified using primers mutL_F_scr and mutL_R_scr-del and mutL_F_scr-del and mutL_R_scr-del. Lane L: NEB 1 kb Plus DNA Ladder. (**B/C**) Schematic diagram of the targeted genomic region showing the primer binding sites in the WT and mutant strains.

To establish if growth under standard culturing conditions had not been compromised in the *mutL* mutant strains, the growth of the mutants was assessed as described in sections 2.9.2. The growth profiles of all mutants were quite similar to WT during the exponential phase, however, during the stationary and death phase the growth profiles of all mutants started to slightly deviate from the WT profile (Figure 4.9). Overall, the *mutL* mutants were able to reach a higher maximum OD_{595} than WT which was followed by a more abrupt death phase. After 24 hours all strains, including WT, had settled at a similar OD_{595}. The ability of the *mutL* mutants to reach a higher maximum OD_{595} could affect the results of the fluctuation assay; a higher maximum cell density allows for more generations in the bacterial population within the same timeframe, thus increasing the chances of acquiring a random mutation that provides resistance to the antibiotic rifampicin. Especially for the complemented strains, where rifampicin resistant CFUs are expected to be low, this could have a significant impact on the calculated mutation frequency and should therefore be taken into consideration when analysing the fluctuation assay data discussed below.

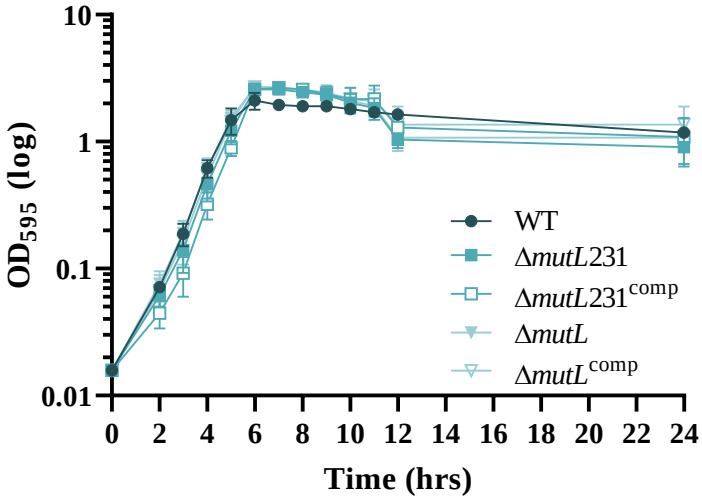


Figure 4.9 Growth profiles of ATCC 19397 *mutL* mutants. Growth profiles of ATCC 19397 Δ*mutL*231, Δ*mutL*231^{comp}, Δ*mutL* and Δ*mutL*^{comp} compared to WT. Strains were cultured in TYG broth and the OD_{595} was monitored for 24 hours. Error bars represent the standard deviation of three independent cultures of each biological replicate.

To assess the effect of the various *mutL* mutations on the genomic stability of ATCC 19397 the mutant strains mutation frequency was established using the previously designed fluctuation assay, as described in section 2.9.6. The WT ATCC 3502^{+} and ATCC 3502 Δ*mutL*231 strains were assayed alongside as a reference. As expected, mutation frequencies significantly increased in the ATCC 19397 *mutL* mutants compared to WT, confirming that a deficiency in MutL activity elevates the mutation rate (Figure 4.10). Complementation of the *mutL* gene essentially returned the mutation frequency back to WT levels. Furthermore, the mutation frequency of strain ATCC 19397 Δ*mutL*231 was similar to strain ATCC 3502 Δ*mutL*231, which supports the hypothesis that the hypermutator phenotype within the WT ATCC 3502 population is caused by this specific deletion within *mutL*. Finally, there was no clear difference between the mutation frequency of the Δ*mutL*231 strains and the ATCC 19397 strains with the full *mutL* deletion, suggesting that the 231 bp deletion in *mutL* results in a non-functional protein.

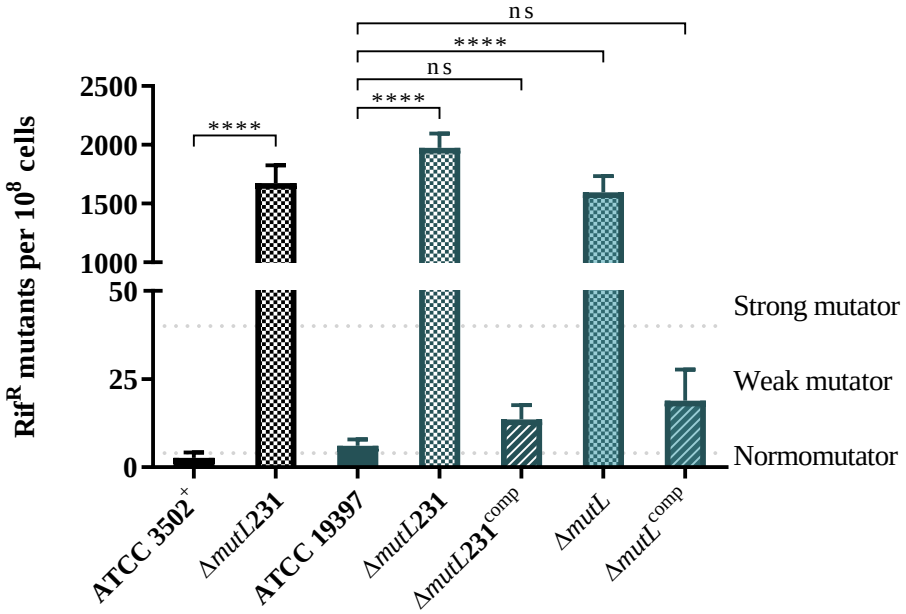


Figure 4.10 Mutation frequency of ATCC 3502 and ATCC 19397 *mutL* mutants. Frequency of spontaneous Rif^R mutants in strains ATCC 3502 WT and Δ*mutL*231 (black) and ATCC 19397 WT, Δ*mutL*231, Δ*mutL*231^{comp}, Δ*mutL* and Δ*mutL*^{comp} (blue). Cells were cultured for 24 hours in TYG broth and subsequently plated onto TYG (total CFU) and TYG supplemented with 25 µg/ml rifampicin (Rif^R CFU). Mutation frequencies were calculated by dividing the Rif^R CFU by the total CFU. The dotted lines indicate the mutator strength categories. Error bars represent the standard error of the mean of 10 single colonies, cultured in triplicates. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, nonparametric Mann-Whitney test.

4.3.5 Competitive fitness: ATCC 3502 mutator *vs* ATCC 3502$^+$

In evolutionary biology, random variation and fitness play a central role; fitter individuals in a certain environment are likely to produce more progeny compared to their less-fit competitors. Consequently, the frequency of these genotypes with higher fitness will increase over time in a given population and can eventually become the dominant type. This effect of fitness variations might have played a pivotal role in the evolutionary trajectory of the Δ*mutL*231 genotype in the ATCC 3502 population. The fitness of the evolved mutator has possibly increased relative to the ancestral competitor, allowing it to become a substantial part of the population.

Although mutator strains experience an increased load of deleterious mutations and are more prone to lethal mutations, it is not uncommon for mutators to reach fixation in a bacterial population.[368, 371, 383] They are regularly identified in laboratory populations and clinical settings, especially in fluctuating and high stress environments.[377-382, 384, 387-389, 393, 405-406, 543-544] Mutator mutations can spontaneously arise and spread in bacterial populations, although, on their own they do not necessarily provide sufficient adaptive advantage to reach high frequencies due to the majority of mutations being deleterious or neutral, thus being an effective fitness cost.[359, 376, 417, 545] Current understanding is that in order for hypermutator variants to become established in a bacterial population the mutator genotype has to "hitchhike" along with genetically linked beneficial mutations.[394, 416, 546-547] The hypermutator's elevated mutation rate increases its chances of acquiring (rare) adaptive mutations when exposed to new environments compared to non-mutators. In addition, specific mutator mutations exhibit their own mutational spectrum, which might influence their fixation probability by yielding mutants with different fitness levels.[375, 548]

Various strategies have been devised to determine the fitness of a mutator compared to their ancestor.[549] The fastest and easiest method to quantify bacterial fitness is by tracking the growth of a strain in a culture on its own over time, and subsequently calculate the maximum growth rate (V_{max}). This method however takes maximum growth rate as the only component of fitness, ignoring other potential forces that effect the fitness of an evolved mutant relative to its ancestor. A second approach revolves around studying how well bacteria can cope and adapt to stressful compounds, such as antibiotics, by quantifying the Minimum Inhibitory Concentration (MIC) of the compound. The same limitations from the first approach apply to this strategy, and thus only provides a proxy for fitness.

The third method to quantify fitness is based on a competition assay, whereby one strain or population directly competes against the other for the same resources. The relative, rather than absolute, fitness is determined by measuring their relative contributions to future generations across the full culture cycle. Now, not only maximum growth rate, but also lag times and stationary phase dynamics, are incorporated in defining strain or population fitness. A notable example of where this assay method has been used extensively to study fitness and evolutionary dynamics is the Long-Term Evolution Experiment (LTEE) with *E. coli*.[352-353, 400, 550] When studying the evolutionary fate of a particular genotype, like in the LTEE, relative fitness provides a more meaningful insight than absolute fitness. Nonetheless, the absolute fitness should remain high enough to prevent extinction.[551] The competition assay was considered the most appropriated method to establish whether the ATCC 3502 *ΔmutL*231 genotype provides a fitness advantage over the ancestral strain under the standard culturing conditions.

Usually competitive fitness assays are started by mixing the two strains or populations of interest in a specific ratio (often an equal volumetric ratio) in fresh broth. Here, however, the two competitors, ATCC 3502^{+} and ATCC 3502 *ΔmutL*231, are already mixed together in a yet to be defined ratio. It was proposed that single colonies of each genotype could be isolated and subsequently used for the assay. The drawback of this approach is that these single colonies might have acquired secondary mutations, especially the mutator ones, which could have a positive or negative effect on the overall fitness of the strain. This would lead to a distorted view of what is actually happening within the ATCC 3502 population. Therefore, to get a genuine impression of the relative fitness of both strains, following this strategy would require many replicate assays, all starting from different single colonies. This was considered not

feasible due to time constraints and *C. botulinum* laboratory related practical limitations (e.g., limited PCR capacity).

Likewise, quantifying if there is a potential fitness advantage associated with the *mutL* deletion using the ATCC 19397 *ΔmutL*231 strains would be meaningless. As mentioned before, mutator genotypes are known to co-evolve (or hitchhike) to higher frequencies together with beneficial mutations. The hypermutable background of the ATCC 19397 mutants did not naturally evolve but was created by directed mutagenesis. The three biological replicates probably differ in their relative fitness to WT, due to secondary mutations that have most likely arisen during the construction of the strain.

Instead, the competitive fitness assays were started from the ATCC 3502 Unilever meat stock (discussed in more detail in section 4.3.6). Starting from the already mixed population meat stock had two advantages: one, there would be no bottlenecks, thus no risk of accidentally selecting an unlucky, less fit, individual for the assay and second, it is the most true representation of what is actually happening within the original population, rather than it being an orchestrated artificial situation.

Unfortunately, there was no easily selectable phenotypic difference (e.g., colony morphology or antibiotic resistance) to distinguish the mutator from the non-mutator. Hence, the only way to discriminate between the two populations was by performing PCR. Ideally, a large number of colonies of at least five replicate assays would have been screened to properly quantify the relative fitness, but as mentioned previously, performing more than three replicate experiments was not feasible due to the limited PCR capacity of the *C. botulinum* laboratory and the labour intensive nature of this screening method. The competitive fitness assay was performed as described in

section 2.9.7, and included three replicates of the static cultures and three replicates of the sub-cultured populations. Cultures were inoculated with cells directly from the meat stock (time 0). The static cultures were incubated for a total of 168 hours and the non-static cultures were sub-cultured every 24 hours for six days. At each time point 32 colonies were screened per culture. Estimating the relative fitness of the two ATCC 3502 populations using such a small dataset would probably be highly inaccurate. Therefore, only the percentage of mutator in the population was calculated for each time point instead of calculating the relative fitness using the estimated final and initial population size of each competitor.[549]

During the first 48 hours of incubation the *ΔmutL*231 genotype slightly increased from roughly 3% at time point 0 to approximately 10% in the static cultures (Figure 4.11). Interestingly, in the following 24 hour the number of mutators in the population started to decline again, returning to a level similar to time point 0. It should be noted that these values are highly inaccurate due to the small number of mutator genotypes (0-3) detected for each static culture at each time point. Nevertheless, these results show that in a static culture there is no obvious net fitness advantage for the mutator genotype in the first 168 hours. Possibly, the inherent degenerative effect of high mutation rates decreased the number of mutators in the population that could form viable spores. After approximately 24 hours of incubation *C. botulinum* cells start to form spores, see section 5.3.2. As discussed in section 4.3.1, spore formation is a complex process which requires the expression of many different genes, and hence, there is a high probability of it being disrupted by random mutations, especially in strains with a high mutation rate. Consequently, the genetically stable WT strain could regain some ground, as probably a larger part of this population produced viable spores compared to the mutator population.

When the mixed ATCC 3502 population was sub-cultured every 24 hours the *ΔmutL*231 genotype increased to on average 50% of the total population by the end of the second sub-culture (72 hours) (Figure 4.11). The population then settled at an approximately 50:50 ratio of mutator and non-mutator. Some variation was observed between the three replicate cultures; in R1 and R2 the mutator settled around 45%, whereas in R2 the mutator became the dominant genotype (70% of the total population). Furthermore, in R3 the *ΔmutL*231 genotype increase was a little slower compared to R1 and R2. This might be caused by evolution taking a slightly different route in each culture due to different mutations emerging in each population by chance.[382, 394]

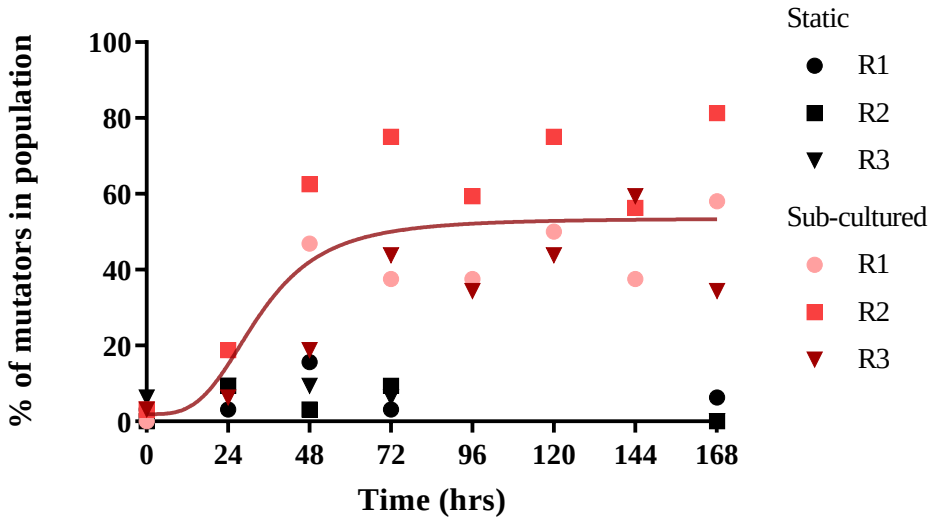


Figure 4.11 Competitive fitness assay of ATCC 3502^{+} *vs* ATCC 3502 Δ*mutL*231. Strains ATCC 3502^{+} and ATCC Δ*mutL*231 were co-cultured in TYG broth and allowed to compete for the same resources, starting from the ATCC 3502$^{+/-}$ (Unilever) meat stock. Cells in the static cultures were cultured for 168 hours in TYG broth and subsequently plated onto TYG. The first sub-cultures (S1) were inoculated (1:100) with cells from the S0 cultures after the first 24 hours of incubation. Subsequently, the S1 cultures were sub-cultured (1:100) after the next 24 hours (S2). This was repeated six times (S6). At time 0 and after every 24 hours 32 colonies per culture were screened by PCR using primers mutL_F_scr and mutL_R_scr to determine the ratio of non-mutator:mutator.

Overall, these results support the hypothesis that under the standard growth conditions, the *ΔmutL*231 genotype has a fitness advantage relative to the non-mutator genotype over the course of a growth cycle, though this advantage seems to diminish when its frequency in the population increases. The simplest explanation for the observed fitness advantage would be that the accidental *mutL* mutator mutation led to the appearance of (rare) beneficial mutations and subsequent hitchhiking of the mutator genotype to a high frequency.[391, 552] Analysis of all the sequenced ATCC 3502 genomes revealed the presence of only one secondary mutation linked to the Δ*mutL*231 genotype (Appendix VIII*).* Interestingly, this SNP at position 1585597 is a T deletion in an intergenic region (in a series of eight Ts), in between gene CBO1451, encoding a GTP-binding and nucleic acid-binding protein (YchF) and gene *mraZ,* encoding a cell division protein (MraZ).

The exact function of YchF is currently unknown; some studies suggest it might negatively regulate the oxidative stress responds, while others show that it interacts with the 70S ribosome and is involved in protein synthesis after translation initiation or late stage ribosome biogenesis.[553-555] In addition, YchF is known to interact with the proteasome, which led to the speculation that it might connect the ribosome and proteasome, facilitating the degradation of proteins damaged during synthesis.

The second protein, MraZ, is a highly conserved transcriptional regulator, regulating the expression of the *dcw* (division and cell wall) cluster.[556] It is thought to repress its own promoter by binding to its operator sites as well as down regulating cell wall synthesis and cell division genes. Overexpression of this regulator gives a clear phenotypic change but in *mraZ* deletion strains no obvious changes have been observed.

Although it seems possible that changes in the expression level of either protein could yield a fitness advantage based on the currently limited understanding of their cellular function, it seems unlikely that such a small genetic change upstream of these proteins, albeit possibly in a regulator binding site, would result in such a significant fitness gain. To establish if SNP 1585597 confers an adaptive advantage, the mutation has to be re-created in the genetically stable ancestor and its effect on fitness determined via a competitive fitness assay.

The lack of an 'obvious' beneficial mutation linked to the *ΔmutL*231 genotype is surprising, as hitchhiking is regarded a prerequisite for mutator genotypes to achieve fixation.[391] Numerous studies have analysed the occurrence of cell lineages with higher mutation rates in populations, whereby the mutation rate modifiers coevolved with adaptive mutations to a high frequency.[348, 392, 394, 502, 545, 557-559] The currently available theoretical frameworks described in literature do not provide an alternative route for fixation of the *ΔmutL*231 mutation. The most likely, although seemingly improbable, explanation is that a beneficial mutation was first acquired by the mutator, due to the higher supply rate of (beneficial) mutations, allowing it to avoid extinction by genetic drift in the early stages, but at a later stage the same mutation was also picked up by the WT subpopulation. Subsequently, the frequency of both subpopulations increased over time, up to a point where the original locus became so rare, it can no longer be detected by sequencing. The effect of multiple lineages carrying beneficial mutations in a population, competing for the same resources, is referred to as clonal interference and is thought to play an important role in the adaptive process driving mutations to fixation.[353, 391] In a recent study, where researchers tried to explain how beneficial and deleterious mutations influence the

fates of mutator genotypes using a model, showed that under specific conditions clonal interference can even enhance the fixation probability of mutators.[560]

Moreover, evolution experiments have provided compelling evidence that clonal interference facilitates the coexistence of multiple adaptive lineages, even if they have different mutation rates.[353, 561] Clonal interference could be one of the factors that led to the *ΔmutL*231 genotype frequency to plateau by the third sub-culture in all three replicate populations after its rapid selective sweep (Figure 4.11). Many studies have reported cases where mutators co-existed alongside non-mutators, for both short and long periods of time, with neither of them being eliminated by the other.[388, 394-395, 560, 562]

The data available to analyse potential clonal interference in the ATCC 3502 population is limited. Only six single colonies of the mutator and ten single colonies of the non-mutator genotype have been sequenced, originating from different experiments (note: none are from the competitive fitness assay). The mixed ATCC 3502 population has been sequenced twice (ATCC 3502 Unilever culture collection stock 70), but both times the average coverage was too low (approximately 60) to detect 'rare' mutations and subpopulations (see Appendix VI Table A 5 and Table A 6). Nine different SNPs were detected in the mixed populations. Six of these were also the most common mutations in the ten mutator isolates sequenced *(*1078297, 1667030, 1736091, 3234007, 3267818, 3629557*),* which suggests that these mutations are the most common and could represent several larger subpopulations. Furthermore, the high frequency of these mutations indicates early occurrence and supports the hypothesis that the Δ*mutL*231 subpopulation evolved from a single deletion event.

Ten mutations were detected in some but not all non-mutator isolates and were not detected in any of the mutator isolates (4174, 393048, 1244400, 2108268, 2111712, 2466355, 2622601, 2651226, 2961144, 3641879; for more details about the individual mutations see appendix VI). These mutations could be in just this specific isolate, acquired by random chance or might actually represent a non-mutator subpopulation. In particular, the few mutations found in various isolates, isolated from different experiments, could represent a non-mutator subpopulation (clade). If they do represent a distinct lineage, it's still difficult to know if these mutations are 'new' (appeared after the emergence of *ΔmutL*231) or 'old' (appeared before the emergence of Δ*mutL*231) and if a specific mutation confers a fitness advantage, is neutral or is slightly deleterious.

Only one mutation occurs in some non-mutators and some mutators (209980/150 bp deletion 209831-209980 upstream *prfB*), and therefore must have evolved independently in both lineages. The mutation overlaps with repetitive sequences present at both ends of the deletion. Such repetitive sequences are known to increases the probability that a specific deletion, like this one, can occur spontaneously multiple times.[361, 365-366] Usually, the independent evolution of the same mutation is used as an argument to support presumed fitness gains, however, this mutation has not affected a coding sequence, so if it provides an adaptive advantage remains up for debate. Another option would be that the *mutL* deletion evolved independently in both lineages. There are, however, no obvious repeat regions upstream and downstream of the *mutL* deletion site, indicating that the independent evolution of the exact same *mutL* deletion was unlikely.

Within the mutator population there appear to be at least three subpopulations (clades) based on the few mutator genomes sequences so far (Appendix VIII). Several colonies

isolated from different experiments carry quite a few identical mutations. Two sets of mutations occurring together have been identified. The probability that the different isolates acquired such a large number of identical mutations independently is very low. Mutator isolates from the same experiment (e.g., the *pyrE*$^{-}$ and *hsdR*1600s::CT mutants) have more SNPs in common, which makes sense because they were created from the same overnight culture.

Although the majority of SNPs identified in each isolate are unique, a few mutations occur in isolates of different clades. This could either suggest that they: (i) confer a fitness advantage, thus are more likely to survive genetic drift; (ii), evolve independently to a high frequency and are subsequently detected in the population, or; (iii) they might have been in a common ancestor but were lost by parts of the population (e.g., because of genetic drift or a negative selection load).

The evolutionary dynamics leading to the fixation and coexistence of mutators with their normomutable ancestor are not fully understood yet. In addition, the long-term fate of mutators in such mixed populations is still unknown.[359] It is, therefore, difficult to predict the evolutionary fate of the Δ*mutL*231 genotype in this ATCC 3502 population. Several theories have been postulated in an attempt to comprehend and described the factors that possibly affect the trajectory of mutators in an evolving population, such as a shifting balance between the advantages of hitchhiking and the burden suffered by a deleterious load, or the effect of clonal interference.[372, 391, 393, 562] Additionally, a high supply rate of adaptive mutations could be susceptible to diminishing-returns epistasis, as the marginal gain from a beneficial genetic modification decreases with surging fitness.[418] It is worth noting that mutator phenotypes are not necessarily a negative selection load, depending on the mutator strength.[376, 417] Moreover, the stability of environmental conditions (large

environmental fluctuations favour a mutator phenotype), anti-mutator mutations, population size, availability of beneficial mutations, clonal interference, genetic drift, genetic draft (co-evolution of beneficial mutations with neutral or slightly deleterious mutations), rate of recombination, horizontal gene transfer and the severity of bottlenecks shape the fate of mutators.[374, 391, 405-406, 560] Overall, there appears to be a complex pool of factors that can influence the fate of mutators. Nonetheless, it is a well-accepted tenet that eventually, when a population becomes will adapted, there will be a negative selection against mutators and lower mutation rates will re-evolve, providing appropriate alleles are available.[372]

Although it is tricky to assert how the *mutL* deletion became fixed in the population using the limited genome sequencing dataset available, it does provide a sneak peek at what might be going on. The SNP profiles suggest that there are at least several different subpopulations in the mutator population. It is unclear whether this is also the case for the non-mutator population. To obtain a better understanding of the evolutionary dynamics that have driven fixation of the Δ*mutL*231 genotype and its future fate, the competitive fitness assay should be repeated. The experiment should include more replicate cultures, a larger number of colonies should be screened per replicate, the duration of the experiment could be extended and the population could be sequenced at different time points using whole-population 'metagenomic' sequencing.[563] Screening a large number of colonies by PCR for the *mutL* deletion is frankly impossible. As an alternative, the full population could be analysed at each time point via 'deep sequencing' of the *mutL* locus, using NGS technology.[564] Deep sequencing of the *mutL* locus will yield thousands of reads for just this specific region and will thus provide a very accurate picture of the mutator non-mutator ratio. Moreover, analysis of older ATCC 3502 populations could provide useful information

about the 'origin' of the *mutL* deletion and how it might have hitchhiked to a high frequency, although getting hold of older mixed ATCC 3502 populations could be challenging.

4.3.6 Going back in time: the origin of mutator ATCC 3502 Δ*mutL*231

One of the first questions which arose after discovering the *mutL* deletion, was when and where did the mutator first emerge. Did it all start at the SBRC or is it more widespread among the *C. botulinum* research community studying ATCC 3502. The latter was considered most likely as the mutator phenotype was present in two independent SBRC stocks which were acquired from different sources (Quadram Institue/IFR and Unilever). Moreover, the deletion was already present in the first published genome of ATCC 3502, which was sent to the Wellcome Trust Sanger Institute for sequencing around the year 2000 and deposited to the NCBI database in 2007.[505] This was long before any of the *C. botulinum* ATCC 3502 stocks were added to the SBRC strain collection.

In an attempt to trace back the mutator's origin all the meat stocks of ATCC 3502 in the SBRC culture collection were checked for presence of the *mutL* deletion or the full *mutL* using flanking primers mutL_F_scr and mutL_R_scr. For the screening process the meat stocks were serial diluted and various dilutions were plated onto TYG agar (the whole process did not include any additional culturing steps, such as an overnight incubation in TYG broth). Subsequently, at least 20 single colonies were screened per stock.

In the meat stocks that were sent to the SBRC in 2006 and 2007 by Quadram Institute only the mutator allele was detected (Figure 4.12A/B). In the Unilever meat stock both the mutator as well as the non-mutator allele were detected in an almost equal ratio. It

should be noted that on all the plates various colony morphologies appeared and that a similar number of each colony morphology was screened, regardless of their frequency. Therefore, the mutator non-mutator ratio is probably inaccurate.

Interestingly, Unilever acquired its ATCC 3502 stock from the NCTC (PHE), but the corresponding genome, published in 2007, carries the mutator *mutL* allele. It is plausible that the stock sequenced was a mixture of both alleles, with the mutator allele being more frequent. Hence, software decided to incorporate the mutator *mutL* sequence instead of the full *mutL* gene.

To verify if the NCTC is in possession of a mixed *C. botulinum* ATCC 3502 (NCTC 13319) population, a new ATCC 3502 stock was acquired directly from the NCTC. Upon arrival, the freeze dried cells were re-suspended in TYG broth plus 10% glycerol and frozen at -80 °C directly, to avoid any possible effects one growth cycle could have on the mutator non-mutator balance. In addition, a fraction of this cell suspension was serial diluted and plated on TYG agar for screening purposes. Surprisingly, all colonies screened had the mutator genotype (Figure 4.12C).

To confirm that the ATCC 3502 populations received from the Quadram Institute and the NCTC exclusively consist of mutator cells, a mutator specific and a WT specific PCR screening were performed on the full population using primer pairs mutL_F2_scr/mutL_del-sp_scr and mutL_F2_scr/mutL_full-sp_scr, respectively (Figure 4.13B). The mutL_del-sp_scr primer anneals to the *mutL* deletion site, with the 3' end annealing to the 14 bp upstream of the deletion site and the 5' end annealing to the 14 bp downstream of the deletion site. No band corresponding to the non-mutator *mutL* allele could be detected in the Quadram Institute meat stocks, the SBRC culture collection stock CB047 (made from the Quadram Institute meat stocks) and

the NCTC 13319 freezer stock (Figure 4.13A). The mutator allele was detected in all samples except culture collection stock CB072 (probably a single colony isolate from the Unilever stock) and the negative control (data not shown).

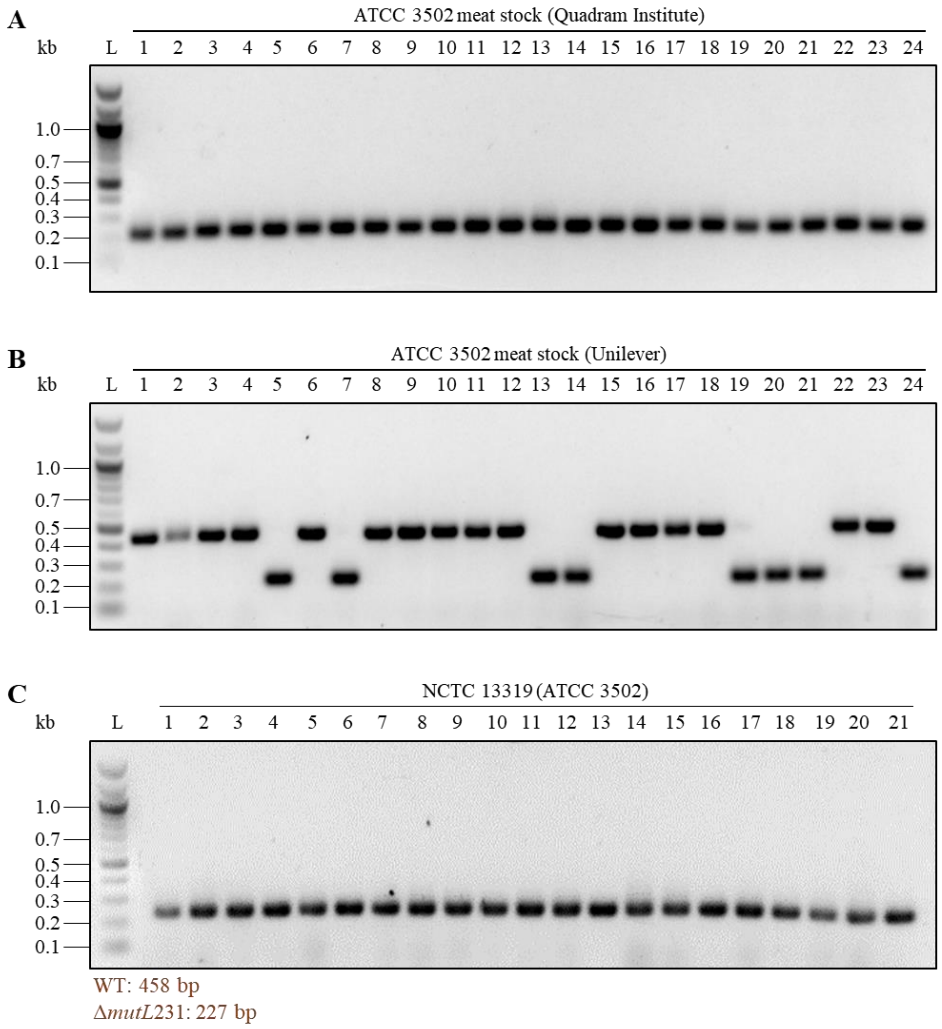


Figure 4.12 PCR screening of ATCC 3502 stocks for Δ*mutL*231. Screening for *mutL* deletion in published mutants (**A**) ATCC 3502 meat stock (Quadram Institute); (**B**) ATCC 3502 meat stock (Unilever); (**C**) NCTC 13319 (ATCC 3502). Products were amplified using flanking primers mutL_F_scr and mutL_R_scr. Lane L: NEB 100 bp DNA Ladder.

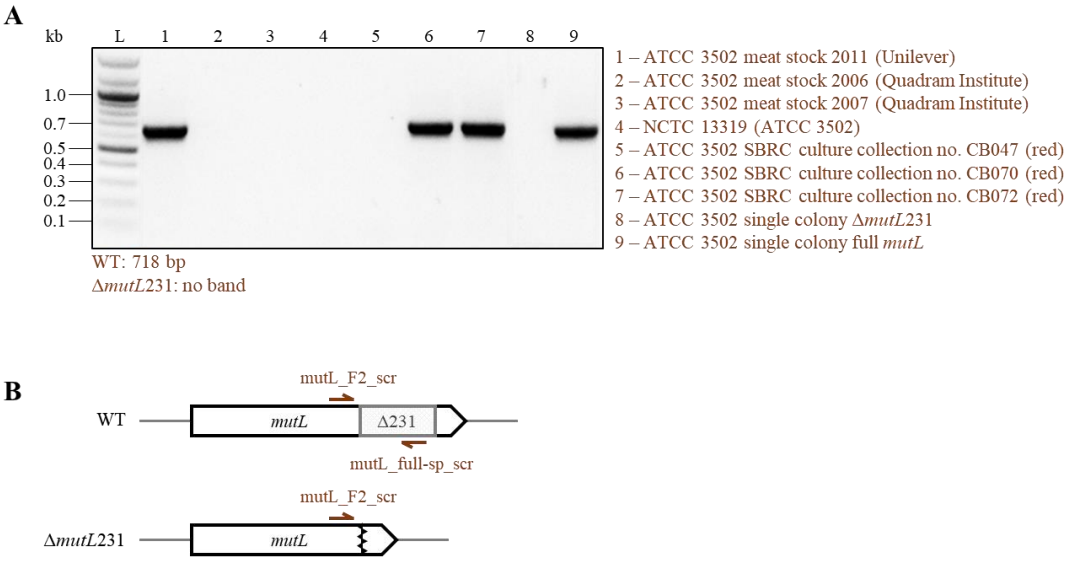


Figure 4.13 PCR screening of ATCC 3502 SBRC stocks for Δ*mutL*231. (**A**) PCR on full population of SBRC ATCC 3502 stocks alongside two single colony controls (Δ*mutL*231 and full *mutL*. Products were amplified using primers specific for the non-mutator genotype; mutL_F2_scr and mutL_full-sp_scr. Lane L: NEB 100 bp DNA Ladder. (**B**) Schematic diagram of the targeted genomic region showing the primer binding sites in the non-mutator and mutator strain.

Furthermore, the NCTC recently re-sequenced part of their culture collection, including ATCC 3502 (NCTC 13319; NCBI accession number NZ_CP046450). As expected, this genome contains the mutated *mutL* gene. The NCTC's policies regarding storage, stock replenishment and which vials get shipped or sequenced are unknown, though it is probably safe to assume that they keep a 'master' stock, which they use to inoculate fresh growth medium in order to generate new stocks for shipment. How they make their stocks (e.g., how many sub-culturing steps, bottlenecks, etc.) could have significantly influenced the stock composition, given that within just one growth cycle the mutator population can increase by approximately 3- to 7-fold. Possibly, the NCTC stock Unilever received in 2010 is from an older, less sub-cultured, batch than the 'new' NCTC stock received by the SBRC in 2018. Further analysis of the NCTC ATCC 3502 culture collection stocks is necessary to understand

why one stock consists of a mixed population whereas the other seems to be solely mutator.

Based on the above information a strain provenance was constructed (Figure 4.14), albeit lacking complete information regarding the Δ*mutL*231 genotype occurrence. Researchers at various institutes, such as Mike Peck (Quadram Institute), Miia Lindström (University of Helsinki) and Marite Bradshaw (University of Wisconsin-Madison) were contacted in an attempt to fill in some of the gaps. Unfortunately, the information received was not sufficient to complete the provenance due to various reasons (e.g., due to lack of sequencing data or unknown source of sequencing data; single colony or population).

No reads from the raw sequencing data of two samples received from the Helsinki laboratory mapped to the full *mutL* gene. Unfortunately, the type of material sequenced (e.g., 'master' stock or single colony) is unknown, therefore the potential presence of the non-mutator genotype cannot be excluded.

The University of Wisconsin-Madison hold a substantial *C. botulinum* strain collection, which includes strains ATCC 3502, 62A and Hall A-Hyper, all acquired from the ATCC. Genomic DNA from several different culture stocks were screened for the presence of the *mutL* deletion using primers mutL_F_scr and mutL_R_scr. All gave the 458 bp band, corresponding to the non-mutator *mutL* allele.

In large asexual populations it takes a long time for a new beneficial mutation or hitchhiking mutator mutation to increase from one individual to a detectable frequency.[359] Lenski *et al*. calculated that it takes approximately the same time for a favoured new genotype to increase from a frequency of 10^{-7} to 10^{-6} as it takes this genotype to increase from 10% to 90%, assuming constant selection.[352] If this true, it

can be logically assumed that the Δ*mutL*231 genotype was already present in the stock(s) at Porton Down.

Porton Down was the first to acquire the ATCC 3502 strain in Europe from the American ATCC in 1987 and subsequently distributed it (indirectly) across Europe (Figure 4.14).[505] Therefore, most likely all laboratories in Europe have been working with a mixed or mutator only. Those pharmaceutical companies that are producing BoNT/A as a therapeutic/cosmetic product from *C. botulinum* strain may need to be aware of the potential occurrence of the *mut* mutations if they are using a type A strain such as ATCC 3502.

To investigate whether the mutator genotype has only spread across Europe or if it is occurring globally, potentially affecting academic research and pharmaceutical companies, ideally the 'oldest' stocks at Porton Down and ATCC are analysed. To date, according to the ATCC website, they have a very limited collection of *C. botulinum* strains and strain ATCC 3502 is missing from this list. The institute was contacted twice but was unfortunately unable to clarify if they still hold ATCC 3502 in their collection. The odd disappearance of the strain might be related to a change in policy regarding *C. botulinum* strains, as they are currently listed as a Schedule 5 pathogen, under the Anti-terrorism, Crime and Security Act 2001.[101-102] It will be very challenging, if not impossible, to get information regarding the ATCC 3502 stocks from both institutions. An easier route might be to ask laboratories across the world to check if they have ATCC 3502 in their culture collection and if so, to sequence the *mutL* locus in their oldest populations via deep sequencing technology.

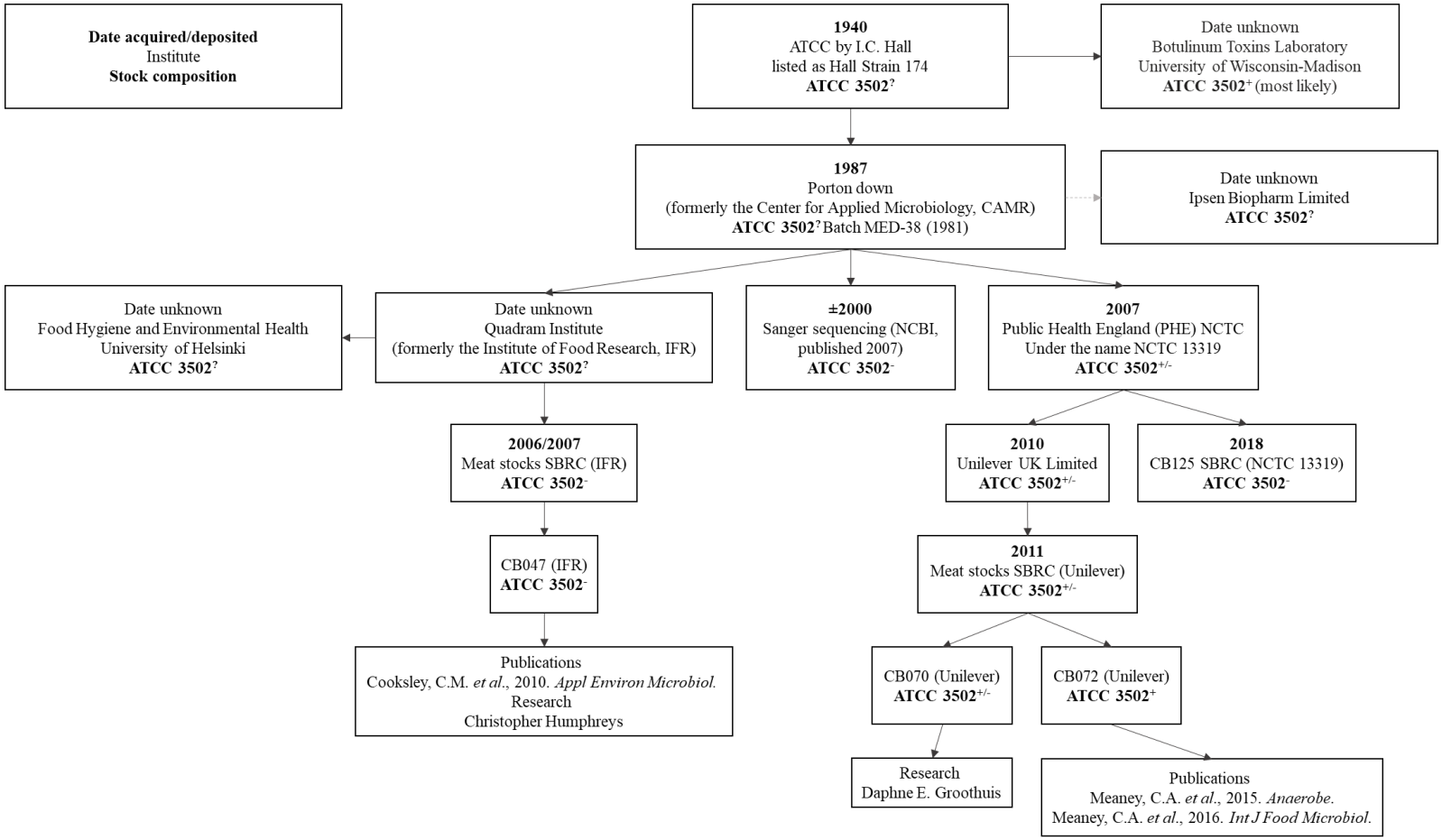


Figure 4.14 ATCC 3502 strain provenance. Flowchart visualising how strain ATCC 3502 was distributed across Europe. Box top: the date a laboratory acquired or deposited a stock of ATCC 3502. Box middle: the institute or company that received the strain on the above date. Box bottom: ATCC 3502 stock composition, determined by PCR screening or NGS; ATCC $3502^{+/-}$: mix of non-mutator and mutator; ATCC 3502^{+}: only non-mutator detected; ATCC 3502^{-}: only mutator detected; ATCC $3502^{?}$: composition unknown.

4.3.7 Published ATCC 3502 research

The established connection between the Δ*mutL*231 genotype, elevated mutation rate and the phenotypic changes resulting from these spontaneous mutations observed previously in ATCC 3502 sheds new light on precursory data. To date, ATCC 3502 is one of the most, if not the most, researched *C. botulinum* type A strain. Its readily genetic accessibility made it the go to organism for engineering based research projects, such as studying the sporulation process or toxin production and its regulation by deleting or inserting genetic material.[170-173, 176, 181-185, 512-514, 565] The question that arises is, did secondary mutations influence the phenotypic characteristics of the mutants described in literature. Is it possible that some of the observed changes have been attributed to specific genetic modifications erroneously, because the phenotypic alternation was actually due to secondary mutations.

The above does not necessarily have to be the case though, if researchers have been following good scientific practices, which includes the use of independent biological replicates, complementation and performed experimental repeats to verify results. However, these standards are not always adhered to, accidentally or intentionally; for example, sometimes the number of experimental repeats is limited due to laboratory limitations or several biological replicates of a mutant strain are isolated from a single mating experiment instead of setting-up independent matings for each replicate. Mutants isolated from the same mating experiment might have originated from a single conjugation event. These isolates are technically clones and not independent biological replicates, thus giving a false sense of consistency/reproducibility.

The *mutL* allele variant in all mutant strains published by the SBRC was determined via PCR screening using primers mutL_F_scr and mutL_R_scr (Figure 4.15). As

expected, the ClosTron mutants studied by Cooksley *et al*., which were generated from the ATCC 3502 (IFR) strain, all carried the mutator allele.[171] It is unknown whether the three biological replicates of each mutant were isolated from a single mating culture or from independent experiments. Despite the mutator background the reported results are probably quite accurate, mainly because similar phenotypic changes were observed for *agrD1* and *agrD2* repression/inactivation in the genetically stable *C. sporogenes* ATCC 15579 strain and *C. botulinum* ATCC 3502. Nonetheless, the published information should be interpreted with caution, especially the reported neurotoxin levels, as the connection between the *agrD* genes and toxin production could not be verified in the non-toxigenic *C. sporogenes* strain.

Surprisingly, all the mutants published by Meaney *et al*., carried the non-mutator *mutL* allele, whereas her WT stock consisted of the mixed ATCC 3502 (Unilever) population. The only logical explanation for this puzzling 'luck' is that all mutants were generated from the pure non-mutator CB072 stock. If this is true, all mutants were generated from a single WT clone. Little variation has been detected between the WT genomes analysed thus far, therefore all mutants having a single clone ancestor is probably irrelevant. Nevertheless, it might be worth to take details like this into consideration when setting up future experiments; even one SNP can have a significant impact on cell processes.

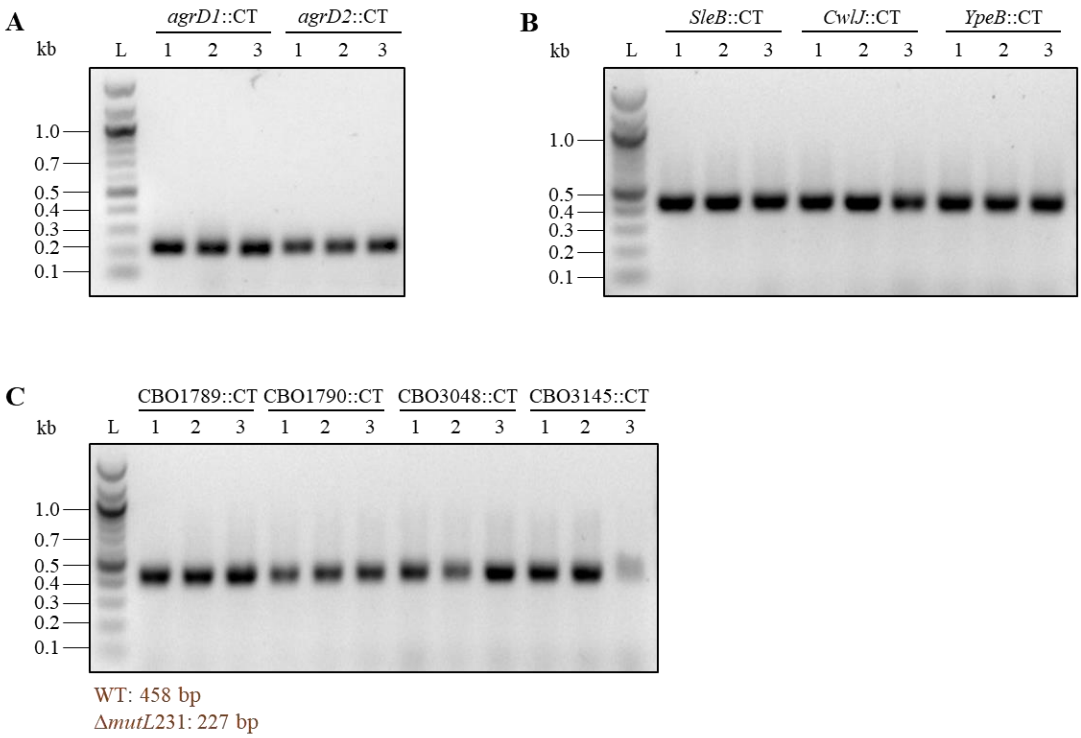


Figure 4.15 PCR screening of published ATCC 3502 mutants for Δ*mutL*231. Screening for *mutL* deletion in published mutants (**A**) *agrD1*::CT and *agrD2*::CT; (**B**) *SleB*::CT, *CwlJ*::CT and *YpeB*::CT; (**C**) CBO1789::CT, CBO1790::CT, CBO3048::CT and CBO3145::CT.[171, 182, 181] Products were amplified using flanking primers mutL_F_scr and mutL_R_scr. Lane L: NEB 100 bp DNA Ladder.

4.3.8 What is a wild type genome?

In biology the term wild type refers to a phenotype, genotype or gene that is the standard/norm in a population. However, with the ever increasing understanding of population genetics, researchers are starting to appreciate that for most organisms a single WT does not exists. In a population a variety of allelic forms exists for most if not all loci, which vary in frequency. The most frequent alleles are deemed WT and non-standard/mutant alleles can become the WT if a genetic shifts occurs within the population, for example due to sudden environmental changes. According to this 'new' definition of WT, both the mutator and non-mutator genotype of ATCC 3502 could be considered WT.

To date however, in molecular microbiology WT usually still refers to a single (published) genome sequence. This WT sequence comprises the most frequent alleles for each locus in a population at a given time. The 'old fashioned' concept of a uniform WT is still being used today mainly for practical reasons; in genetic manipulation experiments, where standard is compared to non-standard, it is very useful for a global research community to know what is referred to as standard and what is considered a mutation that produces a distinctive phenotype.

From a practical standpoint, using the published ATCC 3502 genome sequences (NC_009495 and NZ_CP046450) which were based on a (partial) mutator population as the WT sequence is suboptimal. Therefore, here a new WT genome sequence is proposed, based on the non-mutator whole genome sequences from this study. In Table 4.1 the proposed changes to the published ATCC 3502 (NC_009495) genome sequence are listed.[505] A few mutations were identified in some, but not all, non-mutator genome sequences, which are listed in Table 4.2. More sequencing data is required in order to establish the frequency of these mutations and determine if they should be included in the WT genome sequence. The same applies to the mutations that were detected in only one of the WT isolates (4174, 393048, 2108268, 2111712, 2622601, 2961144 and 3641879).

Table 4.1 List of mutations identified in all non-mutator genome sequences

NC_009495 Reference position	NC_009495 Allele	Proposed WT Allele	Overlapping annotation	Comments
130004	C	A	Gene: *ileS* (Isoleucyl-tRNA synthetase)	Silent mutation
1078297	-	T	Gene: CBO0977 (AraC-type DNA-binding domain)	T insertion restores ORF
1585597	-	T		
1667030	G	A	Gene: CBO1533 (AraC family transcriptional regulator)	Silent mutation
1736091	T	C	Gene: *glcK* (Glucokinase)	Leu71Ser
3234007	-	A	Misc. feature: possible remnant of an integrated prophage	
3267818	-	A	Gene: *hpt* (Hypoxanthine phosphoribosyltransferase)	Asn12 fs
3629557	-	T	Gene: CBO3388 (PDZ domain of trypsin-like serine proteases)	T insertion restores ORF
3724684	C	T	Gene: *tufB* (Elongation factor Tu)	Silent mutation
1914862^1914863	-	For inserted sequence see (1)	Gene: *mutL* (DNA mismatch repair protein)	231 bp inframe insertion

(1) TTTTTTATTTCTTCTTCTTTTAAGTTGTCATTAGCTTTAACTGCGGATTTGCATGCCAGTGTAGCAATGGCATTATATTTTATTGTAGAAGTTTCTTTAG ACTTCATATTTTTTAAATTATAAAGTATATCCATAAAGAGATTTTCTACATTGGGTTTACCTAGAATTAAAGGTACTTCTTTTATATTTATAGTGTATT CTCCAAAAGTTTCCACTGAAAAACCTGAATTC

Table 4.2 List of mutations identified in at least two, but not all non-mutator genome sequences

NC_009495 Reference position	NC_009495 Allele	Proposed WT Allele	Overlapping annotation	Comments
209980	G	-		
1244400	C	T	Gene: CBO1130 (CoA-substrate-specific enzyme activase)	Pro1066Leu
2466355	T	C	Misc. feature	
2651226	G	T	Gene: CBO2514 (hypothetical protein)	Pro25Gln
209831..209980	For sequence see (1)	-		150 bp deletion linked to SNP position 209980 Duplication/repeat region (ATCC 19397 matches NC_009495)
1966492..1966848	For sequence see (2)	-	Partial overlap with gene CBO1845 (hypothetical protein)	357 bp deletion linked to SNP position 2651226 Duplication/repeat region (ATCC 19397 matches NC_009495)

(1) TAAATAACAAGTCAGTATGACAGTCTATTTTGCGTCATACTGCGTCAGCAGAATCCTCCGATAGCACTACTAGCGACGGAACCTGCTTCCTTGGCTG
ACACAAAATATACTGCCATCTTTGACTTGTTATTTATTTTCATGTGCCTAAAG

(2) TTAGCTGTATATTTATCATACCTTTTATGTCTATTCCTAAAGTAATTTCTTTAGTATATTCAATTCTTTTGATTTATTAATATTTAATATATCTTGCATT
GTTAAAAATTTGTTTATTTTTTAATTTATACCTATACGAATTGGTAAAATTATGTTATACTATAAAAGTAAAAAAATTCAAATTCCGTAGAAATGGAT
GGCTACGGTGGTTGTACGTTTTCCGCCTTGGGCTATGGATGATTTTTCATTATAGCTTAAGGGGGTGGTATTGGATATGCTAGAACTTTTAATAAAAG
GAGTTTTTGTGGTTTGTATTATTACTTTAATATTAAAACACAATCTTTTTATTTCCACTAT

4.3.9 Are there other mutators out there?

Mutations that modify the genomic mutation rate of bacteria can occur spontaneously, just like any other beneficial, neutral or deleterious mutation. It is generally assumed that in large bacterial population mutators are constantly being produced from non-mutators, but quickly disappear again due to genetic drift and their associated fitness burden.[369] However, in bacterial populations adapting to fast changing or novel environments, mutators can quickly increase in frequency and even sweep to fixation through genetic hitchhiking with one or several of the many beneficial mutations available.[377, 388, 391, 566] This indirect selection is what might have happened here, where *C. botulinum* ATCC 3502, an environmental isolate, had to suddenly adapt to laboratory conditions, which favoured the *mutL* mutator genotype.

This prompts one to wonder whether environmental *C. botulinum* type A strain isolates are generally prone to these genetic changes and, primarily, what is the reason behind such a propensity. To investigate this, the two common mutator genes *mutS* and *mutL* were analysed and compared to the non-mutator sequence of ATCC 19397. In two type A strains, Hall (NC_009698) and CFSAN064329 (NZ_CP028859, 62A), mutator mutations were identified (Figure 3.14). In the Hall strain the *mutS* gene is truncated due to a SNP resulting in a premature stop codon (Glu87*) and in strain CFSAN064329 (62A) a 152 bp deletion disrupts the *mutS* ORF. Both mutations probably strongly decrease, if not completely abolish, *mutS* activity, and thus elevate the mutation rate.

Parallel evolution is considered a hallmark of adaptation. In addition, the competitive fitness assay showed that the ATCC 3502 mutator had a relative fitness advantage over the non-mutator in laboratory conditions. Overall, based on the findings

presented in this chapter, it is reasonable to assume that the mutators in the three type A populations achieved fixation because they conferred an advantage, either directly or indirectly, to the adapting population.[359]

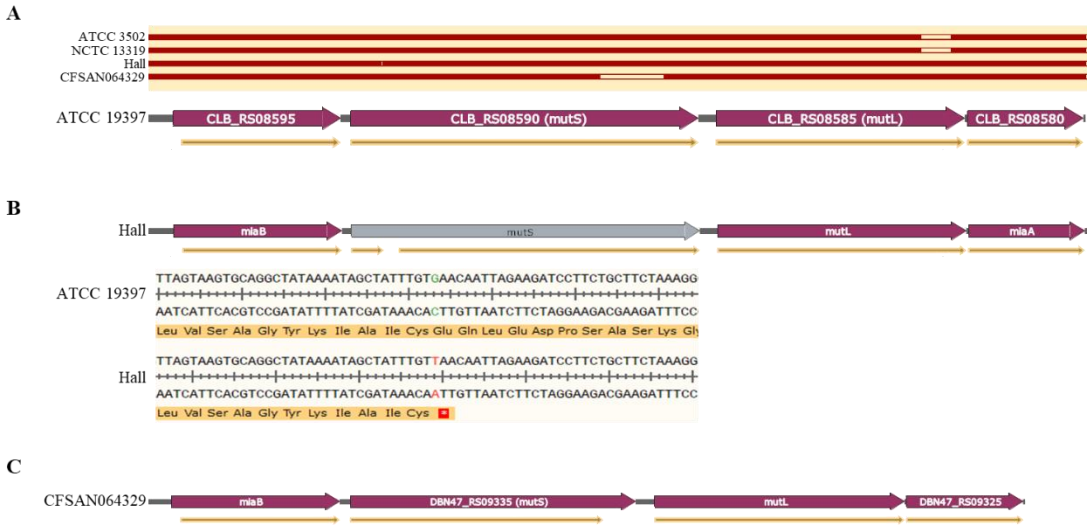


Figure 4.16 Mutations in mutator genes *mutS* and *mutL* in published type A genome sequences. Schematic representations of genomic region containing the mutator genes *mutS* and *mutL*. The depicted sequence alignments were performed using SnapGene. Orange arrows: ORF (**A**) Alignment of ATCC 3502 (NC_009495), NCTC 13319 (NZ_CP046450), Hall (NC_009698) and CFSAN064329 (NZ_CP028859, 62A) to reference ATCC 19397 () sequence. Gaps in reads indicate that this region could not be aligned; deletion or SNP. (**B**) Top: schematic representation of genomic region Hall strain containing *mutS* and *mutL*; bottom: mutation in mutS Hall strain which disrupts the ORF (Glu87*). (**C**) Schematic representation of genomic region CFSAN064329 (62A) strain containing *mutS* and *mutL*.

4.4 Key outcomes

- There is no significant difference in the number of secondary (random) mutations acquired in mutant strains generated using ACE, ClosTron mutagenesis or CRISPR/Cas9 mediated genome engineering.

- Two *C. botulinum* ATCC 3502 subpopulations exist; one with a hypermutator phenotype and one with a normomutator phenotype.

- The hypermutator phenotype of ATCC 3502 is caused by a 231 bp deletion in the *mutL* gene, which encodes a DNA mismatch repair protein and the deletion inactivates MutL.
- The ATCC 3502 Δ*mutL*231 genotype has a relative fitness advantage over the non-mutator ATCC 3502 genotype.
- The *C. botulinum* ATCC 3502 mutator phenotype originated before the strain came to the SBRC, Nottingham. The mutator has spread across European laboratories and possibly occurs in laboratory populations across the world.
- *C. botulinum* type A strains Hall and CFSAN064329 (62A) carry mutator *mutS* alleles.
- A new *C. botulinum* ATCC 3502 WT genome has been assembled to replace the published ATCC 3502 mutator genome.

CHAPTER 5

Generation and characterisation of safe *C. botulinum* strains

5.1 Introduction

5.1.1 The potential applications for non-toxigenic *C. botulinum* strains

The stringent regulations required for the handling of select-agents has severely limited research on *C. botulinum*, which requires certified high containment facilities and highly trained personnel.[71, 99] The use of genetically stable non-toxigenic *C. botulinum* strains that are identical to the WT except for the loss of toxicity would circumvent the current restrictive measures, eliminating biosafety requirements and reduce associated laboratory costs. Hence, such *C. botulinum* 'safe' strains should find wide application in both industrial and research settings.

The food industry is continuously developing their packaging procedures to meet customers' demands, such as less salt or sugars and the use of biodegradable plastics, while ensuring food safety.[72, 139] Despite the enforcement of strict rules to ensure food safety, there is always a potential risk of food poisoning. Toxin producing microorganisms, including *C. botulinum*, are widely distributed in the environment and can easily enter the food chain by, for example, contaminated soil or animal products.[128] Luckily, with the current industry practices the risk of getting foodborne botulism from commercially produced foods is very low. Nonetheless, an outbreak can have devastating effects for both the customers and the company.[131] Moreover, consumers do not always follow the manufacturer's storage recommendations (e.g., temperature abuse). As a consequence, there is continuous pressure on the food industry to develop new intervention and preservation systems in order to make sure that all potentially harmful microorganisms are eliminated.

To confirm the efficacy of new intervention and preservation systems, food challenge tests are performed using the relevant food spoilers.[143, 567-572] Performing such tests

with *C. botulinum* strains is complicated and costly due to the previously mentioned restrictions and are therefore kept to a bare minimum. Additionally, not all thermal food-processing procedures can be replicated in the laboratory, but introducing *C. botulinum* into food-processing facilities is hazardous. Therefore, the industry commonly uses *C. sporogenes*, which is closely related to the proteolytic Group I strains, as a non-toxic surrogate for testing how effectively food-sterilisation methods prevent *C. botulinum* outgrowth in canned food products.[573-574] Although *C. sporogenes* is very similar to *C. botulinum* type A strains, the two species differ in certain key characteristics, most notably spore heat resistance.[575] Furthermore, some of the widely used techniques, such as the botulinum cook, that have been established a long time ago, are still adhered to, in order to minimise process validations with *C. botulinum*. Consequently, there is probably room for improvement and/or optimisation of current intervention and preservation systems that is currently not being explored.

Non-toxigenic *C. botulinum* strains that retain all other characteristics except catalytically active neurotoxin could be used as a surrogate in food challenge studies to avoid the use of the original toxic parent strain.[126, 573] Foodborne botulism in humans can be caused by Group I and Group II strains, which represent physiologically distinct *C. botulinum* species. Even within the groups there is a large diversity in basic characteristics, such as spore heat resistance, growth temperatures and tolerance to additives.[10] Hence, just making one safe strain to represent the whole group would be insufficient. Ideally a library of safe strains is created, comprised of strains that represent the extremes of each group.

In the pharmaceutical industry, safe strains could be used for a variety of applications, including the development of vaccines, as production platforms or can be simply used as a negative control.[149, 156] Generally, pharmaceutical companies are more focused on

the toxin protein and lack the knowhow to engineer *C. botulinum*. Hence, readily available off-the-shelf safe strains would be preferred. For example, non-sporulating non-toxigenic strains can be used as an expression platform for recombinant BoNTs. Furthermore, strains expressing a toxoid could be used for the optimisation of production processes, such as fermentation conditions, at a lower cost.

The number of academic research groups studying *C. botulinum* is very small due to the stringent regulations and the associated costs, hampering research progress. The use of *C. botulinum* safe strains would eliminate the need for high containment facilities, allowing other research groups, without these specialised facilities, to study this intriguing organism. Additionally, a larger variety of research equipment and technologies, such as fluorescence microscopes, will be at researchers' disposal to study the organism when it can be handled in 'normal' laboratory settings. The availability of non-toxigenic *C. botulinum* strains will undoubtedly aid future research activities and accelerate data acquisition.

5.1.2 Currently available safe strains

The first non-toxigenic *C. botulinum* strain was already created in 1997, albeit with a completely different goal; investigating if the neurotoxin gene cluster is located on a mobile genetic element which would aid its transfer across different *Clostridium* spp.[576] The mutants were created using the tetracycline-resistant transposon Tn*916*, which inserted into the genome randomly, and approximately 7,000 colonies had to be screened to find four toxin negative strains. Clearly, this was a very ineffective way to create a specific mutant and it took more than 10 years before the next specific and intentional *C. botulinum* mutant was published; a non-toxigenic type A strain.[126]

The second non-toxigenic *C. botulinum* type A strain was created using ClosTron mutagenesis. The mutants' basic physiological characteristics, including sporulation and growth properties, were similar to the WT strain in laboratory conditions.[126] This strain however, was never re-classified as less hazardous or used for the food challenge studies it was originally created for. The ClosTron disrupts the neurotoxin ORF, but all the genetic material encoding the toxin is technically still present within the cell. Although ClosTron mutants are known to be very stable, there is the potential risk that the mobile genetic elements gets excised, which would restore the ORF and thus yield a toxigenic strain.[228]

The final safe strain currently published was created by Clauwers *et al*. in 2016 using two different techniques; ClosTron-mediated mutagenesis and an ACE-based approach.[529] The non-toxigenic group II NCTC 12219 mutants behaved in a similar way as WT under various conditions and were used in future studies looking at germinant receptors and outgrowth in food matrices.[141] For the ClosTron mutant the same concerns apply as described for the type A safe strain by Bradshaw *et al*. In the Δ*bontE*::*ermB* mutant the full toxin gene is replaced for an *ermB* gene, hence eliminating the risk of the strain reverting back to WT. However, the presence of an antibiotic resistance marker is far from ideal, both for industrial as well as for academic applications.

5.1.3 What is the perfect safe strain?

What constitutes a perfect safe strain depends on the intended application. Ideally, the non-toxigenic *C. botulinum* variants are free of antibiotic resistance markers and cannot easily revert back to WT. The three most obvious mutations to abolish toxicity are: (i) a precise deletion of the *boNT* gene; (ii) a precise deletion of the entire

neurotoxin gene cluster, or; (iii) precise nucleotide changes in the *boNT* gene to create toxoids (Figure 5.1).[225] Depending on the study one of variants might be more appropriate than the others.

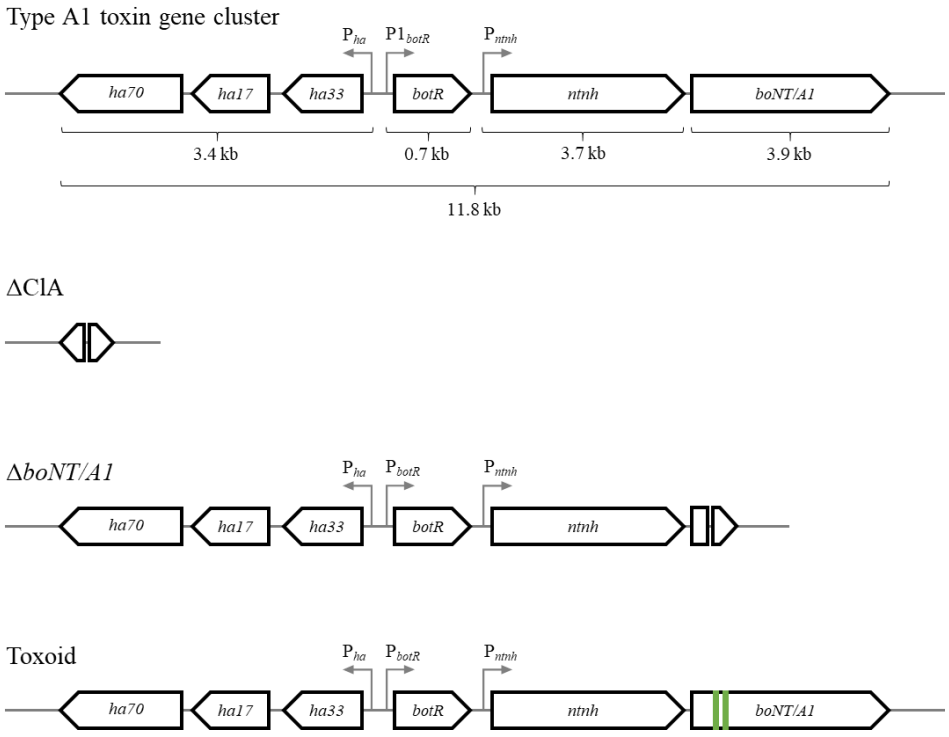


Figure 5.1 Schematic diagram of botulinum neurotoxin gene cluster and possible non-toxic variants. From top to bottom: schematic diagram of (i) full type A toxin gene cluster in ATCC 3502; (ii) full neurotoxin gene cluster deletion; (iii) clean deletion of toxin gene, and; (iv) point mutations in the toxin gene that abolish its catalytic activity. The green lines indicate mutation sites in the light chain.

Neurotoxin gene cluster deletion mutants and neurotoxin deletion mutants are probably the preferred option for food challenge studies. These strains cannot revert back to WT unless they acquire exogenous DNA, which is highly unlikely. Furthermore, these strains could be optimised for the expression and large scale production of BoNT derivatives. For example, a sporulation minus *boNT* deletion mutants could be used for the expression of recombinant BoNTs by inserting the

coding sequence at the original toxin locus. Moreover, these strains can be used as a negative control in a wide variety of experiments.

C. botulinum toxoid strains would be ideal for studying toxin regulation and secretion or what environmental factors influence toxin production. Additionally, toxoids can be used to optimise fermentation conditions. Preferably, these safe strains are made from the original production strains to ensure that discoveries can be translated on a one to one basis to the true production strain. It has been shown that incorperating either the double mutation R363A/Y366F or E224Q/Y366F is sufficient to eliminate any detectable neurotoxin activity in the BoNT/A light chain, while the proteins' structure remains essentially identical to WT.[577-579] BoNT/A light chains carrying all three mutations (E224Q/R363A/Y366F) have been shown to be catalytically inactive and fold correctly.[29] This toxoid variant, named BoNT/Ai, has been used to study the function of NTNH and a similar strategy could be applied to unveil the function of other neurotoxin associated proteins.[37]

5.2 Objectives

i. Find suitable genetically stable *C. botulinum* type A strains for the safe strain library,

ii. Create non-toxigenic *C. botulinum* type A strains that are physiologically identical to WT by; a) removal of the full neurotoxin gene cluster(s); b) removal of the *BoNT* gene(s); and c) introduction of point mutations that make the neurotoxin catalytically inactive using the newly developed RiboCas genome editing tool.

5.3 Results and discussion

5.3.1 Selection of various group I *C. botulinum* type A strains

To proceed with the goal of creating safe *C. botulinum* strains while the investigation of the genomic instability of the ATCC 3502 strain was ongoing, an alternative strain was required. An important factor in the selection process was the requirement for the strain of interest to be easily procurable within the UK (via the NCTC), in order to alleviate the need of acquiring MTAs and/or arranging international shipments of dangerous pathogens.

Table 5.1 shows a list of *C. botulinum* type strains available from the NCTC at the time. Type A strains are highlighted in yellow. Additional information regarding the origin/isolation source of each strain, their genome sequence and restriction/DNA modification systems is also provided.

From the available NCTC collection, strains were selected based on (i) a high degree of similarity with ATCC 3502, and; (ii) a smaller degree of similarity with ATCC 3502, but with comparable features regarding ability to cause disease, toxin production (type A), sporulation and growth. The selection of strains with a smaller degree of similarity to ATCC 3502 was undertaken as a precaution to minimise the risk of picking strains with coincidental genomic instability, given that the underlying cause of the problem had not yet been elucidated.

(i) Strains with high degree of similarity with ATCC 3502:

ATCC 19397: at the time of selection, the strain's genome was published and bore the highest homology to the genome of ATCC 3502 (BLASTn 99.96%, 96% query coverage).[49] However, literature regarding this strain was scarce and there had been no mention of successful DNA transfer.[14, 144, 510, 580-584] Consequently, it was unknown

whether the strain would be amenable to genetic engineering. The REBASE database predicted two potentially active restriction systems for this strain, a type I and a type II system.[308] The type I is 100% identical to the type I found in ATCC 3502, which readily accepts foreign DNA. Thus, this restriction system was anticipated to not be a limiting factor for successful DNA transfer. Contrarily, for the type II system no homologs were found in ATCC 3502 and therefore its impact on foreign DNA uptake was unknown. Fortunately, a toolbox to create the optimal conjugative donor strain for a specific recipient was being developed and this technology could be implemented for this strain if required (unpublished data; R. Mansfield).[338]

62A: the strain's genome and its homology to that of ATCC 3502 was unknown at the time of selection. Nevertheless, the reported sequence of the toxin genes was almost identical, except for one amino acid in BoNT/A1 (A735T) and in NTNH (E333G), to that of ATCC 3502 (BLASTn), indicative of a close relationship between the two strains.[585] With the subsequent publication of 62A's genome, a nucleotide sequence homology of 99.94% (97% query coverage) was calculated using BLASTn between the two strains.[331] Encouragingly, at the time of selection, 62A had been previously engineered and there was good availability of information regarding its sporulation and growth profile.[126, 175, 510, 575, 586-589]

(ii) Strains with smaller degree of similarity with ATCC 3502:

Loch Maree: at the time of selection the strain's genome had been published and although this strain expresses a slightly different type A toxin (BoNT/A3), which is based on a mega-plasmid (267 kb, pCLK), the overall genomic nucleotide sequence homology between ATCC 3502 and Loch Maree calculated by BLASTn was still 98.15% (query coverage of 90%, excluding pCLK).[49] Furthermore, the potential of

DNA transfer had been reported for this strain.[46] The REBASE database listed a type I, type II and type IV restriction system, which could be indicative of poor conjugation efficiencies. However, the creation of a customised donor strain to 'evade' these systems was considered straightforward (as discussed Chapter 3). Additionally, Loch Maree has been previously identified as the causative agent of food poisoning in humans (incident isolate), thus being relevant for food challenge studies.[590]

NCTC 2916: the strain's genome sequence was available at the time of selection, albeit being of poor quality and not complete.[583] The strain also benefitted from availability of information regarding its general phenotypic characteristics.[36, 43, 575, 588, 591] NCTC 2916 was considered a desirable candidate in a safe strain 'cocktail' for food challenge testing, thanks to the presence of the *spoVA*^{2} operon; high spore heat resistance has been attributed to expression of this operon in Bacillus species, although its role is yet to be confirmed for clostridial species.[592-594] The *spoVA*^{2} operon is a characteristic not shared by any of the previous strains. However, NCTC 2916 also encodes toxin B (although it appears to be mutated and inactive), hence rendering it safe would require double knockout of toxins A and B. In addition, the REBASE database only predicted a type IV restriction system, which can be circumvented by using the sExpress donor strain.[338]

Table 5.1 *C. botulinum* strains available from the NCTC (operated by Public Health England).

NCTC No.	Current Name	Other Collection No	Origin/Isolation source	Genome sequencing status	Restriction/DNA modification systems*
NCTC 887	*Clostridium botulinum*	ATCC 441; Type A; 38	K. F. Meyer, California 1921/Crop of chicken with botulism in 1920	X	Unk
NCTC 751	*Clostridium botulinum*	Type B, X	G. S. Burke, California 1919[3]	X	Unk
NCTC 2012	*Clostridium botulinum*	LOCH MAREE; Type A	W. M. Scott, England 1925/Wild duck paste in 1922[590]	Complete genome[49]	Type I, II and IV
NCTC 2916	*Clostridium botulinum*	Type A; 2447	I. C. Hall, University of Colorado 1929/Home canned corn in 1928	49 Contigs	Type IV
NCTC 3805	*Clostridium botulinum*	Type A; 62; CFSAN064329	R. Vawter, Nevada 1919/Liver infarct of cow[595]	Complete genome**	Type I and IV
NCTC 3806	*Clostridium botulinum*	Type A; 97	K. F. Meyer/Canned peas[595]	X	Unk
NCTC 3807	*Clostridium botulinum*	Type B; 115	K. F. Meyer, California 1921/Soil; blue ridge mountains	X	Unk
NCTC 3815	*Clostridium botulinum*	ATCC 443; NEVIN; Type B; 6B	K. F. Meyer, California 1925/Cheese	X	Unk
NCTC 4587	*Clostridium botulinum*	LEDERLE; NCTC 7272; Type A	E. W. Hurst, Lister Institute 1934/Unk	Complete genome[49]	Type I and II
NCTC 7272	*Clostridium botulinum*	ATCC 19397; NCTC 4587; Type A	B. C. Hobbs, PHLS Colindale in 1947/Probably derived from NCTC 4587	Complete genome[49]	Type I and II
NCTC 7273	*Clostridium botulinum*	BEANS; Type B; CECT 4610	B. C. Hobbs, PHLS Colindale in 1947/Unk	X	Unk
NCTC 8265	*Clostridium botulinum*	D 6F; INTNL Type D	K. F. Meyer, 1928/Subculture of strain isolated by E. M. Robinson, S. Africa in 1926-27/Unk	X	Unk
NCTC 8550	*Clostridium botulinum*	PR 15; Type E	A. R. Prevot, Institut Pasteur, Paris 1952/Unk	Complete genome[596]	2x Type II

NCTC 9837	*Clostridium botulinum*	MAURITIUS; Type A	A. J. H. Tomlinson, PHLS County Hall, London in 1955 / Fish preparation causing botulism, Mauritius	Complete genome	Type I
NCTC 10281	*Clostridium botulinum*	LANGELAND; TYPE F	Moller, V. & Scheibel, I., U. of Brit. Columbia Canada 1960, Denmark 1958/Duck pate	Complete genome	2x Type II and 1x Type IV
NCTC 10914	*Clostridium botulinum*	FH 6513/70; Type C (B)	R. J. Gilbert, PHLS Colindale 1971/Chicken feed	X	Unk
NCTC 11199	*Clostridium botulinum*	F199-200/78; Type A	Food Hygiene Laboratory, PHLS Colindale 1978/Faeces, infant botulism (UK)	X	Unk
NCTC 11219	*Clostridium botulinum*	F2561/78; Type E	Food Hygiene Laboratory, PHLS Colindale 1978/Tinned Salmon (US origin)	4 Contigs[596]	Type I and Type IV
NCTC 12265	*Clostridium botulinum*	Type B	Food Hygiene Laboratory, PHLS Colindale 1990 / Hazelnut yogurt (large outbreak of botulism in UK 1989)[597]	X	Unk
NCTC 13319	*Clostridium botulinum*	ATCC 3502 (Hall strain 174)	I. C. Hall, California 1922/Peas	Complete genome [505]	ATCC 3502: Type I NCTC 13319: Type I and IV***

* As detected by REBASE.[308]

** Published Apr 2018, after the strain was acquired from the NCTC for this study.

*** Published Nov 2019, therefore this study was based on the REBASE analysis of the 'old' ATCC 3502 genome published by Sebaihia *et al*., 2007.[505]

5.3.2 Basic characterisation of *C. botulinum* type A strains

To establish whether ATCC 19397, 62A, Loch Maree and NCTC 2916 were suitable safe strains candidates, more knowledge about their basic phenotypic characteristics (e.g., sporulation efficiency, spore heat resistance and genetic accessibility) was required. In addition, culturing the strains under identical laboratory conditions will provide more insight in the degree of similarity between the newly acquired *C. botulinum* strains and the ATCC 3502 strain. This information is important to identify the most appropriate candidate to replace model strain ATCC 3502 for future studies. Initially several standard characterisation experiments were performed; i.e., sporulation efficiency assays, growth curves and conjugation efficiency assays.

First, the sporulation/germination capacity and growth of the four newly acquired *C. botulinum* strains using TYG media was assessed, as described in sections 2.9.4 and 2.9.2, respectively. Figure 5.2A shows that while all strains were capable of producing viable spores, there were distinct differences between their sporulation profiles. As expected, the sporulation/germination capacity of ATCC 19397 was similar to ATCC 3502^{+}, though the formation of viable spores in ATCC 19397 seems to be delayed by approximately 24 hours. Furthermore, NCTC 2916 strongly resembled the ATCC 3502^{+} profile whereas the Loch Maree profile was more similar to ATCC 19397. For these three new strains the difference in the total viable spore count at 120 hours ranged between -1 to -3 fold compared to ATCC 3502^{+}, which was considered non-significant based on a two-tailed unpaired t-test. Interestingly, although a high similarity was anticipated, the 62A sporulation/germination profile was different compared to ATCC 3502^{+}, with the total viable spore count at 120 hours being almost 20-fold lower (significance *, *$p \leq 0.0332$). The poor sporulation/germination capacity observed for this particular 62A WT strain could be due to mutations in genes related

to the sporulation and/or germination cascade. To confirm this further studies would be required, starting with a thorough comparison of all the known genes involved in sporulation/germination in ATCC 19397, ATCC 3502 and 62A. Another possibility is that the growth conditions were suboptimal, effecting the strain's ability to from viable spores or other spore properties essential for spore survival (e.g., heat resistance).

All WT strains grew at approximately the same rate and reached similar total cell densities when cultured in TYG broth (Figure 5.2B). However, during the death phase the decrease in total cell density varied amongst the strains, with ATCC 19397 showing the strongest decline.

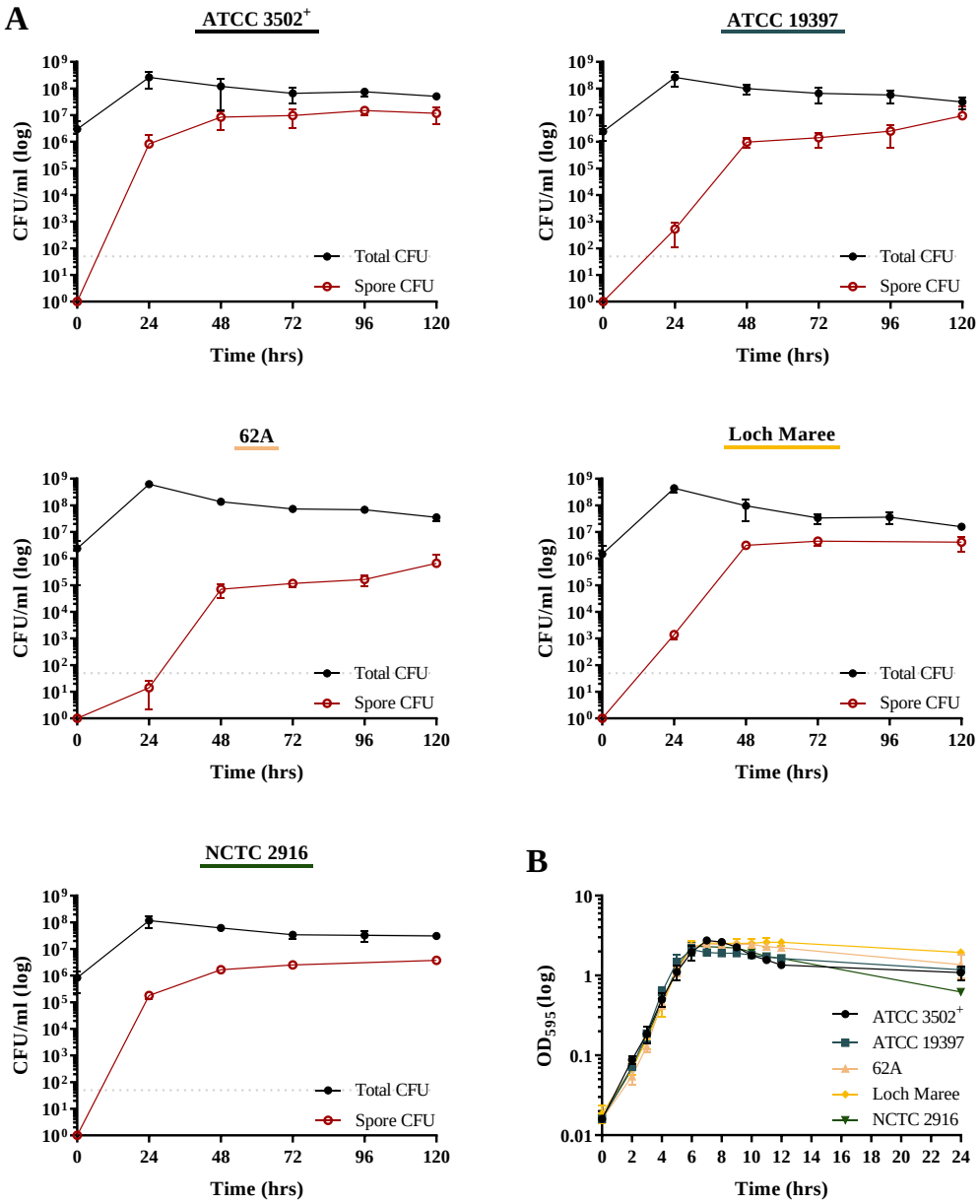


Figure 5.2 Basic phenotypic characterisation of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916. (**A**) Sporulation/germination capacity of WT strains. Strains were cultured for 120 hours in TYG broth and samples were taken every 24 hours. Total CFU corresponds to the total cell count (non-heat treated) and the spore CFU is the number of viable spores withstanding a 15 min, 80°C heat treatment. Error bars represent the standard deviation of three or more independent cultures. The dotted line at 50 CFU/ml represents the detection limit. (**B**) Growth profiles of WT strains. Strains were cultured in TYG broth and the OD_{595} was monitored for 24 hours. Error bars represent the standard deviation of three independent cultures.

As the creation of non-toxigenic mutants using the CRIPSR/Cas9 genome editing tool is dependent on efficient DNA transfer, the conjugation efficiency of the WT strains was assessed. Initially all strains were tested in combination with the standard CA434 *E. coli* donor strain (see Chapter 3, Figure 3.4). As anticipated, for some of the strains this pairing resulted in little or no DNA transfer. Fortunately, by customizing the donor strains to fit the methylation pattern of the specific host, the conjugation efficiency of all strains was significantly improved, allowing the conjugation of plasmids carrying the CRISPR/cas9 genome editing tool components to genetically modify the strains. A summary of the main results presented in Chapter 3 can be found in Table 5.2 and Table 5.3. For more details about the generation and testing of these optimized donor strains see section 3.3).

Table 5.2 Summary of the maximum achieved conjugation efficiencies of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916.

C. botulinum **recipient**	***E. coli*** **donor**	**Methylation plasmid**	**Conjugated plasmid**	**Conjugation efficiency**
ATCC 3502	Stellar-R702^{CA434}	-	pMTL83151	$1.77*10^{-3}$
ATCC 19397	CA434	pM-JTI+JTII	pMTL83151-R1+R2	$7.21*10^{-5}$
62A	sExpress	-	pMTL83151	$1.07*10^{-4}$
Loch Maree	sExpress	-	pMTL83151	$2.65*10^{-5}$
NCTC 2916	sExpress	-	pMTL83151	$8.49*10^{-5}$

Table 5.3 Overview of the theoretically optimal customized donor strains for ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916 to achieve maximum conjugation efficiency.

C. botulinum **recipient**	**Restriction systems in** ***C. botulinum*** **recipient**	***E. coli*** **donor**	**Methylation plasmid**
ATCC 3502	Type I and IV	sExpress	pM-JTI
ATCC 19397	Type I, II and IV	sExpress	pM-JTI+JTII
62A	Type I and IV	sExpress	pM-JTI
Loch Maree	Type I, II and IV	sExpress	Plasmid expressing CLK_RS10775 (*hsdM*) + CLK_RS10770 (*hsdS*)
NCTC 2916	2x Type I and 1x Type IV	sExpress	Plasmid expressing CBN_0725 (*hsdM*) + CBN_0726 (*hsdS*) and CBN_2138 (*hsdM*) + CBN_2137 (*hsdS*)

5.3.3 Whole genome sequencing of *C. botulinum* type A strains

Bacterial strains continuously evolve when grown, albeit slowly, but the more you subculture a strain the bigger the genetic difference with the original 'parent' strain.[352-353, 504] Generally, to distribute strains across laboratories subcultures of a parent strain are created. Between these laboratories culturing conditions as well as storage methods might vary, which can accelerate the genetic differentiation. Consequently, there will always be some genetic variation between WT stocks of the same strain, potentially causing significant differences in the overall behaviour between stocks of a strain. In addition, poor storage practices combined with prolonged culturing episodes often results in strain degeneration. Therefore, the full genome of each strain was re-sequenced and compared to the genome sequence deposited in the NCBI database (ATCC 19397 accession number: NC_009697; 62A accession number: NZ_CP028859; Loch Maree accession number: NC_010520; NCTC 2916 accession number: NZ_ABDO00000001-49). In this case re-sequencing was considered particularly relevant, because *C. botulinum* type A strains seem to have the tendency to evolve towards a mutator phenotype (see Chapter 4, section 4.3.9). The main results of the resequencing analysis are discussed below per strain.

ATCC 19397 (NCTC 7272): in total only four variations were detected; one in a non-coding sequence, two causing amino acid changes and one resulting in a frameshift (Appendix VII Table A 16). Both amino acids substitutions are in a transcriptional regulator, one from the MuR/RpiR family and one belonging to the sigma 54-dependent Fis family. The impact of these mutations on protein functionality is unknown, making it difficult to predict if they effect the general cell metabolism. The fourth mutation (3465907) disrupts the ORF of a putative lactate dehydrogenase (CLB_RS16495), changing the originally 336 amino acid long protein into a small 15

amino acid peptide. This frameshift has probably fully inactivated the dehydrogenase. ATCC 19397 has a second gene annotated as a L-lactate dehydrogenase (CLB_RS07635) which might have a similar function in the cell (catalysing the conversion of lactate to pyruvate), potentially reducing any negative effects of the CLB_RS16495 inactivation. However, no significant homology was detected between the two proteins (BLASTp).

62A (NCTC 3805): because the strain's genome was initially unknown, 99 contigs were assembled based on the illumina MiSeq sequencing in CLC genomics workbench and annotated using the RAST pipeline.[229, 332, 598] The neurotoxin gene cluster as well as several DNA mismatch repair genes were aligned to the ATCC 19397 strain and no abnormalities were detected.

Following the publication of the 62A (CFSAN) genome (CFSAN064329) by the FDA the MiSeq paired end reads were mapped to this published genome and variations identified using the CLC genomics workbench.[331] A striking total of 253 variants were detected; 38 in non-coding regions and 215 within ORFs (Appendix VII Table A 17). Additionally, a 512 bp deletion in the *mutS* gene (2799 bp) of the published 62A strain was identified. As previously described in Chapter 4, the inactivation of *mutS* results in a strong mutation rate increase.[452, 454-455, 457] DNA mismatch repair is probably less efficient, if not completely abolished, in the 62A (CFSAN) strain. The high number of variants between the 62A (CFSAN) and the 62A (NCTC 3805) is most likely due to this specific deletion, accelerating the genetic diversification.

From all the mutations in coding sequences, 59 cause a frameshift, with the majority of them being inside stretches of As or Ts. Interestingly, 55 of the 'frameshift' mutations in the 62A (NCTC 3805) strain restore the ORF. Thus, in the American 62A

strain a total of 55 genes are disrupted, and presumably inactivated, depending on the mutation's location (start, mid or end of the ORF). Surprisingly, one of the frameshifts is in the *boNT/A1* gene, codon 824 (full protein is 1296 amino acids long), so after the light chain domain (± 448 amino acids). It is difficult to predict if this truncated neurotoxin has any residual activity. Furthermore, in two cases the frameshift mutation disrupts the ORF in the UK 62A strain (2041345 and 3243886) and for two frameshift mutations it is unclear whether the 62A (CFSAN) sequence or the 62A (NCTC 3805) is correct (3133402 and 3670192).

Not all the SNVs were checked to establish which allele, the 62A (CFSAN) or the 62A (NCTC 3805), is probably the original WT allele. Only the SNVs in genes with a common name annotation were assessed (see Table 5.4). To determine which one of the two alleles of each gene is most likely the 'original', the gene sequences were compared to the ATCC 3502^{+} and ATCC 19397 alleles. Again, for a majority of the alleles the UK 62A sequence matched with the other two type A strains, whereas the US 62A sequence, with the mutator genotype, was different. In some instances it was difficult to decide which of the two alleles is the original because of a higher sequence heterogeneity between the type A strains. Silent mutations were not excluded from the list, as sometimes codon usage can affect total protein expression and protein folding.[523-524] Overall, the 62A (NCTC 3805) genome sequence is probably the closest to the original sequence of the common ancestor of both strains. Rather than trying to define what is the 'true' WT genome, it could be an option to acknowledge them as different strains and implement the same naming strategy as previously suggested for ATCC 3502 in Chapter 4, section 4.3.3, referring to the 62A (CFSAN) strain with the mutator genotype as $62A^{-}$ and referring to the genetically stable 62A (NCTC 3805) as 62^{+}.

Loch Maree (NCTC 2012): a total of 22 variants were detected, of which 21 were in coding sequences (Appendix VII Table A 18). However, six were in one gene (CLK_RS01880) and nine in another gene (CLK_RS01900). For both the coverage was quite poor (only 11-16 reads), so it is questionable whether the mutations are true or just a result of low sequencing quality. To confirm the sequence of both genes, they should be checked with Sanger sequencing. If the above 15 mutations are ignored, only 7 mutations are left; one in a non-coding region, two frameshift mutations and four SNVs. The frameshift mutations detected in the SBRC sequence restore the ORF, hence it was assumed that these two genes were actually frameshifted in the genome of the published strain (accession number: NC_010520). For the four SNVs no evidence was available to determine the most likely WT allele and, therefore, no further conclusions can be drawn.

NCTC 2916: 9 variants were detected in total; five in coding sequences, of which two resulted in a frameshift (Appendix VII Table A 19). The five base pair insertion at position 241436 of contig 7 in the *ltrA* gene restores the ORF, thus the mutation is actually in the published genome rather than in the NCTC 2916 genome of the SBRC stock. The second frameshift mutation, at position 190 of contig 4, is in a gene annotated as a hypothetical protein. It encodes a peptide consisting of only 39 amino acids. It is unclear which of the two alleles, SBRC or NCBI, is correct. One of the SNVs (165425, contig 2) is in a prophage gene and is thus considered irrelevant. Interestingly, five mutations have quite a low (<85%) frequency, including the insertion at position 241436. The low frequencies might be due to sequencing or alignment errors, though it could also signify a mixed population, with both alleles being common in the NCTC 2916 population. The mixed population theory can be verified by Sanger sequencing the regions of interest of several independent colonies.

Table 5.4 Comparison of SNPs detected between 62A NCBI (CFSAN064329) and 62A NCTC (NCTC 3805) to the published genomes of ATCC 3502 and ATCC 19397.

Reference Position*	Overlapping annotations*	Coding region change**	Amino acid change**	Comparison to ATCC 3502 and ATCC 19397	The proposed WT***
275142	Gene: *rnr*	968 G>A	Ser323Asn	3502/19397 – Asn. The rest of this 62A NCBI gene is 100% identical to 3502/19397.	968A (Asn)
811474	Gene: *gcvT*	392 A>G	Glu131Gly	3502/19397 – Gly. The rest of this 62A NCBI gene is 100% identical to 3502/19397.	392G (Gly)
922830	Gene: *bont*	2203 A>G	Thr735Ala	3502/19397 – Ala. The rest of this 62A NCBI gene is 100% identical to 3502/19397.	2203G (Ala)
1055392	Gene: *cbiG*	167 G>A	Gly56Glu	3502/19397 – Glu. The rest of this 62A NCBI gene is 100% identical to 3502/19397.	167A (Glu)
1618151	Gene: *nadC*	771 C>T		3502/19397 – A (silent mutation). Codon usage: AGG (62A NCBI) is 4%, AGA (62A NCTC) is 24%.	771T
1812816	Gene: *lpdA*	1028 T>C	Val343Ala	3502/19397 – Ala. The rest of this 62A NCBI gene is 100% identical to 3502/19397.	1028C (Ala)
2039076	Gene: *cloSI*	1146 A>G		3502/19397 – A (silent mutation). Codon usage: GAA (62A NCBI) is 60%, GAG (62A NCTC) is 14%. The rest of this 62A NCBI gene is not 100% identical to 3502/19397.	1146A
2382866	Gene: *hypB*	314 A>G	Asn105Ser	3502/19397 – A. The rest of this 62A NCBI gene is not 100% identical to 3502/19397.	314A (Asn)
2503307	Gene: *polC*	1705 G>A	Ala569Thr	3502/19397 – A. The rest of this 62A NCBI gene is not 100% identical to 3502/19397.	1705A (Thr)
2503780	Gene: *polC*	1232 G>A	Gly411Asp	3502/19397 – A. The rest of this 62A NCBI gene is not 100% identical to 3502/19397.	1232A (Asp)
2543138	Gene: *prdB*	10 A>G	Ile4Val	3502/19397 – Val. The rest of this 62A NCBI gene is 100% identical to 3502/19397.	10G (Val)
2921204	Gene: *glsA*	218 C>T	Thr73Ile	3502/19397 – Ile. The rest of this 62A NCBI gene is 100% identical to 3502/19397.	218T (Ile)
3043216	Gene: *hydA*	1341 G>A		Multiple differences in this region between the 62A strains and 3502/19397.	Unknown
3743334	Gene: *fusA*	801 C>T		3502/19397 – T (silent mutation). Codon usage: GGC (62A NCBI) is 5%, GGT (62A NCTC) is 20%. The rest of this 62A NCBI gene is 100% identical to 3502/19397.	801T
3749262	Gene: *rpoB*	3642 A>T	Glu1214Asp	3502/19397 – Asp. The rest of this 62A NCBI gene is 100% identical to 3502/19397. Mutation is not in DNA region associated with Rif^R.	3642T (Asp)

* The reference genome was 62A NCBI (CFSAN064329). The annotations are from the reference strain.

** The first nucleotide/amino acid code is the reference strain and the second nucleotide/amino acid code is the 62A NCTC strain (NCTC 3805).

*** The bases (amino acids) listed in this column are considered the original WT sequence, before mutations occurred due to the *mutS* mutation in 62A NCBI (CFSAN064329).

5.3.4 Genetic stability of *C. botulinum* type A strains

The re-sequencing analysis of the WT strains confirmed that all *mut* family genes were free of detrimental mutations. Even though this was considered encouraging information regarding the genetic stability of the strains, mutations in other proteins important for maintaining genome integrity, for example DNA-polymerases, can significantly affect mutation rates.[392, 453] Therefore, the genetic stability of the strains was experimentally assessed by determining their mutation frequency based on spontaneous arising rifampicin resistant mutants within the population. The fluctuation assay was performed as described in section 2.9.6 and the results are show in Figure 5.3. In general, all strains can be classified as weak mutators ($4 \times 10^{-8} \leq f < 4 \times 10^{-7}$), with ATCC 3502^{+} at the top end of the normomutable ($8 \times 10^{-9} < f < 4 \times 10^{-8}$) category.[384-386, 544]

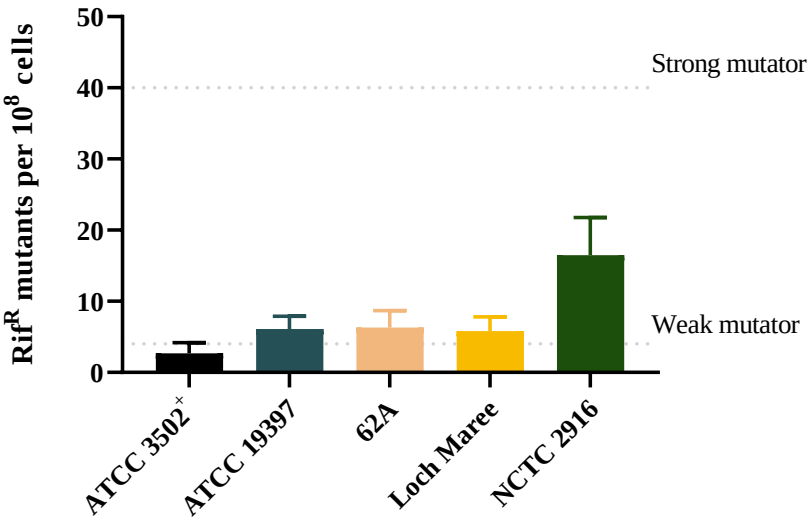


Figure 5.3 Mutation frequency of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916. Frequency of spontaneous Rif^{R} mutants in WT strains. Cells were cultured for 24 hours in TYG broth and subsequently plated onto TYG (total CFU) and TYG supplemented with 25 µg/ml rifampicin (Rif^{R} CFU). Mutation frequencies were calculated by dividing the Rif^{R} CFU by the total CFU. The dotted lines indicate the mutator strength categories. Error bars represent the standard error of the mean of 10 single colonies, cultured in triplicates.

Although for strain engineering purposes mutation frequencies within the normomutable range are preferred, the slightly higher frequencies observed for the type A strains were deemed sufficiently low to generate reliable safe strains. This is mainly because the determined mutation frequencies are very close to the normomutable threshold, and bearing in mind that the accuracy of low mutation frequency values strongly depends on the number of replicates, the calculated mutation frequency might be slightly on the high end considering the limited number of independent cultures per strain used in this study (Figure 5.4).[599] Rosche and Foster recommend that when the estimated number of mutations per culture, *m*, is between 0.5 and 1.5 roughly 65 down to 25 separate cultures (the higher *m*, the lower the number of required cultures) are required to calculate reliable values. In this study only 30 independent cultures per strain was practically feasible due to space restrictions inside the anaerobic cabinet and the time consuming nature of such an experiment. To categorize the mutability of the strains with higher accuracy, the number of replicate cultures could be increased in the future.

As mentioned above, the number of Rif^R colonies per culture was very low, consequently just a few false positives growing due to the high load of cells plated would have a significant impact on the data, resulting in an overestimation of the actual mutation frequency. To confirm that the colonies growing on the rifampicin plates had actually acquired mutations in the *rpoB* gene, associated with rifampicin resistance, various colonies were sent off for Sanger sequencing. Initially the full *rpoB* gene of two colonies per strain was sequenced using six primers, rpoB_scr_F1/F5 and rpoB_scr_R1, which were universal for all strains. All mutations were identified in the sequence covered by the rpoB_scr_F3 primer, therefore the remaining colonies were sequenced using this primer only (Figure 5.5). All the analysed colonies had the

anticipated mutations in the *rpoB* gene, associated with rifampicin resistance in other bacterial species, confirming that the used concentration of rifampicin was sufficient to eliminate false positives.[600-602 539]

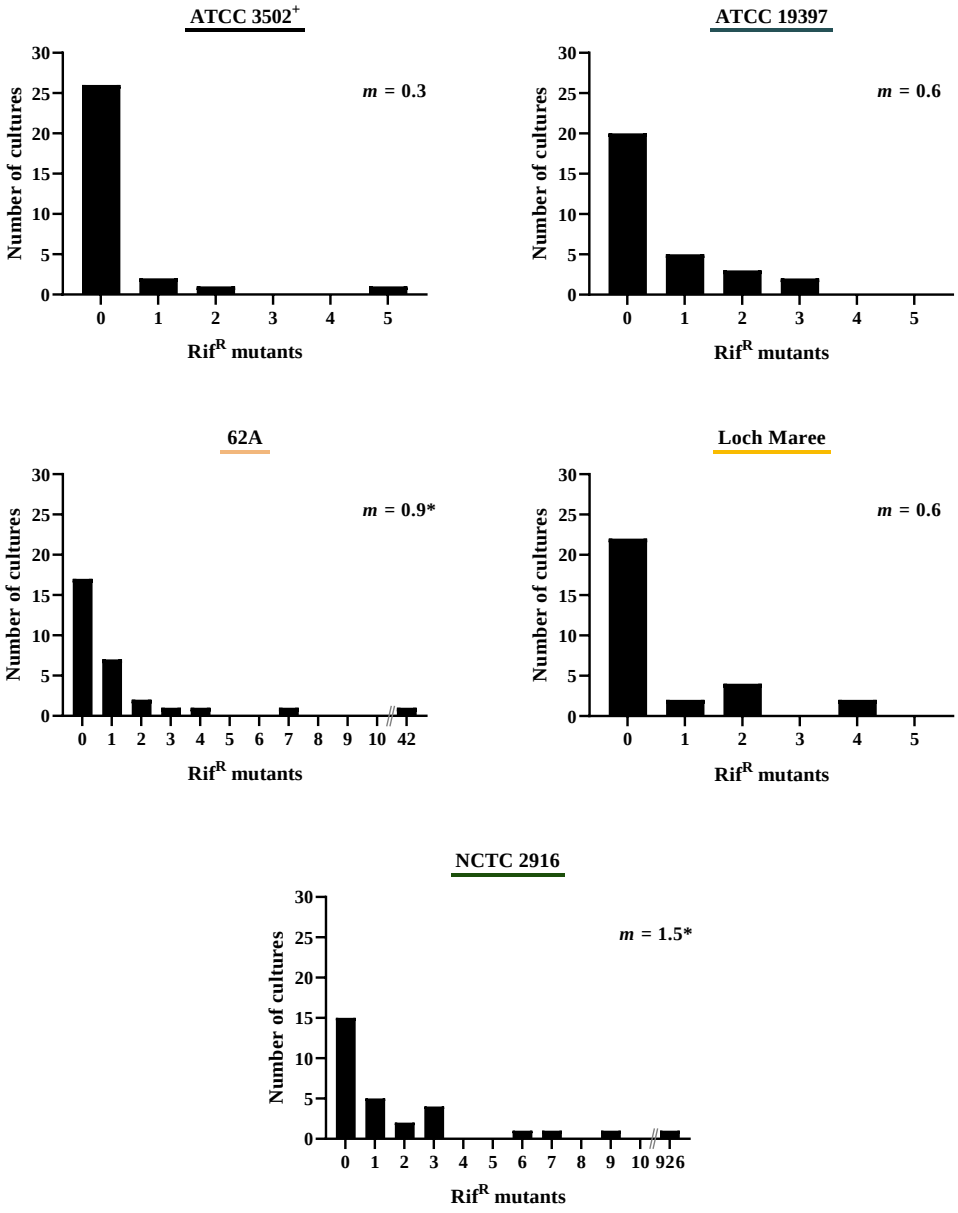


Figure 5.4 Frequency distribution of the number of Rif^R mutants per culture for ATCC 3502$^+$, ATCC 19397, 62A, Loch Maree and NCTC 2916. Histograms of spontaneous Rif^R mutants per culture for a total of 30 cultures per WT strains. *m* is the average number of Rif^R mutants per culture and the * indicates that any outliers (42 and 926 for 62A and NCTC 2916, respectively) were excluded from the calculation. Cells were cultured for 24 hours in TYG broth and subsequently plated onto TYG (total CFU) and TYG supplemented with 25 µg/ml rifampicin (Rif^R CFU).

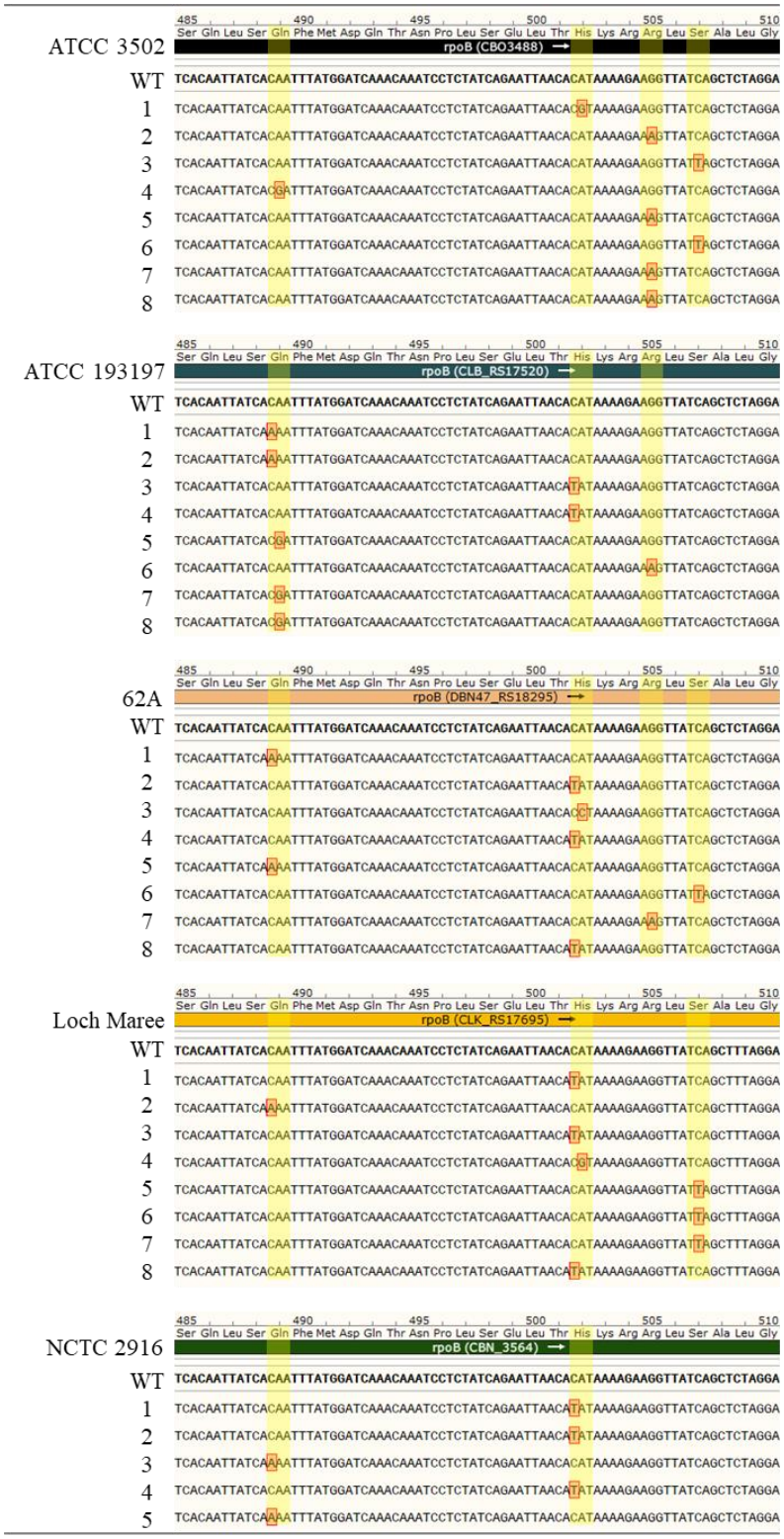


Figure 5.5 Mutations identified in the *rpoB* gene of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916. Sanger sequencing results of the *rpoB* gene in Rif^{R} colonies for all WT strains. The sequencing reads were aligned to the WT sequence using SnapGene. Mutated nucleic acids are highlighted in red and the affected codons are highlighted in yellow.

5.3.5 Spore heat resistance and investigating the role of the *spoVA*^{2} gene cluster

Spore heat resistance was the final characteristic assayed before selecting strains for the safe strain library. To obtain concentrated spore suspensions overnight cultures were plated on CSA medium and incubated for 14 days at 37°C, as described in section 2.9.5.[224] The formed spores were harvested, washed and the final viable spore count in each suspension was determined using the standard heat treatment protocol (15 min at 80°C), as described in section 2.9.4. The spore suspensions were stored at 4°C for 7 days before conducting the heat resistance assay. To establish the heat resistance of the spores, the spore suspensions were split into 50 µl aliquots and incubated at 99°C. Samples were removed from the PCR machine at different time points and plated on TYG agar plates. Depending on the strain, CFUs were enumerated after 24 or 48 hours of incubation at 37°C. For more details about the heat resistance assay see section 2.9.5. Both the spore generation process and the heat treatment method were designed in collaboration with NIZO food research institute.

The spore survival curves are depicted in Figure 5.6 and the linear part of the curves was used to calculate the $D_{99°C}$ values (time required for one decimal reduction) as described in section 2.9.5. As anticipated, the NCTC 2916 spores were the most heat-resistant with a *D*-value of 7.99 min, while for all other strains, lacking the *spoVA*^{2} genes, $D_{99°C}$ values were below 1 min.[592, 594] Spores produced by the Loch Maree strain were the most heat-sensitive, with a *D*-value of 0.51 min. The *D*-values determined for the ATCC 3502$^+$ and ATCC 19397 spores were very similar, 0.66 and 0.62 min, respectively, which again highlights their high degree of similarity. The 62A spores were slightly more heat-resistant, with a *D*-value of 0.74 min. Although it is difficult to compare *D*-values between studies because of the many variables that should be

taken into consideration (e.g., media, heat treatment and pressure), the *D*-values determined in this study are relatively low compared to what has been published so far, yet within the range of what could be expected.[175, 575, 589, 603-604] To increase the accuracy of the calculated *D*-values more data points should be collected.

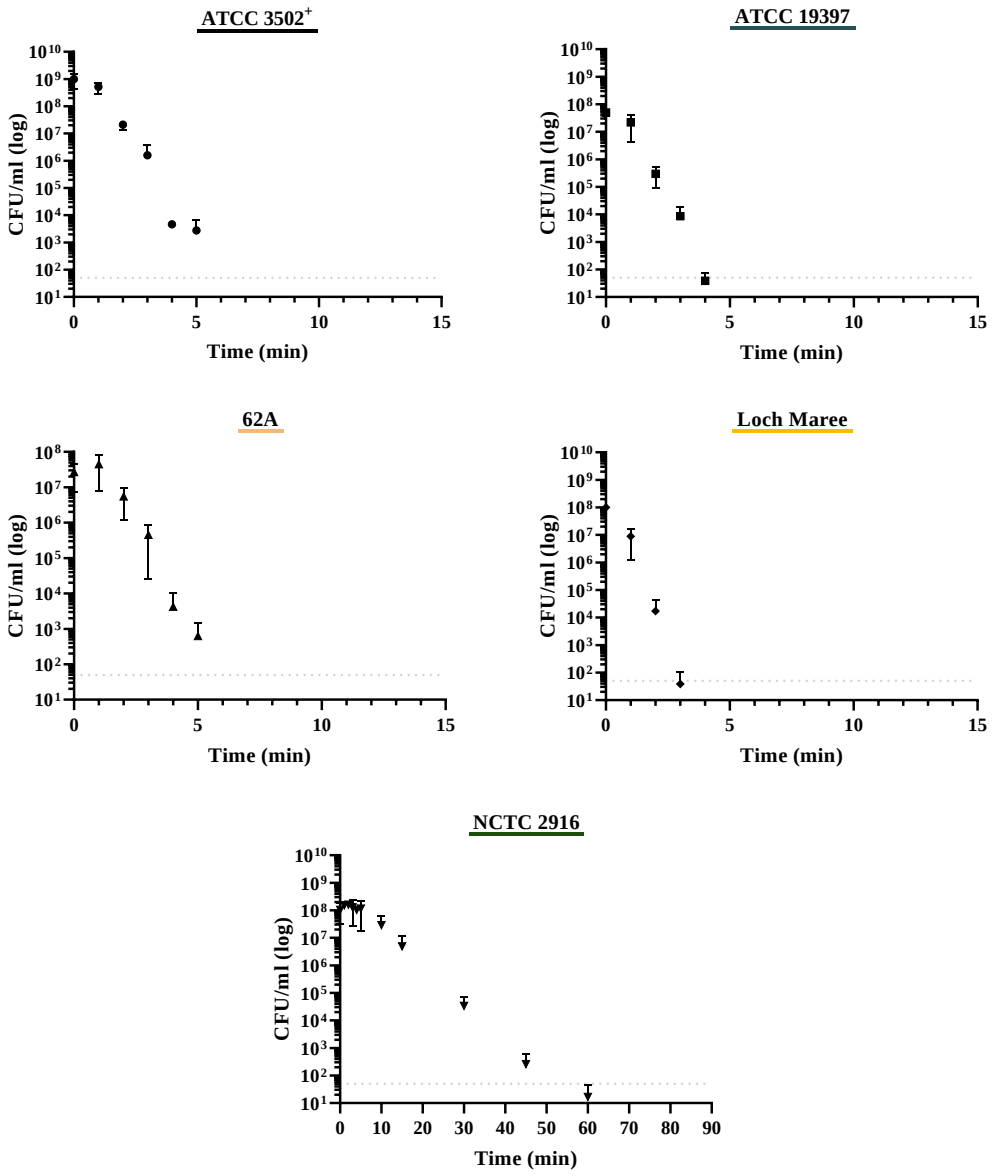


Figure 5.6 Survival of spores of ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916 after 99°C thermal treatment. The number of viable spores withstanding a 99°C heat treatment for a specific amount of time. Spores were harvested after culturing the strains for 14 days on CSA medium at 37°C. Error bars represent the standard deviation of three independent experiments. The dotted line at 50 CFU/ml represents the detection limit.

Surprisingly, in a study published by Bull *et al*. spores from the 62A strain were more heat resistant than NCTC 2916 spores.[575] Their findings contradict the results presented here as well as other studies that show a correlation between the presence of *spoVA*^{2} genes and increased spore heat-resistance.[592-594, 605-606] The most likely explanation for this discrepancy is the different media used. Media composition can strongly effect spore properties, including heat-resistance. Specifically, media lacking particular nutrients or with a very low Ca^{2+} concentration, produce mature spores containing no or low Ca-DPA levels.[607-610] The final Ca-DPA concentrations within mature spores influences their survival when exposed to high temperatures; more Ca-DPA leads to higher heat resistance. The *spoVA*^{2} operon, which confers high-level heat resistance to spores, has been linked to an increased DPA uptake during sporulation.[592] Possibly, the growth conditions used in the study by Bull *et al*. did not facilitate the accumulation of sufficient Ca-DPA and/or the uptake of Ca^{2+} and DPA during the sporulation process, hence the presence of the *spoVA*^{2} operon in the NCTC 2916 strain did not result in spores that were significantly more heat resistant compared to 62A.[575] However, it should be noted that the overall $D_{99°C}$-values determined by Bull *et al*. were much higher in general and the reason for the discrepancies is not necessarily evident, although it could be due to the growth medium.

Furthermore, there was a small increase in the number of viable spores from 0 to 1 min for strains 62A and NCTC 2916. This increase is a common phenomenon, known as spore heat activation.[611-612] ATCC 3502, ATCC 19397 and Loch Maree viable spore counts did not increase after 1 min of heat treatment, though the decline in the first minute is less compared to the decrease in the following minutes. Heat activation probably still occurred but the 1 min incubation at 99°C was ‘too’ lethal, thus masking

the portion of the population initially activated by the heat, before they were inactivation. In addition, the NCTC 2916 survival curve is non-linear in the final section. This frequently observed tailing in survival curves has been attributed to the existence of a subpopulation of more heat-resistant spores.[613]

As mentioned before, studies in *bacillus* species have shown that the presence of one or more *spoVA*^{2} operons increases spore heat resistance and the genetic characterisation of highly heat resistant *C. sporogenes* strains revealed the existence of a homologues operon in *Clostridium.*[592, 594] However, the role of the clostridial *spoVA*^{2} genes in spore heat resistance had not yet been experimentally validated. To do so, CRISPR/Cas9-mediated deletion of the full *spoVA*^{2} operon was carried out as described in section 2.12.3, using plasmid pRECas1-spoVA2-G1, whose construction is outlined in section 2.11.3. The deletion was confirmed by PCR in three biological replicates from three independent conjugations using primers annealing to the chromosome, flanking the editing templates (Figure 5.7). The plasmid was lost from the selected NCTC 2916 Δs*poVA*^{2} mutants, as described in section 2.8. The deletion in the final mutants was subsequently verified by Sanger sequencing.

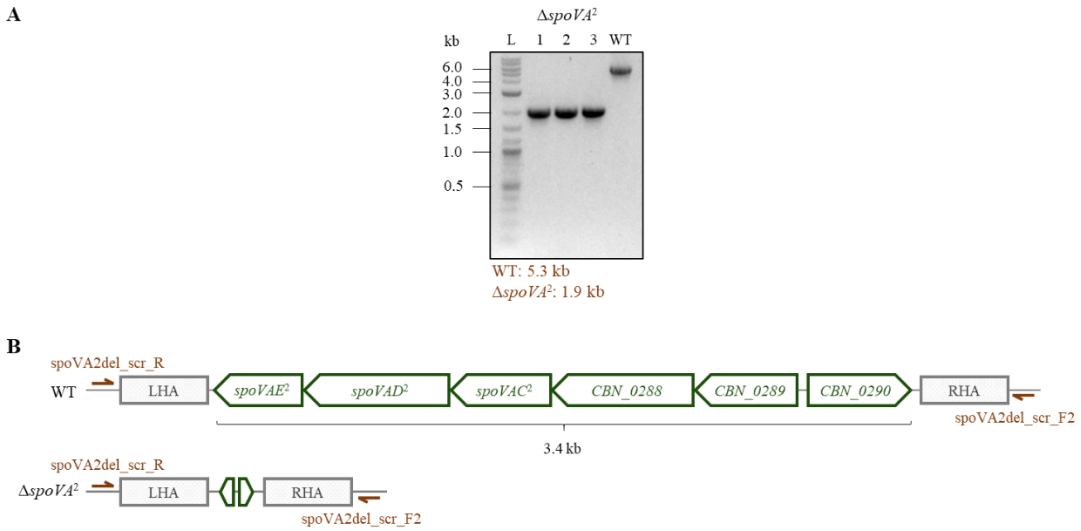


Figure 5.7 Deletion of the *spoVA*^{2} operon in strain NCTC 2916. (**A**) PCR for three biological replicates of NCTC 2916 Δ*spoVA*^{2} mutants alongside WT. Products were amplified using flanking primers spoVA_scr_F2 and spoVA2_scr_R. Lane L: NEB 1 kb Plus DNA Ladder. (**B**) Schematic diagram of the targeted chromosomal region showing the primer binding sites in the WT and mutant strains.

To assess the growth, sporulation/germination capacity and spore heat resistance of the Δ*spoVA*^{2} strains, experiments were performed as described in sections 2.9.2, 2.9.4 and 2.9.5. The growth profile of the mutants was identical to WT (Figure 5.8A). Interestingly, the viable spore count of all three biological replicates was slightly higher (2 to 4-fold) than WT (Figure 5.8B); however, the increase was only statistically significant at 48, 72 and 120 hours but not at 92 hours (two-tailed unpaired t-test, *$p \leq 0.0332$). It suggests that the presence of *spoVA*^{2} imposes a small burden on the sporulation/germination process. It should be noted that there was quite some variation within the dataset and therefore further experiments will be necessary to establish if the presence of *spoVA*^{2} genes has any negative impact on the total number of spores formed or affects the ability of spores to germinate before definitive conclusions can be drawn.

As expected, the Δ*spoVA*^{2} strains yielded spores that were much more readily heat-inactivated compared to WT spores (P<0.0001)(Figure 5.8C); the $D_{99°C}$-value for the mutant spores was 2.24 min while for the WT strain this was 7.99 min.[593-594] Although the deletion of the operon decreased the spore's heat resistance, they were not as heat sensitive as the other type A strains described above. This suggests that whilst the operon plays a major role, there are other factors/genes contributing to the high heat resistance of the NCTC 2916 spores. A more in-depth analysis of the different type A genomes combined with the generation of a large transposon mutant library (i.e., TraDIS library) in the Δ*spoVA*^{2} strain could be used to identify these genes.[330, 614] Subjecting the mutant library to different heat treatments and subsequently comparing the genomes of the survivors as well as to the non-heat treated library should reveal all genes that were essential to survive the more harsh heat treatment.

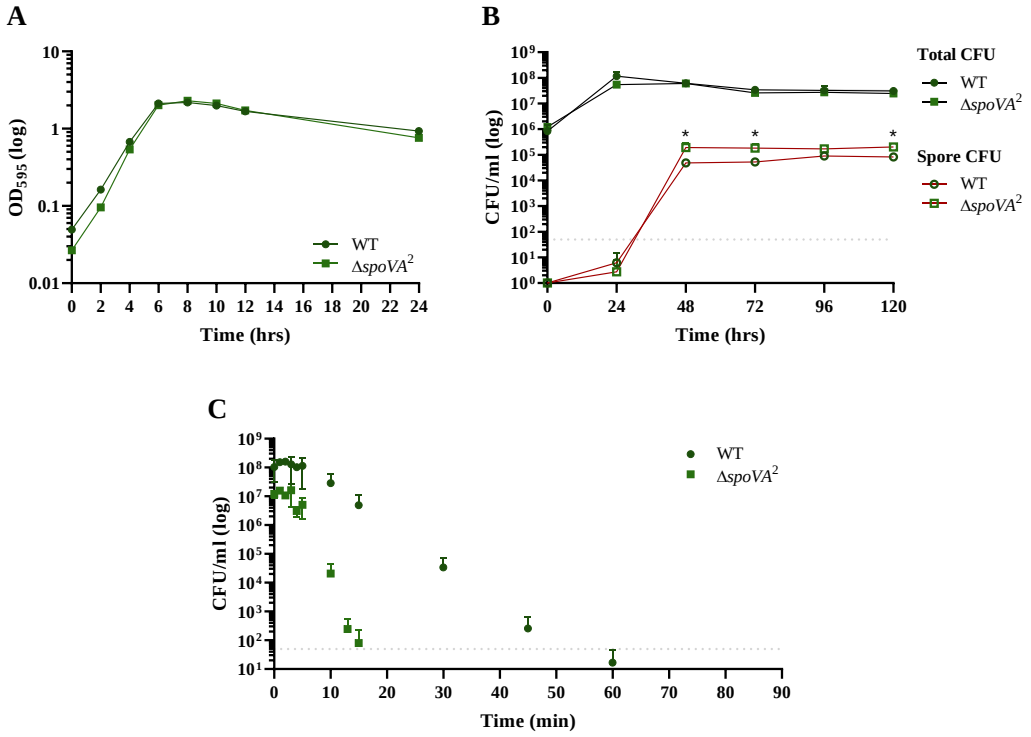


Figure 5.8 Phenotypic characterisation of NCTC 2916 Δ*spoVA*^{2}. (**A**) Growth profiles of Δ*spoVA*^{2} and WT. Strains were cultured in TYG broth and the OD_{595} was monitored for 24 hours. Error bars represent the standard deviation of three independent cultures for WT and of three biological replicates for Δ*spoVA*^{2}. (**B**) Sporulation/germination capacity of Δ*spoVA*^{2} and WT. Strains were cultured for 120 hours in TYG broth and samples were taken every 24 hours. Total CFU corresponds to the total cell count (non-heat treated) and the spore CFU is the number of viable spores withstanding a 15 min, 80°C heat treatment. Error bars represent the standard deviation of three independent cultures for WT and of three biological replicates for Δ*spoVA*^{2}. The dotted line at 50 CFU/ml represents the detection limit. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test. (**C**) Survival curves of Δ*spoVA*^{2} and WT. Spores were harvested after culturing the strains for 14 days on CSA medium at 37°C. Spores were heat-treated for a specific amount of time at 99°C. Error bars represent the standard deviation of three independent cultures for WT and of three biological replicates for Δ*spoVA*^{2} from three independent experiments. The dotted line at 50 CFU/ml represents the detection limit.

5.3.6 Generation and characterisation of non-toxigenic ATCC 19397 strains

Amongst the strains characterised in this chapter, ATCC 19397 was the most similar to ATCC 3502, both genetically and phenotypically. In addition, the strain is genetically stable and conjugal DNA transfer has been established using a customized donor strain, enabling its genetic manipulation (see Chapter 3). Therefore, ATCC

19397 was considered to be the most appropriate strain to substitute the unstable ATCC 3502 for the *C. botulinum* safe strain project as well as for other future studies.

To generate the three safe strain configurations described in section 5.1.3, the previously described theophylline inducible CRISPR/Cas9 genome editing tool was used.[158] The *boNT/A1* gene and the full toxin gene cluster deletion was carried out as described in section 2.12.3, using plasmids pRECas1-19_boNT/A1-G1 and pRECas1-19_ClA-G1, respectively, whose construction is outlined in section 2.11.3. The deletions were confirmed by PCR in three biological replicates from three independent conjugations for each strain using primers annealing to the chromosome, flanking the editing templates (Figure 5.9). The primers flanking the full gene cluster only yield a PCR product when the genes are deleted because the size of the WT product exceeded the maximum capacity of the polymerase utilized. The plasmid was lost from the selected mutants, as described in section 2.8. The deletions in the final mutants were subsequently verified by sequencing.

To create the toxoid strain small changes were made to the original CRISPR/Cas9 genome editing protocol described in section 2.12.3, specifically to the editing template design part. During standard gene deletions or replacements events, the CRISPR/Cas9 target site is removed from the chromosome, preventing its cleavage and thus allowing cell survival. In this case however, only a few point mutations were desired while the remainder of the gene should stay intact. Therefore, the desired mutations (R363A/Y366F) should ideally be within the PAM and/or SEED region of the sgRNA, so that upon homologous recombination the point mutations are introduced and simultaneously the target site is removed. Unfortunately, the region containing the R363 and R366 codons did not overlap with a suitable Cas9 target and an alternative strategy had to be pursued. This involved the selection of a target site

just upstream of these codons and removing the Cas9 target from the editing template by introducing a silent mutation in the PAM sequence (G1053A). For this approach to facilitate efficient mutant selection the Cas9 target site should be as close as possible to the mutation site to minimise the chances of homologous recombination occurring in between the two, which would result in false positives carrying the PAM mutation but lacking the desired codon changes. In this case the Cas9 target site and the point mutations were 32 bp apart.

It is worth nothing that codon changes in a gene, though silent, can influence the rate of translation, RNA stability and protein folding, thereby affecting expression and/or activity levels. Hence when introducing a silent mutation, if possible, codons with a similar frequency are preferred.[523] Here, the PAM sequence partially overlapped with a glutamic acid (351) codon, which is only encoded by two codons with a strong usage bias towards one of them. Therefore selecting a codon with a similar frequency was impossible; the original codon GAG (codon frequency 0.19) was changed into GAA (codon frequency 0.81).

Alongside the editing template described above, with one silent mutation in the PAM, two other editing templates were tested; one with five silent mutations in the SEED sequence and a second combining the SEED and PAM mutations (see Appendix IV). The rationale behind designing three different editing templates was that potentially Cas9 would still target the cleavage site, albeit less efficiently, if only the PAM or SEED sequence was changed, resulting in plasmid degradation and accidental killing of correct mutants. No differences were observed between the efficiencies of the three editing templates (results not shown) and the mutants with only one silent mutant in the PAM sequence were taken forward.

To screen for successful i*boNT/A1*^{RY} (inactive *boNT/A1* R363A/Y366F) mutants a primer annealing to the mutated codons (boNT/A1-toxoid_scr_R) was combined with the flanking primer boNT/A1_scr_F, whereby the primers only yield a product if mutagenesis has been successful (Figure 5.9). The plasmid was lost from the selected mutants, as described in section 2.8. The genomic modifications in the final mutants were subsequently verified by Sanger sequencing.

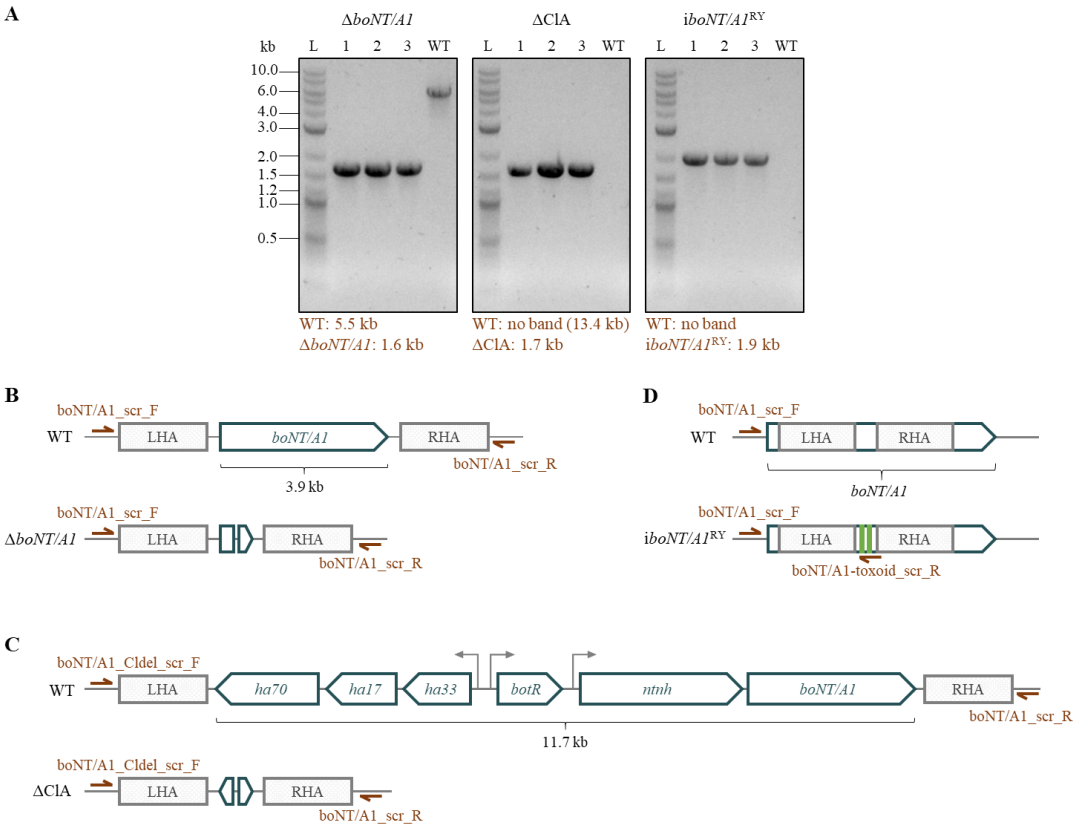


Figure 5.9 Deletion of *boNT/A1*, *boNT/A1* gene cluster (ClA) and generation of i*boNT/A1*^{RY} in strain ATCC 19397. (**A**) PCR for three biological replicates of ATCC 19397 Δ*boNT/A1*, ΔClA and i*boNT/A1*^{RY} mutants alongside WT. Products were amplified using primers boNT/A1_scr_F and boNT/A1_scr_R (Δ*boNT/A1*), boNT/A1_Cldel_scr_F and boNT/A1_scr_R (ΔClA), boNT/A1_scr_F and boNT/A1-toxoid_scr_R (toxoid). Lane L: NEB 1 kb Plus DNA Ladder. (**B/C/D**) Schematic diagram of the targeted genomic region showing the primer binding sites in the WT and mutant strains. (**D**) Green lines indicate the R363A and Y366F mutation sites in the light chain. The silent mutation in the PAM is not indicated in this diagram.

To confirm that the various safe strains retained the characteristics of the parent strain except toxicity, the growth and sporulation/germination capacity of the mutants was assessed as described in sections 2.9.2 and 2.9.4, respectively. The growth profile of all mutants was identical to WT (Figure 5.10A). The sporulation/germination profile of the ΔClA and i*boNT/A1*^{RY} mutant strains was indistinguishable from WT (Figure 5.10C/D). Interestingly, for two Δ*boNT/A1* biological replicates the viable spore count from 48 h onwards was slightly lower (4-fold) and for one replicate (biological no. 1) even 20-fold lower compared to WT (Figure 5.10B and Appendix V). Biological no. 1 was excluded from the dataset.

It is possible that the observed decrease is due to additional randomly acquired mutations elsewhere in the genome, rather than the deletion of the toxin, particularly for Δ*boNT/A1*-no1. To check this, the entire genome of all strains was sequenced and compared to the published WT genome, as described in section 2.10.10. As anticipated, very few or no additional mutations were detected in the ΔClA and i*boNT/A1*^{RY} mutants (Appendix VII Table A 21 and Table A 22). The very low number of additional mutations acquired by these safe strains during the whole mutant generation process support the findings discussed in section 5.3.4 regarding the genetic stability of ATCC 19397.

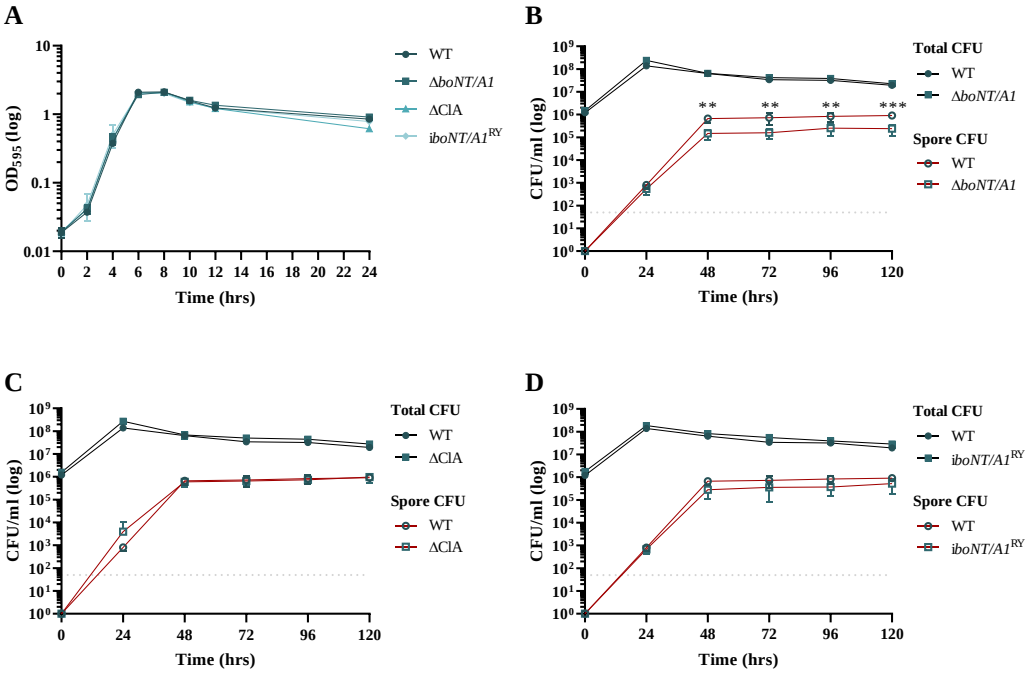


Figure 5.10 Phenotypic characterisation of ATCC 19397 Δ*boNT/A1*, ΔClA and i*boNT/A1*^{RY}. (**A**) Growth profiles of Δ*boNT/A1*, ΔClA, i*boNT/A1*^{RY} and WT. Strains were cultured in TYG broth and the OD_{595} was monitored for 24 hours. Error bars represent the standard deviation of three independent cultures for WT and of three biological replicates for the mutants. (**B/C/D**) Sporulation/germination capacity of Δ*boNT/A1* ΔClA and i*boNT/A1*^{RY} compared to WT. Strains were cultured for 120 hours in TYG broth and samples were taken every 24 hours. Total CFU corresponds to the total cell count (non-heat treated) and the spore CFU is the number of viable spores withstanding a 15 min, 80°C heat treatment. Error bars represent the standard deviation of three independent cultures for WT, of two biological replicates for Δ*boNT/A1* and of three biological replicates for ΔClA and i*boNT/A1*^{RY}. The dotted line at 50 CFU/ml represents the detection limit. ns≤0.1234, *p≤0.0332, **p≤0.0021, ***p≤0.0002, ****p≤0.0001, two-tailed unpaired t-test.

Surprisingly, the same was true for the Δ*boNT/A1* strains. In Δ*boNT/A1*-no1 four additional mutations (all SNVs in coding regions, resulting in an amino acid change) were identified, three being unique (Appendix VII Table A 20). One of the SNVs (position 2061097) was also detected in Δ*boNT/A1*-no2. This mutation might have emerged independently in both strains, but a more likely scenario is that this mutation was already present in a subpopulation of the original WT stock. The three unique SNVs were in genes encoding proteins that have not been previously associated with

sporulation or germination. However, one of the affected proteins is a magnesium-translocating P-type ATPase; magnesium is accumulated in the spore core during the sporulation process and varying concentrations of magnesium in growth medium affects spore yields.[615] It is plausible that the amino acid change in this ATPase negatively impacts its translocation abilities, and thus causing the lower viable spore yield. To establish if SNV 2984334, the other SNVs or possibly a combination of them, decreases spore viability, the mutations should be re-created in the WT strain in a future experiment.

In the other two Δ*boNT/A1* replicates, the small but significant decrease in viable spore counts could not be accounted for by the presence of random additional mutations (Appendix VII Table A 20). Both strains had acquired only one additional mutation, in genes not linked to sporulation or germination; Δ*boNT/A1*-no2 acquired a SNV in a predicted MurR/RpiR transcriptional regulator, resulting in an amino acid change and Δ*boNT/A1*-no3 acquired a silent mutation in an uncharacterized histidine kinase.

It was hypothesized that there could be a direct link between these lower spore yields and the absence of the toxin protein. However, previous studies describing non-toxigenic *C. botulinum* strains generated using directed mutagenesis reported that removal or inactivation of the neurotoxin had no significant impact on the relevant phenotypic characteristics of the parent strain, including sporulation.[126, 529] Furthermore, the sporulation/germination capacity of the ΔClA and i*boNT/A1*^{RY} strains created in this study was not compromised. Hence, there is probably no direct connection between the lack of neurotoxin and the decrease in viable spores. Nevertheless, there might be an indirect correlation, which is yet to be elucidated. Possibly the non-toxic neurotoxin-associated proteins, now unable to form a complex with BoNT/A1, affects other cell processes.

Albeit not completely perfect, the strain Δ*boNT/A1*-no3 was considered a suitable strain for future studies, having acquired only one additional (silent) mutation, an identical growth profile compared to WT and produces sufficient numbers of viable spores.

Finally, the supernatants of the *ΔboNT/A1* and ΔClA strains were analysed by Western blot to confirm the absence of the neurotoxin. The supernatants were prepared and Western blotting was performed as described in section 2.9.8. The ideal time point to harvest samples is when the neurotoxin concentration is at its maximum. Previous studies report that neurotoxin levels generally increase until approximately 24-48 hours and depending on the culturing conditions (e.g., medium composition, temperature) as well as the strain, toxin concentrations remain more or less constant after the first 48 hours.[170, 588, 616] Neurotoxin activity in supernatants of strain ATCC 19397 peaked between 20 and 30 hours when grown at 37°C.[14] To confirm that under the growth conditions applied here type A toxin concentration is also at its maximum after approximately 24 hours, the BoNT/A1 concentration in supernatants at 24 and 48 hours were compared via Western blotting.

The bands corresponding to the BoNT/A1 light chain (50 kDA) were clearly visible at both 24 and 48 hours, but the band from samples harvested after 24 hours appeared to be slightly stronger (Figure 5.11A). In addition, the band at 150 kDa, corresponding to the non-cleaved neurotoxin peptide, was stronger at 24 hours. Based on these results it was decided to harvest the samples for the safe strains at 24 hours. As expected, no BoNT/A1 was detected for the Δ*boNT/A1* and ΔClA strains (Figure 5.11B).

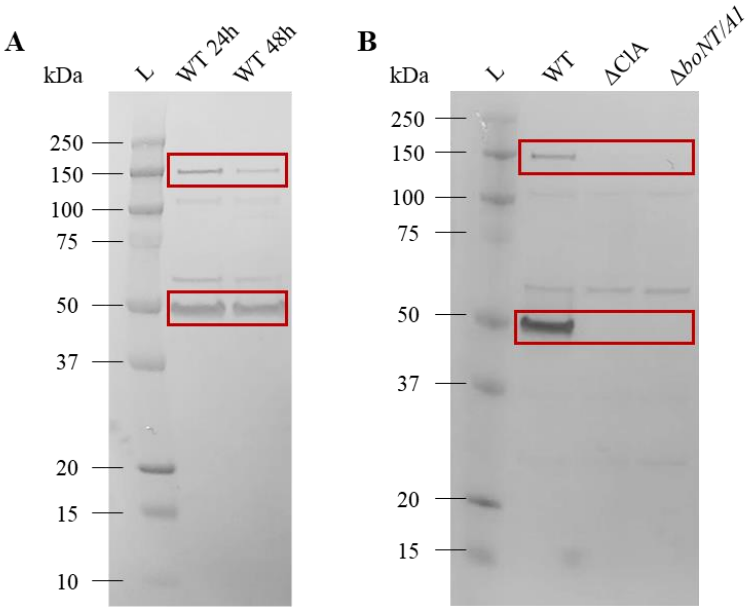


Figure 5.11 Western blot analysis of ATCC 19397 Δ*boNT/A1*, ΔClA and WT supernatants. Supernatants (concentrated 10-fold) were subjected to SDS-PAGE/Western blot with polyclonal *C. botulinum* BoNT-A Light Chain Antibody. (**A**) ATCC 19397 WT, samples were taken after 24 and 48 hours of incubation. (**B**) ATCC 19397 Δ*boNT/A1*, ΔClA and WT, samples were taken after 24 hours of incubation. Lane L: Bio-Rad Precision Plus Protein All Blue Standards ladder; red box at 150 kDa: BoNT-A Light Chain and heavy chain (no proteolytic cleavage); red box at 50 kDa: BoNT-A Light Chain.

The i*boNT/A1*^{RY} supernatants were not subjected to Western blot analysis because it would not provide any insightful information regarding exact expression levels or toxicity. Instead, to confirm that neurotoxin expression has not been compromised in the toxoid strains i*boNT/A1*^{RY} could be quantified using an ELISA assay. Additionally, to asses if the silent PAM mutation changes the toxoid's fold, its crystal structure could be elucidated, though any noticeable effects of this one mutation are highly unlikely.

Although the i*boNT/A1*^{RY} strain created in this study fulfils all requirements to be excluded from the select agent regulations based on existing literature, the way the toxoids were created as well as the inactivated neurotoxin itself could be improved.[156-157, 577, 617] Recently, a new 'bookmark' approach that exploits CRISPR/Cas9 technology was published to engineer clostridia, which avoids the need of ancillary

silent mutations in the PAM and/or SEED sequence of the sgRNA.[618] With this two-step based strategy the toxin would first be replaced by a bookmark sequence, which represents a sgRNA, and subsequently this Cas9 target is substituted with a mutated *boNT/A1* gene. Additionally, this approach avoids the need of mutations to be inside or close to the Cas9 target sequence, facilitating the design and creation of BoNT variants. This new strategy, however, involves two cycles of mutant isolation instead of one, hence being more time consuming and increasing the risk of picking up undesired additional mutation. Therefore, whether the one-step approach relying on silent mutations in Cas9 target sequence or the bookmark strategy is preferred will depend on the strain (i.e., its genetic accessibility and genomic stability) and the location of the desired mutations.

Furthermore, in view of the regulatory hurdles faced by the non-toxigenic *C. botulinum* strains before they will be reclassified, due to the potency of the neurotoxin and its status as a select agent, any modifications to the protein that further minimize safety risks without compromising the toxin's structure, should definitely be considered. Over the last few years quite a few different amino acid substitutions have been identified that reduce the catalytic activity of BoNTs.[156-157] For example, the current i*boNT/A1*^{RY} strain could be improved by including a third mutation, E224Q, in the LC/A.[29, 619] Although the R363A/Y366F double mutation reduces neurotoxicity to undetectable levels at standard concentrations, adding the E224Q mutation results in no catalytic activity being detected at 30,000-fold higher concentrations than WT BoNT/A1.[29] Gu *et al*. also confirmed that introducing this third mutation has no significant impact on the structure of the neurotoxin compared to WT. Depending on the results of the mouse bioassay, the substitution of E224 and/or other amino acids in i*boNT/A1*^{RY} could be explored.

Overall, the ATCC 19397 Δ*boNT/A1*, ΔClA and i*boNT/A1*^{RY} strains created in this study are promising candidates to replace the toxigenic WT strain. Further steps should await *in vivo* evaluation of the safe strains, to establish whether there is no residual toxicity of the strains. Provided that the neurotoxin deficient strains are non-toxigenic, reclassification of these strains by UK authorities should be pursued. Hopefully downregulation will lift *C. botulinum* related research to the next level.

On a side note, because the type A strains discussed in this chapter are very similar, safe strain versions of the ATCC 3502 and 62A strain can easily be created using the same plasmids constructed to generate the non-toxigenic ATCC 19397 strains. Except for the full *boNT/A1* gene cluster deletion in 62A, due to a one nucleotide difference in the left homology arm. 62A is one of the most studied type A *C. botulinum* strains and because researchers probably prefer to continue working with the same strain, three independent biological replicates of the 62A Δ*boNT/A1* strain were created. Viable spore counts of all biologicals were identical to WT. Further characterisation of these strains is pending.

5.3.7 Generation and characterisation of non-toxigenic NCTC 2916 strains

The distinctive high heat resistance of NCTC 2916 spores, combined with its acceptable mutation frequency, decent viable spore counts and receptivity of foreign DNA makes it an excellent candidate for the safe strain collection. However, generating non-toxigenic versions of this strain is slightly more complex due to the presence of two neurotoxin gene clusters (see section 5.3.1). Quite a few non-toxigenic variations can be created and depending on the researchers' interest and its future applications different options might be preferred. Here, although the B neurotoxin is silent and can thus be considered safe, removal of both toxins as well as removal of

the complete toxin gene clusters was preferred for regulatory reasons. Hopefully, creating strains that are essentially indistinguishable from *C. sporogenes* (a non-pathogenic, category 1 organism) will aid the downregulation of these strains by the UK regulatory body (HSE).

To generate the two safe strain configurations the toxins/gene clusters were sequentially deleted as described in section 2.12.3, using the inducible CRISPR/Cas9 genome editing tool.[158] Construction of the four plasmids that were used to create these strains is outlined in section 2.11.3. Briefly, first four single mutants (Δ*boNT/A1*, Δ*boNT/B*, ΔClA and ΔClB) were created. The deletions were confirmed by PCR in three biological replicates from three independent conjugations for each strain using a primer pair of which at least one primer annealed to the chromosome just outside the editing templates (Figure 5.12). The primers flanking the full gene cluster only yield a PCR product when the genes are deleted because the size of the WT product exceeded the maximum capacity of the polymerase utilized. The plasmid was lost from the selected mutants, as described in section 2.8, and the deletions were subsequently verified by Sanger sequencing. The same steps were followed to create the double mutants, starting from the three Δ*boNT/B* and ΔClA strains. Unfortunately, for unknown reasons isolation of the third Δ*boNT/A1* Δ*boNT/B* double mutant was unsuccessful.

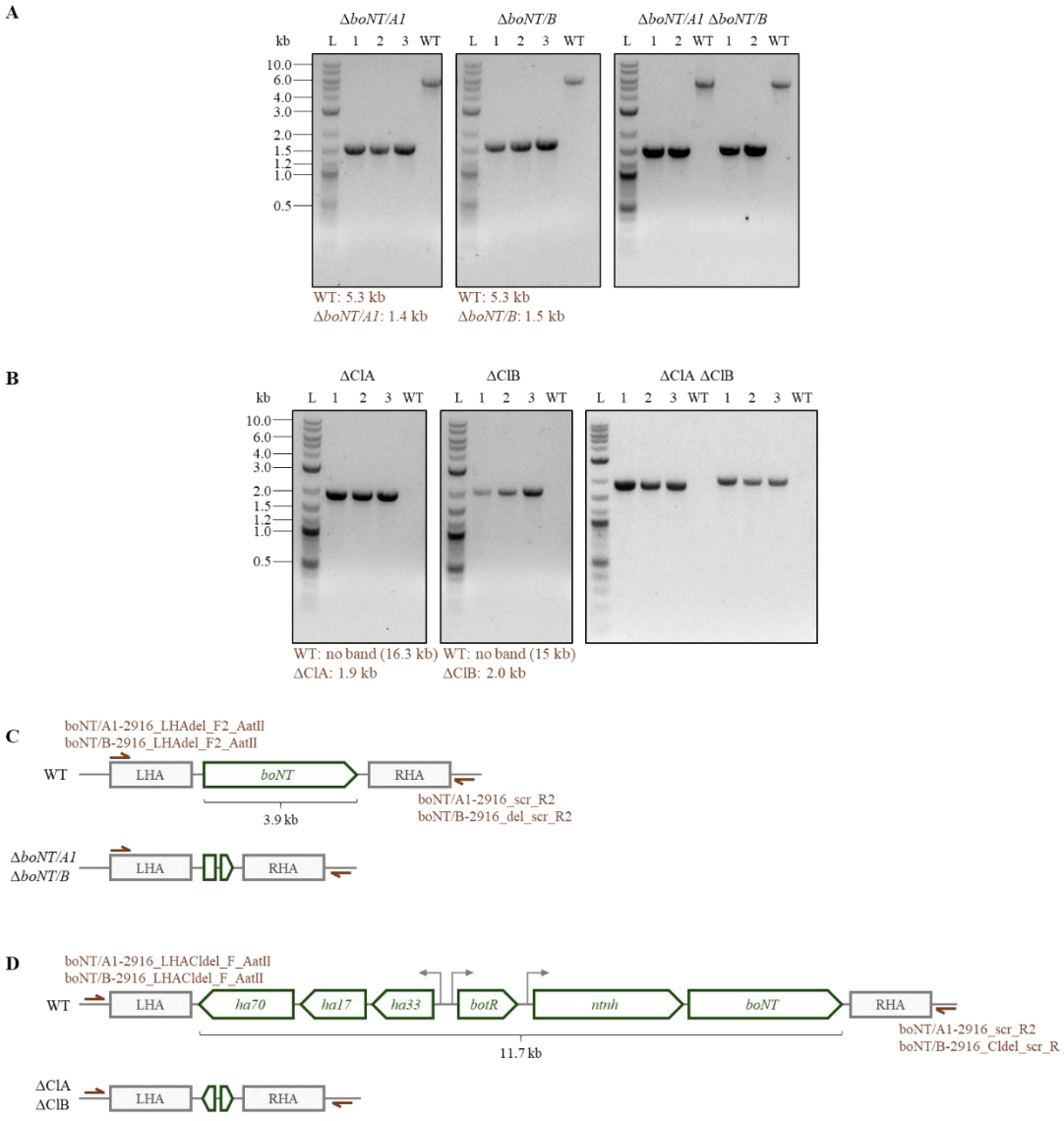


Figure 5.12 Deletion of *boNT/A1*, *boNT/B*, *boNT/A1* gene cluster (ClA) and *boNT/B* gene cluster (ClA) in strain NCTC 2916. (**A**) PCR for three biological replicates of NCTC 2916 Δ*boNT/A1*, Δ*boNT/B* and two biological replicates for the double mutants alongside WT. Products were amplified using primers boNT/A1-2916_LHAdel_F2_AatII and boNT/A1-2916_scr_R2 (Δ*boNT/A1*) and boNT/B-2916_LHAdel_F2_AatII and boNT/B-2916_del_scr_R2 (Δ*boNT/B*). Lane L: NEB 1 kb Plus DNA Ladder. (**B**) PCR for three biological replicates of NCTC 2916 Δ*boNT/A1*, Δ*boNT/B* and the double mutants alongside WT. Products were amplified using primers boNT/A1-2916_LHAClde1_F2_AatII and boNT/A1-2916_scr_R2 (ΔClA) and boNT/B-2916_LHAClde1_F2_AatII and boNT/B-2916_Cldel_scr_R (ΔClB). Lane L: NEB 1 kb Plus DNA Ladder. (**C**/**D**) Schematic diagram of the targeted genomic region showing the primer binding sites in the WT and mutant strains.

To confirm that the basic phenotypic characteristics of the parent strain were not affected by the deletions, the growth and sporulation/germination capacity of the mutants was assessed as described in sections 2.9.2 and 2.9.4, respectively. The growth profiles and the sporulation/germination capacities of all the mutant strains were essentially identical to WT (Figure 5.13).

Although the generation and initial characterisation of the Δ*boNT/A1* Δ*boNT/B* and ΔClA ΔClB strains are pivotal first steps towards downregulation of these strains, several additional experiments will be required to apply for an official downgraded status. Future work should include whole genome sequencing, ELISA and/or Western blot analysis and a mouse bioassay, as previously discussed for the ATCC 19397 safe strains. Additionally, depending on the application of interest and the number of additional mutations detected, further phenotypic characterisation might be necessary; for example, to confirm that spore heat resistance, salt tolerance and growth temperatures have not been compromised.

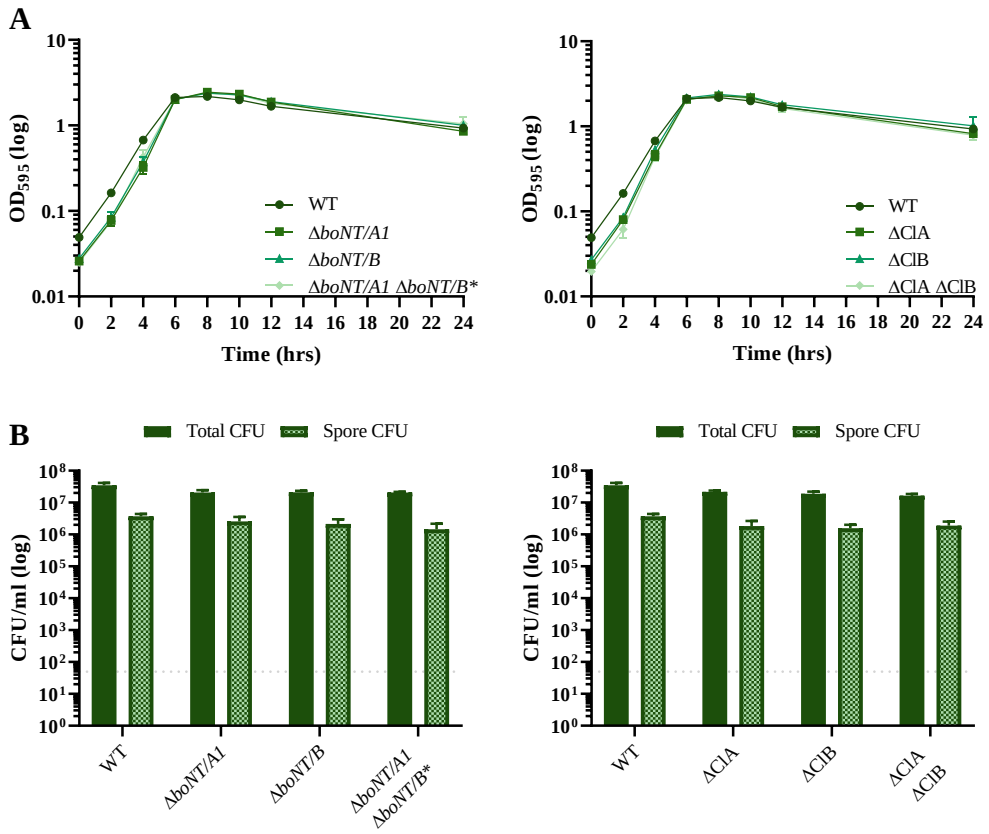


Figure 5.13 Phenotypic characterisation of NCTC 2916 Δ*boNT/A1*, Δ*boNT/B*, ΔClA, ΔClB and double mutants. (**A**) Growth profiles of Δ*boNT/A1*, Δ*boNT/B*, ΔClA, ΔClB, double mutants and WT. Strains were cultured in TYG broth and the OD_{595} was monitored for 24 hours. Error bars represent the standard deviation of three independent cultures for WT and of three biological replicates for the mutants. (**B**) Sporulation/germination capacity of Δ*boNT/A1*, Δ*boNT/B*, ΔClA, ΔClB and double mutants compared to WT. Strains were cultured in TYG broth and samples were taken after 120 hours. Total CFU corresponds to the total cell count (non-heat treated) and the spore CFU is the number of viable spores withstanding a 15 min, 80°C heat treatment. Error bars represent the standard deviation of three independent cultures for WT and of three biological replicates for the mutants. The dotted line at 50 CFU/ml represents the detection limit. * indicates only two biological replicates.

5.4 Key outcomes

- The ATCC 3502^{+}, ATCC 19397, 62A, Loch Maree and NCTC 2916 populations assessed in this study are all genetically stable and thus suitable for a wide range of research applications.

- The CFSAN064329 62A mutator population is genetically significantly different from the NCTC 3805 62A non-mutator population and therefore they should be considered individual strains. Preferably, future strain names include a tag related to the functionality of their *mut* proteins, similar as proposed for strain ATCC 3502 (i.e., $62A^{-}$ and $62A^{+}$, respectively).

- The *spoVA*^{2} operon is a major contributor to the high heat resistance of NCTC 2916 spores.

- Stable non-toxigenic ATCC 19397 and NCTC 2916 strains were generated, which were essentially indistinguishable from WT for the relevant phenotypic characteristics examined.

- Successful implementation of the theophylline inducible riboswitch-Cas9-based editing tool to engineer type A *C. botulinum* strains. Moreover, combining RiboCas with customised *E. coli* donor strains was demonstrated to be a powerful strategy to efficiently modify chromosomal DNA in *C. botulinum*.

CHAPTER 6

General Discussion

6.1 Summary of findings

The work presented in this thesis revolves around the aim of developing a universally applicable strain engineering strategy to aid the creation of safe *C. botulinum* strains. To successfully generate non-toxigenic mutants, foreign DNA uptake by the host, genetic stability and efficient DNA manipulation tools are essential.

Building on previously published findings to establish and improve genetic accessibility of a host, two approaches were attempted: (i) restriction system inactivation, and; (ii) methylation mimicry.[315-316, 338] Inactivation of the type I REase in *C. botulinum* ATCC 3502 proved to be an effective way to improve the conjugation efficiency of plasmids harbouring the type I recognition site. However, inactivation of a host's restriction systems is only possible when there is an initial (low) level of exogenous DNA uptake and therefore cannot be used for strains that are completely inaccessible, such as ATCC 19397.

Methylation mimicry does not require engineering of the hosts and can thus theoretically be used to circumvent the restriction barriers of any strain. Expression of the MTases associated to the host's restriction systems in the donor strain significantly increased the efficiency of conjugative plasmid transfer into the *C. botulinum* type A strains ATCC 3502 and ATCC 19397. Additionally, utilisation of the *dcm*- conjugal donor strain sExpress improved conjugation efficiencies for strains 62A, Loch Maree and NCTC 2916.

To fulfil the requirement that all phenotypic characteristics of the non-toxigenic strains remain identical to WT, except toxicity, a stable genetic background is essential. Previous efforts to create non-toxigenic ATCC 3502 strains in Nottingham failed due to unexpected phenotypic changes, such as the loss of heat resistant spore formation,

caused by the accumulation of a high number of random mutations throughout the mutants' genomes. In an attempt to circumvent the apparently unstable nature of ATCC 3502 (IFR), a different ATCC 3502 stock, acquired via Unilever, was used to create genetically modified strains. However, contrary to what was anticipated, whole genome sequencing revealed that a majority of the mutants had acquired a high number of random secondary mutations, independent of the genetic tool used.

To unveil the reason for this unusual genetic instability, potential mutator genes were identified in ATCC 3502 and its full genome was compared to the genome of the highly similar non-mutator strain ATCC 19397. An in-frame deletion in the *mutL* gene, encoding a DNA mismatch repair protein, was identified and further experimental characterisation of this mutation confirmed it is responsible for the hypermutator phenotype of ATCC 3502. Furthermore, analysis of the various ATCC 3502 stocks led to the discovery that actually two distinct *C. botulinum* ATCC 3502 subpopulations exist; hypermutators with the Δ*mutL*231 genotype and normomutators with a full *mutL* gene. Interestingly, the ATCC 3502 Δ*mutL*231 genotype has a relative fitness advantage over the non-mutator ATCC 3502 genotype, despite its increased load of deleterious mutations. Although it is not uncommon for mutators to reach fixation in a population, they do so by hitchhiking along with beneficial mutations, however, such a mutation could not be identified in the ATCC 3502 Δ*mutL*231 genotype.

Additionally, analysis of the original ATCC 3502 stocks received by the SBRC from Quadram Institute (IFR), Unilever and NCTC confirmed that the mutator genotype was already part of the population before the strain came to Nottingham. Moreover, *C. botulinum* type A strains Hall and CFSAN064329 (62A) also carry mutator alleles. Plausibly, the *C. botulinum* type A strains are poorly adapted to the laboratory

environment and the mutator phenotype confers an advantage in the adapting population.

In parallel to the investigation of ATCC 3502's genomic instability, the type A strains ATCC 19397, 62A, Loch Maree and NCTC 2916 were characterised with the aim of finding more suitable strains for research purposes. All strains have relatively low mutation rates and sporulate efficiently. Additionally, the *spoVA*^{2} operon significantly contributes to the formation of highly heat resistant spores. This knowledge could be used to identify potentially interesting strains in the future to include in the safe strain library for food challenge studies.

Finally, various stable non-toxicogenic ATCC 19397 and NCTC 2916 strains were generated, which were essentially indistinguishable from WT for the relevant phenotypic characteristics examined. Overall, this thesis demonstrates that solid basic strain characterisation combined with customised *E. coli* donor strains and the RiboCas genome editing tool constitutes a powerful strategy to efficiently generate *C. botulinum* mutants.

6.2 Future perspectives

The limited number of genetic tools available to engineer *C. botulinum*, genetic inaccessibility and mostly the stringent select-agent regulations have hampered *C. botulinum* research progress. Replacing the toxigenic strains for non-toxic variants would be a solution to the latter. To do so, however, efficient DNA transfer and effective tools are required. This thesis sought to break this vicious circle to aid future research.

Reliable and efficient DNA transfer is a prerequisite in order to perform site-directed mutagenesis for the creation of safe strains. Unfortunately, a large number of strains

are genetically inaccessible due to the various defence mechanisms they possess that target incoming foreign DNA.[308] Often, this is the first and also the most challenging hurdle to overcome. Here, optimisation of the conjugal donor strain proved to be a very effective method to circumvent host's restriction barriers. To create such 'personalised' conjugal donors for a variety of hosts efficiently in the future, the following four-step customisation pipeline has been proposed: (i) whole genome sequencing of the recipient (using PacBio SMRT); (ii) identification of its restriction systems (using REBASE); (iii) selection of an appropriate conjugal donor strain (Dam/Dcm methylation profile), and; (iv) expression of relevant MTases in the donor.

The work presented in this thesis, however, only included a few type A strains with similar restriction systems. The effectiveness of the employed strategy across the highly diverse *C. botulinum* family, harbouring a wide variety of restriction systems, is yet to be exemplified.[5, 16, 42, 308] In particular, although the expression of the specific MTases used in this study did not negatively impact the *E. coli* donor strain, the expression of non-native MTases can be toxic to the donor cells. Placing the MTase encoding genes downstream of an inducible promoter could mitigate toxicity effects, depending on the level of toxicity. Nevertheless, based on the available literature, the proposed four-step based approach will most likely be effective for a majority of strains from the different *C. botulinum* groups.[327-329, 338]

Furthermore, with the recent development of the RiboCas system and other, less efficient, Cas9-based tools to manipulate DNA *in vivo* in various *Clostridium* spp., generation of the non-toxigenic strains should be relatively straightforward as soon as DNA transfer has been established.[158, 177, 227] Now it is 'just' a matter of finding the right strains for the intended purpose.

Although the non-toxigenic strains created in this study could be used in food challenge studies, the absence of toxin makes it difficult to screen for potential growth. At present, *C. botulinum* growth in food matrices is detected by screening for the presence of neurotoxin.[620-621] Using a strain that produces a toxoid rather than a toxin minus strain could be a solution. The methods available for the detection of catalytically inactive toxin, such as ELISA, are however, generally less sensitive and reliable than the methods used to detect catalytically active neurotoxin, such as the mouse bioassay. To overcome the detection challenge, safe strains could be equipped with a reporter. This concept has been explored by the Nottingham iGEM team; as a proof of concept genes required for the expression of the volatile acetone were placed downstream of the *ntnh* promoter (which is controlled by alternative sigma factor BotR) and expressed in *C. sporogenes* together with *botR* and its native promoter.[622] Future work could include the characterisation of different reporter systems in *C. botulinum* and their subsequent evaluation in various food matrices. Such reporter systems have great potential, also beyond food challenge studies; for example, to study toxin regulation using alternative protein-based reporters (e.g., GusA and FAST).[623]

For other industrial applications, in particular for the pharmaceutical industry, the toxin negative strains presented in this thesis could not only be used as a negative control, but can also serve as a production platform for the expression of recombinant toxins. Ideally, in these strains sporulation is abolished and new coding sequences can be readily inserted into the chromosome. The latter could be achieved by replacing the toxin gene (or full gene cluster) with a bookmark sequence, which contains a CRISPR/Cas9 target site. At the bookmark position genetic material encoding BoNT derivatives can be inserted easily and efficiently.[618] Furthermore, as a safety feature, these recombinant neurotoxins could be placed under the control of a tight inducible

system with a wide dynamic range, such as the RiboLac, specifically developed for *Clostridia* spp.[221] Toxin production will be switched OFF during, for example, the strain generation process or transport and will only be switched ON when toxin production is desired.

Although few research group study *C. botulinum*, over the years a number of different researchers have worked with this organism extensively. It is, therefore, quite surprising that this is the first study revealing the rapid degeneration of strains of this organism. Evidently, these findings might have a significant impact on our current understanding of *C. botulinum*, affecting both the industrial and scientific communities. To highlight this, it is possible that previously published studies have employed degenerated strains whose associated phenotypes may differ compared to those of naturally occurring WT strains. In particular results obtained from studies where the experimental set-up required strong bottlenecks, such as single colony stages to isolate mutants. These population 'representatives' might be more different from the original WT than initially meets the eye. Therefore, previously published results with this organism should be henceforth considered with caution.

Worse yet, there is a risk that botulinum toxin production strains might be (hyper)mutators, potentially impacting unpredictably the quality and safety of products destined for human administration. With the neurotoxin not being essential for cell survival in laboratory conditions it is not inconceivable that occasionally mutations occur in the neurotoxin protein, changing its catalytic properties. Extensive quality control should identify batches with altered activity levels and ensure product safety. Nonetheless, producing neurotoxin with a mutator strain will affect the products homogeneity and potentially its efficacy. Especially because of the very low dosage required and the inherent increased risk of small product variations.

In addition, these findings emphasize the importance of storing strains appropriately, to prevent the emergence and establishment of such phenotypes within a given cellular population. One should always keep in the back of their mind that within a bacterial population there might be significant differences between individual cells that are continuously evolving. Inherently, a WT genome is merely the sum of all the most common alleles of each locus at a given time point, hence defining a 'new' WT genome for strain ATCC 3502 should be done carefully. A population solely consisting of non-mutators can be obtained by selecting a single colony, however, this raises the question, which colony do we decide to be our future WT? In this thesis, a 'new' non-mutator WT genome for strain ATCC 3502 is proposed, using the highly similar strains ATCC 19397 and 62A as a reference whenever the nucleotide call was not clear-cut. Furthermore, the ATCC 3502^{-} and CFSAN064329 62A mutator populations are genetically significantly different from the non-mutator populations and should therefore be considered individual strains. Preferably, future strain names include a tag related to the functionality of their *mut* proteins (i.e., strain^{-} and strain^{+}) to prevent future confusion. Moreover, all institutions that hold any of the potential mutator strains should check their stocks and if the strain happens to be the mutator variant, it would be advisable to reflect on how its genetic instability might have affected past research and if it has any implications for product safety and reliability.

Although the existence of these mutators is mostly 'inconvenient' it is also a very interesting research opportunity. These findings prompt one to wonder whether *C. botulinum* is generally prone to these genetic changes and, primarily, what is the reason behind such a propensity. It is still unclear whether the mutator allele hitchhiked to fixation or if it confers a fitness advantage on its own. In addition, this study only focused on type A strains, but leaves one wondering if this is just common

among Group I strains or if mutator genotypes also frequently occur in the other *C. botulinum* groups.

6.3 Concluding remarks

It is evident that there are plenty of exciting opportunities that could be pursued in the future, which would benefit from - easy to work with - safe strains. The mutants presented here provide a safe alternative for fundamental research studies as well as for food challenge tests. However, to study these organisms in a less restricted laboratory environment the regulatory authorities need to officially recognise that these attenuated strains no longer pose a severe threat to public health and safety. To ultimately prove their safety mouse bioassays will have to be performed in the future. Although the assays to confirm the strains' lack of toxicity are yet to be performed, based on current literature there is no reason to assume that these experiments will prove otherwise. Furthermore, several BoNT toxoids and the amino acids that encode them are already excluded from the requirements of the select agent regulations by the Federal Select Agent Program of the CDC and US Department of Agriculture (USDA; Select Agents and Toxins Exclusions: Nontoxic HHS toxins; section 73.3 (d)(2)). It is therefore realistic to anticipate that non-toxigenic *C. botulinum* strains, such as created in this study, will indeed get downgraded to a Group 1 (or equivalent) organism by the relevant regulatory authorities, and replace the toxigenic ancestors in future research.

Bibliography

1. Erbguth, F. J., Historical notes on botulism, *Clostridium botulinum*, botulinum toxin, and the idea of the therapeutic use of the toxin. *Mov Disord* **2004,** *19 Suppl 8*, S2-6.
2. van Ermengem, E., Classics in infectious diseases. A new anaerobic *bacillus* and its relation to botulism. E. van Ermengem. Originally published as "Ueber einen neuen anaeroben *Bacillus* und seine Beziehungen zum Botulismus" in Zeitschrift fur Hygiene und Infektionskrankheiten 26: 1-56, 1897. *Rev Infect Dis* **1979,** *1* (4), 701-19.
3. Burke, G. S., The Occurrence of *Bacillus botulinu*s in Nature. *Journal of bacteriology* **1919,** *4* (5), 541-553.
4. Burke, G. S., Notes on *Bacillus botulinus*. *J Bacteriol* **1919,** *4* (5), 555-570 1.
5. Peck, M. W.; Smith, T. J.; Anniballi, F.; Austin, J. W.; Bano, L.; Bradshaw, M.; Cuervo, P.; Cheng, L. W.; Derman, Y.; Dorner, B. G.; Fisher, A.; Hill, K. K.; Kalb, S. R.; Korkeala, H.; Lindstrom, M.; Lista, F.; Luquez, C.; Mazuet, C.; Pirazzini, M.; Popoff, M. R.; Rossetto, O.; Rummel, A.; Sesardic, D.; Singh, B. R.; Stringer, S. C., Historical Perspectives and Guidelines for Botulinum Neurotoxin Subtype Nomenclature. *Toxins (Basel)* **2017,** *9* (1).
6. Poulain, B.; Popoff, M. R., Why Are Botulinum Neurotoxin-Producing Bacteria So Diverse and Botulinum Neurotoxins So Toxic? *Toxins (Basel)* **2019,** *11* (1).
7. Williamson, C. H.; Sahl, J. W.; Smith, T. J.; Xie, G.; Foley, B. T.; Smith, L. A.; Fernandez, R. A.; Lindstrom, M.; Korkeala, H.; Keim, P.; Foster, J.; Hill, K., Comparative genomic analyses reveal broad diversity in botulinum-toxin-producing Clostridia. *BMC Genomics* **2016,** *17*, 180.
8. Collins, M. D.; East, A. K., Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *Journal of applied microbiology* **1998,** *84* (1), 5-17.
9. Lindstrom, M.; Korkeala, H., Laboratory diagnostics of botulism. *Clin Microbiol Rev* **2006,** *19* (2), 298-314.
10. Carter, A. T.; Peck, M. W., Genomes, neurotoxins and biology of *Clostridium botulinum* Group I and Group II. *Res Microbiol* **2015,** *166* (4), 303-17.
11. Sakaguchi, Y.; Suzuki, T.; Yamamoto, Y.; Nishikawa, A.; Oguma, K., Genomics of *Clostridium botulinum* group III strains. *Res Microbiol* **2015,** *166* (4), 318-25.
12. Popoff, M. R., Botulinum Neurotoxins: Still a Privilege of Clostridia? *Cell Host Microbe* **2018,** *23* (2), 145-146.
13. Frye, E. A.; Egan, C.; Perry, M. J.; Crouch, E. E.; Burbank, K. E.; Kelly, K. M., Outbreak of botulism type A in dairy cows detected by MALDI-TOF mass spectrometry. *J Vet Diagn Invest* **2020,** *32* (5), 722-726.
14. Peck, M. W., Biology and genomic analysis of *Clostridium botulinum*. *Adv Microb Physiol* **2009,** *55*, 183-265, 320.

15. Stringer, S. C.; Webb, M. D.; Peck, M. W., Contrasting effects of heat treatment and incubation temperature on germination and outgrowth of individual spores of nonproteolytic *Clostridium botulinum* bacteria. *Appl Environ Microbiol* **2009,** *75* (9), 2712-9.

16. Smith, T.; Williamson, C. H. D.; Hill, K.; Sahl, J.; Keim, P., Botulinum Neurotoxin-Producing Bacteria. Isn't It Time that We Called a Species a Species? *mBio* **2018,** *9* (5).

17. Hill, K. K.; Smith, T. J., Genetic diversity within *Clostridium botulinum* serotypes, botulinum neurotoxin gene clusters and toxin subtypes. *Curr Top Microbiol Immunol* **2013,** *364*, 1-20.

18. Kalb, S. R.; Baudys, J.; Webb, R. P.; Wright, P.; Smith, T. J.; Smith, L. A.; Fernandez, R.; Raphael, B. H.; Maslanka, S. E.; Pirkle, J. L.; Barr, J. R., Discovery of a novel enzymatic cleavage site for botulinum neurotoxin F5. *FEBS Lett* **2012,** *586* (2), 109-15.

19. Hill, K. K.; Smith, T. J.; Helma, C. H.; Ticknor, L. O.; Foley, B. T.; Svensson, R. T.; Brown, J. L.; Johnson, E. A.; Smith, L. A.; Okinaka, R. T.; Jackson, P. J.; Marks, J. D., Genetic diversity among Botulinum Neurotoxin-producing clostridial strains. *J Bacteriol* **2007,** *189* (3), 818-32.

20. Smith, T. J.; Lou, J.; Geren, I. N.; Forsyth, C. M.; Tsai, R.; Laporte, S. L.; Tepp, W. H.; Bradshaw, M.; Johnson, E. A.; Smith, L. A.; Marks, J. D., Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization. *Infect Immun* **2005,** *73* (9), 5450-7.

21. Whitemarsh, R. C.; Tepp, W. H.; Bradshaw, M.; Lin, G.; Pier, C. L.; Scherf, J. M.; Johnson, E. A.; Pellett, S., Characterization of botulinum neurotoxin A subtypes 1 through 5 by investigation of activities in mice, in neuronal cell cultures, and in vitro. *Infect Immun* **2013,** *81* (10), 3894-902.

22. Pier, C. L.; Chen, C.; Tepp, W. H.; Lin, G.; Janda, K. D.; Barbieri, J. T.; Pellett, S.; Johnson, E. A., Botulinum neurotoxin subtype A2 enters neuronal cells faster than subtype A1. *FEBS Lett* **2011,** *585* (1), 199-206.

23. Kull, S.; Schulz, K. M.; Weisemann, J.; Kirchner, S.; Schreiber, T.; Bollenbach, A.; Dabrowski, P. W.; Nitsche, A.; Kalb, S. R.; Dorner, M. B.; Barr, J. R.; Rummel, A.; Dorner, B. G., Isolation and functional characterization of the novel *Clostridium botulinum* neurotoxin A8 subtype. *PLoS One* **2015,** *10* (2), e0116381.

24. Kozaki, S.; Kamata, Y.; Nishiki, T.; Kakinuma, H.; Maruyama, H.; Takahashi, H.; Karasawa, T.; Yamakawa, K.; Nakamura, S., Characterization of *Clostridium botulinum* type B neurotoxin associated with infant botulism in japan. *Infect Immun* **1998,** *66* (10), 4811-6.

25. Strotmeier, J.; Gu, S.; Jutzi, S.; Mahrhold, S.; Zhou, J.; Pich, A.; Eichner, T.; Bigalke, H.; Rummel, A.; Jin, R.; Binz, T., The biological activity of botulinum neurotoxin type C is dependent upon novel types of ganglioside binding sites. *Mol Microbiol* **2011,** *81* (1), 143-56.

26. Peng, L.; Tepp, W. H.; Johnson, E. A.; Dong, M., Botulinum neurotoxin D uses synaptic vesicle protein SV2 and gangliosides as receptors. *PLoS Pathog* **2011,** *7* (3), e1002008.

27. Smith, T. J.; Williamson, C. H. D.; Hill, K. K.; Johnson, S. L.; Xie, G.; Anniballi, F.; Auricchio, B.; Fernandez, R. A.; Caballero, P. A.; Keim, P.; Sahl, J. W., The Distinctive Evolution of orfX *Clostridium parabotulinum* Strains and Their Botulinum Neurotoxin Type A and F Gene Clusters Is Influenced by Environmental Factors and Gene Interactions via Mobile Genetic Elements. *Front Microbiol* **2021,** *12*, 566908.

28. Jahn, R., Neuroscience. A neuronal receptor for botulinum toxin. *Science* **2006,** *312* (5773), 540-1.

29. Gu, S.; Rumpel, S.; Zhou, J.; Strotmeier, J.; Bigalke, H.; Perry, K.; Shoemaker, C. B.; Rummel, A.; Jin, R., Botulinum neurotoxin is shielded by NTNHA in an interlocked complex. *Science* **2012,** *335* (6071), 977-81.

30. Matsumura, T.; Jin, Y.; Kabumoto, Y.; Takegahara, Y.; Oguma, K.; Lencer, W. I.; Fujinaga, Y., The HA proteins of botulinum toxin disrupt intestinal epithelial intercellular junctions to increase toxin absorption. *Cell Microbiol* **2008,** *10* (2), 355-64.

31. Lee, K.; Zhong, X.; Gu, S.; Kruel, A. M.; Dorner, M. B.; Perry, K.; Rummel, A.; Dong, M.; Jin, R., Molecular basis for disruption of E-cadherin adhesion by botulinum neurotoxin A complex. *Science* **2014,** *344* (6190), 1405-10.

32. Fujinaga, Y.; Inoue, K.; Watarai, S.; Sakaguchi, Y.; Arimitsu, H.; Lee, J. C.; Jin, Y.; Matsumura, T.; Kabumoto, Y.; Watanabe, T.; Ohyama, T.; Nishikawa, A.; Oguma, K., Molecular characterization of binding subcomponents of *Clostridium botulinum* type C progenitor toxin for intestinal epithelial cells and erythrocytes. *Microbiology (Reading)* **2004,** *150* (Pt 5), 1529-1538.

33. Mazuet, C.; Legeay, C.; Sautereau, J.; Ma, L.; Bouchier, C.; Bouvet, P.; Popoff, M. R., Diversity of Group I and II *Clostridium botulinum* Strains from France Including Recently Identified Subtypes. *Genome Biol Evol* **2016,** *8* (6), 1643-60.

34. Connan, C.; Deneve, C.; Mazuet, C.; Popoff, M. R., Regulation of toxin synthesis in *Clostridium botulinum* and *Clostridium tetani*. *Toxicon* **2013,** *75*, 90-100.

35. Raffestin, S.; Dupuy, B.; Marvaud, J. C.; Popoff, M. R., BotR/A and TetR are alternative RNA polymerase sigma factors controlling the expression of the neurotoxin and associated protein genes in *Clostridium botulinum* type A and *Clostridium tetani*. *Mol Microbiol* **2005,** *55* (1), 235-49.

36. Marvaud, J. C.; Gibert, M.; Inoue, K.; Fujinaga, Y.; Oguma, K.; Popoff, M. R., botR/A is a positive regulator of botulinum neurotoxin and associated non-toxin protein genes in *Clostridium botulinum* A. *Mol Microbiol* **1998,** *29* (4), 1009-18.

37. Kalb, S. R.; Baudys, J.; Smith, T. J.; Smith, L. A.; Barr, J. R., Characterization of Hemagglutinin Negative Botulinum Progenitor Toxins. *Toxins (Basel)* **2017,** *9* (6).

38. Couesnon, A.; Raffestin, S.; Popoff, M. R., Expression of botulinum neurotoxins A and E, and associated non-toxin genes, during the transition phase and stability at high temperature: analysis by quantitative reverse transcription-PCR. *Microbiology (Reading)* **2006,** *152* (Pt 3), 759-770.

39. Hutson, R. A.; Zhou, Y.; Collins, M. D.; Johnson, E. A.; Hatheway, C. L.; Sugiyama, H., Genetic characterization of *Clostridium botulinum* type A containing silent type B neurotoxin gene sequences. *J Biol Chem* **1996,** *271* (18), 10786-92.

40. Dover, N.; Barash, J. R.; Hill, K. K.; Davenport, K. W.; Teshima, H.; Xie, G.; Arnon, S. S., *Clostridium botulinum* strain Af84 contains three neurotoxin gene clusters: *bont/A2*, *bont/F4* and *bont/F5*. *PLoS One* **2013,** *8* (4), e61205.

41. Popoff, M. R.; Bouvet, P., Genetic characteristics of toxigenic Clostridia and toxin gene evolution. *Toxicon* **2013,** *75*, 63-89.

42. Hill, K. K.; Xie, G.; Foley, B. T.; Smith, T. J., Genetic diversity within the botulinum neurotoxin-producing bacteria and their neurotoxins. *Toxicon* **2015,** *107* (Pt A), 2-8.

43. Raffestin, S.; Marvaud, J. C.; Cerrato, R.; Dupuy, B.; Popoff, M. R., Organization and regulation of the neurotoxin genes in *Clostridium botulinum* and *Clostridium tetani*. *Anaerobe* **2004,** *10* (2), 93-100.

44. Nawrocki, E. M.; Bradshaw, M.; Johnson, E. A., Botulinum neurotoxin-encoding plasmids can be conjugatively transferred to diverse clostridial strains. *Sci Rep* **2018,** *8* (1), 3100.

45. Skarin, H.; Segerman, B., Horizontal gene transfer of toxin genes in *Clostridium botulinum*: Involvement of mobile elements and plasmids. *Mob Genet Elements* **2011,** *1* (3), 213-215.

46. Marshall, K. M.; Bradshaw, M.; Johnson, E. A., Conjugative botulinum neurotoxin-encoding plasmids in *Clostridium botulinum*. *PLoS One* **2010,** *5* (6), e11087.

47. Hill, K. K.; Xie, G.; Foley, B. T.; Smith, T. J.; Munk, A. C.; Bruce, D.; Smith, L. A.; Brettin, T. S.; Detter, J. C., Recombination and insertion events involving the botulinum neurotoxin complex genes in *Clostridium botulinum* types A, B, E and F and *Clostridium butyricum* type E strains. *BMC Biol* **2009,** *7*, 66.

48. Smith, T. J.; Hill, K. K.; Xie, G.; Foley, B. T.; Williamson, C. H. D.; Foster, J. T.; Johnson, S. L.; Chertkov, O.; Teshima, H.; Gibbons, H. S.; Johnsky, L. A.; Karavis, M. A.; Smith, L. A., Genomic sequences of six botulinum neurotoxin-producing strains representing three clostridial species illustrate the mobility and diversity of botulinum neurotoxin genes. *Infect Genet Evol* **2015,** *30*, 102-113.

49. Smith, T. J.; Hill, K. K.; Foley, B. T.; Detter, J. C.; Munk, A. C.; Bruce, D. C.; Doggett, N. A.; Smith, L. A.; Marks, J. D.; Xie, G.; Brettin, T. S., Analysis of the neurotoxin complex genes in *Clostridium botulinum* A1-A4 and B1 strains: BoNT/A3, /Ba4 and /B1 clusters are located within plasmids. *PLoS One* **2007,** *2* (12), e1271.

50. Carter, A. T.; Austin, J. W.; Weedmark, K. A.; Peck, M. W., Evolution of Chromosomal *Clostridium botulinum* Type E Neurotoxin Gene Clusters: Evidence Provided by Their Rare Plasmid-Borne Counterparts. *Genome Biol Evol* **2016,** *8* (3), 540-55.

51. Fischer, A.; Montal, M., Crucial role of the disulfide bridge between botulinum neurotoxin light and heavy chains in protease translocation across membranes. *J Biol Chem* **2007,** *282* (40), 29604-11.

52. Montecucco, C.; Schiavo, G., Structure and function of tetanus and botulinum neurotoxins. *Q Rev Biophys* **1995,** *28* (4), 423-72.

53. de Paiva, A.; Poulain, B.; Lawrence, G. W.; Shone, C. C.; Tauc, L.; Dolly, J. O., A role for the interchain disulfide or its participating thiols in the internalization of botulinum neurotoxin A revealed by a toxin derivative that binds to ecto-acceptors and inhibits transmitter release intracellularly. *J Biol Chem* **1993,** *268* (28), 20838-44.

54. Das Gupta, B. R.; Sugiyama, H., Role of a protease in natural activation of *Clostridium botulinum* neurotoxin. *Infect Immun* **1972,** *6* (4), 587-90.

55. Simpson, L., The life history of a botulinum toxin molecule. *Toxicon* **2013,** *68*, 40-59.

56. Fujinaga, Y.; Popoff, M. R., Translocation and dissemination of botulinum neurotoxin from the intestinal tract. *Toxicon* **2018,** *147*, 13-18.

57. Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; Polverino de Laureto, P.; DasGupta, B. R.; Montecucco, C., Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **1992,** *359* (6398), 832-5.

58. Binz, T.; Rummel, A., Cell entry strategy of clostridial neurotoxins. *J Neurochem* **2009,** *109* (6), 1584-95.

59. Surana, S.; Tosolini, A. P.; Meyer, I. F. G.; Fellows, A. D.; Novoselov, S. S.; Schiavo, G., The travel diaries of tetanus and botulinum neurotoxins. *Toxicon* **2018,** *147*, 58-67.

60. Cai, S.; Kukreja, R.; Shoesmith, S.; Chang, T. W.; Singh, B. R., Botulinum neurotoxin light chain refolds at endosomal pH for its translocation. *Protein J* **2006,** *25* (7-8), 455-62.

61. Koriazova, L. K.; Montal, M., Translocation of botulinum neurotoxin light chain protease through the heavy chain channel. *Nat Struct Biol* **2003,** *10* (1), 13-8.

62. Rowland, L. P., Stroke, spasticity, and botulinum toxin. *N Engl J Med* **2002,** *347* (6), 382-3.

63. Schiavo, G.; Matteoli, M.; Montecucco, C., Neurotoxins affecting neuroexocytosis. *Physiol Rev* **2000,** *80* (2), 717-66.

64. Pantano, S.; Montecucco, C., The blockade of the neurotransmitter release apparatus by botulinum neurotoxins. *Cell Mol Life Sci* **2014,** *71* (5), 793-811.

65. Montecucco, C.; Schiavo, G.; Pantano, S., SNARE complexes and neuroexocytosis: how many, how close? *Trends Biochem Sci* **2005,** *30* (7), 367-72.

66. Tsai, Y. C.; Kotiya, A.; Kiris, E.; Yang, M.; Bavari, S.; Tessarollo, L.; Oyler, G. A.; Weissman, A. M., Deubiquitinating enzyme VCIP135 dictates the duration of botulinum neurotoxin type A intoxication. *Proc Natl Acad Sci U S A* **2017,** *114* (26), E5158-E5166.

67. Pellett, S.; Bradshaw, M.; Tepp, W. H.; Pier, C. L.; Whitemarsh, R. C. M.; Chen, C.; Barbieri, J. T.; Johnson, E. A., The Light Chain Defines the Duration of Action of Botulinum Toxin Serotype A Subtypes. *mBio* **2018,** *9* (2).

68. Wongtanate, M.; Sucharitchan, N.; Tantisiriwit, K.; Oranrigsupak, P.; Chuesuwan, A.; Toykeaw, S.; Suputtamongkol, Y., Signs and symptoms predictive of respiratory failure in patients with foodborne botulism in Thailand. *Am J Trop Med Hyg* **2007,** *77* (2), 386-9.

69. Bossi, P.; Tegnell, A.; Baka, A.; Werner, A.; van Loock, F.; Hendriks, J.; Maidhof, H.; Gouvras, G., Bichat guidelines for the clinical management of botulism and bioterrorism-related botulism. *Euro Surveill* **2004,** *9* (12), 31-32.

70. Gill, D. M., Bacterial toxins: a table of lethal amounts. *Microbiol Rev* **1982,** *46* (1), 86-94.

71. Arnon, S. S.; Schechter, R.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Hauer, J.; Layton, M.; Lillibridge, S.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Swerdlow, D. L.; Tonat, K.; Working Group on Civilian, B., Botulinum toxin as a biological weapon: medical and public health management. *JAMA* **2001,** *285* (8), 1059-70.

72. Peck, M. W., *Clostridium botulinum* and the safety of minimally heated, chilled foods: an emerging issue? *Journal of applied microbiology* **2006,** *101* (3), 556-70.

73. Johnson, E. A., Clostridial toxins as therapeutic agents: benefits of nature's most toxic proteins. *Annu Rev Microbiol* **1999,** *53*, 551-75.

74. Schantz, E. J.; Johnson, E. A., Properties and use of botulinum toxin and other microbial neurotoxins in medicine. *Microbiol Rev* **1992,** *56* (1), 80-99.

75. Kirk, M. D.; Pires, S. M.; Black, R. E.; Caipo, M.; Crump, J. A.; Devleesschauwer, B.; Dopfer, D.; Fazil, A.; Fischer-Walker, C. L.; Hald, T.; Hall, A. J.; Keddy, K. H.; Lake, R. J.; Lanata, C. F.; Torgerson, P. R.; Havelaar, A. H.; Angulo, F. J., World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLoS Med* **2015,** *12* (12), e1001921.

76. Luquez, C.; Bianco, M. I.; de Jong, L. I.; Sagua, M. D.; Arenas, G. N.; Ciccarelli, A. S.; Fernandez, R. A., Distribution of botulinum toxin-producing clostridia in soils of Argentina. *Appl Environ Microbiol* **2005,** *71* (7), 4137-9.

77. Smith, L. D., The occurrence of *Clostridium botulinum* and *Clostridium tetani* in the soil of the United States. *Health Lab Sci* **1978,** *15* (2), 74-80.

78. Myllykoski, J.; Nevas, M.; Lindstrom, M.; Korkeala, H., The detection and prevalence of *Clostridium botulinum* in pig intestinal samples. *Int J Food Microbiol* **2006,** *110* (2), 172-7.

79. Dahlenborg, M.; Borch, E.; Radstrom, P., Development of a combined selection and enrichment PCR procedure for *Clostridium botulinum* Types B, E, and F and its use to determine prevalence in fecal samples from slaughtered pigs. *Appl Environ Microbiol* **2001,** *67* (10), 4781-8.

80. McCallum, N.; Gray, T. J.; Wang, Q.; Ng, J.; Hicks, L.; Nguyen, T.; Yuen, M.; Hill-Cawthorne, G. A.; Sintchenko, V., Genomic Epidemiology of *Clostridium botulinum* Isolates from Temporally Related Cases of Infant Botulism in New South Wales, Australia. *J Clin Microbiol* **2015,** *53* (9), 2846-53.

81. Viray, M. A.; Wamala, J.; Fagan, R.; Luquez, C.; Maslanka, S.; Downing, R.; Biggerstaff, M.; Malimbo, M.; Kirenga, J. B.; Nakibuuka, J.; Ddumba, E.; Mbabazi, W.; Swerdlow, D. L., Outbreak of type A foodborne botulism at a boarding school, Uganda, 2008. *Epidemiol Infect* **2014,** *142* (11), 2297-301.

82. Lindstrom, M.; Myllykoski, J.; Sivela, S.; Korkeala, H., *Clostridium botulinum* in cattle and dairy products. *Crit Rev Food Sci Nutr* **2010,** *50* (4), 281-304.

83. Vosloo, M. N.; Opperman, C. J.; Geyer, H. D. W.; Setshedi, G. M.; Allam, M.; Kwenda, S.; Ismail, A.; Khumalo, Z. T. H.; Brink, A. J.; Frean, J. A.; Rossouw, J., First confirmed case of infant botulism in Africa, caused by a dual-toxin-producing *Clostridium botulinum* strain. *Int J Infect Dis* **2021,** *103*, 164-166.

84. Souillard, R.; C, L. E. M.; Balaine, L.; Rouxel, S.; Poezevara, T.; Ballan, V.; Chemaly, M.; S, L. E. B., Manure contamination with *Clostridium botulinum* after avian botulism outbreaks: management and potential risk of dissemination. *Vet Rec* **2020,** *187* (6), 233.

85. Leclair, D.; Pagotto, F.; Farber, J. M.; Cadieux, B.; Austin, J. W., Comparison of DNA fingerprinting methods for use in investigation of type E botulism outbreaks in the Canadian Arctic. *J Clin Microbiol* **2006,** *44* (5), 1635-44.

86. Lindstrom, M.; Nevas, M.; Kurki, J.; Sauna-aho, R.; Latvala-Kiesila, A.; Polonen, I.; Korkeala, H., Type C botulism due to toxic feed affecting 52,000 farmed foxes and minks in Finland. *J Clin Microbiol* **2004,** *42* (10), 4718-25.

87. Yule, A. M.; Barker, I. K.; Austin, J. W.; Moccia, R. D., Toxicity of *Clostridium botulinum* type E neurotoxin to Great Lakes fish: implications for avian botulism. *J Wildl Dis* **2006,** *42* (3), 479-93.

88. Rao, A. K.; Sobel, J.; Chatham-Stephens, K.; Luquez, C., Clinical Guidelines for Diagnosis and Treatment of Botulism, 2021. *MMWR Recomm Rep* **2021,** *70* (2), 1-30.

89. Brett, M. M.; Hallas, G.; Mpamugo, O., Wound botulism in the UK and Ireland. *J Med Microbiol* **2004,** *53* (Pt 6), 555-561.

90. Harris, R. A.; Anniballi, F.; Austin, J. W., Adult Intestinal Toxemia Botulism. *Toxins (Basel)* **2020,** *12* (2).

91. Antonucci, L.; Locci, C.; Schettini, L.; Clemente, M. G.; Antonucci, R., Infant botulism: an underestimated threat. *Infect Dis (Lond)* **2021**, 1-14.

92. Koepke, R.; Sobel, J.; Arnon, S. S., Global occurrence of infant botulism, 1976-2006. *Pediatrics* **2008,** *122* (1), e73-82.

93. Anniballi, F.; Fiore, A.; Lofstrom, C.; Skarin, H.; Auricchio, B.; Woudstra, C.; Bano, L.; Segerman, B.; Koene, M.; Baverud, V.; Hansen, T.; Fach, P.; Tevell Aberg, A.; Hedeland, M.; Olsson Engvall, E.; De Medici, D., Management of animal botulism outbreaks: from clinical suspicion to practical countermeasures to prevent or minimize outbreaks. *Biosecur Bioterror* **2013,** *11 Suppl 1*, S191-9.

94. Leonardi, L.; Haggiag, S.; Petrucci, A.; Lispi, L., Electrophysiological abnormalities in iatrogenic botulism: Two case reports and review of the literature. *J Clin Neurosci* **2019,** *60*, 138-141.

95. Chertow, D. S.; Tan, E. T.; Maslanka, S. E.; Schulte, J.; Bresnitz, E. A.; Weisman, R. S.; Bernstein, J.; Marcus, S. M.; Kumar, S.; Malecki, J.; Sobel, J.;

Braden, C. R., Botulism in 4 adults following cosmetic injections with an unlicensed, highly concentrated botulinum preparation. *JAMA* **2006,** *296* (20), 2476-9.

96. Coban, A.; Matur, Z.; Hanagasi, H. A.; Parman, Y., Iatrogenic botulism after botulinum toxin type A injections. *Clin Neuropharmacol* **2010,** *33* (3), 158-60.

97. Ghasemi, M.; Norouzi, R.; Salari, M.; Asadi, B., Iatrogenic botulism after the therapeutic use of botulinum toxin-A: a case report and review of the literature. *Clin Neuropharmacol* **2012,** *35* (5), 254-7.

98. Yiannakopoulou, E., Serious and long-term adverse events associated with the therapeutic and cosmetic use of botulinum toxin. *Pharmacology* **2015,** *95* (1-2), 65-9.

99. Cenciarelli, O.; Riley, P. W.; Baka, A., Biosecurity Threat Posed by Botulinum Toxin. *Toxins (Basel)* **2019,** *11* (12).

100. Berger, T.; Eisenkraft, A.; Bar-Haim, E.; Kassirer, M.; Aran, A. A.; Fogel, I., Toxins as biological weapons for terror-characteristics, challenges and medical countermeasures: a mini-review. *Disaster Mil Med* **2016,** *2*, 7.

101. Shukla, H. D.; Sharma, S. K., *Clostridium botulinum*: a bug with beauty and weapon. *Crit Rev Microbiol* **2005,** *31* (1), 11-8.

102. Villar, R. G.; Elliott, S. P.; Davenport, K. M., Botulism: the many faces of botulinum toxin and its potential for bioterrorism. *Infect Dis Clin North Am* **2006,** *20* (2), 313-27, ix.

103. Janik, E.; Ceremuga, M.; Saluk-Bijak, J.; Bijak, M., Biological Toxins as the Potential Tools for Bioterrorism. *Int J Mol Sci* **2019,** *20* (5).

104. Turton, K.; Chaddock, J. A.; Acharya, K. R., Botulinum and tetanus neurotoxins: structure, function and therapeutic utility. *Trends Biochem Sci* **2002,** *27* (11), 552-8.

105. Schussler, E.; Sobel, J.; Hsu, J.; Yu, P.; Meaney-Delman, D.; Grammer, L. C., 3rd; Nowak-Wegrzyn, A., Workgroup Report by the Joint Task Force Involving American Academy of Allergy, Asthma & Immunology (AAAAI); Food Allergy, Anaphylaxis, Dermatology and Drug Allergy (FADDA) (Adverse Reactions to Foods Committee and Adverse Reactions to Drugs, Biologicals, and Latex Committee); and the Centers for Disease Control and Prevention Botulism Clinical Treatment Guidelines Workgroup-Allergic Reactions to Botulinum Antitoxin: A Systematic Review. *Clin Infect Dis* **2017,** *66* (suppl_1), S65-S72.

106. Pirazzini, M.; Rossetto, O.; Eleopra, R.; Montecucco, C., Botulinum Neurotoxins: Biology, Pharmacology, and Toxicology. *Pharmacol Rev* **2017,** *69* (2), 200-235.

107. Pirazzini, M.; Rossetto, O., Challenges in searching for therapeutics against Botulinum Neurotoxins. *Expert Opin Drug Discov* **2017,** *12* (5), 497-510.

108. Sundeen, G.; Barbieri, J. T., Vaccines against Botulism. *Toxins (Basel)* **2017,** *9* (9).

109. Rusnak, J. M.; Smith, L. A., Botulinum neurotoxin vaccines: Past history and recent developments. *Hum Vaccin* **2009,** *5* (12), 794-805.

110. Rasetti-Escargueil, C.; Popoff, M. R., Antibodies and Vaccines against Botulinum Toxins: Available Measures and Novel Approaches. *Toxins (Basel)* **2019,** *11* (9).

111. Holzer, E., Botulism caused by inhalation. *Med Klin* **1962,** *57*, 1735-8.

112. Dressler, D., Clinical applications of botulinum toxin. *Curr Opin Microbiol* **2012,** *15* (3), 325-36.

113. Simpson, D. M.; Hallett, M.; Ashman, E. J.; Comella, C. L.; Green, M. W.; Gronseth, G. S.; Armstrong, M. J.; Gloss, D.; Potrebic, S.; Jankovic, J.; Karp, B. P.; Naumann, M.; So, Y. T.; Yablon, S. A., Practice guideline update summary: Botulinum neurotoxin for the treatment of blepharospasm, cervical dystonia, adult spasticity, and headache: Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology* **2016,** *86* (19), 1818-26.

114. Nicacio, G. M.; Luna, S. P. L.; Cavaleti, P.; Cassu, R. N., Intra-articular botulinum toxin A (BoNT/A) for pain management in dogs with osteoarthritis secondary to hip dysplasia: A randomized controlled clinical trial. *J Vet Med Sci* **2019,** *81* (3), 411-417.

115. Wijnberg, I. D.; Hardeman, L. C.; van der Meij, B. R.; Veraa, S.; Back, W.; van der Kolk, J. H., The effect of *Clostridium botulinum* toxin type A injections on motor unit activity of the deep digital flexor muscle in healthy sound Royal Dutch sport horses. *Vet J* **2013,** *198 Suppl 1*, e147-51.

116. Borodic, G., Botulinum toxin, immunologic considerations with long-term repeated use, with emphasis on cosmetic applications. *Facial Plast Surg Clin North Am* **2007,** *15* (1), 11-6, v.

117. Naumann, M.; Albanese, A.; Heinen, F.; Molenaers, G.; Relja, M., Safety and efficacy of botulinum toxin type A following long-term use. *Eur J Neurol* **2006,** *13 Suppl 4*, 35-40.

118. Ramirez-Castaneda, J.; Jankovic, J., Long-term efficacy, safety, and side effect profile of botulinum toxin in dystonia: a 20-year follow-up. *Toxicon* **2014,** *90*, 344-8.

119. Montecucco, C.; Molgo, J., Botulinal neurotoxins: revival of an old killer. *Curr Opin Pharmacol* **2005,** *5* (3), 274-9.

120. Davletov, B.; Bajohrs, M.; Binz, T., Beyond BOTOX: advantages and limitations of individual botulinum neurotoxins. *Trends Neurosci* **2005,** *28* (8), 446-52.

121. Carruthers, J. D.; Carruthers, J. A., Treatment of glabellar frown lines with *C. botulinum*-A exotoxin. *J Dermatol Surg Oncol* **1992,** *18* (1), 17-21.

122. Guo, Y.; Lu, Y.; Liu, T.; Zhou, Y.; Yang, P.; Zhu, J.; Chen, L.; Yang, Q., Efficacy and Safety of Botulinum Toxin Type A in the Treatment of Glabellar Lines: A Meta-Analysis of Randomized, Placebo-Controlled, Double-Blind Trials. *Plast Reconstr Surg* **2015,** *136* (3), 310e-318e.

123. Naumann, M.; Boo, L. M.; Ackerman, A. H.; Gallagher, C. J., Immunogenicity of botulinum toxins. *J Neural Transm (Vienna)* **2013,** *120* (2), 275-90.

124. Naumann, M.; Dressler, D.; Hallett, M.; Jankovic, J.; Schiavo, G.; Segal, K. R.; Truong, D., Evidence-based review and assessment of botulinum neurotoxin for the treatment of secretory disorders. *Toxicon* **2013,** *67*, 141-52.

125. Hussain, M. A.; Dawson, C. O., Economic Impact of Food Safety Outbreaks on Food Businesses. *Foods* **2013,** *2* (4), 585-589.

126. Bradshaw, M.; Marshall, K. M.; Heap, J. T.; Tepp, W. H.; Minton, N. P.; Johnson, E. A., Construction of a nontoxigenic *Clostridium botulinum* strain for food challenge studies. *Appl Environ Microbiol* **2010,** *76* (2), 387-93.

127. Rosen, H. E.; Kimura, A. C.; Crandall, J.; Poe, A.; Nash, J.; Boetzer, J.; Tecle, S.; Mukhopadhyay, R.; McAuley, K.; Kasirye, O.; Garza, A.; Shahkarami, M.; Chaturvedi, V.; Kiang, D.; Vidanes, J.; McCoy, K.; Barcellos, M.; Derby, T.; Jain, S.; Vugia, D. J., Foodborne Botulism Outbreak Associated With Commercial Nacho Cheese Sauce From a Gas Station Market. *Clin Infect Dis* **2020,** *70* (8), 1695-1700.

128. Pernu, N.; Keto-Timonen, R.; Lindstrom, M.; Korkeala, H., High prevalence of *Clostridium botulinum* in vegetarian sausages. *Food Microbiol* **2020,** *91*, 103512.

129. Nevas, M.; Hielm, S.; Lindstrom, M.; Horn, H.; Koivulehto, K.; Korkeala, H., High prevalence of *Clostridium botulinum* types A and B in honey samples detected by polymerase chain reaction. *Int J Food Microbiol* **2002,** *72* (1-2), 45-52.

130. Barker, G. C.; Malakar, P. K.; Plowman, J.; Peck, M. W., Quantification of Nonproteolytic *Clostridium botulinum* Spore Loads in Food Materials. *Appl Environ Microbiol* **2016,** *82* (6), 1675-85.

131. Peck, M. W.; Webb, M. D.; Goodburn, K. E., Assessment of the risk of botulism from chilled, vacuum/modified atmosphere packed fresh beef, lamb and pork held at 3 degrees C-8 degrees C. *Food Microbiol* **2020,** *91*, 103544.

132. Sobel, J.; Tucker, N.; Sulka, A.; McLaughlin, J.; Maslanka, S., Foodborne botulism in the United States, 1990-2000. *Emerg Infect Dis* **2004,** *10* (9), 1606-11.

133. Brown, K. L., Control of bacterial spores. *Br Med Bull* **2000,** *56* (1), 158-71.

134. Eklund, M. W.; Wieler, D. I.; Poysky, F. T., Outgrowth and toxin production of nonproteolytic type B *Clostridium botulinum* at 3.3 to 5.6 °C. *J Bacteriol* **1967,** *93* (4), 1461-2.

135. Gassara, F.; Kouassi, A. P.; Brar, S. K.; Belkacemi, K., Green Alternatives to Nitrates and Nitrites in Meat-based Products-A Review. *Crit Rev Food Sci Nutr* **2016,** *56* (13), 2133-48.

136. Flores, M.; Toldra, F., Chemistry, safety, and regulatory considerations in the use of nitrite and nitrate from natural origin in meat products - Invited review. *Meat Sci* **2021,** *171*, 108272.

137. Galvez, A.; Abriouel, H.; Lopez, R. L.; Ben Omar, N., Bacteriocin-based strategies for food biopreservation. *Int J Food Microbiol* **2007,** *120* (1-2), 51-70.

138. Yildirim, S.; Rocker, B.; Pettersen, M. K.; Nilsen-Nygaard, J.; Ayhan, Z.; Rutkaite, R.; Radusin, T.; Suminska, P.; Marcos, B.; Coma, V., Active Packaging Applications for Food. *Compr Rev Food Sci Food Saf* **2018,** *17* (1), 165-199.

139. Al-Tayyar, N. A.; Youssef, A. M.; Al-Hindi, R., Antimicrobial food packaging based on sustainable Bio-based materials for reducing foodborne Pathogens: A review. *Food Chem* **2020,** *310*, 125915.

140. Leistner, L., Basic aspects of food preservation by hurdle technology. *Int J Food Microbiol* **2000,** *55* (1-3), 181-6.

141. Gonzalez-Angulo, M.; Clauwers, C.; Harastani, R.; Tonello, C.; Jaime, I.; Rovira, J.; Michiels, C. W., Evaluation of factors influencing the growth of non-toxigenic *Clostridium botulinum* type E and *Clostridium sp*. in high-pressure processed and conditioned tender coconut water from Thailand. *Food Res Int* **2020,** *134*, 109278.

142. Brunt, J.; Plowman, J.; Gaskin, D. J.; Itchner, M.; Carter, A. T.; Peck, M. W., Functional characterisation of germinant receptors in *Clostridium botulinum* and *Clostridium sporogenes* presents novel insights into spore germination systems. *PLoS Pathog* **2014,** *10* (9), e1004382.

143. Lebrun, S.; Van Nieuwenhuysen, T.; Crevecoeur, S.; Vanleyssem, R.; Thimister, J.; Denayer, S.; Jeuge, S.; Daube, G.; Clinquart, A.; Fremaux, B., Influence of reduced levels or suppression of sodium nitrite on the outgrowth and toxinogenesis of psychrotrophic *Clostridium botulinum* Group II type B in cooked ham. *Int J Food Microbiol* **2020,** *334*, 108853.

144. Ihekwaba, A. E.; Mura, I.; Peck, M. W.; Barker, G. C., The pattern of growth observed for *Clostridium botulinum* type A1 strain ATCC 19397 is influenced by nutritional status and quorum sensing: a modelling perspective. *Pathogens and disease* **2015,** *73* (9), ftv084.

145. Ihekwaba, A. E.; Mura, I.; Malakar, P. K.; Walshaw, J.; Peck, M. W.; Barker, G. C., New Elements To Consider When Modeling the Hazards Associated with Botulinum Neurotoxin in Food. *J Bacteriol* **2016,** *198* (2), 204-11.

146. Juneja, V. K.; Purohit, A. S.; Golden, M.; Osoria, M.; Glass, K. A.; Mishra, A.; Thippareddi, H.; Devkumar, G.; Mohr, T. B.; Minocha, U.; Silverman, M.; Schaffner, D. W., A predictive growth model for *Clostridium botulinum* during cooling of cooked uncured ground beef. *Food Microbiol* **2021,** *93*, 103618.

147. Graham, A. F.; Mason, D. R.; Peck, M. W., Predictive model of the effect of temperature, pH and sodium chloride on growth from spores of non-proteolytic *Clostridium botulinum*. *Int J Food Microbiol* **1996,** *31* (1-3), 69-85.

148. Peck, M. W.; van Vliet, A. H., Impact of *Clostridium botulinum* genomic diversity on food safety. *Curr Opin Food Sci* **2016,** *10*, 52-59.

149. Rasetti-Escargueil, C.; Popoff, M. R., Engineering Botulinum Neurotoxins for Enhanced Therapeutic Applications and Vaccine Development. *Toxins (Basel)* **2020,** *13* (1).

150. Sikorra, S.; Donald, S.; Elliott, M.; Schwede, S.; Coker, S. F.; Kupinski, A. P.; Tripathi, V.; Foster, K.; Beard, M.; Binz, T., Engineering an Effective Human

SNAP-23 Cleaving Botulinum Neurotoxin A Variant. *Toxins (Basel)* **2020,** *12* (12).

151. Tao, L.; Peng, L.; Berntsson, R. P.; Liu, S. M.; Park, S.; Yu, F.; Boone, C.; Palan, S.; Beard, M.; Chabrier, P. E.; Stenmark, P.; Krupp, J.; Dong, M., Engineered botulinum neurotoxin B with improved efficacy for targeting human receptors. *Nat Commun* **2017,** *8* (1), 53.

152. Sikorra, S.; Litschko, C.; Muller, C.; Thiel, N.; Galli, T.; Eichner, T.; Binz, T., Identification and Characterization of Botulinum Neurotoxin A Substrate Binding Pockets and Their Re-Engineering for Human SNAP-23. *J Mol Biol* **2016,** *428* (2 Pt A), 372-384.

153. Masuyer, G.; Davies, J. R.; Moore, K.; Chaddock, J. A.; Ravi Acharya, K., Structural analysis of *Clostridium botulinum* neurotoxin type D as a platform for the development of targeted secretion inhibitors. *Sci Rep* **2015,** *5*, 13397.

154. Ferrari, E.; Maywood, E. S.; Restani, L.; Caleo, M.; Pirazzini, M.; Rossetto, O.; Hastings, M. H.; Niranjan, D.; Schiavo, G.; Davletov, B., Re-assembled botulinum neurotoxin inhibits CNS functions without systemic toxicity. *Toxins (Basel)* **2011,** *3* (4), 345-55.

155. Chen, S.; Barbieri, J. T., Engineering botulinum neurotoxin to extend therapeutic intervention. *Proc Natl Acad Sci U S A* **2009,** *106* (23), 9180-4.

156. Webb, R. P., Engineering of Botulinum Neurotoxins for Biomedical Applications. *Toxins (Basel)* **2018,** *10* (6).

157. Fonfria, E.; Elliott, M.; Beard, M.; Chaddock, J. A.; Krupp, J., Engineering Botulinum Toxins to Improve and Expand Targeting and SNARE Cleavage Activity. *Toxins (Basel)* **2018,** *10* (7).

158. Canadas, I. C.; Groothuis, D.; Zygouropoulou, M.; Rodrigues, R.; Minton, N. P., RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. *ACS Synth Biol* **2019,** *8* (6), 1379-1390.

159. Pellett, S.; Tepp, W. H.; Johnson, E. A., Critical Analysis of Neuronal Cell and the Mouse Bioassay for Detection of Botulinum Neurotoxins. *Toxins (Basel)* **2019,** *11* (12).

160. Hobbs, R. J.; Thomas, C. A.; Halliwell, J.; Gwenin, C. D., Rapid Detection of Botulinum Neurotoxins-A Review. *Toxins (Basel)* **2019,** *11* (7).

161. Whitemarsh, R. C.; Tepp, W. H.; Johnson, E. A.; Pellett, S., Persistence of botulinum neurotoxin a subtypes 1-5 in primary rat spinal cord cells. *PLoS One* **2014,** *9* (2), e90252.

162. Whitemarsh, R. C.; Strathman, M. J.; Chase, L. G.; Stankewicz, C.; Tepp, W. H.; Johnson, E. A.; Pellett, S., Novel application of human neurons derived from induced pluripotent stem cells for highly sensitive botulinum neurotoxin detection. *Toxicol Sci* **2012,** *126* (2), 426-35.

163. McNutt, P.; Celver, J.; Hamilton, T.; Mesngon, M., Embryonic stem cell-derived neurons are a novel, highly sensitive tissue culture platform for botulinum research. *Biochem Biophys Res Commun* **2011,** *405* (1), 85-90.

164. Brunt, J.; van Vliet, A. H. M.; Carter, A. T.; Stringer, S. C.; Amar, C.; Grant, K. A.; Godbole, G.; Peck, M. W., Diversity of the Genomes and Neurotoxins of

Strains of *Clostridium botulinum* Group I and *Clostridium sporogenes* Associated with Foodborne, Infant and Wound Botulism. *Toxins (Basel)* **2020,** *12* (9).

165. Skarin, H.; Segerman, B., Plasmidome interchange between *Clostridium botulinum*, *Clostridium novyi* and *Clostridium haemolyticum* converts strains of independent lineages into distinctly different pathogens. *PLoS One* **2014,** *9* (9), e107777.

166. Moreira, C., Jr.; Ferreira, M. R. A.; Finger, P. F.; Magalhaes, C. G.; Cunha, C. E. P.; Rodrigues, R. R.; Otaka, D. Y.; Galvao, C. C.; Salvarani, F. M.; Moreira, A. N.; Conceicao, F. R., Protective efficacy of recombinant bacterin vaccine against botulism in cattle. *Vaccine* **2020,** *38* (11), 2519-2526.

167. Lam, K. H.; Qi, R.; Liu, S.; Kroh, A.; Yao, G.; Perry, K.; Rummel, A.; Jin, R., The hypothetical protein P47 of *Clostridium botulinum* E1 strain Beluga has a structural topology similar to bactericidal/permeability-increasing protein. *Toxicon* **2018,** *147*, 19-26.

168. Selby, K.; Mascher, G.; Somervuo, P.; Lindstrom, M.; Korkeala, H., Heat shock and prolonged heat stress attenuate neurotoxin and sporulation gene expression in group I *Clostridium botulinum* strain ATCC 3502. *PLoS One* **2017,** *12* (5), e0176944.

169. Fredrick, C. M.; Lin, G.; Johnson, E. A., Regulation of Botulinum Neurotoxin Synthesis and Toxin Complex Formation by Arginine and Glucose in *Clostridium botulinum* ATCC 3502. *Appl Environ Microbiol* **2017,** *83* (13).

170. Zhang, Z.; Dahlsten, E.; Korkeala, H.; Lindstrom, M., Positive regulation of botulinum neurotoxin gene expression by CodY in *Clostridium botulinum* ATCC 3502. *Appl Environ Microbiol* **2014,** *80* (24), 7651-8.

171. Cooksley, C. M.; Davis, I. J.; Winzer, K.; Chan, W. C.; Peck, M. W.; Minton, N. P., Regulation of neurotoxin production and sporulation by a Putative agrBD signaling system in proteolytic *Clostridium botulinum*. *Appl Environ Microbiol* **2010,** *76* (13), 4448-60.

172. Soderholm, H.; Derman, Y.; Lindstrom, M.; Korkeala, H., Functional csdA is needed for effective adaptation and initiation of growth of *Clostridium botulinum* ATCC 3502 at suboptimal temperature. *Int J Food Microbiol* **2015,** *208*, 51-7.

173. Derman, Y.; Soderholm, H.; Lindstrom, M.; Korkeala, H., Role of csp genes in NaCl, pH, and ethanol stress response and motility in *Clostridium botulinum* ATCC 3502. *Food Microbiol* **2015,** *46*, 463-470.

174. Mascher, G.; Derman, Y.; Kirk, D. G.; Palonen, E.; Lindstrom, M.; Korkeala, H., The CLO3403/CLO3404 two-component system of *Clostridium botulinum* E1 Beluga is important for cold shock response and growth at low temperatures. *Appl Environ Microbiol* **2014,** *80* (1), 399-407.

175. Marshall, K. M.; Nowaczyk, L., 2nd; Morrissey, T. R.; Loeza, V.; Halik, L. A.; Skinner, G. E.; Reddy, N. R.; Fleischman, G. J.; Larkin, J. W., Effect of sporulation temperature on the resistance of *Clostridium botulinum* type A spores to thermal and high pressure processing. *J Food Prot* **2015,** *78* (1), 146-50.

176. Kirk, D. G.; Zhang, Z.; Korkeala, H.; Lindstrom, M., Alternative sigma factors SigF, SigE, and SigG are essential for sporulation in *Clostridium botulinum* ATCC 3502. *Appl Environ Microbiol* **2014,** *80* (16), 5141-50.

177. Mertaoja, A.; Nowakowska, M. B.; Mascher, G.; Heljanko, V.; Groothuis, D.; Minton, N. P.; Lindstrom, M., CRISPR-Cas9-Based Toolkit for *Clostridium botulinum* Group II Spore and Sporulation Research. *Front Microbiol* **2021,** *12*, 617269.

178. Engstrom, M. D.; Pfleger, B. F., Transcription control engineering and applications in synthetic biology. *Synth Syst Biotechnol* **2017,** *2* (3), 176-191.

179. Heap, J. T.; Pennington, O. J.; Cartman, S. T.; Carter, G. P.; Minton, N. P., The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *J Microbiol Methods* **2007,** *70* (3), 452-64.

180. Heap, J. T.; Pennington, O. J.; Cartman, S. T.; Minton, N. P., A modular system for *Clostridium* shuttle plasmids. *J Microbiol Methods* **2009,** *78* (1), 79-85.

181. Meaney, C. A.; Cartman, S. T.; McClure, P. J.; Minton, N. P., The role of small acid-soluble proteins (SASPs) in protection of spores of *Clostridium botulinum* against nitrous acid. *Int J Food Microbiol* **2016,** *216*, 25-30.

182. Meaney, C. A.; Cartman, S. T.; McClure, P. J.; Minton, N. P., Optimal spore germination in *Clostridium botulinum* ATCC 3502 requires the presence of functional copies of SleB and YpeB, but not CwlJ. *Anaerobe* **2015,** *34*, 86-93.

183. Kirk, D. G.; Dahlsten, E.; Zhang, Z.; Korkeala, H.; Lindstrom, M., Involvement of *Clostridium botulinum* ATCC 3502 sigma factor K in early-stage sporulation. *Appl Environ Microbiol* **2012,** *78* (13), 4590-6.

184. Soderholm, H.; Lindstrom, M.; Somervuo, P.; Heap, J.; Minton, N.; Linden, J.; Korkeala, H., *cspB* encodes a major cold shock protein in *Clostridium botulinum* ATCC 3502. *Int J Food Microbiol* **2011,** *146* (1), 23-30.

185. Derman, Y.; Isokallio, M.; Lindstrom, M.; Korkeala, H., The two-component system CBO2306/CBO2307 is important for cold adaptation of *Clostridium botulinum* ATCC 3502. *Int J Food Microbiol* **2013,** *167* (1), 87-91.

186. Perutka, J.; Wang, W.; Goerlitz, D.; Lambowitz, A. M., Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J Mol Biol* **2004,** *336* (2), 421-39.

187. Ng, Y. K.; Ehsaan, M.; Philip, S.; Collery, M. M.; Janoir, C.; Collignon, A.; Cartman, S. T.; Minton, N. P., Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using *pyrE* alleles. *PLoS One* **2013,** *8* (2), e56051.

188. Heap, J. T.; Ehsaan, M.; Cooksley, C. M.; Ng, Y. K.; Cartman, S. T.; Winzer, K.; Minton, N. P., Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res* **2012,** *40* (8), e59.

189. Cartman, S. T.; Kelly, M. L.; Heeg, D.; Heap, J. T.; Minton, N. P., Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. *Appl Environ Microbiol* **2012,** *78* (13), 4683-90.

190. Ehsaan, M.; Kuit, W.; Zhang, Y.; Cartman, S. T.; Heap, J. T.; Winzer, K.; Minton, N. P., Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824. *Biotechnol Biofuels* **2016,** *9*, 4.

191. Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J.; Charpentier, E.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J.; Terns, R. M.; Terns, M. P.; White, M. F.; Yakunin, A. F.; Garrett, R. A.; van der Oost, J.; Backofen, R.; Koonin, E. V., An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **2015,** *13* (11), 722-36.

192. Mojica, F. J.; Diez-Villasenor, C.; Garcia-Martinez, J.; Soria, E., Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **2005,** *60* (2), 174-82.

193. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P., CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007,** *315* (5819), 1709-12.

194. Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L. A., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **2013,** *31* (3), 233-9.

195. Mali, P.; Esvelt, K. M.; Church, G. M., Cas9 as a versatile tool for engineering biology. *Nat Methods* **2013,** *10* (10), 957-63.

196. Selle, K.; Barrangou, R., Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol* **2015,** *23* (4), 225-32.

197. Sapranauskas, R.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V., The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res* **2011,** *39* (21), 9275-82.

198. Matthews, L. A.; Simmons, L. A., Bacterial nonhomologous end joining requires teamwork. *J Bacteriol* **2014,** *196* (19), 3363-5.

199. Pyne, M. E.; Bruder, M. R.; Moo-Young, M.; Chung, D. A.; Chou, C. P., Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci Rep* **2016,** *6*, 25666.

200. Choi, K. R.; Lee, S. Y., CRISPR technologies for bacterial systems: Current achievements and future directions. *Biotechnol Adv* **2016,** *34* (7), 1180-1209.

201. Bhaya, D.; Davison, M.; Barrangou, R., CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* **2011,** *45*, 273-97.

202. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012,** *337* (6096), 816-21.

203. Mojica, F. J. M.; Diez-Villasenor, C.; Garcia-Martinez, J.; Almendros, C., Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* **2009,** *155* (Pt 3), 733-740.

204. Li, Q.; Chen, J.; Minton, N. P.; Zhang, Y.; Wen, Z.; Liu, J.; Yang, H.; Zeng, Z.; Ren, X.; Yang, J.; Gu, Y.; Jiang, W.; Jiang, Y.; Yang, S., CRISPR-based genome

editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnol J* **2016,** *11* (7), 961-72.

205. Sternberg, S. H.; Doudna, J. A., Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol Cell* **2015,** *58* (4), 568-74.

206. Minton, N. P.; Ehsaan, M.; Humphreys, C. M.; Little, G. T.; Baker, J.; Henstra, A. M.; Liew, F.; Kelly, M. L.; Sheng, L.; Schwarz, K.; Zhang, Y., A roadmap for gene system development in *Clostridium*. *Anaerobe* **2016,** *41*, 104-112.

207. Adam M Guss, D. G. O., Nicky C Caiazza and Lee R Lynd, Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. *Biotechnology for Biofuels* **2012,** *5* (30).

208. Wang, Y.; Zhang, Z. T.; Seo, S. O.; Choi, K.; Lu, T.; Jin, Y. S.; Blaschek, H. P., Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J Biotechnol* **2015,** *200*, 1-5.

209. Wang, S.; Hong, W.; Dong, S.; Zhang, Z. T.; Zhang, J.; Wang, L.; Wang, Y., Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clin Microbiol Infect* **2018,** *24* (10), 1095-1099.

210. Wang, Y.; Zhang, Z. T.; Seo, S. O.; Lynn, P.; Lu, T.; Jin, Y. S.; Blaschek, H. P., Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "Clean" Mutant Selection in *Clostridium beijerinckii* as an Example. *ACS Synth Biol* **2016,** *5* (7), 721-32.

211. Wasels, F.; Jean-Marie, J.; Collas, F.; Lopez-Contreras, A. M.; Lopes Ferreira, N., A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *J Microbiol Methods* **2017,** *140*, 5-11.

212. McAllister, K. N.; Bouillaut, L.; Kahn, J. N.; Self, W. T.; Sorg, J. A., Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Sci Rep* **2017,** *7* (1), 14672.

213. Bruder, M. R.; Pyne, M. E.; Moo-Young, M.; Chung, D. A.; Chou, C. P., Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus *Clostridium*. *Appl Environ Microbiol* **2016,** *82* (20), 6109-6119.

214. Zhang, J.; Zong, W.; Hong, W.; Zhang, Z. T.; Wang, Y., Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab Eng* **2018,** *47*, 49-59.

215. Maikova, A.; Kreis, V.; Boutserin, A.; Severinov, K.; Soutourina, O., Using an Endogenous CRISPR-Cas System for Genome Editing in the Human Pathogen *Clostridium difficile*. *Appl Environ Microbiol* **2019,** *85* (20).

216. Walker, J. E.; Lanahan, A. A.; Zheng, T.; Toruno, C.; Lynd, L. R.; Cameron, J. C.; Olson, D. G.; Eckert, C. A., Development of both type I-B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium *Clostridium thermocellum*. *Metab Eng Commun* **2020,** *10*, e00116.

217. Dong, H.; Tao, W.; Zhang, Y.; Li, Y., Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium*

acetobutylicum: A useful tool for strain engineering. *Metab Eng* **2012,** *14* (1), 59-67.

218. Zhang, Y.; Xu, S.; Chai, C.; Yang, S.; Jiang, W.; Minton, N. P.; Gu, Y., Development of an inducible transposon system for efficient random mutagenesis in *Clostridium acetobutylicum*. *FEMS Microbiol Lett* **2016,** *363* (8).

219. Hartman, A. H.; Liu, H.; Melville, S. B., Construction and characterization of a lactose-inducible promoter system for controlled gene expression in *Clostridium perfringens*. *Appl Environ Microbiol* **2011,** *77* (2), 471-8.

220. Banerjee, A.; Leang, C.; Ueki, T.; Nevin, K. P.; Lovley, D. R., Lactose-inducible system for metabolic engineering of *Clostridium ljungdahlii*. *Appl Environ Microbiol* **2014,** *80* (8), 2410-6.

221. Blasco, I. C. C. Novel Genetic Tools for Clostridia and a Synthetic Biology Approach to Control Spore Formation in Process Organisms. University of Nottingham, 2018.

222. Topp, S.; Reynoso, C. M.; Seeliger, J. C.; Goldlust, I. S.; Desai, S. K.; Murat, D.; Shen, A.; Puri, A. W.; Komeili, A.; Bertozzi, C. R.; Scott, J. R.; Gallivan, J. P., Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl Environ Microbiol* **2010,** *76* (23), 7881-4.

223. Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P., A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chem Biol* **2007,** *14* (2), 173-84.

224. Shao, Y. W.; Ramaswamy, H. S., *Clostridium sporogenes*-ATCC 7955 Spore Destruction Kinetics in Milk Under High Pressure and Elevated Temperature Treatment Conditions. *Food Bioprocess Tech* **2011,** *4* (3), 458-468.

225. Humphreys, C. M. The Generation of Neurotoxin-minus Reporter Strains of *Clostridium botulinum* to Facilitate the Development of Intervention and Preservation Systems in Food. University of Nottingham, 2013.

226. Horton, R. M., In vitro recombination and mutagenesis of DNA. SOEing together tailor-made genes. *Methods Mol Biol* **1997,** *67*, 141-9.

227. Ingle, P.; Groothuis, D.; Rowe, P.; Huang, H.; Cockayne, A.; Kuehne, S. A.; Jiang, W.; Gu, Y.; Humphreys, C. M.; Minton, N. P., Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system. *Sci Rep* **2019,** *9* (1), 8123.

228. Heap, J. T.; Kuehne, S. A.; Ehsaan, M.; Cartman, S. T.; Cooksley, C. M.; Scott, J. C.; Minton, N. P., The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *J Microbiol Methods* **2010,** *80* (1), 49-55.

229. Brettin, T.; Davis, J. J.; Disz, T.; Edwards, R. A.; Gerdes, S.; Olsen, G. J.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G. D.; Shukla, M.; Thomason, J. A., 3rd; Stevens, R.; Vonstein, V.; Wattam, A. R.; Xia, F., RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* **2015,** *5*, 8365.

230. Ochman, H.; Lawrence, J. G.; Groisman, E. A., Lateral gene transfer and the nature of bacterial innovation. *Nature* **2000,** *405* (6784), 299-304.

231. Dubnau, D., DNA uptake in bacteria. *Annu Rev Microbiol* **1999,** *53*, 217-44.

232. Redfield, R. J., Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. *J Hered* **1993,** *84* (5), 400-4.

233. Davies, J.; Davies, D., Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **2010,** *74* (3), 417-33.

234. Slager, J.; Kjos, M.; Attaiech, L.; Veening, J. W., Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* **2014,** *157* (2), 395-406.

235. Smith, T. J.; Hill, K. K.; Raphael, B. H., Historical and current perspectives on *Clostridium botulinum* diversity. *Res Microbiol* **2015,** *166* (4), 290-302.

236. Johnston, C.; Martin, B.; Fichant, G.; Polard, P.; Claverys, J. P., Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* **2014,** *12* (3), 181-96.

237. Claverys, J. P.; Prudhomme, M.; Martin, B., Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol* **2006,** *60*, 451-75.

238. Canchaya, C.; Fournous, G.; Brussow, H., The impact of prophages on bacterial chromosomes. *Mol Microbiol* **2004,** *53* (1), 9-18.

239. Chen, I.; Christie, P. J.; Dubnau, D., The ins and outs of DNA transfer in bacteria. *Science* **2005,** *310* (5753), 1456-60.

240. Griffith, F., The Significance of Pneumococcal Types. *J Hyg (Lond)* **1928,** *27* (2), 113-59.

241. Avery, O. T.; Macleod, C. M.; McCarty, M., Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type Iii. *J Exp Med* **1944,** *79* (2), 137-58.

242. Chen, I.; Dubnau, D., DNA uptake during bacterial transformation. *Nat Rev Microbiol* **2004,** *2* (3), 241-9.

243. Thomas, C. M.; Nielsen, K. M., Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **2005,** *3* (9), 711-21.

244. Hanahan, D.; Jessee, J.; Bloom, F. R., Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol* **1991,** *204*, 63-113.

245. Dower, W. J.; Miller, J. F.; Ragsdale, C. W., High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **1988,** *16* (13), 6127-45.

246. Tu, Q.; Yin, J.; Fu, J.; Herrmann, J.; Li, Y.; Yin, Y.; Stewart, A. F.; Muller, R.; Zhang, Y., Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency. *Sci Rep* **2016,** *6*, 24648.

247. Fernandez, L.; Rodriguez, A.; Garcia, P., Phage or foe: an insight into the impact of viral predation on microbial communities. *ISME J* **2018,** *12* (5), 1171-1179.

248. Krupovic, M.; Prangishvili, D.; Hendrix, R. W.; Bamford, D. H., Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev* **2011,** *75* (4), 610-35.

249. Sorensen, S. J.; Bailey, M.; Hansen, L. H.; Kroer, N.; Wuertz, S., Studying plasmid horizontal transfer in situ: a critical review. *Nat Rev Microbiol* **2005,** *3* (9), 700-10.

250. Hu, B.; Khara, P.; Christie, P. J., Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli*. *Proc Natl Acad Sci U S A* **2019,** *116* (28), 14222-14227.

251. Lawley, T. D.; Klimke, W. A.; Gubbins, M. J.; Frost, L. S., F factor conjugation is a true type IV secretion system. *FEMS Microbiology Letters* **2003,** *224* (1), 1-15.

252. Fu, Y. H.; Tsai, M. M.; Luo, Y. N.; Deonier, R. C., Deletion analysis of the F plasmid oriT locus. *J Bacteriol* **1991,** *173* (3), 1012-20.

253. Samson, J. E.; Magadan, A. H.; Sabri, M.; Moineau, S., Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **2013,** *11* (10), 675-87.

254. van Houte, S.; Buckling, A.; Westra, E. R., Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol Mol Biol Rev* **2016,** *80* (3), 745-63.

255. Bernheim, A.; Sorek, R., The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol* **2020,** *18* (2), 113-119.

256. Stern, A.; Sorek, R., The phage-host arms race: shaping the evolution of microbes. *Bioessays* **2011,** *33* (1), 43-51.

257. Makarova, K. S.; Wolf, Y. I.; Snir, S.; Koonin, E. V., Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol* **2011,** *193* (21), 6039-56.

258. Ibanez de Aldecoa, A. L.; Zafra, O.; Gonzalez-Pastor, J. E., Mechanisms and Regulation of Extracellular DNA Release and Its Biological Roles in Microbial Communities. *Front Microbiol* **2017,** *8*, 1390.

259. Labrie, S. J.; Samson, J. E.; Moineau, S., Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **2010,** *8* (5), 317-27.

260. Rostol, J. T.; Marraffini, L., (Ph)ighting Phages: How Bacteria Resist Their Parasites. *Cell Host Microbe* **2019,** *25* (2), 184-194.

261. Dedrick, R. M.; Jacobs-Sera, D.; Bustamante, C. A.; Garlena, R. A.; Mavrich, T. N.; Pope, W. H.; Reyes, J. C.; Russell, D. A.; Adair, T.; Alvey, R.; Bonilla, J. A.; Bricker, J. S.; Brown, B. R.; Byrnes, D.; Cresawn, S. G.; Davis, W. B.; Dickson, L. A.; Edgington, N. P.; Findley, A. M.; Golebiewska, U.; Grose, J. H.; Hayes, C. F.; Hughes, L. E.; Hutchison, K. W.; Isern, S.; Johnson, A. A.; Kenna, M. A.; Klyczek, K. K.; Mageeney, C. M.; Michael, S. F.; Molloy, S. D.; Montgomery, M. T.; Neitzel, J.; Page, S. T.; Pizzorno, M. C.; Poxleitner, M. K.; Rinehart, C. A.; Robinson, C. J.; Rubin, M. R.; Teyim, J. N.; Vazquez, E.; Ware, V. C.; Washington, J.; Hatfull, G. F., Prophage-mediated defence against viral attack and viral counter-defence. *Nat Microbiol* **2017,** *2*, 16251.

262. Wozniak, R. A.; Waldor, M. K., Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* **2010,** *8* (8), 552-63.

263. Possoz, C.; Gagnat, J.; Sezonov, G.; Guerineau, M.; Pernodet, J. L., Conjugal immunity of *Streptomyces* strains carrying the integrative element pSAM2 is

due to the pif gene (pSAM2 immunity factor). *Mol Microbiol* **2003,** *47* (5), 1385-93.

264. Pohlman, R. F.; Genetti, H. D.; Winans, S. C., Entry exclusion of the IncN plasmid pKM101 is mediated by a single hydrophilic protein containing a lipid attachment motif. *Plasmid* **1994,** *31* (2), 158-65.

265. Haase, J.; Kalkum, M.; Lanka, E., TrbK, a small cytoplasmic membrane lipoprotein, functions in entry exclusion of the IncP alpha plasmid RP4. *J Bacteriol* **1996,** *178* (23), 6720-9.

266. Hegge, J. W.; Swarts, D. C.; van der Oost, J., Prokaryotic Argonaute proteins: novel genome-editing tools? *Nat Rev Microbiol* **2018,** *16* (1), 5-11.

267. Swarts, D. C.; Hegge, J. W.; Hinojo, I.; Shiimori, M.; Ellis, M. A.; Dumrongkulraksa, J.; Terns, R. M.; Terns, M. P.; van der Oost, J., Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Res* **2015,** *43* (10), 5120-9.

268. Swarts, D. C.; Jore, M. M.; Westra, E. R.; Zhu, Y.; Janssen, J. H.; Snijders, A. P.; Wang, Y.; Patel, D. J.; Berenguer, J.; Brouns, S. J. J.; van der Oost, J., DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **2014,** *507* (7491), 258-261.

269. Kuzmenko, A.; Oguienko, A.; Esyunina, D.; Yudin, D.; Petrova, M.; Kudinova, A.; Maslova, O.; Ninova, M.; Ryazansky, S.; Leach, D.; Aravin, A. A.; Kulbachinskiy, A., DNA targeting and interference by a bacterial Argonaute nuclease. *Nature* **2020,** *587* (7835), 632-637.

270. Fu, L.; Xie, C.; Jin, Z.; Tu, Z.; Han, L.; Jin, M.; Xiang, Y.; Zhang, A., The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria. *Nucleic Acids Res* **2019,** *47* (7), 3568-3579.

271. Jolly, S. M.; Gainetdinov, I.; Jouravleva, K.; Zhang, H.; Strittmatter, L.; Bailey, S. M.; Hendricks, G. M.; Dhabaria, A.; Ueberheide, B.; Zamore, P. D., *Thermus thermophilus* Argonaute Functions in the Completion of DNA Replication. *Cell* **2020,** *182* (6), 1545-1559 e18.

272. Hille, F.; Richter, H.; Wong, S. P.; Bratovic, M.; Ressel, S.; Charpentier, E., The Biology of CRISPR-Cas: Backward and Forward. *Cell* **2018,** *172* (6), 1239-1259.

273. Adli, M., The CRISPR tool kit for genome editing and beyond. *Nat Commun* **2018,** *9* (1), 1911.

274. Pourcel, C.; Salvignol, G.; Vergnaud, G., CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading)* **2005,** *151* (Pt 3), 653-663.

275. van der Oost, J.; Westra, E. R.; Jackson, R. N.; Wiedenheft, B., Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **2014,** *12* (7), 479-92.

276. Marraffini, L. A.; Sontheimer, E. J., CRISPR interference limits horizontal gene transfer in *staphylococci* by targeting DNA. *Science* **2008,** *322* (5909), 1843-5.

277. Brouns, S. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J.; Snijders, A. P.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der Oost, J., Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **2008,** *321* (5891), 960-4.

278. Sternberg, S. H.; LaFrance, B.; Kaplan, M.; Doudna, J. A., Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **2015,** *527* (7576), 110-3.

279. Dimitriu, T.; Szczelkun, M. D.; Westra, E. R., Evolutionary Ecology and Interplay of Prokaryotic Innate and Adaptive Immune Systems. *Curr Biol* **2020,** *30* (19), R1189-R1202.

280. Gophna, U.; Kristensen, D. M.; Wolf, Y. I.; Popa, O.; Drevet, C.; Koonin, E. V., No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. *ISME J* **2015,** *9* (9), 2021-7.

281. Oliveira, P. H.; Touchon, M.; Rocha, E. P., The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res* **2014,** *42* (16), 10618-31.

282. Bickle, T. A.; Kruger, D. H., Biology of DNA restriction. *Microbiol Rev* **1993,** *57* (2), 434-50.

283. Wilson, G. G.; Murray, N. E., Restriction and modification systems. *Annu Rev Genet* **1991,** *25*, 585-627.

284. Rocha, E. P.; Danchin, A.; Viari, A., Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Res* **2001,** *11* (6), 946-58.

285. Rusinov, I.; Ershova, A.; Karyagina, A.; Spirin, S.; Alexeevski, A., Lifespan of restriction-modification systems critically affects avoidance of their recognition sites in host genomes. *BMC Genomics* **2015,** *16*, 1084.

286. Vasu, K.; Nagaraja, V., Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* **2013,** *77* (1), 53-72.

287. Jeltsch, A., Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene* **2003,** *317* (1-2), 13-6.

288. Moser, D. P.; Zarka, D.; Kallas, T., Characterization of a restriction barrier and electrotransformation of the *Cyanobacterium Nostoc* PCC 7121. *Arch Microbiol* **1993,** *160* (3), 229-37.

289. Purdy, D.; O'Keeffe, T. A.; Elmore, M.; Herbert, M.; McLeod, A.; Bokori-Brown, M.; Ostrowski, A.; Minton, N. P., Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol* **2002,** *46* (2), 439-52.

290. Pinedo, C. A.; Smets, B. F., Conjugal TOL transfer from *Pseudomonas putida* to *Pseudomonas aeruginosa*: effects of restriction proficiency, toxicant exposure, cell density ratios, and conjugation detection method on observed transfer efficiencies. *Appl Environ Microbiol* **2005,** *71* (1), 51-7.

291. Tock, M. R.; Dryden, D. T., The biology of restriction and anti-restriction. *Curr Opin Microbiol* **2005,** *8* (4), 466-72.

292. Roberts, R. J.; Belfort, M.; Bestor, T.; Bhagwat, A. S.; Bickle, T. A.; Bitinaite, J.; Blumenthal, R. M.; Degtyarev, S.; Dryden, D. T.; Dybvig, K.; Firman, K.; Gromova, E. S.; Gumport, R. I.; Halford, S. E.; Hattman, S.; Heitman, J.; Hornby, D. P.; Janulaitis, A.; Jeltsch, A.; Josephsen, J.; Kiss, A.; Klaenhammer, T. R.; Kobayashi, I.; Kong, H.; Kruger, D. H.; Lacks, S.; Marinus, M. G.; Miyahara, M.; Morgan, R. D.; Murray, N. E.; Nagaraja, V.; Piekarowicz, A.; Pingoud, A.; Raleigh, E.; Rao, D. N.; Reich, N.; Repin, V. E.; Selker, E. U.; Shaw, P. C.; Stein, D. C.; Stoddard, B. L.; Szybalski, W.; Trautner, T. A.; Van Etten, J. L.; Vitor, J. M.; Wilson, G. G.; Xu, S. Y., A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **2003,** *31* (7), 1805-12.

293. Roer, L.; Aarestrup, F. M.; Hasman, H., The EcoKI type I restriction-modification system in *Escherichia coli* affects but is not an absolute barrier for conjugation. *J Bacteriol* **2015,** *197* (2), 337-42.

294. Murray, N. E., Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **2000,** *64* (2), 412-34.

295. Janscak, P.; MacWilliams, M. P.; Sandmeier, U.; Nagaraja, V.; Bickle, T. A., DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J* **1999,** *18* (9), 2638-47.

296. Seidel, R.; van Noort, J.; van der Scheer, C.; Bloom, J. G.; Dekker, N. H.; Dutta, C. F.; Blundell, A.; Robinson, T.; Firman, K.; Dekker, C., Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nat Struct Mol Biol* **2004,** *11* (9), 838-43.

297. Kennaway, C. K.; Taylor, J. E.; Song, C. F.; Potrzebowski, W.; Nicholson, W.; White, J. H.; Swiderska, A.; Obarska-Kosinska, A.; Callow, P.; Cooper, L. P.; Roberts, G. A.; Artero, J. B.; Bujnicki, J. M.; Trinick, J.; Kneale, G. G.; Dryden, D. T., Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes Dev* **2012,** *26* (1), 92-104.

298. Pingoud, A.; Fuxreiter, M.; Pingoud, V.; Wende, W., Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci* **2005,** *62* (6), 685-707.

299. Pingoud, A.; Wilson, G. G.; Wende, W., Type II restriction endonucleases-a historical perspective and more. *Nucleic Acids Res* **2014,** *42* (12), 7489-527.

300. Dryden, D. T.; Murray, N. E.; Rao, D. N., Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res* **2001,** *29* (18), 3728-41.

301. Butterer, A.; Pernstich, C.; Smith, R. M.; Sobott, F.; Szczelkun, M. D.; Toth, J., Type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res* **2014,** *42* (8), 5139-50.

302. Crampton, N.; Roes, S.; Dryden, D. T.; Rao, D. N.; Edwardson, J. M.; Henderson, R. M., DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *EMBO J* **2007,** *26* (16), 3815-25.

303. Loenen, W. A.; Raleigh, E. A., The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res* **2014,** *42* (1), 56-69.

304. Bair, C. L.; Black, L. W., A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J Mol Biol* **2007,** *366* (3), 768-78.

305. Ishikawa, K.; Fukuda, E.; Kobayashi, I., Conflicts targeting epigenetic systems and their resolution by cell death: novel concepts for methyl-specific and other restriction systems. *DNA Res* **2010,** *17* (6), 325-42.

306. Aertsen, A.; Michiels, C. W., Mrr instigates the SOS response after high pressure stress in *Escherichia coli*. *Mol Microbiol* **2005,** *58* (5), 1381-91.

307. Aune, T. E.; Aachmann, F. L., Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl Microbiol Biotechnol* **2010,** *85* (5), 1301-13.

308. Roberts, R. J., Vincze, T., Posfai, J., Macelis, D., REBASE-a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **2015,** *43*, D298-D299.

309. Pyne, M. E.; Moo-Young, M.; Chung, D. A.; Chou, C. P., Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnol Biofuels* **2013,** *6* (1), 50.

310. Schafer, A.; Kalinowski, J.; Puhler, A., Increased fertility of *Corynebacterium glutamicum* recipients in intergeneric matings with *Escherichia coli* after stress exposure. *Appl Environ Microbiol* **1994,** *60* (2), 756-9.

311. Zeng, X.; Ardeshna, D.; Lin, J., Heat Shock-Enhanced Conjugation Efficiency in Standard *Campylobacter jejuni* Strains. *Appl Environ Microbiol* **2015,** *81* (13), 4546-52.

312. Kirk, J. A.; Fagan, R. P., Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe* **2016,** *42*, 1-5.

313. Johnston, C. D.; Cotton, S. L.; Rittling, S. R.; Starr, J. R.; Borisy, G. G.; Dewhirst, F. E.; Lemon, K. P., Systematic evasion of the restriction-modification barrier in bacteria. *Proc Natl Acad Sci U S A* **2019,** *116* (23), 11454-11459.

314. Hampton, H. G.; Watson, B. N. J.; Fineran, P. C., The arms race between bacteria and their phage foes. *Nature* **2020,** *577* (7790), 327-336.

315. Monk, I. R.; Shah, I. M.; Xu, M.; Tan, M. W.; Foster, T. J., Transforming the untransformable: application of direct transformation to manipulate genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. *mBio* **2012,** *3* (2).

316. Croux, C.; Nguyen, N. P.; Lee, J.; Raynaud, C.; Saint-Prix, F.; Gonzalez-Pajuelo, M.; Meynial-Salles, I.; Soucaille, P., Construction of a restriction-less, marker-less mutant useful for functional genomic and metabolic engineering of the biofuel producer *Clostridium acetobutylicum*. *Biotechnol Biofuels* **2016,** *9*, 23.

317. Chung, D.; Farkas, J.; Westpheling, J., Overcoming restriction as a barrier to DNA transformation in *Caldicellulosiruptor* species results in efficient marker replacement. *Biotechnol Biofuels* **2013,** *6* (1), 82.

318. Dong, H.; Zhang, Y.; Dai, Z.; Li, Y., Engineering *clostridium* strain to accept unmethylated DNA. *PLoS One* **2010,** *5* (2), e9038.

319. Gonzalez-Ceron, G.; Miranda-Olivares, O. J.; Servin-Gonzalez, L., Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases. *FEMS Microbiol Lett* **2009,** *301* (1), 35-43.

320. Davies, G. P.; Powell, L. M.; Webb, J. L.; Cooper, L. P.; Murray, N. E., EcoKI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities. *Nucleic Acids Res* **1998,** *26* (21), 4828-36.

321. Janscak, P.; Sandmeier, U.; Bickle, T. A., 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 Res* **1999,** *27* (13), 2638-43.

322. Rusinov, I. S.; Ershova, A. S.; Karyagina, A. S.; Spirin, S. A.; Alexeevski, A. V., Avoidance of recognition sites of restriction-modification systems is a widespread but not universal anti-restriction strategy of prokaryotic viruses. *BMC Genomics* **2018,** *19* (1), 885.

323. Wilkins, B. M.; Chilley, P. M.; Thomas, A. T.; Pocklington, M. J., Distribution of restriction enzyme recognition sequences on broad host range plasmid RP4: molecular and evolutionary implications. *J Mol Biol* **1996,** *258* (3), 447-56.

324. Pleska, M.; Guet, C. C., Effects of mutations in phage restriction sites during escape from restriction-modification. *Biol Lett* **2017,** *13* (12).

325. Zhang, G.; Wang, W.; Deng, A.; Sun, Z.; Zhang, Y.; Liang, Y.; Che, Y.; Wen, T., A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria. *PLoS Genet* **2012,** *8* (9), e1002987.

326. Kim, J. Y.; Wang, Y.; Park, M. S.; Ji, G. E., Improvement of transformation efficiency through in vitro methylation and SacII site mutation of plasmid vector in *Bifidobacterium longum* MG1. *J Microbiol Biotechnol* **2010,** *20* (6), 1022-6.

327. Jennert, K. C. B.; Tardif, C.; Young, D. I.; Young, M., Gene transfer to *Clostridium cellulolyticum* ATCC 35319. *Microbiology (Reading)* **2000,** *146 Pt 12*, 3071-3080.

328. Davis, T. O.; Henderson, I.; Brehm, J. K.; Minton, N. P., Development of a transformation and gene reporter system for group II, non-proteolytic *Clostridium botulinum* type B strains. *J Mol Microbiol Biotechnol* **2000,** *2* (1), 59-69.

329. Lesiak, J. M.; Liebl, W.; Ehrenreich, A., Development of an in vivo methylation system for the solventogen *Clostridium saccharobutylicum* NCP 262 and analysis of two endonuclease mutants. *J Biotechnol* **2014,** *188*, 97-9.

330. van Opijnen, T.; Camilli, A., Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* **2013,** *11* (7), 435-42.

331. Wentz, T. G.; Yao, K.; Schill, K. M.; Reddy, N. R.; Skinner, G. E.; Morrissey, T. R.; Wang, Y.; Muruvanda, T.; Manickam, G.; Pillai, C. A.; Thirunavukkarasu, N.; Hoffmann, M.; Hammack, T. S.; Brown, E. W.; Allard, M. W.; Sharma, S. K., Closed Genome Sequence of *Clostridium botulinum* Strain CFSAN064329 (62A). *Genome Announc* **2018,** *6* (26).

332. Aziz, R. K.; Bartels, D.; Best, A. A.; DeJongh, M.; Disz, T.; Edwards, R. A.; Formsma, K.; Gerdes, S.; Glass, E. M.; Kubal, M.; Meyer, F.; Olsen, G. J.; Olson, R.; Osterman, A. L.; Overbeek, R. A.; McNeil, L. K.; Paarmann, D.; Paczian, T.; Parrello, B.; Pusch, G. D.; Reich, C.; Stevens, R.; Vassieva, O.; Vonstein, V.; Wilke, A.; Zagnitko, O., The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **2008,** *9*, 75.

333. Datsenko, K. A.; Wanner, B. L., One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **2000,** *97* (12), 6640-5.

334. Jeong, H.; Sim, Y. M.; Kim, H. J.; Lee, S. J., Unveiling the Hybrid Genome Structure of *Escherichia coli* RR1 (HB101 RecA(+)). *Front Microbiol* **2017,** *8*, 585.

335. Kolek, J.; Sedlar, K.; Provaznik, I.; Patakova, P., Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnol Biofuels* **2016,** *9*, 14.

336. Russell, D. W.; Zinder, N. D., Hemimethylation prevents DNA replication in *E. coli*. *Cell* **1987,** *50* (7), 1071-9.

337. Sitaraman, R.; Leppla, S. H., Methylation-dependent DNA restriction in *Bacillus anthracis*. *Gene* **2012,** *494* (1), 44-50.

338. Woods, C.; Humphreys, C. M.; Rodrigues, R. M.; Ingle, P.; Rowe, P.; Henstra, A. M.; Kopke, M.; Simpson, S. D.; Winzer, K.; Minton, N. P., A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe* **2019,** *59*, 184-191.

339. Marinus, M. G., DNA methylation and mutator genes in *Escherichia coli* K-12. *Mutat Res* **2010,** *705* (2), 71-6.

340. Smets, B. F.; Rittmann, B. E.; Stahl, D. A., Quantification of the effect of substrate concentration on the conjugal transfer rate of the TOL plasmid in short-term batch mating experiments. *Lett Appl Microbiol* **1995,** *21* (3), 167-72.

341. Smets, B. F.; Rittmann, B. E.; Stahl, D. A., The specific growth rate of *Pseudomonas putida* PAW1 influences the conjugal transfer rate of the TOL plasmid. *Appl Environ Microbiol* **1993,** *59* (10), 3430-7.

342. Licht, T. R.; Christensen, B. B.; Krogfelt, K. A.; Molin, S., Plasmid transfer in the animal intestine and other dynamic bacterial populations: the role of community structure and environment. *Microbiology (Reading)* **1999,** *145 (Pt 9)*, 2615-2622.

343. Darwin, C., *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* 1860/04/01 ed.; John Murray: 1859; Vol. 25, p 367-404.

344. Spencer, H., *The principles of biology*. Williams and Norgate: 1864; Vol. 1.

345. Chew, M. K.; Laubichler, M. D., Perceptions of science. Natural enemies-metaphor or misconception? *Science* **2003,** *301* (5629), 52-3.

346. Niles Eldredge, J. N. T., Paul M. Brakefield, Sergey Gavrilets, David Jablonski, Jeremy B. C. Jackson, Richard E. Lenski, Bruce S. Lieberman, Mark A. McPeek

and William Miller, The dynamics of evolutionary stasis. *Paleobiology* **2005,** *31* (2), 133-145.

347. Wiens, J. J.; Ackerly, D. D.; Allen, A. P.; Anacker, B. L.; Buckley, L. B.; Cornell, H. V.; Damschen, E. I.; Jonathan Davies, T.; Grytnes, J. A.; Harrison, S. P.; Hawkins, B. A.; Holt, R. D.; McCain, C. M.; Stephens, P. R., Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol Lett* **2010,** *13* (10), 1310-24.

348. Barrick, J. E.; Yu, D. S.; Yoon, S. H.; Jeong, H.; Oh, T. K.; Schneider, D.; Lenski, R. E.; Kim, J. F., Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **2009,** *461* (7268), 1243-7.

349. Allison, S. D.; Martiny, J. B., Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A* **2008,** *105 Suppl 1*, 11512-9.

350. Cordero, O. X.; Polz, M. F., Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* **2014,** *12* (4), 263-73.

351. Trindade, S.; Perfeito, L.; Gordo, I., Rate and effects of spontaneous mutations that affect fitness in mutator *Escherichia coli*. *Philos Trans R Soc Lond B Biol Sci* **2010,** *365* (1544), 1177-86.

352. Lenski, R. E.; Travisano, M., Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci U S A* **1994,** *91* (15), 6808-14.

353. Good, B. H.; McDonald, M. J.; Barrick, J. E.; Lenski, R. E.; Desai, M. M., The dynamics of molecular evolution over 60,000 generations. *Nature* **2017,** *551* (7678), 45-50.

354. Bodi, Z.; Farkas, Z.; Nevozhay, D.; Kalapis, D.; Lazar, V.; Csorgo, B.; Nyerges, A.; Szamecz, B.; Fekete, G.; Papp, B.; Araujo, H.; Oliveira, J. L.; Moura, G.; Santos, M. A. S.; Szekely, T., Jr.; Balazsi, G.; Pal, C., Phenotypic heterogeneity promotes adaptive evolution. *PLoS Biol* **2017,** *15* (5), e2000644.

355. Crispo, E., Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. *J Evol Biol* **2008,** *21* (6), 1460-9.

356. Veening, J. W.; Smits, W. K.; Kuipers, O. P., Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **2008,** *62*, 193-210.

357. Ackermann, M., A functional perspective on phenotypic heterogeneity in microorganisms. *Nat Rev Microbiol* **2015,** *13* (8), 497-508.

358. Partridge, S. R.; Kwong, S. M.; Firth, N.; Jensen, S. O., Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin Microbiol Rev* **2018,** *31* (4).

359. Wielgoss, S.; Barrick, J. E.; Tenaillon, O.; Wiser, M. J.; Dittmar, W. J.; Cruveiller, S.; Chane-Woon-Ming, B.; Medigue, C.; Lenski, R. E.; Schneider, D., Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci U S A* **2013,** *110* (1), 222-7.

360. Lynch, M., Evolution of the mutation rate. *Trends Genet* **2010,** *26* (8), 345-52.

361. Darmon, E.; Leach, D. R., Bacterial genome instability. *Microbiol Mol Biol Rev* **2014,** *78* (1), 1-39.

362. Vandecraen, J.; Chandler, M.; Aertsen, A.; Van Houdt, R., The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit Rev Microbiol* **2017,** *43* (6), 709-730.

363. Schurch, A. C.; Arredondo-Alonso, S.; Willems, R. J. L.; Goering, R. V., Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect* **2018,** *24* (4), 350-354.

364. Land, M.; Hauser, L.; Jun, S. R.; Nookaew, I.; Leuze, M. R.; Ahn, T. H.; Karpinets, T.; Lund, O.; Kora, G.; Wassenaar, T.; Poudel, S.; Ussery, D. W., Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* **2015,** *15* (2), 141-61.

365. Tenaillon, O.; Barrick, J. E.; Ribeck, N.; Deatherage, D. E.; Blanchard, J. L.; Dasgupta, A.; Wu, G. C.; Wielgoss, S.; Cruveiller, S.; Medigue, C.; Schneider, D.; Lenski, R. E., Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **2016,** *536* (7615), 165-70.

366. Bryant, J.; Chewapreecha, C.; Bentley, S. D., Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiol* **2012,** *7* (11), 1283-1296.

367. Aarestrup, F. M.; Brown, E. W.; Detter, C.; Gerner-Smidt, P.; Gilmour, M. W.; Harmsen, D.; Hendriksen, R. S.; Hewson, R.; Heymann, D. L.; Johansson, K.; Ijaz, K.; Keim, P. S.; Koopmans, M.; Kroneman, A.; Lo Fo Wong, D.; Lund, O.; Palm, D.; Sawanpanyalert, P.; Sobel, J.; Schlundt, J., Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerg Infect Dis* **2012,** *18* (11), e1.

368. Funchain, P.; Yeung, A.; Stewart, J. L.; Lin, R.; Slupska, M. M.; Miller, J. H., The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* **2000,** *154* (3), 959-70.

369. Desai, M. M.; Fisher, D. S., The balance between mutators and nonmutators in asexual populations. *Genetics* **2011,** *188* (4), 997-1014.

370. Perfeito, L.; Fernandes, L.; Mota, C.; Gordo, I., Adaptive mutations in bacteria: high rate and small effects. *Science* **2007,** *317* (5839), 813-5.

371. Eyre-Walker, A.; Keightley, P. D., The distribution of fitness effects of new mutations. *Nat Rev Genet* **2007,** *8* (8), 610-8.

372. Couce, A.; Caudwell, L. V.; Feinauer, C.; Hindre, T.; Feugeas, J. P.; Weigt, M.; Lenski, R. E.; Schneider, D.; Tenaillon, O., Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proc Natl Acad Sci U S A* **2017,** *114* (43), E9026-E9035.

373. Drake, J. W., A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A* **1991,** *88* (16), 7160-4.

374. Lynch, M.; Ackerman, M. S.; Gout, J. F.; Long, H.; Sung, W.; Thomas, W. K.; Foster, P. L., Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* **2016,** *17* (11), 704-714.

375. Miller, J. H., Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. *Annu Rev Microbiol* **1996,** *50*, 625-43.

376. Tenaillon, O.; Toupance, B.; Le Nagard, H.; Taddei, F.; Godelle, B., Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **1999,** *152* (2), 485-93.

377. Zhao, S.; Lieberman, T. D.; Poyet, M.; Kauffman, K. M.; Gibbons, S. M.; Groussin, M.; Xavier, R. J.; Alm, E. J., Adaptive Evolution within Gut Microbiomes of Healthy People. *Cell Host Microbe* **2019,** *25* (5), 656-667 e8.

378. Matic, I.; Radman, M.; Taddei, F.; Picard, B.; Doit, C.; Bingen, E.; Denamur, E.; Elion, J., Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* **1997,** *277* (5333), 1833-4.

379. LeClerc, J. E.; Li, B.; Payne, W. L.; Cebula, T. A., High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **1996,** *274* (5290), 1208-11.

380. Couce, A.; Alonso-Rodriguez, N.; Costas, C.; Oliver, A.; Blazquez, J., Intrapopulation variability in mutator prevalence among urinary tract infection isolates of *Escherichia coli*. *Clin Microbiol Infect* **2016,** *22* (6), 566 e1-7.

381. Oliver, A.; Canton, R.; Campo, P.; Baquero, F.; Blazquez, J., High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **2000,** *288* (5469), 1251-4.

382. Sniegowski, P. D.; Gerrish, P. J.; Lenski, R. E., Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **1997,** *387* (6634), 703-5.

383. Boe, L.; Danielsen, M.; Knudsen, S.; Petersen, J. B.; Maymann, J.; Jensen, P. R., The frequency of mutators in populations of *Escherichia coli*. *Mutat Res* **2000,** *448* (1), 47-55.

384. Wang, S.; Wang, Y.; Shen, J.; Wu, Y.; Wu, C., Polymorphic mutation frequencies in clinical isolates of *Staphylococcus aureus*: the role of weak mutators in the development of fluoroquinolone resistance. *FEMS Microbiol Lett* **2013,** *341* (1), 13-7.

385. Baquero, M. R.; Nilsson, A. I.; Turrientes Mdel, C.; Sandvang, D.; Galan, J. C.; Martinez, J. L.; Frimodt-Moller, N.; Baquero, F.; Andersson, D. I., Polymorphic mutation frequencies in *Escherichia coli*: emergence of weak mutators in clinical isolates. *J Bacteriol* **2004,** *186* (16), 5538-42.

386. Turrientes, M. C.; Baquero, M. R.; Sanchez, M. B.; Valdezate, S.; Escudero, E.; Berg, G.; Canton, R.; Baquero, F.; Galan, J. C.; Martinez, J. L., Polymorphic mutation frequencies of clinical and environmental *Stenotrophomonas maltophilia* populations. *Appl Environ Microbiol* **2010,** *76* (6), 1746-58.

387. Karami-Zarandi, M.; Douraghi, M.; Vaziri, B.; Adibhesami, H.; Rahbar, M.; Yaseri, M., Variable spontaneous mutation rate in clinical strains of multidrug-resistant *Acinetobacter baumannii* and differentially expressed proteins in a hypermutator strain. *Mutat Res* **2017,** *800-802*, 37-45.

388. Ramiro, R. S.; Durao, P.; Bank, C.; Gordo, I., Low mutational load and high mutation rate variation in gut commensal bacteria. *PLoS Biol* **2020,** *18* (3), e3000617.

389. Swings, T.; Van den Bergh, B.; Wuyts, S.; Oeyen, E.; Voordeckers, K.; Verstrepen, K. J.; Fauvart, M.; Verstraeten, N.; Michiels, J., Adaptive tuning of mutation rates allows fast response to lethal stress in *Escherichia coli*. *Elife* **2017,** *6*.

390. Sprouffske, K.; Aguilar-Rodriguez, J.; Sniegowski, P.; Wagner, A., High mutation rates limit evolutionary adaptation in *Escherichia coli*. *PLoS Genet* **2018,** *14* (4), e1007324.

391. Raynes, Y.; Sniegowski, P. D., Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity (Edinb)* **2014,** *113* (5), 375-80.

392. Loh, E.; Salk, J. J.; Loeb, L. A., Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc Natl Acad Sci U S A* **2010,** *107* (3), 1154-9.

393. Giraud, A.; Matic, I.; Tenaillon, O.; Clara, A.; Radman, M.; Fons, M.; Taddei, F., Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* **2001,** *291* (5513), 2606-8.

394. Notley-McRobb, L.; Seeto, S.; Ferenci, T., Enrichment and elimination of mutY mutators in *Escherichia coli* populations. *Genetics* **2002,** *162* (3), 1055-62.

395. Turrientes, M. C.; Baquero, F.; Levin, B. R.; Martinez, J. L.; Ripoll, A.; Gonzalez-Alba, J. M.; Tobes, R.; Manrique, M.; Baquero, M. R.; Rodriguez-Dominguez, M. J.; Canton, R.; Galan, J. C., Normal mutation rate variants arise in a Mutator (MutS) *Escherichia coli* population. *PLoS One* **2013,** *8* (9), e72963.

396. Sniegowski, P. D.; Gerrish, P. J.; Johnson, T.; Shaver, A., The evolution of mutation rates: separating causes from consequences. *Bioessays* **2000,** *22* (12), 1057-66.

397. de Visser, J., The fate of microbial mutators. *Microbiology (Reading)* **2002,** *148* (Pt 5), 1247-1252.

398. Sung, W.; Ackerman, M. S.; Dillon, M. M.; Platt, T. G.; Fuqua, C.; Cooper, V. S.; Lynch, M., Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life. *G3 (Bethesda)* **2016,** *6* (8), 2583-91.

399. Lynch, M., The lower bound to the evolution of mutation rates. *Genome Biol Evol* **2011,** *3*, 1107-18.

400. Wiser, M. J.; Ribeck, N.; Lenski, R. E., Long-term dynamics of adaptation in asexual populations. *Science* **2013,** *342* (6164), 1364-7.

401. Bell, G.; Gonzalez, A., Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. *Science* **2011,** *332* (6035), 1327-30.

402. Blazquez, J., Hypermutation as a factor contributing to the acquisition of antimicrobial resistance. *Clin Infect Dis* **2003,** *37* (9), 1201-9.

403. Revitt-Mills, S. A.; Robinson, A., Antibiotic-Induced Mutagenesis: Under the Microscope. *Front Microbiol* **2020,** *11*, 585175.

404. Sheng, H.; Huang, J.; Han, Z.; Liu, M.; Lu, Z.; Zhang, Q.; Zhang, J.; Yang, J.; Cui, S.; Yang, B., Genes and Proteomes Associated With Increased Mutation Frequency and Multidrug Resistance of Naturally Occurring Mismatch Repair-Deficient *Salmonella* Hypermutators. *Front Microbiol* **2020,** *11*, 770.

405. Travis, J. M.; Travis, E. R., Mutator dynamics in fluctuating environments. *Proc Biol Sci* **2002,** *269* (1491), 591-7.

406. Pal, C.; Macia, M. D.; Oliver, A.; Schachar, I.; Buckling, A., Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **2007,** *450* (7172), 1079-81.

407. Wielgoss, S.; Bergmiller, T.; Bischofberger, A. M.; Hall, A. R., Adaptation to Parasites and Costs of Parasite Resistance in Mutator and Nonmutator Bacteria. *Mol Biol Evol* **2016,** *33* (3), 770-82.

408. Bosshard, L.; Dupanloup, I.; Tenaillon, O.; Bruggmann, R.; Ackermann, M.; Peischl, S.; Excoffier, L., Accumulation of Deleterious Mutations During Bacterial Range Expansions. *Genetics* **2017,** *207* (2), 669-684.

409. Lynch, M.; Burger, R.; Butcher, D.; Gabriel, W., The mutational meltdown in asexual populations. *J Hered* **1993,** *84* (5), 339-44.

410. Peischl, S.; Kirkpatrick, M.; Excoffier, L., Expansion load and the evolutionary dynamics of a species range. *Am Nat* **2015,** *185* (4), E81-93.

411. Covert, A. W., 3rd; Lenski, R. E.; Wilke, C. O.; Ofria, C., Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proc Natl Acad Sci U S A* **2013,** *110* (34), E3171-8.

412. Lehe, R.; Hallatschek, O.; Peliti, L., The rate of beneficial mutations surfing on the wave of a range expansion. *PLoS Comput Biol* **2012,** *8* (3), e1002447.

413. Gralka, M.; Stiewe, F.; Farrell, F.; Mobius, W.; Waclaw, B.; Hallatschek, O., Allele surfing promotes microbial adaptation from standing variation. *Ecol Lett* **2016,** *19* (8), 889-98.

414. Desai, M. M.; Fisher, D. S.; Murray, A. W., The speed of evolution and maintenance of variation in asexual populations. *Curr Biol* **2007,** *17* (5), 385-94.

415. Bull, J. J.; Wilke, C. O., Lethal mutagenesis of bacteria. *Genetics* **2008,** *180* (2), 1061-70.

416. Shaver, A. C.; Dombrowski, P. G.; Sweeney, J. Y.; Treis, T.; Zappala, R. M.; Sniegowski, P. D., Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* **2002,** *162* (2), 557-66.

417. Taddei, F.; Radman, M.; Maynard-Smith, J.; Toupance, B.; Gouyon, P. H.; Godelle, B., Role of mutator alleles in adaptive evolution. *Nature* **1997,** *387* (6634), 700-2.

418. Arjan, J. A.; Visser, M.; Zeyl, C. W.; Gerrish, P. J.; Blanchard, J. L.; Lenski, R. E., Diminishing returns from mutation supply rate in asexual populations. *Science* **1999,** *283* (5400), 404-6.

419. Shaver, A. C.; Sniegowski, P. D., Spontaneously arising mutL mutators in evolving *Escherichia coli* populations are the result of changes in repeat length. *J Bacteriol* **2003,** *185* (20), 6076-82.

420. Lenhart, J. S.; Schroeder, J. W.; Walsh, B. W.; Simmons, L. A., DNA repair and genome maintenance in *Bacillus subtilis*. *Microbiol Mol Biol Rev* **2012,** *76* (3), 530-64.

421. Aravind, L.; Walker, D. R.; Koonin, E. V., Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* **1999,** *27* (5), 1223-42.

422. Hsieh, P., Molecular mechanisms of DNA mismatch repair. *Mutat Res* **2001,** *486* (2), 71-87.

423. Foster, P. L., Stress-induced mutagenesis in bacteria. *Crit Rev Biochem Mol Biol* **2007,** *42* (5), 373-97.

424. Guan, N.; Li, J.; Shin, H. D.; Du, G.; Chen, J.; Liu, L., Microbial response to environmental stresses: from fundamental mechanisms to practical applications. *Appl Microbiol Biotechnol* **2017,** *101* (10), 3991-4008.

425. Eisen, J. A.; Hanawalt, P. C., A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* **1999,** *435* (3), 171-213.

426. Erill, I.; Campoy, S.; Barbe, J., Aeons of distress: an evolutionary perspective on the bacterial SOS response. *FEMS Microbiol Rev* **2007,** *31* (6), 637-56.

427. Kreuzer, K. N., DNA damage responses in prokaryotes: regulating gene expression, modulating growth patterns, and manipulating replication forks. *Cold Spring Harb Perspect Biol* **2013,** *5* (11), a012674.

428. Mesak, L. R.; Miao, V.; Davies, J., Effects of subinhibitory concentrations of antibiotics on SOS and DNA repair gene expression in *Staphylococcus aureus*. *Antimicrob Agents Chemother* **2008,** *52* (9), 3394-7.

429. Maslowska, K. H.; Makiela-Dzbenska, K.; Fijalkowska, I. J., The SOS system: A complex and tightly regulated response to DNA damage. *Environ Mol Mutagen* **2019,** *60* (4), 368-384.

430. Au, N.; Kuester-Schoeck, E.; Mandava, V.; Bothwell, L. E.; Canny, S. P.; Chachu, K.; Colavito, S. A.; Fuller, S. N.; Groban, E. S.; Hensley, L. A.; O'Brien, T. C.; Shah, A.; Tierney, J. T.; Tomm, L. L.; O'Gara, T. M.; Goranov, A. I.; Grossman, A. D.; Lovett, C. M., Genetic composition of the *Bacillus subtilis* SOS system. *J Bacteriol* **2005,** *187* (22), 7655-66.

431. Goranov, A. I.; Kuester-Schoeck, E.; Wang, J. D.; Grossman, A. D., Characterization of the global transcriptional responses to different types of DNA damage and disruption of replication in *Bacillus subtilis*. *J Bacteriol* **2006,** *188* (15), 5595-605.

432. Cox, M. M., Regulation of bacterial RecA protein function. *Crit Rev Biochem Mol Biol* **2007,** *42* (1), 41-63.

433. Groban, E. S.; Johnson, M. B.; Banky, P.; Burnett, P. G.; Calderon, G. L.; Dwyer, E. C.; Fuller, S. N.; Gebre, B.; King, L. M.; Sheren, I. N.; Von Mutius, L. D.; O'Gara, T. M.; Lovett, C. M., Binding of the *Bacillus subtilis* LexA protein to the SOS operator. *Nucleic Acids Res* **2005,** *33* (19), 6287-95.

434. Butala, M.; Zgur-Bertok, D.; Busby, S. J., The bacterial LexA transcriptional repressor. *Cell Mol Life Sci* **2009,** *66* (1), 82-93.

435. Cirz, R. T.; Jones, M. B.; Gingles, N. A.; Minogue, T. D.; Jarrahi, B.; Peterson, S. N.; Romesberg, F. E., Complete and SOS-mediated response of *Staphylococcus aureus* to the antibiotic ciprofloxacin. *J Bacteriol* **2007,** *189* (2), 531-9.

436. van der Veen, S.; van Schalkwijk, S.; Molenaar, D.; de Vos, W. M.; Abee, T.; Wells-Bennik, M. H. J., The SOS response of *Listeria monocytogenes* is involved in stress resistance and mutagenesis. *Microbiology (Reading)* **2010,** *156* (Pt 2), 374-384.

437. Jochmann, N.; Kurze, A. K.; Czaja, L. F.; Brinkrolf, K.; Brune, I.; Huser, A. T.; Hansmeier, N.; Puhler, A.; Borovok, I.; Tauch, A., Genetic makeup of the *Corynebacterium glutamicum* LexA regulon deduced from comparative transcriptomics and in vitro DNA band shift assays. *Microbiology (Reading)* **2009,** *155* (Pt 5), 1459-1477.

438. Nishimura, T.; Teramoto, H.; Inui, M.; Yukawa, H., Gene expression profiling of *Corynebacterium glutamicum* during Anaerobic nitrate respiration: induction of the SOS response for cell survival. *J Bacteriol* **2011,** *193* (6), 1327-33.

439. Nuyts, S.; Van Mellaert, L.; Barbe, S.; Lammertyn, E.; Theys, J.; Landuyt, W.; Bosmans, E.; Lambin, P.; Anne, J., Insertion or deletion of the Cheo box modifies radiation inducibility of *Clostridium promoters*. *Appl Environ Microbiol* **2001,** *67* (10), 4464-70.

440. Walter, B. M.; Rupnik, M.; Hodnik, V.; Anderluh, G.; Dupuy, B.; Paulic, N.; Zgur-Bertok, D.; Butala, M., The LexA regulated genes of the *Clostridium difficile*. *BMC Microbiol* **2014,** *14*, 88.

441. Walter, B. M.; Cartman, S. T.; Minton, N. P.; Butala, M.; Rupnik, M., The SOS Response Master Regulator LexA Is Associated with Sporulation, Motility and Biofilm Formation in *Clostridium difficile*. *PLoS One* **2015,** *10* (12), e0144763.

442. Simmons, L. A.; Goranov, A. I.; Kobayashi, H.; Davies, B. W.; Yuan, D. S.; Grossman, A. D.; Walker, G. C., Comparison of responses to double-strand breaks between *Escherichia coli* and *Bacillus subtilis* reveals different requirements for SOS induction. *J Bacteriol* **2009,** *191* (4), 1152-61.

443. Fijalkowska, I. J.; Schaaper, R. M.; Jonczyk, P., DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair. *FEMS Microbiol Rev* **2012,** *36* (6), 1105-21.

444. Kunkel, T. A.; Bebenek, K., DNA replication fidelity. *Annu Rev Biochem* **2000,** *69*, 497-529.

445. Sanders, G. M.; Dallmann, H. G.; McHenry, C. S., Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Mol Cell* **2010,** *37* (2), 273-81.

446. Dervyn, E.; Suski, C.; Daniel, R.; Bruand, C.; Chapuis, J.; Errington, J.; Janniere, L.; Ehrlich, S. D., Two essential DNA polymerases at the bacterial replication fork. *Science* **2001,** *294* (5547), 1716-9.

447. Fijalkowska, I. J.; Schaaper, R. M., Mutants in the Exo I motif of *Escherichia coli* dnaQ: defective proofreading and inviability due to error catastrophe. *Proc Natl Acad Sci U S A* **1996,** *93* (7), 2856-61.

448. Bruck, I.; Goodman, M. F.; O'Donnell, M., The essential C family DnaE polymerase is error-prone and efficient at lesion bypass. *J Biol Chem* **2003,** *278* (45), 44361-8.

449. Le Chatelier, E.; Becherel, O. J.; d'Alencon, E.; Canceill, D.; Ehrlich, S. D.; Fuchs, R. P.; Janniere, L., Involvement of DnaE, the second replicative DNA polymerase from *Bacillus subtilis*, in DNA mutagenesis. *J Biol Chem* **2004,** *279* (3), 1757-67.

450. Klocko, A. D.; Schroeder, J. W.; Walsh, B. W.; Lenhart, J. S.; Evans, M. L.; Simmons, L. A., Mismatch repair causes the dynamic release of an essential DNA polymerase from the replication fork. *Mol Microbiol* **2011,** *82* (3), 648-63.

451. Kunkel, T. A.; Erie, D. A., DNA mismatch repair. *Annu Rev Biochem* **2005,** *74*, 681-710.

452. Prunier, A. L.; Leclercq, R., Role of *mutS* and *mutL* genes in hypermutability and recombination in *Staphylococcus aureus*. *J Bacteriol* **2005,** *187* (10), 3455-64.

453. Negishi, K.; Loakes, D.; Schaaper, R. M., Saturation of DNA mismatch repair and error catastrophe by a base analogue in *Escherichia coli*. *Genetics* **2002,** *161* (4), 1363-71.

454. Lenhart, J. S.; Pillon, M. C.; Guarne, A.; Biteen, J. S.; Simmons, L. A., Mismatch repair in Gram-positive bacteria. *Res Microbiol* **2016,** *167* (1), 4-12.

455. Cox, E. C.; Degnen, G. E.; Scheppe, M. L., Mutator gene studies in *Escherichia coli*: the *mutS* gene. *Genetics* **1972,** *72* (4), 551-67.

456. Oliver, A.; Baquero, F.; Blazquez, J., The mismatch repair system (*mutS*, *mutL* and *uvrD* genes) in *Pseudomonas aeruginosa*: molecular characterization of naturally occurring mutants. *Mol Microbiol* **2002,** *43* (6), 1641-50.

457. Ginetti, F.; Perego, M.; Albertini, A. M.; Galizzi, A., *Bacillus subtilis mutS mutL* operon: identification, nucleotide sequence and mutagenesis. *Microbiology (Reading)* **1996,** *142 (Pt 8)*, 2021-9.

458. Dupes, N. M.; Walsh, B. W.; Klocko, A. D.; Lenhart, J. S.; Peterson, H. L.; Gessert, D. A.; Pavlick, C. E.; Simmons, L. A., Mutations in the *Bacillus subtilis* beta clamp that separate its roles in DNA replication from mismatch repair. *J Bacteriol* **2010,** *192* (13), 3452-63.

459. Pillon, M. C.; Lorenowicz, J. J.; Uckelmann, M.; Klocko, A. D.; Mitchell, R. R.; Chung, Y. S.; Modrich, P.; Walker, G. C.; Simmons, L. A.; Friedhoff, P.; Guarne, A., Structure of the endonuclease domain of MutL: unlicensed to cut. *Mol Cell* **2010,** *39* (1), 145-51.

460. Yang, H.; Yung, M.; Li, L.; Hoch, J. A.; Ryan, C. M.; Kar, U. K.; Souda, P.; Whitelegge, J. P.; Miller, J. H., Evidence that YycJ is a novel 5'-3' double-stranded DNA exonuclease acting in *Bacillus anthracis* mismatch repair. *DNA Repair (Amst)* **2013,** *12* (5), 334-46.

461. Rossolillo, P.; Albertini, A. M., Functional analysis of the *Bacillus subtilis yshD* gene, a mutS paralogue. *Mol Gen Genet* **2001,** *264* (6), 809-18.

462. Burby, P. E.; Simmons, L. A., MutS2 Promotes Homologous Recombination in *Bacillus subtilis*. *J Bacteriol* **2017,** *199* (2).

463. Pinto, A. V.; Mathieu, A.; Marsin, S.; Veaute, X.; Ielpi, L.; Labigne, A.; Radicella, J. P., Suppression of homologous and homeologous recombination by the bacterial MutS2 protein. *Mol Cell* **2005,** *17* (1), 113-20.

464. Fukui, K.; Kosaka, H.; Kuramitsu, S.; Masui, R., Nuclease activity of the MutS homologue MutS2 from *Thermus thermophilus* is confined to the Smr domain. *Nucleic Acids Res* **2007,** *35* (3), 850-60.

465. Baute, J.; Depicker, A., Base excision repair and its role in maintaining genome stability. *Crit Rev Biochem Mol Biol* **2008,** *43* (4), 239-76.

466. Dalhus, B.; Laerdahl, J. K.; Backe, P. H.; Bjoras, M., DNA base repair-recognition and initiation of catalysis. *FEMS Microbiol Rev* **2009,** *33* (6), 1044-78.

467. Hsu, G. W.; Ober, M.; Carell, T.; Beese, L. S., Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* **2004,** *431* (7005), 217-21.

468. Qi, Y.; Spong, M. C.; Nam, K.; Banerjee, A.; Jiralerspong, S.; Karplus, M.; Verdine, G. L., Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* **2009,** *462* (7274), 762-6.

469. Fowler, R. G.; White, S. J.; Koyama, C.; Moore, S. C.; Dunn, R. L.; Schaaper, R. M., Interactions among the *Escherichia coli* mutT, mutM, and mutY damage prevention pathways. *DNA Repair (Amst)* **2003,** *2* (2), 159-73.

470. Sasaki, M.; Kurusu, Y., Analysis of spontaneous base substitutions generated in mutator strains of *Bacillus subtilis*. *FEMS Microbiol Lett* **2004,** *234* (1), 37-42.

471. Castellanos-Juarez, F. X.; Alvarez-Alvarez, C.; Yasbin, R. E.; Setlow, B.; Setlow, P.; Pedraza-Reyes, M., YtkD and MutT protect vegetative cells but not spores of *Bacillus subtilis* from oxidative stress. *J Bacteriol* **2006,** *188* (6), 2285-9.

472. Rotman, E.; Kuzminov, A., The mutT defect does not elevate chromosomal fragmentation in *Escherichia coli* because of the surprisingly low levels of MutM/MutY-recognized DNA modifications. *J Bacteriol* **2007,** *189* (19), 6976-88.

473. Cox, E. C., Mutator gene studies in *Escherichia coli*: the *mutT* gene. *Genetics* **1973,** *73*, Suppl 73:67-80.

474. Truglio, J. J.; Croteau, D. L.; Van Houten, B.; Kisker, C., Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev* **2006,** *106* (2), 233-52.

475. Kint, N.; Janoir, C.; Monot, M.; Hoys, S.; Soutourina, O.; Dupuy, B.; Martin-Verstraete, I., The alternative sigma factor sigma(B) plays a crucial role in adaptive strategies of *Clostridium difficile* during gut infection. *Environ Microbiol* **2017,** *19* (5), 1933-1958.

476. Million-Weaver, S.; Samadpour, A. N.; Moreno-Habel, D. A.; Nugent, P.; Brittnacher, M. J.; Weiss, E.; Hayden, H. S.; Miller, S. I.; Liachko, I.; Merrikh, H., An underlying mechanism for the increased mutagenesis of lagging-strand genes in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **2015,** *112* (10), E1096-105.

477. Deaconescu, A. M.; Chambers, A. L.; Smith, A. J.; Nickels, B. E.; Hochschild, A.; Savery, N. J.; Darst, S. A., Structural basis for bacterial transcription-coupled DNA repair. *Cell* **2006,** *124* (3), 507-20.

478. Pybus, C.; Pedraza-Reyes, M.; Ross, C. A.; Martin, H.; Ona, K.; Yasbin, R. E.; Robleto, E., Transcription-associated mutation in *Bacillus subtilis* cells under stress. *J Bacteriol* **2010,** *192* (13), 3321-8.

479. Ramirez-Guadiana, F. H.; Del Carmen Barajas-Ornelas, R.; Ayala-Garcia, V. M.; Yasbin, R. E.; Robleto, E.; Pedraza-Reyes, M., Transcriptional coupling of DNA repair in sporulating *Bacillus subtilis* cells. *Mol Microbiol* **2013,** *90* (5), 1088-99.

480. Ayora, S.; Rojo, F.; Ogasawara, N.; Nakai, S.; Alonso, J. C., The Mfd protein of *Bacillus subtilis* 168 is involved in both transcription-coupled DNA repair and DNA recombination. *J Mol Biol* **1996,** *256* (2), 301-18.

481. Bejerano-Sagie, M.; Oppenheimer-Shaanan, Y.; Berlatzky, I.; Rouvinski, A.; Meyerovich, M.; Ben-Yehuda, S., A checkpoint protein that scans the chromosome for damage at the start of sporulation in *Bacillus subtilis*. *Cell* **2006,** *125* (4), 679-90.

482. Yeeles, J. T.; Dillingham, M. S., The processing of double-stranded DNA breaks for recombinational repair by helicase-nuclease complexes. *DNA Repair (Amst)* **2010,** *9* (3), 276-85.

483. Ayora, S.; Carrasco, B.; Cardenas, P. P.; Cesar, C. E.; Canas, C.; Yadav, T.; Marchisone, C.; Alonso, J. C., Double-strand break repair in bacteria: a view from *Bacillus subtilis*. *FEMS Microbiol Rev* **2011,** *35* (6), 1055-81.

484. Cox, M. M., Relating biochemistry to biology: how the recombinational repair function of RecA protein is manifested in its molecular properties. *Bioessays* **1993,** *15* (9), 617-23.

485. Michod, R. E.; Bernstein, H.; Nedelcu, A. M., Adaptive value of sex in microbial pathogens. *Infect Genet Evol* **2008,** *8* (3), 267-85.

486. Sanchez, H.; Kidane, D.; Reed, P.; Curtis, F. A.; Cozar, M. C.; Graumann, P. L.; Sharples, G. J.; Alonso, J. C., The RuvAB branch migration translocase and RecU Holliday junction resolvase are required for double-stranded DNA break repair in *Bacillus subtilis*. *Genetics* **2005,** *171* (3), 873-83.

487. West, S. C., Processing of recombination intermediates by the RuvABC proteins. *Annu Rev Genet* **1997,** *31*, 213-44.

488. McGregor, N.; Ayora, S.; Sedelnikova, S.; Carrasco, B.; Alonso, J. C.; Thaw, P.; Rafferty, J., The structure of *Bacillus subtilis* RecU Holliday junction resolvase and its role in substrate selection and sequence-specific cleavage. *Structure* **2005,** *13* (9), 1341-51.

489. Carrasco, B.; Cozar, M. C.; Lurz, R.; Alonso, J. C.; Ayora, S., Genetic recombination in *Bacillus subtilis* 168: contribution of Holliday junction processing functions in chromosome segregation. *J Bacteriol* **2004,** *186* (17), 5557-66.

490. Fernandez, S.; Sorokin, A.; Alonso, J. C., Genetic recombination in *Bacillus subtilis* 168: effects of recU and recS mutations on DNA repair and homologous recombination. *J Bacteriol* **1998,** *180* (13), 3405-9.

491. Sanchez, H.; Kidane, D.; Castillo Cozar, M.; Graumann, P. L.; Alonso, J. C., Recruitment of *Bacillus subtilis* RecN to DNA double-strand breaks in the absence of DNA end processing. *J Bacteriol* **2006,** *188* (2), 353-60.

492. Krishnamurthy, M.; Tadesse, S.; Rothmaier, K.; Graumann, P. L., A novel SMC-like protein, SbcE (YhaN), is involved in DNA double-strand break repair and competence in *Bacillus subtilis*. *Nucleic Acids Res* **2010,** *38* (2), 455-66.

493. Alonso, J. C.; Stiege, A. C.; Luder, G., Genetic recombination in *Bacillus subtilis* 168: effect of recN, recF, recH and addAB mutations on DNA repair and recombination. *Mol Gen Genet* **1993,** *239* (1-2), 129-36.

494. Kobayashi, K.; Ehrlich, S. D.; Albertini, A.; Amati, G.; Andersen, K. K.; Arnaud, M.; Asai, K.; Ashikaga, S.; Aymerich, S.; Bessieres, P.; Boland, F.; Brignell, S. C.; Bron, S.; Bunai, K.; Chapuis, J.; Christiansen, L. C.; Danchin, A.; Debarbouille, M.; Dervyn, E.; Deuerling, E.; Devine, K.; Devine, S. K.; Dreesen, O.; Errington, J.; Fillinger, S.; Foster, S. J.; Fujita, Y.; Galizzi, A.; Gardan, R.; Eschevins, C.; Fukushima, T.; Haga, K.; Harwood, C. R.; Hecker, M.; Hosoya, D.; Hullo, M. F.; Kakeshita, H.; Karamata, D.; Kasahara, Y.; Kawamura, F.; Koga, K.; Koski, P.; Kuwana, R.; Imamura, D.; Ishimaru, M.; Ishikawa, S.; Ishio, I.; Le Coq, D.; Masson, A.; Mauel, C.; Meima, R.; Mellado, R. P.; Moir, A.; Moriya, S.; Nagakawa, E.; Nanamiya, H.; Nakai, S.; Nygaard, P.; Ogura, M.; Ohanan, T.; O'Reilly, M.; O'Rourke, M.; Pragai, Z.; Pooley, H. M.; Rapoport, G.; Rawlins, J. P.; Rivas, L. A.; Rivolta, C.; Sadaie, A.; Sadaie, Y.; Sarvas, M.; Sato, T.; Saxild, H. H.; Scanlan, E.; Schumann, W.; Seegers, J. F.; Sekiguchi, J.; Sekowska, A.; Seror, S. J.; Simon, M.; Stragier, P.; Studer, R.; Takamatsu, H.; Tanaka, T.; Takeuchi, M.; Thomaides, H. B.; Vagner, V.; van Dijl, J. M.; Watabe, K.; Wipat, A.; Yamamoto, H.; Yamamoto, M.; Yamamoto, Y.; Yamane, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.; Zuber, U.; Ogasawara, N., Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci U S A* **2003,** *100* (8), 4678-83.

495. Carrasco, B.; Yadav, T.; Serrano, E.; Alonso, J. C., *Bacillus subtilis* RecO and SsbA are crucial for RecA-mediated recombinational DNA repair. *Nucleic Acids Res* **2015,** *43* (12), 5984-97.

496. Bruand, C.; Farache, M.; McGovern, S.; Ehrlich, S. D.; Polard, P., DnaB, DnaD and DnaI proteins are components of the *Bacillus subtilis* replication restart primosome. *Mol Microbiol* **2001,** *42* (1), 245-55.

497. Briggs, G. S.; Smits, W. K.; Soultanas, P., Chromosomal replication initiation machinery of low-G+C-content Firmicutes. *J Bacteriol* **2012,** *194* (19), 5162-70.

498. Sargentini, N. J.; Smith, K. C., Quantitation of the involvement of the *recA*, *recB*, *recC*, *recF*, *recJ*, *recN*, *lexA*, *radA*, *radB*, *uvrD*, and *umuC* genes in the repair of X-ray-induced DNA double-strand breaks in *Escherichia coli*. *Radiat Res* **1986,** *107* (1), 58-72.

499. Bellio, P.; Mancini, A.; Di Pietro, L.; Cracchiolo, S.; Franceschini, N.; Reale, S.; de Angelis, F.; Perilli, M.; Amicosante, G.; Spyrakis, F.; Tondi, D.; Cendron,

L.; Celenza, G., Inhibition of the transcriptional repressor LexA: Withstanding drug resistance by inhibiting the bacterial mechanisms of adaptation to antimicrobials. *Life Sci* **2020,** *241*, 117116.

500. Duigou, S.; Ehrlich, S. D.; Noirot, P.; Noirot-Gros, M. F., Distinctive genetic features exhibited by the Y-family DNA polymerases in *Bacillus subtilis*. *Mol Microbiol* **2004,** *54* (2), 439-51.

501. Rocha, E. P.; Matic, I.; Taddei, F., Over-representation of repeats in stress response genes: a strategy to increase versatility under stressful conditions? *Nucleic Acids Res* **2002,** *30* (9), 1886-94.

502. Labat, F.; Pradillon, O.; Garry, L.; Peuchmaur, M.; Fantin, B.; Denamur, E., Mutator phenotype confers advantage in *Escherichia coli* chronic urinary tract infection pathogenesis. *FEMS Immunol Med Microbiol* **2005,** *44* (3), 317-21.

503. Peng, M.; Liang, Z., Degeneration of industrial bacteria caused by genetic instability. *World J Microbiol Biotechnol* **2020,** *36* (8), 119.

504. Levin, B. R.; Perrot, V.; Walker, N., Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* **2000,** *154* (3), 985-97.

505. Sebaihia, M.; Peck, M. W.; Minton, N. P.; Thomson, N. R.; Holden, M. T.; Mitchell, W. J.; Carter, A. T.; Bentley, S. D.; Mason, D. R.; Crossman, L.; Paul, C. J.; Ivens, A.; Wells-Bennik, M. H.; Davis, I. J.; Cerdeno-Tarraga, A. M.; Churcher, C.; Quail, M. A.; Chillingworth, T.; Feltwell, T.; Fraser, A.; Goodhead, I.; Hance, Z.; Jagels, K.; Larke, N.; Maddison, M.; Moule, S.; Mungall, K.; Norbertczak, H.; Rabbinowitsch, E.; Sanders, M.; Simmonds, M.; White, B.; Whithead, S.; Parkhill, J., Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome Res* **2007,** *17* (7), 1082-92.

506. Zhang, W.; Ma, J.; Zang, C.; Song, Y.; Liu, P., The fur transcription regulator and fur-regulated genes in *Clostridium botulinum* A ATCC 3502. *J Biomed Biotechnol* **2011,** *2011*, 934756.

507. Liang, W. D.; Bi, Y. T.; Wang, H. Y.; Dong, S.; Li, K. S.; Li, J. S., Gene expression profiling of *Clostridium botulinum* under heat shock stress. *Biomed Res Int* **2013,** *2013*, 760904.

508. Kirk, D. G.; Palonen, E.; Korkeala, H.; Lindstrom, M., Evaluation of normalization reference genes for RT-qPCR analysis of spo0A and four sporulation sigma factor genes in *Clostridium botulinum* Group I strain ATCC 3502. *Anaerobe* **2014,** *26*, 14-9.

509. Dahlsten, E.; Isokallio, M.; Somervuo, P.; Lindstrom, M.; Korkeala, H., Transcriptomic analysis of (group I) *Clostridium botulinum* ATCC 3502 cold shock response. *PLoS One* **2014,** *9* (2), e89958.

510. Alberto, F.; Broussolle, V.; Mason, D. R.; Carlin, F.; Peck, M. W., Variability in spore germination response by strains of proteolytic *Clostridium botulinum* types A, B and F. *Lett Appl Microbiol* **2003,** *36* (1), 41-5.

511. Artin, I.; Mason, D. R.; Pin, C.; Schelin, J.; Peck, M. W.; Holst, E.; Radstrom, P.; Carter, A. T., Effects of carbon dioxide on growth of proteolytic *Clostridium*

botulinum, its ability to produce neurotoxin, and its transcriptome. *Appl Environ Microbiol* **2010,** *76* (4), 1168-72.

512. Selby, K.; Lindstrom, M.; Somervuo, P.; Heap, J. T.; Minton, N. P.; Korkeala, H., Important role of class I heat shock genes hrcA and dnaK in the heat shock response and the response to pH and NaCl stress of group I *Clostridium botulinum* strain ATCC 3502. *Appl Environ Microbiol* **2011,** *77* (9), 2823-30.

513. Lindstrom, M.; Dahlsten, E.; Soderholm, H.; Selby, K.; Somervuo, P.; Heap, J. T.; Minton, N. P.; Korkeala, H., Involvement of two-component system CBO0366/CBO0365 in the cold shock response and growth of group I (proteolytic) *Clostridium botulinum* ATCC 3502 at low temperatures. *Appl Environ Microbiol* **2012,** *78* (15), 5466-70.

514. Dahlsten, E.; Kirk, D.; Lindstrom, M.; Korkeala, H., Alternative sigma factor SigK has a role in stress tolerance of group I *Clostridium botulinum* strain ATCC 3502. *Appl Environ Microbiol* **2013,** *79* (12), 3867-9.

515. Dahlsten, E.; Zhang, Z.; Somervuo, P.; Minton, N. P.; Lindstrom, M.; Korkeala, H., The cold-induced two-component system CBO0366/CBO0365 regulates metabolic pathways with novel roles in group I *Clostridium botulinum* ATCC 3502 cold tolerance. *Appl Environ Microbiol* **2014,** *80* (1), 306-19.

516. Poole, K., Bacterial stress responses as determinants of antimicrobial resistance. *J Antimicrob Chemother* **2012,** *67* (9), 2069-89.

517. Saint-Ruf, C.; Pesut, J.; Sopta, M.; Matic, I., Causes and consequences of DNA repair activity modulation during stationary phase in *Escherichia coli*. *Crit Rev Biochem Mol Biol* **2007,** *42* (4), 259-70.

518. Setlow, P., Spore germination. *Curr Opin Microbiol* **2003,** *6* (6), 550-6.

519. Broussolle, V.; Alberto, F.; Shearman, C. A.; Mason, D. R.; Botella, L.; Nguyen-The, C.; Peck, M. W.; Carlin, F., Molecular and Physiological Characterisation of Spore Germination in *Clostridium botulinum* and *C. sporogenes*. *Anaerobe* **2002,** *8* (3), 89-100.

520. Alberto, F.; Botella, L.; Carlin, F.; Nguyen-The, C.; Broussolle, V., The *Clostridium botulinum* GerAB germination protein is located in the inner membrane of spores. *FEMS Microbiol Lett* **2005,** *253* (2), 231-5.

521. Ramirez-Guadiana, F. H.; Meeske, A. J.; Wang, X.; Rodrigues, C. D. A.; Rudner, D. Z., The *Bacillus subtilis* germinant receptor GerA triggers premature germination in response to morphological defects during sporulation. *Mol Microbiol* **2017,** *105* (5), 689-704.

522. Feucht, A.; Evans, L.; Errington, J., Identification of sporulation genes by genome-wide analysis of the sigmaE regulon of *Bacillus subtilis*. *Microbiology (Reading)* **2003,** *149* (Pt 10), 3023-3034.

523. Plotkin, J. B.; Kudla, G., Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **2011,** *12* (1), 32-42.

524. Hanson, G.; Coller, J., Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* **2018,** *19* (1), 20-30.

525. Serrano, M.; Hovel, S.; Moran, C. P., Jr.; Henriques, A. O.; Volker, U., Forespore-specific transcription of the *lonB* gene during sporulation in *Bacillus subtilis*. *J Bacteriol* **2001,** *183* (10), 2995-3003.

526. Tsilibaris, V.; Maenhaut-Michel, G.; Van Melderen, L., Biological roles of the Lon ATP-dependent protease. *Res Microbiol* **2006,** *157* (8), 701-13.

527. Ribis, J. W.; Fimlaid, K. A.; Shen, A., Differential requirements for conserved peptidoglycan remodeling enzymes during *Clostridioides difficile* spore formation. *Mol Microbiol* **2018,** *110* (3), 370-389.

528. Huang, H.; Chai, C.; Li, N.; Rowe, P.; Minton, N. P.; Yang, S.; Jiang, W.; Gu, Y., CRISPR/Cas9-Based Efficient Genome Editing in *Clostridium ljungdahlii*, an Autotrophic Gas-Fermenting Bacterium. *ACS Synth Biol* **2016,** *5* (12), 1355-1361.

529. Clauwers, C.; Vanoirbeek, K.; Delbrassinne, L.; Michiels, C. W., Construction of Nontoxigenic Mutants of Nonproteolytic *Clostridium botulinum* NCTC 11219 by Insertional Mutagenesis and Gene Replacement. *Appl Environ Microbiol* **2016,** *82* (10), 3100-3108.

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

531. Saujet, L.; Pereira, F. C.; Serrano, M.; Soutourina, O.; Monot, M.; Shelyakin, P. V.; Gelfand, M. S.; Dupuy, B.; Henriques, A. O.; Martin-Verstraete, I., Genome-wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. *PLoS Genet* **2013,** *9* (10), e1003756.

532. Andreev, J.; Dibrov, P. A.; Klein, D.; Braun, S., Chemotaxis, sporulation, and larvicide production in *Bacillus sphaericus* 2362. The influence of L-ethionine, and of aminophenylboronic acid. *FEBS Lett* **1994,** *349* (3), 416-9.

533. Lee, K. Y.; Juang, T. C.; Lee, K. C., Effect of metal ions on growth and sporulation of *Clostridium perfringens* in a synthetic medium. *Zhonghua Min Guo Wei Sheng Wu Xue Za Zhi* **1978,** *11* (2), 50-61.

534. Jee, J.; Rasouly, A.; Shamovsky, I.; Akivis, Y.; Steinman, S. R.; Mishra, B.; Nudler, E., Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature* **2016,** *534* (7609), 693-6.

535. Robert, L.; Ollion, J.; Robert, J.; Song, X.; Matic, I.; Elez, M., Mutation dynamics and fitness effects followed in single cells. *Science* **2018,** *359* (6381), 1283-1286.

536. Lee, H.; Popodi, E.; Tang, H.; Foster, P. L., Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A* **2012,** *109* (41), E2774-83.

537. Krasovec, R.; Richards, H.; Gomez, G.; Gifford, D. R.; Mazoyer, A.; Knight, C. G., Measuring Microbial Mutation Rates with the Fluctuation Assay. *J Vis Exp* **2019,** (153).

538. Foster, P. L., Methods for determining spontaneous mutation rates. *Methods Enzymol* **2006,** *409*, 195-213.

539. Campbell, E. A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S. A., Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **2001,** *104* (6), 901-12.

540. Horst, J. P.; Wu, T. H.; Marinus, M. G., *Escherichia coli* mutator genes. *Trends Microbiol* **1999,** *7* (1), 29-36.

541. Fernandez, S.; Kobayashi, Y.; Ogasawara, N.; Alonso, J. C., Analysis of the *Bacillus subtilis recO* gene: RecO forms part of the RecFLOR function. *Mol Gen Genet* **1999,** *261* (3), 567-73.

542. Iyer, R. R.; Pluciennik, A.; Burdett, V.; Modrich, P. L., DNA mismatch repair: functions and mechanisms. *Chem Rev* **2006,** *106* (2), 302-23.

543. Richardson, A. R.; Yu, Z.; Popovic, T.; Stojiljkovic, I., Mutator clones of *Neisseria meningitidis* in epidemic serogroup A disease. *Proc Natl Acad Sci U S A* **2002,** *99* (9), 6103-7.

544. Matsushima, A.; Takakura, S.; Fujihara, N.; Saito, T.; Ito, I.; Iinuma, Y.; Ichiyama, S., High prevalence of mutators among *Enterobacter cloacae* nosocomial isolates and their association with antimicrobial resistance and repetitive detection. *Clin Microbiol Infect* **2010,** *16* (9), 1488-93.

545. Gentile, C. F.; Yu, S. C.; Serrano, S. A.; Gerrish, P. J.; Sniegowski, P. D., Competition between high- and higher-mutating strains of *Escherichia coli*. *Biol Lett* **2011,** *7* (3), 422-4.

546. Mao, E. F.; Lane, L.; Lee, J.; Miller, J. H., Proliferation of mutators in A cell population. *J Bacteriol* **1997,** *179* (2), 417-22.

547. Elena, S. F.; Lenski, R. E., Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* **2003,** *4* (6), 457-69.

548. Couce, A.; Guelfo, J. R.; Blazquez, J., Mutational spectrum drives the rise of mutator bacteria. *PLoS Genet* **2013,** *9* (1), e1003167.

549. Wiser, M. J.; Lenski, R. E., A Comparison of Methods to Measure Fitness in *Escherichia coli*. *PLoS One* **2015,** *10* (5), e0126210.

550. Sleight, S. C.; Lenski, R. E., Evolutionary adaptation to freeze-thaw-growth cycles in *Escherichia coli*. *Physiol Biochem Zool* **2007,** *80* (4), 370-85.

551. Bell, G., Evolutionary rescue and the limits of adaptation. *Philos Trans R Soc Lond B Biol Sci* **2013,** *368* (1610), 20120080.

552. Denamur, E.; Matic, I., Evolution of mutation rates in bacteria. *Mol Microbiol* **2006,** *60* (4), 820-7.

553. Balasingam, N.; Brandon, H. E.; Ross, J. A.; Wieden, H. J.; Thakor, N., Cellular roles of the human Obg-like ATPase 1 (hOLA1) and its YchF homologs. *Biochem Cell Biol* **2020,** *98* (1), 1-11.

554. Teplyakov, A.; Obmolova, G.; Chu, S. Y.; Toedt, J.; Eisenstein, E.; Howard, A. J.; Gilliland, G. L., Crystal structure of the YchF protein reveals binding sites for GTP and nucleic acid. *J Bacteriol* **2003,** *185* (14), 4031-7.

555. Fernebro, J.; Blomberg, C.; Morfeldt, E.; Wolf-Watz, H.; Normark, S.; Normark, B. H., The influence of in vitro fitness defects on pneumococcal ability to colonize and to cause invasive disease. *BMC Microbiol* **2008,** *8*, 65.

556. Eraso, J. M.; Markillie, L. M.; Mitchell, H. D.; Taylor, R. C.; Orr, G.; Margolin, W., The highly conserved MraZ protein is a transcriptional regulator in *Escherichia coli*. *J Bacteriol* **2014,** *196* (11), 2053-66.

557. Lang, G. I.; Rice, D. P.; Hickman, M. J.; Sodergren, E.; Weinstock, G. M.; Botstein, D.; Desai, M. M., Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **2013,** *500* (7464), 571-4.

558. de Visser, J. A.; Rozen, D. E., Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics* **2006,** *172* (4), 2093-100.

559. Blount, Z. D.; Barrick, J. E.; Davidson, C. J.; Lenski, R. E., Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **2012,** *489* (7417), 513-8.

560. Good, B. H.; Desai, M. M., Evolution of Mutation Rates in Rapidly Adapting Asexual Populations. *Genetics* **2016,** *204* (3), 1249-1266.

561. Maddamsetti, R.; Lenski, R. E.; Barrick, J. E., Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with *Escherichia coli*. *Genetics* **2015,** *200* (2), 619-31.

562. Katz, S.; Avrani, S.; Yavneh, M.; Hilau, S.; Gross, J.; Hershberg, R., Dynamics of adaptation during three years of evolution under long-term stationary phase. *Mol Biol Evol* **2021**.

563. Cvijovic, I.; Nguyen Ba, A. N.; Desai, M. M., Experimental Studies of Evolutionary Dynamics in Microbes. *Trends Genet* **2018,** *34* (9), 693-703.

564. McElroy, K.; Thomas, T.; Luciani, F., Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions. *Microb Inform Exp* **2014,** *4* (1), 1.

565. Zhang, Z.; Korkeala, H.; Dahlsten, E.; Sahala, E.; Heap, J. T.; Minton, N. P.; Lindstrom, M., Two-component signal transduction system CBO0787/CBO0786 represses transcription from botulinum neurotoxin promoters in *Clostridium botulinum* ATCC 3502. *PLoS Pathog* **2013,** *9* (3), e1003252.

566. Raynes, Y.; Sniegowski, P. D.; Weinreich, D. M., Migration promotes mutator alleles in subdivided populations. *Evolution* **2019,** *73* (3), 600-608.

567. Erickson, M. C.; Ma, L. M.; Doyle, M. P., Clostridium botulinum Toxin Production in Relation to Spoilage of Atlantic Salmon (Salmo salar) Packaged in Films of Varying Oxygen Permeabilities and with Different Atmospheres. *J Food Prot* **2015,** *78* (11), 2006-18.

568. Newell, C. R.; Doyle, M.; Ma, L., Inability of non-proteolytic *Clostridium botulinum* to grow in mussels inoculated via immersion and packaged in high oxygen atmospheres. *Food Microbiol* **2015,** *46*, 204-209.

569. Ma, L.; Zhang, G.; Sobel, J.; Doyle, M. P., Evaluation of the effect of acetylsalicylic acid on *Clostridium botulinum* growth and toxin production. *J Food Prot* **2007,** *70* (12), 2860-3.

570. Clavero, M. R. S.; Brackett, R. E.; Beuchat, L. R.; Doyle, M. P., Influence of water activity and storage conditions on survival and growth of proteolytic

Clostridium botulinum in peanut spread. *Food Microbiology* **2000,** *17* (1), 53-61.

571. Doyle, M. P., Evaluating the Potential Risk from Extended-Shelf-Life Refrigerated Foods by *Clostridium Botulinum* Inoculation Studies. *Food Technol. Chicago* **1991,** *45* (4), 154-&.

572. Glass, K. A.; Doyle, M. P., Relationship between Water Activity of Fresh Pasta and Toxin Production by Proteolytic *Clostridium Botulinum*. *J Food Protect* **1991,** *54* (3), 162-165.

573. Brown, J. L.; Tran-Dinh, N.; Chapman, B., *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*. *J Food Prot* **2012,** *75* (4), 779-92.

574. Janganan, T. K.; Mullin, N.; Tzokov, S. B.; Stringer, S.; Fagan, R. P.; Hobbs, J. K.; Moir, A.; Bullough, P. A., Characterization of the spore surface and exosporium proteins of *Clostridium sporogenes*; implications for *Clostridium botulinum* group I strains. *Food Microbiol* **2016,** *59*, 205-12.

575. Bull, M. K.; Olivier, S. A.; van Diepenbeek, R. J.; Kormelink, F.; Chapman, B., Synergistic inactivation of spores of proteolytic *Clostridium botulinum* strains by high pressure and heat is strain and product dependent. *Appl Environ Microbiol* **2009,** *75* (2), 434-45.

576. Johnson, E. A.; Lin, W. J.; Zhou, Y. T.; Bradshaw, M., Characterization of neurotoxin mutants in *Clostridium botulinum* type A. *Clin Infect Dis* **1997,** *25 Suppl 2*, S168-70.

577. Binz, T.; Bade, S.; Rummel, A.; Kollewe, A.; Alves, J., Arg(362) and Tyr(365) of the botulinum neurotoxin type a light chain are involved in transition state stabilization. *Biochemistry* **2002,** *41* (6), 1717-23.

578. Breidenbach, M. A.; Brunger, A. T., Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* **2004,** *432* (7019), 925-9.

579. Gu, S.; Jin, R., Assembly and function of the botulinum neurotoxin progenitor complex. *Curr Top Microbiol Immunol* **2013,** *364*, 21-44.

580. Deatherage Kaiser, B. L.; Hill, K. K.; Smith, T. J.; Williamson, C. H. D.; Keim, P.; Sahl, J. W.; Wahl, K. L., Proteomic analysis of four *Clostridium botulinum* strains identifies proteins that link biological responses to proteomic signatures. *PLoS One* **2018,** *13* (10), e0205586.

581. Cooper, L. E.; O'Leary, S. E.; Begley, T. P., Biosynthesis of a thiamin antivitamin in *Clostridium botulinum*. *Biochemistry* **2014,** *53* (14), 2215-7.

582. Fang, P. K.; Raphael, B. H.; Maslanka, S. E.; Cai, S.; Singh, B. R., Analysis of genomic differences among *Clostridium botulinum* type A1 strains. *BMC Genomics* **2010,** *11*, 725.

583. Ng, V.; Lin, W. J., Comparison of assembled *Clostridium botulinum* A1 genomes revealed their evolutionary relationship. *Genomics* **2014,** *103* (1), 94-106.

584. Shin, N. R.; Shin, J. H.; Chun, J. H.; Yoon, S. Y.; Kim, B. S.; Oh, H. B.; Rhie, G. E., Determination of neurotoxin gene expression in *Clostridium botulinum* type A by quantitative RT-PCR. *Mol Cells* **2006,** *22* (3), 336-42.

585. Singh, A. K.; Brown, E. W.; Sharma, S. K., Comparative sequence analyses of the neurotoxin complex genes in *Clostridium botulinum* serotypes A, B, E, and F. *Microbiology Research* **2012,** *3* (2).

586. Bradshaw, M.; Goodnough, M. C.; Johnson, E. A., Conjugative transfer of the *Escherichia coli-Clostridium perfringens* shuttle vector pJIR1457 to *Clostridium botulinum* type A strains. *Plasmid* **1998,** *40* (3), 233-7.

587. Billon, C. M.; McKirgan, C. J.; McClure, P. J.; Adair, C., The effect of temperature on the germination of single spores of *Clostridium botulinum* 62A. *Journal of applied microbiology* **1997,** *82* (1), 48-56.

588. Bradshaw, M.; Dineen, S. S.; Maks, N. D.; Johnson, E. A., Regulation of neurotoxin complex expression in *Clostridium botulinum* strains 62A, Hall A-hyper, and NCTC 2916. *Anaerobe* **2004,** *10* (6), 321-33.

589. Reddy, N. R.; Marshall, K. M.; Morrissey, T. R.; Loeza, V.; Patazca, E.; Skinner, G. E.; Krishnamurthy, K.; Larkin, J. W., Combined high pressure and thermal processing on inactivation of type A and proteolytic type B spores of *Clostridium botulinum*. *J Food Prot* **2013,** *76* (8), 1384-92.

590. Leighton, G., Botulism and food preservation (The Loch Maree Tragedy). *Nature* **1923,** *111*, 737.

591. Henderson, I.; Whelan, S. M.; Davis, T. O.; Minton, N. P., Genetic characterisation of the botulinum toxin complex of *Clostridium botulinum* strain NCTC 2916. *FEMS Microbiol Lett* **1996,** *140* (2-3), 151-8.

592. Berendsen, E. M.; Boekhorst, J.; Kuipers, O. P.; Wells-Bennik, M. H., A mobile genetic element profoundly increases heat resistance of bacterial spores. *ISME J* **2016,** *10* (11), 2633-2642.

593. Berendsen, E. M.; Koning, R. A.; Boekhorst, J.; de Jong, A.; Kuipers, O. P.; Wells-Bennik, M. H., High-Level Heat Resistance of Spores of *Bacillus amyloliquefaciens* and *Bacillus licheniformis* Results from the Presence of a *spoVA* Operon in a Tn1546 Transposon. *Front Microbiol* **2016,** *7*, 1912.

594. Butler, R. R., 3rd; Schill, K. M.; Wang, Y.; Pombert, J. F., Genetic Characterization of the Exceptionally High Heat Resistance of the Non-toxic Surrogate *Clostridium sporogenes* PA 3679. *Front Microbiol* **2017,** *8*, 545.

595. Sommer, F. W., Heat Resistaxce of the Spores of *Clostridium Botulinum*. *The Journal of Infectious Diseases* **1930,** *46* (2), 85-114.

596. Clauwers, C.; Briers, Y.; Lavigne, R.; Michiels, C. W., Two Complete and One Draft Genome Sequence of Nonproteolytic *Clostridium botulinum* Type E Strains NCTC 8266, NCTC 8550, and NCTC 11219. *Genome Announc* **2015,** *3* (2).

597. M. O'Mahony, E. M., R. J. Gilbert, D. N. Hutchinson, N. T. Begg, J. C. Rodhouse, and J. E. Morris, An outbreak of foodborne botulism associated with contaminated hazelnut yoghurt. *Epidemiol Infect.* **1990,** *104* (3), 389-395.

598. Overbeek, R.; Olson, R.; Pusch, G. D.; Olsen, G. J.; Davis, J. J.; Disz, T.; Edwards, R. A.; Gerdes, S.; Parrello, B.; Shukla, M.; Vonstein, V.; Wattam, A. R.; Xia, F.; Stevens, R., The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **2014,** *42*, D206-14.

599. Rosche, W. A.; Foster, P. L., Determining mutation rates in bacterial populations. *Methods* **2000,** *20* (1), 4-17.

600. Brandis, G.; Pietsch, F.; Alemayehu, R.; Hughes, D., Comprehensive phenotypic characterization of rifampicin resistance mutations in *Salmonella* provides insight into the evolution of resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* **2015,** *70* (3), 680-5.

601. Jin, D. J.; Gross, C. A., Mapping and sequencing of mutations in the *Escherichia coli rpoB* gene that lead to rifampicin resistance. *J Mol Biol* **1988,** *202* (1), 45-58.

602. Morlock, G. P.; Plikaytis, B. B.; Crawford, J. T., Characterization of spontaneous, In vitro-selected, rifampin-resistant mutants of *Mycobacterium tuberculosis* strain H37Rv. *Antimicrob Agents Chemother* **2000,** *44* (12), 3298-301.

603. Reddy, N. R.; Solomon, H. M.; Tetzloff, R. C.; Rhodehamel, E. J., Inactivation of *Clostridium botulinum* type A spores by high-pressure processing at elevated temperatures. *J Food Prot* **2003,** *66* (8), 1402-7.

604. Diao, M. M.; Andre, S.; Membre, J. M., Meta-analysis of D-values of proteolytic *Clostridium botulinum* and its Surrogate strain *Clostridium sporogenes* PA 3679. *International Journal of Food Microbiology* **2014,** *174*, 23-30.

605. Berendsen, E. M.; Zwietering, M. H.; Kuipers, O. P.; Wells-Bennik, M. H., Two distinct groups within the *Bacillus subtilis* group display significantly different spore heat resistance properties. *Food Microbiol* **2015,** *45* (Pt A), 18-25.

606. Wang, Z.; Li, P.; Luo, L.; Simpson, D. J.; Ganzle, M. G., Daqu Fermentation Selects for Heat-Resistant *Enterobacteriaceae* and *Bacilli*. *Appl Environ Microbiol* **2018,** *84* (21).

607. Paredes-Sabja, D.; Setlow, B.; Setlow, P.; Sarker, M. R., Characterization of *Clostridium perfringens* spores that lack SpoVA proteins and dipicolinic acid. *J Bacteriol* **2008,** *190* (13), 4648-59.

608. Li, Y.; Davis, A.; Korza, G.; Zhang, P.; Li, Y. Q.; Setlow, B.; Setlow, P.; Hao, B., Role of a SpoVA protein in dipicolinic acid uptake into developing spores of *Bacillus subtilis*. *J Bacteriol* **2012,** *194* (8), 1875-84.

609. Setlow, B.; Atluri, S.; Kitchel, R.; Koziol-Dube, K.; Setlow, P., Role of dipicolinic acid in resistance and stability of spores of *Bacillus subtilis* with or without DNA-protective alpha/beta-type small acid-soluble proteins. *J Bacteriol* **2006,** *188* (11), 3740-7.

610. Magge, A.; Granger, A. C.; Wahome, P. G.; Setlow, B.; Vepachedu, V. R.; Loshon, C. A.; Peng, L.; Chen, D.; Li, Y. Q.; Setlow, P., Role of dipicolinic acid in the germination, stability, and viability of spores of *Bacillus subtilis*. *J Bacteriol* **2008,** *190* (14), 4798-807.

611. Luu, S.; Cruz-Mora, J.; Setlow, B.; Feeherry, F. E.; Doona, C. J.; Setlow, P., The effects of heat activation on *Bacillus* spore germination, with nutrients or under high pressure, with or without various germination proteins. *Appl Environ Microbiol* **2015,** *81* (8), 2927-38.

612. Montville, T. J., Effect of plating medium on heat activation requirement of *Clostridium botulinum* spores. *Appl Environ Microbiol* **1981,** *42* (4), 734-6.

613. Kort, R.; O'Brien, A. C.; van Stokkum, I. H.; Oomes, S. J.; Crielaard, W.; Hellingwerf, K. J.; Brul, S., Assessment of heat resistance of bacterial spores from food product isolates by fluorescence monitoring of dipicolinic acid release. *Appl Environ Microbiol* **2005,** *71* (7), 3556-64.

614. Langridge, G. C.; Phan, M. D.; Turner, D. J.; Perkins, T. T.; Parts, L.; Haase, J.; Charles, I.; Maskell, D. J.; Peters, S. E.; Dougan, G.; Wain, J.; Parkhill, J.; Turner, A. K., Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res* **2009,** *19* (12), 2308-16.

615. Bressuire-Isoard, C.; Broussolle, V.; Carlin, F., Sporulation environment influences spore properties in *Bacillus*: evidence and insights on underlying molecular and physiological mechanisms. *FEMS Microbiol Rev* **2018,** *42* (5), 614-626.

616. Zhao, Y.; Kang, L.; Gao, S.; Gao, X.; Xin, W.; Wang, J., PEG precipitation coupled with chromatography is a new and sufficient method for the purification of botulinum neurotoxin type B [corrected]. *PLoS One* **2012,** *7* (6), e39670.

617. Pier, C. L.; Tepp, W. H.; Bradshaw, M.; Johnson, E. A.; Barbieri, J. T.; Baldwin, M. R., Recombinant holotoxoid vaccine against botulism. *Infect Immun* **2008,** *76* (1), 437-42.

618. Seys, F. M.; Rowe, P.; Bolt, E. L.; Humphreys, C. M.; Minton, N. P., A Gold Standard, CRISPR/Cas9-Based Complementation Strategy Reliant on 24 Nucleotide Bookmark Sequences. *Genes (Basel)* **2020,** *11* (4).

619. Przedpelski, A.; Tepp, W. H.; Zuverink, M.; Johnson, E. A.; Pellet, S.; Barbieri, J. T., Enhancing toxin-based vaccines against botulism. *Vaccine* **2018,** *36* (6), 827-832.

620. Thirunavukkarasu, N.; Johnson, E.; Pillai, S.; Hodge, D.; Stanker, L.; Wentz, T.; Singh, B.; Venkateswaran, K.; McNutt, P.; Adler, M.; Brown, E.; Hammack, T.; Burr, D.; Sharma, S., Botulinum Neurotoxin Detection Methods for Public Health Response and Surveillance. *Front Bioeng Biotechnol* **2018,** *6*, 80.

621. Sharma, S. K.; Whiting, R. C., Methods for detection of *Clostridium botulinum* toxin in foods. *J Food Prot* **2005,** *68* (6), 1256-63.

622. S. Crouch, M. M., A Hodson, F. Kemm, J. Gausden, M. Johnson, S. Rahman, Y. Tengur, D. Vaughan, J. Abbott, R. Griffin, N. Minton NoTox.

623. Streett, H. E.; Kalis, K. M.; Papoutsakis, E. T., A Strongly Fluorescing Anaerobic Reporter and Protein-Tagging System for *Clostridium* Organisms Based on the Fluorescence-Activating and Absorption-Shifting Tag Protein (FAST). *Appl Environ Microbiol* **2019,** *85* (14).

APPENDIX

List of Appendix Figures

Figure A 1 Phenotypic characterisation of ATCC 19397 Δ*boNT/A1* mutants. 299

List of Appendix Tables

Table A 1 List of *E. coli* strains used in this study 284

Table A 2 List of *C. botulinum* strains used in this study.....285

Table A 3 List of plasmids used in this study 287

Table A 4 List of primers used in this study 291

Table A 5 List of variants detected in ATCC 3502$^{+/-}$ (Unilever; green SBRC culture collection stock) compared to the published reference genome (accession number: NC_009495).....301

Table A 6 List of variants detected in ATCC 3502$^{+/-}$ (Unilever; red SBRC culture collection stock) compared to the published reference genome (accession number: NC_009495).....302

Table A 7 List of variants detected in *hsdR*1600s::CT 1 compared to the published reference genome (accession number: NC_009495) 303

Table A 8 List of variants detected in *hsdR*1600s::CT 2 compared to the published reference genome (accession number: NC_009495) 306

Table A 9 List of variants detected in *hsdR*1600s::CT 3 compared to the published reference genome (accession number: NC_009495) 308

Table A 10 List of variants detected in *pyrE*$^{-}$ 1 compared to the published reference genome (accession number: NC_009495) 310

Table A 11 List of variants detected in *pyrE*$^{-}$ 2 compared to the published reference genome (accession number: NC_009495) 313

Table A 12 List of variants detected in *pyrE*$^{-}$ 3 compared to the published reference genome (accession number: NC_009495) 317

Table A 13 List of variants detected in Δ*boNT/A1* 1 compared to the published reference genome (accession number: NC_009495) 321

Table A 14 List of variants detected in Δ*boNT/A1* 2 compared to the published reference genome (accession number: NC_009495) 324

Table A 15 List of variants detected in Δ*boNT/A1* 3 compared to the published reference genome (accession number: NC_009495) 327

Table A 16 List of variants detected in ATCC 19397 WT compared to the published reference genome (accession number: NC_009697) 329

Table A 17 List of variants detected in 62A WT compared to the published reference genome (accession number: NZ_CP028859).....330

Table A 18 List of variants detected in Loch Maree WT compared to the published reference genome (accession number: NC_010520.1) 341

Table A 19 List of variants detected in NCTC 2916 WT compared to the published reference genome (accession number: NZ_ABDO00000000.2) 343

Table A 20 List of variants detected in ATCC 19397 Δ*boNT/A1* compared to the published reference genome (accession number: NC_009697).....344

Table A 21 List of variants detected in ATCC 19397 ΔClA compared to the published reference genome (accession number: NC_009697).....346

Table A 22 List of variants detected in ATCC 19397 i*boNT/A1*^{RY} compared to the published reference genome (accession number: NC_009697).....348

Table A 23 List of all variants detected in ATCC 3502 strains used in this study ..350

I. Bacterial strains

Table A 1 List of *E. coli* strains used in this study

Strain	Description	Source
TOP10	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lac*X74 *nupG recA1 araD139* Δ(*ara-leu*)7697 *galE15 galK16 rpsL*(Str^r) *endA1* λ$^-$	Invitrogen
CA434	Kan^R; *E. coli* HB101 harbouring R702 (F$^-$ *mcrB mrr hsdS20*(rB$^-$ mB$^-$) *recA13 leuB6 ara*-14 *proA2 lacY1 galK2 xyl*-5 *mtl*-1 *rpsL20*(*Smr*) *glnV44*λ$^-$)	Williams *et al*., 1990
CA434 pM-JTI	Apr^R; CA434 harbouring plasmid pM-JTI	this study
CA434 pM-JTII	Apr^R; CA434 harbouring plasmid pM-JTII	this study
CA434 pM-JTI+TII	Apr^R; CA434 harbouring plasmid pM-JTI+TII	this study
CA434 pM-BTI	Apr^R; CA434 harbouring plasmid pM-BTI	this study
CA434 pM-BTII	Apr^R; CA434 harbouring plasmid pM-BTII	this study
RE-CA434	Cm^R; pKD46 lambda red recombineering plasmid	Rodrigues, 2018 (unpublished)
CA434 R702TI+TII	Kan^R/Apr^R; CA434 harbouring the recombineered R factor R702TI+TII	this study
Stellar	F$^–$, *endA1*, *supE44*, *thi*-1, *recA1*, *relA1*, *gyrA96*, *phoA*, Φ80d *lacZ*Δ M15, Δ(*lacZYA-argF*) U169, Δ(*mrr-hsdRMS-mcrBC*), Δ*mcrA*, λ$^–$	Clontech
InterStellar	Stellar harbouring R factor R702	Millard, 2016 (unpublished)
InterStellar pM-JTI+TII	Apr^R; InterStellar harbouring pM-JTI+TII	this study
RE-InterStellar	Cm^R; InterStellar harbouring pKD46 lambda red recombineering plasmid	Millard, 2017 (unpublished)
InterStellar R702apr	Kan^R/Apr^R; InterStellar harbouring the recombineered R factor R702apr	Mansfield, 2018 (unpublished)
InterStellar R702TI+TII	Kan^R/Apr^R; InterStellar harbouring the recombineered R factor R702TI+TII	this study
HB101	(F$^-$ *mcrB mrr hsdS20*(rB$^-$ mB$^-$) *recA13 leuB6 ara*-14 *proA2 lacY1 galK2 xyl*-5 *mtl*-1 *rpsL20*(*Smr*) *glnV44*λ$^-$)	unknown
HB101 R702	Kan^R: HB101 harbouring R factor R702	this study
HB101 R702TI+TII	Kan^R/Apr^R; HB101 harbouring the recombineered R factor R702TI+TII	this study

sExpress	Kan^{R}; *E. coli* NEB Express harbouring R702 (*fhuA2* [*lon*] *ompT gal sulA11 R*(*mcr*-73::miniTn10-*TetS*)2 [*dcm*] *R*(*zgb*-210::Tn10-*TetS*) *endA1* Δ(*mcrC-mrr*)114::IS10)	Woods *et al*., 2019
LE392	*hsdR514*(rk^{-} mk^{+}) *glnV*(*supE44*) *tryT*(*supF58*) *lacY1* or Δ(*lacIZY*)6 *galK2 galT22 metB1 trpR55*	Promega

Table A 2 List of *C. botulinum* strains used in this study

Strain	Description	Source
General		
C. botulinum ATCC 3502$^{+/-}$ (Unilever)	wildtype isolate (mixed population)	Unilever UK
C. botulinum ATCC 3502^{+}	non-mutator isolate from ATCC 3502$^{+/-}$ (Unilever)	this study
C. botulinum ATCC 3502^{-}	mutator isolate from ATCC 3502$^{+/-}$ (Unilever)	this study
C. botulinum ATCC 3502 (IFR)	wildtype isolate	Peck, IFR (Quadram Institute)
C. botulinum NCTC 13319	wildtype isolate	NCTC
C. botulinum ATCC 19397	wildtype isolate	NCTC
C. botulinum 62A	wildtype isolate	NCTC
C. botulinum Loch Maree	wildtype isolate	NCTC
C. botulinum NCTC 2916	wildtype isolate	NCTC
Chapter 3		
C. botulinum ATCC 3502 Unilever *hsdR*1600s::CT	ClosTron inserted in *hsdR* gene; Erm^{R}	this study
Chapter 4		
C. botulinum ATCC 3502 (IFR) *pyrE*-	*pyrE* deficient mutant via ACE	Humphreys, 2013 (unpublished)
C. botulinum ATCC 3502 (IFR) Δ*boNT/A1 pyrE*-	*boNT/A1* deletion in *pyrE* deficient strain via ACE	Humphreys, 2013 (unpublished)
C. botulinum ATCC 3502 (IFR) Δ*boNT/A1*	*boNT/A1* deletion in *pyrE* restored strain via ACE	Humphreys, 2013 (unpublished)
C. botulinum ATCC 3502 (IFR) ΔClA *pyrE*-	botulinum neurotoxin gene cluster deletion in *pyrE* deficient strain via ACE	Humphreys, 2013 (unpublished)
C. botulinum ATCC 3502 (Unilever) *pyrE*-	*pyrE* deficient mutant via ACE	this study

C. botulinum ATCC 3502 (Unilever) Δ*boNT/A1*	*boNT/A1* deletion via CRISPR/Cas9	this study
C. botulinum ATCC 19397 Δ*mutL*231	231 bp deletion in *mutL* via CRISPR/Cas9	this study
C. botulinum ATCC 19397 Δ*mutL*231^{comp}	complementation of 231 bp deletion in *mutL* via CRISPR/Cas9	this study
C. botulinum ATCC 19397 Δ*mutL*	*mutL* deletion via CRISPR/Cas9	this study
C. botulinum ATCC 19397 Δ*mutL*^{comp}	complementation of *mutL* deletion via CRISPR/Cas9	this study
C. botulinum ATCC 3502 (IFR) *agrD1*::CT	ClosTron inserted in *agrD1* gene; Erm^{R}	Cooksley *et al*., 2010
C. botulinum ATCC 3502 (IFR) *agrD2*::CT	ClosTron inserted in *agrD2* gene; Erm^{R}	Cooksley *et al*., 2010
C. botulinum ATCC 3502 (Unilever) *SleB*::CT	ClosTron inserted in *SleB* gene; Erm^{R}	Meaney *et al*., 2015
C. botulinum ATCC 3502 (Unilever) *CwlJ*::CT	ClosTron inserted in *CwlJ*: gene; Erm^{R}	Meaney *et al*., 2015
C. botulinum ATCC 3502 (Unilever) *YpeB*::CT	ClosTron inserted in *YpeB* gene; Erm^{R}	Meaney *et al*., 2015
C. botulinum ATCC 3502 (Unilever) CBO1789::CT	ClosTron inserted in CBO1789 gene; Erm^{R}	Meaney *et al*., 2016
C. botulinum ATCC 3502 (Unilever) CBO1790::CT	ClosTron inserted in CBO1790 gene; Erm^{R}	Meaney *et al*., 2016
C. botulinum ATCC 3502 (Unilever) CBO3048::CT	ClosTron inserted in CBO3048 gene; Erm^{R}	Meaney *et al*., 2016
C. botulinum ATCC 3502 (Unilever) CBO3145::CT	ClosTron inserted in CBO3145 gene; Erm^{R}	Meaney *et al*., 2016
Chapter 5		
C. botulinum NCTC 2916 Δ*spoVA*^{2}	*spoVA*^{2} deletion via CRISPR/Cas9	this study
C. botulinum ATCC 19397 Δ*boNT/A1*	*boNT/A1* deletion via CRISPR/Cas9	this study
C. botulinum ATCC 19397 ΔClA	botulinum neurotoxin gene cluster type A deletion via CRISPR/Cas9	this study
C. botulinum ATCC 19397 i*boNT/A1*^{RY}	toxoid; R363A/Y366F mutations in *boNT/A1* via CRISPR/Cas9	this study
C. botulinum NCTC 2916 Δ*boNT/A1*	*boNT/A1* deletion via CRISPR/Cas9	this study
C. botulinum NCTC 2916 Δ*boNT/B*	*boNT/B* deletion via CRISPR/Cas9	this study
C. botulinum NCTC 2916 Δ*boNT/A1* Δ*boNT/B*	*boNT/A1 boNT/B* deletion via CRISPR/Cas9	this study
C. botulinum NCTC 2916 ΔClA	botulinum neurotoxin gene cluster type A deletion via CRISPR/Cas9	this study
C. botulinum NCTC 2916 ΔClB	botulinum neurotoxin gene cluster type B deletion via CRISPR/Cas9	this study
C. botulinum NCTC 2916 ΔClA ΔClB	botulinum neurotoxin gene cluster type A and B deletion via CRISPR/Cas9	this study

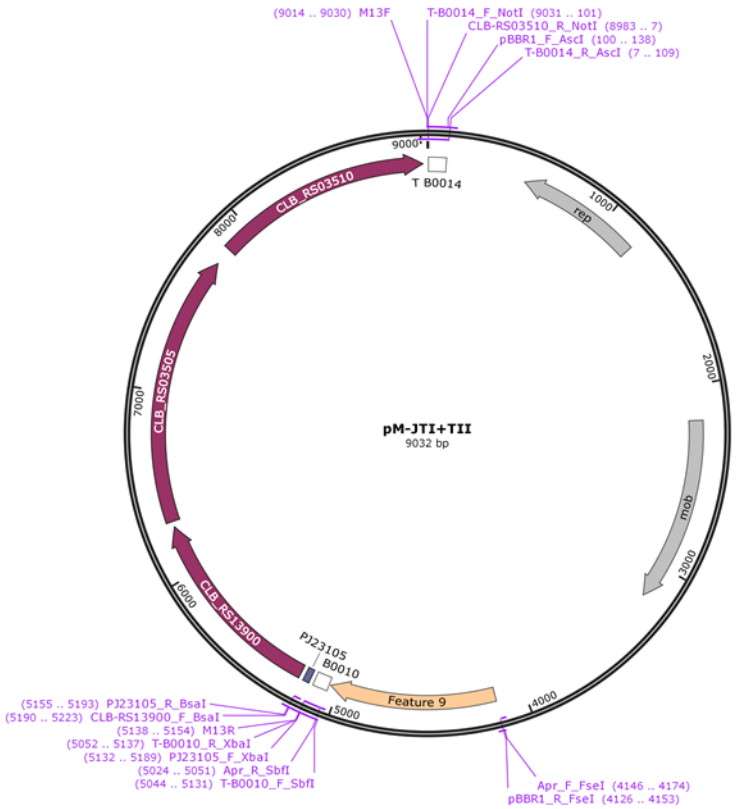
II. Plasmids

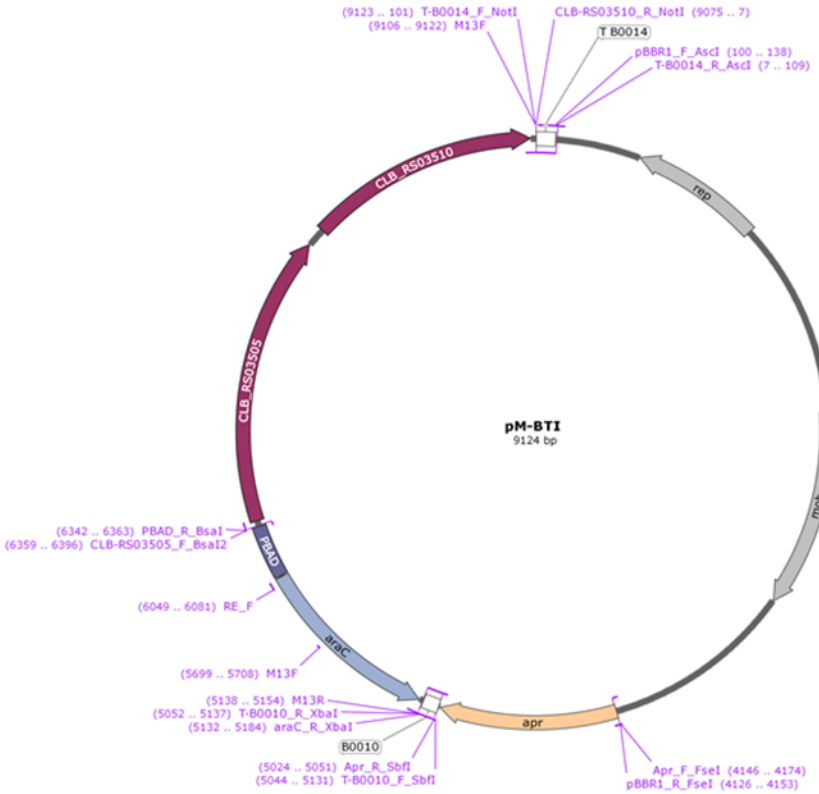
Table A 3 List of plasmids used in this study

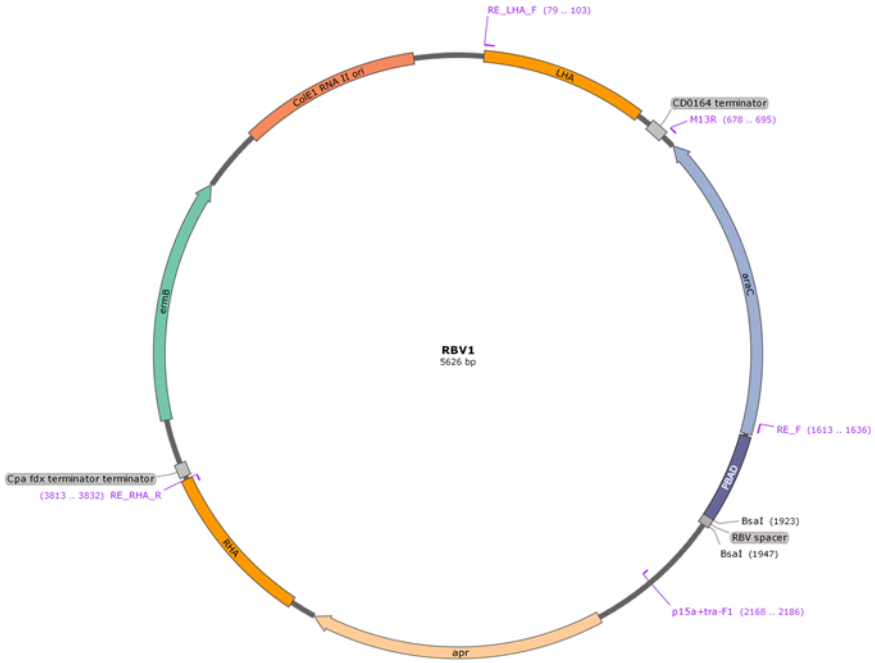
Plasmid	Description	Source
General		
pMTL83151	Cm^R/Tm^R; *catP*, pCB102/ColE1+tra replicons, MCS	SBRC
pMTL431511	Cm^R/Tm^R; *catP*, pCB102/ColE1+tra replicons, P_{fdx}-fs*cas9*, P_{araE}	Ingle *et. al*., 2019
pRECas1-MSC	Cm^R/Tm^R; *catP*, pCB102/ColE1+tra replicons, P_{fdx}*E-*cas9*, P_{araE}	Cañadas *et al.*, 2019
pRECas2-MSC	Cm^R/Tm^R; *catP*, pCB102/ColE1+tra replicons, P_{fdx}*E-*cas9*, P_{J23119}	Cañadas *et al.*, 2019
R702	$Kan^R/Strp^R$; conjugative multidrug resistant plasmid (R factor)	SBRC
Chapter 3		
RBV1	Erm^R/Apr^R; *ermB*, *apr*, ColE1 replicon, P_{BAD}-*araC*, HAs *aadA1*	Mansfield, 2018 (unpublished)
RBV1-BTI+TII	Erm^R/Apr^R; targeting *aadA1* locus R702 and inserting P_{BAD}-*araC*/CLB_RS13900/*hsdM*/*hsdS*	this study
pKD46	Cm^R; Lambda Red recombinase expression plasmid	Datsenko & Wanner, 2000
pMTL83151-R1	Cm^R; ATCC 19397 type I RM system recognition site R1	this study
pMTL83151-R2	Cm^R; ATCC 19397 type I RM system recognition site R2	this study
pMTL83151-R1+R2	Cm^R; ATCC 19397 type I RM system recognition site R1 and R2	this study
pMTL007C-E2-hsdR-1600s (pMTL007C-E2:240608)	Cm^R/Tm^R; ClosTron plasmid; *catP*, ColE1+tra/pCB101 replicon, group II intron targeting *hsdR C. botulinum* ATCC 3502	ATUM
pM-JTI	Apr^R; *apr*, pBBR1 replicon, P_{J23105}-*hsdM*/*hsdS*	this study
pM-JTII	Apr^R; *apr*, pBBR1 replicon, P_{J23105}-*hsdM*/*hsdS*	this study
pM-JTI+II	Apr^R; *apr*, pBBR1 replicon, P_{J23105}- CLB_RS13900/*hsdM*/*hsdS*	this study
pM-BTI	Apr^R; *apr*, pBBR1 replicon, P_{BAD}-*araC*/*hsdM*/*hsdS*	this study
pM-BTII	Apr^R; *apr*, pBBR1 replicon, P_{BAD}-*araC*/CLB_RS13900	this study
R702TI+TII	Kan^R/Apr^R; P_{BAD}-*araC*/CLB_RS13900/*hsdM*/*hsdS* at *aadA1* locus	this study

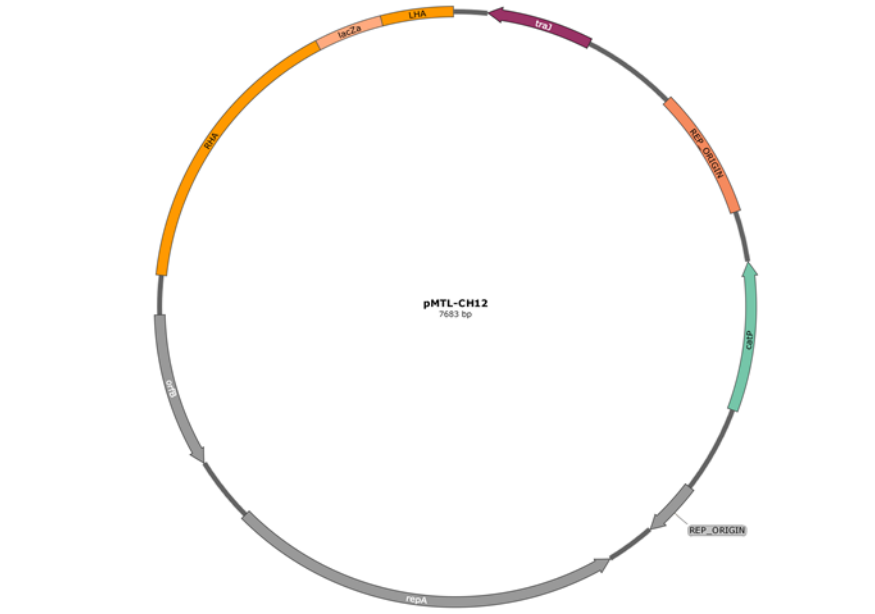
Chapter 4		
pMTL-CH12	Cm^R/Tm^R; *catP*, pCD6/ColE1+tra replicons; ACE plasmid for the creation of *pyrE*$^-$ in *C. botulinum* ATCC 3502	Humphreys, 2013 (unpublished)
pMTL431511_boNT/A1-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of *boNT/A1* in *C. botulinum* ATCC 3502	this study
pRECas2-mutL-G1d	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of *mutL* in *C. botulinum* ATCC 19397	this study
pRECas2-mutL231-G1d	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of 231 bp in *mutL* in *C. botulinum* ATCC 19397	this study
pRECas2-mutL_comp-G2c	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the insertion of *mutL* in *C. botulinum* ATCC 19397 Δ*mutL* at the *mutL* locus	this study
pRECas2-mutL_comp-G1c	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the insertion of *mutL* in *C. botulinum* ATCC 19397 Δ*mutL*231 at the *mutL* locus	this study
Chapter 5		
pRECas1-spoVA2-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of *spoVA*^{2} in *C. botulinum* NCTC 2916	this study
pMTL-CH15	Cm^R/Tm^R; *catP*, pCD6/ColE1+tra replicons; ACE plasmid for the deletion of the neurotoxin gene cluster in *C. botulinum* ATCC 3502	Humphreys, 2013 (unpublished)
pRECas1-19_boNT/A1-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of *boNT/A1* in *C. botulinum* ATCC 19397	this study
pRECas1-19_ClA-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of neurotoxin gene cluster in *C. botulinum* ATCC 19397	this study
pRECas1-19_toxoid_P-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for mutating R363A/Y366F in *boNT/A1* in *C. botulinum* ATCC 19397	this study
pRECas1-29_boNT/A1-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of *boNT/A1* in *C. botulinum* NCTC 2916	this study
pRECas1-29_boNT/B-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of *boNT/*B in *C. botulinum* NCTC 2916	this study
pRECas1-29_ClA-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of neurotoxin gene cluster A in *C. botulinum* NCTC 2916	this study
pRECas1-29_ClB-G1	Cm^R/Tm^R; CRISPR/Cas9 plasmid for the deletion of neurotoxin gene cluster B in *C. botulinum* NCTC 2916	this study

III. Representative plasmid maps






RE_LHA_F (79 .. 103)
LHA
CD0164 terminator
M13R (678 .. 695)
araC
RE_F (1613 .. 1636)
PBAD
BsaI (1923)
RBV spacer
BsaI (1947)
p15a+tra-F1 (2168 .. 2186)
apr
RHA
Cpa fdx terminator terminator
(3813 .. 3832) RE_RHA_R
ermB
CoIE1 RNA II ori
RBV1
5626 bp


LHA
lacZa
traJ
RHA
REP_ORIGIN
cat-P
oriB
repA
REP_ORIGIN
pMTL-CH12
7683 bp

IV. Primers

Table A 4 List of primers used in this study

Primer	Sequence	Description
General		
M13F	TGTAAAACGACGGCCAGT	screening/sequencing
M13R	CAGGAAACAGCTATGACC	screening/sequencing
ColE1-F1	CGTATTACCGCCTTTGAGTG	screening/sequencing
MCS-R1	GGTCATAGCTGTTTCCTG	screening/sequencing
pCB102-R1	CTGTTATGCCTTTTGACTATC	screening/sequencing
83XXXL_R	ACGGCTTGATGTGTTGGTAG	screening/sequencing
p15a+tra-F1	GGGTTCGTGTAGACTTTCC	screening/sequencing
ParaE_F	ATATATTCTAGATTTATATTTAGTCCCTTGCC	screening/sequencing
sgRNA_R_AatII_Universal	GTATCAGACGTCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCC	sgRNAs universal reverse primer
Chapter 3		
TI_R1_Cas9_F	TATGCTAAATACGATGAAGC	RM type I recognition site R1 ATCC 19397
TI_R1_Cas9_R	GGCCGCTTCATCGTATTTAGCA	RM type I recognition site R1 ATCC 19397
TI_R2_Cas9_F	TATGCTAGTCATTATGAAGC	RM type I recognition site R2 ATCC 19397
TI_R2_Cas9_R	GGCCGCTTCATAATGACTAGCA	RM type I recognition site R2 ATCC 19397
TI-R1+2_Cas9_NdeI_F	GCTATACATATGCTAAATACGATGAAAATGATAAACTTATTCG	RM type I recognition site R1+R2 ATCC 19397
TI-R1+2_Cas9_NotI_R	GCATATGCGGCCGCTTCATAATGACTAGCTAAATATAAAAAATTCAC	RM type I recognition site R1+R2 ATCC 19397
Cbot_T1R_F	GATTTAATTATTGTTGATGAATCCCATAGAAG	screening/sequencing *hsdR* deletion ATCC 3502

Cbot_T1R_R	GCCACTGATTAATAGCAGAACTGTA	screening/sequencing *hsdR* deletion ATCC 3502
RE_F	GCCATTCAGAGAAGAAACCAATTG	screening/sequencing recombineering cassette insertion in R702
RE_R	ATTCAGAATGCCGAACACC	screening/sequencing recombineering cassette insertion in R702
RE_LHA_F	TTCTACGGCACGTTTGAAGGCGCGC	amplification recombineering cassette
RE_RHA_R	GCAGCGACTTCCACGATGGG	amplification recombineering cassette
CLB-RS13900_F_BsaI	GAATTCGGTCTCGCATTCCCAAAAAAACGGGTATG	CLB-RS13900 ATCC 19397
CLB-RS13900_R_BsaI	GTACTTGGTCTCTCCCCCTTATAGATTAACATTACATTCTACAC	CLB-RS13900 ATCC 19397
CLB-RS03505_F_BsaI	GTATCTGGTCTCTGGGGTATAACTAATGATGGAGGAAACATAATGATTAC	CLB-RS03505 (*hsdM*) ATCC 19397
CLB-RS03510_R_BsaI	GTACTTGGTCTCTATCGCTATTTAATTAAACAACTCTCCTTTAAATGC	CLB-RS03510 (*hsdS*) ATCC 19397
CLB_RS03505_S1_R	GCAAGTACTAGAGTATATTTTTCACTGTCC	sequencing
CLB_RS03510_S1_R	CTGGAAATCCCCTTTTCATCCAAG	sequencing
CLB_RS03510_S2_R	GTTAATATATCTCCTAATTCGCACAAC	sequencing
T-B0014_F_NotI	GTAACTGCGGCCGCTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATATACTAGAGAGAGAATATAAAAAGCCAGATTATTAATCCGGCTTTTTTATTATTT	terminator B0014
T-B0014_R_AscI	GTATTAGGCGCGCCAAATAATAAAAAAGCCGGATTAATAATCTGGCTTTTTATATTCTCTCTCTAGTATATAAACGCAGAAAGGCCCACCCGAAGGTGAGCCAGTGTGA	terminator B0014
T-B0010_F_SbfI	GATTATCCTGCAGGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTC	terminator B0010
T-B0010_R_XbaI	GATACTTCTAGAGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGATGCCTGG	terminator B0010
pBBR1_F_AscI	GATATTGGCGCGCCCGTGCTACGCCTGAATAAGTGATAATAAG	replicon pBBR1

pBBR1_R_FseI	GTATATTGGCCGGCCCTGAGGTCTGCCTCGTGAAG	replicon pBBR1
Apr_F_FseI	GTATATTGGCCGGCCGTGCAGCTCCATCAGCAAAAG	apramycin resistance gene
Apr_R_SbfI	TTTTATCCTGCAGGTCAGCCAATCGACTGGCGAG	apramycin resistance gene
CLB-RS03505_F_BsaI2	GTATCTGGTCTCGAATGTATAACTAATGATGGAGGAAACATAATGATTAC	CLB-RS03505 (*hsdM*) ATCC 19397
CLB-RS03510_R_NotI	GATATTAGCGGCCGCGTAAAACGACGGCCAGTCTATTTAATTAAACAACTCTCCTTTAAATGC	CLB-RS03510 (*hsdS*) ATCC 19397
CLB-RS13900_R_NotI	GATATTAGCGGCCGCCGTAAAACGACGGCCAGTCCCCTTATAGATTAACATTACATTCTACAC	CLB-RS13900 ATCC 19397
PBAD_R_BsaI	GAATTCGGTCTCGCATTCCCAAAAAAACGGGTATG	inducible promoter system P_{BAD}-*araC*
araC_R_XbaI	GATCAATCTAGACAGGAAACAGCTATGACTTATGACAACTTGACGGCTACATCATTCAC	inducible promoter system P_{BAD}-*araC*
PJ23105_F_XbaI	GATCAATCTAGACAGGAAACAGCTATGACTTTACGGCTAGCTCAGTCCTAGGTACTATGCTAGC	P_{j23105}
PJ23105_R_BsaI	GAATTCGGTCTCGCATTGCTAGCATAGTACCTAGGACTGAGCTAGCCGTAAA	P_{j23105}
Chapter 4		
CbopyrE/Fd	CCGGTTATAGGTATAGGAGG	screening/sequencing *pyrE* truncation
boNT/A1_scr_F	GTTTTCAATATGATAGATTCTAATGGAAATG	screening/sequencing *boNT/A1* deletion
boNT/A1_scr_R	CGAAGAATGAAGTTGAATATAAACAG	screening/sequencing *boNT/A1* deletion
boNT/A1_RHA_F	AAATATGCCACCACTGTAATTAATCTCAAACTAC	RHA *boNT/A1* deletion ATCC 3502
boNT/A1_RHA_R_AscI	ATATCTGGCGCGCCCGCATTTACAGATTAATGAAG	RHA *boNT/A1* deletion ATCC 3502
boNT/A1_LHA_AsiSI_F	GTATACGCGATCGCGGCACTATATAACTATATCTGTAGATCG	LHA *boNT/A1* deletion ATCC 3502
boNT/A1_LHA_R	ATTACAGTGGTGGCATATTTAACACCTCTTA	LHA *boNT/A1* deletion ATCC 3502
sgRNA1_boNT/A1_F_SalI	TTTTCGTCGACGTTAACAGAGATTTACACAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	sgRNA targeting *boNT/A1* ATCC 3502

sgRNA_R_AsiSI	CGCGCGCGGCGATCGCATAAAAATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTTATAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTG	sgRNA universal reverse primer
mutL_F_scr	CTTTATATAATAGATCAACATGCGGCTCATG	screening/sequencing *mutL*231 deletion
mutL_R_scr	GTAGGTCTTCCATGGGGACAAG	screening/sequencing *mutL*231 deletion
mutL_del-sp_scr	GCATGTCTTCTATTAATTTAAAAATATCCTTG	screening/sequencing *mutL*231 deletion specific
mutL_full-sp_scr	CTTTAACTGCGGATTTGCATGC	screening/sequencing full *mutL* specific
mutL_F2_scr-del	GGTAATAATATAAATAAACTACCTAACAATACAGAACTTC	screening/sequencing full *mutL*
Gd1mutL_SpeI_F	GTATCAACTAGTGTTGTCATTAGCTTTAACTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG	sgRNA targeting *mutL*
G1c_CLB-RS08580_SpeI_F	GTATCAACTAGTCCATAATAGCAGGAGGTACAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG	sgRNA targeting CLB-RS08580 for complementation *mutL*231 ATCC 19397
G2c_CLB-RS08580_SpeI_F	GTATCAACTAGTATAAATTTGCATAGAGTCAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG	sgRNA targeting CLB-RS08580 for complementation *mutL* ATCC 19397
mutL_LHA_F_AatII	GTAGCCGACGTCGCTAATATTTCTATATGTGATAAGATATTTACAAGAATAG	LHA full *mutL* deletion ATCC 19397
mutL_LHA_R_BsaI	GCTACTGGTCTCAGTATCCTCATTTAAATCACCACATAGC	LHA full *mutL* deletion ATCC 19397
mutL_RHA_F_BsaI	GCTATGGTCTCCATACAATGAGGGGGATAAGGAAAAATG	RHA full *mutL* deletion ATCC 19397
mutL_RHA_R_AscI	GTTCATGGCGCGCCCTTATTCTTATATAATTCTATAAATAAATTAAAAATCCTTTTAATTC	RHA full *mutL* deletion ATCC 19397
mutL_LHAdel231_F_AatII	GTAGCCGACGTCCAATAAGTGAAAACCTTATAAACTTCGAAAG	LHA *mutL*231 deletion ATCC 19397
mutL_LHAdel231_R_BsaI	GCTACTGGTCTCAATTTAAAAATATCCTTGTTTTCTTCGTATATATTAAATTCATC	LHA *mutL*231 deletion ATCC 19397
mutL_RHAdel231_F_BsaI	GCTATGGTCTCCAAATTAATAGAAGACATGCTTATATTAAATAATCC	RHA *mutL*231 deletion ATCC 19397

mutL_RHAdel231_R_AscI	GTTCATGGCGCGCCGATCTTTTATCTTTTCTAAACCAAGTAAGTTG	RHA *mutL*231 deletion ATCC 19397
mutL_Comp_G1c_R_BsaI	GCTACTGGTCTCATAATGGGCAACTTTCCTCTGTTCCATATATC	LHA *mutL* complementation in *mutL* locus with silent mutations in the SEED sequence ATCC 19397
mutL_Comp_G1c_F_BsaI	GCTATGGTCTCCATTATTGCTGGTGGAACTGGTTTATATATTAATTCTTTAATATACAATTATGATTTTACAGATGCAGATAG	RHA full *mutL* insertion in *mutL* locus with silent mutations in the SEED sequence ATCC 19397
mutL_Comp_G2c_R_BsaI	GCTACTGGTCTCACTGATATTATCTCTCCATTTAGTTTTTCTGCAAGTTTTATG	LHA full *mutL* insertion in *mutL* locus with silent mutations in the SEED sequence ATCC 19397
mutL_Comp_G2c_F_BsaI	GCTATGGTCTCCTCAGCAGATAGTATGCAGATATACAAATACATGGATATAGGTTCAGCCAAAATAAC	RHA full *mutL* insertion in *mutL* locus with silent mutations in the SEED sequence ATCC 19397
mutL_F_scr-del	GCAGGTAGCCCTAATTACAATTATGG	screening/sequencing full *mutL* deletion
mutL_R_scr-del	CAACTTTAGTCATATAACCTTTCCCTC	screening/sequencing full *mutL* deletion
mutL_S1_F	CAATATGACCCCAATGGAAG	Sequencing *mutL*
Chapter 5		
rpoB_scr_F1	GAGAAGCTATAAAAGCTAAGTTAGAAG	sequencing *rpoB*
rpoB_scr_F2	AGTAAGCCAATTAGTTAGATCACCAG	sequencing *rpoB*
rpoB_scr_F3	AGAAGATATATATGCTACCATAAATTATG	sequencing *rpoB*
rpoB_scr_F4	GCAACTGCAATGATACCATTCC	sequencing *rpoB*
rpoB_scr_F5	GAGTTAATAAATTAGTTAGATGCTATATAGCTC	sequencing *rpoB*
rpoB_scr_R1	CTGGTGAAGCTAAACCTATTTG	sequencing *rpoB*
G1spoVA2D_SalI_F	TTTTCGTCGACAAAGACCACCTACTGCTCAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	sgRNA targeting *spoVA*^{2} NCTC 2916
G2spoVA2D_SalI_F	TTTTCGTCGACTGGAGCAGCTATGGCACCAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	sgRNA targeting *spoVA*^{2} NCTC 2916
spoVA2_LHA_F2_AatII	GTATATGACGTCCCGTATAAGTCATATCCTGCTTTTCTC	LHA *spoVA*^{2} deletion NCTC 2916

spoVA2_LHA_R2_BsaI	GTATATGGTCTCTCTTTAATAAATAATTAATATACTCCCATATCTAGTTAAATTTATATTTTATAG	LHA *spoVA*^{2} deletion NCTC 2916
spoVA2_RHA_F2_BsaI	GTATATGGTCTCTAAAGGTTAAGCATATGAAAAAATATGTTAC	RHA *spoVA*^{2} deletion NCTC 2916
spoVA2_RHA_R_AscI	GATATAGGCGCGCCCTTAAGCTTTGCCAGTCGTGAAG	RHA *spoVA*^{2} deletion NCTC 2916
spoVA2del_scr_R	GCTGCTTGTGTTGTAGATGCAGG	screening/sequencing
spoVA2del_scr_F2	CTCATTTACAACTCTTATGCCACC	screening/sequencing
G1boNT/A1_F_SalI	TTTTCGTCGACGTTAACAGAGATTTACACAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	sgRNA targeting *boNT/A1* (ATCC 19397/NCTC 2916)
G1boNT/B_F_SalI	TTTTCGTCGACGCAAGTATATTTAATAGACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	sgRNA targeting *boNT/B*
boNT/A1_LHA_AatII_F	GTACTAGACGTCGGCACTATATAACTATATCTGTAGATCG	LHA *boNT/A1* deletion ATCC 19397
boNT/A1_ClDel_AatII_F	GTACTAGACGTCGATATTTAGTTATACTTATTTTGATTATACTACC	LHA ClA deletion ATCC 19397
boNT/A1-toxoid_LHA_AatII_F	GTACTAGACGTCGGAATACCATTTTGGGGTGGAAG	LHA *boNT/A1* toxoid ATCC 19397
MutboATox_R_P	CTTCTGTGTAAATCTCTGTTAACATTTTGTATAACTTATCAAATTTTAATTTATC	LHA *boNT/A1* toxoid ATCC 19397 with a silent mutation in the PAM sequence
MutboATox_R_S	CCTCAGTATAAATCTCAGTAAGCATTTTGTATAACTTATCAAATTTTAATTTATC	LHA *boNT/A1* toxoid ATCC 19397 with a silent mutation in the SEED sequence
MutboATox_R_PS	CTTCAGTATAAATCTCAGTAAGCATTTTGTATAACTTATCAAATTTTAATTTATC	LHA *boNT/A1* toxoid ATCC 19397 with a silent mutation in the PAM and SEED sequence
MutboATox_F_P	GTTAACAGAGATTTACACAGAAGATAATTTTGTTAAGTTTTTTAAAGTACTTAACGCTAAAACATTTTTGAATTTTGATAAAG	RHA *boNT/A1* toxoid ATCC 19397 with a silent mutation in the PAM sequence
MutboATox_F_S2	GCTTACTGAGATTTATACTGAGGATAATTTTGTTAAGTTTTTTAAAGTACTTAACGCTAAAACATTTTTGAATTTTGATAAAG	RHA *boNT/A1* toxoid ATCC 19397 with a silent mutation in the SEED sequence
MutboATox_F_PS	GCTTACTGAGATTTATACTGAAGATAATTTTGTTAAGTTTTTTAAAGTACTTAACGCTAAAACATTTTTGAATTTTGATAAAG	RHA *boNT/A1* toxoid ATCC 19397 with a silent mutation in the PAM and SEED sequence
boNT/ATox_AscI_R	ATATCTGGCGCGCCCATATACTAATTGTTCTACCCAGC	RHA *boNT/A1* toxoid ATCC 19397

boNT/A1_scr_F	GTTTTCAATATGATAGATTCTAATGGAAATG	screening/sequencing
boNT/A1_scr_R	CGAAGAATGAAGTTGAATATAAACAG	screening/sequencing
boNT/A1_Cldel_scr_F	TATTGCTAATTTCTGATATTAAAGGTC	screening/sequencing
boNT/A1-toxoid_scr_R	GGCTTTATCAAAATTCAAAAATGTTTTAGC	screening/sequencing
boNT/A1-2916_LHAdel_F2_AatII	GTATATGACGTCACTTGATGAGGTAATAATTTCTGTATTAGATAATATGG	LHA *boNT/A1* deletion NCTC 2916
boNT/A1-2916_LHAdel_R_BsaI	GTATATGGTCTCAGTGGTTGCATATTTAACACCTCTTATAATATTTAAAATTAATTAAATTATTAATTATTTAAATG	LHA *boNT/A1* deletion NCTC 2916
boNT/A1-2916_RHA_F_BsaI	GTATATGGTCTCACCACTGTAATTAATCTCAAACTACATGAG	RHA *boNT/A1* deletion NCTC 2916
boNT/A1-2916_RHA_R2_AscI	GTATATGGCGCGCCCCTAAATTTTTCTCATAGTTATATCTTTCATAATAAAAAAGAAAATGTATG	RHA *boNT/A1* deletion NCTC 2916
boNT/B-2916_LHAdel_F2_AatII	GATATAGACGTCGGGATGAAGGAGTAGTTTGTTTATTAGGAG	LHA *boNT/B* deletion NCTC 2916
boNT/B-2916_LHAdel_R_BsaI	GATATGGTCTCACAGTTGGCATAAATATTCTCCTCCTTATTTTAAGTTTATATCC	LHA *boNT/B* deletion NCTC 2916
boNT/B-2916_RHAdel_F_BsaI	GATATTGGTCTCAACTGAATAATATAACTATATGCTCAGCAAACC	RHA *boNT/B* deletion NCTC 2916
boNT/B-2916_RHAdel_R2_AscI	GATATAGGCGCGCCCAAAATATGATTTTGGTTTTTCTTCATAACCTGCAG	RHA *boNT/B* deletion NCTC 2916
boNT/A1-2916_LHACldel_F_AatII	GTATATGACGTCGTCTTATCAGAATAATTATTTTTATTTATTGGCACAG	LHA ClA deletion NCTC 2916
boNT/A1-2916_LHA-Cldel_R_BsaI	GTATATGGTCTCAGTGGAATACTTAATTATATATTTGTGCTAAATATTATTTTAAATTCCC	LHA ClA deletion NCTC 2916
boNT/A1-2916_RHA_F_BsaI	GTATATGGTCTCACCACTGTAATTAATCTCAAACTACATGAG	RHA ClA deletion NCTC 2916
boNT/A1-2916_RHA_R_AscI	GTATATGGCGCGCCCATAGCAGCTTATAAATTTAATAGGTGTATTTATAC	RHA ClA deletion NCTC 2916
boNT/B-2916_LHACldel_F_AatII	GATATAGACGTCGATATTAATCTATATATATTTGATGAAGAATATAATTCATTTG	LHA ClB deletion NCTC 2916
boNT/B-2916_LHA-Cldel_R_BsaI	GATATGGTCTCATGAATCTGATTGAAGTACTGCGGC	LHA ClB deletion NCTC 2916

boNT/B-2916_RHA-Cldel_F_BsaI	GATATTGGTCTCATTCATTTAATAGTTCTATATGTCTTTGAACTCTAG	RHA ClB deletion NCTC 2916
boNT/B-2916_RHA-Cldel_R_AscI	GATATAGGCGCGCCGTAACTTATTGCCACATTTTCTGC	RHA ClB deletion NCTC 2916
boNT/A1-2916_scr_R2	CAGCTCTTAAAATAGCAAATTGTATTCC	screening/sequencing
boNT/B-2916_del_scr_R2	AATAAAGATTACAATACTGAATTTGCAATTCC	screening/sequencing
boNT/B-2916_Cldel_scr_R	CATCATTAAAAATTTCATTATCCAAAAGCTTATC	screening/sequencing

V. Phenotypic characterisation of ATCC 19397 Δ*boNT/A1* mutants

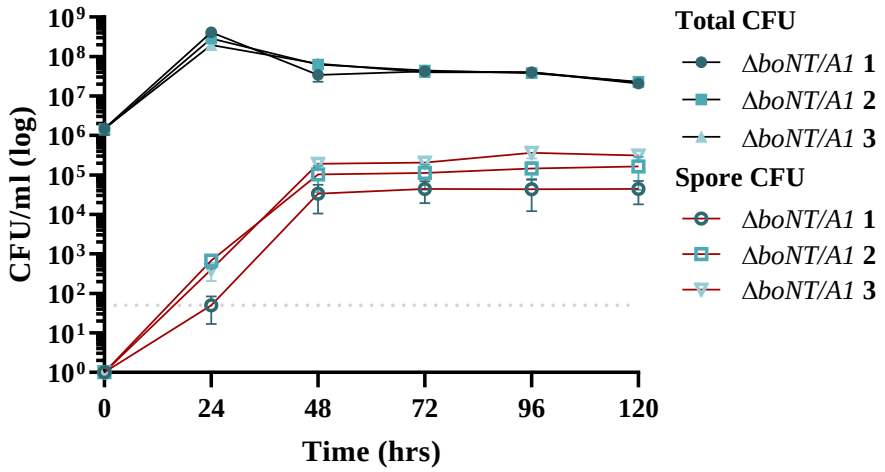


Figure A 1 Phenotypic characterisation of ATCC 19397 Δ*boNT/A1* mutants. (**A**) Growth profiles of Δ*boNT/A1* mutants. Strains were cultured in TYG broth and the OD_{595} was monitored for 24 hours. Error bars represent the standard deviation of three independent cultures for WT and of three biological replicates for the mutants. (**B/C/D**) Sporulation/germination capacity of Δ*boNT/A1* mutants. Strains were cultured for 120 hours in TYG broth and samples were taken every 24 hours. Total CFU corresponds to the total cell count (non-heat treated) and the spore CFU is the number of viable spores withstanding a 15 min, 80°C heat treatment. Error bars represent the standard deviation of three independent cultures. The dotted line at 50 CFU/ml represents the detection limit.

VI. Whole genome sequencing and variant detection Chapter 4

Comparison of WT genomes:

- Light grey: low coverage and/or frequency
- Red: mutation is only present in one of the two WT genomes

Comparison of mutant genomes to both WT genomes:

- Light grey: low coverage and/or frequency
- Yellow: mutation is also present in one of the two WT genomes
- Light Blue: present in multiple mutants but not in one of the WT genomes
- Dark Blue: mutations that strains *pyrE*$^{-}$ 2 and *pyrE*$^{-}$ 3 have many in common (isolated from a single conjugal mixture).
- Green: mutation in gene related to sporulation or germination

Table A 5 List of variants detected in ATCC 3502$^{+/-}$ (Unilever; green SBRC culture collection stock) compared to the published reference genome (accession number: NC_009495)

ATCC 3502$^{+/-}$ WT (Unilever; green stock)									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotation**	**Coding region change**	**Amino acid change**
130004	SNV	C	A	60	60	100	Gene: *ileS* (isoleucyl-tRNA synthetase)	1617 C>A	
209980	Deletion	G	-	27	49	55.1			
570086	Deletion	A	-	39	64	60.9	Gene: CBO0481 (Hypothetical protein: no conserved domains detected)	93 del T	.Lys32 fs
1667030	SNV	G	A	64	69	92.8	Gene: CBO1533 (AraC family transcriptional regulator)	696 G>A	
1736091	SNV	T	C	54	62	87.1	Gene: *glcK* (Glucokinase)	212 T>C	Leu71Ser
3234007	Insertion	-	A	82	98	83.7			
3267818	Insertion	-	A	51	60	85	Gene: *hpt* (Hypoxanthine phosphoribosyltransferase)	35 dup	Asn12 fs
3629557	Insertion	-	T	43	60	71.7	Gene: CBO3388 (PDZ domain of trypsin-like serine proteases, involved in heat-shock response, chaperone function, and apoptosis)	WT with fs, T insertion restores ORF	
3724684	SNV	C	T	37	64	57.8	Gene: *tufB* (Elongation factor Tu)	381 G>A	

Table A 6 List of variants detected in ATCC 3502$^{+/-}$ (Unilever; red SBRC culture collection stock) compared to the published reference genome (accession number: NC_009495)

ATCC 3502$^{+/-}$ WT (Unilever; red stock)									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotation**	**Coding region change**	**Amino acid change**
130004	SNV	C	A	150	150	100	Gene: *ileS* (isoleucyl-tRNA synthetase)	1617 C>A	
209980	Deletion	G	-	62	105	59			
1667030	SNV	G	A	96	114	84.2	Gene: CBO1533 (AraC family transcriptional regulator)	696 G>A	
1736091	SNV	T	C	108	116	93.1	Gene: *glcK* (Glucokinase)	212 T>C	Leu71Ser
2804573	Insertion	-	C	65	122	53.3	Gene: CBO2653 (Flagellar protein FlbD)	196 dup	Glu66 fs
3234007	Insertion	-	A	126	145	86.9			
3267818	Insertion	-	A	144	165	87.3	Gene: *hpt* (Hypoxanthine phosphoribosyltransferase)	35 dup	Asn12 fs
3629557	Insertion	-	T	117	136	86	Gene: CBO3388 (PDZ domain of trypsin-like serine proteases, involved in heat-shock response, chaperone function, and apoptosis)	WT with fs, T insertion restores ORF	
3724684	SNV	C	T	51	85	60	Gene: *tufB* (Elongation factor Tu)	381 G>A	

Table A 7 List of variants detected in *hsdR*1600s::CT 1 compared to the published reference genome (accession number: NC_009495)

***hsdR*1600s::CT 1**									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotation**	**Coding region change**	**Amino acid change**
130004	SNV	C	A	90	90	100	Gene: *ileS* (Isoleucyl-tRNA synthetase)	1617 C>A	
385352	SNV	A	G	86	86	100	Gene: CBO0332 (*argB*: accessory gene regulator B)	578 A>G	Gln193Arg
392977	Insertion	-	A	77	79	97.5	Gene: CBO0340 (Signal transduction histidine kinase)	563 dup	Asn188 fs
524885	Insertion	-	A	65	69	94.2			
589301	SNV	A	G	75	75	100	Gene: CBO0586 (LolE superfamily: ABC-type transport system, involved in lipoprotein release, permease component, cell wall/membrane/envelope biogenesis)	319 T>C	Trp107Arg
683064	SNV	C	T	67	67	100		572 C>T	Ala191Val
826887	Insertion	-	A	68	70	97.1			
895995	Deletion	A	-	88	91	96.7			
1224668	Deletion	A	-	72	75	96	Gene: CBO1113 (TRX family group I protein)	295 del A	Ile99 fs
1224963	Insertion	-	A	58	61	95.1	Gene: CBO1114 (Hypothetical protein: no conserved domains detected)	176 dup	Asn59 fs
1371019	Deletion	T	-	82	86	95.4			
1466519	Deletion	A	-	62	64	96.9	Gene: CBO1342 (MurR/RpiR family: DNA-binding transcriptional regulator)	61 del A	Ile21 fs
1500731	Insertion	-	A	68	69	98.6	Gene: CBO1377 (L-type amino acid transporter)	35 dup	Asn12 fs
1537837	SNV	G	A	59	59	100	Gene: CBO1410 (Nicotinate phosphoribosyltransferase)	819 C>T	
1539098	Deletion	G	-	63	70	90			

1735600	Deletion	A	-	64	69	92.8			
1765830	SNV	T	C	52	52	100	Gene: *acoR* (Transcriptional regulator of acetoin/glycerol metabolism)	180 T>C	
1930729	SNV	G	A	80	80	100			
2289691	Insertion	-	A	66	66	100	Gene: CBO2146 (FtsX-like permease family: predicted permeases and hypothetical transmembrane proteins)	881 dup	Leu294 fs
2391417	SNV	C	T	62	63	98.4	Gene: CBO2232 (M20_pepD: Specificity for beta-alanyl-L-histidine dipeptide)	486 G>A	
2392655	Deletion	A	-	61	64	95.3	Gene: CBO2233 (Beta-eliminating lyase)	834 del T	Phe278 fs
2652170	SNV	A	G	65	65	100	Gene: CBO2515 (Hypotetical protein: no conserved domains detected)	24 T>C	
2966347	Insertion	-	A	66	68	97.1			
2991115	SNV	C	T	88	89	98.9			
3062301	Insertion	-	T	73	81	90.1	Gene: *purC* (Phosphoribosylaminoimidazole-succinocarboxamide synthase)	191 dup	Val66 fs
3093768	Deletion	T	-	67	70	95.7	Gene: CBO2904 (helD: DNA helicase IV)	1859 del A	Asn620 fs
3315239	SNV	A	G	81	81	100	Gene: CBO3117 (Hypothetical protein: no conserved domains detected)	246 T>C	
3553722	SNV	A	G	71	71	100	Gene: CBO3335 (Cyclic beta 1-2 glucan synthetase)	7443 T>C	
3637040	Insertion	-	T	87	88	98.9	Gene: *pemK* (Type II toxin-antitoxin system)	315 dup	Val106 fs
3671591	Insertion	-	T	93	98	94.9	Gene: CBO3430 (Hypothetical protein: no conserved domains detected)	831 dup	Val278 fs
3724684	SNV	C	T	59	59	100	Gene: *tufB* (Elongation factor Tu)	381 G>A	
3743752	SNV	G	A	43	43	100	Gene: *fusB* (Elongation factor G)	2054 C>T	Ser685 Phe
3866544	Insertion	-	TA	61	64	95.3	Gene: CBO3623 (gdPP/DHH family protein: phosphodiesterase)	1817-1818 dup	Ser607 fs

Reference position	Overlapping annotation	Insertion/deletion sequence
3743691..3743756	Downstream *fusB*	Deletion: AATATAGTATAAAAATCACCTAGTTTCTAGGTGATTTTTATACTATATTTTATTTTTACTAGATAA

Table A 8 List of variants detected in *hsdR*1600s::CT 2 compared to the published reference genome (accession number: NC_009495)

***hsdR*1600s::CT 2**									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotation**	**Coding region change**	**Amino acid change**
130004	SNV	C	A	90	90	100	Gene: *ileS* (Isoleucyl-tRNA synthetase)	1617 C>A	
209980	Deletion	G	-	85	88	96.6			
1078297	Insertion	-	T	132	143	92.3	Gene: CBO0977 (AraC-type DNA-binding domain)	WT with fs, T insertion restores ORF	Leu189Thr
1244400	SNV	C	T	93	93	100	Gene: CBO1130 (CoA-substrate-specific enzyme activase)	3197 C>T	Pro1066Leu
1585597	Insertion	-	T	111	113	98.2			
1667030	SNV	G	A	103	103	100	Gene: CBO1533 (AraC family transcriptional regulator)	696 G>A	
1736091	SNV	T	C	84	84	100	Gene: *glcK* (Glucokinase)	212 T>C	Leu71Ser
2466355	SNV	T	C	45	46	97.8			
3234007	Insertion	-	A	146	152	96.1			
3267818	Insertion	-	A	107	114	93.9	Gene: *hpt* (Hypoxanthine phosphoribosyltransferase)	35 dup	Asn12 fs
3629557	Insertion	-	T	86	93	92.5	Gene: CBO3388 (PDZ domain of trypsin-like serine proteases, involved in heat-shock response, chaperone function, and apoptosis)	WT with fs, T insertion restores ORF	Asn258Lys
3724684	SNV	C	T	92	92	100	Gene: *tufB* (Elongation factor Tu)	381 G>A	

Reference position	Overlapping annotation	Insertion/deletion sequence
209831..209980	Upstream *prfB*	Deletion: TAAATAACAAGTCAGTATGACAGTCTATTTTGCGTCATACTGCGTCAGCAGAATCCTCCGATAG CACTACTAGCGACGGAACCTGCTTCCTTGGCTGACACAAAATATACTGCCATCTTTGACTTGTTA TTTATTTTCATGTGCCTAAAG
1914862^1914863	*mutL*	Insertion: TTTTTTATTTCTTCTTCTTTTAAGTTGTCATTAGCTTTAACTGCGGATTTGCATGCCAGTGTAGCA ATGGCATTATATTTTATTGTAGAAGTTTCTTTAGACTTCATATTTTTTAAATTATAAAGTATATCC ATAAAGAGATTTTCTACATTGGGTTTACCTAGAATTAAAGGTACTTCTTTTATATTTATAGTGTA TTCTCCAAAAGTTTCCACTGAAAAACCTGAATTC

Table A 9 List of variants detected in *hsdR*1600s::CT 3 compared to the published reference genome (accession number: NC_009495)

***hsdR*1600s::CT 3**									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotations**	**Coding region change**	**Amino acid change**
130004	SNV	C	A	90	90	100	Gene: *ileS* (isoleucyl-tRNA synthetase)	1617 C>A	
196143	Deletion	A	-	168	175	96	Gene: CBO0161 (Peptidase family M23)	12 delA	Lys4 fs
244279	Insertion	-	A	149	155	96.1	Gene: *cysK1* (Cysteine synthase A)	456 dup	His153 fs
424790	Insertion	-	AT	114	120	95	Gene: CBO0366 (Sensor histidine kinase)	24-25 dup	Lys9 fs
681349	Insertion	-	A	108	114	94.7	Gene: CBO0584 (Lipoprotein)	117 dup	Glu40 fs
946308	Insertion	-	A	105	112	93.8	Gene: CBO0837 (Hypothetical protein: no conserved domains detected)	535 dup	Ser179 fs
1061539	SNV	C	T	160	160	100	Gene: *modA* (Molybdate-binding periplasmic protein precursor)	81 C>T	
1078297	Insertion	-	T	131	138	94.9	Gene: CBO0977 (AraC-type DNA-binding domain)	WT with fs, T insertion restores ORF	Leu189Thr
1342138	Deletion	T	-	101	104	97.1	Gene: CBO1221A (Hypothetical protein: no conserved domains detected)	260 del A	Asn87 fs
1524096	Insertion	-	A	84	91	92.3	Gene: CBO1397 (Peptide ABC transporter Pep4E family/ permease)	1354 dup	Ser452 fs
1691311	Insertion	-	A	123	126	97.6	Gene: CBO1558 (MA superfamily: Methyl-accepting chemotaxis protein)	925 dup	Ser309 fs
1806932	SNV	A	G	112	112	100	Gene: CBO1663 (Hypothetical protein: no conserved domains detected)	672 A>G	
1938356	Insertion	-	T	110	118	93.2	Gene: CBO1816 (Glycosyltransferase)	864 dup	Ala289 fs

2101364	Insertion	-	T	90	97	92.8	Gene: CBO1967 (BaeS superfamily: sensor histidine kinase)	136 dup	Met46 fs
2222256	SNV	C	T	109	109	100	CDS: CBO2086, DNA repair helicase	2074 G>A	Val692Ile
2289691	Insertion	-	A	97	101	96	Gene: CBO2146 (FtsX superfamily: ABC transporter permease)	881 dup	Leu294 fs
2788099	Insertion	-	A	141	146	96.6			
2983244	Insertion	-	T	132	139	95	Gene: *nrdE* (Ribonucleotide-diphosphate reductase subunit alpha)	67 dup	Met23 fs
3157040	Insertion	-	T	111	115	96.5	Gene: *spoIIP* (stage II sporulation protein)	485 dup	Lys163 fs
3299428	SNV	C	T	134	134	100	Gene: CBO3109 (FlaA1 superfamily: Exopolysaccharide biosynthesis protein, Cell wall/membrane/envelope biogenesis, Posttranslational modification, protein turnover, chaperones)	1687 G>A	Asp563Asn
3438345	Insertion	-	T	134	139	96.4			
3439378	SNV	T	C	162	162	100	Gene: *lonB* (lon superfamily: ATP-dependent protease)	714 A>G	
3721362	SNV	T	C	143	144	99.3			
3724684	SNV	C	T	97	97	100	Gene: *tufB* (Elongation factor Tu)	381 G>A	
3743752	SNV	G	A	91	91	100	Gene: *fusB* (Elongation factor G)	2054 C>T	Ser685Phe

Reference position	**Overlapping annotation**	**Insertion/deletion sequence**
3743687..3743752	Downstream *fusB*	Deletion: ATAAAATATAGTATAAAAATCACCTAGTTTCTAGGTGATTTTTATACTATATTTTATTTTTACTAG

Table A 10 List of variants detected in *pyrE*⁻ 1 compared to the published reference genome (accession number: NC_009495)

***pyrE*⁻ 1**									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotations**	**Coding region change**	**Amino acid change**
130004	SNV	C	A	90	90	100	Gene: *ileS* (isoleucyl-tRNA synthetase)	1617 C>A	
202999	Deletion	A	-	74	75	98.7	Gene: CBO0169 (RecD/TraA superfamily: helicase)	865 del A	Ile289 fs
314422	SNV	A	G	76	76	100	Gene: CBO0266 (GlcD superfamily: FAD-dependent oxidoreductase)	426 T>C	
392621	Insertion	-	T	41	44	93.2	Gene: CBO0340 (VicK superfamily: two-component sensor kinase)	207 dup	Thr70 fs
552405	SNV	A	G	165	167	98.8	Gene: CBO0464 (Hypothetical protein)	96 A>G	
760208	SNV	C	T	53	53	100			
879645	Deletion	A	-	77	79	97.5			
981108	Insertion	-	T	77	81	95.1	Gene: CBO0878 (Nitroreductase)	478 dup	Arg160 fs
1078221	Insertion	-	T	38	40	95	Gene: CBO0977 (AraC-type DNA-binding domain)	WT with fs, T insertion retores ORF	Leu189Thr
1165048	Insertion	-	T	56	58	96.6			
1243791	SNV	C	T	53	53	100	Gene: CBO1130 (YjiL/HGD-D superfamily: CoA-substrate-specific enzyme activase)	2588 C>T	Pro863Leu
1246905	Insertion	-	A	47	50	94	Gene: *opuD* (Choline/carnitine/betaine transporter)	177 dup	Ile60 fs
1358201	Deletion	T	-	24	24	100	Gene: CBO1240 (DNA-binding protein)	42 del A	Gly15 fs
1466519	Deletion	A	-	34	35	97.1	Gene: CBO1342 (MurR/RpiR family: DNA-binding transcriptional regulator)	61 del A	Ile21 fs

1467064	Insertion	-	A	47	49	95.9	Gene: CBO1342 (MurR/RpiR family: DNA-binding transcriptional regulator)	605 dup	Asn202 fs
1535502	Deletion	G	-	57	61	93.4	Gene: CBO1409 (C4-dicarboxylate transporter)	9 del G	Ile4 fs
1719199	SNV	C	T	52	52	100	Gene: *arcD* (Arginine/ornithine antiporter)	361 C>T	Arg121Cys
1736091	SNV	T	C	38	38	100	Gene: *glcK* (glucokinase)	212 T>C	Leu71Ser
2129295	Deletion	T	-	56	56	100	Gene: CBO1991 (PTS system fructose family transporter subunit IIBC, feneral metabolism)	712 del A	Met238 fs
2267736	SNV	T	C	64	64	100	Gene: CBO2129 (Hypothetical protein: no conserved domains detected)	287 A>G	Tyr96Cys
2278781	SNV	A	G	68	68	100	Gene: CBO2138 (Hypothetical protein: GluZincin superfamily)	512 T>C	Phe171Ser
2333003	SNV	C	T	115	116	99.1	Gene: CBO2185 (Hypothetical protein: no conserved domains detected0	348 C>T	
2411493	SNV	A	G	72	72	100	Gene: CBO2253 (Hydroxyethylthiazole kinase)	509 T>C	Met170Thr
2586961	Insertion	-	A	30	33	90.9			
2643525	SNV	T	A	40	40	100	Gene: CBO2507 (Ribosomal RNA small subunit methyltransferase B)	453 A>T	Lys151Asn
2715781	Deletion	TATATT	-	14	16	87.5			
2755021	Deletion	T	-	31	32	96.9			
2806558	Deletion	T	-	50	52	96.2	Gene: *flgD* (Flagellar hook capping protein)	795 del A	Lys265 fs
3142799	SNV	C	A	58	58	100	Gene: *spoIV* (Stage IV sporulation protein YqfD)	61 G>T	Glu21*
3234007	Insertion	-	A	48	51	94.1			
3391836	Insertion	-	C	58	60	96.7			
3591693	Insertion	-	C	61	66	92.4			

3724684	SNV	C	T	94	96	97.9	Gene: *tufB* (Elongation factor Tu)	381 G>A

Table A 11 List of variants detected in *pyrE*⁻ 2 compared to the published reference genome (accession number: NC_009495)

***pyrE*⁻ 2**									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotations**	**Coding region change**	**Amino acid change**
34221	Deletion	A	-	68	72	94.4	Gene: CBO0028 (Sensory box sigma-54 dependent transcriptional regulator)	240 del A	Lys80 fs
85454	SNV	A	T	63	64	98.4	Gene: CBO0052 (Amidohydrolase)	1130 T>A	Val377Glu
130004	SNV	C	A	90	90	100	Gene: *ileS* (isoleucyl-tRNA synthetase)	1617 C>A	
199497	SNV	T	G	82	83	98.8	Gene: CBO0165 (Hypothetical protein: tRNA 2-thiocytidine biosynthesis protein TtcA)	299 A>C	Asn100Thr
331795	Insertion	-	A	29	30	96.7	Gene: CBO0280 (MurR/RpiR family: Transcriptional regulator)	172 dup	Ile58 fs
392621	Insertion	-	T	22	22	100	Gene: CBO0340 (VicK superfamily: two-component sensor kinase)	207 dup	Thr70 fs
520957	Insertion	-	AT	62	65	95.4			
542165	SNV	A	G	72	72	100	Gene: CBO0453 (High-affinity branched-chain amino acid ABC transporter amino acid-binding protein)	261 A>G	
600787	Deletion	A	-	43	43	100	Gene: CBO0508 (Methyl-accepting chemotaxis protein)	1088 del A	Asn363 fs
606873	Insertion	-	A	34	36	94.4	Gene: CBO0511 (ATP-dependent helicase)	1818 dup	Gly607 fs
671367	Insertion	-	T	43	45	95.6	Gene: CBO0571 (Pentapeptide repeat-containing protein)	189 dup	Gly64 fs
860317	Deletion	T	-	61	64	95.3			
1043719	SNV	T	C	49	49	100	Gene: *cbiT* (Precorrin-6Y C5,15-methyltransferase subunit)	366 T>C	

1059859	Deletion	T	-	43	44	97.7			
1116813	Insertion	-	T	43	44	97.7	Gene: CBO1014 (Aspartate aminotransferase)	84 dup	Glu29 fs
1243791	SNV	C	T	51	51	100	Gene: CBO1130 (YjiL/HGD-D superfamily: CoA-substrate-specific enzyme activase)	2588 C>T	Pro863Leu
1466519	Deletion	A	-	29	29	100	Gene: CBO1342 (MurR/RpiR family: DNA-binding transcriptional regulator)	61 del A	Ile21 fs
1476063	Deletion	A	-	36	38	94.7	Gene: CBO1351 (LytB superfamily: Cell-surface protein)	509 del A	Asn170 fs
1613471	Insertion	-	A	36	39	92.3	Gene: CBO1481 (Methyltransferase)	1065 dup	Leu356 fs
1614684	Deletion	A	-	43	46	93.5	Gene: modA (Molybdate-binding periplasmic protein precursor)	186 del A	Lys62 fs
1736091	SNV	T	C	32	32	100	Gene: *glcK* (glucokinase)	212 T>C	Leu71Ser
2068908	Deletion	T	-	53	54	98.2	Gene: moeA2 (LysR family transcriptional regulator)	208 del A	Thr70 fs
2108692	Deletion	T	-	50	51	98	Gene: *gerAB1* (Spore germination protein)	815 del T	Leu272 fs
2129278	Insertion	-	T	41	44	93.2	Gene: CBO1991 (PTS system fructose family transporter subunit IIBC, feneral metabolism)	729 dup	Leu244 fs
2129295	Deletion	T	-	44	45	97.8	Gene: CBO1991 (PTS system fructose family transporter subunit IIBC, feneral metabolism)	712 del A	Met238 fs
2140175	Insertion	-	T	34	36	94.4			
2144065	SNV	G	A	48	48	100	Gene: CBO2006 (Xanthine/uracil permease)	104 C>T	Ala35Val
2198989	Insertion	-	C	96	110	87.3			
2261816	Deletion	T	-	61	66	92.4	Gene: CBO2124 (Hypothetical protein: no conserved domains detected)	77 del A	Lys26 fs
2333003	SNV	C	T	94	94	100	Gene: CBO2185 (Hypothetical protein: no conserved domains detected)	348 C>T	

2411493	SNV	A	G	61	61	100	Gene: CBO2253 (Hydroxyethylthiazole kinase)	509 T>C	Met170Thr
2480917	SNV	T	C	72	73	98.6	Gene: CBO2331 (Prophage: tail fiber protein)	715 A>G	Lys239Glu
2518104	Insertion	-	T	56	60	93.3	Gene: CBO2386 (Prophage: hypothetical protein/Class I aldolase)	168 dup	Leu57 fs
2608531	Insertion	-	A	65	66	98.5	Gene: CBO2472 (Amino acid permease: putative glutamate/gamma-aminobutyrate antiporter)	670 dup	Thr224 fs
2646907	SNV	G	A	41	41	100	Gene: CBO2511 (Primosome assembly protein PriA)	1763 C>T	Ala588Val
2741807	SNV	G	T	43	43	100	Gene: CBO2596 (Hypothetical protein: Putative threonine/serine exporter)	753 G>T	Leu251Phe
2853374	SNV	G	A	31	31	100	Gene: CBO2706 (Asparagine synthetase)	1424 C>T	Pro475Leu
2887469	Insertion	-	T	48	49	98	Gene: *fliS* (Flagellar protein FliS)	303 dup	Asp102 fs
2923837	Insertion	-	A	29	29	100			
3060299	SNV	C	T	55	55	100	Gene: *purF* (Amidophosphoribosyltransferase)	1397 G>A	Gly466Asp
3081166	Deletion	A	-	59	61	96.7	Gene: CBO2891 (Hypothetical protein: molybdenum cofactor cytidylyltransferase)	159 del A	Val54 fs
3088488	SNV	T	A	53	53	100	Gene: CBO2900 (Hypothetical protein: YdcF-like superfamily)	76 A>T	Ile26Phe
3126216	Insertion	-	A	52	70	74.3			
3176312	SNV	A	T	46	49	93.9	Gene: *rluD2* (Pseudouridine synthase)	35 T>A	Ile12Asn
3199689	Insertion	-	T	54	60	90	Gene: CBO3006 (Hypothetical protein: no conserved domains detected)	141 dup	Gly48 fs
3234007	Insertion	-	A	76	77	98.7			
3307378	Insertion	-	T	40	42	95.2			

3351372	SNV	G	A	72	72	100	Gene: *oppB* (Oligopeptide/dipeptide ABC transporter permease)	242 C>T	Ala81Val
3465157	Deletion	C	-	140	148	94.6			
3591693	Insertion	-	C	64	68	94.1			
3672989	Deletion	A	-	55	57	96.5	Gene: CBO3432 (Cell wall hydrolase)	320 del A	Asn107 fs
3724684	SNV	C	T	127	127	100	Gene: *tufB* (Elongation factor Tu)	381 G>A	
3836626	Insertion	-	A	56	58	96.6	Gene: CBO3592 (LysR family transcriptional regulator)	194 dup	Leu65 fs

Table A 12 List of variants detected in *pyrE*⁻ 3 compared to the published reference genome (accession number: NC_009495)

***pyrE*⁻ 3**									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotations**	**Coding region change**	**Amino acid change**
34221	Deletion	A	-	68	72	94.4	Gene: CBO0028 (Sensory box sigma-54 dependent transcriptional regulator)	240 del A	Lys80 fs
85454	SNV	A	T	63	64	98.4	Gene: CBO0052 (Amidohydrolase)	1130 T>A	Val377Glu
130004	SNV	C	A	90	90	100	Gene: *ileS* (isoleucyl-tRNA synthetase)	1617 C>A	
199497	SNV	T	G	82	83	98.8	Gene: CBO0165 (Hypothetical protein: tRNA 2-thiocytidine biosynthesis protein TtcA)	299 A>C	Asn100Thr
331795	Insertion	-	A	29	30	96.7	Gene: CBO0280 (MurR/RpiR family: Transcriptional regulator)	172 dup	Ile58 fs
392621	Insertion	-	T	22	22	100	Gene: CBO0340 (VicK superfamily: two-component sensor kinase)	207 dup	Thr70 fs
520957	Insertion	-	AT	62	65	95.4			
542165	SNV	A	G	72	72	100	Gene: CBO0453 (High-affinity branched-chain amino acid ABC transporter amino acid-binding protein)	261 A>G	
600787	Deletion	A	-	43	43	100	Gene: CBO0508 (Methyl-accepting chemotaxis protein)	1088 del A	Asn363 fs
606873	Insertion	-	A	34	36	94.4	Gene: CBO0511 (ATP-dependent helicase)	1818 dup	Gly607 fs
671367	Insertion	-	T	43	45	95.6	Gene: CBO0571 (Pentapeptide repeat-containing protein)	189 dup	Gly64 fs
860317	Deletion	T	-	61	64	95.3			

1043719	SNV	T	C	49	49	100	Gene: *cbiT* (Precorrin-6Y C5,15-methyltransferase subunit)	366 T>C	
1116813	Insertion	-	T	43	44	97.7	Gene: CBO1014 (Aspartate aminotransferase)	84 dup	Glu29 fs
1243791	SNV	C	T	51	51	100	Gene: CBO1130 (YjiL/HGD-D superfamily: CoA-substrate-specific enzyme activase)	2588 C>T	Pro863Leu
1419605	Deletion	A	-	50	54	92.6	Gene: CBO1304 (Copper-translocating P-type ATPase)	150 del A	Lys50 fs
1466519	Deletion	A	-	29	29	100	Gene: CBO1342 (MurR/RpiR family: DNA-binding transcriptional regulator)	61 del A	Ile21 fs
1476063	Deletion	A	-	36	38	94.7	Gene: CBO1351 (LytB superfamily: Cell-surface protein)	509 del A	Asn170 fs
1613471	Insertion	-	A	36	39	92.3	Gene: CBO1481 (Methyltransferase)	1065 dup	Leu356 fs
1614684	Deletion	A	-	43	46	93.5	Gene: modA (Molybdate-binding periplasmic protein precursor)	186 del A	Lys62 fs
1736091	SNV	T	C	32	32	100	Gene: *glcK* (glucokinase)	212 T>C	Leu71Ser
2068908	Deletion	T	-	53	54	98.2	Gene: *moeA2* (LysR family transcriptional regulator)	208 del A	Thr70 fs
2129278	Insertion	-	T	41	44	93.2	Gene: CBO1991 (PTS system fructose family transporter subunit IIBC, feneral metabolism)	729 dup	Leu244 fs
2129295	Deletion	T	-	44	45	97.8	Gene: CBO1991 (PTS system fructose family transporter subunit IIBC, feneral metabolism)	712 del A	Met238 fs
2140175	Insertion	-	T	34	36	94.4			
2144065	SNV	G	A	48	48	100	Gene: CBO2006 (Xanthine/uracil permease)	104 C>T	Ala35Val
2198989	Insertion	-	C	96	110	87.3			
2261816	Deletion	T	-	61	66	92.4	Gene: CBO2124 (Hypothetical protein: no conserved domains detected)	77 del A	Lys26 fs

2333003	SNV	C	T	94	94	100	Gene: CBO2185 (Hypothetical protein: no conserved domains detected)	348 C>T	
2411493	SNV	A	G	61	61	100	Gene: CBO2253 (Hydroxyethylthiazole kinase)	509 T>C	Met170Thr
2480917	SNV	T	C	72	73	98.6	Gene: CBO2331 (Prophage: tail fiber protein)	715 A>G	Lys239Glu
2518104	Insertion	-	T	56	60	93.3	Gene: CBO2386 (Prophage: hypothetical protein/Class I aldolase)	168 dup	Leu57 fs
2608531	Insertion	-	A	65	66	98.5	Gene: CBO2472 (Amino acid permease: putative glutamate/gamma-aminobutyrate antiporter)	670 dup	Thr224 fs
2646907	SNV	G	A	41	41	100	Gene: CBO2511 (Primosome assembly protein PriA)	1763 C>T	Ala588Val
2741807	SNV	G	T	43	43	100	Gene: CBO2596 (Hypothetical protein: Putative threonine/serine exporter)	753 G>T	Leu251Phe
2802690	SNV	A	T	63	91	69.2	Gene: *fliL* (Flagellar basal body-associated protein)	328 T>A	Leu110Ile
2853374	SNV	G	A	31	31	100	Gene: CBO2706 (Asparagine synthetase)	1424 C>T	Pro475Leu
2887469	Insertion	-	T	48	49	98	Gene: *fliS* (Flagellar protein FliS)	303 dup	Asp102 fs
2923837	Insertion	-	A	29	29	100			
3060299	SNV	C	T	55	55	100	Gene: *purF* (Amidophosphoribosyltransferase)	1397 G>A	Gly466Asp
3081166	Deletion	A	-	59	61	96.7	Gene: CBO2891 (Hypothetical protein: molybdenum cofactor cytidylyltransferase)	159 del A	Val54 fs
3088488	SNV	T	A	53	53	100	Gene: CBO2900 (Hypothetical protein: YdcF-like superfamily)	76 A>T	Ile26Phe
3126216	Insertion	-	A	52	70	74.3			
3176312	SNV	A	T	46	49	93.9	Gene: *rluD2* (Pseudouridine synthase)	35 T>A	Ile12Asn
3234007	Insertion	-	A	76	77	98.7			

3307378	Insertion	-	T	40	42	95.2			
3465157	Deletion	C	-	140	148	94.6			
3591693	Insertion	-	C	64	68	94.1			
3672989	Deletion	A	-	55	57	96.5	Gene: CBO3432 (Cell wall hydrolase)	320 del A	Asn107 fs
3724684	SNV	C	T	127	127	100	Gene: *tufB* (Elongation factor Tu)	381 G>A	
3836626	Insertion	-	A	56	58	96.6	Gene: CBO3592 (LysR family transcriptional regulator)	194 dup	Leu65 fs

Table A 13 List of variants detected in Δ*boNT/A1* 1 compared to the published reference genome (accession number: NC_009495)

Δ*boNT/A1* 1									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotations**	**Coding region change**	**Amino acid change**
84820	SNV	T	C	64	65	98.5			
130004	SNV	C	A	90	90	100	Gene: *ileS* (Isoleucyl-tRNA synthetase)	1617 C>A	
603757	SNV	A	G	103	104	99	Gene: *serS* (Seryl-tRNA synthetase)	551 A>G	Tyr184Cys
779151	Deletion	T	-	38	38	100	Gene: *hsdS* (Type I restriction enzyme S subunit)	780 del T	Phe260 fs
828172	SNV	A	G	42	42	100	Gene: CBO0719 (Methyl-accepting chemotaxis protein)	5 A>G	Asp2Gly
838149	Insertion	-	A	46	50	92	Gene: *chiA* (Chitinase A)	706 dup	Ile236 fs
895995	Deletion	A	-	30	30	100			
945079	Insertion	-	A	67	68	98.5			
976250	Insertion	-	A	89	91	97.8	Gene: CBO0873 (Hypothetical protein: predicted DNA-binding transcriptional regulator YafY)	489 dup	Asp164 fs
1011187	Deletion	A	-	45	47	95.7	Gene: CBO0906 (Two component sensor kinase)	662 del A	Asn221 fs
1011866	SNV	C	T	41	41	100			
1117314	Insertion	-	A	24	26	92.3			
1188145	SNV	A	G	51	51	100	Gene: *feoB* (Ferrous iron transport protein B)	288 T>C	
1417459	SNV	G	A	54	54	100			
1487032	Insertion	-	A	48	50	96	Gene: CBO1361 (Manganese/zinc/iron chelate uptake ABC transporter permease)	520 dup	Ser174 fs
1541595	Deletion	T	-	36	37	97.3	Gene: CBO1414 (ATP-dependent DNA helicase)	249 del T	Phe83 fs

1634573	Insertion	-	A	44	45	97.8	Gene: CBO1503 (FAD/FMN-binding oxidoreductase)	954 dup	Ala319 fs
1792890	Deletion	A	-	76	79	96.2	Gene: CBO1646 (Large repeat generated by the insertion of prophage)		
2365974	SNV	A	G	41	41	100			
2483935	SNV	A	G	53	53	100	Gene: CBO2335 (Hypothetical protein: no conserved domains detected, Misc. feature: prophage)	248 T>C	Leu83Ser
2529607	SNV	T	C	57	57	100	Gene: CBO2402 (Phosphoribose diphosphate:decaprenyl-phosphate phosphoribosyltransferase)	683 A>G	Gln228Arg
2728995	Insertion	-	TA	44	46	95.7	Gene: CBO2584 (Acyl-CoA synthetase)	33-34 dup	Thr12 fs
2927721	Insertion	-	T	80	91	87.9	Gene: *gyrA* (DNA gyrase subunit A)	2868 dup	*957 fs
3115016	Deletion	T	-	73	76	96.1	Gene: CBO2927 (Phage infection protein)	1653 del A	Lys551 fs
3126216	Insertion	-	A	55	76	72.4			
3134274	SNV	C	T	42	42	100	Gene: *ppdK* (Pyruvate phosphate dikinase)	64 G>A	Gly22Arg
3217092	Insertion	-	T	41	44	93.2	Gene: CBO3025 (Hypothetical protein: no conserved domains detected)	31 dup	Thr11 fs
3231859	SNV	G	A	69	69	100	Gene: CBO3042 (Misc. feature: possible remnant of an integrated prophage)	622 C>T	Gln208*
3341788	SNV	G	A	83	83	100	Gene: CBO3140 (Hypothetical protein: no conserved domains detected)	445 G>A	Gly149Arg
3457861	Deletion	T	-	63	68	92.6	Gene: CBO3250 (Hypothetical protein: no conserved domains detected)	294 del A	Ala99 fs
3472648	Insertion	-	T	45	49	91.8	Gene: CBO3268 (Integral membrane transport protein)	1177 dup	Met393 fs
3531415	SNV	A	T	71	74	95.9			
3641967	Deletion	T	-	85	87	97.7	Gene: CBO3401 (Hypothetical protein: no conserved domains detected)	165 del A	Glu56 fs

3724684	SNV	C	T	126	128	98.4	Gene: *tufB* (Elongation factor Tu)	381 G>A
3886591	SNV	C	T	39	39	100		

Table A 14 List of variants detected in Δ*boNT/A1* 2 compared to the published reference genome (accession number: NC_009495)

Δ*boNT/A1* 2									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotations**	**Coding region change**	**Amino acid change**
130004	SNV	C	A	90	90	100	Gene: *ileS* (Isoleucyl-tRNA synthetase)	1617 C>A	
206761	Insertion	-	A	47	55	85.5	Gene: CBO0172 (Ribosomal subunit interface protein)	268 dup	Thr90 fs
209980	Deletion	G	-	21	21	100			
248507	Insertion	-	A	13	16	81.3	Gene: CBO0210 (Malic enzyme)	438 dup	Leu147 fs
334845	Deletion	T	-	24	24	100			
554147	Deletion	A	-	135	145	93.1	Gene: CBO0465 (NUDIX family: hydrolase)	711 del A	Lys237 fs
570086	Deletion	A	-	48	48	100	Gene: CBO0481 (Hypothetical protein: no conserved domains detected)	93 del T	Lys32 fs
640892	Deletion	T	-	19	20	95			
881185	SNV	G	A	65	65	100	Gene: CBO0780 (Sensor histidine kinase)	577 G>A	Gly193Arg
1028572	Insertion	-	T	25	26	96.2	Gene: *hemD* (Uroporphyrinogen III synthase/methyltransferase)	945 dup	Asp316 fs
1224781	Insertion	-	G	19	20	95			
1246898	Insertion	-	A	14	16	87.5	Gene: *opuD* (Choline/carnitine/betaine transporter)	184 dup	Cys62 fs
1325374	Deletion	A	-	43	48	89.6	Gene: *tyrR* (Tyrosyl-tRNA synthetase)		
1335638	Deletion	A	-	53	54	98.1	Gene: CBO1213 (Sensory box-containing diguanylate cyclase; Cell surface adhesion)	339 del A	Lys113 fs
1364760	Insertion	-	T	30	33	90.9	Gene: CBO1247 (Hypothetical protein: no conserved domains detected)	156 dup	Gly53 fs
1427514	Insertion	-	G	19	19	100			

1649593	Insertion	-	A	15	16	93.8	Gene: CBO1515 (Carbohydrate uptake ABC transporter CUT1 family, ATP-binding protein)	1067 dup	Asn356 fs
1667030	SNV	G	A	36	36	100	Gene: CBO1533 (AraC family transcriptional regulator)	696 G>A	
1710271	SNV	T	C	29	29	100	Gene: CBO1580 (Pyridine nucleotide-disulfide oxidoreductase)	1097 T>C	Val366Ala
1736091	SNV	T	C	22	22	100	Gene: *glcK* (Glucokinase)	212 T>C	Leu71Ser
1736348	SNV	G	A	20	20	100	Gene: *glcK* (Glucokinase)	469 G>A	Gly157Arg
2038493	Insertion	-	T	15	17	88.2			
2087885	SNV	T	C	60	61	98.4	Gene: CBO1958 (Secreted carbohydrate-binding protein)	1034 T>C	Val345Ala
2101478	Deletion	T	-	41	43	95.3	Gene: CBO1967 (Sensor histidine kinase)	22 del A	Ser8 fs
2129295	Deletion	T	-	27	27	100	Gene: CBO1991 (PTS system fructose family transporter subunit IIBC; general metabolism)	712 del A	Met238 fs
2148763	SNV	A	T	23	25	92			
2573235	Deletion	A	-	21	21	100	Gene: CBO2439 (Mg chelatase-like protein)	225 del A	Glu76 fs
2741807	Deletion	G	-	22	22	100	Gene: CBO2596 (Hypothetical protein: Putative threonine/serine exporter)	758 del G	Gly253 fs
2804573	Insertion	-	C	25	29	86.2	Gene: CBO2653 (flagellar protein FlbD)	196 dup	Glu66 fs
3104415	SNV	G	A	29	29	100	Gene: *purF* (Amidophosphoribosyltransferase)	686 C>T	Pro229Leu
3126216	Insertion	-	A	37	43	86.0			
3233766	Insertion	-	T	49	50	98.0			
3234007	Insertion	-	A	48	51	94.1			
3267818	Insertion	-	A	18	18	100	Gene: *hpt* (Hypoxanthine phosphoribosyltransferase)	35 dup	Asn12 fs
3503687	Insertion	-	A	36	36	100			

3540627	Deletion	A	-	28	30	93.3	Gene: CBO3321 (Glycoprotease (YeaZ); Protein synthesis, tRNA and rRNA base modification)	691 del A	Ile231 fs
3629557	Insertion	-	T	15	18	83.3	Gene: CBO3388 (PDZ domain of trypsin-like serine proteases, involved in heat-shock response, chaperone function, and apoptosis)	WT with fs, T insertion restores ORF	Asn258Lys
3630200	Deletion	T	-	16	17	94.1	Gene: CBO3388 (PDZ domain of trypsin-like serine proteases, involved in heat-shock response, chaperone function, and apoptosis).	130 del T	Leu90*
3665439	Deletion	T	-	23	24	95.8			
3752302	SNV	C	T	23	26	88.5	Gene: 23S_rRNA		
3872323	Deletion	A	-	20	21	95.2			

Reference position	**Overlapping annotation**	**Insertion/deletion sequence**
209831..209980	Upstream *prfB*	Deletion: TAAATAACAAGTCAGTATGACAGTCTATTTTGCGTCATACTGCGTCAGCAGAATCCTCCGATAG CACTACTAGCGACGGAACCTGCTTCCTTGGCTGACACAAAATATACTGCCATCTTTGACTTGTTA TTTATTTTCATGTGCCTAAAG

Table A 15 List of variants detected in Δ*boNT/A1* 3 compared to the published reference genome (accession number: NC_009495)

Δ*boNT/A1* 3									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotations**	**Coding region change**	**Amino acid change**
130004	SNV	C	A	90	90	100	Gene: *ileS* (Isoleucyl-tRNA synthetase)	1617 C>A	
209980	Deletion	G	-	47	47	100			
1078297	Insertion	-	T	73	75	97.3	Gene: CBO0977 (AraC-type DNA-binding domain)	WT with fs, T insertion restores ORF	Leu189Thr
1244400	SNV	C	T	76	77	98.7	Gene: CBO1130 (CoA-substrate-specific enzyme activase)	3197 C>T	Pro1066Leu
1585597	Insertion	-	T	70	71	98.6			
1667030	SNV	G	A	74	74	100	Gene: CBO1533 (AraC family transcriptional regulator)	696 G>A	
1736091	SNV	T	C	76	76	100	Gene: *glcK* (Glucokinase)	212 T>C	Leu71Ser
2466355	SNV	T	C	37	37	100			
3234007	Insertion	-	A	95	97	97.9			
3267818	Insertion	-	A	76	82	92.7	Gene: *hpt* (Hypoxanthine phosphoribosyltransferase)	35 dup	Asn12 fs
3629557	Insertion	-	T	77	79	97.5	Gene: CBO3388 (PDZ domain of trypsin-like serine proteases, involved in heat-shock response, chaperone function, and apoptosis)	WT with fs, T insertion restores ORF	Asn258Lys
3641879	SNV	C	A	77	78	98.7	Gene: CBO3401 (Hypothetical protein: no conserved domains detected)	253 G>T	Glu85*
3724684	SNV	C	T	56	56	100	Gene: *tufB* (Elongation factor Tu)	381 G>A	

Reference position	Overlapping annotation	Insertion/deletion sequence
209831..209980	Upstream *prfB*	Deletion: TAAATAACAAGTCAGTATGACAGTCTATTTTGCGTCATACTGCGTCAGCAGAATCCTCCGATAG CACTACTAGCGACGGAACCTGCTTCCTTGGCTGACACAAAATATACTGCCATCTTTGACTTGTTA TTTATTTTCATGTGCCTAAAG
1914862^1914863	*mutL*	Insertion: TTTTTTATTTCTTCTTCTTTTAAGTTGTCATTAGCTTTAACTGCGGATTTGCATGCCAGTGTAGCA ATGGCATTATATTTTATTGTAGAAGTTTCTTTAGACTTCATATTTTTTAAATTATAAAGTATATCC ATAAAGAGATTTTCTACATTGGGTTTACCTAGAATTAAAGGTACTTCTTTTATATTTATAGTGTA TTCTCCAAAAGTTTCCACTGAAAAACCTGAATT

VII. Whole genome sequencing and variant detection Chapter 5

WT SBRC genomes compared to published reference genomes

Table A 16 List of variants detected in ATCC 19397 WT compared to the published reference genome (accession number: NC_009697)

Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
2061071	SNV	A	C	47	53	88.7	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	318 T>G	Phe106Leu
3312582	SNV	T	C	69	70	98.6			
3380286	SNV	C	G	58	58	100	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	60	61	98.4	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs

Table A 17 List of variants detected in 62A WT compared to the published reference genome (accession number: NZ_CP028859)

Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
1530	Insertion	-	CTA	35	39	89.7			
8296	SNV	A	G	82	82	100	Gene: DBN47_RS00035	1958 A>G	Glu653Gly
53714	SNV	G	A	101	101	100	Gene: DBN47_RS00295	515 C>T	Ser172Leu
83479	Insertion	-	T	63	70	90	Gene: DBN47_RS00420	928 dup	Arg310 fs
91460	Insertion	-	A	84	95	88.4	Gene: DBN47_RS00465	116 dup	Arg40 fs
92988	Deletion	A	-	79	82	96.3	Gene: *rsmI*	310 del A	Met104 fs
154044	SNV	C	T	87	88	98.9	Gene: DBN47_RS00755	1263 C>T	
160591	SNV	T	C	72	74	97.3	Gene: DBN47_RS00790	1045 T>C	Ser349Pro
160798	SNV	T	C	86	88	97.7	Gene: DBN47_RS00790	1252 T>C	Ser418Pro
173042	SNV	A	G	95	95	100	Gene: DBN47_RS00845	1136 A>G	Lys379Arg
187683	Insertion	-	A	76	83	91.6	Gene: DBN47_RS00940	519 dup	*174 fs
215526	SNV	T	C	63	63	100	Gene: DBN47_RS01075	538 T>C	
275142	SNV	G	A	53	54	98.1	Gene: *rnr* (Ribonucleotide reductase)	968 G>A	Ser323Asn
277714	SNV	G	A	76	76	100			
291881	SNV	G	A	105	106	99.1			
328729	SNV	T	C	72	72	100	Gene: DBN47_RS01590	1365 T>C	
336091	Insertion	-	AT	36	39	92.3	Gene: DBN47_RS01635	421-422 dup	*141 fs
352648	SNV	G	A	73	74	98.7	Gene: DBN47_RS01705	1386 G>A	
384533	Deletion	A	-	61	62	98.4	Gene: DBN47_RS01860	108 del A	Lys36 fs
392587	Deletion	A	-	30	32	93.8			
393082	SNV	G	A	39	39	100	Gene: DBN47_RS01910	298 G>A	Gly100Arg
394197	SNV	C	T	43	43	100	Gene: DBN47_RS01910	1413 C>T	

406577	Insertion	-	A	68	74	91.9			
439891	Insertion	-	A	79	81	97.5	Gene: DBN47_RS02105	1262 dup	Leu422 fs
464012	Deletion	AT	-	45	47	95.7			
487828	SNV	G	A	71	71	100			
561593	SNV	T	C	88	88	100	Gene: DBN47_RS02630	486 T>C	
564374	Deletion	A	-	55	56	98.2	Gene: DBN47_RS02640	784 del A	Arg262 fs
604690	Insertion	-	T	75	80	93.8	Gene: DBN47_RS02860	1589 dup	Asn530 fs
658635	SNV	A	G	56	56	100	Gene: DBN47_RS03100	474 A>G	
664875	SNV	G	C	66	66	100	regulatory: TPP riboswitch		
668757	SNV	T	C	67	67	100	Gene: DBN47_RS03140	548 T>C	Ile183Thr
691161	SNV	A	G	20	20	100	Gene: DBN47_RS03275	1399 A>G	Thr467Ala
720000	SNV	T	C	70	72	97.2	Gene: DBN47_RS03425	420 T>C	
731791	SNV	G	A	41	41	100	Gene: DBN47_RS03475	83 G>A	Ser28Asn
749208	Deletion	A	-	76	76	100	Gene: DBN47_RS03585	670 del A	Ser224 fs
777390	SNV	T	C	61	61	100	Gene: DBN47_RS03720	871 T>C	*291Gln
809133	SNV	T	C	82	82	100	Gene: DBN47_RS03865	1372 A>G	Lys458Glu
811474	SNV	A	G	56	56	100	Gene: *gcvT*	392 A>G	Glu131Gly
821269	SNV	A	G	45	45	100	Gene: DBN47_RS03910	319 A>G	Thr107Ala
850324	SNV	A	G	50	50	100			
856538	Deletion	A	-	55	57	96.5	Gene: DBN47_RS04110	168 del A	Lys56 fs
870352	SNV	G	A	54	54	100	Gene: DBN47_RS04210	205 G>A	Asp69Asn
914859	SNV	G	A	28	28	100	Gene: DBN47_RS04445	276 C>T	
922830	SNV	A	G	48	49	98.0	Gene: *bont*	2203 A>G	Thr735Ala
923091	Insertion	-	TA	36	40	90	Gene: *bont*	2469-2470 dup	*824 fs
930184	SNV	T	G	69	70	98.6	Gene: *brnQ*	16 T>G	Ser6Ala

930207	SNV	G	T	58	58	100	Gene: *brnQ*	39 G>T	
930228	SNV	C	T	56	56	100	Gene: *brnQ*	60 C>T	
930240	SNV	A	C	59	60	98.3	Gene: *brnQ*	72 A>C	
930249	SNV	A	C	59	60	98.3	Gene: *brnQ*	81 A>C	
930252	SNV	C	T	61	61	100	Gene: *brnQ*	84 C>T	
930255	SNV	G	A	58	59	98.3	Gene: *brnQ*	87 G>A	
930264	SNV	T	G	58	58	100	Gene: *brnQ*	96 T>G	Ile32Met
930270	SNV	T	A	55	58	94.8	Gene: *brnQ*	102 T>A	Asn34Lys
930282	MNV	CA	TC	56	56	100	Gene: *brnQ*	114-115 del CA ins TC	Lys39Gln
930288	SNV	C	A	59	59	100	Gene: *brnQ*	120 C>A	
930292	MNV	AT	GC	54	54	100	Gene: *brnQ*	124-125 del AT ins GC	Ile42Ala
930297	SNV	C	A	54	55	98.2	Gene: *brnQ*	129 C>A	
930300	SNV	T	A	55	55	100	Gene: *brnQ*	132 T>A	
930303	SNV	T	A	55	55	100	Gene: *brnQ*	135 T>A	
930312	SNV	G	A	53	53	100	Gene: *brnQ*	144 G>A	
930318	SNV	T	C	52	53	98.1	Gene: *brnQ*	150 T>C	
930328	SNV	C	T	51	51	100	Gene: *brnQ*	160 C>T	
932517	SNV	G	A	69	69	100	Gene: *brnQ*	897 G>A	
959722	SNV	T	C	92	94	97.9	Gene: DBN47_RS04635	174 A>G	Ile58Met
988439	SNV	T	C	66	66	100	Gene: DBN47_RS04815	140 T>C	Leu47Ser
1017100	Insertion	-	A	48	52	92.3	Gene: DBN47_RS04950	661 dup	Ile221 fs
1046819	Insertion	-	A	78	81	96.3	Gene: *cbiD* (Cobalamin biosynthesis)	281 dup	Asn94 fs
1055392	SNV	G	A	68	70	97.1	Gene: *cbiG* (Cobalamin biosynthesis)	167 G>A	Gly56Glu
1077480	SNV	G	A	67	68	98.5	Gene: DBN47_RS05295	489 C>T	

1085212	SNV	G	A	51	51	100			
1088290	SNV	A	G	43	43	100	Gene: DBN47_RS05350	759 A>G	Ile253Met
1098936	SNV	T	C	49	49	100	Gene: DBN47_RS05405	116 A>G	*39Trp
1132009	Deletion	A	-	47	48	97.9			
1205102	Insertion	-	T	57	63	90.5	Gene: DBN47_RS05905	80 dup	Leu27 fs
1213267	SNV	T	C	73	75	97.3	Gene: DBN47_RS05940	25 A>G	Thr9Ala
1217576	Insertion	-	A	60	66	90.9			
1228329	SNV	G	A	20	20	100	Gene: DBN47_RS06010	805 G>A	Val269Met
1230331	SNV	T	C	42	42	100	Gene: DBN47_RS06020	422 T>C	Val141Ala
1238833	SNV	A	G	40	42	95.2	Gene: DBN47_RS06070	654 A>G	
1247274	SNV	G	A	77	77	100			
1278055	Insertion	-	T	48	48	100			
1323010	SNV	T	C	66	67	98.5	Gene: DBN47_RS06440	714 T>C	
1346059	SNV	T	C	42	43	97.7	Gene: DBN47_RS06550	56 A>G	Glu19Gly
1360213	Insertion	-	C	51	59	86.4	Gene: DBN47_RS06645	356 dup	Ile120 fs
1401997	Deletion	A	-	61	63	96.8	Gene: DBN47_RS06895	252 del A	Lys84 fs
1405045	SNV	G	A	64	65	98.5	Gene: DBN47_RS06905	1840 G>A	Glu614Lys
1405836	Insertion	-	T	58	63	92.1	Gene: DBN47_RS06910	128 dup	Leu43 fs
1430351	Deletion	A	-	33	35	94.3	Gene: DBN47_RS07000	508 del A	Ser170 fs
1519307	Deletion	T	-	59	59	100			
1532147	SNV	A	G	53	53	100	Gene: DBN47_RS07575	397 A>G	Lys133Glu
1539700	Deletion	A	-	62	64	96.9	Gene: DBN47_RS07610	792 del A	Lys264 fs
1555377	Deletion	A	-	61	62	98.4	Gene: DBN47_RS07685	12 del A	Val5 fs
1610790	SNV	T	C	66	67	98.5	Gene: DBN47_RS07950	1141 A>G	Arg381Gly
1618151	SNV	G	A	63	64	98.4	Gene: *nadC*	771 C>T	
1634112	SNV	T	C	58	59	98.3	Gene: DBN47_RS08035	625 T>C	Ser209Pro

1666451	SNV	T	C	43	45	95.6	Gene: DBN47_RS08205	450 T>C	
1667369	SNV	T	C	55	56	98.2	Gene: DBN47_RS08210	21 A>G	
1674893	SNV	T	C	63	63	100	Gene: DBN47_RS08250	663 T>C	
1724285	SNV	A	G	41	42	97.6	Gene: DBN47_RS08480	201 T>C	
1747171	Deletion	A	-	60	62	96.8	Gene: DBN47_RS08615	76 del A	Arg26 fs
1750401	SNV	A	G	60	61	98.4	Gene: DBN47_RS08625	115 A>G	Thr39Ala
1769858	SNV	G	A	48	48	100			
1778639	SNV	T	C	64	64	100	Gene: DBN47_RS08740	602 T>C	Leu201Pro
1784026	SNV	T	C	43	43	100			
1812816	SNV	T	C	63	63	100	Gene: *lpdA*	1028 T>C	Val343Ala
1830627	SNV	T	C	52	52	100	Gene: DBN47_RS08950	209 T>C	Val70Ala
1839758	Deletion	A	-	71	74	95.9	Gene: DBN47_RS09010	631 del A	Ile211 fs
1850878	SNV	T	C	87	89	97.8	Gene: DBN47_RS09055	26 T>C	Leu9Pro
1852678	Insertion	-	T	54	57	94.7			
1854772	SNV	G	A	36	36	100	Gene: DBN47_RS09070	1845 G>A	
1881284	SNV	A	G	59	60	98.3	Gene: DBN47_RS09205	347 T>C	Val116Ala
1888397	SNV	G	A	38	38	100	Gene: DBN47_RS09235	541 G>A	Val181Met
1892314	Deletion	A	-	38	38	100	Gene: DBN47_RS09240	2865 del A	Lys955 fs
1897925	SNV	T	A	59	59	100	Gene: DBN47_RS09245	3613 T>A	Cys1205Ser
1921051	SNV	T	C	21	21	100	Gene: DBN47_RS09360	700 A>G	Ser234Gly
1940110	SNV	G	A	47	47	100	Gene: DBN47_RS09435	35 C>T	Thr12Ile
1946069	SNV	A	G	33	33	100	Gene: DBN47_RS09480	115 T>C	Phe39Leu
1963018	Insertion	-	A	67	71	94.4			
1966919	SNV	G	A	54	55	98.2	Gene: DBN47_RS09585	1076 C>T	Thr359Met
1969942	SNV	C	T	59	59	100	Gene: DBN47_RS09595	459 G>A	
1970154	SNV	T	C	67	67	100	Gene: DBN47_RS09595	247 A>G	Ile83Val

2020129	SNV	C	T	24	24	100	Gene: DBN47_RS09865	388 G>A	Asp130Asn
2034990	SNV	T	C	63	64	98.4	Gene: DBN47_RS09950	1206 T>C	
2037873	Deletion	T	-	47	50	94	Gene: DBN47_RS09960	305 del A	Asn102 fs
2039076	SNV	T	C	56	56	100	Gene: *cloSI*	1146 A>G	
2041345	Insertion	-	G	35	38	92.1	Gene: DBN47_RS09970	651 dup	Asn218 fs
2052031	SNV	C	T	22	22	100	Gene: DBN47_RS10025	1556 G>A	Arg519Gln
2060752	Deletion	TT	-	69	71	97.2	Gene: DBN47_RS10060	1505-1506 del AA	Lys502 fs
2068708	SNV	A	G	65	65	100	Gene: DBN47_RS10100	652 T>C	*218Gln
2079568	Deletion	T	-	47	50	94.0	Gene: DBN47_RS10145	30 del A	Val11 fs
2079820	Insertion	-	A	57	62	91.9	Gene: DBN47_RS10150	115 dup	Met39 fs
2091810	SNV	T	A	79	80	98.8			
2103126	SNV	A	G	37	37	100	Gene: DBN47_RS10245	198 T>C	
2107879	Insertion	-	T	22	24	91.7	Gene: DBN47_RS10265	290 dup	Ser98 fs
2109598	SNV	A	G	50	52	96.2	Gene: DBN47_RS10270	928 A>G	Lys310Glu
2113556	SNV	C	T	62	62	100	Gene: DBN47_RS10300	360 C>T	
2121906	SNV	A	G	65	65	100	Gene: DBN47_RS10320	1237 A>G	Thr413Ala
2129708	SNV	T	G	60	60	100	Gene: DBN47_RS10355	95 A>C	His32Pro
2147077	SNV	C	T	62	62	100	Gene: DBN47_RS10435	825 C>T	
2155285	Deletion	T	-	39	40	97.5	Gene: *larA*	841 del A	Ile281 fs
2169407	Deletion	A	-	81	83	97.6			
2173809	SNV	G	A	41	41	100	Gene: DBN47_RS10570	1872 C>T	
2193497	SNV	A	G	56	57	98.2	Gene: DBN47_RS10645	985 A>G	Ile329Val
2196046	Insertion	-	C	60	67	89.6			
2201937	Insertion	-	T	66	71	93.0	Gene: DBN47_RS10705	285 dup	Cys96 fs
2204948	SNV	C	T	61	62	98.4	Gene: DBN47_RS10725	430 G>A	Gly144Ser

2218553	SNV	T	C	45	45	100	Gene: DBN47_RS10810	2832 A>G	
2256380	SNV	C	T	61	61	100	Gene: DBN47_RS10990	195 G>A	
2259007	Insertion	-	TA	59	64	92.2			
2274894	SNV	T	A	57	59	96.6	Gene: DBN47_RS11085	1087 A>T	Ile363Leu
2278876	Deletion	T	-	66	72	91.7	Gene: DBN47_RS11095	47 del A	Lys16 fs
2279609	Deletion	T	-	65	71	91.5	Gene: DBN47_RS11100	699 del A	Ala234 fs
2326962	Insertion	-	T	77	79	97.5	Gene: DBN47_RS11310	525 dup	Tyr176 fs
2347983	SNV	C	T	59	59	100	Gene: DBN47_RS11385	128 G>A	Cys43Tyr
2361541	SNV	T	C	74	74	100	Gene: DBN47_RS11435	176 A>G	Asn59Ser
2380991	Deletion	T	-	63	72	87.5	Gene: *cooS* (Carbon monoxide dehydrogenase)	578 del A	Lys193 fs
2382866	SNV	A	G	53	53	100	Gene: *hypB*	314 A>G	Asn105Ser
2388633	SNV	C	T	64	64	100			
2403084	SNV	A	G	73	73	100	Gene: DBN47_RS11620	557 T>C	Ile186Thr
2427760	SNV	A	G	67	68	98.5	Gene: DBN47_RS11770	84 T>C	
2428020	Deletion	G	-	58	58	100			
2451735	SNV	T	C	58	58	100	Gene: DBN47_RS11915	763 A>G	Thr255Ala
2469684	SNV	C	T	54	54	100	Gene: DBN47_RS12020	160 G>A	Val54Ile
2488477	SNV	T	C	61	61	100	Gene: DBN47_RS12110	1270 A>G	Thr424Ala
2503307	SNV	C	T	64	65	98.5	Gene: *polC*	1705 G>A	Ala569Thr
2503780	SNV	C	T	77	77	100	Gene: *polC*	1232 G>A	Gly411Asp
2513623	Insertion	-	T	73	77	94.8			
2543138	SNV	T	C	86	86	100	Gene: *prdB*	10 A>G	Ile4Val
2547157	Deletion	T	-	104	108	96.3			
2595584	SNV	C	T	41	41	100	Gene: DBN47_RS12650	516 G>A	
2604039	Insertion	-	A	62	66	93.9			

2662734	SNV	T	C	66	67	98.5	Gene: DBN47_RS13025	48 A>G	*16Trp
2720580	SNV	C	T	66	66	100	Gene: DBN47_RS13275	1035 G>A	
2732486	SNV	A	G	54	54	100	Gene: DBN47_RS13330	279 T>C	
2751420	SNV	T	A	78	78	100	Gene: DBN47_RS13425	590 A>T	Glu197Val
2778776	Insertion	-	A	18	18	100	Gene: DBN47_RS13575	524 dup	Leu175 fs
2783516	SNV	A	G	45	45	100	Gene: DBN47_RS13600	371 T>C	Leu124Pro
2829233	SNV	C	T	73	73	100	Gene: DBN47_RS13810	489 G>A	
2907975	SNV	T	G	48	48	100	Gene: DBN47_RS14160	385 A>C	Ile129Leu
2909319	SNV	T	C	51	52	98.1	Gene: DBN47_RS14165	128 A>G	*43Trp
2910209	SNV	G	C	63	64	98.4	Gene: DBN47_RS14170	715 C>G	Gln239Glu
2913108	SNV	A	G	71	71	100	Gene: DBN47_RS14190	575 T>C	Leu192Pro
2918346	Deletion	T	-	53	54	98.1			
2921204	SNV	G	A	63	63	100	Gene: *glsA* (Glutaminase)	218 C>T	Thr73Ile
2942193	SNV	C	T	37	38	97.4	Gene: DBN47_RS14310	1863 G>A	Met621Ile
2943970	Deletion	T	-	44	46	95.7	Gene: DBN47_RS14310	86 del A	Asn29 fs
2950667	SNV	A	G	64	65	98.5	Gene: DBN47_RS14350	181 A>G	Asn61Asp
2954771	Insertion	-	T	52	59	88.1			
2991277	SNV	T	C	71	71	100	Gene: DBN47_RS14530	999 A>G	
2991438	SNV	C	T	91	91	100	Gene: DBN47_RS14530	838 G>A	Val280Met
3003514	SNV	A	G	42	42	100	Gene: DBN47_RS14600	606 T>C	
3005731	SNV	A	G	29	29	100	Gene: DBN47_RS14610	101 T>C	Val34Ala
3007083	Deletion	T	-	67	69	97.1	Gene: DBN47_RS14615	557 del A	Lys186 fs
3022486	Deletion	T	-	64	66	97.0	Gene: *mgtA* (Magnesium transporter A)	777 del A	Lys259 fs
3037793	SNV	T	C	72	72	100	Gene: DBN47_RS14735	748 A>G	Thr250Ala
3043216	SNV	C	T	94	94	100	Gene: *hydA* (Ferredoxin hydrogenase)	1341 G>A	
3050558	Insertion	-	A	56	62	90.3	Gene: *mocA*	158 dup	Phe54 fs

3061683	SNV	C	T	92	94	97.9	Gene: DBN47_RS14830	5 G>A	Gly2Asp
3063070	Deletion	T	-	51	53	96.2	Gene: DBN47_RS14835	1860 del A	Lys620 fs
3063931	SNV	A	G	54	54	100	Gene: DBN47_RS14835	999 T>C	
3083533	SNV	A	G	69	69	100	Gene: DBN47_RS14935	138 T>C	
3104272	SNV	A	G	71	71	100	Gene: DBN47_RS15060	235 T>C	Ser79Pro
3116357	Insertion	-	T	36	39	92.3	Gene: DBN47_RS15150	252 dup	*85 fs
3127697	SNV	G	A	54	55	98.2	Gene: DBN47_RS15230	412 C>T	His138Tyr
3133402	Deletion	A	-	72	75	96	Gene: DBN47_RS15275	16 del A	Met6 fs
3137236	SNV	C	T	68	68	100	Gene: DBN47_RS15295	94 G>A	Val32Ile
3138620	SNV	A	G	66	66	100	Gene: DBN47_RS15305	833 T>C	Leu278Pro
3139768	SNV	A	G	85	86	98.8	Gene: DBN47_RS15310	1482 T>C	
3141528	Deletion	T	-	77	79	97.5			
3212289	Deletion	T	-	68	70	97.1	Gene: DBN47_RS15655	429 del A	Lys143 fs
3243886	Deletion	T	-	48	49	98.0	Gene: DBN47_RS15850	718 del A	Arg240 fs
3245644	SNV	C	T	56	57	98.2	Gene: DBN47_RS15865	282 C>T	
3246351	Deletion	A	-	69	74	93.2			
3263876	Insertion	-	AT	63	71	88.7	Gene: *secD* (Protein-export membrane protein)	557-558 dup	Gly187 fs
3287471	Deletion	T	-	40	45	88.9			
3297926	SNV	C	T	44	44	100	Gene: DBN47_RS16145	144 G>A	
3312258	Insertion	-	AT	55	57	96.5	Gene: DBN47_RS16205	978-979 dup	*327 fs
3328645	Deletion	TA	-	53	58	91.4			
3366052	SNV	C	T	69	69	100	Gene: DBN47_RS16430	389 G>A	Cys130Tyr
3372663	SNV	A	G	75	76	98.7			
3391129	SNV	A	G	57	57	100	Gene: DBN47_RS16530	2939 T>C	Val980Ala
3401854	SNV	A	G	77	80	96.3	Gene: DBN47_RS16575	499 A>G	Ser167Gly
3408923	SNV	A	G	52	54	96.3	Gene: DBN47_RS16610	673 T>C	Cys225Arg

3466046	SNV	C	T	81	81	100	Gene: DBN47_RS16890	285 G>A	
3488709	SNV	G	A	97	98	99.0	Gene: DBN47_RS17135	888 C>T	
3514394	SNV	C	T	64	64	100	Gene: DBN47_RS17250	1228 G>A	Glu410Lys
3533221	SNV	C	T	72	72	100	Gene: DBN47_RS17320	555 G>A	
3560380	Insertion	-	T	53	56	94.6			
3595037	SNV	C	T	51	51	100	Gene: DBN47_RS17535	1273 G>A	Asp425Asn
3606031	SNV	A	G	53	54	98.1	Gene: DBN47_RS17585	1427 T>C	Val476Ala
3614278	SNV	C	T	80	80	100	Gene: DBN47_RS17615	535 G>A	Asp179Asn
3645325	Insertion	-	A	40	44	90.9	Gene: DBN47_RS17735	775 dup	Ser259 fs
3670192	Insertion	-	T	54	58	93.1	Gene: DBN47_RS17855	1192 dup	Ile398 fs
3715940	SNV	T	C	68	68	100	Gene: DBN47_RS18050	38 A>G	Asn13Ser
3720152	Deletion	A	-	63	64	98.4	Gene: DBN47_RS18070	121 del T	*41 fs
3725529	SNV	A	G	54	54	100	Gene: DBN47_RS18110	688 T>C	Tyr230His
3734057	SNV	C	T	85	86	98.8	Gene: DBN47_RS18195	420 G>A	
3743334	SNV	G	A	68	72	94.4	Gene: *fusA* (Elongation factor G)	801 C>T	
3749262	SNV	T	A	85	87	97.7	Gene: *rpoB* (β subunit RNA polymerase)	3642 A>T	Glu1214Asp
3763755	SNV	A	G	75	75	100	Gene: DBN47_RS18380	361 T>C	Tyr121His
3800000	SNV	A	G	81	83	97.6	Gene: DBN47_RS18540	200 T>C	Leu67Pro
3806505	Insertion	-	AT	62	74	83.8	Gene: *spoIIE*	253-254 dup	Phe86 fs
3815306	Deletion	T	-	74	79	93.7			
3825817	SNV	C	T	41	41	100	Gene: DBN47_RS18690	49 G>A	Val17Met
3825867	SNV	G	A	40	40	100			
3859158	Insertion	-	TA	68	75	90.7	Gene: DBN47_RS18875	200-201 dup	Arg68 fs
3880810	Insertion	-	A	62	69	89.9	Gene: DBN47_RS18985	425 dup	Leu142 fs
3884207	SNV	A	G	74	74	100	Gene: DBN47_RS19000	687 A>G	
3903194	Deletion	A	-	85	87	97.7			

3910299	SNV	C	T	32	32	100	Gene: DBN47_RS19190	114 G>A	
3919188	SNV	A	G	48	48	100	Gene: DBN47_RS19240	53 T>C	Leu18Pro

Light blue: NCTC probably parent WT allele (based on genome sequence ATCC 3502 and 19397; and if ORF is actually corrected/restored or broken)
Dark blue: NCBI probably parent WT allele (based on genome sequence ATCC 3502 and 19397; and if ORF is actually corrected/restored or broken)
Light grey: probably sequencing/alignment error (brnQ is dublicated in genome, both genes are slightly different in nucleotide sequence)
Light red: unclear if NCBI or NCTC is the parent WT allele
White: not checked (only mutations causing fs and genes with name annotation were checked)

Table A 18 List of variants detected in Loch Maree WT compared to the published reference genome (accession number: NC_010520.1)

Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
400153	MNV	TA	AG	12	12	100	Gene: CLK_RS01880 (Sensor histidine kinase)	1449_1450 del TA ins AG	Ile484Val
400156	SNV	T	A	12	12	100	Gene: CLK_RS01880 (Sensor histidine kinase)	1452 T>A	
400166	SNV	C	T	11	11	100	Gene: CLK_RS01880 (Sensor histidine kinase)	1462 C>T	Leu488Phe
400168	SNV	T	A	11	11	100	Gene: CLK_RS01880 (Sensor histidine kinase)	1464 T>A	
400171	SNV	A	T	11	11	100	Gene: CLK_RS01880 (Sensor histidine kinase)	1467 A>T	
400174	MNV	CT	TG	11	11	100	Gene: CLK_RS01880 (Sensor histidine kinase)	1470_1471 del CT ins TG	Ser491Ala
403559	SNV	G	A	10	10	100	Gene: CLK_RS01900 (Sensor histidine kinase)	1387 G>A	Asp463Asn
403564	SNV	T	G	13	13	100	Gene: CLK_RS01900 (Sensor histidine kinase)	1392 T>G	
403570	SNV	A	G	13	14	92.9	Gene: CLK_RS01900 (Sensor histidine kinase)	1398 A>G	
403573	SNV	G	A	13	14	92.9	Gene: CLK_RS01900 (Sensor histidine kinase)	1401 G>A	
403576	SNV	T	C	15	15	100	Gene: CLK_RS01900 (Sensor histidine kinase)	1404 T>C	
403579	SNV	G	T	15	15	100	Gene: CLK_RS01900 (Sensor histidine kinase)	1407 G>T	Met469Ile
403582	SNV	T	C	15	15	100	Gene: CLK_RS01900 (Sensor histidine kinase)	1410 T>C	
403585	SNV	G	A	16	16	100	Gene: CLK_RS01900 (Sensor histidine kinase)	1413 G>A	
403589	SNV	C	A	16	16	100	Gene: CLK_RS01900 (Sensor histidine kinase)	1417 C>A	Pro473Thr
1805495	SNV	G	T	71	72	98.6			
1808085	SNV	G	T	55	55	100	Gene: CLK_RS08380 (LysR family transcriptional regulator)	658 C>A	Leu220Ile
1830764	Deletion	TTATCTT	-	19	24	79.2	Gene: CLK_RS08465 (ABC transporter permease)	470_476 del CTTTTAT	Ser157 fs
2224669	SNV	A	G	48	48	100	Gene: CLK_RS10250 (Hypothetical protein - spore germination B3/GerAC like)	274 A>G	Thr92Ala

2245534	SNV	G	A	48	48	100	Gene: CLK_RS10330 (RpiR family transcriptional regulator)	245 C>T	Ala82Val
2876468	Insertion	-	T	61	62	98.4	Gene: CLK_RS13300 (Flagellar biosynthesis protein FlgC)	203 dup	Asn68 fs
3580345	SNV	G	A	68	68	100	Gene: CLK_RS16590/brnQ (Branched-chain amino acid transport protein)	854 C>T	Ala285Val

Light blue: NCTC probably parent WT allele (ORF is corrected/restored)

Light grey: low coverage

Table A 19 List of variants detected in NCTC 2916 WT compared to the published reference genome (accession number: NZ_ABDO00000000.2)

Contig	Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
7	241058	SNV	G	A	87	149	58.4	Gene: *ltrA* (CBN_2340) (Group II intron reverse transcriptase/maturase)	818 G>A	Cys273Tyr
7	241436	Insertion	-	TATCG	31	58	53.4	Gene: *ltrA* (CBN_2340) (Group II intron reverse transcriptase/maturase)	1195_1196 ins TATCG	Lys399 fs
4	190	Replacement	G	CA	13	13	100	Gene: CBN_1988 (Hypothetical protein)	del C ins TG	Ala21 fs
4	193	SNV	G	A	14	14	100	Gene: CBN_1988 (Hypothetical protein)	2 C>T	Ser20Phe
3	112644	Insertion	-	GTATTTTATTG	31	60	51.7			
2	165425	SNV	G	A	104	105	99.0	Gene: CBN_3421 (Site-specific recombinase, phage integrase family)	320 C>T	Ser107Phe
2	165790	Deletion	G	-	31	55	56.4			
2	165952	Insertion	-	C	44	68	64.7			
1	798852	SNV	A	G	76	77	98.7			

Light blue: NCTC probably parent WT allele (ORF is corrected/restored)

Light red: unclear if NCBI or NCTC is the parent WT allele

Light grey: low coverage

Safe strain genome sequences compared to published reference genome and WT SBRC genome

Table A 20 List of variants detected in ATCC 19397 Δ*boNT/A1* compared to the published reference genome (accession number: NC_009697)

Δ*boNT/A1* biological replicate 1									
Reference Position	**Type**	**Reference**	**Allele**	**Count**	**Coverage**	**Frequency**	**Overlapping annotations**	**Coding region change**	**Amino acid change**
2061071	SNV	A	C	54	54	100	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	318 T>G	Phe106Leu
2061097	SNV	C	T	52	52	100	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	292 G>A	Glu98Lys
2389771	SNV	T	A	56	57	98.3	Gene: CLB_RS11150 (Membrane protein)	410 T>A	Phe137Tyr
2984334	SNV	C	A	49	49	100	Gene: CLB_RS14020 (Magnesium-translocating P-type ATPase)	2485 G>T	Ala829Ser
3312582	SNV	T	C	75	75	100			
3380286	SNV	C	G	67	67	100	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	61	61	100	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs
3799877	SNV	G	A	71	71	100	Gene: CLB_RS18085 (YjjW famile glycyl radical enzyme activase)	341 G>A	Gly114Asp
Δ*boNT/A1* biological replicate 2									
Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
2061071	SNV	A	C	57	57	100	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	318 T>G	Phe106Leu
2061097	SNV	C	T	59	59	100	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	292 G>A	Glu98Lys
3312582	SNV	T	C	59	59	100			

3380286	SNV	C	G	43	44	97.7	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	47	47	100	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs
Δ*boNT/A1* biological replicate 3									
Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
982949	SNV	C	A	68	68	100	Gene: CLB_RS04645 (HAMP-domain containing histidine kinase)	90 C>A	
3312582	SNV	T	C	79	79	100			
3380286	SNV	C	G	53	53	100	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	60	60	100	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs

Yellow: present in WT

Table A 21 List of variants detected in ATCC 19397 ΔClA compared to the published reference genome (accession number: NC_009697)

ΔClA biological replicate 1									
Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
2061071	SNV	A	C	61	61	100	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	318 T>G	Phe106Leu
3312582	SNV	T	C	66	66	100			
3380286	SNV	C	G	67	68	98.5	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	75	76	98.7	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs
ΔClA biological replicate 2									
Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
686745	SNV	C	A	80	80	100	Gene: CLB_RS03185 (ATP-binding protein)	772 G>T	Gly258Cys
1003068	SNV	A	C	61	122	50.0			
2061071	SNV	A	C	60	60	100	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	318 T>G	Phe106Leu
3312582	SNV	T	C	67	68	98.5			
3380286	SNV	C	G	77	77	100	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	72	72	100	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs
ΔClA biological replicate 3									
Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change

165266	SNV	C	A	29	77	37.7	Gene: CLB_RS00805 (Transcription termination factor Rho)	413 C>A	Ser138Tyr
1003068	SNV	A	C	54	115	47.0			
1563755	SNV	A	C	83	83	100	Gene: CLB_RS07355	48 A>C	
2061071	SNV	A	C	61	61	100	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	318 T>G	Phe106Leu
2061145	SNV	C	A	57	58	98.3	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	244 G>T	Val82Leu
3312582	SNV	T	C	76	77	98.7			
3380286	SNV	C	G	71	72	98.6	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	53	55	96.4	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs

Yellow: present in WT

Table A 22 List of variants detected in ATCC 19397 i*boNT/A1*^{RY} compared to the published reference genome (accession number: NC_009697)

Toxoid i*boNT/A1*^{RY} biological replicate 1									
Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
883088	SNV	G	A	89	89	100	Gene: *bont*	1053 G>A	
883122	MNV	AGA	GCT	82	82	100	Gene: *bont*	1087-1089 del/ins GCT	Arg363Ala
883132	SNV	A	T	82	82	100	Gene: *bont*	1097 A>T	Tyr366Phe
3312582	SNV	T	C	72	72	100			
3380286	SNV	C	G	72	72	100	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	62	63	98.4	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs
Toxoid i*boNT/A1*^{RY} biological replicate 2									
Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
883088	SNV	G	A	94	94	100	Gene: *bont*	1053 G>A	
883122	MNV	AGA	GCT	88	88	100	Gene: *bont*	1087-1089 del/ins GCT	Arg363Ala
883132	SNV	A	T	88	88	100	Gene: *bont*	1097 A>T	Tyr366Phe
2061071	SNV	A	C	56	56	100	Gene: CLB_RS09580 (MurR/RpiR family transcriptional regulator)	318 T>G	Phe106Leu
3312582	SNV	T	C	71	71	100			
3380286	SNV	C	G	69	69	100	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	69	70	98.6	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs

Toxoid i*boNT/A1*^{RY} biological replicate 3									
Reference Position	Type	Reference	Allele	Count	Coverage	Frequency	Overlapping annotations	Coding region change	Amino acid change
883088	SNV	G	A	75	76	98.7	Gene: *bont*	1053 G>A	
883122	MNV	AGA	GCT	73	74	98.6	Gene: *bont*	1087-1089 del/ins GCT	Arg363Ala
883132	SNV	A	T	76	76	100	Gene: *bont*	1097 A>T	Tyr366Phe
1644272	SNV	G	A	77	78	98.7	Gene: CLB_RS07730 (peptide MFS transporter)	417 G>A	
3312582	SNV	T	C	70	70	100			
3380286	SNV	C	G	67	68	98.5	Gene: CLB_RS15955 (Sigma-54-dependent Fis family transcriptional regulator)	1118 G>C	Arg373Thr
3465907	Insertion	-	A	69	72	95.8	Gene: CLB_RS16495 (Lactate dehydrogenase)	31 dup	Tyr11 fs

Yellow: present in WT

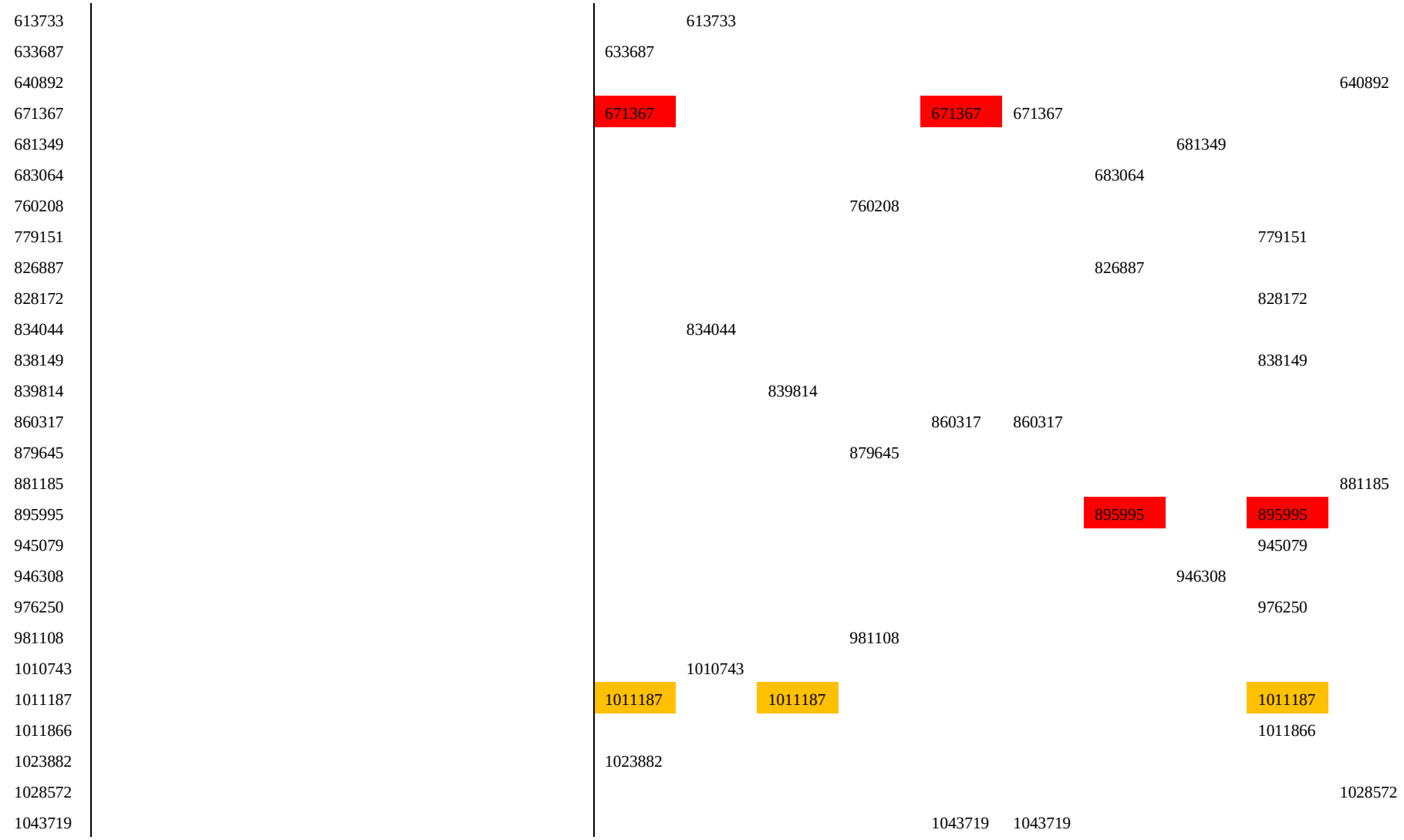
Green: anticipated change

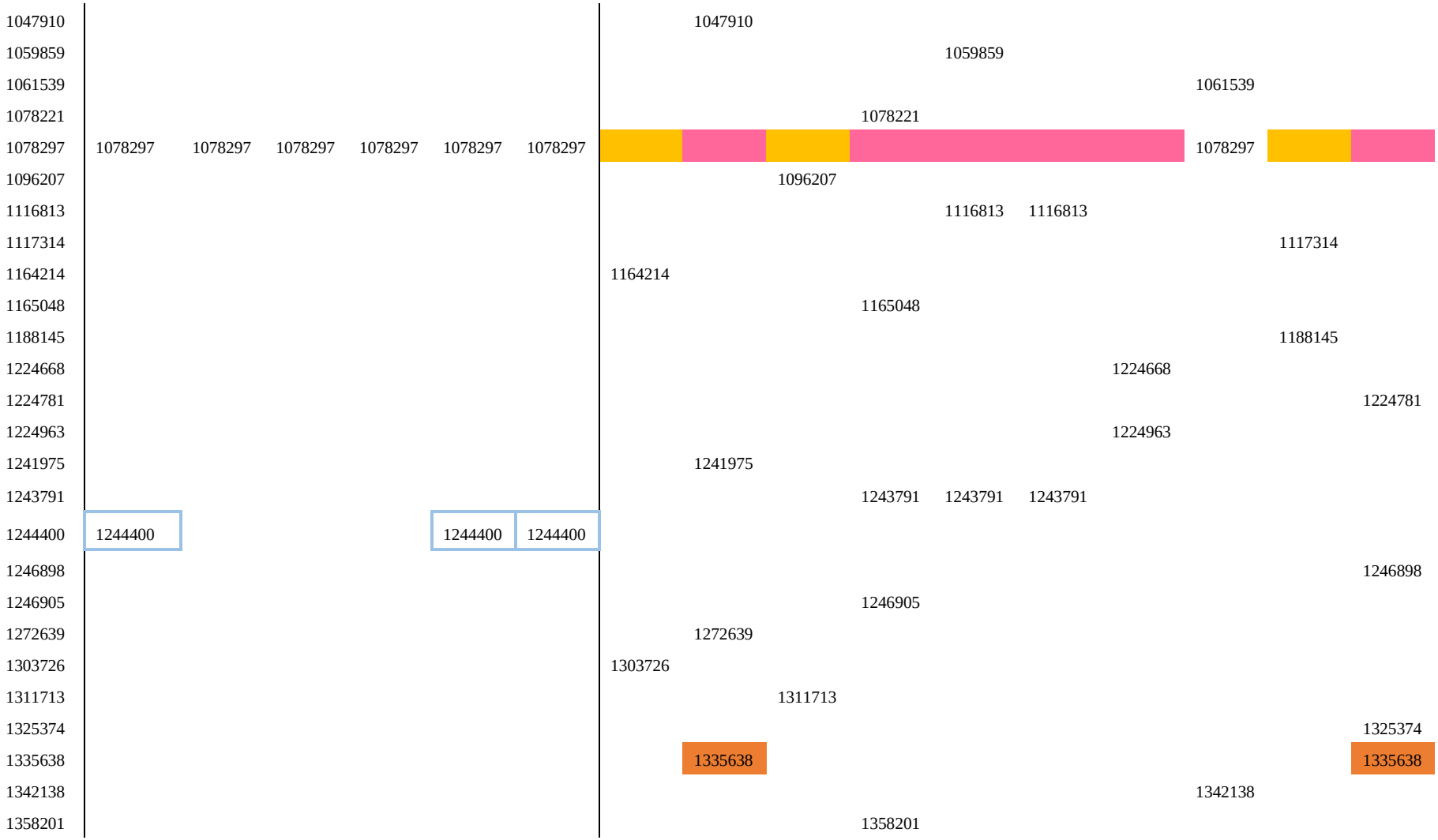
VIII. SNPs/SNVs comparison table for ATCC 3502

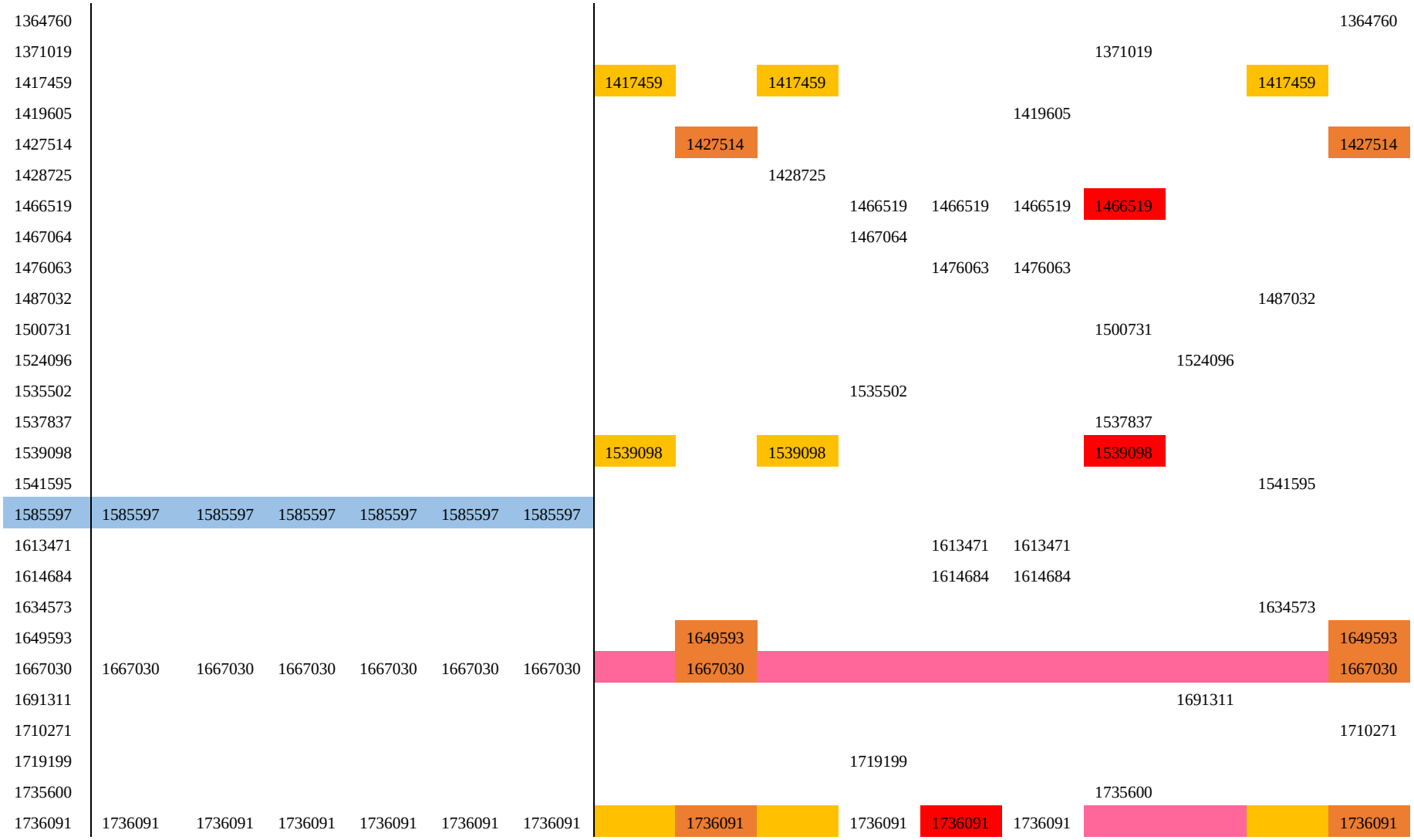
Table A 23 List of all variants detected in ATCC 3502 strains used in this study

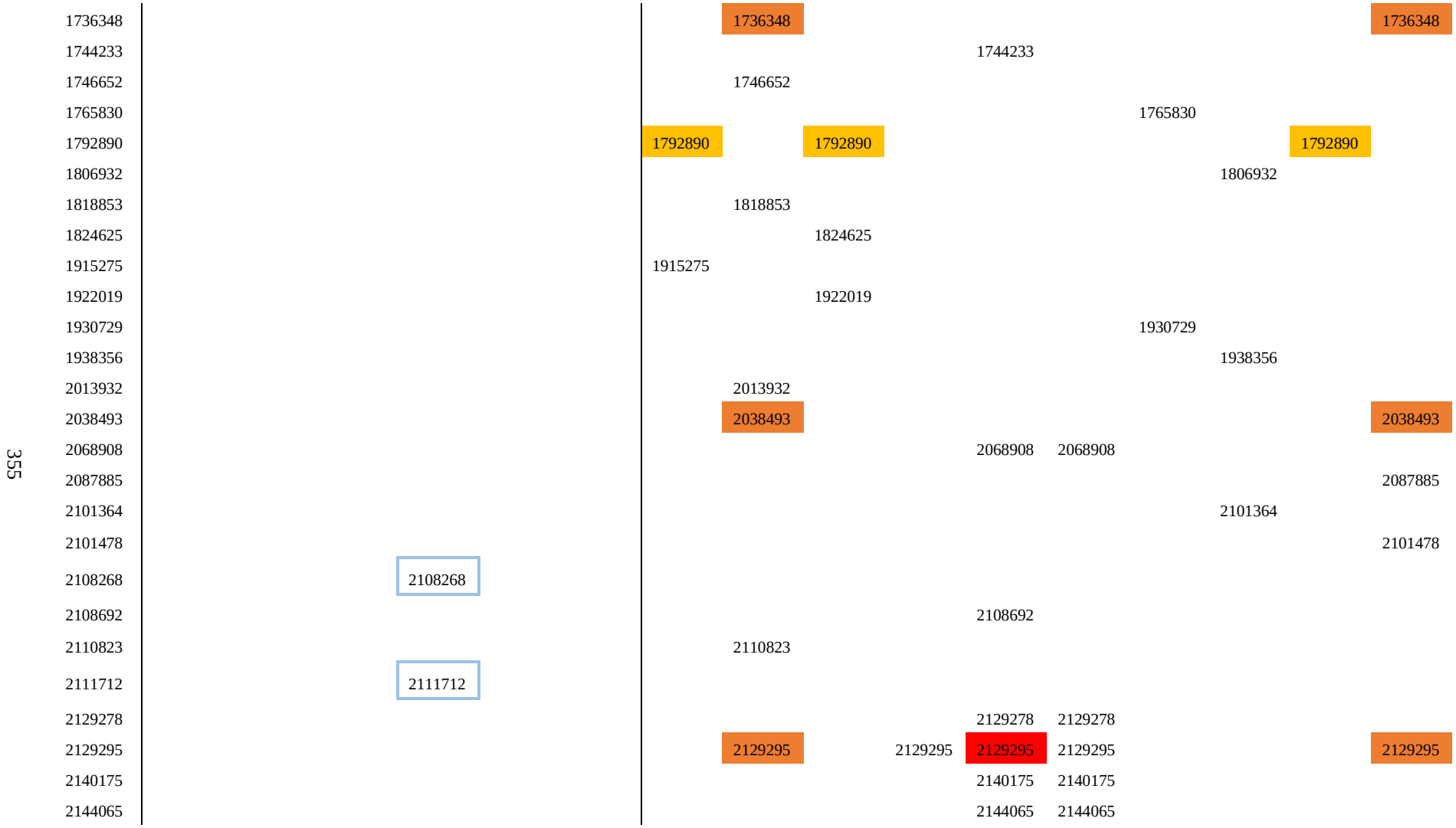
Position	+ ATCC 3502^{+} 1	+ ATCC 3502^{+} 2	+ ATCC 3502^{+} 3	+ ATCC 3502^{+} 4	+ *hsdR*1600s::CT 2	+ Δ*boNT/A1* 3	- ATCC 3502^{-} 1	- ATCC 3502^{-} 2	- ATCC 3502^{-} 3	- *pyrE*$^{-}$ 1	- *pyrE*$^{-}$ 2	- *pyrE*$^{-}$ 3	- *hsdR*1600s::CT 1	- *hsdR*1600s::CT 3	- Δ*boNT/A1* 1	- Δ*boNT/A1* 2
4174				4174												
34221											34221	34221				
41154								41154								
84820															84820	
85454											85454	85454				
122731									122731							
130004	130004	130004	130004	130004	130004	130004	130004	130004	130004	130004	130004	130004	130004	130004	130004	130004
148878									148878							
148885									148885							
148891									148891							
152849							152849									
196143														196143		
199497											199497	199497				
202999										202999						
206761																206761
209980	209980				209980	209980		209980								209980

244279									244279		
248507											248507
275721			275721								
278091			278091								
314422					314422						
331795						331795	331795				
334845											334845
385352								385352			
392621					392621	392621	392621				
392977								392977			
393048	393048										
413266		413266									
424790									424790		
520957						520957	520957				
524885								524885			
542165						542165	542165				
543190		543190									
552405					552405						
554147											554147
570086			570086								570086
589301								589301			
590039			590039								
598476				598476							
600787						600787	600787				
603757										603757	
606873						606873	606873				

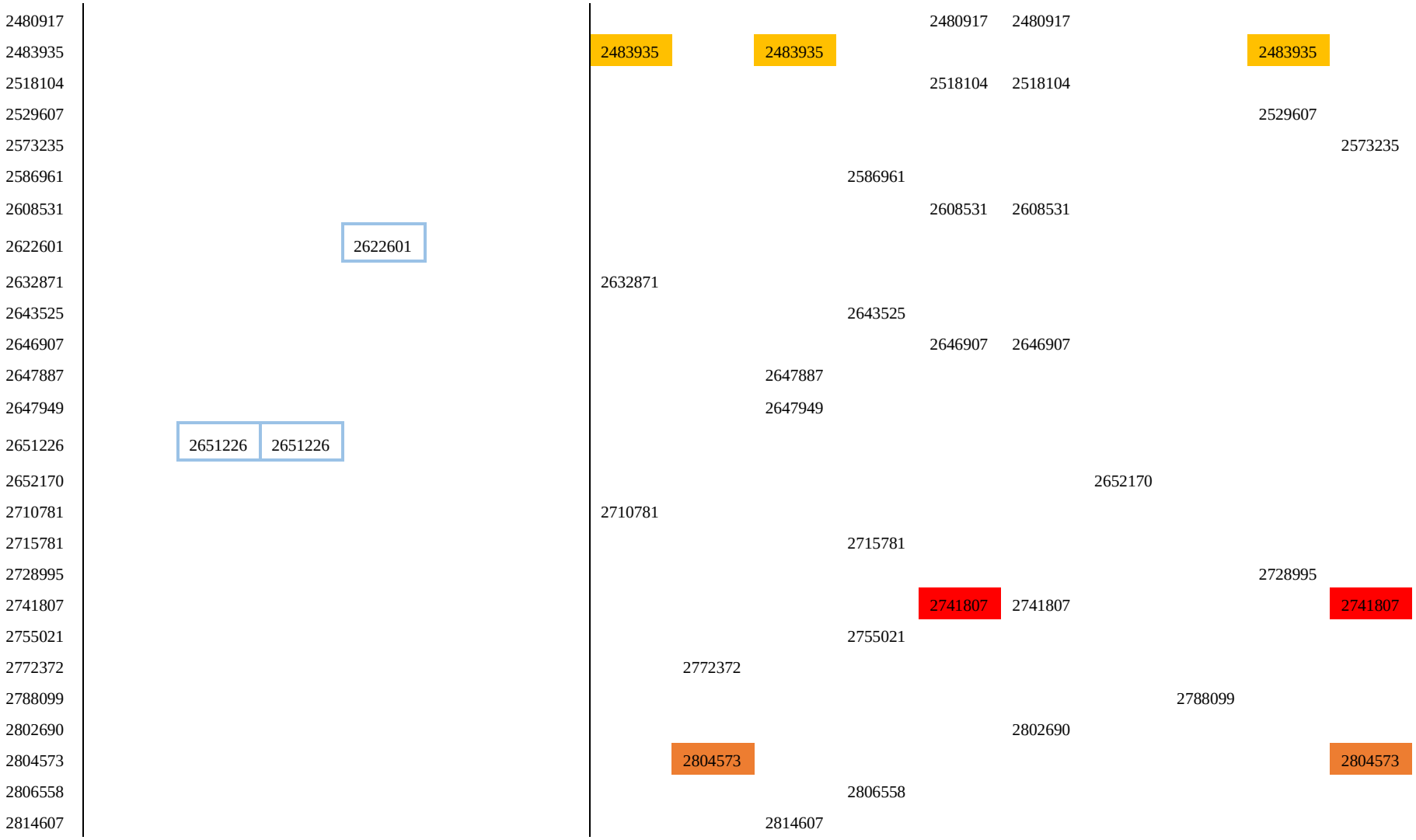

613733
633687
640892
671367
681349
683064
760208
779151
826887
828172
834044
838149
839814
860317
879645
881185
895995
945079
946308
976250
981108
1010743
1011187
1011866
1023882
1028572
1043719

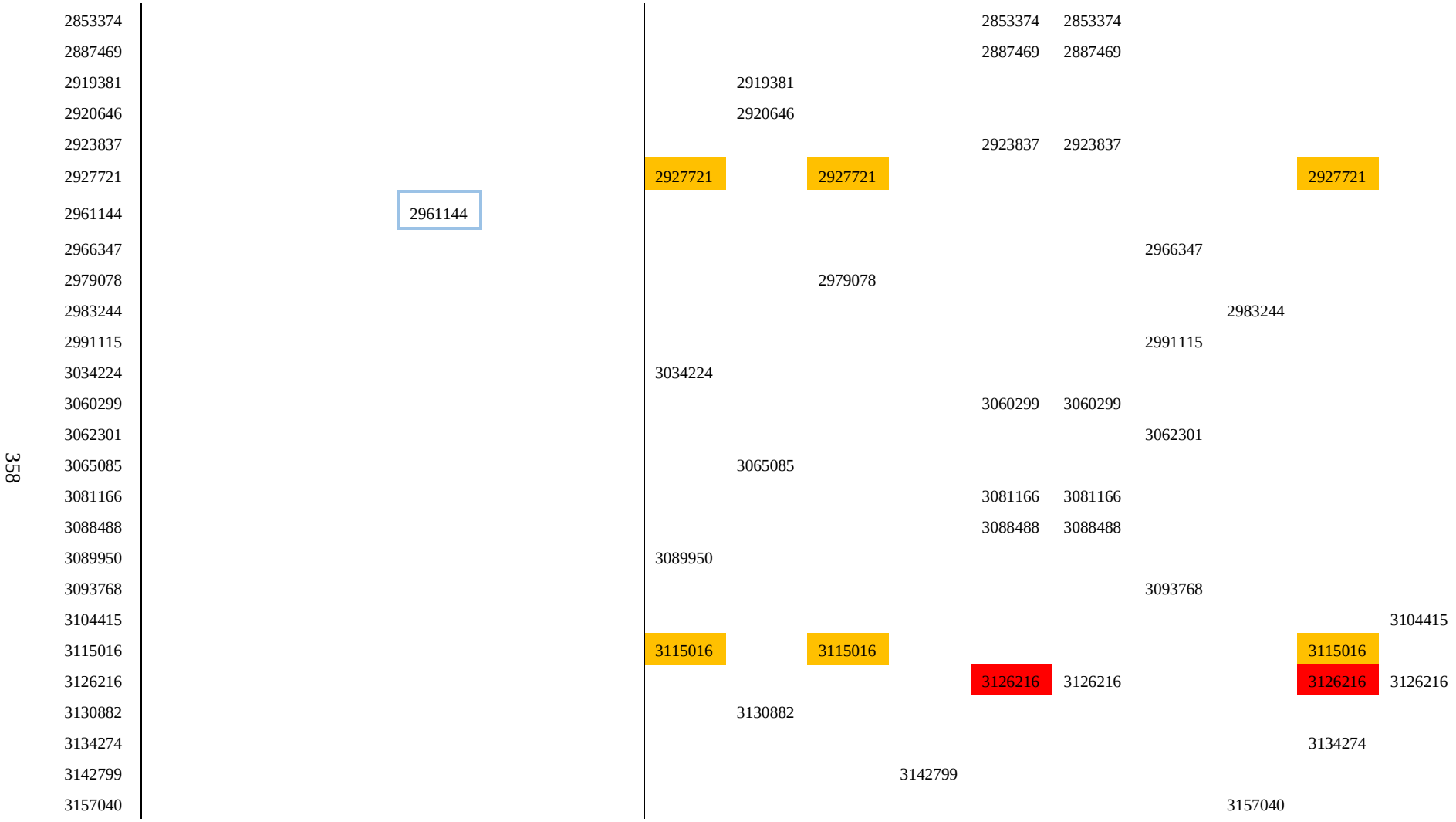

1047910
1059859
1061539
1078221
1078297
1096207
1116813
1117314
1164214
1165048
1188145
1224668
1224781
1224963
1241975
1243791
1244400
1246898
1246905
1272639
1303726
1311713
1325374
1335638
1342138
1358201
1078297
1078297
1078297
1078297
1078297
1078297
1244400
1244400
1244400

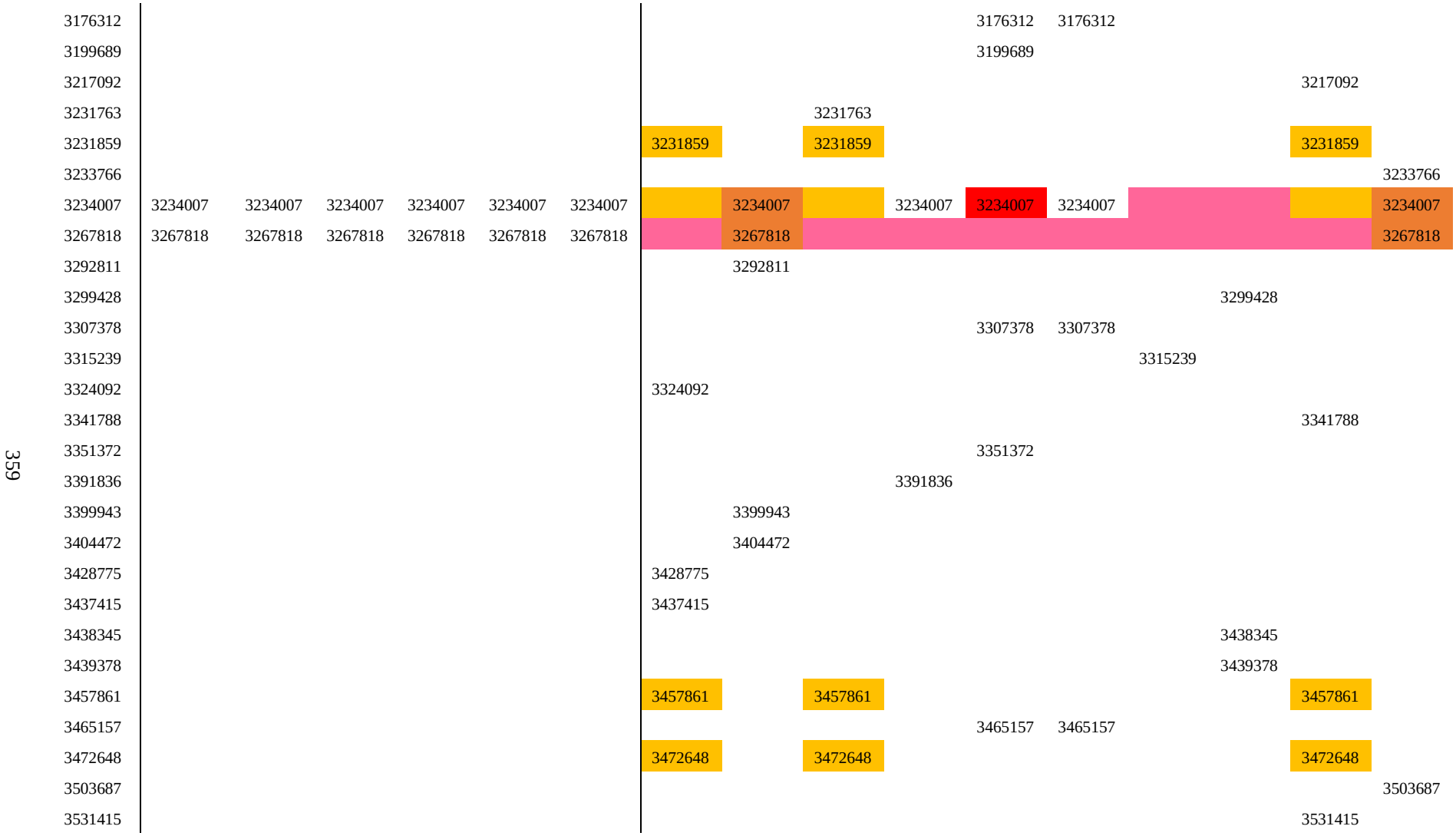

1364760
1371019
1417459
1419605
1427514
1428725
1466519
1467064
1476063
1487032
1500731
1524096
1535502
1537837
1539098
1541595
1585597
1613471
1614684
1634573
1649593
1667030
1691311
1710271
1719199
1735600
1736091


1736348
1744233
1746652
1765830
1792890
1806932
1818853
1824625
1915275
1922019
1930729
1938356
2013932
2038493
2068908
2087885
2101364
2101478
2108268
2108692
2110823
2111712
2129278
2129295
2140175
2144065
1736348
1744233
1746652
1765830
1792890
1792890
1792890
1806932
1818853
1824625
1915275
1922019
1930729
1938356
2013932
2038493
2038493
2068908
2068908
2087885
2101364
2101478
2108268
2108692
2110823
2111712
2129278
2129278
2129295
2129295
2129295
2129295
2140175
2140175
2144065
2144065

2148763												2148763
2182829			2182829									
2189526									2189526	2189526		
2189534									2189534	2189534		
2189543									2189543	2189543		
2189557									2189557	2189557		
2198989							2198989	2198989				
2222256										2222256		
2231813			2231813									
2261816							2261816	2261816				
2267736						2267736						
2270854			2270854									
2278781						2278781						
2289691									2289691	2289691		
2333003						2333003	2333003	2333003				
2335069			2335069									
2347194					2347194							
2354425				2354425								
2365974											2365974	
2391417									2391417			
2392655									2392655			
2411493						2411493	2411493	2411493				
2437504					2437504							
2466355	2466355	2466355										
2469901				2469901								
2480001					2480001							


2480917
2483935
2518104
2529607
2573235
2586961
2608531
2622601
2632871
2643525
2646907
2647887
2647949
2651226
2652170
2710781
2715781
2728995
2741807
2755021
2772372
2788099
2802690
2804573
2806558
2814607
2622601
2651226
2651226
2480917
2480917
2483935
2483935
2483935
2518104
2518104
2529607
2573235
2586961
2608531
2608531
2632871
2643525
2646907
2646907
2647887
2647949
2652170
2710781
2715781
2728995
2741807
2741807
2741807
2755021
2772372
2788099
2802690
2804573
2804573
2806558
2814607


2853374
2887469
2919381
2920646
2923837
2927721
2961144
2966347
2979078
2983244
2991115
3034224
3060299
3062301
3065085
3081166
3088488
3089950
3093768
3104415
3115016
3126216
3130882
3134274
3142799
3157040
2961144
2853374
2853374
2887469
2887469
2919381
2920646
2923837
2923837
2927721
2927721
2927721
2966347
2979078
2983244
2991115
3034224
3060299
3060299
3062301
3065085
3081166
3081166
3088488
3088488
3089950
3093768
3104415
3115016
3115016
3115016
3126216
3126216
3126216
3126216
3130882
3134274
3142799
3157040


3176312
3199689
3217092
3231763
3231859
3233766
3234007
3267818
3292811
3299428
3307378
3315239
3324092
3341788
3351372
3391836
3399943
3404472
3428775
3437415
3438345
3439378
3457861
3465157
3472648
3503687
3531415

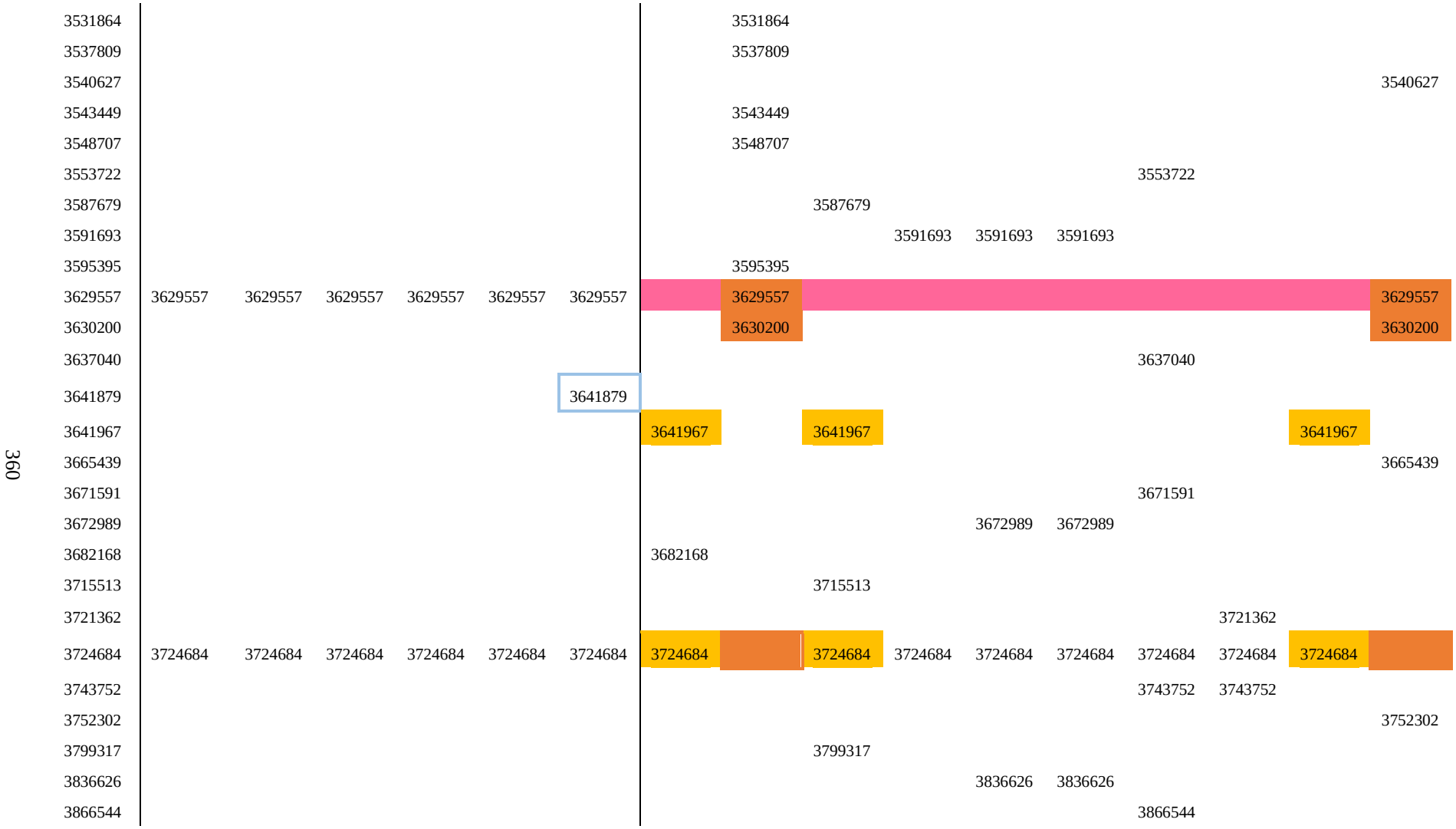

3531864
3537809
3540627
3543449
3548707
3553722
3587679
3591693
3595395
3629557
3630200
3637040
3641879
3641967
3665439
3671591
3672989
3682168
3715513
3721362
3724684
3743752
3752302
3799317
3836626
3866544



Table legend

- In all strains
- Only in all non-mutator strains
- Only, but not all non-mutator strains
- Clade
- Clade
- Between clades
- In majority of mutator strains