



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Interaction of *Bacteroides fragilis* with Host Proteins and Effects of
Nitrogen Limitation on the *B. fragilis* Transcriptome**



Aparna Shankar

Thesis presented for the degree of Doctor of Philosophy

**Institute of Cell Biology
The University of Edinburgh
November 2015**

Declaration

The author performed all of the investigations and procedures presented in this thesis, unless otherwise stated.



APARNA SHANKAR

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to each and everyone who made this four year journey worthwhile and unforgettable. To begin with, I thank my supervisor Dr. Garry Blakely for giving me an opportunity to pursue doctoral research in his lab. It was definitely his constant encouragement, valuable suggestions and enthusiasm for science that kept me going. I would also like to thank my second supervisor, Dr. Maurice Gallagher and our collaborator, Prof. Sheila Patrick for their timely inputs that led me in the right direction in my project.

I extend my thanks to the Edinburgh Genomics team who supported me with the sequencing work and Ms. Persistent Bioinformatics for assisting me in the analysis. In particular, I thank Dr. Anamika and Srikant for patiently answering all my queries regarding the in silico analysis, which helped me in the project completion. I also thank the members of the Darwin media kitchen for their organised work which helped me plan my daily laboratory experiments without hindrance. A big thank you to all my past lab members and the lovely floor mates of the old Darwin building for the informative talks and discussions on science. A special hug goes out to Keerthi, Pradeep, Daksh, Sanju and Sharanya for spreading goodwill and cheer which helped me de-stress after a bad 'science' day at work! I would also like to thank all my dear friends back home who were just a call away whenever I needed them.

I am extremely indebted to my family, especially my parents, mother-in-law and father-in-law for believing in me and being there by my side. I owe it to my mother for instilling a passion for science and research in me. I am short of words to thank my ever loving and caring husband, Raman, for his round-the-clock support, technical help and above all, for patiently putting up with my 'thesis-induced' panic!

Table of Contents

Chapter 1. Introduction: The Human Microbiome	5
1.1. Commensalism in the Human Intestine.....	6
1.2. <i>Bacteroides fragilis</i>	10
1.2.1. Polysaccharide synthesis.....	11
1.2.2. Polysaccharide A	13
1.2.3. Outer membrane vesicles.....	15
1.2.4. Pathogenic factors.....	19
Chapter 2. Materials and Methods	23
2.1. Strains	23
2.2. Bacterial culture conditions.....	23
2.3. Oligonucleotides.....	24
2.4. Bacterial techniques	28
2.4.1. <i>B. fragilis</i> growth curves.....	28
2.4.2. Preparation of chemically competent cells	28
2.4.3. Transformation: Heat-shock method	29
2.4.4. Transformation: Electroporation method.....	29
2.4.5. Conjugation of <i>E. coli</i> S17-1 λ pir with <i>B. fragilis</i> NCTC 9343	30
2.4.6. Screening for deletion genotypes.....	30
2.5. DNA techniques	30
2.5.1. Genomic DNA extraction	30
2.5.2. Plasmid DNA extraction	30
2.5.3. Polymerase chain reactions.....	31
2.5.4. Agarose gel electrophoresis	31
2.5.5. DNA purification	31
2.5.6. DNA quantification.....	32
2.5.7. Restriction digestion	32
2.5.8. Cloning.....	32

2.5.9. Single colony assay	32
2.5.10. Sanger-based plasmid sequencing	33
2.6. RNA sequencing.....	33
2.6.1. RNA extraction	33
2.6.2. Library preparation	34
2.7. Protein techniques.....	35
2.7.1. Preparation of whole cell protein extracts	35
2.7.2. Preparation of concentrated culture supernatants	35
2.7.3. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).....	35
2.7.4. Staining/destaining of gels	36
2.7.5. Zymography	36
2.7.6. Preparation of samples for immunoblot analysis.....	37
2.7.7. Transfer of proteins by wet electroblotting.....	38
2.7.8. Development of western blots and far-western blots.....	38
2.7.9. Stripping antibodies from immunoblots	39
2.7.10. Protein expression.....	39
2.7.11. Protein purification by nickel affinity chromatography.....	41
2.7.12. Protein estimation by bradford assay.....	42
2.7.13. Immunofluorescence microscopy (IFM)	43
2.8. Bioinformatics.....	44
2.8.1. Quality control of RNA extracts	44
2.8.2. RNA sequencing data, assembly and annotation.....	44
Chapter 3. Binding of Host Proteins by <i>Bacteroides fragilis</i>	47
3.1. Introduction	47
3.1.1. Structure of the extracellular matrix (ECM).....	48
3.1.2. Microbial surface components recognised by adhesive matrix molecules (MSCRAMMs) in bacteria	55
3.1.3. MSCRAMMs expressed in intestinal microbiota.....	61
3.1.4. Aims.....	67

3.2. Choice of bacterial strains	68
3.3. Generation of markerless deletions in <i>Bacteroides fragilis</i> NCTC 9343 genome	70
3.4. Fibrinogen binding studies previously conducted in <i>B. fragilis</i>	74
3.5. Generation of a markerless deletion of BF1705 encoding a putative fibrinogen-binding protein in <i>B. fragilis</i>	75
3.5.1. Primer design for amplifying the upstream and downstream regions of BF1705 gene	75
3.5.2. Generation of the BF1705 deletion construct.....	76
3.5.3. Cloning the BF1705 deletion construct into pGB910	78
3.5.4. Conjugal transfer of the Δ BF1705 construct into <i>B. fragilis</i> NCTC 9343 81	
3.5.5. Resolution of partial diploid <i>B. fragilis</i> strains carrying the deletion construct.....	84
3.6. IFM analysis of fibrinogen binding by the Δ BF1705 strain.....	87
3.7. Investigating the role of capsular polysaccharides in fibrinogen binding.....	89
3.7.1. Generation of a markerless double deletion of BF1705 and BF1708.....	89
3.7.2. IFM analysis of fibrinogen binding in Δ BF1705 Δ BF1708 cells.....	90
3.8. Expression and Purification of the BF1705 protein	95
3.8.1. Transformation of <i>E. coli</i> BL21 cells with the pET-15b expressing BF1705.....	95
3.8.2. BF1705 protein expression	98
3.8.3. Protein purification	98
3.9. Analysis of fibrinogen binding by <i>B. fragilis</i> concentrated culture supernatants in immunoblots.....	102
3.9.1. Choice of sample preparation methods for immunoblot analyses.....	102
3.9.2. Detection of BF1705 protein in concentrated supernatant samples	104
3.9.3. Far-western analysis of fibrinogen binding in Δ BF1705.....	106
3.9.4. Analysis of fibrinogen binding by Δ BF1705 Δ BF1708 cells in immunoblots	111
3.10. Detection of fibrinogen degradation by the Δ BF1705 strain	114
3.11. Binding potential of BF1705 protein to fibronectin by far-western analysis	116

3.12. Discussion.....	119
Chapter 4. Degradation of Fibrinogen by <i>Bacteroides fragilis</i>	131
4.1. Introduction	131
4.1.1. Proteases	131
4.1.2. Microbial proteases.....	132
4.1.3. Enzymatic activities of <i>Bacteroides</i> spp.....	136
4.1.3 Aims.....	146
4.2. Fibrinogen degradation detected in <i>B. fragilis</i> strain NCTC 9343.....	146
4.3. Generation of markerless protease-encoding gene deletion mutants in <i>B. fragilis</i>	148
4.3.1. Primer design for amplifying the upstream and downstream regions of the selected protease-encoding genes	151
4.3.2. Generation of the deletion constructs	153
4.3.3. Cloning the deletion constructs into pGB910.....	156
4.3.4. Conjugal transfer of gene deletion constructs into <i>B. fragilis</i> strain NCTC 9343.....	160
4.3.5. Resolution of partial diploids carrying the deletion constructs	166
4.4. Analysis of fibrinogen degradation by protease-encoding gene deletion mutants.....	172
4.5. Analysis of fibrinogen degradation by the generation of multiple protease-encoding gene deletion mutants in <i>B. fragilis</i>	173
4.5.1. Detection of fibrinogen degradation by the multiple deletion mutants ..	179
4.6. Identification, cloning and expression of metalloproteases secreted by <i>B. fragilis</i>	181
4.6.1. Identification of secreted proteases and peptidases in <i>B. fragilis</i>	181
4.6.2. Choice of vector for the expression of the selected proteases in <i>E. coli</i> . 187	
4.6.3. Primer design and cloning of expression construct into pET-100 vector188	
4.6.4. Expression of recombinant pET-TOPO vector.....	195
4.7. Analysis of fibrinogen degradation by the expressed proteases/peptidases in fibrinogen zymography.....	196
4.8. Discussion.....	198

Chapter 5. Effect of Nitrogen Limitation on the transcriptome of <i>Bacteroides fragilis</i>	201
5.1. Introduction	201
5.1.1. Metabolic regulation of <i>E. coli</i> in nitrogen-sufficient and nitrogen-limiting environments.....	205
5.1.2. The link between nitrogen and carbon metabolism	209
5.1.3. Aims.....	210
5.2. Choice of nitrogen-limiting medium for the growth of <i>B. fragilis</i>	210
5.2.1. Effect of glutamine as the sole nitrogen source on growth rate.....	214
5.3. RNA sequencing.....	220
5.3.1. Validation of RNA samples prior to sequencing	222
5.3.2. Library construction and Illumina-based sequencing	230
5.3.3. Validation of RNA-Seq reads post RNA sequencing.....	233
5.3.4. Alignment against the reference genome.....	255
5.3.5. Gene quantification.....	258
5.3.6. Differential expression.....	262
5.4. Analysis of differential gene expression with respect to nitrogen limitation	280
5.4.1. Nitrogen assimilation and metabolism	281
5.4.2. Host matrix adhesins and secreted proteases	296
5.4.3. Response to shock/stress.....	302
5.4.4. Drug/antibiotic resistance	307
5.5. Discussion	310
5.5.1. Glutamine in the nitrogen-limiting medium	310
5.5.2. Inability to grow on glutamic acid as the sole nitrogen source	312
5.5.3. Choice of time points for RNA-Seq	314
5.5.4. Analysis of differentially regulated genes under nitrogen-limiting conditions.....	315
Chapter 6. General Discussion	331
Chapter 7. Bibliography	341
Appendices enclosed as soft copy in CD at the back	

List of Figures

Figure 2.1: BSA standard curve using the Bradford assay for protein estimation	42
Figure 3.1: Schematic of the extracellular matrix	49
Figure 3.2: Structural organization of FnBps and ClfA of <i>Staphylococcus aureus</i>	60
Figure 3.3: Electron micrograph of <i>B. fragilis</i> NCTC 9343	68
Figure 3.4: Schematic of the generation of markerless chromosomal deletions in <i>B. fragilis</i>	73
Figure 3.5: Graphic representation of BF1705 and the flanking regions	76
Figure 3.6: Amplification of flanking regions and the deletion construct of BF170577	
Figure 3.7: Vector map of pGB910	79
Figure 3.8: Restriction profiling of pGB910 for cloning experiments	80
Figure 3.9: Confirmation of the BF1705 deletion construct in the recombinant vector	81
Figure 3.10: Amplification of BF1705 and the flanking regions in the partial diploid	83
Figure 3.11: Vector map of pGB920	85
Figure 3.12: Generation of the Δ BF1705 deletion mutant	86
Figure 3.13: Immunofluorescence microscopy analysis of wild-type and Δ BF1705 cells incubated with fibrinogen	88
Figure 3.14: Generation of the Δ BF1705 Δ BF1708 double deletion mutant from the Δ BF1705	93
Figure 3.15: Immunofluorescence microscopy analysis of wild-type and Δ BF1705 Δ BF1708 cells incubated with fibrinogen	95
Figure 3.16: Confirmation of the recombinant vector carrying the BF1705 expression construct	97
Figure 3.17: Expression of BF1705 in <i>E. coli</i>	100
Figure 3.18: Purification of BF1705 by nickel column chromatography	100
Figure 3.19: Purification of BF1705 by gel filtration column chromatography	101
Figure 3.20: Analysis of concentrated culture supernatants and BF1705	103
Figure 3.21: Detection of BF1705 in concentrated culture supernatants	105

Figure 3.22: Analysis of fibrinogen binding by the wild-type and Δ BF1705 culture supernatants	108
Figure 3.23: Detection of fibrinogen binding by BF1705	109
Figure 3.24: Analysis of fibrinogen binding by the wild-type and differing BF1705 concentrations	110
Figure 3.25: Analysis of fibrinogen binding (0.1 mg/ml) by whole cell lysates of wild-type and deletion mutants	113
Figure 3.26: Analysis of fibrinogen binding (0.015 mg/ml) by whole cell lysates of wild-type and deletion mutants	114
Figure 3.27: Analysis of fibrinogen degradation by the wild-type and Δ BF1705 cultures	115
Figure 3.28: Analysis of fibronectin binding by BF1705 and wild-type and Δ BF1705 culture supernatants	118
Figure 4.1: Detection of fibrinogen degradation by the wild-type 9343	148
Figure 4.2: Graphic representation of the selected protease encoding genes and flanking regions	153
Figure 4.3: Amplification of flanking regions of the protease encoding genes	154
Figure 4.4: Amplification of deletion constructs of the protease encoding genes	155
Figure 4.5: Confirmation of the deletion constructs present in the recombinant vector	159
Figure 4.6: Amplification of protease encoding genes and flanking regions in the respective partial dipolids	165
Figure 4.7: Generation of the protease encoding gene deletion mutants	171
Figure 4.8: Analysis of fibrinogen degradation of the deletion mutants by zymography	173
Figure 4.9: Generation of the Δ BF1979 Δ BF3775 double deletion mutant from the Δ BF3775	176
Figure 4.10: Generation of the Δ BF0275 Δ BF1979 Δ BF3775 triple deletion mutant from the Δ BF1979 Δ BF3775	177

Figure 4.11: Generation of the Δ BF0657 Δ BF0275 Δ BF1979 Δ BF3775 quadruple deletion mutant from the Δ BF0275 Δ BF1979 Δ BF3775	178
Figure 4.12: Analysis of fibrinogen degradation by the quadruple deletion mutant	180
Figure 4.14: Graphic representation of the expression construct model	189
Figure 4.15: Amplification of the protease/peptidase encoding genes selected for expression	190
Figure 4.16: Confirmation of the clones positive for the protease-encoding gene expression constructs	193
Figure 4.17: Expression of the selected protease-encoding genes in <i>E. coli</i>	196
Figure 4.18: Analysis of fibrinogen degradation by the protease-encoding genes expressed in <i>E. coli</i>	197
Figure 5.1: GS/GOGAT pathway	205
Figure 5.2: GDH pathway	206
Figure 5.3: Nitrogen assimilation pathway in <i>E. coli</i>	208
Figure 5.4: Effect of altered nitrogen (ammonium sulphate) content on <i>B. fragilis</i> growth	213
Figure 5.5: Comparison of bacterial growth in glutamine-containing glucose-DM and glutamic acid-containing glucose-DM	217
Figure 5.6: Comparison of <i>B. fragilis</i> growth in glutamic acid-containing glucose-DM and normal glucose-DM	218
Figure 5.7: Effect of nitrogen limitation on <i>B. fragilis</i> growth	219
Figure 5.8: Schematic of RNA-Seq workflow	221
Figure 5.9: RNA extracted from <i>B. fragilis</i> grown on normal glucose-DM and GNM	223
Figure 5.10: RIN-based quantification of the extracted RNA samples	226
Figure 5.11: RIN-based quantification of the RNA samples post rRNA depletion	229
Figure 5.12: Outline of the Illumina (Solexa) sequencing platform	232
Figure 5.13: Screenshot providing the summary of the FastQC report of a read pair file	234

Figure 5.14: Display of results of the modules performed by the FastQC tool on N1_1 fastq file	235
Figure 5.15: Distance plots of the normalized RNA-Seq reads	264
Figure 5.16: Principal component analysis of the normalized RNA-Seq reads	265
Figure 5.17: Representation of the normalized expression values	270
Figure 5.18: Qualitative analysis of the differentially expressed genes in the GA set with reference to the N set	272
Figure 5.19: Qualitative analysis of the differentially expressed genes in the GA set with reference to the G1 set	274
Figure 5.20: Qualitative analysis of the differentially expressed genes in the G1 set with reference to the N set	276
Figure 5.21: Qualitative analysis of the differentially expressed genes involved in nitrogen metabolism in the GA and G1 sets with reference to the N set	288
Figure 5.22: Synthesis of asparagine by glutamine hydrolysis	291
Figure 5.23: Krebs-Henseleit pathway-mediated formation of arginine as an intermediate in aspartate metabolism	292
Figure 5.24: Arginine deiminase pathway for ammonia generation	293
Figure 5.25: Qualitative analysis of the differentially expressed genes involved in general amino acid metabolism and transport in GA and G1 with reference to N	295
Figure 5.26: Qualitative analysis of the differentially expressed genes involved in host matrix adhesion and proteolytic activity in GA and G1 with reference to N	301
Figure 5.27: Qualitative analysis of the differentially expressed genes involved in response to shock and oxidative stress in GA and G1 with reference to N	306
Figure 5.28: Qualitative analysis of the differentially expressed genes involved drug/antibiotic resistance in GA and G1 with reference to N	309

List of tables

Table 2.1: List of strains	23
Table 2.2: List of oligonucleotides	25
Table 2.3: List of read pairs after trimming the adapter sequences	45
Table 3.1: Description of BF1705 using artemis genome browser (Carver et al., 2008)	75
Table 3.2: Description of BF1708 using artemis genome browser	91
Table 4.1: Description of the selected protease encoding genes using artemis genome browser	150
Table 4.2: List of the genes encoding putative secreted proteases and peptidases in NCTC 9343 genome	183
Table 5.1: Nanodrop-based quantification of RNA samples prior to sequencing	223
Table 5.2: Fast QC report of the read pair files of the G1, GA and N sets	252
Table 5.3: List of the unique overrepresented sequences	253
Table 5.4: List of read pairs aligned against the <i>B. fragilis</i> NCTC 9343 genome	256
Table 5.5: List of the number of read pairs assigned to genes	259
Table 5.6: List of the number of genes differentially expressed in N vs GA, G1 vs GA and N vs G1 analyses at q value < 0.05	279
Table 5.7: List of the number of genes differentially expressed in N vs GA, G1 vs GA and N vs G1 analyses at q value < 0.05 and log2 FC cut-off of 1	279

Abbreviations

aa	Amino acid
AATGal	Acetamido-amino-2,4,6-trideoxy galactose
ADI	Arginine deiminase
AhpC	Alkyl hydroperoxide reductase
AIEC	Adherent and invasive <i>E. coli</i>
Ang 4	Angiogenin 4
APS	Ammonium persulphate
ATase	Adenylyltransferase
BCG 85	Bacillus Calmette Guérin antigen 85
Bcp	Bacterioferritin comigratory protein
BF-FBP	<i>Bacteroides fragilis</i> fibrinogen binding protein
BFT	<i>Bacteroides fragilis</i> toxin
BHI-S	Supplemented Brain heart infusion
BLAST	Basic local alignment search tool
BLASTN	Nucleotide basic local alignment search tool
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
CbpI	Collagen binding protein
CCD	Charge-coupled device
cDNA	complementary DNA
ClfA	Clumping factor A
CMP	Cytidine 5'-monophosphate
CP1	Type 1 <i>Streptococcus pneumoniae</i> capsular polysaccharide
CPS	Capsular polysaccharide synthesis
CRP	cAMP receptor protein

DC	Dendritic cell
DEPC	Diethyl pyrocarbonate
DFP	Diisopropyl fluorophosphate
DM	Defined medium
dNTP	Deoxyribonucleotide triphosphate
ds	Double-stranded
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
dUTP	Deoxyuridine triphosphate
ECF	Extracytoplasmic function
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDL	Electron dense layer
EDTA	Ethylenediaminetetraacetic acid
Efb	Extracellular fibrinogen binding protein
ER	Endoplasmic reticulum
ETBF	Enterotoxigenic <i>Bacteroides fragilis</i>
FDR	False discovery rate
Fg	Fibrinogen
FITC	Fluorescein isothiocyanate
FKP	L-fucokinase/guanosine5'-diphosphate-L-fucose pyrophosphorylase
FnBPA	Fibronectin Binding Protein A
FnBPB	Fibronectin Binding Protein B
Fnr	Fumarate and nitrate reductase
Gadd45 α	Growth arrest and DNA-damage-inducible
GAG	Glycosaminoglycan
GAM	Glutamic acid-containing glucose-DM

GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GATase	Glutamine amidotransferase
GDH	Glutamate dehydrogenase
GEM	Glutamine-containing glucose-DM
GF	Germ-free
GPRPNA	Glycylproline p-nitroanilide
GS/GOGAT	Glutamine synthetase/Glutamate synthase
GTP	Guanosine triphosphate
HMMP	High molecular mass polysaccharide
IBD	Inflammatory bowel disease
ICAM-1	Intercellular adhesion molecule-1
IEC	Intestinal epithelial cell
IFM	Immunofluorescence microscopy
IFN-gamma	Interferon-gamma
IL-4	Interleukin-4
IL-10	Interleukin-10
IL-12	Interleukin-12
iNOS	Inducible nitric oxide synthase
IPTG	Isopropyl-beta-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
KatA	Catalase A
KatB	Catalase B
LB	Luria-Bertani
LC	Large capsule
LIR	Left inside reverse
LOF	Left outside forward
log2 FC	log2 fold-change

LPNA	Leucine p-nitroanilide
LPS	Lipopolysaccharide
LRR	Leucine-rich repeat
mAb	Monoclonal antibody
MAC	Membrane attack complex
MC	Micro-capsule
MCS	Multiple cloning site
MHC II	Major histocompatibility complex II
MMC	Murine mesothelial cell
MOPS	3- (N-morpholino) propanesulfonic acid
mRNA	messenger RNA
MS	Mass spectrometry
MSCRAMM	Microbial surface components recognizing adhesive matrix molecule
MWCO	Molecular weight cut off
NA	Neuraminidase
Nac	Nitrogen assimilation control
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	National center for biotechnology information
ncRNA	non-coding RNA
NCTC	National collection of type cultures
NEB	New England biolabs
NF- κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequencing
Nim	Nitroimidazole resistance
Ni-NTA	Nickel-nitrilotriacetic acid
Nmr	Nitrogen metabolite repression

NTBF	Non enterotoxin-producing <i>B. fragilis</i>
Ntr	Nitrogen regulated response
OD	Optical density
OMP	Outer membrane protein
OMV	Outer membrane vesicle
ORF	Open reading frame
OTCase	Ornithine carbamoyltransferase
PBS	Phosphate-buffered saline
PbSp	<i>Paracoccidioides brasiliensis</i> serine protease
PCA	Principal component analysis
PCR	Polymerase chain reaction
PMNL	Polymorphonuclear leukocyte
PMSF	Phenylmethylsulfonyl fluoride
PPAR- γ	Proliferators-activated receptor gamma
ppGpp	Guanosine tetraphosphate
PS A	Polysaccharide A
PS B	Polysaccharide B
PS G	Polysaccharide G
PTS	Phosphotransferase system
PUL	Polysaccharide utilization loci
PVDF	Polyvinylidene fluoride
RDA	Recommended dietary allowance
R/M	Restriction/modification
RelA	v-rel avian reticuloendotheliosis viral oncogene homolog A
RIF	Right inside forward
RIN	RNA Integrity Number
RNA	Ribonucleic acid

RNA-seq	RNA sequencing
ROR	Right outside reverse
ROS	Reactive oxygen species
rRNA	ribosomal RNA
RNase	Ribonuclease
RT-PCR	Reverse transcription polymerase chain reaction
SAM	Sequence alignment/map format
SCF	Single colony forming
SCFA	Short chain fatty acid
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SOC	Super optimal broth with catabolite repression
Sod	Superoxide dismutase
SpeB	Streptococcal pyrogenic exotoxin B
SPR	Surface plasmon resonance
sRNA	small RNA
Stat	Signal transducer and activator of transcription protein
STSS	Streptococcal toxic shock syndrome
SurA	Survival proteinA
Sus	Starch utilisation system
TAE	Tris acetate EDTA
TBDT	TonB-dependent transporter
TBS-T	Tris buffered saline-tween 20
TCA	Trichloroacetic acid
TCR	T cell receptor
T _d	Doubling time
TEM	Transmission-electron microscopy
TEMED	Tetramethylethylenediamine

TFB	Transformation buffer
Th cells	T-helper cells
TLR2	Toll-like receptor 2
T _m	Melting temperature
TNF-alpha	Tumor necrosis factor-alpha
Tpx	Thiol-dependent peroxidase
TSA	Thiol specific antioxidant
TSS	Transcription start site
UMP	Uridine monophosphate
UTase/UR	Uridylyl transferase /uridylyl-removing enzyme
UTR	Untranslated region
UV	Ultraviolet
VAPNA	Valyl-alanine p-nitroanilide
WBC	White blood cell
ZPS	Zwitterionic polysaccharides

Abstract

Bacteroides fragilis is a member of the normal microbiota that resides in the human lower gastrointestinal tract. This bacterium is of clinical significance because it is the most frequently isolated Gram-negative obligate anaerobe from peritoneal abscesses and bloodstream infections. Human fibrinogen is a hexameric-glycoprotein that is important for fibrin-mediated abscess formation and limiting the spread of infection. *B. fragilis* can bind and degrade fibrinogen which may aid in its escape from abscesses into the bloodstream, thereby promoting bacteraemia.

In addition to fibrinogen, binding of *B. fragilis* to fibronectin, a component of the extracellular matrix, found in association with fibrinogen at wound sites, has also been reported. An outer membrane protein, BF1705, expressed by *B. fragilis* was found to share homology with BspA from *Tannerella forsythia* which is known to bind fibrinogen. The gene encoding BF1705 was deleted from the *B. fragilis* NCTC 9343 genome in the present work using a markerless gene deletion technology. Proteins derived from the outer membranes of wild-type *B. fragilis* were able to bind fibronectin and fibrinogen in far-western blots. Similar protein extracts from the Δ BF1705 strain did not bind fibrinogen and fibronectin, which confirms the role of BF1705 in adhesive interactions with proteins of the host extracellular matrix.

The possible involvement of BF1705 in fibrinogen degradation was ruled out because the Δ BF1705 strain still degraded fibrinogen. To identify the proteases involved in degradation of fibrinogen, four genes encoding putative extracellular metallo- and serine proteases in the size range 45-50 kDa were deleted from the NCTC 9343 genome. All of the single and multiple mutants defective in these selected proteases were still capable of degrading fibrinogen as determined by zymography. Expression of eight *B. fragilis* proteases in *E. coli* did not lead to detectable degradation of fibrinogen. These observations suggest that these proteases

alone cannot degrade fibrinogen and either that an unidentified protease is responsible for degradation or that there is redundancy in the proteases involved.

Under conditions of nitrogen limitation bacteria resort to scavenging nitrogen from the environment to replenish the depleting intracellular nitrogen content. By examining the differential regulation of the *B. fragilis* transcriptome under nitrogen replete and depleting conditions, a potential role for BF1705 and secreted proteases in nutrient binding and assimilation were studied. Growth on conventional glucose defined medium with ammonia as the nitrogen source was compared to growth in defined medium with glutamine as nitrogen source. A reduced doubling time and diauxic growth in the medium containing glutamine indicated nitrogen limitation. Comparison of the transcriptome derived from cultures of *B. fragilis* grown on either ammonia or glutamine by RNA-Seq did not reveal a significant upregulation of BF1705 in response to nitrogen limitation. This observation in conjunction with its inability to degrade fibrinogen suggests that the primary role of BF1705 might be as an adhesin and does not act directly in nutrient binding and degradation. Nevertheless, nitrogen limitation was found to induce the expression of four protease-encoding genes by over a 2-fold (adjusted p value < 0.05). The molecular weight of three of these proteases were identified to be within the size range of 45-55 kDa which corresponded to the lysis bands detected by fibrinogen zymography with wild-type *B. fragilis* protein extracts. Therefore the possible involvement of these three proteases in fibrinogen degradation could be assessed. A 155-fold upregulation (adjusted p value < 0.05) in *asnB*, encoding a homologue of asparagine synthetase B, under conditions of nitrogen limitation suggest a previously uncharacterised aspartate metabolism pathway for ammonia generation via arginine catabolism in *B. fragilis*. Ammonia thus formed might aid in sustaining *B. fragilis* growth under nitrogen deprived conditions. In addition to nitrogen assimilation, significant upregulation was observed in the expression of genes involved in regulation of oxidative stress and metronidazole resistance. The observed changes in the transcriptome will add to our understanding of the *B. fragilis* metabolism and potential assist with unravelling the mechanisms of infection mediated by this important opportunistic pathogen.

Lay Abstract

Bacteroides fragilis is a bacterium found in the human gut and is a member of the commensal microbiota, or 'good' bacteria, which have a positive impact on human health and nutrition. However, *B. fragilis* is also the most frequently isolated obligate anaerobe from peritoneal abscesses and bloodstream infections when it accidentally escapes from the gastrointestinal tract, following injury. Human fibrinogen is a plasma glycoprotein and is converted by thrombin into fibrin, which helps in the formation of blood clots and abscesses, thereby curtailing infection. *B. fragilis* can bind and degrade fibrinogen which might aid in the escape of the bacterium from abscesses, leading to infection dissemination and bacteraemia.

B. fragilis can also bind fibronectin, a component of the host extracellular matrix, which function in wound healing in association with fibrinogen. BF1705, an outer membrane protein expressed by *B. fragilis* was found to share homology with previously annotated fibrinogen binding protein in another bacteria. The gene encoding BF1705 was deleted from the genome of the *B. fragilis* strain NCTC 9343 using a markerless gene deletion technology. Proteins derived from the outer membranes of wild-type *B. fragilis*, by concentrating the cell-free culture supernatants, were observed to bind to fibrinogen and fibronectin. On the contrary, similar protein extracts from the gene deletion mutant, Δ BF1705, did not bind fibrinogen and fibronectin, which confirms the role of BF1705 in adhesive interaction with proteins of the extracellular matrix.

BF1705 was not involved in fibrinogen degradation by *B. fragilis* because in situ degradation of fibrinogen was observed when the protein was incubated with the Δ BF1705 strain. Zymography experiments with fibrinogen as the substrate revealed a zone of lysis at ~ 45 kDa for wild-type *B. fragilis*. Therefore, four genes encoding putative proteases in the size range 45-50 kDa were deleted from the NCTC 9343 genome to identify the proteases involved in degradation of fibrinogen. Single and

multiple mutants defective in the selected proteases were generated and the deletion mutants were still able to degrade fibrinogen. Expression of eight additional *B. fragilis* proteases in *E. coli* also did not lead to detectable degradation of fibrinogen. Therefore, the proteases analysed in the present study cannot degrade fibrinogen on their own. Either the fibrinogenolytic protease is yet to be identified or there is redundancy in the proteases involved.

Growth under nitrogen limiting conditions urges the bacteria to 'search' for nitrogen from the environment to restore the depleting intracellular nitrogen content, which is essential for survival and biosynthetic processes. Therefore the differential regulation of the *B. fragilis* transcriptome under nitrogen abundant and nitrogen limiting conditions were examined to find a potential role for BF1705 and secreted proteases in nutrient binding and nitrogen assimilation. Whole transcriptome analysis under the two conditions was performed by RNA Sequencing, which reveals a snapshot of RNA presence and quantity from a genome at a given point in time. Although BF1705 was not significantly differentially regulated in response to nitrogen limitation, expression of four protease-encoding genes were induced over a 2-fold in the nitrogen limiting condition when compared to the nitrogen abundant condition. The molecular weight of three of these induced proteases were within the range of 45-55 kDa which corresponded to the lysis bands detected by fibrinogen zymography with wild-type *B. fragilis* protein extracts and therefore might potentially be involved in degrading fibrinogen. An alternative pathway for ammonia generation under nitrogen limiting conditions was suggested owing to a 155-fold upregulation of a gene encoding an asparagine synthetase B homologue. Formation of ammonia through the aspartate metabolism pathway via arginine degradation has not been previously characterised in *B. fragilis* and might account for the sustenance of *B. fragilis* growth under nitrogen limiting conditions. Moreover, the significant upregulation observed in the expression of genes involved in regulation of oxidative stress and metronidazole resistance will potentially assist with unravelling the mechanisms of infection mediated by this important opportunistic pathogen.

Chapter 1 Introduction: The Human Microbiome

The adult human body harbours a wide range of microorganisms, the cells of which ($\sim 10^{14}$) outnumber that of the human body ($\sim 10^{13}$) by an order of ten (Hattori and Taylor, 2009). The statistics are further magnified at the genetic level where the close-knit interactions between human and microbiome system could be considered as a human 'superorganism' model. Albeit the human microbiota is primarily composed of bacteria, a large number of diverse viruses, fungi and protozoa are also present in the human body (Breitbart and Rohwer, 2005; Furuse et al., 1983). Following a natural childbirth, infants obtain their initial microbial inoculum from the mother followed by a progression of events resulting in the development of a child's own microbiome. By three years of age, the microbiota stabilizes in composition and numbers while discretely being involved in shaping the host immune system (Proctor, 2014). One of the theories that explains the basis of the successful host-bacterial partnership is the hologenome theory. The theory advocates the evolution of an organism along with its microbiome, comprising parasites, mutualists, synergists and amenalists, as a result of co-development of the host and symbionts. Although the human microbiome is colon-biased, each region in the human body is defined by its own distinct microbial community. The human microbiome project has identified *Staphylococcus epidermidis* as a frequent colonizer of the skin with a higher prevalence of *Propionibacterium acnes* on facial skin and nose whereas *Streptococcus* spp. dominate the oral cavity. Among the intra-body microbiota, *Lactobacillus* spp. are predominant in the vagina and the genus *Bacteroides* is the most abundant in the gut of healthy subjects (Bakhtiar et al., 2013).

The adult human intestine contains an estimated 10^{13} - 10^{14} bacteria and archaea in association with $\sim 10^{14}$ - 10^{15} bacteriophages, viruses and parasitic eukaryotic microbes (Aziz, 2009). Examination of human faeces has revealed $\sim 10^{12}$ viable bacterial cells per gram of the luminal content, representing >1000 different species (Eckburg et al.,

2005; Hooper and Gordon, 2001). 16S rRNA analysis, performed as part of the Human Microbiome Project, has revealed that the majority of the human intestinal microbiota belong to four bacterial divisions : Firmicutes (64%), which includes the genera *Clostridium*, *Eubacterium* etc.; Bacteroidetes (23%), which includes the genus *Bacteroides*; Proteobacteria (8%) and Actinobacteria (3%) (Hattori and Taylor, 2009; Seksik et al., 2003). The remaining minor taxonomic divisions display great inter-individual diversity. The infant intestinal microbiota differs from the adult and is mostly composed of genera such as *Staphylococcus*, *Streptococcus*, *Bifidobacterium* and *Enterobacteria*; certain species of which are known to be transmitted from mother to baby (Hattori and Taylor, 2009; Ley et al., 2006). The composition of the microbiota within each adult is distinct and surprisingly stable over a normal healthy life span (Dethlefsen et al., 2007).

1.1. Commensalism in the Human Intestine

The gut microbiome, which is estimated to contain more than 100 times the number of genes in the human genome, endows the host with metabolic activities that humans are incapable of (Bäckhed et al., 2005). For instance, the host does not possess the enzymatic ability to degrade complex non-starch polysaccharides (dietary fiber) that enter the colon (Comstock, 2009). The anaerobic bacteria residing in the gut degrade these polysaccharides, including plant-derived pectin and cellulose, thus indicating a commensal relationship where the host gains carbon and energy, through the production of fermentative end products in the form of short-chain fatty acids (SCFA), and the microbes are provided with glycans and a protected anoxic environment (Bäckhed et al., 2005).

The *Bacteroides* spp. have evolved enzymatic systems to harvest both complex plant polysaccharides from the human diet and host muco-polysaccharides, derived from the mucus layer overlaying the intestinal epithelial surface. The genome of *Bacteroides thetaiotaomicron* contains polysaccharide utilization loci (PULs) that encode proteins involved in sensing, importing and degrading specific glycans of the colonic ecosystem (Martens et al., 2008). For instance, the *sus* locus involved in

starch utilization is comprised of eight genes encoding SusR and SusA-SusG, that include a regulatory protein, five outer membrane proteins involved in starch binding, degradation and import into the periplasm and two periplasmic glycohydrolases with neopullulanase and alpha-glucosidase activity (D'Elia and Salyers, 1996). Although different PULs are dedicated to different repertoires of functional products involved in the utilization of specific polysaccharides, in general, PULs contain genes encoding hybrid two-component histidine kinase response regulators, Extracytoplasmic Function (ECF)-type sigma factors and anti-sigma factors, outer membrane proteins involved in nutrient binding and import (SusD and SusC paralogues) and glycohydrolases that enzymatically cleave the glycosidic linkages of specific glycans (Martens et al., 2008). The genome of *B. thetaiotaomicron* strain VPI-8254 which contains 88 PULs efficiently utilises the host-derived glycans heparin, chondroitin and hyaluronan (Xu et al., 2007). Diet-associated changes in glycan foraging behaviour of *B. thetaiotaomicron* are accompanied by changes in the expression of its capsular polysaccharide synthesizing loci. This indicates that *B. thetaiotaomicron* is able to change its carbohydrate surface depending upon nutrient availability (Bäckhed et al., 2005). Other *Bacteroides* species possess glycolytic capabilities absent in *B. thetaiotaomicron*, such as the large complement of enzymes expressed by *B. vulgatus* that target pectin (Xu et al., 2007). In *Bacteroides fragilis*, the dietary or host glycans first bind to a glycan-specific SusD-like outer surface lipoprotein. The outer surface glycohydrolases degrade the polymer into smaller oligosaccharides which are then transported into the periplasm by the outer membrane TonB-dependent β -barrel SusC-like proteins. In the periplasm, additional glycohydrolases further degrade the molecules to monosaccharide components that are transported to the cytoplasm by sugar-specific permeases. The uncharged monosaccharides thus formed can either be destined for incorporation into bacterial glycans or for catabolism (Comstock, 2009). The *B. fragilis* genome encodes at least two enzymes, Fkp and a putative CMP-sialic acid synthetase, that can directly convert unphosphorylated host-derived monosaccharides (fucose and sialic acid) into their

nucleotide-activated forms for incorporation into bacterial glycans (Coyne et al., 2005). Fkp, a characteristic feature of intestinal Bacteroidales, is a protein which shares homology to mammalian L-fucose-1-P-guanyl transferase in N terminus and to L-fucose kinases in the C terminus, and is involved in generating GDP-L-fucose from exogenous L-fucose followed by its incorporation into capsular polysaccharide (CPS) glycan structures thereby linking L-fucose availability in the organism's intestinal habitat to CPS capsular structure (Coyne et al., 2005). The capsular polysaccharides, extracellular polysaccharides and glycoproteins thus formed, are essential for bacterial survival in the intestine and can also provide immunomodulatory properties to the host when they are released as outer membrane vesicles by the bacteria (Patrick et al., 1996). The other uncharged monosaccharides are acted upon by enzymes, such as RokA and HexA cytoplasmic hexokinases which phosphorylate hexoses on the sixth carbon (Brigham and Malamy, 2005). Subsequent to phosphorylation, the monosaccharides could be converted to their nucleotide-activated forms and directed into synthesis of bacterial glycans. Alternatively, the monosaccharides can be catabolised by different enzymes to provide SCFAs as an end product which is excreted into the lumen and utilised by host cells as an energy source. Butyrate, an SCFA has been demonstrated to have anticancer properties (Hamer et al., 2008). The participation of the monosaccharides in either glycan synthesis or catabolism is governed by the enzymes competing to divert substrates into these pathways.

Methanogenic archaea, such as *Methanobrevibacter smithii* provide the last link in the metabolic chain of polysaccharide processing where methanogens lower the partial pressure of any hydrogen formed by generating methane. The reduction in pressure may also increase microbial fermentation rates (Miller et al., 1982).

The indigenous gut inhabitants are also important for host defence because they inhibit the growth of potentially pathogenic microorganisms by the depletion of nutrients, prevention of access to adherence sites and production of inhibitory substances or conditions (Donskey, 2004). Certain *Lactobacillus* strains, which are

natural inhabitants of the gastrointestinal tract possess probiotic qualities and are beneficial in the treatment of murine colitis and against tetanus toxin (Ahrné et al., 1998; Grangette et al., 2002; Steidler et al., 2000).

Paneth cells, and dendritic cells present below the mammalian epithelia are sources of endogenous antimicrobial substances including lysozymes, phospholipases and various antimicrobial peptides, such as alpha-defensins, angiogenin 4 (Ang 4) and RegIII γ . *B. thetaiotaomicron* can stimulate production of Ang 4 that can kill pathogenic organisms such as *Listeria monocytogenes* (Hooper et al., 2003). It has also been shown that *B.thetaiotaomicron* promotes angiogenesis and post natal development of small intestinal villi in newborn mice (Stappenbeck et al., 2002). *B.thetaiotaomicron* induced peroxisome proliferators-activated receptor γ (PPAR γ)-mediated cytoplasmic redistribution of the Nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) subunit ν -rel avian reticuloendotheliosis viral oncogene homolog A (RelA) in intestinal cells, thereby attenuating the inflammatory response and promoting the intestinal survival of the microbe (Kelly et al., 2004). *Enterococcus* strains have also been shown to regulate phosphorylation of PPAR γ to induce the expression of interleukin (IL)-10, an anti-inflammatory cytokine (Are et al., 2008). In addition to anti-cancer properties, butyrate, one of the end products of glycan utilisation by intestinal microbes induces the expression of antimicrobial cathelicidin, LL-37. Induction of LL-37 controls unnecessary inflammatory responses by preventing excessive bacterial contact with the epithelium (Pamer, 2007).

Host interactions with the commensal gut microbiota, that aid in nutritional and immunomodulatory functions, are exquisitely equilibrated between pro-inflammatory and anti-inflammatory responses. The host preserves intestinal microbes while still being able to sense bacteria that penetrate intestinal borders. Selective pressures exerted on the intestinal microbiota by dietary constituents and antibiotic treatment can alter its stability. The growth of opportunistic pathogens, such as *Clostridium difficile*, under unstable conditions, can lead to undesirable

health outcomes (Donskey, 2004). Studies in humans and germfree mice have also shown that the respective levels of the two main intestinal phyla, the Bacteroidetes and the Firmicutes, are linked to obesity by carbohydrate metabolism (Gill et al., 2006; Ley et al., 2005; Turnbaugh et al., 2006). The microbiota of obese individuals are more heavily enriched with bacteria of the phylum Firmicutes and less with Bacteroidetes, suggesting that this mix is more efficient at extracting energy from a given diet than the microbiota of lean individuals. The genome of *Eubacterium rectale* (a member of Firmicutes division) is significantly enriched for glycoside hydrolases compared to several genomes of *Bacteroides* sp, which might explain a more efficient energy extraction by the former (Turnbaugh et al., 2006).

1.2. *Bacteroides fragilis*

The cultured human gut microbiota categorised within the Gram-negative Bacteroidetes phylum belong chiefly to the order Bacteroidales and comprises two predominant genera, namely *Bacteroides* and *Parabacteroides* (Comstock, 2009). *Bacteroides* spp. account for ~ 30% of the 10^{12} viable cells per gram colonic content that constitute the faecal microbiota (Hooper and Gordon, 2001; Salyers, 1984). *B. vulgatus* predominates the faecal microbial content at ~ 45% of the total bacteria present (Patrick et al., 1995a). The prevalence of *B. fragilis* in the faecal isolates ranges from 4-13% and is 10 to 100-fold lower than those of other intestinal *Bacteroides* spp. (Namavar et al., 1989; Salyers, 1984; Willis, 1991). However, *B. fragilis* is the most studied among the Gram-negative anaerobic gut bacteria owing to their significant role in glycan-synthesis and utilisation, and host immunomodulation. *B. fragilis* becomes part of the human microbiota from the earliest stages of life since they are passed from mother to child during vaginal birth (Turnbaugh et al., 2006). Although *B. fragilis* is poorly represented in the colonic microbiota, studies by Namavar et al. (1989) suggest that they comprise 42% of the adherent colonic mucosa which might explain why *B. fragilis* is the most frequent isolates from clinical specimens (Sheila Patrick et al. 1995; Wexler 2007). Bowel surgery is the most common cause of the accidental release of *B. fragilis* from its natural niche

into the peritoneal cavity, thus triggering abscess formation (Cohen-Poradosu et al., 2011).

1.2.1. Polysaccharide Synthesis

The great expansion of the PUL (polysaccharide utilisation loci) in the genomes of many members of *Bacteroides* genome is complemented by an expansion of loci involved in bacterial glycan synthesis. The *B. fragilis* NCTC 9343 genome encodes ~80 predicted glycosyltransferases and dedicates ~215,000 kb of its genome to the synthesis and regulation of glycans (Cerdeño-Tárraga et al., 2005; Coyne et al., 2008). The genome contains ten annotated regions (PS A-J) for the synthesis of polysaccharides, comprising antigenically distinct, within-strain variable, large capsule (extracellular polysaccharide), small capsule and micro-capsules (electron dense layer) (Cerdeño-Tárraga et al., 2005; Patrick et al., 1986, 2009). Although the *B. thetaiotaomicron* genome possesses a more extensive starch utilisation system with 172 glycosylhydrolases and 163 homologues of starch binding proteins when compared to *B. fragilis* genome, the latter is more adept at phase variation of the antigenic surface polysaccharides (Kuwahara et al., 2004). Variable micro-capsule expression is related to eight loci (PS A-H) in the genome where each locus is a set of 11-21 genes under the control of a single promoter. Phase variation is attributed to 7 of the 8 loci which are switched on and off by site-specific inversion of promoter sequences mediated by serine recombinases (Coyne et al., 2003; Liu et al., 2008; Patrick et al., 2003). Therefore a strain of *B. fragilis* is composed of subsets of populations simultaneously expressing a variable combination of capsular polysaccharides at any given point in time. Theoretically, 256 different surface profiles are possible due to within-strain phase variant micro-capsule synthesis which renders *B. fragilis* competent for host intestinal survival (Krinis et al., 2001). On the contrary, a situation where a sub-population of *B. fragilis* experiences the phasing off of all eight micro-capsules would not be encountered under wild-type conditions since the PS C promoter does not invert. Also, PS C is not expressed when other polysaccharides are locked on (Coyne et al., 2003). The inability of a stable

acapsular mutant, which is deficient in the synthesis of sugars associated with micro-capsule polysaccharides, to effectively colonise the intestine of a mouse has been reported by Coyne et al. (2008). Therefore, PS C might pose as a 'fool-proof' mechanism which ensures that the bacterium produces at least one capsule all the time. Nine of the ten polysaccharide biosynthesis loci contain genes predicted to encode Wzx and Wzy proteins which are flippases and polymerases (Patrick et al., 2009). In *E. coli* and *Salmonella enterica*, these proteins transport oligosaccharides across the inner membrane and polymerise the polysaccharides required for the production of lipopolysaccharides (LPS) and capsules (Whitfield, 2006). Post polymerisation, the polysaccharides are either ligated to a lipid-A-core by WaaL for LPS production or transported directly beyond the outer membrane by Wza for capsular polysaccharide production. The *Bacteroides fragilis* genome, however, encodes only one potential Wza homologue and no identifiable homologue of WaaL (Patrick et al., 2009). Deletion studies involving genes encoding a putative Wzz protein and WbaP protein in *B. fragilis* conducted by Patrick et al. (2009) indicated the dependence of the micro-capsules on Wzz, which suggested a link to LPS biosynthesis. Similarities in the mechanisms of synthesis of micro-capsules and phase-variable large capsule to the respective lipid A-linked K_{LPS}/O-antigen and to colanic acid-related group 1 and 4 capsules encoded by *E. coli* describes the variation in *B. fragilis*-mediated polysaccharide synthesis.

The synthesis of multiple phase-variable capsular polysaccharides is a conserved feature of the intestinal species within the order Bacteroidales which confers them with an advantage over other species for survival in the intestine (Coyne et al., 2008). Genomic comparisons of the *B. fragilis* strains, NCTC 9343, 638R and YCH46 revealed 28 divergent polysaccharide biosynthesis loci amongst the three strains. YCH46 shares one locus with NCTC 9343 and another with 638R (Patrick et al., 2010). This level of diversity in polysaccharide biosynthesis is unprecedented.

1.2.2. Polysaccharide A

Polysaccharide A (PS A) and PS B are two of the micro-capsule (MC) polysaccharides synthesised by *B. fragilis*. PS A and PS B along with several other MC-associated polysaccharides belong to a class of zwitterionic polysaccharides (ZPS) which are characterised by the presence of both a positive and a negative charge within each repeating unit (Tzianabos et al., 1993). ZPS have also been identified from other bacterial species, including type 1 *Streptococcus pneumoniae* capsular polysaccharide (CP1) and type 5 and type 8 *Staphylococcus aureus* capsular polysaccharides, however, PS A is the best characterised ZPS (Tzianabos et al., 2001).

ZPS are capable of inducing a CD4⁺ T cell-dependent immune response in a similar manner to protein antigens that evoke an IgG response. Since the activation of CD4⁺ T cells requires the specific interaction of the alpha-beta T cell receptor (TCR) on the T cell with the antigen presented by major histocompatibility complex II (MHC II) on a professional antigen-presenting cell, PS A is chemically processed in the endocytic pathway and presented on an MHC II molecule (Watts and Powis, 1999). Studies on oral treatment of mice with fluorescently labelled PS A have shown that PS A specifically associates with CD11c⁺ dendritic cells, but not with CD4⁺ T cells or CD19⁺ B cells in mesenteric lymph nodes (Mazmanian et al., 2005). Activation of CD11c⁺ dendritic cells proceeds through Toll-like receptor 2 (TLR2) stimulation, thereby upregulating MHC II as well as co-stimulatory CD86. Presentation of the antigen on MHC II in the presence of appropriate co-stimulatory molecules is essential for the subsequent activation of T cells (Pasare and Medzhitov, 2004; Troy and Kasper, 2010). Mono-colonisation of previously germ-free (GF) mice with *B. fragilis* alone corrected the CD4⁺ T cell deficiencies and resulted in a more defined splenic architecture with larger lymphoid follicles. Thus commensal bacteria have a positive impact on development of gut-associated lymphoid tissue, intraepithelial lymphocytes and mucosal IgA production. Experiments have shown that this effect is mediated by PS A which diminishes the CD4⁺ T cell deficiency in the spleen of GF mouse (Mazmanian et al., 2005). CD4⁺ T cells are comprised of

sub-types known as T-helper 1, T-helper 2 and T-helper 17 cells (Th1, Th2 and Th17), each of which induce distinct and opposing immune responses. Th1 cells promote the development of cell-mediated immunity and host defense against viral and bacterial pathogens. Th2 cells mediate host defence against helminths (Anthony et al., 2007; Murphy and Reiner, 2002). Th17 are important mediators of inflammation and protect the host from extracellular bacteria and fungi (Chen and O'Shea, 2008; Matsuzaki and Umemura, 2007). Imbalanced Th1 or Th2 responses may result in autoimmune diseases while an overactive Th2 response has been implicated in asthma and predisposition to allergy (Kidd, 2003; Wills-Karp et al., 2001).

Th1/Th2 imbalances in germ-free mice can be corrected through bacterial colonisation. Th1 cells are characterised by the production of interferon (IFN)-gamma and Th2 by the production of IL-4. The majority of mammalian immune systems including that of GF mice are intrinsically skewed towards Th2 cytokine production. Stimulation of the GF mouse immune system with PS A generated a Th1 response that corrected the imbalance and restored IFN-gamma production to levels found in conventionally colonized mice (Mazmanian et al., 2005). The ability of PS A to stimulate IFN-gamma production by T cells represents a Th1 response for which signalling by both IL-12 and signal transducer and activator of transcription protein (Stat4) (Th1-specific transcription factors) is required. TLR2 stimulation by PS A produces IL-12 from activated dendritic cells (Wang et al., 2006). IL-12 binds the IL-12 receptor on T cells and signals the activation of Stat4 (Macatonia et al., 1995). IL-10, a potent anti-inflammatory cytokine produced by CD4⁺ T cells is synthesized following signalling through TLR2 and is required for protection against inflammation. Therefore, TLR2 activation which was previously considered to promote immunity is resultant of PS A detection by the immune system thereby restraining anti-bacterial response (Round et al., 2011). Animal models of diseases replicating uncontrolled inflammatory responses, such as Inflammatory Bowel Disease (IBD) reveal that PS A administration prevents these diseases by reducing the production of pro-inflammatory cytokines, such as IL-23, tumor necrosis factor

(TNF)-alpha and IL-17 which, is a result of repression of Th17 cells and activation of IL-10 production (Mazmanian et al., 2008). A subset of CD4⁺ T cells known as foxp3⁺ regulatory T cells are required for *B. fragilis*-mediated Th17 suppression.

1.2.3. Outer Membrane Vesicles

Outer membrane vesicles (OMV) have been studied extensively in Gram-negative bacteria and contain biologically active proteins, lipoproteins, phospholipids and LPS, mostly of outer membrane and periplasmic origin, which perform diverse biological processes such as horizontal gene transfer, intercellular communication, transfer of contents to host cells and eliciting immune responses (reviewed in Kulp & Kuehn 2010; Kulkarni & Jagannadham 2014). OMVs released by Gram-negative bacteria are small spherical structures ranging from 10-300 nm in diameter (Kulp and Kuehn, 2010; McBroom and Kuehn, 2005). They are released by a mechanical process of 'bulging away and pinching off' from the outer membrane of the cell, which is independent of ATP/GTP hydrolysis (Kulp and Kuehn, 2010). Although vesicles have been best characterised in pathogens with respect to their toxin delivery and other virulence factors, they do perform a protective role in bacteria by reducing the levels of toxic compounds and aiding in the release of attacking phages, thereby promoting bacterial survival in a hostile environment (reviewed in Kuehn & Kesty 2005). OMVs enable bacteria to secrete insoluble molecules in combination with soluble material which are destined to distant, otherwise inaccessible, targets in the host in a concentrated, protected and targeted form (Kulp and Kuehn, 2010). The inability to generate non-vesiculating mutants in bacteria to date, except for modulation of OMV-release, emphasizes the significance of OMV-mediated functions in the growth of a bacterium (McBroom et al., 2006). Release of OMV by *B. fragilis* was first suggested by Macfarlane et al. (1992) who analysed protease synthesis by the bacterium and detected ~ 10% of the extracellular proteolytic activity to be association with particulate matter. Immuno-analysis with monoclonal antibodies, flow cytometric detection and further enzymatic analysis by Lutton et al. (1991) and Patrick et al. (1996) confirmed the production of OMV by the electron

dense layer (EDL)-enriched population of *B. fragilis*. EDL refers to a type of capsule expressed by *B. fragilis* which is not visible by light microscopy and accounts for <5% of the laboratory culture population. Labelling with monoclonal antibodies (mAbs) specific for surface polysaccharides indicated the presence of outer membrane epitopes in *B. fragilis* OMVs. Later, vesicles with diameter ranging from 30-80 nm was observed in *B. fragilis* by transmission-electron microscopy (TEM), and these OMV were smaller when compared to the vesicular sizes of many other Gram-negative bacteria (Elhenawy et al., 2014). Arousing the mammalian immune system is an inevitable consequence of OMVs since they deliver antigenic OM components, LPS and toxins (Kondo et al., 1993). Therefore certain bacteria, both pathogens and non-pathogens, have incorporated proteins into their OMVs which aid in immune suppression and promote bacterial survival in the host. UspA1/A2 virulence factors found in *Moraxella catarrhalis* OMVs were observed to bind and inactivate C3 complement protein and *Neisseria* OMVs removed bactericidal factors found in serum. OMVs released by *P. gingivalis*, a periodontal pathogen which is evolutionarily close to *B. fragilis*, caused CD14 degradation on human macrophages by incorporating gingipain proteases into their vesicular content (Duncan et al., 2004).

Similar to the immunosuppression exhibited by other bacteria, Shen et al. (2012) detected immunomodulatory polysaccharide PS A and PS B in the OMVs secreted by *B. fragilis*. A TLR2-mediated sensing of OMV-associated PS A presented by dendritic cells (DCs) activated regulatory T cells and IL-10 anti-inflammatory cytokine production in the presence of Growth Arrest and DNA-Damage-Inducible (*Gadd45 α*) protein. Expression of *Gadd45 α* protein by DCs has previously been shown to promote T cell responses (Jirmanova et al., 2007). Shen et al., (2012) suggested a therapeutic role of *B. fragilis* OMVs where OMV-treated DCs prevented experimental colitis through the production of IL-10 anti-inflammatory cytokine and the suppression of both TNF α production and Th17 development. However, it was

observed that the absence of PS A blocked the ability of *B. fragilis* OMV to promote regulatory T cells by dendritic cells (Shen et al., 2012).

Another mechanism by which OMVs support the survival of bacteria in the host environment is by the sequestration and inactivation of antibiotics. OMVs can pose as decoy targets that protect cells by titration of phages and antibiotics (Manning and Kuehn, 2011). In addition, proteins that degrade antibiotics have been identified in bacterial OMVs, such as β -lactamase in the OMVs released by *M. catarrhalis* and the Gram-positive bacterium, *S. aureus* (Lee et al., 2009; Schaar et al., 2011). Such enzymatic functions with respect to antibiotics also account for community-associated functions by which the OMVs provide transient antibiotic resistance for neighbouring cells, thereby promoting their survival within the host. OMVs facilitate long-term drug resistance in a population by horizontal gene transfer. The identification of cephalosporinase (BtCepA), which was exposed on the surface of vesicles produced by a majority of *Bacteroides* spp., was observed to degrade the antibiotic, cefotaxime, a third-generation cephalosporin. It was also demonstrated that these cephalosporinase-loaded OMVs could protect *Salmonella* typhimurium and *Bifidobacterium breve* from β -lactam activity (Stentz et al., 2015). Thus *Bacteroides* spp. which constitute a major part of the human gastrointestinal tract microbiota confer antibiotic resistance to enteric pathogens and commensals. Since there has been growing evidence of the adverse effects of antibiotic treatments which lead to intestinal dysbiosis and increased susceptibility to IBD, cephalosporinases secreted by *Bacteroides* might play a significant role in maintaining a balanced intestinal microbiota through resisting the inhibitory responses of cephalosporin antibiotic.

OMVs also play a role in bacterial nutrition, often exhibiting a community-related role by facilitating inter- and intra-species nutrient transfer. In addition to enzymes that degrade complex biomolecules and allow nutrient uptake, OMVs collect and concentrate metal ions for bacterial consumption (Kulp and Kuehn, 2010). Participation of OMVs released by marine bacteria in carbon flux between several

species of bacteria and cyanobacteria was demonstrated by Biller et al. (2014). In similar terms, a model proposed by Rakoff-Nahoum et al. (2014) suggested that OMVs secreted by certain species of the intestinal microbiota are able to break down polysaccharides for the benefit of other species in the community. Recent Mass Spectrometry (MS) analysis of the proteome of the outer membrane and OMV released by *B. fragilis* and *B. thetaiotaomicron* has revealed the presence of more than 40 proteins that are exclusive to the OMV. 24 of the proteins present in OMVs were hydrolases and 80% of the OMV-specific proteins were acidic in nature. The OMV-proteome was in stark contrast with the outer membrane proteins which contained very few hydrolases with only 2 out of the 33 OM-specific proteins possessing an acidic isoelectric point. The selective sorting of proteins in OMVs confirms the biogenesis of *B. fragilis* OMV as a directed process (Elhenawy et al., 2014). Similar patterns of cargo selection have been observed in *Pseudomonas aeruginosa* and *Porphyromonas gingivalis* by which particular LPS subtypes are enriched in OMVs to enhance their virulence-associated functions (Haurat et al., 2011; Kadurugamuwa and Beveridge, 1995). In addition to A-band LPS, OMVs released by *P.gingivalis* were enriched in a class of proteases known as gingipains which enabled degradation of host proteins. The regulation of genes encoding preferentially packed OMV proteins depends on the host environment. Enrichment of aminopeptidase was observed in OMVs released from clinical isolates of *P. aeruginosa* which facilitated association of vesicles with cultured epithelial cells when compared to lab strain-derived OMVs (reviewed in Kuehn & Kesty 2005). A similar upregulation in the activity of α -L fucosidase, an OMV-exclusive sugar hydrolase, was observed in *B. fragilis* in the presence of fucose. Since fucose metabolism, in addition to its utilisation in polysaccharides, is a major function of the *Bacteroides* spp., regulation of the OMV content to optimise fucose breakdown is of advantage to the species. Additionally, the glycosidases exclusive to *B. fragilis* OMVs can degrade a variety of polysaccharides which could be utilised by another bacterium deficient in glycolytic activity (Elhenawy et al., 2014). Furthermore, recent studies indicate that the sugars liberated by *B. thetaiotaomicron* nourishes

pathogens such as *S. enterica* and *C. difficile* and promotes their colonisation in the gut (Ng et al., 2013). Since Elhenawy et al. (2014) also suggested that the sorting mechanism is conserved within the genus despite variability among related species, a similar function of sugars liberated by OMVs might exist in *B. fragilis* as that observed in *B. thetaiotaomicron*. Thus the selective packing of acidic hydrolases along with the ability to agglutinate erythrocytes suggests a virulent role for *B. fragilis* OMVs which is discussed in detail in the next section (Patrick et al., 1996).

1.2.4. Pathogenic factors

B. fragilis accounts for 50%-70% of the intestinal anaerobes isolated from opportunistic infections including peritonitis, intra-abdominal sepsis, soft tissue abscesses, brain abscesses and bacteraemia (Lobo et al., 2013; Patrick, 2002; Patrick and Duerden, 2006; Patrick et al., 1995b). The *B. fragilis* group is responsible for 45-65% of nosocomial and community acquired anaerobic bacteraemia with a reported mortality rate of 24%-31% (Blairon et al., 2006; Brook, 1990). *B. fragilis* dons the role of an opportunistic pathogen when it escapes from the gastrointestinal tract into the peritoneal cavity owing to surgical, traumatic or a disease-induced disruption of the intestinal mucosal integrity (Coyne et al., 2001). *B. fragilis* accounts for only 4%-13% of the normal faecal microbiota and is isolated from 63%-80% infections caused by *Bacteroides* spp., whereas *B. thetaiotaomicron* which constitutes 15%-29% of the luminal content accounts for only 13%-17% of the infection cases (Cerdeño-Tárraga et al., 2005).

The adherent nature of the *B. fragilis* capsular polysaccharide surface which was evident in the sodium periodate-sensitive adherence to Caco-2 cells, allows the bacterium to establish at the site of an infection (Ferreira et al., 2002). In addition to adherence, the respective roles of large capsule and micro-capsules in resistance to phagocytic-uptake and serum bactericidal activity have been reported earlier by Reid & Patrick (1984). Therefore the presence of multiple intra-strain phase-variant polysaccharides ensure there is always a sub-population of *B. fragilis* that can circumvent the host immune response and persist in infections (Coyne et al., 2008).

It was suggested that incomplete opsonisation of the large capsule with limited amounts of an essential opsonin, such as IgM, led to impaired phagocytosis (Reid and Patrick, 1984). The most distinct and characteristic feature of the pathogenic phase of *B. fragilis* is abscess formation. Studies by Gibson et al. (1998) on mouse models identified the role of the capsular polysaccharide complex in coordinating cellular events that resulted in abscess development. Adherence of *B. fragilis* surface polysaccharides to murine mesothelial cells (MMCs) stimulated TNF- α production by peritoneal macrophages. Cytokine stimulation led to Intercellular Adhesion Molecule-1 (ICAM-1) production by MMCs which supported increased polymorphonuclear leukocyte (PMNL) binding to the site, thus inducing an abscess. Later on, several studies including generation of gene-deletion mutants narrowed down the sole abscess-inducing region of *B. fragilis* strain 9343 to the unusual free amino sugar residue, acetamido-amino-2,4,6-trideoxy galactose (AATGal), on the zwitterionic PS A repeating unit (Coyne et al., 2001). Each repeating unit of PS A contains both positively charged and negatively charged groups and this zwitterionic nature of the polysaccharide is essential for abscess induction. Conservation of the regions flanking each side of the PS A locus was evident in the Polymerase Chain Reaction (PCR) analysis of fifty *B. fragilis* isolates. Downstream conserved region includes two terminal PS A biosynthesis genes, *wcfR* which is likely responsible for amino sugar transfer on the AATGal residue and, *wcfS*, a putative transferase responsible for transfer of AATGal nucleotide precursor to an undecaprenyl-phosphate lipid carrier marking the first step in synthesis of the PS A repeating unit. Conservation of these genes suggested that this sugar is present in the PS A of all serotypes and might explain the abscessogenic nature of this bacterium across many strains. The study also predicted the presence of AATGal in a similar, but more complex PS A locus of *B. fragilis* strain 638R. However, whole genome sequence comparisons conducted by Patrick et al. (2010) on 3 strains, NCTC 9343, 638R and YCH46 revealed that the PS A/2 locus is not conserved and is exclusive to the NCTC 9343 genome. Nonetheless, the presence of putative transferases and dehydrogenases

in the PS loci of 638R and YCH46 suggested the possible production of a different cluster of abscess-producing polysaccharides in these strains.

As mentioned in the last section, PS A is delivered to the host by means of OMVs released by *B. fragilis* (Shen et al., 2012). Therefore the accidental displacement of the bacterium into the peritoneal cavity leads to the release of OMVs that promote abscess formation. Several studies have demonstrated a key role played by OMVs in various stages of infection. Hypervesiculating mutants were observed to combat conditions of stress in a better manner owing to the release of misfolded proteins, thereby safeguarding bacterial integrity (reviewed in Kulkarni & Jagannadham 2014). In events of infection-induced stress, bacteria responded by upregulating OMV production which was accompanied by increase in toxicity and activation of host immune responses (Kulp and Kuehn, 2010). Studies in enterotoxigenic and non-enterotoxigenic *E. coli* suggest that pathogenic bacteria release more vesicles than their non-pathogenic counterparts (Horstman and Kuehn, 2002). On comparison of various *B. fragilis* clinical isolates, Pumbwe et al. (2006) observed an accelerated OMV production in certain isolates which might suggest a pathogenic role. Enrichment of virulence factors in OMVs have been observed in pathogens, such as, *P. gingivalis* which excludes a multitude of outer membrane proteins (OMP) and preferentially packs gingipains, a proteolytic virulence factor in its OMVs (Haurat et al., 2011). Adherence of OMVs released by *Pseudomonas aeruginosa*, a respiratory pathogen, to the cultured lung epithelial cells was found to be strengthened by the OMV enrichment of aminopeptidase (reviewed in Kuehn & Kesty 2005). Proteome analysis in *B. fragilis* has identified a sorting mechanism by which acidic hydrolases are preferentially packed into OMVs. Although this accounts for a beneficial role wherein the glycolytic end products released by the OMVs promote self and inter-species nutrition in the gut, it could turn unfavourable when it enhances the survival rate of enteric pathogens (Elhenawy et al., 2014). Additionally, it has been observed that OMVs released by *Myxococcus xanthus* exhibit a predatory role and lytic activity against *E. coli* owing to the presence of a high number of digestive hydrolases in vesicles released by *M. xanthus* (Kahnt et al., 2010; Whitworth, 2011).

The selective packing of hydrolases in *B. fragilis* might also target the host mucus glycans and mucin protein which aid in epithelial invasion. Previous studies which demonstrate haemagglutination, sialidase, as well as other enzymatic activities with respect to *B. fragilis* OMVs suggest that the vesicles attack host cells following an initial attachment (Patrick et al., 1996; Tanaka et al., 1992). Vesicles purified from the oral bacterium *Treponema denticola* have been found to possess a protease, dentilysin, which allows the bacterium to penetrate the epithelial barrier by disrupting tight junctions (Chi et al., 2003). Furthermore, studies have been initiated to identify a possible role of OMVs released by colonic bacteria in depression-related disorders with regard to the 'leaky gut' syndrome. Previous studies have reported that OMVs can affect the blood-brain-barrier permeability which provides a lead to its role in entering systemic circulation and being delivered to the brain which might contribute to depression-related disorders (Muraca et al., 2015).

B. fragilis strains isolated from intestinal infections secreted very high levels of an antibacterial chemical, bacteriocin, and were resistant to bacteriocins secreted by other strains of *B. fragilis*. Bacteriocins are known to attack closely-related species or strains of the producers, thus competing for a similar ecological niche. However, non-intestinal isolates such as, strains isolated from intra-abdominal infections, bacteraemia and aquatic environments secreted lower amounts of bacteriocin and were more susceptible to that produced by other strains. A bacteriocin-producing *B. fragilis* strain which inhibited RNA polymerase specifically of other *B. fragilis* strains in vivo was found to have a slightly wider activity spectrum in vitro (reviewed in Wexler 2007). In addition to the factors discussed above, a wide range of potential virulence mechanisms are prevalent in *B. fragilis* which might promote the dissemination of opportunistic infections. The contributing factors include adherent surface proteins, fimbriae, enterotoxin, haemolysins, neuraminidase, and OMV-dependent and independent enzymatic activities which are explained in detail in the following chapters.

Chapter 2 Materials and Methods

2.1. Strains

The strains used in this study and their sources are shown in Table 2.1.

Table 2.1: List of strains

Strain	Genotype	Source
<i>E. coli</i> BL21 Star (DE3)	F ⁻ <i>ompT hsdS_B (r_B m_B) gal dcm rne131</i> (DE3)	Invitrogen
<i>E. coli</i> MG1655	F ⁻ λ ⁻ <i>ilvG- rfb-50 rph-1</i>	Maurice Gallagher
<i>E. coli</i> S17-1 (λpir)	<i>recA thi pro hsdRM+ RP4::2-Tc::mu::Km Tn 7 λ pir</i>	Simon <i>et al</i> 1983
<i>E. coli</i> TOP10	F ⁻ <i>mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (Str^R) endA1 nupG</i>	Invitrogen
<i>B. fragilis</i> ΔBF1708	NCTC 9343 ΔBF1708	David Roberts
<i>B. fragilis</i> NCTC 9343	clinical isolate	Sheila Patrick

2.2. Bacterial culture conditions

B. fragilis strain NCTC 9343 was grown in a MiniMacs anaerobic work station (Don Whitley Scientific, UK) at 37°C with an anaerobic gas mix (10% Hydrogen 10% Carbon dioxide and 80% Nitrogen), in Brain Heart Infusion broth (BHI) (Difco, USA) supplemented with 5% cysteine, 10% sodium bicarbonate, 50 µg/ml haemin and 0.5 µg/ml menadione or defined medium (DM) (Van Tassell & Wilkins, 1978) supplemented with 10% glucose or 2% fucose. *E. coli* strains were grown in Luria-Bertani Broth (LB) (Difco, USA) shaken at 200 rpm, or on LB agar at 37 °C. For analysing the effect of nitrogen limitation, *B. fragilis* was grown in glucose-DM containing +5x, -5x and -10x times the normal concentration of ammonium sulphate. Experiments for analysing nitrogen limitation were also performed by growing *E. coli* and *B. fragilis* in glucose-DM where ammonium sulphate was substituted with

with 0.2% glutamine or glutamic acid and sub-culturing in 0.1% glutamine or glutamic acid. Strains were stored in a final concentration of 50% glycerol at -80°C.

2.3. Oligonucleotides

Oligonucleotides used in this study are shown in Table 2.2. The Artemis genome browser was used to design oligonucleotide primers to amplify coding regions or flanking regions (Carver et al., 2008). PCR amplification was carried out as described in 2.5.3.

Table 2.2: List of oligonucleotides

Oligonucleotide primers designed for gene deletion and gene amplification along with their loci in artemis genome browser. 4 sets of primers were designed to amplify the flanking regions of each gene to be deleted. Regions of the primer sequence indicated in green are for compatibility between inside primers designed for each gene. Regions of the primer sequence indicated in blue are for compatibility with the cloning vector. Forward and reverse primers were designed for each protease-encoding gene to be amplified. Region of the forward primer sequence indicated in red is for compatibility with the expression vector used. Commercial primers used in library construction and sequencing experiments are also shown.

Primer (locus in artemis)	Sequence
0275 left outside forward (304517..304536)	ATCGATAAGCTTGATAAGACTACACCGGTAGAACC
0275 left inside reverse (305002..305024)	AGCCATGGCTCCGGACTTATCAGAATCAAAGAAAGCG
0275 right inside forward (306285..306304)	TTTGATTCTGATAAGTCCGGAGCCATGGCTGAACC
0275 right outside reverse (306796..306815)	CTGCAGGAATTCGATGTGGCAATCCAGATCAAAC C
0657 left outside forward (788674...788693)	ATCGATAAGCTTGATTCTCCGTTCCGCCACTACC
0657 left inside reverse (789182.....789203)	AATCCTTTTCGTTAGGGTGTTCCTTTTCATGATCG G
0657 right inside forward (790170....790194)	TGAAAAGAAAACACCCTAACGAAAAGGATTATTAA GAACG
0657 right outside reverse (790690.....790711)	CTGCAGGAATTCGATTGATACAACCGGTAAGTTTC GG
1979 left outside forward (2312254.....231227	ATCGATAAGCTTGATGTGAAGATAGTAGTAATATTT CG

6)	
1979 left inside reverse (2312804....2312825)	TATCCGTTATTCCACGACTTGTCAGTATAAGAGTTCC
1979 right inside forward (2314026....2314048)	TTATACTGACAAGTCGTGGAATAACGGATAAGTAA TCG
1979 right outside reverse (2314582....2314604)	CTGCAGGAATTCGATGTAAGCAATAGATAGATATCTCG
3775 left outside forward (4457633...4457652)	ATCGATAAGCTTGATGCAACTGATGGCAGAAATCG
3775 left inside reverse (4458178....4458197)	TTAAAGTATCGCATCGTGCCCGATCTGTGGAAAGC
3775 right inside forward (4459425....4459446)	CCACAGATCGGGCACGATGCGATACTTTAACGTATCC
3775 right outside reverse (4460026...4460046)	CTGCAGGAATTCGATAGACTATCGCGATTTTCTTCG
1705 left outside forward (1984400-1984422)	ATCGATAAGCTTGATTAATCAATCGATCAAGAAGTAGC
1705 left inside reverse (1984913-1984932)	CCAAAGACTCTCCAAAAGCCTGGAAGCCATTCCG C
1705 right inside forward (1986372-1986393)	TGGCTTCCAGGCTTTTGGAGAGTCTTTGGAACATGC
1705 right outside reverse (1986914-	CTGCAGGAATTCGATGTATTGAAACAGCCGAAGGC

1986934)	
0935 forward 1170883...1170897	CACCATGCAGCAAATGCCGCCC
0935 reverse 1168138...1168164	TGCAAGGTCTTATTATTTCTTTTCCGG
1335 forward 1598108...1598125	CACCATGAAAATCGGTGGAAAGAAA
1335 reverse 1597397...1597423	TTACTTGTTCTTAGTATAGTTGTCTGC
1496 forward 1757435...1757452	CACCATGGGAAAGGAAGTACTCGAC
1496 reverse 1757915...1757941	TCTATTTCTGTTCAATCGTCATAGCC
2517 forward 2932159...2932173	CACCATGGGAGAAGCACGGTTG
2517 reverse 2929000...2929026	ATAAGTTTATTTACTAAAATCTCTCGG
3021 forward 3512952...3512966	CACCATGACCGATGCCGGAAAC
3021 reverse 3510915...3510941	ATACAAGCCGATCTATAGTCCCCT
3247 forward 3822630...3822644	CACCATGACAAACACTGCTGAG
3247 reverse 3820662...3820685	TTAATCATTTATAGAAGGCAGCGG
3732 forward 4402046...4402060	CACCATGCAGGATTTGATTGCC
3732 reverse 4401236...4401259	GTAGTATAACTATTATGAACAGCG
3918 forward 4617550...4617564	CACCATGAAGAAAGAGGCCGTA
3918 reverse 4615588...4615611	AATCTATTACCAAATTGACACCCG

TruSeq Universal Adapter	AATGATACGGCGACCACCGAGATCTACACTCTTTC CCTACACGACGCTCTTCCGATCT
Illumina Multiplexing Read2 Sequencing Primer	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

2.4. Bacterial techniques

2.4.1. *B. fragilis* growth curves

A single colony was used to inoculate 5 ml BHI-S broth or DM, containing antibiotics if appropriate and incubated overnight at 37°C anaerobically. A 1:100 dilution was made into fresh pre-reduced BHI-S broth or DM without antibiotic selection and incubated anaerobically at 37°C. Optical density readings at 600 nm (OD₆₀₀) were taken at 45 or 60 min intervals for strains grown in BHI-S and DM, respectively. Before readings were recorded, samples were diluted in the appropriate medium to ensure the OD₆₀₀ readings did not exceed 1. Measurements were multiplied by the dilution factor to obtain the optical density. Readings were taken until stationary phase was reached. All growth experiments were performed in three or six biological replicates.

2.4.2. Preparation of chemically competent cells

Solutions

Transformation Buffer 1 (TFB1): 30 mM Potassium acetate, 10 mM CaCl₂, 50 mM MnCl₂, 100 mM RbCl, 15% v/v glycerol (pH 5.8) in 100 ml dH₂O

Transformation Buffer 2 (TFB2): 10 mM MOPS, 75 mM CaCl₂, 10 mM RbCl, 15% v/v glycerol (pH 6.5) in 10 ml dH₂O

Competent *E. coli* cells were prepared using the rubidium chloride method. 20 mM MgSO₄ was added to a 1:100 dilution of overnight culture in pre-warmed LB. The cells were grown at 37°C until an OD₆₀₀ of 0.4. The cells were pelleted by

centrifugation at 4500xg at 4°C for 5 min and resuspended in 100 ml ice-cold TFB1 before incubation on ice for 5 min. Cells were again pelleted by centrifugation and resuspended in 10 ml ice-cold TFB2. Cells were incubated on ice for an hour before snap freezing in liquid N₂ and storing at -80°C in 100 or 200 µl aliquots.

2.4.3. Transformation: Heat-shock method

For transformation, 2-5 µl of DNA were added to 100 µl competent cells. The transformation mix was incubated on ice for 30 min before giving a heat shock at 37°C for 2 min followed by incubation on ice for 5 min. The cells were revived on LB for an hour. The revived cells were spread onto LB plates containing the antibiotic of selection and incubated overnight at 37°C. Individual colonies were patched and streaked on to LB plates containing the antibiotic and incubated overnight at 37°C.

2.4.4. Transformation: Electroporation method

The overnight culture of *B. fragilis* NCTC 9343 was diluted 1:40 in pre-warmed and pre-reduced BHI-S broth and grown anaerobically at 37°C to an OD₆₀₀ of 0.4. 10 ml of the culture were centrifuged at 2600xg for 15 min at 4°C. The pelleted cells were washed in 10 ml ice-cold pre-reduced dH₂O. Washing was repeated three times and the cells were finally resuspended in 100 µl dH₂O. The washed cells were left on ice until required. 40µl of cells were placed in a pre-chilled 0.1 cm gap Gene Pulser® electroporation cuvette (Bio-Rad), to which 1 µl of DNA was added. The cells were electroporated in a MicroPulser™ (Bio-Rad) at a voltage of 2kV. 1ml of pre-warmed BHI-S was quickly added to the cells followed by anaerobic incubation at 37°C overnight. The cultures were centrifuged in a 5415 D centrifuge (Eppendorf) at 13000 rpm for 2 min and then resuspended in 200µl of pre-warmed BHI-S. The cells were spread on BHI-S plates containing appropriate antibiotics and incubated for 1-2 days in an anaerobic atmosphere at 37°C. The antibiotic resistant colonies were re-streaked on BHI-S plates containing the specific antibiotic to obtain single colonies.

2.4.5. Conjugation of *E. coli* S17-1 λ pir with *B. fragilis* NCTC 9343

Cultures of the donor *E. coli* S17-1 λ pir and *B. fragilis* were grown overnight. *E. coli* was sub-cultured 1:50 into fresh LB and *B. fragilis* 1:40 into fresh BHI-S without antibiotic selection. Both the cultures were grown to an OD₆₀₀ of 0.2 and donor to recipient were mixed in a ratio of 1:10. 200-300 μ l of the mixed cultures were spotted onto BHI-S plates and incubated aerobically for 1-2 hours before transferring into the anaerobic cabinet for overnight incubation. The cells were washed and scraped from the plate using 1 ml BHI-S. 200 μ l aliquots were spread onto selective BHI-S plates containing gentamicin and the appropriate antibiotic, and incubated overnight at 37°C in the anaerobic cabinet. Resistant colonies were picked and restreaked on BHI-S agar plates containing gentamicin and the appropriate antibiotic.

2.4.6. Screening for Deletion genotypes

The colonies resistant to the appropriate antibiotic were streaked on DM containing fucose (0.2%) and the antibiotic, and incubated anaerobically for 1-2 days at 37°C. The individual colonies were picked and patched on BHI-S plates and BHI-S plates containing the sensitive antibiotic using sterile toothpicks. The sensitive colonies represented strains that had resolved to either wild-type or deletion genotype.

2.5. DNA techniques

2.5.1. Genomic DNA extraction

Genomic DNA was extracted from 5 ml overnight cultures using the Wizard Genomic DNA purification kit (Promega) following the manufacturer's instructions. The extracted DNA was resuspended in 100 μ l rehydration solution.

2.5.2. Plasmid DNA extraction

Cultures were grown overnight in 5 ml broth, containing antibiotics if appropriate. Plasmid DNA was extracted using QIAfilter plasmid mini kit (Qiagen) following the manufacturer's instructions.

2.5.3. Polymerase Chain Reactions

Oligonucleotides were synthesised by MWG Operon (Eurofins, Germany) and the sequences are shown (Table 2.2). PCR reactions were performed in a Techne Progene Thermal Cycler (Bibby Scientific Limited, UK) or a Mastercycler gradient cycler (Eppendorf). A standard 50 µl reaction contained 1 µl forward primer (10 pmol/µl), 1 µl reverse primer (10 pmol/µl), 1 µl dNTP's (10mM), 10 µl 5x reaction buffer, 0.5 µl Phusion polymerase (2000 U/ml) (New England Biolabs, UK), 1 µl ~0.1 ng/µl template DNA and 35.5 µl dH₂O. The thermal program was 1 cycle at 98°C for 1 min, 30 cycles at 98°C for 30 seconds, x°C for 30 seconds, 72°C for 45 seconds per kb, followed by a final extension of 72°C for 15 min, where x represents the annealing temperature of each primer pair.

2.5.4. Agarose Gel Electrophoresis

Solutions

50X Tris Acetate EDTA (TAE) electrophoresis buffer: 242g Tris base, 57.1 ml Glacial acetic acid, 100 ml 0.5M EDTA (pH 8.0) in 1l dH₂O

Agarose gel electrophoresis of DNA fragments was carried out in 1% (w/v) agarose gels (Invitrogen, USA), with 1x Tris-acetate EDTA buffer (TAE). 1 x DNA loading buffer (New England Biolabs, UK) was added to DNA samples before loading onto the gel alongside molecular weight markers, 100 bp or 1 kb DNA ladder (New England Biolabs, UK). Electrophoresis was carried out at 80V for 60 min. Gels were stained in 1 µg/ml ethidium bromide for 10 min and de-stained in dH₂O for 10 min. Gels were visualised using a UVP VisiDoc-It Imaging system (UVP LLC, USA).

2.5.5. DNA purification

PCR products were purified using the Wizard SV Gel and PCR Clean-up System (Promega, UK) following the manufacturer's instructions and stored at -20°C.

2.5.6. DNA quantification

DNA concentration (ng/μl) was determined using a nanodrop ND-1000 spectrophotometer (Thermo Scientific, USA) by measuring the absorbance of the sample at a wavelength of 260 nm according to the manufacturer's instructions.

2.5.7. Restriction Digestion

Digestions were carried out using restriction enzymes and their appropriate buffers as outlined in the manufacturer's instructions (New England Biolabs, UK). A standard 50 μl digestion contained: 0.5 μl BSA, 5μl 10X reaction buffer, 3 μl restriction enzyme, 10 μl plasmid DNA and 33.5 μl dH₂O. The reaction mixes were incubated at 37°C overnight for digestion.

2.5.8. Cloning

Cloning of the deletion constructs into pGB910 vector was performed in a ligation-independent manner using the Infusion cloning kit (Clontech). The vector and insert were mixed in appropriate ratios according to the manufacturer's instructions along with the enzyme mix. The expression constructs of the protease-encoding genes were cloned into pET100/D-TOPO vector using Champion pET Directional TOPO Expression Kit (Invitrogen) by following directions outlined in the manufacturer's instructions.

2.5.9. Single Colony Assay

Solutions

Single Colony Forming (SCF) buffer: 2.5% Ficoll, 1.25% SDS, 0.015% Bromophenol blue, 10 μg/ml RNase A

Cells from a patch plate were re-suspended in 200 μl SCF buffer and incubated at room temperature for 20 min. Samples were then centrifuged at 13000 rpm for 30 min. 30 μl of supernatant were loaded onto 1% agarose gel, and the appropriate supercoiled plasmid DNA were loaded in the first and last lane as size reference markers.

2.5.10. Sanger-based Plasmid Sequencing

5µl clean DNA (33-83 ng/µl) and 1µl sequencing primer (3.2 pmol/µl) were mixed in 0.2 ml tubes. The plasmids were sequenced by Sanger sequencing at the Edinburgh Genomics facility.

2.6. RNA Sequencing

2.6.1. RNA Extraction

Solutions

Nuclease-free water (Qiagen)

DiEthyl PyroCarbonate (DEPC) (Sigma-Aldrich)

Ethanol

β-Mercaptoethanol

15 mg/ml lysozyme (Sigma-Aldrich)

Proteinase K (Qiagen)

Tris EDTA (30 mM Tris from 1M DNase- and RNase-free Tris-HCl (Sigma-Aldrich) and 1 mM EDTA from 0.3 M EDTA stock concentration) pH 8.0

RNase ZAP (Sigma-Aldrich)

Prior to RNA extraction, water and other solutions were treated with 0.1% DEPC to inactivate RNases by covalent modification. 0.1 ml DEPC solution were added to 100 ml solution, shaken vigorously and incubated for 12h at 37°C. DEPC treated solutions were autoclaved at 100°C for 15 min to remove DEPC traces. Glasswares were made RNase-free by treating with 0.1% DEPC-water for 12h at 37°C followed by heating to 100°C for 15 min to remove residual DEPC. Nuclease-free water was used in processes requiring pure water. The work-bench and solid surfaces associated

with RNA extraction were wiped with RNase ZAP to remove RNase contamination. Hand gloves were changed frequently to inhibit RNase introduction.

Total RNA was extracted from *B. fragilis* cultures grown to appropriate time points by following enzymatic lysis and proteinase K digestion using RNA protect Bacteria Reagent (Qiagen) and purified using RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Extracted RNA samples were quantified using Nanodrop (2.5.6). In-solution DNase digestion of the RNA samples was performed using RNase-free DNase kit (Qiagen) according to the manufacturer's instructions. Final clean-up and concentration of the RNA samples was carried out using RNA Clean and Concentrator-5 (Zymo Research) following manufacturer's instructions. The concentrated RNA was eluted in 30 µl elution buffer and quantified again using Nanodrop. rRNA was depleted from the extracted RNA using Ribo-Zero-rRNA Removal Kit for Bacteria (Illumina) as per the manufacturer's instructions at the Edinburgh Genomics Facility.

2.6.1.1. Agarose Gel Electrophoresis of RNA

1X TAE was prepared in DEPC-treated water since adding DEPC directly to Tris buffers inactivates primary amines. The electrophoresis tank and combs were wiped with RNase ZAP to remove RNase contamination. 10 µl RNA was treated with 2 µl 6X DNA dye (New England Biolabs, UK) and 12 µl formamide. The RNA samples were heated at 65-70°C for 10 min and electrophoresed on 1% agarose-TAE gels at 70V for 30 min.

2.6.2. Library Preparation

Stranded libraries were constructed for each RNA sample using TruSeq Stranded Total RNA Library Prep Kit (Illumina) at the Edinburgh Genomics facility. 10 cycles of PCR were performed using TruSeq Universal Adapter (Table 2.2) according to the manufacturer's instructions. The libraries were pooled and paired-end sequencing was performed in HiSeq v4 HO sequencer (Illumina) using Illumina Multiplexing Read2 Sequencing Primer (Table 2.2) according to the manufacturer's instructions.

2.7. Protein techniques

2.7.1. Preparation of Whole cell protein extracts

B. fragilis and *E. coli* strains were grown overnight in BHI-S or glucose-DM. Cells were pelleted by centrifugation at 16000xg for 5 min and re-suspended in 500 µl 3x SDS Sample Buffer (New England Biolabs, UK). Samples were stored at -20°C.

2.7.2. Preparation of Concentrated culture supernatants

Concentrated supernatants were prepared based on a previously described method (Horstman and Kuehn, 2000) with modifications for *B. fragilis*. *B. fragilis* was grown overnight in BHI-S or glucose-DM (200-250 ml), were aseptically transferred to polypropylene containers and cells were harvested by centrifugation at 10000xg at 4°C for 10 min using a Sorvall RC-5B Refrigerated Superspeed Centrifuge (Dupont Instruments). The supernatant was separated from the cell pellet and centrifuged for a further 15 min at 4°C. The supernatant was filtered using 0.45 µm low protein-binding syringe filters (Millipore) following which the filtrate was aseptically transferred to 100 kDa Molecular Weight Cut-Off (MWCO) Centricon Plus-70 centrifugal filters (Millipore), sealed using parafilm tape and concentrated following the manufacturer's instructions. Supernatant (~25 ml at a time) was concentrated at 4°C by centrifugation at 3500xg for 15-40 min until 200 ml of culture supernatant were concentrated to ~2 ml volume. The concentrate was recovered by spinning the inverted concentrate cup at 1000xg for 2 min. 5µl of 3x SDS-PAGE sample buffer (New England Biolabs, UK) were added and mixed with 20 µl of each concentrated supernatant sample. Samples were stored at -20°C.

2.7.3. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Solutions

10X Tris-glycine running buffer: 30g Tris base, 144g Glycine in 1l dH₂O

1X Tris-glycine running buffer (Working concentration): 100 ml 10X stock, 10 ml 10% SDS, 890 ml dH₂O

Separating gel buffer: 1.5 M Tris-HCl pH 8.8

Stacking gel buffer: 0.5 M Tris-HCl pH 6.8

12% Separating gel solution: 4.8 ml 30% acrylamide/BIS, 3 ml separating gel buffer, 120 μ l 10% SDS, 4 ml dH₂O, 8 μ l TEMED, 100 μ l 10% APS

5% Stacking gel solution: 1.5 ml 30% acrylamide/BIS, 2.25 ml stacking gel buffer, 90 μ l 10% SDS, 5 ml dH₂O, 8 μ l TEMED, 100 μ l 10% APS

Concentrated supernatants and other protein samples were analysed using SDS-PAGE. Glass plates were cleaned with ethanol and dH₂O, and assembled according to manufacturer's instructions (Bio-Rad Mini Protean III). Gels measuring 70x85x0.5mm and consisting of separating gel and stacking gel were electrophoresed in parallel for 60 min at 125V.

2.7.4. Staining/Destaining of gels

Solutions

Staining solution: 25% Methanol, 10% Acetic acid, 64.9% ultrapure water, 0.1% w/v Coomassie brilliant blue R-250

Destaining solution: 25% Methanol, 10% Acetic acid, 65% ultrapure water

Gels were briefly rinsed in distilled water and stained for 2h at room temperature with gentle agitation in staining solution. After 2h, the protein gels were destained for 16h to 24h at room temperature with gentle agitation in 300 ml destaining solution. When clear bands were observed, the gel images were digitally saved using the Image Lab Software in a ChemiDoc instrument (Bio-Rad) following manufacturer's instructions.

2.7.5. Zymography

Solutions

Fibrinogen solution (Stock): 2 mg/ml human fibrinogen (Sigma-Aldrich) was layered on top of 0.85% w/v NaCl and placed in 37°C water bath for 4-6h with frequent

gentle agitation. The solution was then filtered through a 0.22 µm syringe filter (Millipore) and stored at 4°C.

10% Separating gel solution: 3.333 ml 30% acrylamide/BIS, 2.5 ml separating gel buffer, 0.1 ml 10% SDS, 4.5 ml 2.0 mg/ml plasminogen-free human fibrinogen dissolved in ultra-pure water, 5µl TEMED, 50 µl 10% APS

4% Stacking gel solution: 0.666 ml 30% acrylamide/BIS, 1.25ml stacking gel buffer, 50 µl 10% SDS, 3 ml ultra-pure water, 2.5 µl TEMED, 50 µl 10% APS

Serine protease activating buffer: 50 mM tris-HCl, 150 mM NaCl, 5 mM CaCl₂ pH 8.2

Metalloprotease activating buffer: 50 mM tris-HCl, 200 mM NaCl, 5 mM CaCl₂, 1 µM ZnCl₂, 0.02% Brij-35 pH 7.6

Zymography was performed based on a previously described method (Houston et al., 2010). Following SDS-PAGE using 10% Separating gel solution and 4% Stacking gel solution, the gels were washed to remove SDS which prevents proteases from renaturing and exerting their proteolytic activity on the copolymerised fibrinogen. Each zymogram was washed 2x15 min at room temperature with gentle agitation on a rocker (Model R100 Luckham) in 250 ml of 2.5% triton-X-100 dissolved in ultra pure water. The zymograms were then washed 2x15 min at room temperature with gentle agitation on the rocker in 250 ml of 2.5% triton-X-100 dissolved in 50 mM tris-HCl (pH 7.5). Finally, washes were performed in 50 mM tris-HCl (pH 7.5) as described above.

The washed gels were incubated in appropriate activation buffers for 48h at 37°C with gentle agitation on a rocker.

2.7.6. Preparation of samples for Immunoblot analysis

Appropriate concentrations of whole cell pellets, concentrated supernatants and purified protein were treated with 1X SDS sample buffer containing 0.1 M DTT. Samples and Color Prestained Protein Standard (New England Biolabs, UK) were

boiled for 10 min at 98°C. The treated samples were electrophoresed on a SDS PAGE, comprising 12% separating gel and 5% stacking gel for 60 min at 150V.

2.7.7. Transfer of proteins by Wet Electroblothing

Solutions

1X Transfer buffer: 14.4 g Glycine, 3.02 g Tris base, 100 ml Methanol in 1 l dH₂O

1X Tris Buffered Saline-Tween 20 (TBS-T) buffer: 6.06 g Tris, 8.77 g NaCl, 500 µl Tween 20 in 1 l dH₂O

Blocking Buffer: 5% dried skimmed milk powder (Marvel) in TBS-T

Polyvinylidene fluoride (PVDF) membrane was cut to the size of the SDS protein gel and wetted in methanol followed by soaking in cold transfer buffer. Fibrous pads and filter papers were also soaked in cold transfer buffer. The protein gel was then placed in the 'transfer sandwich' (filter paper-gel-membrane-filter paper), cushioned by fibrous pads and pressed together by a cassette. The gel and membrane were inserted into a transfer tank (Bio-Rad) filled with transfer buffer. The electroblotting was performed at 100V, 350A for 60 min. Ice packs were placed in the tank to mitigate the heat produced. After the transfer, the membrane was blocked overnight at 4°C in 5% skimmed milk blocking buffer.

2.7.8. Development of Western blots and Far-western blots

The blocked membrane was washed 3x15 min in TBS-T. Far-western blotting was performed by incubating the washed membrane blot in TBS-T containing 10 µg/ml fibrinogen or fibronectin (Sigma-Aldrich) at room temperature for an hour. The membrane was rinsed in TBS-T to remove the unbound fibrinogen. Protein incubation step in far-western blotting was followed by steps in western blotting wherein the membrane was incubated in appropriate dilutions of primary antibody (1:10000/1:20000) in blocking buffer for an hour on a rocker (Model R100 Luckham). The primary antibodies used were Anti-Fibrinogen antibody produced in goat (Sigma-Aldrich), Anti-Fibronectin antibody produced in rabbit (Sigma-Aldrich)

and Anti-1705 antibody produced in rabbit (Patrick Lab). The unbound primary antibody was rinsed off with TBS-T and the membrane was incubated for an hour in HRP-conjugated secondary antibody (1: 40000) for detection. Anti-Goat IgG (whole molecule)-Peroxidase antibody produced in rabbit and Anti-Rabbit IgG (whole molecule)-Peroxidase antibody produced in goat were the respective secondary antibodies used. Excess antibody was rinsed off and the membrane blot was treated with a 1:1 mix of detection reagent 1 and detection reagent 2 from Amersham Enhanced chemiluminescence (ECL) Prime Western Blotting Detection Reagent (GE Healthcare) in the developing room. Excess reagent was shaken off after 5 min. The membrane blot was wrapped in a plastic saran wrap and the air pockets were smoothed out carefully. Lights were switched off and an Amersham imaging hyperfilm ECL (GE Healthcare) was placed on top of the blot. The film was exposed for 1 min for initial detection, and was developed in an X-Ray film developer. The exposure was repeated, varying the time for optimal detection of signals.

2.7.9. Stripping Antibodies from Immunoblots

Stripping Buffer: 1.5g Glycine, 1.24g Tris, 0.58g NaCl, 0.5g SDS, 1 ml Tween 20 in 100 ml dH₂O pH 2.2

Antibodies were stripped off to facilitate reprobing of the PVDF membrane blots with a different set of antibodies. The membrane blots were removed from saran wrap and submerged in pure H₂O. The blots were placed on a rocker for 15 min followed by washing in TBS-T for 2x15 min. TBS-T was then replaced with stripping buffer and placed on the rocker for 30 min at room temperature. Blots were washed in H₂O for 3x5 min. Finally, H₂O was replaced with blocking buffer and kept at 4°C without shaking for at least 1h.

2.7.10. Protein Expression

Solutions

Binding Buffer: 5 mM Imidazole, 50 mM Hepes pH 7.5, 500 mM NaCl, 5% Glycerol

The protease-encoding genes were expressed in *E. coli* BL21 cells using the Champion pET Directional TOPO Expression Kit (Invitrogen) as per the instructions outlined by the manufacturer.

For BF1705 expression, a single colony of the *E. coli* strain transformed with pET15b recombinant vector carrying the BF1705 insert was inoculated into 1 ml of LB containing 50 µg/ml ampicillin, which was then immediately inoculated into 10 ml of LB containing ampicillin and incubated aerobically in a shaker at 37°C, 225 rpm until the OD₆₀₀ reached 0.8. Once the required growth was achieved, the whole 10 ml of culture were inoculated into 1l of pre-warmed LB containing 50µg/ml ampicillin in a 2.5l capacity ful-baf Tunair shake flask capped with dri-gauze filter lining (Sigma-Aldrich). The large scale culture was incubated at 37°C and 230 rpm in an orbital incubator (Gallenkamp, England) until the OD₆₀₀ reached between 0.9-1.0. The culture was then cooled on ice to 15-18°C. 500 µl of 1M Isopropyl-β-D-thiogalactoside (IPTG), giving a final concentration of 0.5mM, were added to the culture to induce expression of the recombinant protein. Protein induction and overexpression was performed between 15-18°C for 16h at 200 rpm in an orbital incubator (Gallenkamp, England). Recombinant protein expression was carried out at a low temperature in order to facilitate folding, solubilisation and to allow the host cells to direct most of their energy to recombinant protein expression. 10 ml of overnight growth were stored at -20°C for quality control analysis. The remaining 990 ml of culture were harvested by centrifugation at 6000xg at 4°C for 20 min in a Sorvall RC-5B Refrigerated Superspeed Centrifuge (Dupont Instruments). The supernatant was discarded and the pellets were resuspended by vortexing in 35 ml of cold sterile filtered binding buffer. The cell suspension was then poured into four 50 ml Falcon tubes and treated with 2 tablets of protease inhibitor cocktail (Sigma-Aldrich) by mixing. The suspension was then flash frozen in dry ice layered with ethanol for 5-10 min and stored at -80°C until required.

2.7.11. Protein Purification by Nickel Affinity Chromatography

Solutions

500 ml loading buffer: 1.56g NaH₂PO₄, 14.61g NaCl, 0.68g Imidazole, 1 ml PMSF (0.05 M PMSF dissolved in 99% ice-cold ethanol), 6 protease inhibitor tablets pH7.5

100 ml elution buffer: 0.312g NaH₂PO₄, 2.92g NaCl, 3.4g Imidazole pH 7.4

1l Gel filtration buffer: 0.271g NaH₂PO₄, 1.405g Na₂HPO₄, 8.0g NaCl, 0.8 ml DTT, 0.292g EDTA, 100 ml Glycerol pH 7.4

Frozen samples were thawed on ice and the cells were disrupted by passing the sample through a French press along with 35 ml cold sterile filtered loading buffer. The pressed samples were pelleted at 14000 rpm for 60 min at 4°C and the supernatant was filtered through 0.2 µm filters (Millipore). The filtered samples were stored on ice. 5 ml nickel solution were added to a ready-made Nickel-nitrilotriacetic acid (Ni-NTA) sepharose column (Qiagen) and equilibrated with 5X loading buffer (25 ml). The filtered sample was loaded onto the column and the flow-through was collected in a Falcon tube. The column was washed a second time with 25 ml loading buffer and the flow-through collected. Filtered elution buffer was added to the column and the flow-through was collected in 1.5 ml micro centrifuge tubes. After protein elution, the column was washed with loading buffer, sealed and stored in the refrigerator.

Approximately 5 ml of the nickel-column purified sample was subjected to higher grade purification by gel-filtration column chromatography. Pre-packed Hi-Prep Sephacryl S-200 HR gel filtration column (GE Healthcare Life Sciences) of 120 ml volume (1.6 cm x 60 cm), with a separation range of 5 kDa-250 kDa was used to obtain purified BF1705 protein. Prior to chromatography the column was equilibrated with sterile filtered gel filtration buffer. The elutes from column chromatography were electrophoresed on 12% SDS-PAGE.

2.7.12. Protein Estimation by Bradford assay

The quantification of proteins was performed by Bradford assay (Bio-Rad). Standards (0-12 $\mu\text{g/ml}$) were prepared using Bovine Serum Albumin (BSA) (800 μl) and Bradford reagent (200 μl) to a final volume of 1 ml. The unknown samples were prepared in 1:100 or 1:1000 dilution ratio and made up to 1 ml using Bradford reagent (Bio-Rad). The samples were mixed and left in the dark for 15 min. The OD was measured at 595 nm. Absorbance was plotted against concentration of the known BSA samples. The concentrations of unknown samples were calculated using the standard curve equation (Figure 2.1).

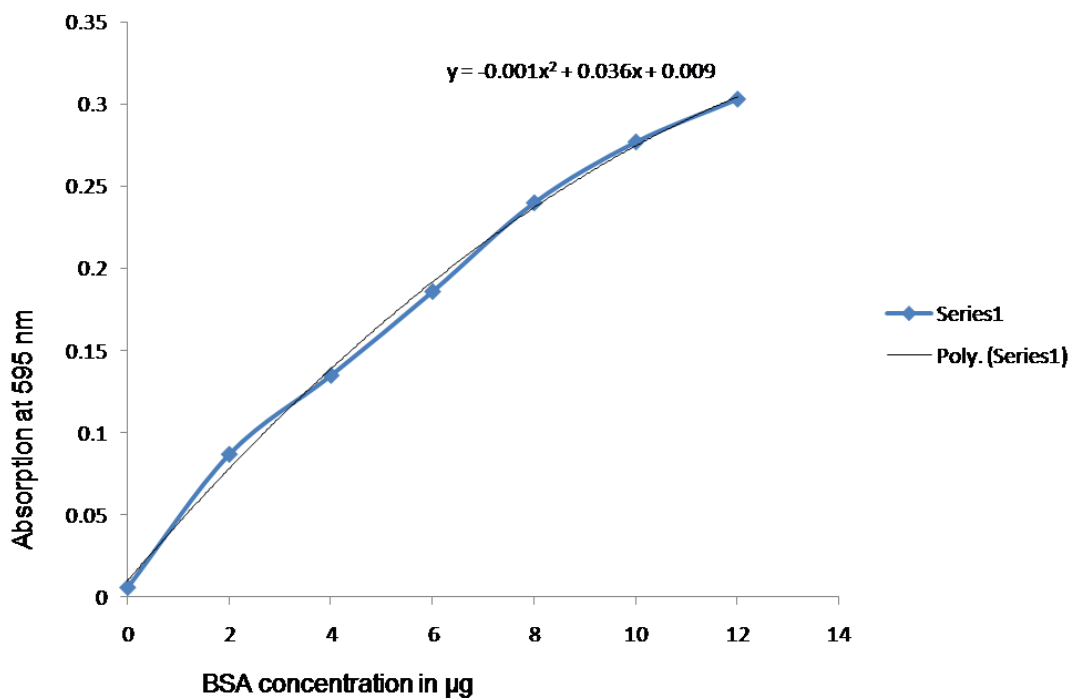


Figure 2.1: BSA standard curve using the Bradford assay for protein estimation

Readings at OD_{600} was obtained for 0-12 $\mu\text{g/ml}$ BSA concentrations plotted in a log-based graph to derive standard curve polynomial equation for protein estimation.

2.7.13. Immunofluorescence Microscopy (IFM)

IFM experiments were performed based on a previously described method (Houston et al., 2010) with slight modification. *B. fragilis* NCTC 9343 was grown to late exponential phase (16h) in BHI-S broth. 10 ml of culture were then harvested by centrifugation at 1610xg for 25 min at room temperature. The cell pellet was washed 3 x 20 min in autoclaved PBS by centrifugation at 1610xg at room temperature. The pellet was then resuspended in sterile PBS to OD₆₀₀ 0.3.

2.7.13.1. Assay 1 (On-slide fibrinogen incubation)

20 µl of resuspended cells (OD₆₀₀ 0.3) were pipetted onto 12 well Teflon poly-L-lysine coated slides and air-dried at 37°C. Slides were then fixed in 100% methanol at -20°C for 20 min. The fixed slides were dried at room temperature and washed in sterile Phosphate-buffered saline (PBS) bath for 20 min. The fixed cells were incubated with 20 µl 1.0 mg/ml or 0.1 mg/ml fibrinogen for 12-14h. Excess fibrinogen was washed off by placing the cells in a PBS bath. Care was taken in the subsequent steps to prevent the slides from drying. To inhibit non-specific binding, slides were blocked by adding 10 µl 2% BSA-PBS to each well and incubating at room temperature for 15 min. BSA-PBS was aspirated and 20 µl of primary antibody (1:300/1:3000 dilution) were added to the slides and incubated in a humidified box at 37°C for 1h. Anti-Fibrinogen antibody produced in goat (Sigma-Aldrich F8512) was the primary antibody used. Excess antibody was washed off using a PBS bath and the region in between the wells were wiped dry using a filter paper. The slides were then incubated with Fluorescein isothiocyanate (FITC)-conjugated secondary antibody to detect fluorescence (1:100 dilution). Anti-Goat IgG (whole molecule)-FITC antibody produced in rabbit was the secondary antibody used. Secondary antibody incubation was followed by PBS wash to remove excess antibody. The slides were mounted with glycerol-PBS and sealed with a cover slip using varnish. The sealed slides were then examined under a microscope in 100X oil emulsion and in fluorescence mode.

Two sets of negative controls were used : Negative control 1 (No fibrinogen added): 20 µl of resuspended cells (OD₆₀₀ 0.3) were fixed onto 12 well Teflon coated slide.

Negative control 2: 20 µl of 100 µg/ml plasminogen-free human fibrinogen in ultrapure water were applied to 12 well Teflon coated slide, dried for 3h and fixed.

2.7.13.2. Assay 2 (*Pre-incubation with fibrinogen in vitro*)

Resuspended cells (OD₆₀₀ 0.3) were incubated with 100 µg/ml and 1 mg/ml plasminogen-free human fibrinogen at 37°C for 2h and 12h with occasional agitation. Following incubation, excess fibrinogen was removed by washing the cells using centrifugation at 1610xg for 3x10 min in sterile PBS. 20 µl of resuspended cells were applied and fixed to 12 well Teflon coated slides and immunolabelled as described above (2.7.13.1).

2.8. Bioinformatics

2.8.1. Quality control of RNA extracts

The quality of the extracted RNA samples were analysed before and after rRNA depletion using RIN software (Agilent Technologies) at the Edinburgh Genomics Facility.

2.8.2. RNA Sequencing data, Assembly and Annotation

The quality of the read pair sequences obtained in the fastq format from the Edinburgh Genomics Facility were assessed using the FastQC tool (FastQC v0.11.3). The adapter sequencers from the read pair files were trimmed using the cutadapt tool (cutadapt 1.8.1). Reads shorter than 20 nucleotides and longer than 50 nucleotides were discarded. Total number of reads processed and the number of reads that passed the filters are shown in Table 2.3. The trimmed reads were aligned against the reference genome using Bowtie tool (Bowtie2-2.2.5). The parameters chosen for alignment were 'end-to-end' in the 'sensitive' mode. The aligned output from Bowtie was obtained in SAM format. The SAM files were converted to BAM format by running the Samtools (SAMtools-1.0). The Samtools command 'sort' was used to sort the alignment file by name and by position. Index command was used to index the alignment file. The aligned BAM files were viewed in artemis genome

browser. The aligned reads were counted to quantitate the genes using htseq-count (HTSeq-0.6.1). The information about the differentially expressed genes were obtained using deseq tool (DESeq2_1.6.1). The test type and dispersion method were set as 'Wald' and 'parametric', respectively.

Table 2.3: List of read pairs after trimming the adapter sequences

Trimming of the adapter sequences was performed using cutadapt tool. The table indicates the number of read pairs processed from each read pair input file, read1 and read2 with adapter, pairs that were too short and too long, and number of read pairs that passed the specified filters.

Input files	Total read pairs processed	Read1 with adapter	Read2 with adapter	Pairs that were too short	Pairs that were too long	Pairs that passed filters
Gu1_1.fastq.gz, Gu1_2.fastq.gz	19,747,613	1,024,191	1,131,848	37,374	0	19,710,239
Gu2_1.fastq.gz, Gu2_2.fastq.gz	15,194,819	758,531	879,063	23,096	0	15,171,723
Gu3_1.fastq.gz, Gu3_2.fastq.gz	18,350,715	997,195	1,061,692	23,385	0	18,327,330
Gu4_1.fastq.gz, Gu4_2.fastq.gz	17,762,906	911,064	1,015,511	40,811	0	17,722,095
Gu5_1.fastq.gz, Gu5_2.fastq.gz	15,553,594	829,198	876,360	51,302	0	15,502,292
GuB_1.fastq.gz, GuB_2.fastq.gz	18,321,211	897,982	1,033,971	65,964	0	18,255,247
GuC_1.fastq.g	17,509,499	868,934	975,750	58,125	0	17,451,

z,GuC_2.fastq.gz						374
GuD_1.fastq.gz,GuD_2.fastq.gz	16,127,604	836,730	906,408	112,927	0	16,014,677
GuE_1.fastq.gz,GuE_2.fastq.gz	14,148,510	712,967	794,959	80,767	0	14,067,743
GuF_1.fastq.gz,GuF_2.fastq.gz	10,563,179	521,522	590,523	56,645	0	10,506,534
N1_1.fastq.gz, N1_2.fastq.gz	15,425,050	810,697	879,416	28,172	0	15,396,878
N3_1.fastq.gz, N3_2.fastq.gz,	20,837,400	1,140,003	1,182,767	39,175	0	20,798,225
N4_1.fastq.gz, N4_2.fastq.gz	19,525,023	975,698	1,112,828	35,228	0	19,489,795
N5_1.fastq.gz, N5_2.fastq.gz	14,952,840	799,977	848,364	29,354	0	14,923,486
N6_1.fastq.gz, N6_2.fastq.gz	23,240,069	1,262,839	1,331,587	52,223	0	23,187,846

Chapter 3 Binding of Host Proteins by *Bacteroides fragilis*

3.1. Introduction

Completion of the human microbiome project in 2012 unveiled the plethora of microbes that thrive within the human body, occupying different habitats such as mouth, gut, skin and vagina. This vast microbiota is subject to variation within and between individuals (Huttenhower et al., 2012). Microbes exist in a symbiotic relationship with the host, however, occasional breaches of which, resulting in unfavourable health conditions have been documented (Winkler et al., 1987). Bacteria constitute the majority of the microbial inhabitants, along with viruses and bacteriophages (Breitbart et al., 2003). The presence of smaller proportions of archae and eukarya has also been reported (Eckburg et al., 2003).

Bacteria thrive in a wide range of microhabitats and metabolic niches in mucus layer lining the gut and mucosa, which require adherence to the mucus layer and deep penetrations into the crypt channels to withstand the rapid turnover of epithelial cells and fast rate of peristalsis. They must also be competent to overcome mechanical defence mechanisms operating at tissue surfaces (Smith 1995). Certain bacteria, such as those belonging to the genera *Enterococcus*, *Bifidobacterium* and *Bacteroides* can flourish in biofilms, wherein individual groups of bacteria form microcolonies or associate with other species and inhabit the mucosal layer and particulate surfaces in the gut lumen facilitating nutrient harvest and exchange (Macfarlane and Dillon, 2007). Extracellular matrix (ECM) is a major structural component compartmentalising and safeguarding the epithelium, lining the various multicellular structures from the surrounding endothelial cells and connective tissue. Although the commensal microbiota adheres to the mucosal layer and ECM for effective colonisation, ECM adhesive mechanisms employed by pathogens are of major therapeutic interest. Such interactions have been studied extensively with respect to both endogenous opportunistic pathogens and exogenous pathogens seeking passage through the ECM basement membrane lining the epithelium for spreading infections.

ECM adhesins range from carbohydrate-protein interactions to protein-protein bivalent soluble target proteins (Westerlund and Korhonen, 1993).

This chapter deals with the ability of *Bacteroides fragilis*, an opportunistic pathogen associated with blood-related infections, to bind host fibrinogen, explained in terms of pathogenic microbial interactions with the ECM macromolecules observed in various bacteria.

3.1.1. Structure of the Extracellular Matrix (ECM)

Extracellular matrix, consisting of a basement membrane and an interstitial matrix, is an important structural component underlying epithelial cells, endothelial cells and surrounding connective tissue cells (Hay, 1991) (Figure 3.1). ECM is comprised of three major classes of macromolecules, namely proteoglycans, adhesive glycoproteins and fibres (Szöke et al., 1996). Resident cells secrete these molecules into the matrix via exocytosis and fill the interstitial space by forming an interlocking network which acts as a shock-absorber in the matrix. The basement membranes are sheet-like depositions of ECM present in close proximity to the epithelial cells (Figure 3.1). Although the composition of ECM varies between multi cellular structures, the common functions include segregating different cell types, providing tissue structure and elasticity and serving as a reservoir for growth factors by isolating and protecting them in the microenvironment. ECM also acts as a structural barrier, preventing the entry of invading pathogens into the endothelium and bloodstream. Components of the ECM engage in interactions with specific adhesion receptors on cell surfaces and promote multiple cell functions including adhesion, migration, proliferation and differentiation (Hay, 1991).

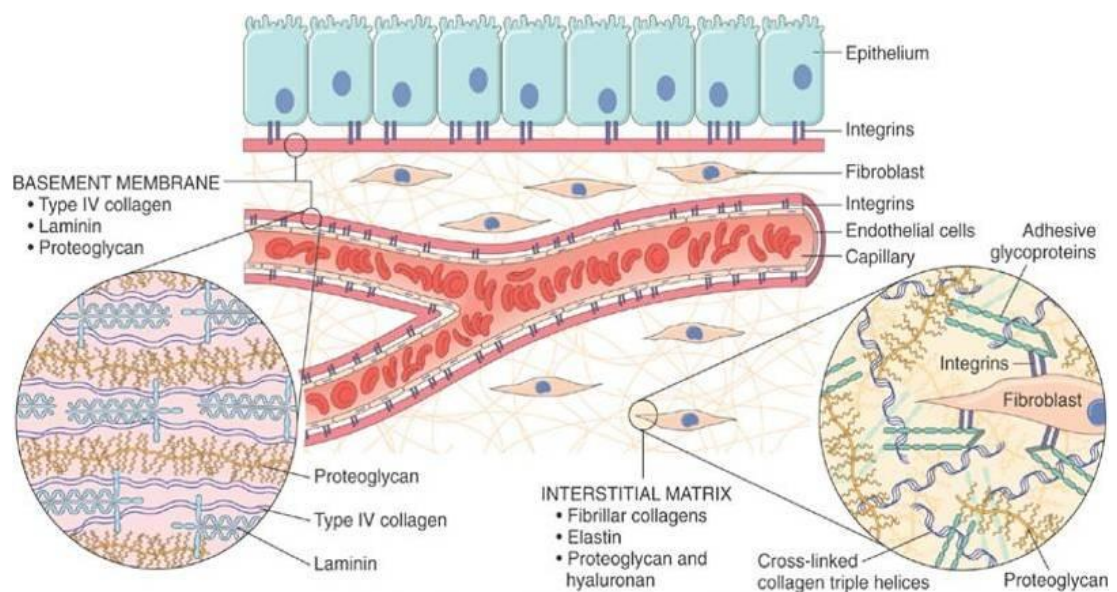


Figure 3.1: Schematic of the extracellular matrix

Basement membrane and interstitial matrix are components of the extracellular matrix underlying the epithelial cells, endothelial cells and surrounding connective tissue cells. Collagen, laminin and proteoglycan that constitute the basement membrane, which is linked to the epithelium and endothelium via integrins are illustrated. The interstitial matrix is composed of fibrillar collagens, elastin, proteoglycan and hyaluronan. (www.studyblue.com, Rosalind Franklin University)

ECM proteins exhibit functional motifs which recognise and adhere to specific cells and proteins. The adhesive glycoproteins also share the ability to form fibrils and aid in structural support. The major adhesive proteins in the ECM and basement membrane include:

3.1.1.1. Fibronectin

Fibronectin is a 440 kDa dimeric glycoprotein, consisting of repeating units of amino acids forming three domains, namely, type I, II and III present in 2 similar disulphide linked subunits, each of ~250 kDa in size. The domains enable fibronectin to interact with a variety of cells through both integrin and non integrin receptors. Alternative mRNA splicing of the type III domain leads to 20 isoforms of fibronectin in humans (Schwarzbaauer, 1991). Formation of mature mRNA encoding fibronectin from pre mRNA via alternative splicing varies depending on the cell type. Presence of

variable amounts of alternative splice forms results in the presence of two types of fibronectin in vertebrates. These are, soluble plasma fibronectin, a component of blood plasma produced by hepatocytes, and insoluble cellular fibronectin, a major ECM component, secreted locally by fibroblasts as a soluble dimer and later immobilized into the matrix (Romberger, 1997). Assembling fibronectin into matrix fibrils requires the interaction of the former with activated integrin receptors on cell surfaces (Wu et al., 1995). Fibronectin contains discrete binding sites for extracellular molecules, such as fibrin, heparin and collagen (Joh et al., 1999). Migration of cells through the ECM is brought about by fibronectin mediated cellular connection with collagen fibers. At a site of tissue injury, fibronectin binds to platelets via integrins and facilitates cellular movement for wound healing (Pereira et al., 2002).

3.1.1.2. Collagen

Collagen is a fibrous protein accounting for 30% of total proteins in mammalian bodies (Watanabe, 2004) . Various types of collagen are secreted by connective tissue as well as other cell types and constitute a major structural component of skin and bone (Alberts et al., 2002; Watanabe, 2004). Collagens are long, stiff proteins formed by three α polypeptide chains helically wound around one another in a rope-like superhelix (Alberts et al., 2002). Pro- α chains, precursors of individual collagen polypeptide chains are synthesised on membrane-bound ribosomes and injected into the lumen of endoplasmic reticulum (ER). Apart from the amino terminal signal peptide, pro- α chains also possess additional amino acids, called propeptides, at their N- and C- terminal ends. In the lumen, selected prolines and lysines undergo hydroxylation to form hydroxyproline and hydroxylysine, respectively, followed by glycosylation of specific hydroxyprolines. Each pro- α chain then combines with two others to form hydrogen-bonded, triple-stranded helical molecule known as procollagen. Post secretion, propeptides of fibrillary procollagen molecules are removed by specific extracellular proteolytic enzymes. Collagen molecules thus formed assemble in ECM to form larger collagen fibrils. Polypeptide chains of

collagen molecules are composed of numerous repeats of a tripeptide unit, Gly-Pro-X, where X is usually a post-translationally modified hydroxy-proline. High content of proline is a distinct feature of collagens (Watanabe, 2004). Glycine, which is the second most abundant amino acid residue in collagens, contributes to the tight structure of the protein. 20 types of collagen have been identified, which are formed from 25 distinct collagen α chains, each encoded by a separate gene. Connective tissue secretes types I, II, V and XI collagens, type I being the principle collagen present in skin and bone. Type I is the most common form of collagen and is a representative of fibril-forming collagens, which, after being secreted into the extracellular space, assemble into higher-order collagen fibrils (Figure 3.1). The fibrils then cluster into larger cable-like bundles known as collagen fibers. Collagen type IX and XII form the fibril-associated collagens, which constitute the surface of collagen fibrils and connect fibrils to one another and other ECM components. Type IV and VII are network-forming collagens. Type IV collagen forms a meshwork assembly in the basal laminae whereas type VII assemble into structures known as anchoring fibrils in the skin (Alberts et al., 2002) (Figure 3.1).

3.1.1.3. Laminin

Laminins are heterotrimeric glycoproteins of high molecular weight (~400 kDa), containing α , β and γ chains, and are found in the basal laminae where they intersect to form cross-like structures that bind to other ECM molecules, such as, type IV collagen (Figure 3.1). Laminins are large, multi-domain matrix glycoproteins occurring in 5, 4 or 3 genetic variants that are named according to the chain composition. 15 isoforms of laminin have been identified, with laminin-I being the classical laminin composed of 3 long polypeptide chains (α , β and γ) arranged in an asymmetric cross and held together by disulphide bonds. Several isoforms of each type of laminin chain can associate in different combinations to form large family of laminins. Like type IV collagen, laminins can self-assemble in vitro into felt-like sheet through interactions with laminin ends which might play a role in formation of laminin network in the basement membrane (Alberts et al., 2002). Laminin helps in

cell adhesion and regulates cell behaviour by integrin interaction through several functional domains, which may help in organizing the assembly of basal lamina (De et al., 2006).

3.1.1.4. Elastin

Elastin is the major constituent of insoluble elastic fibers and are about 750 amino acids long (Alberts et al., 2002; Vrhovski and Weiss, 1998). Two major components that contribute to the complex structure of elastic fibers include an amorphous component, consisting of extensively cross linked elastin protein which makes up more than 90% of elastin fiber and a fibrillar component, composed of microfibrils rich in acidic glycoproteins and structured into 8 nm-16 nm fibrils of beaded appearance. Elastin is rich in non-glycosylated hydroxy-proline and glycine. Soluble tropoelastin, precursor of elastin, is secreted into extracellular space and assembled into insoluble elastic fibers near the plasma membrane (Figure 3.1). Crosslinking between lysine residues makes the mature elastin protein insoluble. 11 human tropoelastin splice variants have been identified with respect to 6 exons. The highly cross-linked mature elastin fibers provide elasticity and resilience to tissues which require ability to deform repetitively and reversibly. Tissues rich in elastin include aorta and major vascular vessels, lung, elastic ligaments, tendon and skin. Long, inelastic collagen fibres are interwoven with elastic fibers to restrict extent of stretching and prevent tissue tearing.

3.1.1.5. Glycosaminoglycans

Glycosaminoglycans (GAGs) are unbranched polysaccharide chains composed of repeating disaccharide units. One of the 2 sugars in repeating disaccharide is an amino sugar (N acetyl glucosamine/galactosamine), which is sulphated in most cases and a second uronic acid sugar (glucuronic/ iduronic). GAGs are highly negatively charged owing to the abundance of sulphate or carboxyl groups on most of the sugars. GAGs can be classified into four main groups depending upon their sugars, type of linkage between sugars and the number and location of sulphate groups, namely hyaluronan, chondroitin sulphate and dermatan sulphate, heparan sulphate, and

keratan sulphate. The rigid nature of polysaccharide chains block them from being folded up into globular structures of polypeptide chains and promotes adoption of highly extended conformations occupying huge volume and forming gels at very low concentrations. Due to the high density of negative charges on GAGs, large amounts of water gets sucked into the matrix which creates a turgor pressure and helps to withstand compressive forces. GAGs contribute to less than 10% of the weight of fibrous proteins in connective tissue. However, they fill most of extracellular space by porous hydrated gel formation and provide mechanical support to the tissue.

i. Hyaluronic acid

Hyaluronan is the simplest GAG and are disaccharides consisting of up to 25000 alternating residue repeats of D-glucuronic acid and non sulphated N-acetylglucosamine. They are found in variable amounts in all adult tissues and fluids, but are abundant in early embryos. In hyaluronan, the disaccharide units are generally not covalently linked to a core protein. Hyaluronans are synthesised directly on cell surface by an enzyme complex embedded in plasma membrane. Functions of hyaluronan depend on specific interactions with other ECM molecules. They resist compressive forces in tissues and joints and act as space filler during embryonic development. Hyaluronan regulates wound healing and is present in joint fluid as a lubricant.

ii. Proteoglycans

Proteoglycans are GAG chains covalently linked to a core protein, which is synthesised on membrane-bound ribosomes and threaded into ER lumen. The assembly of polysaccharide chains on the core protein takes place in golgi apparatus and is initialised by a tetrasaccharide link on a serine side chain followed by glycosyltransferase mediated addition of sugars. Proteoglycans differ from glycoproteins on the nature of sugar side chains. Atleast one of the sugar side chains in a proteoglycan has to be a GAG. Glycoproteins contain 1-60% carbohydrate by weight, made up of numerous, relatively short, branched oligosaccharide chains,

whereas proteoglycans contain 95% carbohydrate by weight composed mostly of long, unbranched GAG chains. The core proteins can vary greatly in the number and types of attached GAG chains. Thus, proteoglycans can be regarded as highly glycosylated glycoproteins whose functions are mediated by both core proteins and GAG chains that constitute them. Three proteoglycans present in vascular and epithelial basement membrane of mammalian organisms include perlecan, agrin and bamacan.

Heparan sulphate is a linear polysaccharide present in all animal tissues and is closely related to heparin. It consists of variably sulphated repeating disaccharide units, the most common unit being glucuronic acid linked to N-acetylglucosamine which contributes to 50% of the total disaccharide units. In its proteoglycan form, 2-3 heparan sulphate chains are attached in close proximity to ECM proteins and regulates biological activities such as developmental processes, angiogenesis, blood coagulation and tumour metastasis. Perlecan is a heparan sulphate proteoglycan present in basal lamina of kidney glomerulus and filters molecules passing into urine from bloodstream. Heparin sulphate chains also bind to fibroblast growth factors, stimulates cell proliferation and oligomerises growth factor molecules enabling them to cross link and activate cell-surface receptors. These proteoglycans immobilise chemokines on endothelial surface of a blood vessel at inflammatory sites, thereby stimulating White Blood Cells (WBC) migration to injury sites. Syndecans are plasma membrane proteoglycans carrying up to three chondroitin sulphate and heparin sulphate GAG chains. They are located on the surface of many types of cells including fibroblasts and epithelial cells, and serve as receptors for matrix proteins. In fibroblasts, syndecans modulate integrin function by interacting with fibronectin on cell surface and with signalling and cytoskeletal proteins within the cell. Chondroitin sulphate chains provide the tensile strength to cartilage, tendons, ligaments, aorta walls and are also implicated in regulation of cell migration and pattern formation in developing peripheral nervous system (Iozzo, 1998). Bamacan is a chondroitin sulphate proteoglycan and is located in basement membrane of Bowman's capsule and within mesangial matrix (McCarthy et al., 1994). Keratan

sulphate chains have a variable sulphate content and are present in cornea and bones. Aggrecan is a keratin sulphate proteoglycan, which is a major component of cartilaginous tissues (Iozzo, 1998).

3.1.2. Microbial Surface Components Recognised by Adhesive Matrix Molecules (MSCRAMMs) in Bacteria

Owing to their adhesive properties, the components of ECM act as major substrates for microbial attachment. Adherence to the ECM is facilitated by MSCRAMMs (Microbial Surface Components Recognised by Adhesive Matrix Molecules) present on the bacterial surface (Patti et al., 1994). For a component to be termed as an MSCRAMM, it has to be:

- i. Present on the bacterial surface
- ii. The microbial component must recognise a macromolecular component of the ECM
- iii. The recognition must be of high affinity and specificity without the interference of other molecules

Binding of microbial cell surface receptors to adhesive glycoproteins of the ECM plays a dual role in host-microbial relationships. Adherence mechanisms of commensal bacteria enable members of the normal microbiota to transform into opportunistic pathogens in different parts of the human body (Nagy et al., 1994). Following tissue injury, the components of ECM get exposed on the epithelial surface which allows recognition and adhesion by microbial components (Patti et al., 1994). Bacterial binding to plasma proteins like fibronectin mask the immunogenic epitopes and prevent recognition of bacteria by the host immune system (Dinkla et al., 2003). The bacterial colonisation of ECM might initiate translocation through endothelial tissue, invasion of adjacent tissues and eventually reaching the bloodstream, thus playing a major role in aetiology of infection dissemination (Szöke et al., 1996). Therefore cell surface adhesins expressed by

bacteria can promote adhesion, invasion and evasion of host defences and antimicrobial agents (Patti et al., 1994).

3.1.2.1. MSCRAMMs in Staphylococcal and Streptococcal groups

Several MSCRAMMs involved in the recognition of the same matrix molecule have been studied in some bacterial species, such as, fibronectin binding to Fibronectin Binding Protein A (FnBPA) and Fibronectin Binding Protein B (FnBPB) proteins expressed by *Staphylococcus aureus*, the Gram-positive Firmicute inhabiting the skin and respiratory tract (Westerlund and Korhonen, 1993). The gene encoding the 200 KDa FnBPA was the first gene of a bacterial ECM binding protein to be characterised and fibronectin was the first ECM protein to be experimentally shown as a eukaryotic substrate for adhesion (Flock et al., 1987; Guan, 1990). Sequence analysis revealed similarities in functional domains and fibronectin binding mechanisms adopted by proteins FnBPA and FnBPB of *S. aureus*, SfbI of *S. pyogenes*, FnBA and FnBB of *S. dysgalactiae* and FSE of *S. equisimilis*, respectively (Westerlund and Korhonen, 1993). Around a dozen fibronectin binding proteins have since been identified in the *Staphylococcus* and *Streptococcus* genera (Joh et al., 1999). The C terminal regions of these MSCRAMMs encode tandem repeats of 35-45 highly conserved amino acid residues such as, D1,D2 and D3 in FnBPA. In addition, an incomplete repeat D4 and a less conserved fifth Du unit, located next to the D1 unit are present in FnBPA. Each of these units bind individually to the 29 kDa N-terminal region of fibronectin, which contains a string of five type I F1 modules (1 F1- 5 F1). There are additional binding units that facilitate weak interactions with other parts of fibronectin like the UR sequence located upstream of the tandem repeats in fibronectin binding protein F1/SfbI of *S. pyogenes*. UR sequence binds to the 40kDa collagen binding region in fibronectin, which is adjacent to the 29 kDa N terminal binding region. The collagen binding region and the C terminal heparin binding region in fibronectin are also characterised by F1 module pairs. The unorganised residues of the bacterial MSCRAMMs attain conformational order only on binding to fibronectin. Nuclear magnetic resonance studies have revealed that the

D3 subunit of *S. aureus* FnBPA undergoes extension upon binding to 4F1-5F1 fibronectin modules (Penkett et al., 2000). Gunasekaran et al. (2003) have hypothesised that disordered proteins are able to form larger interfaces with the ligands when compared to stably folded proteins. FnBP-mediated fibronectin binding in Staphylococcal and Streptococcal groups might prove an example for this proposed mechanism.

Bacterial uptake into non-phagocytic mammalian cells is promoted by fibronectin binding which enables *S. aureus* and *S. pyogenes* to invade host tissue, 'hide' from phagocytes and spread infection. Invasion is achieved via a molecular bridge between FnBPs on the bacterial surface and integrins on the host cell (Joh et al., 1999). Internalisation of *S. aureus* into endothelial cells, osteoblasts, keratinocytes, mammary gland cells, corneal epithelial cells and skin fibroblasts has been observed (reviewed in Schwarz-Linek et al. 2004). Fibronectin binding integrins $\alpha_v\beta_3$ and β_1 promoted invasion of cultured human epithelial cells by *S. pyogenes* mediated by F1 fibronectin binding protein (Joh et al., 1999). $\alpha_5\beta_1$ integrin has been implicated in FnBP mediated invasion of epithelial cells in *S. aureus* and *S. pyogenes* (Dziewanowska et al., 1999; Ozeri et al., 1998). The bridging mechanism might be mediated by FnBPA binding to heat shock protein Hsp60 which act as a co-receptor with β_1 integrins linked through fibronectin, thus enhancing bacterial internalisation by epithelial cells (Dziewanowska et al., 2000).

Binding of *S. aureus* to cartilage, rich in type II collagen, allowed the identification of a collagen binding protein which shares functional similarity with the fibronectin binding molecule. The addition of soluble plasma fibronectin promoted binding of *S. suis* to type IV collagen present in basement membranes (Esgleas et al., 2005). In a similar manner, SfbI protein, an allelic variant of F1 protein, of *S. pyogenes* interacts with collagen via surface bound fibronectin, thereby forming bacterial aggregates embedded in a fibronectin-collagen matrix. This interaction blocks Polymorphonuclear leukocyte (PMN) binding to *Streptococci* and protects the bacteria from PMN-mediated ingestion and killing in a serum free environment

(Dinkla et al., 2003). Their role in infection has been shown by mutant strains which are less virulent than the wild-type. However, the Streptococcal Toxic Shock Syndrome (STSS) inducing GrpA *Streptococci* harbour a more virulent M1 protein that blocks phagocytosis (Macheboeuf et al., 2011).

3.1.2.2. MSCRAMMs in other bacteria

Although fibronectin binding has been studied extensively in Staphylococcal and Streptococcal species, it has been found in other microbes like Mycobacterial spp. and *E. coli*. Three distinct classes of fibronectin binding proteins have been identified in Mycobacterial spp. : i) BCG 85, an antigen complex, comprising 3 closely related components of 30-31 kDa molecular mass, secreted by *Mycobacterium bovis* that binds to soluble fibronectin (Westerlund and Korhonen, 1993). ii) a 55-60 kDa fibronectin binding protein (Abou-Zeid et al., 1988). iii) a 120 kDa β_1 integrin on the surface of *M. avium* and *M. intracellulare* that adhere immobilised fibronectin, laminin and type I collagen, and cross-react with polyclonal antibodies against human-laminin binding integrin $\alpha_3\beta_1$ and human-fibronectin binding antibody $\alpha_5\beta_1$ (Rao et al., 1992). Enterobacterial fimbriae has also been implicated in ECM adherence. *E. coli* fimbriae bind fibronectin and also adhere to fibroblasts via fibronectin. Fimbrial adhesion is mediated by protein subunits present at the tip of these structures. P fimbriae binds to immobilised fibronectin and sequence analysis has detected similarity to fibronectin binding repeats of Staphylococcal and Streptococcal MSCRAMMs (Joh et al., 1999). Type I and S fimbriae in *E. coli*, associated with meningitis, bind to oligosaccharides of laminin and promote proteolytic penetration through the basement membrane into the circulatory system (Westerlund and Korhonen, 1993).

3.1.2.3. Fibrinogen Binding

In addition to fibronectin binding, *S. aureus* FnBPA has also been implicated in fibrinogen binding via the N terminal A region of the MSCRAMM (Wann et al., 2000). Fibrinogen is an essential clotting plasma glycoprotein of 340 kDa, secreted primarily by hepatocytes, found at a concentration of $\sim 9 \mu\text{M}$ in blood and consists of

pairs of A α -, B β - and γ -chains in a symmetrical, dimeric arrangement. Platelet adherence at injury sites is mediated by this clotting protein via $\alpha_{11b}\beta_3$ platelet integrin interaction, during which the fibrinogen concentration in blood increases two fold (Herrick et al., 1999). This glycoprotein is incorporated into the extracellular matrix fibrils during active fibronectin polymer elongation. Fibrinogen deposition alters ECM topology and provides a surface for cell migration and matrix remodelling during tissue repair (Pereira et al., 2002). Fibrinogen undergoes α thrombin proteolytic cleavage, with the release of fibrinopeptides A and B from the amino termini of the A α - and B β -chains, respectively to form insoluble fibrin, which is a major component in blood clots. FnBPA binding has been localised to the C terminal γ region in fibrinogen, a domain which is also recognised by ClfA (clumping factor A) protein of *S. aureus* that binds fibrinogen. The surface adhesin FnBPA of *S. aureus* mediates fibrinogen and fibronectin binding which is required for enhancing the infection in an experimental model of endocarditis. Attachment and internalisation into eukaryotic cells is mediated by fibronectin binding, however, fibrinogen binding accelerates these processes to enable valve colonisation and disease induction (Piroth et al., 2008). The N terminal A region of FnBPA shares ~46% amino acid identity to the minimum region for fibrinogen binding in ClfA and thus competes with ClfA for fibrinogen binding. However, the recombinant protein, Clf40, containing the fibrinogen binding region of ClfA was a more potent inhibitor of ClfA-mediated binding to immobilised fibrinogen than its counterpart, rFnBPA, derived from FnBPA. Clf40 also inhibited ClfA-mediated bacterial cell clumping more efficiently than rFnBPA (Figure 3.2) (Wann et al., 2000). ClfA-mediated bacterial adherence to fibrinogen promoted platelet activation which is regarded as an important virulence factor in the aetiology of infective endocarditis. Platelet activation required the binding of IgG molecules to the A domain in ClfA (Loughman et al., 2005).

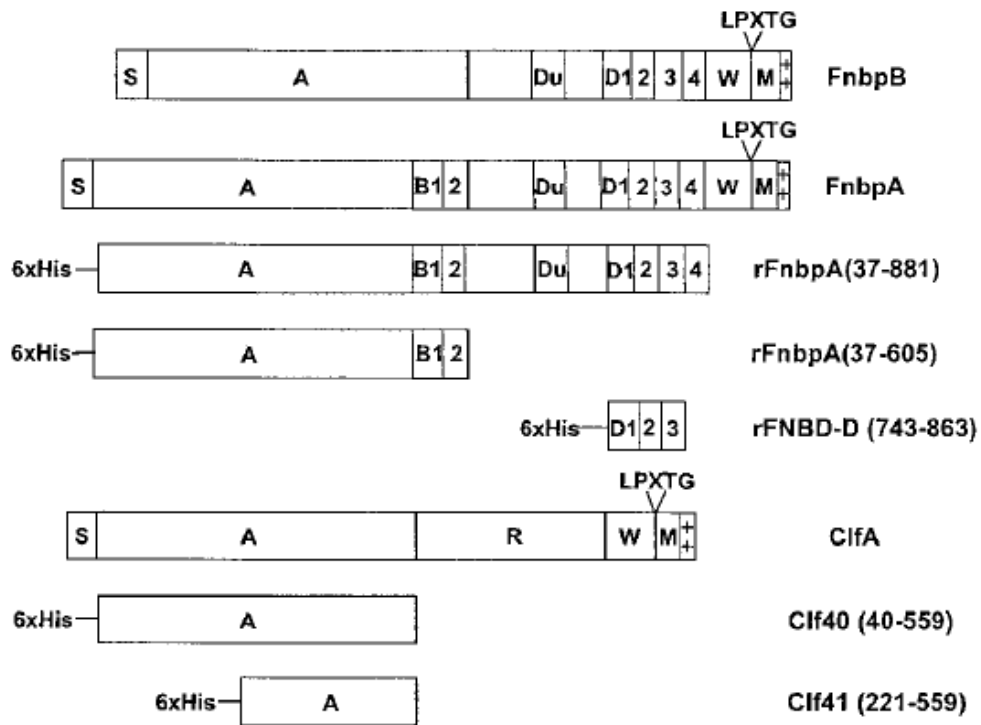


Figure 3.2: Structural organization of FnBps and ClfA of *Staphylococcus aureus*

S, signal sequence; A, a unique N terminal region is known to bind fibrinogen in ClfA; Du-D4, Fn-binding repeat units; R, serine-aspartate repeat region; W, wall-spanning region; M, membrane-spanning region. The LPXTG motifs that anchor the proteins to the cell wall peptidoglycan are indicated. The recombinant proteins (with N terminal His tags) and the amino acid residues (in parantheses) contained in each construct are indicated. (Wann et al., 2000)

Interactions of *Streptococcus* spp. with fibrinogen is of pathogenic interest and has been studied in great detail. The M1 protein expressed by GrpA *Streptococci* plays a role in the resistance of phagocytic killing in blood through the interaction with D fragments of fibrinogen, generated by plasmin degradation of the fibrin matrix. M1 organises the D fragments of four fibrinogen molecules into a cross-like pattern, which is essential for neutrophil activation, that induces sepsis and circulatory failure (Macheboeuf et al., 2011). Fibrinogen binding has also been proposed as a mechanism to evade complement-mediated extracellular host defenses in tularaemia disease caused by the pathogen, *Francisella tularensis*. This bacterium is able to

replicate in host cells, including phagocytes, with the eventual microbial release causing flu-like symptoms. *F. tularensis* binds plasminogen on its surface, which is converted into plasmin and interacts with fibrinogen. The bound fibrinogen appears to recruit host factor H which inhibits formation of C3 convertase enzyme and serves as a co-factor for complement inhibitory host factor I binding which subsequently mediates the cleavage of C3b into inhibitory C3bi and C3d. These subunits prevent Membrane Attack Complex (MAC)-mediated membrane lysis and promote opsonophagocytosis by binding to the CR3 receptor on the surface of phagocytes. Ingestion by phagocytes promotes effective subversion of host defenses by the bacterium and facilitates replication in a non-optimal manner (Jones et al., 2012).

3.1.3. MSCRAMMs expressed in Intestinal Microbiota

The gastrointestinal tract of a healthy mammalian body harbours a wide range of microorganisms, which benefit the host in terms of nutrient utilisation and immunomodulation (Sonnenburg et al. 2004). The estimated density is $\sim 10^{12}$ microbes per ml of luminal contents in the distal gut, belonging to 500-1000 diverse species (Savage, 1977; Xu and Gordon, 2003). The epithelium lining the human intestine is highly dynamic in structure owing to the constant motion of semi-digested food by peristalsis and of excretory materials for cleansing purposes (Joh et al., 1999; Sonnenburg et al., 2004). These processes lead to a large number of epithelial cells being shed every hour, in the colon and small intestine, respectively (Xu and Gordon, 2003). Therefore, the gut colonizers must develop strong interactions with the layers of the intestinal epithelium to ensure their survival in the intestine (Patti et al., 1994). It has been shown that *B. fragilis* can specifically bind to mucin present in the mucosa, a viscous layer separating epithelial cells from the gut lumen and preventing pathogenic infections (Huang et al., 2011). This interaction might be achieved through the expression of phase variable SusC/SusD protein homologues that binds starch present in the mucin glycoprotein (Nakayama-Imahiji et al., 2009). Such physical interactions help the bacteria in long-term gut colonization and enable utilisation of host-derived nutrients for growth (Huang et al.,

2011). MSCRAMMs expressed on the surfaces of commensal members of intestinal microbiota aid in proper colon colonisation as seen in *B. vulgatus* binding to fibronectin. However, MSCRAMMs have also been associated with the spread of opportunistic infections in *Yersinia* spp. and *B. fragilis* via ECM adherence.

3.1.3.1. ECM binding in *Yersinia* spp.

A single MSCRAMM can be implicated in the binding of several ECM ligands as seen in the Gram-negative intestinal pathogenic bacterium *Yersinia enterocolitica*. This bacterium expresses YadA protein that binds to collagen (types I,III,IV and V), laminin and immobilised cellular fibronectin, with the highest affinity for collagen binding, conferring a multi-adhering nature to the protein. YadA is a 200-240 kDa protein expressed by the *yadA* gene located in a 70 kb virulence plasmid. In addition to epithelial cell adhesion, YadA might play a role in attaching the bacterium to the surface of phagocytes for the efficient delivery of anti phagocytic effector Yop proteins into these cells. Human serum resistance which is a common attribute of many pathogens is also mediated by YadA in *Y. enterocolitica*. YadA binds host serum factor H which leads to reduced deposition of C3b and C9, components of the complement system on the bacterium, thereby evading the cell-killing membrane attack complex (El Tahir and Skurnik, 2001).

3.1.3.2. Surface adhesins of *B. fragilis*

The environmental conditions defining the microhabitat in which a bacterium resides has an influence on the bacterium's surface adhesins that mediate binding to ECM components. In vitro experiments, such as agglutination assays, immuno blotting and affinity column chromatography suggest that binding of *B. fragilis* to ECM proteins, laminin and fibronectin, was more intense in cultures grown under oxidising conditions compared to those grown under reducing conditions (Ferreira et al., 2008; Pauer et al., 2009). There was also an increase in general protein expression when OMP electrophoretic profiles from the two conditions were compared. These observed differences in laminin and fibronectin binding, coupled with the fact that *B. fragilis* is the most commonly isolated intestinal bacterium from intra-abdominal

infections and soft tissue abscesses, provide valuable insight into the conditions that facilitate the shift of *B. fragilis* from an intestinal commensal to an opportunistic pathogen in vivo. Lesions, injury and other factors that disrupt the tissue integrity lead to an oxidising environment near the intestinal epithelium. In this oxidised condition, an increase in expression of ECM binding proteins by *B. fragilis* might promote the dissemination of infection (Ferreira et al., 2008; Pauer et al., 2009). In 1993, Goldner et al. demonstrated the ability of *B. fragilis* to penetrate into HeLa cells at high redox levels, whereas lower redox levels did not support tissue cell invasion. Therefore outer membrane surface adhesins pose as essential virulence factors in *B. fragilis*. Moreover, proteases, capsular polysaccharide complex, fimbriae, enterotoxin, neuraminidases and hemolysins play additional roles in pathogenesis (Guzmán et al., 1990; Hofstad, 1992; Myers et al., 1984; Robertson et al., 2006).

The majority of *B. fragilis* strains screened were able to bind laminin I, independent of integrin (De et al., 2006). Affinity column chromatography conducted on cultures grown in both reduced and oxidised media identified a laminin I binding outer membrane protein of 60 kDa, most probably a member of the glycoprotein family (Ferreira et al., 2008). Sequence analysis identified 98% identity to a putative plasminogen binding protein. It was shown experimentally that the laminin binding protein was also able to bind to the circulating precursor, plasminogen, and convert it to active serine protease, plasmin, via the tissue type plasminogen activator. The inherent function of this protease is to solubilise fibrin present in blood clots but has also been implicated in ECM protein degradation in many pathogens expressing plasminogen receptors (Ferreira et al., 2009). Owing to this host-dependent proteolytic ability inherent to *B. fragilis*, the bacterium is able to induce tissue injury and also activate other matrix degrading proteolytic enzymes. As laminin is found in close association with collagen in the basement membrane, the plasmin degrades fibrin, collagen and other tissue components resulting in the destruction of ECM basement membrane barrier (Ferreira et al., 2013). Collagen binding was observed in 87% of *B. fragilis* isolates studied. In strain GSH18, Collagen type I binding was

associated with a 34 kDa glycosylated Collagen Binding Protein (CbpI), but glycosylation was not a prerequisite for binding. Adhesion to collagen type I was detected by the protein encoded by BF0586 in *B. fragilis* NCTC 9343. Analogue of BF0586 is also present in YCH46 and 638R *B. fragilis*. The study also indicated that more than one surface component contributes to collagen binding since no marked difference was observed between the collagen binding ability of whole cells of wild-type and *cbpI* mutant strain (Galvão et al., 2014).

Since the early 1990s, non-covalent fibronectin binding via the N terminal domain of fibronectin has been identified in clinical samples of *B. fragilis* isolated from healthy individuals and from wounds. Fibronectin binding is mostly accompanied by binding to vitronectin, another ECM glycoprotein in this bacterium (Nagy et al., 1994). Among the strains studied, *B. fragilis* 1405 which was isolated from a bacteraemia case exhibited the greatest difference when plasmatic fibronectin adherence levels were compared between strains grown under oxidising and reducing conditions. Electrophoretic profiles of protein expression were higher under oxidising conditions and increased binding to fibronectin was detected under these conditions. Western blot analysis demonstrated a 102 kDa fibronectin binding protein encoded by the *bf1991* gene which was identified as a TonB-dependent outer membrane protein and was detected in all the strains studied. In *E. coli*, TonB functions as an energy transducer that couples proton motive force for active transport of iron siderophores across the outer membrane. The bulk of the TonB protein extends into the periplasm. Therefore, in *B. fragilis* this protein might play a role in the transport of tightly bound ligands into the periplasmic space, which might suggest host interaction (Pauer et al., 2009). However, the TonB-dependent protein was not distributed uniformly on the cell surface and a mutant harbouring an inactivated *bf1991* exhibited increased adhesion to fibronectin. These findings suggest that there are other fibronectin binding proteins encoded by the *B. fragilis* genome. Nonetheless, the TonB-dependent protein plays a major role in pathogenesis as the mutant strain was phagocytosed more readily by peritoneal macrophages than the wild-type (Pauer et al., 2013).

Once the epithelium of the intestine is breached and *B. fragilis* escapes into the peritoneal cavity, the bacterium is confronted by the peritoneal macrophages, which are the first line of host immunological defence mechanisms. Macrophages present antigens and secrete cytokines such as TNF- α and IL-1. Microbicidal function is promoted by activated macrophages through the production of cytotoxic radicals, one of them being IFN- γ -mediated Inducible Nitric Oxide Synthase (iNOS) expression, which is co-localised with the microfilament network of macrophages. Interaction with *B. fragilis* results in macrophage membrane rupture which render the cytoplasm of macrophages disorganised and granulated owing to cytoplasmic extrusions through pore-like structures in the membrane (Vieira et al., 2005). These events culminate in a necrosis-like condition in the macrophages (Vieira et al., 2009). Significant decrease in nitric oxide production was observed as the actin filament guided iNOS outside the cell via the pores formed, thus enabling *B. fragilis* cells to evade macrophage killing and persist in the peritoneal cavity.

Capsule PS A of *B. fragilis* promotes formation of intra abdominal abscess which helps in the containment of polymicrobial infections, such as peritonitis, caused by *B. fragilis* along with other opportunistic pathogens residing in the intestinal cavity. Abscess formation is a host-mediated mechanism to isolate bacteria, leucocytes, tissue debris and inflammatory exudates in fibrin-mediated abscess walls to prevent the dissemination of infection (Tally and Ho, 1987). To invade neighbouring tissues and spread blood related infections, it is essential to escape the fibrin mediated abscess walls. Also, fibrinogen, the soluble form of fibrin, is incorporated into the ECM by active deposition of fibronectin and many pathogens are capable of fibrinogen interactions for better chances of survival in the host. Studies by Houston et al. (2010) have suggested that *B. fragilis* is able to bind and degrade fibrinogen. This interaction can be regarded as a mechanism by which the bacterium inhibits the formation, or rupture, of abscesses and enhances the spread of infection. Fibrinogen adhesion is brought about by *Bacteroides fragilis* Fibrinogen Binding Protein (BF-FBP), a putative surface protein, encoded by BF1705 in NCTC 9343 (Houston et al., 2010) (Figure 3.3). BF-FBP shares homology with a 98 kDa BspA surface protein of

Tannerella forsythia, a periodontitis causing oral pathogen in humans, which binds to fibrinogen and fibronectin (Houston et al., 2010). It is an important virulence factor which helps during epithelial cell invasion and specific antibodies against BspA have been detected in adult patients suffering from periodontitis (Sharma et al., 1998).

Following escape from abscesses, *B. fragilis* needs to combat the MAC-mediated bactericidal activity of the serum and neutrophil-mediated opsonophagocytosis and cause bacteraemia. Rotimi & Eke (1984) observed that some *B. fragilis* isolates possess an increased ability to inhibit killing by normal human serum and serum from infected patients when compared to other members of the *Bacteroides* genus. Resistance to phagocytic uptake and killing by neutrophils in bacterial concentrations comparable to those in naturally occurring in vivo in abscesses, was also observed in *B. fragilis* studies conducted by Ingham et al., (1981). In polymicrobial abscesses, the obligate anaerobe competed with facultative anaerobes like *E. coli* for opsonins in a synergistic manner, thereby impairing opsonophagocytosis and intracellular killing. Succinate, the end product of *B. fragilis* fermentative metabolism, has been implicated in preventing neutrophil migration and killing activity in conditions mimicking the extracellular environment (Rotstein et al., 1987). The antigenically distinct and phase variable capsular polysaccharide surface of *B. fragilis* has been studied extensively with regard to its role in evasion of neutrophil-mediated as well as serum killing (Kasper, 1976; Onderdonk et al., 1977; Reid and Patrick, 1984; Simon et al., 1982). Examination of fibrinogen binding to the surface of *B. fragilis* for a potential role in preventing complement mediated bactericidal activity by a similar mechanism as proposed for *Francisella tularensis* would open new research frontiers.

3.1.4. Aims

- To investigate the role of BF1705 protein, present in the *B. fragilis* NCTC 9343 genome, in fibrinogen binding by expression and reverse genetics
- To examine the possible role of capsular polysaccharides in fibrinogen binding by markerless gene deletion and Immunofluorescence Microscopy studies
- To determine binding of fibrinogen by whole protein extracts and protein extracts from the outer membrane of *B. fragilis* wild-type and deletion mutants in immunoblots
- To test a possible involvement of the Δ BF1705 in degradation of fibrinogen
- To analyse the binding potential of BF1705 protein to the components of the extracellular matrix

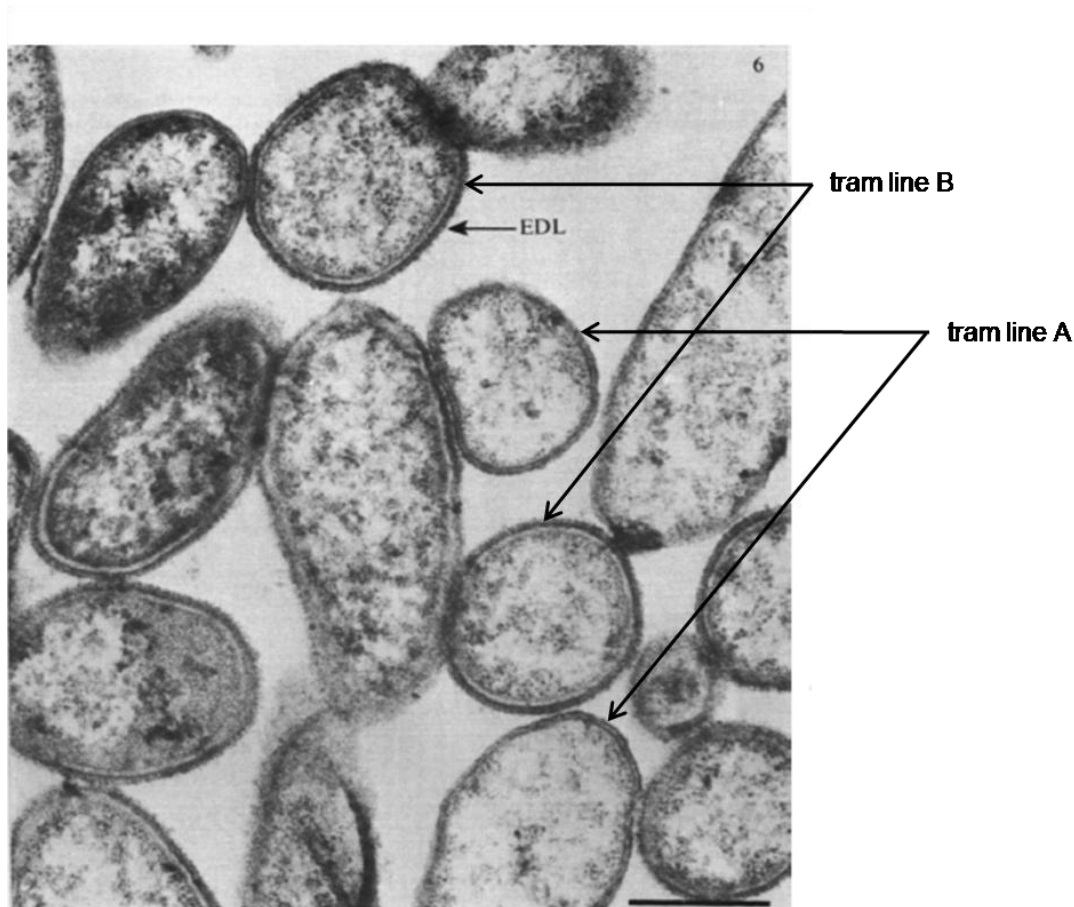


Figure 3.3: Electron micrograph of *B. fragilis* NCTC 9343

The location of BF1705, a putative surface protein associated with fibrinogen binding, is indicated (tram line A). The electron dense layer (EDL and tram line B) is observed on the surface of some cells (1.2.1). BF1708 protein contributes to the assembly of high molecular mass polysaccharides into a long-chain-length polymer, thereby involving in the production of EDL and is discussed later in this chapter (3.7). BF1705 and BF1708 proteins are likely to be associated with the outer membrane based on bioinformatics analysis. (Patrick et al., 1986)

3.2. Choice of bacterial strains

All experiments were conducted using the non enterotoxin-producing *B. fragilis* (NTBF) national collection of type cultures (NCTC) 9343 strain, which was originally isolated from a human peritoneal abscess at St Bartholomew's Hospital, London, UK in 1955 (Culture collection, Public Health England). The complete genome sequence of NCTC 9343 is available, which renders molecular experiments

involving genetic manipulation of the bacterium amenable (Cerdeño-Tárraga et al., 2005). NCTC 9343 comprises a single circular chromosome of 5,205,140 bp, predicted to encode 4274 genes, and a plasmid pBF9343. Shotgun assembly of the genome revealed that specific inversions of sequences occurred at high frequency within the DNA extracted from bacterial population used for sequencing (Cerdeño-Tárraga et al., 2005). These invertible regions can be grouped into two depending on the inverted repeat sequences that flank them. Identified invertible regions contain a consensus promoter, which suggests they control the expression of downstream genes. Seven of the group I invertible promoters have been associated with 7 of the 10 polysaccharide biosynthesis gene clusters. Thus, expression of within-strain phase and antigenic variation of capsular polysaccharide surface components, which is a characteristic feature of *B. fragilis*, is controlled by upstream invertible regions. Apart from NCTC 9343, diversity in polysaccharide loci has also been identified in the published complete genome sequences of two other non toxigenic strains, YCH46, a blood culture isolate from Japan and 638R, a plasmid-free spontaneous rifampicin-resistant clinical isolate from Chicago, USA (Kuwahara et al., 2004; Patrick et al., 2010). This demonstrates the diversity of the *B. fragilis* pan-genome. In addition to promoter sequences, *B. fragilis* harbours intergenic shufflons that are involved in the inversion of complete and partial coding sequences. Therefore, DNA inversion events mediated by *B. fragilis* play an important role in controlling the expression of surface and secreted components, regulatory molecules and restriction-modification proteins in the bacterium. Recently, Nikitina et al. (2015) have published the complete genome sequence of an enterotoxigenic *B. fragilis* (ETBF) clinical isolate, BOB25, from a stool sample of a dysbiosis patient. A pathogenicity island identified within a conjugative transposon is a unique feature of this ETBF strain which was not present in previously published sequences of NTBF strains and appears to contribute to the pathogenicity of ETBF strain.

3.3. Generation of Markerless deletions in *Bacteroides fragilis* NCTC 9343 genome

Reverse genetics was employed to analyse the interaction of specific proteins encoded by the bacterium with components of the host. Construction of gene deletion mutants provides an effective way to characterise gene functionally. However, restriction/modification (R/M) systems that recognise and cleave foreign invading bacteriophage DNA can also destroy DNA introduced by techniques such as electroporation, thereby posing a barrier to genetic manipulation of many bacteria, including *B. fragilis*. One of the intergenic shufflons present in *B. fragilis* NCTC 9343 generates eight different HsdS proteins that provide DNA target specificity for a type I R/M system. Two further type I and two type III R/M systems are also present in NCTC 9343 genome. The ability to transform 638R explains why this strain has mostly been involved in the production of mutations in *B. fragilis* in the past. The *B. fragilis* 638R genome specifies two type I, one type IIS and one type III R/M systems. In addition to the classical R/M systems, NCTC 9343 encodes proteins sharing homology with the type IV restriction endonuclease McrBC, which cleaves specific DNA sequences modified with 5-methylcytosine (Patrick et al., 2010). The absence of these genes in 638R might aid in the effective transformation of 638R strain using Dcm methyltransferase-modified DNA from *E. coli* K12. To make deletions, Coyne et al. (2003); Coyne et al. (2008) developed a method that depends on integration of an altered sequence into the NCTC 9343 chromosome followed by screening for spontaneous resolution of the diploid to yield the desired product. This technique is not only time consuming, but also fails to generate the required mutant if the diploid is stable. Alternatively, trimethoprim resistance has been utilised for enriching resolved diploids by introducing a chromosomal *thyA* mutation into *B. fragilis* which is complemented with a functional copy of *thyA* in a suicide plasmid (Baughn and Malamy, 2002). This method leads to a *thyA* mutation in all resulting strains due to the mutation in the primary engineered *B. fragilis* strain, which might cause complications in future studies, similar to the SOS response induced in *E. coli*

thyA mutants when not supplied with exogenous thymine (Begg and Donachie, 1978).

The present study adopts a modification of a reliable method developed by Patrick et al. (2009) for generating deletion mutants in *B. fragilis* NCTC 9343. The method was originally developed for mutational analysis of genes involved in *B. fragilis* LPS and capsular polysaccharide biosynthesis. Furthermore, these non-polar gene deletions can be complemented, which is important for confirming phenotypes in *B. fragilis* genetic studies. In the polysaccharide studies conducted by Patrick et al. (2009), the mutants generated were marked with an erythromycin resistance determinant, whereas the present work relies on generation of markerless gene deletions. To generate deletions, approximately 500 bp of nucleotides, flanking (upstream and downstream regions) the sequence to be deleted, were amplified by polymerase chain reactions. Markerless deletion constructs were prepared by fusing the amplified flanks using cross-over PCR. The amplicon was cloned into the multiple cloning site of an RP4-based conjugative vector, pGB910, which contains an ISce-I meganuclease recognition sequence. Recombinant pGB910 vector carrying the markerless deletion constructs were then recovered in *E. coli* S17-1 λ pir. *E. coli* transformants positive for the recombinant vector were selected for chloramphenicol resistance and the presence of deletion constructs were confirmed by PCR. The deletion constructs were mobilized from *E. coli* S17-1 λ pir to *B. fragilis* NCTC 9343 (Figure 3.4 A). Transconjugants were grown in the presence of erythromycin to select for the integrated recombinant pGB910 plasmid. Diploids were resolved by transforming strains, carrying the plasmid integrant, with pGB920 that expresses the I-SceI meganuclease under the control of the *B. fragilis fucR* promoter, with selection for tetracycline resistance (Figure 3.4 B). Selected transformants were grown on defined medium containing fucose, which induces an ISce-I mediated double-strand break (Figure 3.4 C), resulting in the resolution of diploid by homologous recombination into either wild-type (Figure 3.4 E) or deletion genotype (Figure 3.4 D). Colonies were then screened for loss of resistance to erythromycin and the presence of a deletion was confirmed by PCR. Natural resistance of *B.*

fragilis to many antibiotics restricts the number of selection markers that can be employed for genetic studies in this bacterium. An advantage of markerless gene deletions over marked deletions is that they can be complemented using plasmids bearing the same antibiotic resistance determinant as the one used for integrant selection.

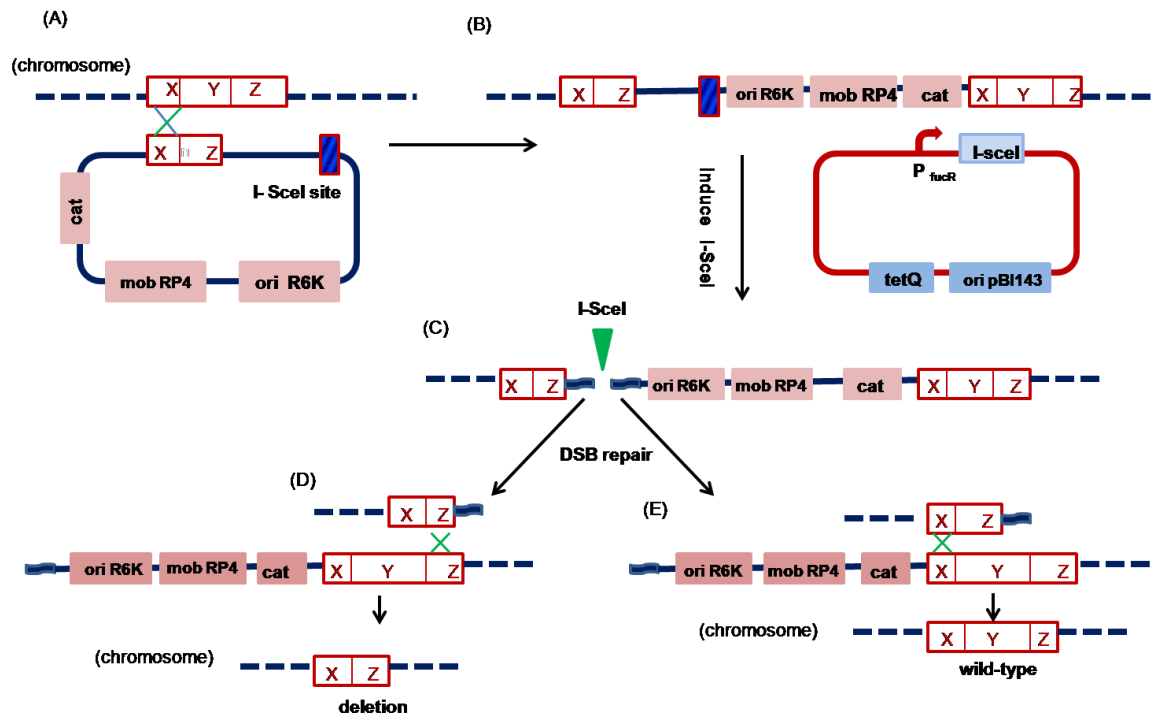


Figure 3.4: Schematic of the generation of markerless chromosomal deletions in *B. fragilis*

(A) A suicide plasmid, containing an I-SceI recognition site (blue box) and sequences homologous to chromosomal DNA (X and Z), replacing the gene to be deleted (Y), are introduced into *B. fragilis* by conjugation. Plasmid integrants are selected for resistance to erythromycin. (B) Diploids are transformed with a plasmid expressing I-SceI under the control of the fucose-inducible promoter *P_{fucR}*. (C) Growth on defined medium in the presence of fucose leads to an I-SceI-mediated double strand break (▼) and resolution of the diploid to generate either the deletion (D) or wild-type (E) genotypes by homologous recombination. Chromosomal DNA is indicated as dashed blue lines and plasmid DNA is indicated as solid blue lines.

3.4. Fibrinogen binding studies previously conducted in *B. fragilis*

The present study aims to apply reverse genetics by generation of markerless BF1705 gene deletion mutants to confirm the fibrinogen binding function of BF1705 protein and also attempts to identify other cell surface components involved in host fibrinogen associations. Interaction of *B. fragilis* cells with human fibrinogen, a major clotting protein, via one or more cell surface components, was initially identified by Houston et al. (2010) using IFM studies. Subsequent sequence analysis revealed an ~ 57 kDa cell surface protein encoded by BF1705 as a putative fibrinogen binding protein (Table 3.1). The BF1705 protein contains a signal sequence and a membrane lipoprotein lipid attachment site. Leucine-rich repeats (LRRs) present in the protein sequence suggested probable protein-protein interactions. NCBI protein BLAST analysis predicted an E value of 1.1 e-17 with 32.8% sequence identity in 295 aa to the LRR-containing surface antigen BspA of *Tannerella forsythia*, a periodontal disease causing bacterium, which binds human fibrinogen. Binding of purified BF1705 protein to fibrinogen was observed in a dose-dependent manner in dot blots, while far-western analysis identified strong interaction of BF1705 protein with B β -chain and weakly with A α - and γ -chains of fibrinogen (Houston et al., 2010). However, the possible involvement of other cell surface components in fibrinogen binding detected by IFM studies, was not investigated in the previous study. Also, the analysis of BF1705 regulation under nitrogen limiting conditions would reveal a potential role of the protein encoded by BF1705 in nutrient binding, thereby facilitating nitrogen replenishment through proteolysis.

Table 3.1: Description of BF1705 using artemis genome browser (Carver et al., 2008)

The protein sequence (520 aa) shared 32.8% identity in 295 aa to BspA, an outer membrane surface protein of *Tannerella forsythia* with an E value of 1.1 e-17.

Feature Name	CDS	Size	Hits
BF1705 <i>B. fragilis</i> cell surface protein	complement 1984894...1986453	1560 bp 520 aa 56.987 kDa	<i>Tannerella forsythia</i> Outer membrane surface protein BspA (1081 aa)

3.5. Generation of a markerless deletion of BF1705 encoding a putative fibrinogen-binding protein in *B. fragilis*

3.5.1. Primer design for amplifying the upstream and downstream regions of BF1705 gene

A primer set, comprising four individual primers (LOF, LIR, RIF and ROR), was designed to amplify the region flanking BF1705 using the Artemis genome browser (2.3). Artemis provides a graphic interpretation of the gene arrangement of the NCTC 9343 genome along with nucleotide and amino-acid sequences, which facilitated primer design (Carver et al., 2008). The left outside forward (LOF) and left inside reverse (LIR) primers were designed to amplify ~500 bp upstream of the BF1705 gene whereas the right inside forward (RIF) primer and right outside reverse (ROR) primers were generated to amplify the region flanking the sequence downstream (Figure 3.5). Primer sequences were in the range of 20-24 bases in length and care was taken to avoid poly T sequences and GC rich areas. The melting temperature (T_m) values of the primer sequences were restricted to 60-66°C. Primers designed for the outside region of flanks (left outside and right outside primers) were attached with ~15 bp overlap from the vector sequence in order to ensure compatibility while cloning. In a similar manner, the left inside and right inside primers were prefixed with reverse complements of primer sequences from each other to facilitate fusion PCR.

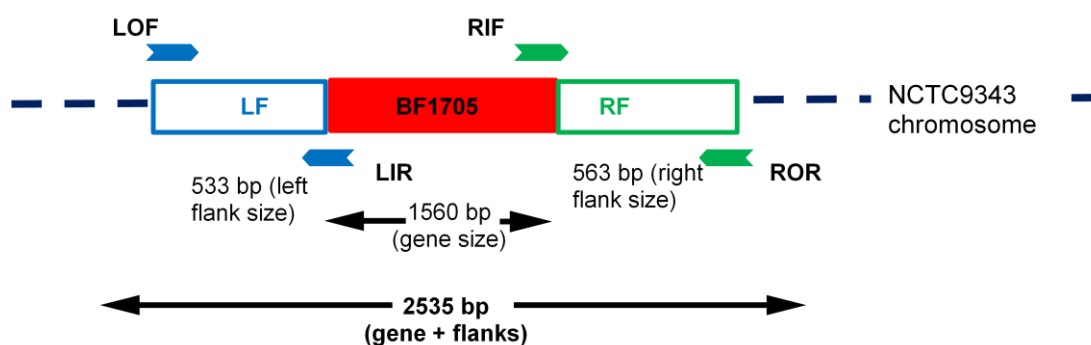


Figure 3.5: Graphic representation of BF1705 and the flanking regions

BF1705 (red box) with left outside (LOF) and left inside (LIR) primers designed for amplifying the upstream region (denoted in blue) and right inside (RIF) and right outside (ROR) primers designed for amplifying the downstream region (denoted in green).

3.5.2. Generation of the BF1705 deletion construct

To generate the deletion construct, the left upstream and right downstream regions of BF1705 gene were amplified following appropriate PCR conditions (2.5.3) using specific primers (2.3) and NCTC 9343 genomic DNA template (2.5.1). The amplified left and right flanks of BF1705 were visualised after electrophoresis through 1% agarose gels and were detected at ~500 bp and ~550 bp in size (Figure 3.6 A lane 1 and Figure 3.6 B lane 1). The observed amplicons were in agreement with expected sizes of 533 bp and 563 bp, respectively. The amplified flanks were then subjected to fusion PCR using left outside forward primer and right outside reverse primer to fuse together the left and right flanking regions and generate a 1096 bp deletion construct. Agarose gel analysis of the fusion PCR product confirmed a band at ~1000 bp (Figure 3.6 C lane 1).

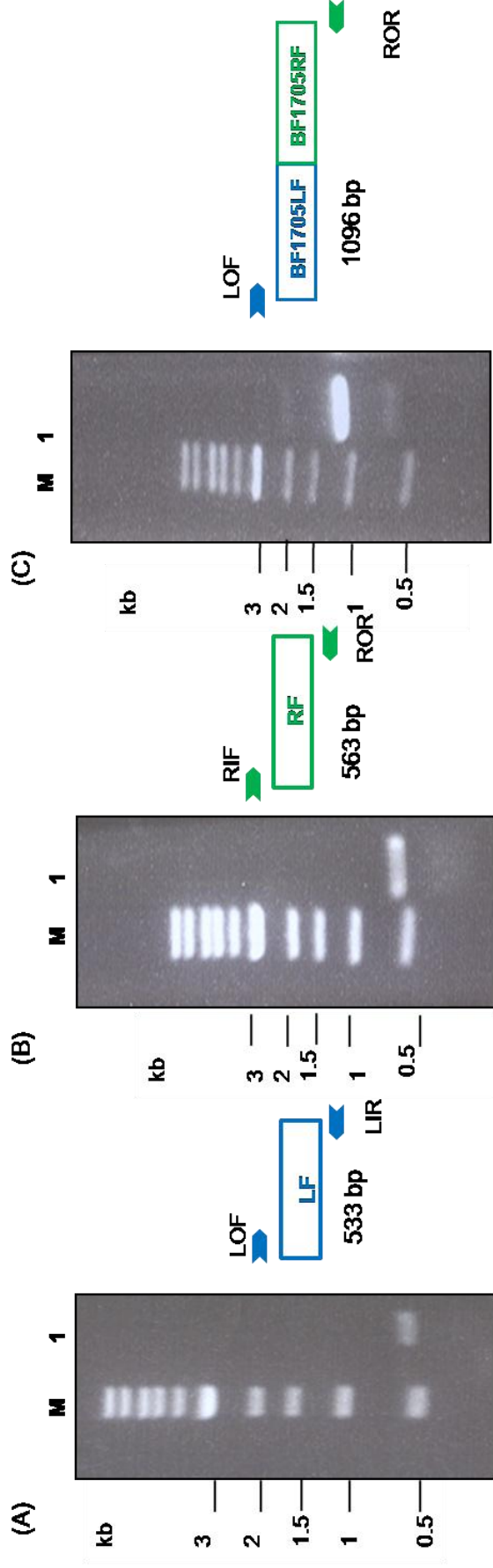


Figure 3.6: Amplification of flanking regions and the deletion construct of BF1705

Agarose gel analysis of (A) upstream region of BF1705 (left flank) (lane 1) using LOF and LIR primers (B) downstream region of BF1705 (right flank) (lane 1) using RIF and ROR primers and (C) deletion construct of BF1705 by fusing the left flank and right flank together (lane 1) using LOF and ROR primers. Arrows and numbers on the left side indicate size of DNA in kilo bases (kb) in a 1kb DNA ladder (M). The BF1705 left and right flanks (A and B lane 1) were of ~500 bp in size while the fused deletion construct (C lane 1) was of ~1000 bp.

3.5.3. Cloning the BF1705 deletion construct into pGB910

The deletion amplicon was cloned into a NotI restriction site in the multiple cloning site (MCS) region of linearised pGB910 by Infusion cloning, a ligation independent cloning method (Clontech; 2.5.8). The Infusion enzymes fuse DNA fragments or PCR generated sequences with linearised vectors, efficiently and precisely by recognising a 15 bp overlap at their ends. The infusion cloning reaction was performed by incubating the BF1705 deletion construct (flanked by a 15 bp vector sequence), linearised vector and the infusion enzymes (2.5.8). Plasmid pGB910 is a 5305 bp RP4-based conjugative vector (Patrick et al., 2009) (Figure 3.7) generated by inserting an *ermF* cassette in the KpnI site of pGB909, a pEP185.2 derivative containing an I-SceI meganuclease recognition sequence inserted at the SacI site. pGB910 also harbours *cat*, a chloramphenicol resistance marker inserted at the EcoRI site. Resistance to chloramphenicol is required for selection of transformants in *E. coli*. *ermF* allows selection of *B. fragilis* transconjugants carrying the plasmid integrant and also allows screening of recombinants that lost the plasmid derived *ermF* post resolution. pGB910 plasmid is maintained in an *E.coli* S17-1 λ pir strain. In the present study, the vector was linearised by digesting with NotI restriction enzyme (Figure 3.8 lane 2).

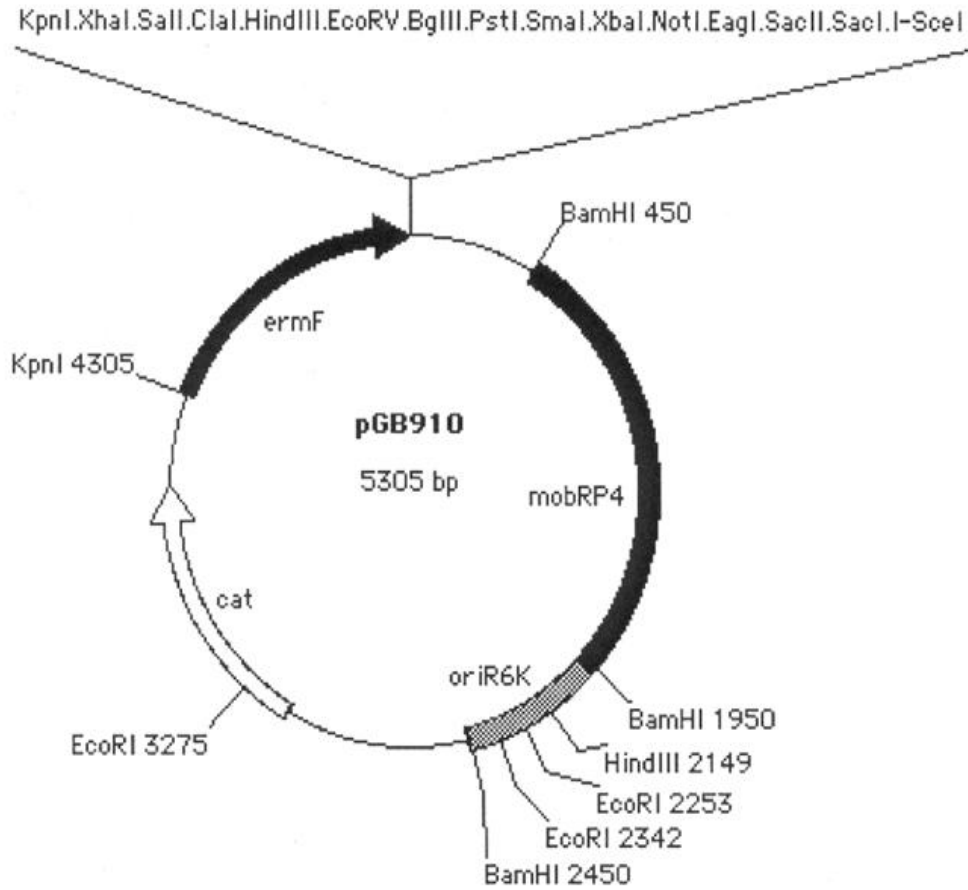


Figure 3.7: Vector map of pGB910

pGB910 was constructed by Blakely lab for bacterial conjugation experiments. Size of the vector is 5305 bp. *ermF* is inserted in the KpnI site. The vector contains multiple BamHI, EcoRI and HindIII sites and is mobilised by RP4 integrated in the chromosome.

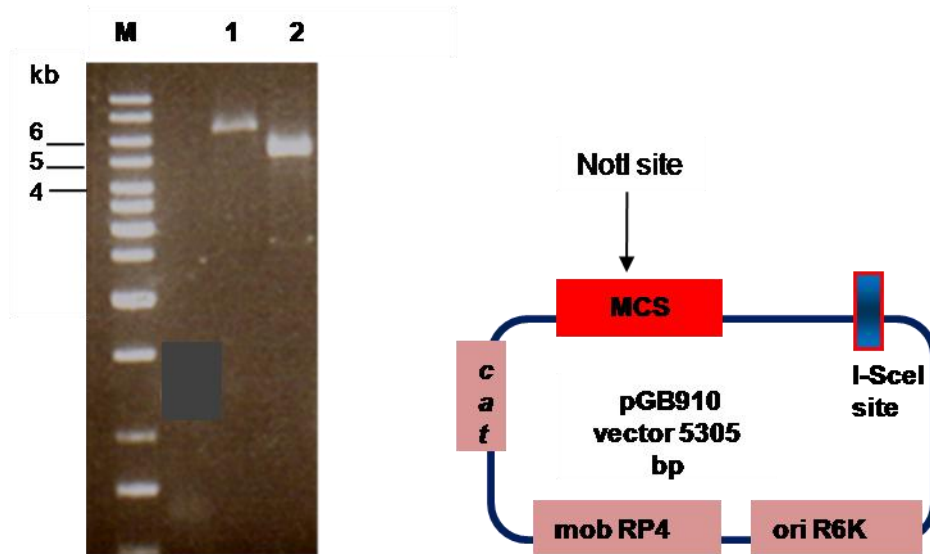


Figure 3.8: Restriction profiling of pGB910 for cloning experiments

The supercoiled pGB910 (lane 1) was linearised using NotI restriction enzyme. Arrows and numbers on the left side indicate size of DNA in kilo bases (kb) in a 1kb DNA ladder (M). The linearised pGB 910 (lane 2) was of ~ 5000 bp in size which is in agreement with the expected size of pGB910 mentioned in the graphic representation.

The infusion cloning mixture was used to transform *E.coli* S17-1 λ pir (2.4.3) with selection for chloramphenicol resistance. The *E. coli* S17-1 strain contains the *pir* gene, a replication initiator for plasmid R6K (Rakowski and Filutowicz, 2013). The S17-1 strain has chromosomally integrated RP4-based conjugal transfer functions. Chloramphenicol resistant colonies carrying the recombinant pGB910 vector were confirmed by PCR amplification of the deletion construct. Amplified bands corresponding to ~1000 bp on an agarose gel (Figure 3.9 lane 1 and 2) confirmed the presence of the BF1705 deletion construct in the recombinant vector. Sanger-based sequencing of the recombinant vector was also performed to check for the presence of the construct (2.5.10).

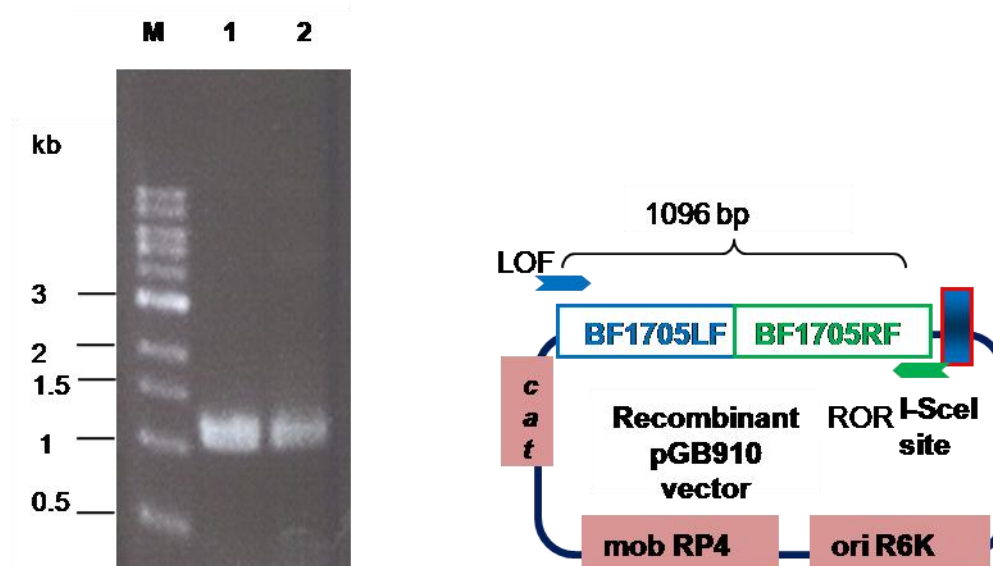


Figure 3.9: Confirmation of the BF1705 deletion construct in the recombinant vector

Amplification was performed using LOF and ROR primers. Arrows and numbers on the left side indicate size of DNA in kb in a 1kb DNA marker (M). The BF1705 deletion construct was of ~1000 bp in size (lane 1 and 2) and is in agreement with the expected size of 1096 bp indicated in the graphic representation.

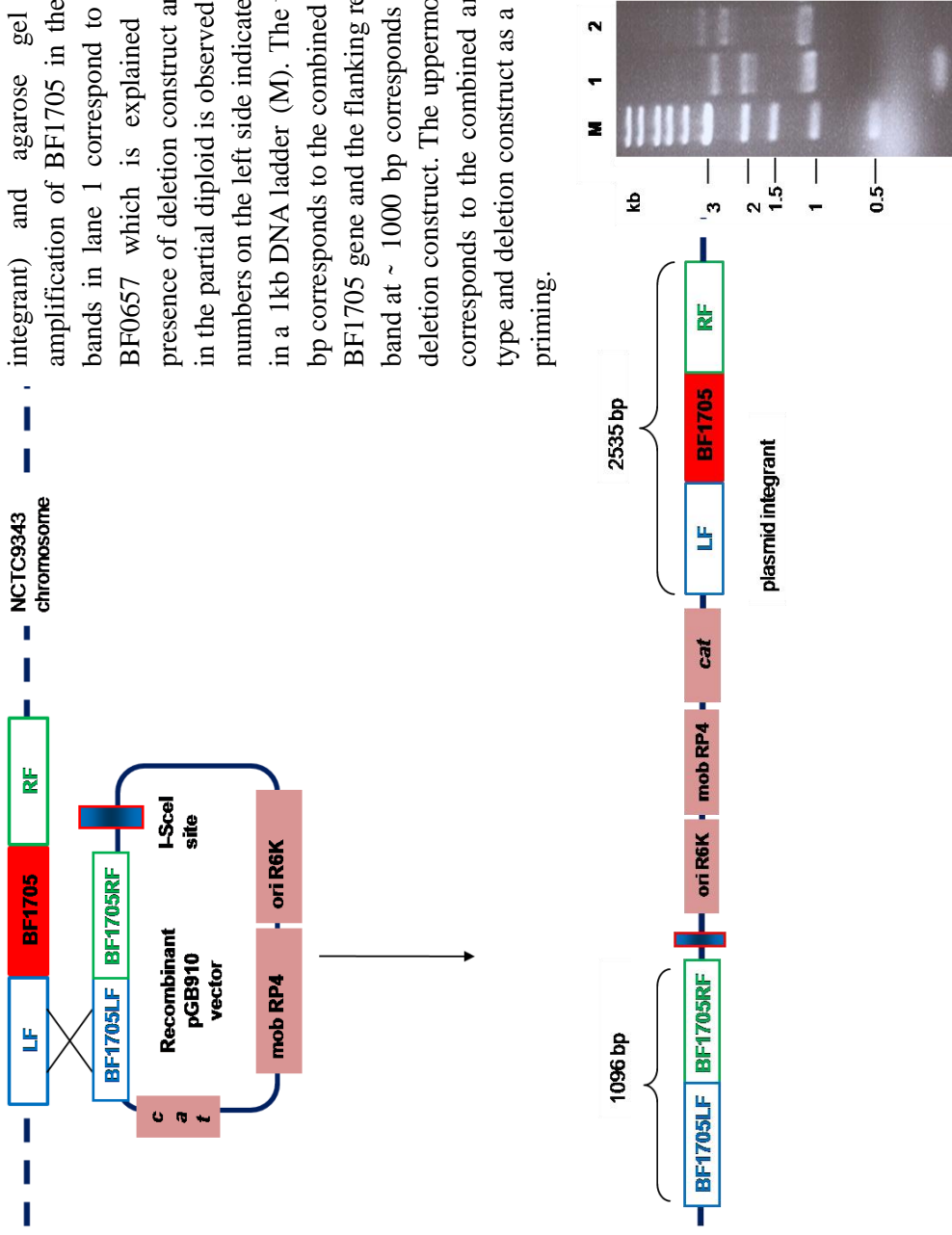
3.5.4. Conjugal transfer of the Δ BF1705 construct into *B. fragilis* NCTC 9343

The recombinant vector carrying the Δ BF1705 construct was mobilized from *E. coli* S17-1 λ pir into *B. fragilis* NCTC 9343 by conjugation (2.4.5). The transfer was originally based on a filter mating method developed by Valentine et al. (1988) for mobilisation of *Bacteroides* plasmids by *Bacteroides* conjugal elements. Post conjugation, the strains were plated onto BHI-S agar containing gentamicin sulphate, to counter-select *E. coli*, and erythromycin, to select *B. fragilis* transconjugants carrying the recombinant pGB910 vector. The plated cells were grown anaerobically for 2 days prior to confirming the presence of integrated plasmids in the erythromycin resistant colonies by PCR using left and right outside primers. Following conjugation, the *B. fragilis* strains carrying the integrated plasmid is in a

partial diploid state and therefore PCR amplification with left outside forward and right outside reverse primers resulted in amplicons corresponding to the size of the deletion construct (1096 bp) as well as the combined size of the wild-type gene and the flanking region (2535 bp) as observed on 1% agarose gels (Figure 3.10 lane 2). The partial diploids carrying the integrated plasmids were thus generated successfully.

Figure 3.10: Amplification of BF1705 and the flanking regions in the partial diploid

Graphic representation of the mobilization of the plasmid carrying the BF1705 deletion construct into *B. fragilis* to form the partial diploid (plasmid integrant) and agarose gel analysis of PCR amplification of BF1705 in the partial diploid. The bands in lane 1 correspond to the amplification of BF0657 which is explained in Chapter 4. The presence of deletion construct and wild-type BF1705 in the partial diploid is observed (lane 2). Arrows and numbers on the left side indicate size of DNA in kb in a 1kb DNA ladder (M). The upper band at ~ 2500 bp corresponds to the combined size of wild -type BF1705 gene and the flanking regions and the lower band at ~ 1000 bp corresponds to the BF1705 gene deletion construct. The uppermost band at ~ 3500 bp corresponds to the combined amplification of wild-type and deletion construct as a result of non specific priming.



3.5.5. Resolution of partial diploid *B. fragilis* strains carrying the deletion construct

The partial diploid strains were transformed by electroporation with pGB920 (2.4.4). The I-SceI expressing pGB920 plasmid was generated from pLYL01 (Figure 3.11) and consists of an I-SceI coding region under the control of the *B. fragilis fucR* promoter with selection for resistance to tetracycline. Tetracycline resistant colonies were selected and streaked on defined medium containing 2% fucose, which triggered an I-SceI mediated double-strand break. Thus, diploids were resolved by homologous recombination into either wild-type or deletion genotypes. Colonies grown on fucose were patched onto plain BHI-S medium and BHI-S medium containing erythromycin to screen for loss of erythromycin resistance. Chromosomal DNA from the erythromycin sensitive colonies (2.5.1) were subjected to PCR amplification using left outside and right outside primers to confirm the presence of deletion constructs at ~1000 bp (Figure 3.12 lanes 1 and 4). However, the diploids that resolved into wild-type resulted in bands corresponding to the size of the wild-type gene along with flanking regions of ~2500 bp (Figure 3.12 lanes 5 and 7). The expected size of the wild-type gene + flanks is 2534 bp which is in agreement with the observed bands.

I-SceI mediated break enhances the resolution of diploids. Therefore most transformants were resolved into either mutant or wild-type thus eliminating the need to repeatedly subculture diploid strains and screen large number of colonies for resolution by random recombination events.

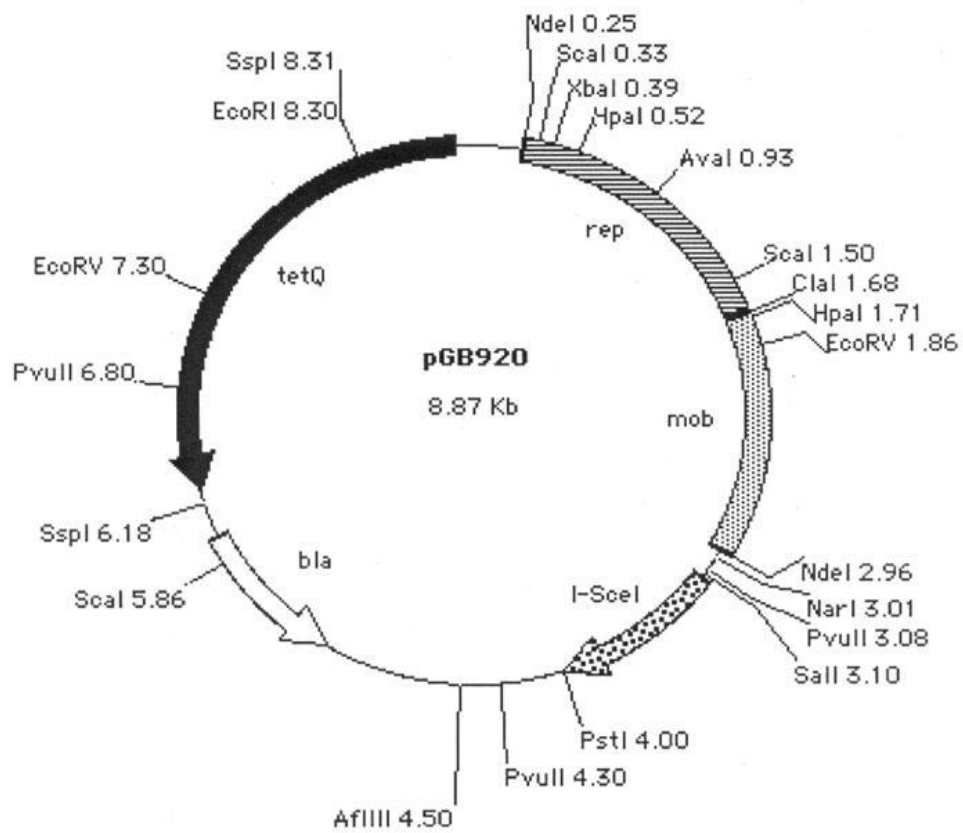
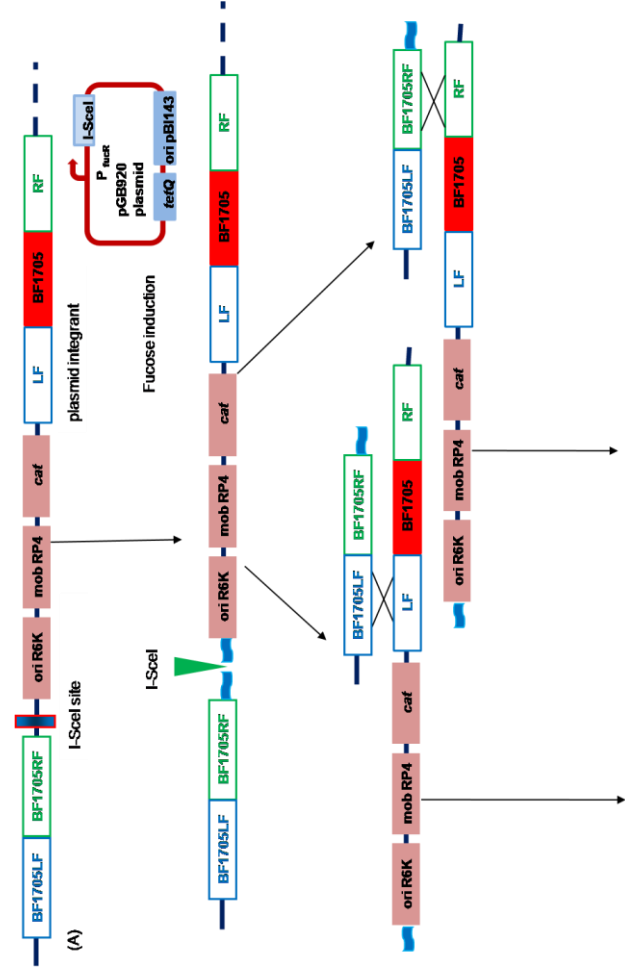


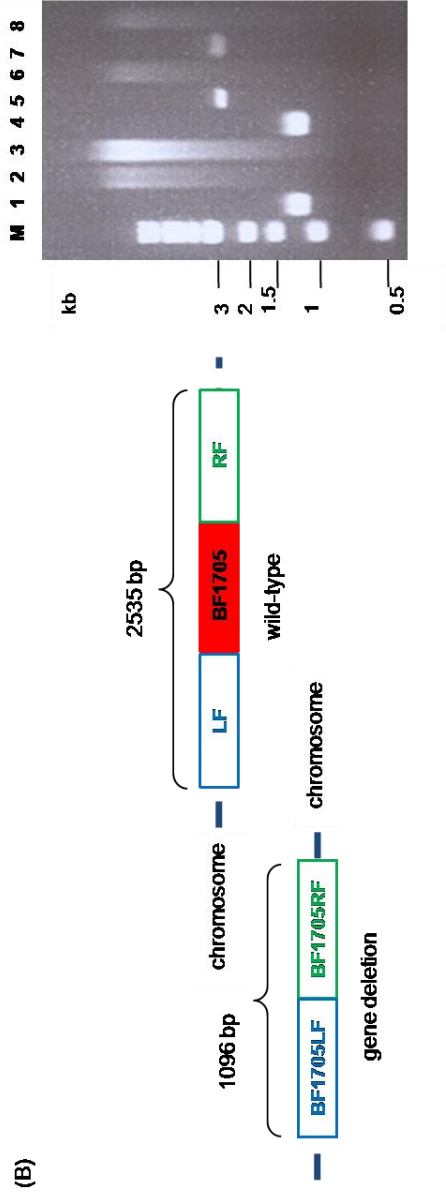
Figure 3.11: Vector map of pGB920

pGB920 was constructed by Blakely lab for I-SceI expression. Expression is controlled by a promoter derived from the fucose operon and is regulated by FucR. *tetQ* is the tetracycline selection marker. pGB920 is used for diploid resolution initiated by fucose induced I-SceI double-strandbreak.

Figure 3.12: Generation of the Δ BF1705 deletion mutant



(A) Graphic representation of the partial diploid carrying the BF1705 deletion construct prior to the fucose-mediated I-SceI double-strand break followed by electroporation with pGB920 and fucose induction which resolves the partial diploid into either a wild-type or a BF1705 gene deletion mutant. (B) Agarose gel analysis of PCR amplification of resolution of diploid strains into deletion mutant (lane 1 and 4) of ~1000 bp (expected size 1096 bp) or wild-type (lane 5 and 7) of ~2500 bp (expected size 2535 bp). Some of the tests failed to produce amplifications due to poor quality of the DNA extract used (lanes 2,3,6 and 8). Arrows and numbers on the left side indicate size of DNA in kb in a 1kb ladder



3.6. IFM analysis of fibrinogen binding by the Δ BF1705 strain

Binding of fibrinogen by wild-type NCTC 9343 cells and Δ BF1705 cells was examined by IFM. IFM slides were prepared by incubating cells fixed on glass slides with 0.1 mg/ml fibrinogen for 12 h followed by washing and labelling with 1^o antibody, anti-fibrinogen produced in goat and anti-goat FITC conjugated 2^o antibody to detect fluorescence of fibrinogen bound cells (2.7.13.1). The prepared slides were viewed using microscope in phase contrast and fluorescence mode. Fluorescence was observed in wild-type as well as Δ BF1705 cells. Adherence of fibrinogen to wild-type cells as indicated by fluorescence (Figure 3.13 B), agreed with the results of IFM studies conducted by Patrick et al. (2009), where fibrinogen binding by wild-type cells had attained saturation levels within a 2h exposure. Δ BF1705 cells exhibited a similar adhesion pattern as observed in the wild-type (Figure 3.13 D). Bright regions of fluorescence which lit up the cell outline was observed against a dark background. However, no fluorescence was detected in *E. coli* Top10 cells, which was used as a negative control (Figure 3.13 F). Background fluorescence was observed in *E. coli* Top10-immobilised slides, in spite of blocking the free binding sites on slides and cells with bovine serum albumin-phosphate-buffered saline (BSA-PBS) blocking agent to prevent non-specific binding. Further, 2x10 minute washes in PBS were also performed after each antibody incubation to wash away unbound antibodies. There has been no evidence that *E. coli* cells of the K12 genotype bind fibrinogen, and hence the choice of the same as negative control. This supports the specificity of the observed fibrinogen binding to *B. fragilis* wild-type and mutant cells. IFM results suggest that the fibrinogen-binding ability of *B. fragilis* is retained even in the absence of BF1705, an identified fibrinogen-binding protein.

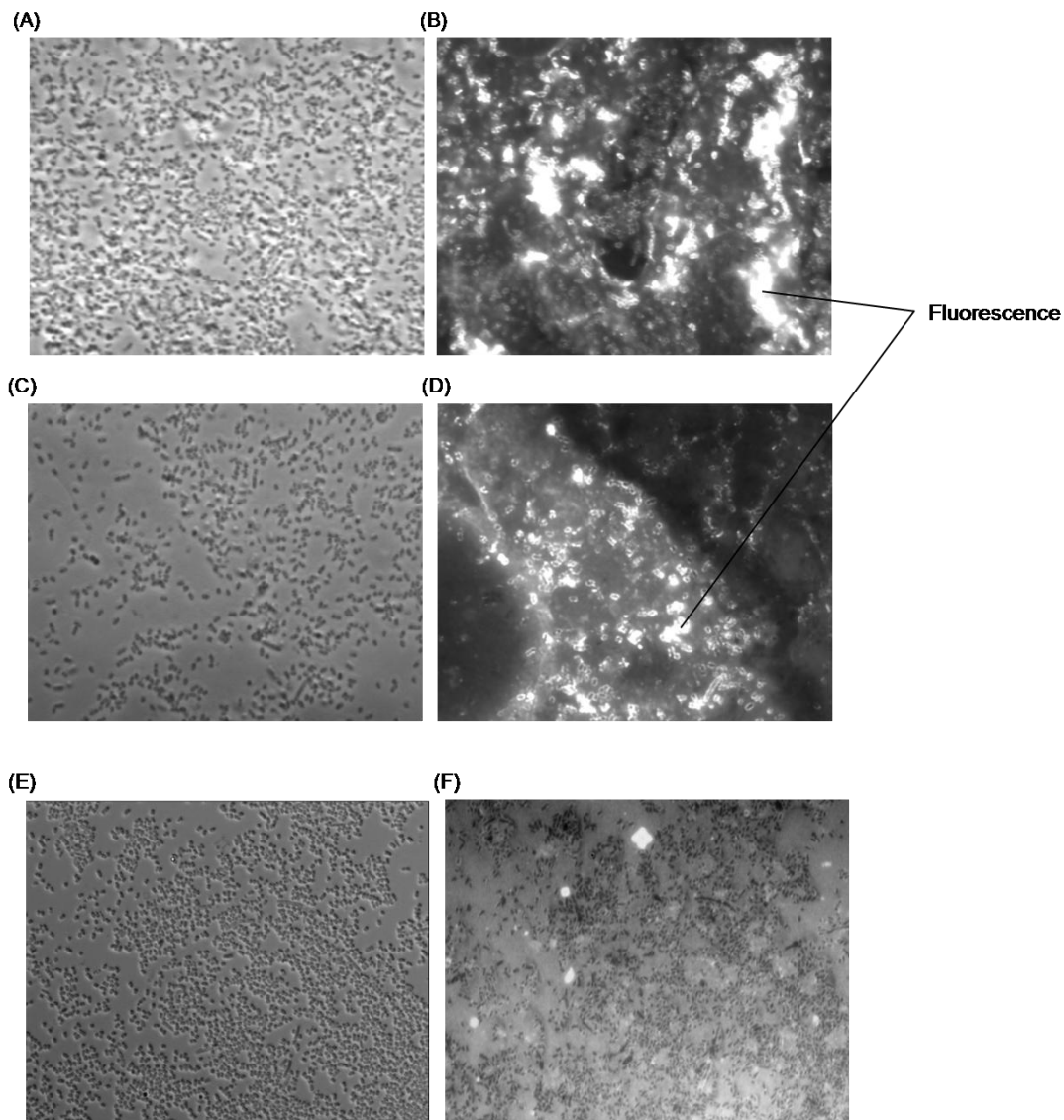


Figure 3.13: Immunofluorescence microscopy analysis of wild-type and Δ BF1705 cells incubated with fibrinogen

(A) Diluted wild-type NCTC 9343 cells incubated with 0.1 mg/ml fibrinogen for 12 h in phase contrast mode and (B) fluorescence mode. (C) Diluted Δ BF1705 cells incubated with 0.1 mg/ml fibrinogen for 12h in phase contrast mode and (D) fluorescence mode. (E) Diluted *E.coli* Top10 cells (negative control) incubated with 0.1 mg/ml fibrinogen for 12h in phase contrast mode and (F) fluorescence mode. Fluorescence observed in (B) and (D) indicates fibrinogen binding by the cells.

3.7. Investigating the role of capsular polysaccharides in fibrinogen binding

The observation that Δ BF1705 could still bind fibrinogen suggested that other surface components were capable of binding fibrinogen. A distinct feature of the *B. fragilis* surface is the expression of antigenically distinct large, small and micro-capsule polysaccharides (Lutton et al., 1991). The micro-capsule polysaccharides (electron-dense layer) have been implicated in colonization of the intestinal tract and in complement-mediated resistance by *B. fragilis* (Coyne et al., 2008; Reid and Patrick, 1984) (Figure 3.3). Sequence homology identified ten regions (PS A-J) in the *B. fragilis* genome annotated for capsular polysaccharide production, out of which eight are related to the production of phase variable micro-capsules (Cerdeño-Tárraga et al., 2005; Patrick et al., 2003). Patrick et al. (2009) observed that a marked deletion of BF1708 prevented assembly of High Molecular Mass Polysaccharides (HMMPs) into a long-chain-length polymer, thereby blocking micro-capsule production (Figure 3.3). BF1708 has been identified as a chain-length determining protein of ~41 kDa in *B. fragilis* and shared 20% identity to *E. coli* WzzB, a member of the polysaccharide copolymerase1 class that modulates the polymerase activity of the Wzy proteins involved in LPS biosynthesis to produce a distinctive modal distribution of O-antigen chain lengths (Franco et al., 1998). BF1708 also shares ~25% identity in 183 aa to *Thermoanaerobacter tengcongensis* capsular polysaccharide biosynthesis with an e value of 0.00032 (Table 3.2). The sequence homology of BF1708 to Wzz coupled with the inability of Δ BF1708 to produce micro-capsule layers confirmed the role of BF1708 in polysaccharide copolymerisation. Since micro-capsule expression has been associated with host attachment, colonisation and evasion of immune responses, the potential role of *B. fragilis* capsular polysaccharide in interacting with fibrinogen was examined using a Δ BF1708 deletion mutant.

3.7.1. Generation of a markerless double deletion of BF1705 and BF1708

For the generation of Δ BF1705 Δ BF1708, a markerless gene deletion construct for BF1708 of ~1200 bp in size, was generated by fusion-PCR, cloned in pGB910 and

maintained in *E. coli* S17-1 λ pir strain. The BF1708 deletion construct was mobilised from *E. coli* into the Δ BF1705 *B. fragilis* strain by conjugation (2.4.5). Erythromycin resistant transconjugants were selected and then electroporated with pGB920 followed by streaking tetracycline resistant colonies on fucose-containing defined medium. The resolved diploids were screened for the loss of erythromycin resistance and the Δ BF1705 Δ BF1708 double mutants were confirmed by PCR using primers specific for BF1705 deletion construct and BF1708 deletion construct (Figure 3.14 lane 1 and 2). PCR-amplified products were of ~1000 bp in size which were in agreement with the expected sizes of deletion constructs for BF1705 and BF1708.

3.7.2. IFM analysis of fibrinogen binding in Δ BF1705 Δ BF1708 cells

The role of micro-capsule polysaccharide as a fibrinogen-binding surface component was examined in the absence of BF1705. IFM slides were prepared by incubating Δ BF1705 Δ BF1708 double mutant cells, fixed on glass slides using 100% methanol, with 0.1 mg/ml fibrinogen for 12h followed by washing and incubation in anti-fibrinogen 1^o antibody and anti-goat FITC conjugated 2^o antibody to detect fluorescence of fibrinogen bound cells (2.7.13.1). The prepared slides were viewed using microscope in phase contrast and fluorescence mode. Clumping was observed in the Δ BF1705 Δ BF1708 cells in phase contrast mode (Figure 3.15 A) when compared to the cell distribution pattern in wild-type 9343 and Δ BF1705 IFM slides (Figure 3.15 D and F). In addition to clumping, flocculation of double mutant cells were also observed in liquid culture medium. These distinct phenotypic features could be attributed to the altered cell-surface properties brought about by BF1708 deletion (Patrick et al., 2009). However, the Δ BF1705 Δ BF1708 cells still bound fibrinogen as indicated from the fluorescence observed (Figure 3.15 C). Binding of fibrinogen was also detected in the positive control, wild-type 9343 cells, and Δ BF1705 cells (Figure 3.15 E and G). Binding of fibrinogen by the Δ BF1705 Δ BF1708 cells indicates that *B. fragilis* can interact with fibrinogen in the absence of both BF1705 fibrinogen-binding protein and micro-capsule polysaccharides. The absence of an electron-dense layer formed by micro-capsules

might enhance the accessibility of other surface components involved in fibrinogen interaction as observed in the non-capsulated strain of *S. aureus* which facilitated fibrinogen interaction with bound coagulase to prevent phagocytic uptake (Kapral, 1966).

Table 3.2: Description of BF1708 using artemis genome browser

BF1708 is a chain-length determining protein in *B. fragilis* with sequence homology to *Thermoanaerobacter tengcongensis* capsular polysaccharide biosynthesis (E value of 0.00032 and 25.13% identity in 183 aa), present in NCTC 9343 genome, using artemis genome browser.

Feature Name	CDS	Size	Hits
BF1708 <i>B. fragilis</i> chain-length determining protein	1992099...1993208	1110 bp 370 aa 40.917 kDa	Limited similarity to <i>Thermoanaerobacter tengcongensis</i> capsular polysaccharide biosynthesis TTE0704 (293 aa)

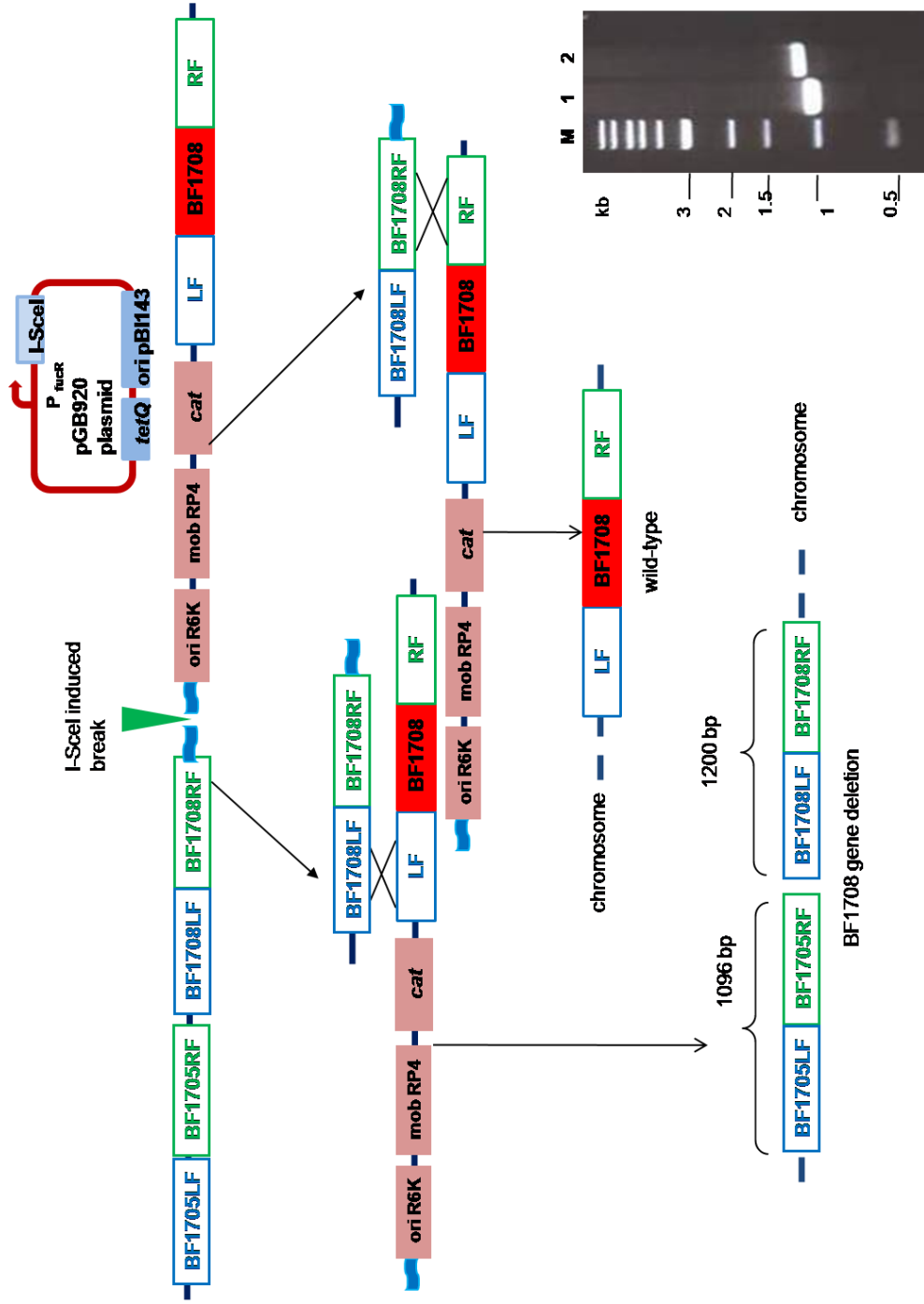
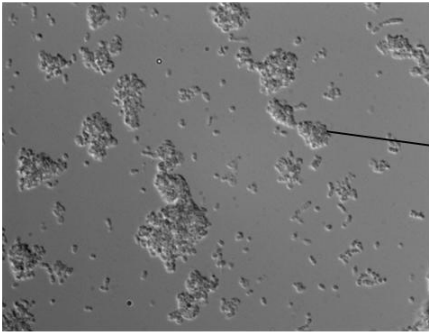


Figure 3.14: Generation of the Δ BF1705 Δ BF1708 double deletion mutant from the Δ BF1705

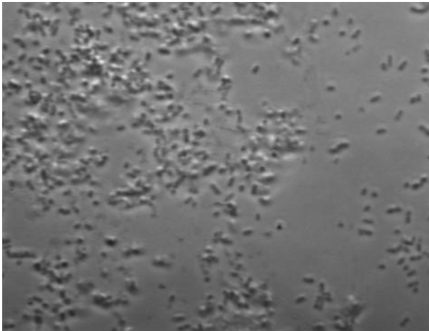
Graphic representation of the generation of Δ BF1705 Δ BF1708 by subjecting the partial diploid carrying the BF1708 deletion construct in Δ BF1705 strain to pGB920 electroporation followed by fucose induction. The resulting I-SceI double-strand break resolves the partial diploid into either BF1708 wild-type or deletion genotype by homologous recombination. Agarose gel analysis of PCR amplification of the double gene deletion mutants carrying deletion constructs for BF 1708 (lane 1) and BF 1705 (lane 2) of ~ 1000 bp (estimated sizes of 1200 bp and 1096 bp for BF1708 and BF1705, respectively). Arrows and numbers on the left side indicate size of DNA in kb in a 1kb DNA ladder (M).

(A)

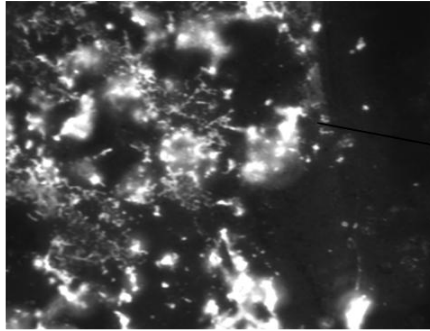


Cell-clumping

(B)

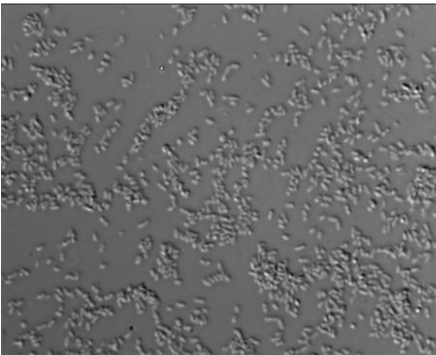


(C)

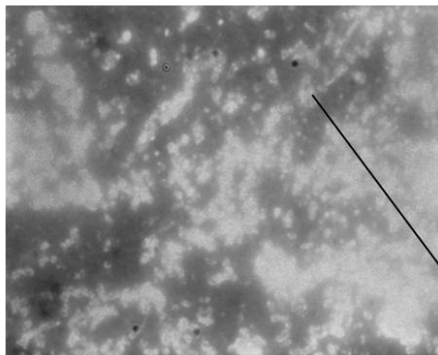


Fluorescence

(D)

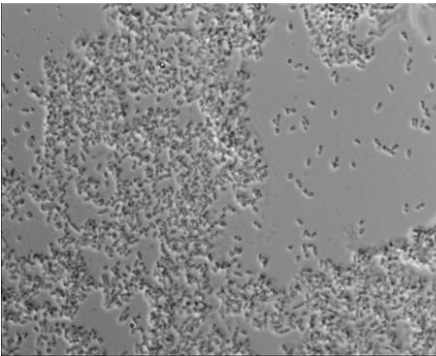


(E)



Fluorescence

(F)



(G)

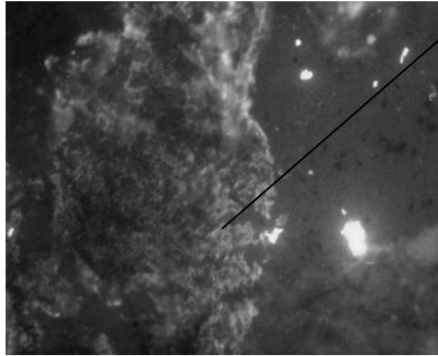


Figure 3.15: Immunofluorescence microscopy analysis of wild-type and Δ BF1705 Δ BF1708 cells incubated with fibrinogen

(A) Δ BF1705 Δ BF1708 cells incubated with 0.1mg/ml fibrinogen for 12 h in phase contrast mode. Clumping of cells can be observed. (B) Δ BF1705 Δ BF1708 cells incubated with 0.1mg/ml fibrinogen for 12 h in phase contrast mode and (C) fluorescence mode. (D) Wild-type 9343 cells incubated with 0.1mg/ml fibrinogen for 12 h in phase contrast mode and (E) fluorescence mode. (F) Δ BF1705 cells incubated with 0.1mg/ml fibrinogen for 12 h in phase contrast mode and (G) fluorescence mode. Fluorescence observed in (C), (E) and (G) indicate fibrinogen binding by the cells.

3.8. Expression and Purification of the BF1705 protein

Studies by Houston et al. (2010) have confirmed the ability of purified BF1705 protein to bind fibrinogen. Binding of fibrinogen by biotinylated his-tagged recombinant BF1705 was demonstrated using dot-blots and far-western-analysis. For expression, the mature BF1705 coding sequence (signal sequence and lipoprotein lipid attachment sites truncated) was amplified from *B. fragilis* NCTC 9343 genome and ligated into the MCS region of the pET-15b N terminal 6xHis-tag expression vector (Novagen). The expression vector was used to transform *E. coli* BL21(DE3) cells and the resultant clone was confirmed by DNA sequencing. The expression construct obtained from this work was used for BF1705 protein expression and purification in the present study. The purpose of purification was to use BF1705 as a positive control to detect fibrinogen binding in the subsequent blot analyses involving wild-type and Δ BF1705 concentrated cell-free supernatant extracts.

3.8.1. Transformation of *E. coli* BL21 cells with the pET-15b expressing BF1705

Recombinant pET-15b, carrying the BF1705 expression construct (gifted by Prof. Sheila Patrick, Queen's University, Belfast, UK), was recovered in *E. coli* BL21 (2.4.3). The resulting ampicillin resistant transformants were checked for the presence of recombinant pET15b by single colony analysis, restriction digestion analysis on 1% agarose gels and Sanger-based sequencing. Single colony analysis was performed by incubating the ampicillin resistant transformants in appropriate buffers (2.5.9) and confirming a molecular weight difference compared to the size of

the vector-backbone, due to the presence of BF1705 expression construct, on agarose gels (Figure 3.16 A). The presence of the expression construct in the isolated recombinant vectors were further confirmed by gel electrophoresis of extracted plasmids in comparison to the empty pET-15b vector backbone. Backbone and recombinant vectors were digested with *ApaI* enzyme and the linearised plasmids were checked for size by electrophoresis. A size difference corresponding to ~1.5 kb confirming the presence of BF1705 construct was observed in pET15b recombinant vectors (Figure 3.16 B lanes 1-4). Recombinant plasmids were subjected to Sanger-based sequencing and analysed by nucleotide BLAST. Two stretches of matches with 94% and 96% sequence identity confirmed the presence of BF1705 expression construct in the vector.

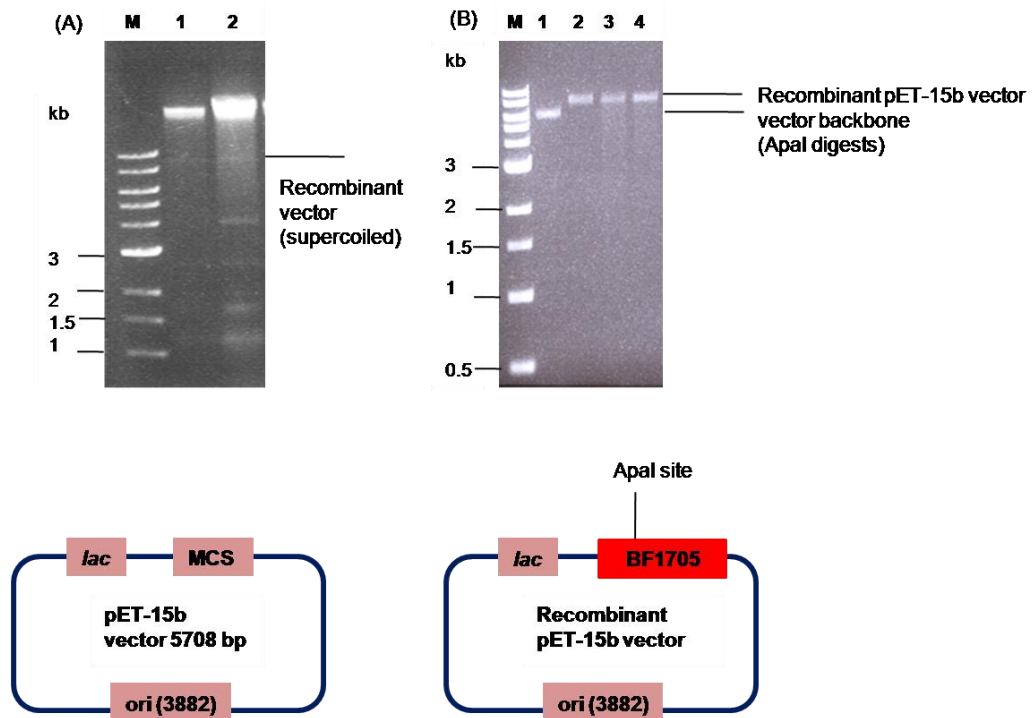


Figure 3.16: Confirmation of the recombinant vector carrying the BF1705 expression construct

(A) Agarose gel analysis of single colony assay showing the recombinant vector in supercoiled form (lanes 1 and 2) (B) Agarose gel analysis of ApaI digests of recombinant vector (lanes 2-4) in comparison with the vector backbone (lane 1). A shift in the size of recombinant pET-15b can be observed when compared to vector backbone (5708 bp). Arrows and numbers on the left side indicate size of DNA in kb (M). pET-15b backbone and recombinant vector carrying the BF1705 expression construct have been represented in graphics.

3.8.2. BF1705 protein expression

Expression of the BF1705 protein was brought about by growing ampicillin resistant BL21 derivatives carrying BF1705 expression construct in LB followed by induction at appropriate cell density using 0.5mM IPTG (2.7.10). Whole cell extracts were electrophoresed on 12% SDS- polyacrylamide gels and the induced BF1705 sample was compared with the uninduced sample. The expression band of the mature protein was confirmed at ~54 kDa, since the signal peptide sequence was eliminated while designing the expression construct (Figure 3.17 lanes 1,3,5).

3.8.3. Protein purification

A volume of 1l culture of BL21 derivative carrying BF1705 expression construct was induced with IPTG for 16h at 17°C in a 200 rpm shaker. 250 ml of the culture, expressing the BF1705 protein was centrifuged, and the pellet was resuspended in appropriate buffer and treated with protease inhibitor cocktail to prevent protein degradation (2.7.10). Cell pellet containing the crude protein extract was mixed with 35 ml loading buffer and passed through a french pressure cell press, in order to disrupt the cell membrane and recover the expressed protein. The sample was centrifuged and the supernatant containing the expressed protein was filter-sterilised. Primary protein purification was performed on 5 ml Ni-NTA agarose column that was washed 5 times with loading buffer (2.7.11). Samples were loaded and collected, followed by a column-wash and finally, elution buffer was loaded to the column to recover the his-tagged protein from the nickel column in 1ml aliquots. The loaded, washed and eluted samples were checked on SDS gels along with the induced culture and partially purified protein was identified in the eluted sample (Figure 3.18 A lanes 4-9 and B lanes 7,8,9). ~5 ml volume aliquots of partially purified BF1705 protein sample corresponding to high-concentration bands in the SDS gel (Figure 3.18 A lanes 4-7) were subjected to gel-filtration column chromatography. Pre-packed gel filtration columns (Hi-Prep Sephacryl S-200 HR) of 120 ml volume (1.6 cm x 60 cm), with a separation range of 5 kDa-250 kDa was used to obtain purified BF1705 protein. The protein sample was passed through the column (pre-equilibrated with

buffer) at a maximum back pressure of 0.15 MPa and recovery flow rate of 0.5 ml/min. The protein samples were examined on SDS PAGE and highly concentrated single bands were observed at the expected size of ~54kDa which corresponded to a single intensity peak in the UV₂₈₀ absorbance graph (Figure 3.19 A, B and C). Since the purified sample was of high concentration, fractions corresponding to 1D1, 1D3 and 1D4 in the absorbance graph (Figure 3.19 C) were electrophoresed on SDS gel (Figure 3.19 A lane 3,2 and B lane 1) and concentrations were determined as 0.38 mg/ml, 0.42 mg/ml and 0.4 mg/ml, respectively by nanodrop. These three individual fractions from the purified protein sample were of optimal concentrations and were used as positive controls in the subsequent experiments.

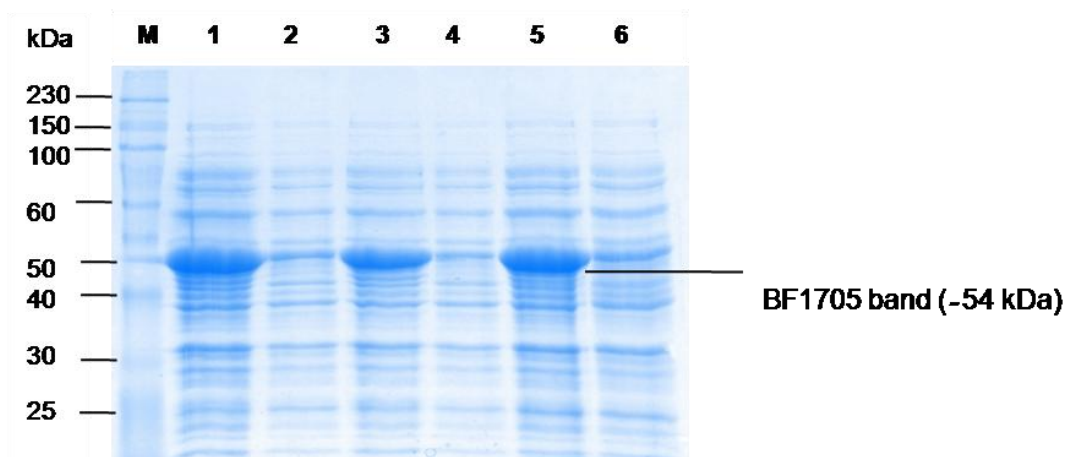


Figure 3.17: Expression of BF1705 in *E. coli*

SDS-PAGE analysis of IPTG-induced *E. coli* BL21 cells expressing the BF 1705 protein (lanes 1,3,5) at ~54 kDa in comparison with uninduced cultures (lanes 2,4,6). The positions of molecular weight markers (M) in kDa are shown adjacent to the gel.

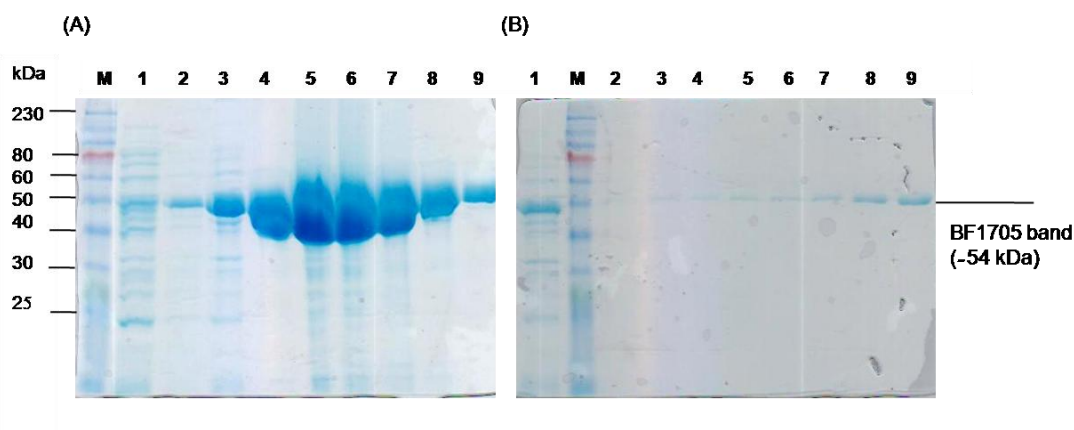


Figure 3.18: Purification of BF1705 by nickel column chromatography

SDS-PAGE analysis of (A) load buffer (lane 1), wash buffer (lane 2), induced culture (lane 3) eluted his-tagged BF1705 protein (lanes 4-9) and (B) uninduced culture (lane 1), eluted his-tagged BF1705 protein (lanes 2-9). The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. The purified BF1705 protein was observed in the eluted fraction at ~54kDa which corresponds to the expected size of the mature BF1705 protein.

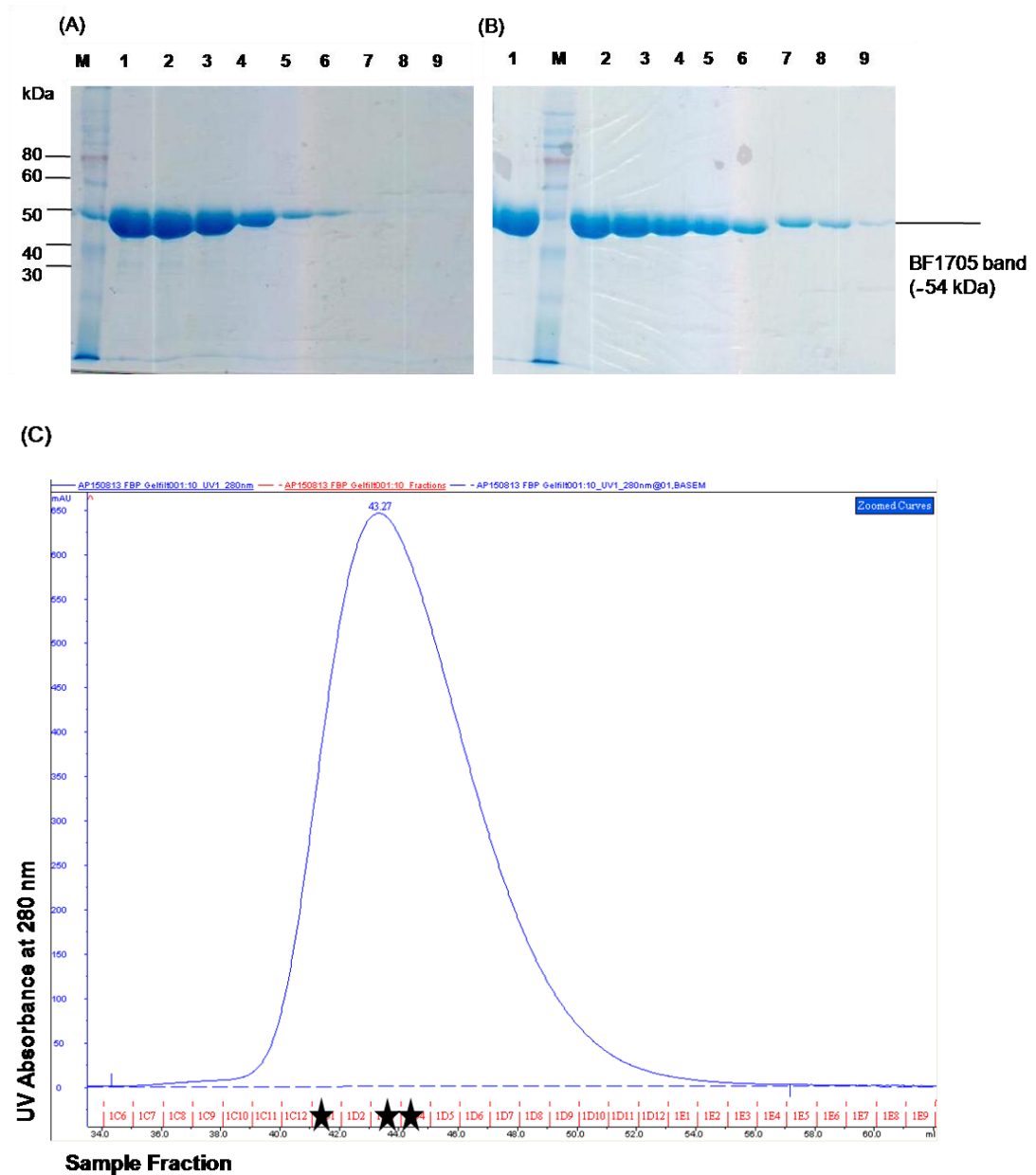


Figure 3.19: Purification of BF1705 by gel filtration column chromatography

SDS-PAGE analysis of gel filtration column purified fractions of BF1705 protein. Highly purified concentrated fractions were observed in (A) (lanes 1-4) and (B) (lane 1, lanes 2-5). The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. (C) Graphical representation of UV absorbance at 280 nm vs fraction of the purified BF1705 protein. Fractions 1D1-1D8, corresponding to the peak in the graph were highly concentrated as observed in (A) 1D1 (lane 1) 1D2 (lane 2) 1D3 (lane 3) and (B) 1D4 (lane 1) 1D5 (lane 2) 1D6 (lane 3) 1D7 (lane 4) and 1D8 (lane 5). Fractions 1D1 (0.38 mg/ml), 1D3 (0.4 mg/ml) and 1D4 (0.42 mg/ml) denoted by ★ were chosen for subsequent experiments.

3.9. Analysis of fibrinogen binding by *B. fragilis* concentrated culture supernatants in Immunoblots

Binding of fibrinogen by wild-type 9343, Δ BF1705 and Δ BF1705 Δ BF1708 whole cells had been previously observed by IFM (3.7.2). Western blot experiments were designed to further analyse the IFM results by determining the sizes of individual surface proteins of wild-type and mutant cells involved in fibrinogen interaction. Concentrated culture supernatants and whole cell fractions of wild-type and mutant strains were electrophoresed on SDS-gels and transferred onto PVDF membranes by wet blotting (2.7.3 and 2.7.7). Immunoblots on PVDF membrane was blocked using dried skimmed-milk to prevent non-specific binding. The blocked membranes were incubated in appropriate concentrations of fibrinogen protein and antibody solutions. Fibrinogen binding to fractionated proteins blotted on the PVDF membrane was detected as signals on X ray films by chemical luminescence of the blots.

3.9.1. Choice of sample preparation methods for Immunoblot analyses

BF1705 has been identified as a protein expressed on the surface of *B. fragilis*. Interaction with the host is facilitated by the outer membrane and OMVs secreted by bacteria. Therefore, preparation of samples in a manner that emphasises the outer membrane of the cell would be effective in detecting fibrinogen binding by BF1705 and potentially, by other surface proteins. Cell-free culture supernatants of stationary phase cultures of Δ BF1705 and wild-type 9343 strains grown in glucose DM medium were concentrated using 100 kDa MWCO centrifugal filters (2.7.2). The concentrated supernatants of Δ BF1705 and wild-type 9343 stationary phase cultures were analysed on SDS-polyacrylamide gels (Figure 3.20 lane 1 and 3) along with BF1705 uninduced, induced cultures of *E. coli* BL21 (Figure 3.20 lane 4 and 5) and purified BF1705 protein (Figure 3.20 lane 6) to confirm protein size and concentration before commencing western blot analyses of the samples. Concentrated culture supernatants provide a representation of the outer membrane of the cell with respect to secreted proteins, enzymes, OMVs and surface associated proteins. OMVs contain biologically active proteins, lipoproteins and phospholipids,

and previous studies have emphasized their role in processes such as intercellular communication, transfer of contents to host cells and eliciting immune responses (Kulkarni and Jagannadham, 2014). Therefore, concentrated supernatants were the preferred choice for analysing interactions of fibrinogen protein with the cell surface of *B. fragilis*. Concentration of purified protein and culture supernatants were determined by Bradford protein assay (2.7.12).

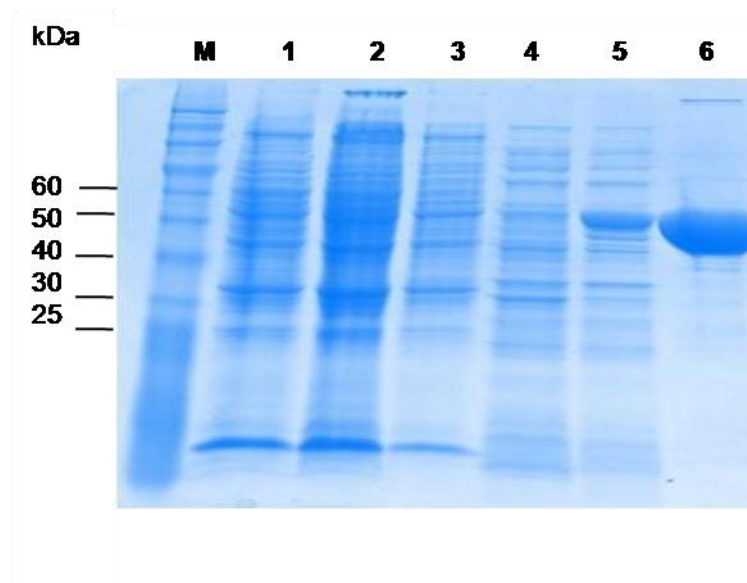


Figure 3.20: Analysis of concentrated culture supernatants and BF1705

SDS-PAGE analysis of concentrated culture supernatants of Δ BF1705 (lane 1), Δ BF1705 Δ BF1708 (lane 2; will be discussed later), wild-type 9343 (lane 3), uninduced and induced culture for BF1705 expression (lane 4 and 5), purified BF1705 protein (lane 6). The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. A band at ~54 kDa in the induced culture and purified protein indicate the presence of BF1705 protein.

3.9.2. Detection of BF1705 protein in concentrated supernatant samples

Identification of the BF1705 protein in concentrated wild-type 9343 supernatant samples would provide an indication of the presence of surface proteins in the prepared concentrated culture supernatants. Wild-type 9343 and Δ BF1705 strains, and dilutions of purified BF1705 protein were analysed by incubating the PVDF membrane blots of the electrophoresed samples (Figure 3.21 A) with 1^o antibody, polyclonal anti-1705 produced in rabbit and anti rabbit hrp-conjugated 2^o antibody (2.7.7 and 2.7.8). Visualisation of PVDF blots on X-ray film at two different exposure times revealed a single band at ~54 kDa for concentrated wild-type 9343 supernatant (Figure 3.21 B lane 3 and C lane 3) whereas no bands were observed for the Δ BF1705 concentrated supernatant. High intensity signals were detected for the positive control, BF1705 protein samples, even at a low concentration of 0.4 μ g and 0.06 μ g of the protein (Figure 3.21 B lane 1, 4 and C lane 1, 4) which confirms specificity of the protein to the antibody. The banding patterns of BF1705 at its expected size in the concentrated wild-type 9343 supernatant suggest the presence of the protein in OMV released by *B. fragilis*. Absence of bands in concentrated Δ BF1705 supernatant is consistent with deletion of BF1705 in this strain.

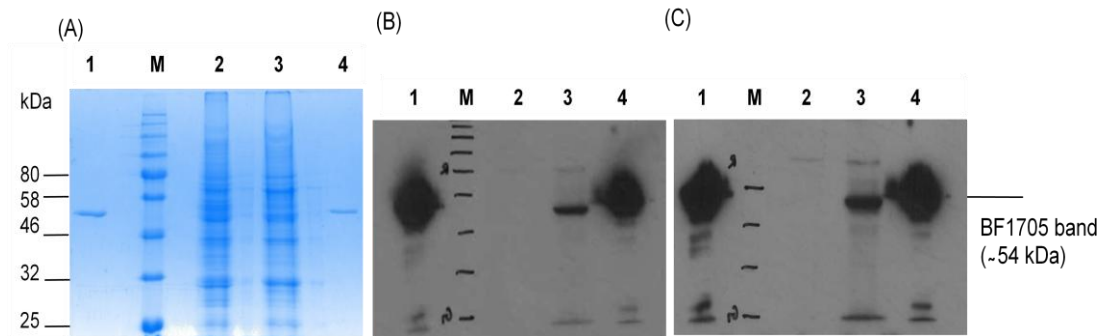


Figure 3.21: Detection of BF1705 in concentrated culture supernatants

(A) SDS-PAGE analysis of diluted BF1705 purified protein (0.4 µg) (lane 1), concentrated culture supernatants of Δ BF1705 (35 µg) (lane 2) and wild-type 9343 (35 µg) (lane 3) and diluted BF1705 protein (0.06 µg) (lane 4). (B) and (C) Immunoblots of the corresponding samples using anti-1705 antibody for (A) 2s and (B) 10s exposure. The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. A band was observed at ~54 kDa in immunoblots of BF1705 protein (B) (lane 1, 4) and (C) (lane 1, 4) and of wild-type 9343 concentrated supernatant (B) (lane 3) (C) (lane 3).

3.9.3. Far-western analysis of fibrinogen binding in Δ BF1705

Further investigation of the binding results obtained from IFM (3.6) was done using far-western blots to analyse fibrinogen binding by concentrated cell supernatants and the positive control, BF1705 protein. PVDF membrane blots were blocked in dried, semi-skimmed milk to prevent non-specific binding and incubated in 10 μ g/ml fibrinogen solution. Fibrinogen binding to the blotted samples was detected by incubation with 1^o antibody, anti-fibrinogen produced in goat and anti goat hrp-conjugated 2^oantibody (2.7.7). Visual analysis of PVDF blots on X-ray films (2.7.8) revealed multiple bands at ~54 kDa and ~57 kDa for 1x and 2x dTT treated concentrated wild-type 9343 supernatants (Figure 3.22 B lane 1 and 2). Bands were not present in concentrated culture supernatants of Δ BF1705 (Figure 3.22 B lane 3). Also, fibrinogen binding was not observed in the purified BF1705 protein at a concentration of ~ 2 μ g (Figure 3.22 B lane 4). However, a band corresponding to ~54 kDa was observed in purified BF1705 protein of ~5 μ g (Figure 3.23 B lane 1, Figure 3.24 B lane 3 and C lane 3). Multiple bands were still observed for 9343 concentrated supernatants on replicate analyses (Figure 3.23 B lane 3 and Figure 3.24 B lane 4). Fibrinogen binding was not detected in concentrated Δ BF1705 supernatant as no bands were observed (Figure 3.23 B lane 5 and Figure 3.24 B lane 6). *E. coli* Top10 whole protein extract was used as a negative control which did not show any binding (Figure 3.23 B lane 4 and Figure 3.24 B lane 5). Top10 strain is derived from *E. coli* K12 and no fibrinogen binding has been reported in this strain. The inability to bind fibrinogen by Top10 cells has also been observed in IFM conducted in the present study. On increasing the exposure time of X-ray film to 5 min, faint bands were observed at the expected size for diluted samples of BF1705 protein (Figure 3.24 C lane 1 and 2). The far-western analysis of purified protein confirm fibrinogen binding to BF1705 at high concentrations of the surface protein (> 4 μ g) (Figure 3.23 B lane 1 and Figure 3.24 B lane 3). This indicates that fibrinogen binding by BF1705 protein is concentration-dependent and is facilitated only at relatively higher concentrations of the latter. Multiple bands observed in the extract from wild-type strain might suggest the binding of fibrinogen by the mature

protein at ~54 kDa as well as the native protein at ~57 kDa. The involvement of disulphide bonds in the observed multiple banding is not likely since the concentrated supernatant extracts were treated with varying concentrations of dTT to reduce disulphide bridges. The lower bands might be fibrinogen binding to the degradation products of BF1705 in concentrated wild-type supernatants. The probability of non-specific binding can be ruled out since the blots were blocked in dried skimmed milk and no binding was observed in the mutant extract. The absence of bands in Δ BF1705 extract also indicate the absence of other surface proteins involved in fibrinogen binding.

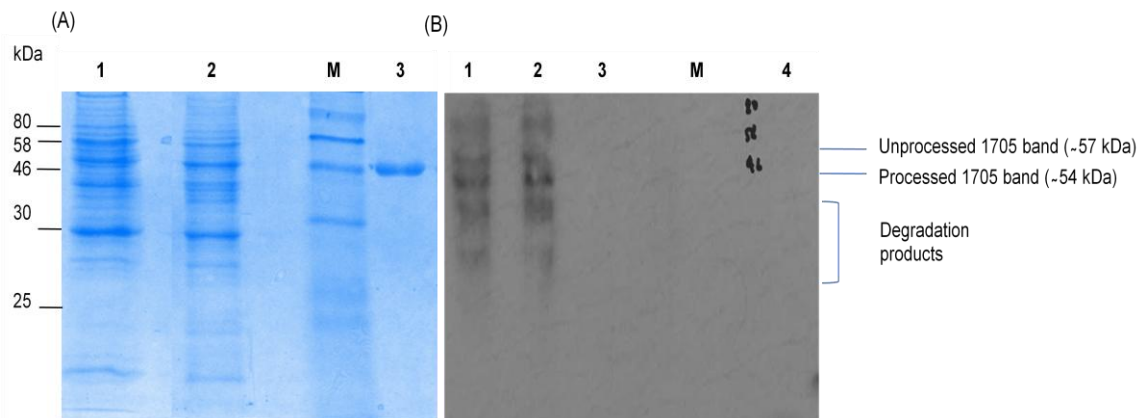


Figure 3.22: Analysis of fibrinogen binding by the wild-type and Δ BF1705 culture supernatants

(A) SDS-PAGE analysis of concentrated culture supernatants of wild-type 9343 (70 μ g) (lane 1), Δ BF1705 (70 μ g) (lane 2) and purified 1705 protein (2 μ g) (lane 3). (B) Immunoblot of 1x dTT treated and 2x dTT treated concentrated culture supernatants of wild-type 9343 (70 μ g) (lane 1 and 2), concentrated culture supernatant of Δ BF1705 (70 μ g) (lane 3) and purified BF1705 protein (2 μ g) (lane 4) incubated in fibrinogen followed by labelling with anti-fibrinogen antibody. The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. Multiple bands were observed at ~ 57 kDa, ~54 kDa and at lower molecular sizes in 1xdTT treated (B) (lane 1) and 2xdTT treated (B) (lane 2) concentrated culture supernatants of wild-type 9343.

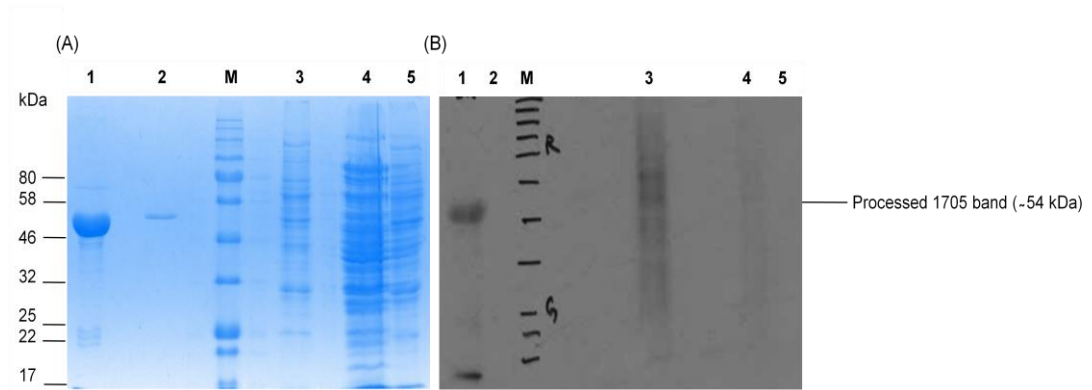


Figure 3.23: Detection of fibrinogen binding by BF1705

(A) SDS-PAGE analysis of concentrated purified BF1705 protein (5 µg) (lane 1), diluted BF1705 protein (0.3 µg) (lane 2), concentrated culture supernatant of wild-type 9343 (35 µg) (lane 3), *E. coli* Top10 whole cell fraction (lane 4) and concentrated culture supernatant of Δ BF1705 (35 µg) (lane 5) and (B) Immunoblot of corresponding samples incubated with fibrinogen followed by labelling with anti-fibrinogen antibody (lanes 1,2, 3-5) for 10 min exposure. The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. A band was observed at ~ 54 kDa in concentrated purified BF1705 protein (B) (lane 1) and a smeared band was observed in concentrated wild-type culture supernatant (B) (lane 3).

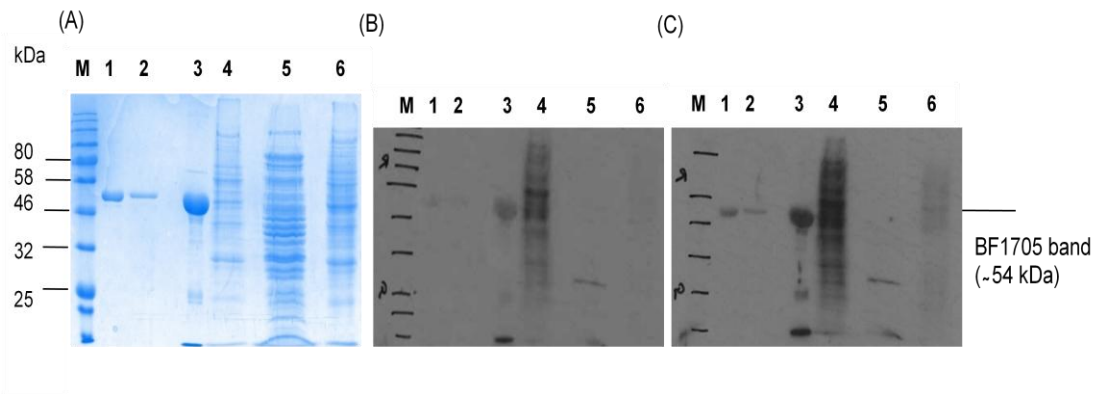


Figure 3.24: Analysis of fibrinogen binding by the wild-type and differing BF1705 concentrations

(A) SDS-PAGE analysis of dilutions of purified BF1705 protein (2 μg) (lane 1) (0.35 μg) (lane 2) , concentrated BF1705 protein (5 μg) (lane 3), concentrated culture supernatant of wild-type 9343 (35 μg) (lane 4), *E.coli* Top10 whole cell lysate (lane 5) and concentrated culture supernatant of ΔBF1705 (35 μg) (lane 6) and (B) Immunoblots of the corresponding samples incubated in fibrinogen followed by labelling with anti-fibrinogen antibody (lanes 1-6) for (B) 2 min. and (C) 5 min. exposure. The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. Bands were observed at ~54 kDa in concentrated BF1705 protein (B) (lane 3) (C) (lane 3) and smeared multiple bands in concentrated wild-type culture supernatant (B) (lane 4) (C) (lane 4). On extending exposure time to 5 min., bands were observed for dilutions of purified BF1705 protein (C) (lane 1 and 2).

3.9.4. Analysis of fibrinogen binding by Δ BF1705 Δ BF1708 cells in Immunoblots

The binding of fibrinogen observed by IFM in mutant cells fixed on slides contradicted the far-western results, in which no binding was observed in the mutant extract. This might be due to the presence of fibrinogen-binding surface components on whole cells observed in IFM which were absent in concentrated supernatant extracts that were fractionated on SDS-polyacrylamide gels. However, non-specific background fluorescence observed in a few IFM slides, even after blocking the slides and repeated washes, could pose a hindrance in accurate interpretation of binding observed by IFM. Therefore, owing to the sensitivity and relative reliability of far-western analysis, diluted whole cells incubated with fibrinogen were washed, electrophoresed and blotted on to PVDF membranes. The blotted membranes were analysed for presence of fibrinogen by incubating in specific antibodies. Fibrinogen protein was used as a positive control and *E. coli* Top10 cells were used as negative control. Interactions of 9343, Δ BF1705 and Δ BF1705 Δ BF1708 whole cells with fibrinogen were analysed by incubating 0.5 ml diluted cultures with 0.1 mg/ml fibrinogen. After 10h, excess fibrinogen was washed off by centrifuging the treated cells thrice in 1 ml PBS. SDS-polyacrylamide electrophoresed samples were transferred by membrane blotting and incubated with fibrinogen-specific antibody followed by incubation with hrp-conjugated 2^o antibody for detection. Fibrinogen bands, corresponding to the three chains, were observed even in the negative control, *E. coli* Top10 (Figure 3.25 B lane 4). Although fibrinogen binding has been reported in a few clinical isolates of *E.coli* strains implicated in bacteraemia, it has not been detected in the K12 genotype to which *E.coli* Top10 belongs. Also, fibrinogen binding by *E.coli* Top10 cells was not observed in IFM and far-western analysis performed in the present study. Therefore, the presence of fibrinogen observed in the immunoblot of *E.coli* whole cell fraction could be regarded as an artefact. The presence of fibrinogen in the residual volume after the PBS washes performed to remove excess fibrinogen might have accounted for a positive signal in the *E.coli* whole cell fraction. The observed artefact, owing to the sensitivity of the technique adopted, could be addressed by using lower concentrations of fibrinogen for binding

and increasing the number of washes to remove unbound fibrinogen. Therefore, the immunoblots were repeated by incubating whole cells with 0.015 mg/ml fibrinogen and increasing the number of PBS washes to 5 times. The signals were not over exposed due to the lower concentration of fibrinogen used. However, fibrinogen binding by *E. coli* was still observed (Figure 3.26 B lane 4). This might still account for a false-positive band or could suggest fibrinogen binding by *E. coli* that was not detected by less sensitive techniques. The evidence that ~20% of *E.coli* genome accounts for unknown genes coupled with the fact that *E. coli* co-exists with *B. fragilis* in the gastrointestinal tract and has been isolated from polymicrobial infections suggests that *E. coli* might harbour surface components involved in fibrinogen binding. Moreover, Shen et al. (1995) observed that *E. coli* strains from patients with colonic disorders expressed binding of fibrinogen which may promote pathogenesis of colonic diseases.

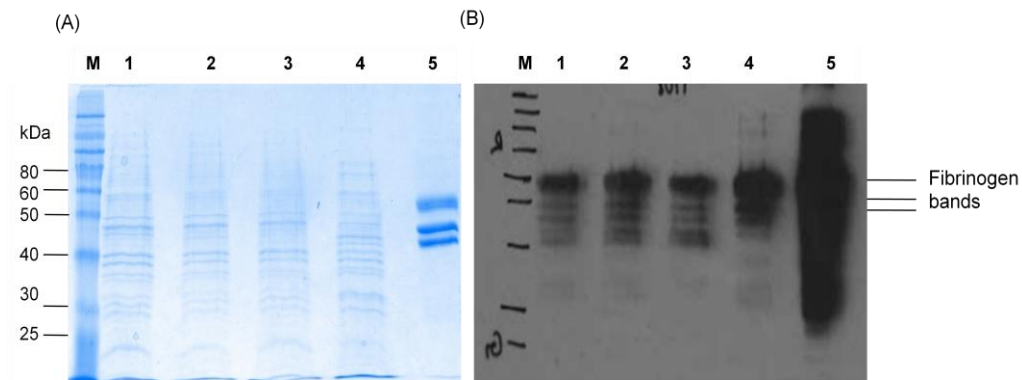


Figure 3.25: Analysis of fibrinogen binding (0.1 mg/ml) by whole cell lysates of wild-type and deletion mutants

(A) SDS-PAGE analysis of diluted lysates of wild-type 9343 (lane1), BFΔ1705 (lane 2), BFΔ1705Δ1708 (lane 3), *E. coli* Top10 (negative control) (lane 4) incubated with 0.1 mg/ml fibrinogen and 0.1 mg/ml fibrinogen (positive control) (lane 5) and (B) Immunoblot of corresponding samples labelled with anti-fibrinogen antibody (lane 1-4) for 10s exposure. The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. Bands are observed at ~60 kDa, ~50 kDa and ~40kDa in wild-type 9343 lysate (B) (lane 1) BFΔ1705 lysate (B) (lane 2) BFΔ1705Δ1708 lysate (B) (lane 3) and *E.coli* Top10 lysate (B) (lane 4) and over-exposure is observed in fibrinogen (B) (lane 5).

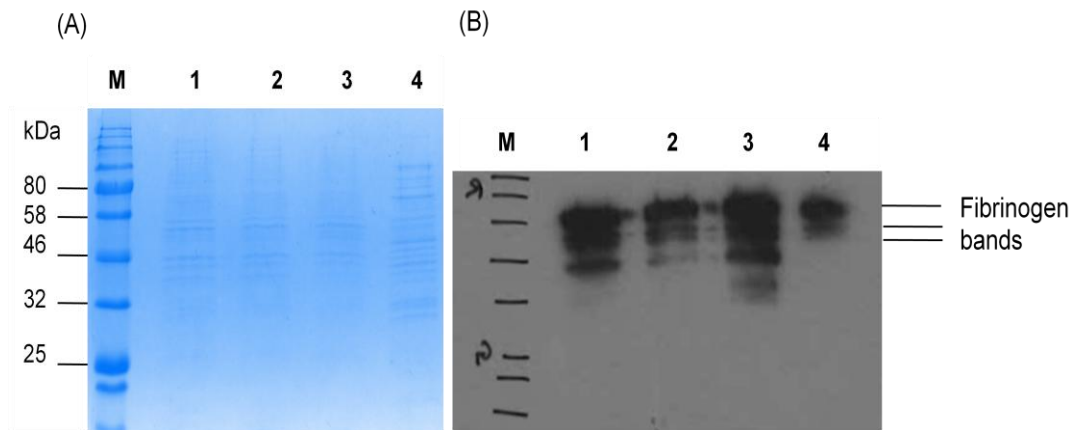


Figure 3.26: Analysis of fibrinogen binding (0.015 mg/ml) by whole cell lysates of wild-type and deletion mutants

(A) SDS-PAGE analysis of diluted cultures of wild-type 9343 (lane1), BF Δ 1705 (lane 2), BF Δ 1705 Δ 1708 (lane 3) and *E. coli* Top10 (lane 4) incubated with 0.015 mg/ml fibrinogen and (B) Immunoblot of corresponding samples labelled with anti-fibrinogen antibody (lanes 1-4) for 5s exposure. The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. Bands are observed at ~60 kDa, ~50 kDa and ~40kDa in wild-type 9343 lysate (B) (lane 1) BF Δ 1705 lysate (B) (lane 2) BF Δ 1705 Δ 1708 lysate (B) (lane 3) and *E.coli* Top10 lysate (B) (lane 4).

3.10. Detection of fibrinogen degradation by the Δ BF1705 strain

Lantz, Allen, Vail, et al. (1991) identified a 150-kDa cell surface protein in the periodontal pathogen, *Porphyromonas gingivalis*, that mediated binding and degradation of human fibrinogen. It was hypothesised that the fibrinogen binding component was a non-catalytic form of fibrinogen degrading component. It has been demonstrated that *B. fragilis* can degrade fibrinogen, in addition to binding fibrinogen (Houston et al. 2010). However, the role of the identified fibrinogen-binding protein, BF1705 in *B. fragilis*-mediated degradation of fibrinogen was not investigated. The genetic similarity between the two bacteria might suggest a similar mechanism for fibrinogen degradation in *B. fragilis* as observed in *P. gingivalis* albeit there is no significant similarity between the two surface proteins. Binding of fibrinogen followed by its degradation could provide an effective strategy for *B. fragilis* adhesion and invasion of host cells and prolonging fibrin-mediated clotting

and abscess formation. Ability of BF1705 protein to degrade fibrinogen was analysed by incubating 0.1mg/ml fibrinogen with Δ BF1705 strain and 9343 wild-type strain in two separate tubes of BHI-S followed by analysis of the culture supernatants at 3 h and 48 h time periods in SDS-polyacrylamide gels for degradation of fibrinogen bands. Fibrinogen incubated in plain BHI-S served as the control sample. Complete degradation of the A α -chain, as observed by the disappearance of the band, and partial degradation of B β - and γ -chains of fibrinogen were observed in both the cultures after 48h (Figure 3.27 lane 2 and 6) when compared to intact bands in the control-BHI-S sample (Figure 3.27 lane 8). Partial degradation of the A α -chain was observed in the 3h culture samples (Figure 3.27 lane 1 and 5). BF1705 protein does not play a role in the ability of *B. fragilis* to degrade fibrinogen as the wild-type and mutant strains degraded fibrinogen when compared to the control sample, which is evident from the disappearance of bands corresponding to the A α -chain. These observations suggest that fibrinogen degradation by *B. fragilis* might take place independently of fibrinogen binding. Fibrinogen degradation by *B. fragilis* has to be investigated in detail to support this hypothesis.

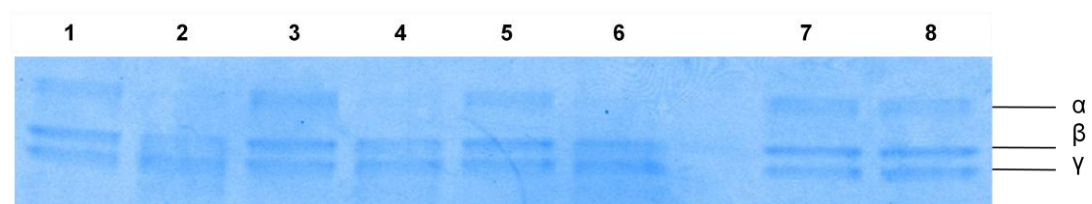


Figure 3.27: Analysis of fibrinogen degradation by the wild-type and Δ BF1705 cultures

SDS-PAGE analysis of in situ fibrinogen degradation by 3h (lane 1) and 48h (lane 2) grown cultures of wild-type 9343 and 3h (lane 5) and 48h (lane 6) grown cultures of Δ BF1705 on comparison to fibrinogen in control BHI-S medium (lanes 7 and 8). Disappearance of fibrinogen α -band is observed in 48h grown culture of wild-type 9343 (lane 2) and 48h grown culture of Δ BF1705 (lane 6). Intact bands are observed in control BHI-S medium (lane 7 and 8).

3.11. Binding potential of BF1705 protein to fibronectin by far-western analysis

Fibronectin is a glycoprotein present in the ECM that binds other components of the ECM, such as, collagen, fibrin and heparan sulphate, in addition to membrane spanning integrins. Pereira et al. (2002) observed that incorporation of fibrinogen into ECM during events of tissue repair required active elongation of fibronectin polymers into matrix fibrils. This observation provided a link between fibrinogen and fibronectin functions in the host. Interestingly, studies by Lantz, Allen, Vail, et al. (1991); Lantz, Allen, Duck, et al. (1991) suggested that fibrinogen and fibronectin are recognised and modified by the same 150 kDa cell surface component on *P. gingivalis*. However, the affinity of *P. gingivalis* for fibrinogen was ten-fold higher than that observed for fibronectin. In a similar manner, a bifunctional surface protein, FnbpA, has been identified in *Staphylococcus aureus* that binds fibrinogen in addition to fibronectin and interdependently promote endothelial invasion by *S. aureus* (Piroth et al., 2008; Wann et al., 2000). Studies by Pauer et al. (2009) identified a 102-kDa TonB-dependent protein as a potential fibronectin binding protein in *B. fragilis* strain 1405. The putative fibronectin binding protein was examined by reverse genetics. Enhanced binding of fibronectin by a gene-inactivated mutant (Δ bf1991) in *B. fragilis* strain 638R was observed in latex agglutination assays when compared to the wild-type. Western blot analysis detected an increase in the expression of an unknown 30 kDa protein in the absence of the TonB-dependent fibronectin binding protein. Increased fibronectin binding observed in the TonB-dependent protein encoding gene-inactivated mutant suggests the role of other components that might mediate fibronectin binding by *B. fragilis* (Pauer et al., 2013). These observations regarding microbial interaction with fibronectin suggests a possible role played by BF1705 in fibronectin binding by *B. fragilis*.

An attempt was made to identify potential fibronectin adhering functions of BF1705 by performing far-western analysis using human fibronectin. Concentrated culture supernatants of 9343 and Δ BF1705, concentrated BF1705 protein, human fibronectin (positive control) and BSA (negative control) were electrophoresed on 12% SDS-polyacrylamide gel and transferred onto PVDF membrane by blotting.

Membrane blot was blocked and incubated with 10µg/ml fibronectin, followed by labelling with 1^o antibody, anti-fibronectin produced in rabbit and anti-rabbit hrp conjugated 2^o antibody (2.7.7 and 2.7.8). Visualisation of the incubated blot on X-ray films revealed an intense band for the positive control, fibronectin (Figure 3.28 B lane 2), a faint band at ~54 kDa for BF1705 protein sample (Figure 3.28 B lane 3) and a smeared band with a distinct intensity at ~32 kDa for the wild-type concentrated supernatant (Figure 3.28 B lane 4). No fibronectin binding was observed for the negative control BSA (Figure 3.28 B lane 1) and for ΔBF1705 concentrated culture supernatant (Figure 3.28 B lane 5). These observations indicate that BF1705 protein can bind fibronectin in addition to fibrinogen. Although, the band at ~32 kDa in 9343 suggest fibronectin binding by surface components other than BF1705 (Figure 3.28 B lane 4), the absence of BF1705 seems to inhibit fibronectin binding by the concentrated culture supernatant (Figure 3.28 B lane 5). However, a band at 102 kDa, corresponding to the previously studied putative fibronectin-binding protein was absent in our study. The absence of the TonB-dependent fibronectin-binding protein might be due to the difference in *B. fragilis* strains used between the two studies. The band at ~32 kDa in the present study could even be a degradation product of BF1705 in the wild-type concentrated supernatant which binds fibronectin. The probability of non-specific fibronectin binding by BF1705 protein could be ruled out since the blots were treated with dried skimmed milk to block the free binding sites. Alternatively, the band observed at ~32 kDa might be similar to the band observed by Pauer et al. (2013) in the fibronectin binding far-western analysis of gene-inactivated mutant.

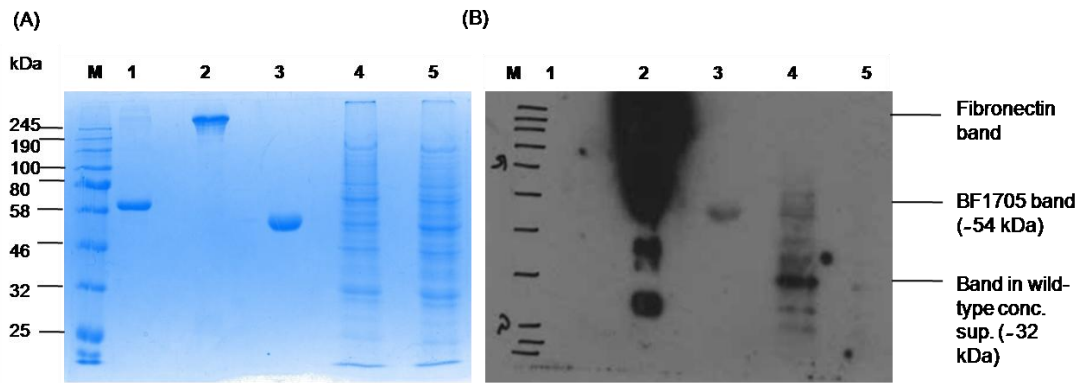


Figure 3.28: Analysis of fibronectin binding by BF1705 and wild-type and Δ BF1705 culture supernatants

(A) SDS-PAGE analysis of BSA negative control (2 μ g) (lane 1), fibronectin (6 μ g) (lane 2), purified BF1705 protein (4 μ g) (lane 3), concentrated culture supernatants of wild-type 9343 (35 μ g) (lane 4) and BF Δ 1705 (35 μ g) (lane 5) and (B) Immunoblots of corresponding samples incubated in 10 μ g/ml fibronectin and labelled with anti fibronectin antibody for 2.5 min. exposure. The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. Over-exposure is observed in fibronectin positive control (B) (lane 2), band at ~54 kDa was observed in purified BF1705 protein (B) (lane 3) and a smeared band with intensity at ~32 kDa is observed in concentrated wild-type 9343 supernatant (B) (lane 4).

3.12. Discussion

Binding of host proteins by commensals of the gastrointestinal tract has been regarded as an effective mechanism for microbial survival. Bacterial adherence to the host epithelium is essential for obtaining nourishment and for evading adverse host immune responses (Sonnenburg et al., 2004). However, binding of opportunistic pathogens to the host matrix-associated proteins has been implicated in infections (Westerlund and Korhonen, 1993). In the present study, the fibrinogen binding function of BF1705 protein in *B. fragilis* was confirmed by generating BF1705 gene-deletion mutants in the NCTC 9343 strain. The reverse genetic studies conducted also provided an opportunity to identify other surface components in *B. fragilis* that might bind fibrinogen in the absence of BF1705.

Houston et al. (2010) demonstrated the ability of *B. fragilis* whole cells to bind human fibrinogen with the identification of a putative fibrinogen-binding protein (BF-FBP) encoded by BF1705 in NCTC 9343, by virtue of its homology to fibrinogen binding antigen BspA of *Tannerella forsythia*. The gene encoding the protein was also identified in the genomes of *B. fragilis* strains 638R and YCH46 isolated from abscess and bacteraemia patients (Patrick et al., 2010).

Fibrinogen binding might interfere with the process of fibrin-mediated blood clotting and abscess formation, thereby promoting the survival of *B. fragilis* in abscesses and dissemination of infections. Interaction of fibrinogen with the coagulase bound on the cell surface of non-encapsulated *Staphylococcus aureus* causes clumping and has been reported to block phagocytosis (Kapral, 1966). Clumping of cells facilitated by fibrinogen might have a protective effect on the bacterium in early abscess formation and during later stages of persistent infections (Dominiecki and Weiss, 1999). Many fibrinogen-binding MSCRAMMs have been identified on the surface of *S. aureus*. One of them, Clumping factor A (ClfA) which binds within the C terminus of fibrinogen γ -chain mediates platelet aggregation and has been postulated as a virulence determinant in animal models of septic arthritis and infective endocarditis (Ganesh et al., 2008; McDevitt et al., 1997). The importance of ClfA-fibrinogen interactions in assessing early host survival in conditions of acute septicemia were

demonstrated by an impeded virulence state exhibited by mice expressing mutant forms of fibrinogen lacking the motif recognised by ClfA (Flick et al., 2013). More recent studies have identified an additional binding protein in *S. aureus*, the extracellular fibrinogen binding protein (Efb) which expresses a distinctively higher affinity for fibrinogen than the previously characterised MSCRAMMs. Fibrinogen binding by Efb blocks the interaction of fibrinogen with the leucocyte integrin receptor $\alpha_M\beta_2$, thereby evading an inflammatory immune response and promoting bacterial virulence (Ko et al., 2011). Additionally, a model has been proposed in *Francisella tularensis*, an airborne pathogen, in which the bacterium indirectly recruits fibrinogen to the outer membrane to escape complement-mediated membrane attack complex (Jones et al., 2012). Therefore the specific interactions of *B. fragilis* with the host fibrinogen identified by Houston et al. (2010) is of significance and might be related to the higher frequency of isolation of this obligate anaerobe in clinical samples of soft tissue abscesses, peritonitis and bacteraemia.

In the present study, it was observed that *B. fragilis* harbours additional components on the cell surface that might be involved in fibrinogen binding in the absence of BF1705, since Δ BF1705 cells retained the ability to bind fibrinogen in IFM analysis (Figure 3.13). Markerless gene deletions were generated by a method adopted from Patrick et al. (2009) which involved development of the deletion constructs on a suicide plasmid in *E. coli* and integration of the construct into the chromosome of *B. fragilis* NCTC 9343 via conjugal transfer followed by resolution of the partial diploids (Figure 3.4). The difficulties associated with the successful genetic manipulation of *Bacteroides* spp. have already been reported (Smith, 1995a). Introduction of foreign DNA into *B. fragilis* using methods such as electroporation are cumbersome owing to the protective R-M systems operating in the bacterium. Three type I and two type III R-M systems have been identified in the genome of the strain NCTC 9343. One of the *hsds* genes, encoding a subunit of the type I R-M system, which dictates DNA binding specificity of methyl transferase and endonuclease, is subjected to a shufflon-mediated site-specific rearrangement of the target recognition domains. The shufflon contains four sets of inverted repeats

resulting in eight different possible polypeptide combinations, producing an equal number of DNA recognition specificities for the HsdS subunit. Moreover, the NCTC 9343 genome encodes proteins sharing homology to type IV restriction endonuclease McrBC which cleaves specific DNA sequences modified with 5-methylcytosine (Patrick et al., 2010). The successful deletion of BF1705 in the present study was hence confirmed by PCR and immunoblot analysis (Figure 3.12 and Figure 3.21). The observed binding ability of Δ BF1705 strain is not surprising as many pathogens mediate interactions with the host through multiple microbial components, as observed in *S. aureus*. The possible ability of within-strain phase and antigenically variable micro-capsule (MC) polysaccharides (electron dense layer) present on the outer membrane of *B. fragilis* to bind fibrinogen was analysed by generating Δ BF1705 Δ BF1708 strains (Figure 3.14).

The genome sequence of three strains of *B. fragilis* identified 28 separate divergent polysaccharide (PS) loci each potentially dedicated to the expression of a distinct polysaccharide, of which only two were conserved between two strains (Patrick et al., 2010). Variable MC expression in the NCTC 9343 genome is assigned to 8 (A-I) out of the 10 identified PS loci and 7 of them are controlled by site-specific inversion of promoter sequences (Patrick et al., 2003). Micro-capsules form a marginal electron-dense layer of ~ 35 nm in size outwith the outer membrane and are invisible by light microscopy (Patrick et al., 1986). Sodium periodate-sensitive adhesion of *B. fragilis* to human colon carcinoma cells, Caco-2, identified the involvement of capsular polysaccharides in cellular adherence. Also, the strongest adherence to polarized Caco-2 cells was exhibited by the strain with the highest percentage of capsulated cells (Ferreira et al., 2002). Micro-capsules have been associated with the colonization of the gastrointestinal tract since a stable acapsular mutant which was deficient in synthesising all eight micro-capsules failed to colonise the mouse intestine when compared to the wild-type *B. fragilis* (Coyne et al., 2008). In addition to being implicated in complement resistance, the MC specific to PS A/2 locus has been involved in abscess formation (Coyne et al., 2001; Reid and Patrick, 1984). Apart from MCs, antigenically distinct within-strain variable large capsule and small

capsule that are visible by light microscopy are also expressed by *B. fragilis* (Patrick and Lutton, 1990). Large capsule has been reported to resist phagocytic uptake and killing of *B. fragilis* by PMNLs in vitro. The presence of common epitopes indicates co-expression of large capsule (LC) with the MCs (Lutton et al., 1991; Reid and Patrick, 1984). These adherent and evasive functions associated with the capsular polysaccharides urged us to analyse the role of capsules in fibrinogen binding (Patrick and Lutton, 1990). BF1708 present at J/10 in the PSJ locus is conserved among all three sequenced strains and encodes a putative PS chain-length determining protein, the only Wzz protein homologue identified in *B. fragilis* (Patrick et al., 2010). Deletion of BF1708 in NCTC 9343 genome gave rise to a population of cells expressing a single microcapsule polysaccharide specific to PSD locus, instead of an antigenically mixed population identified by the different MC-specific monoclonal antibodies (Lutton et al., 1991; Patrick et al., 2009). The Δ BF1708 strain was also characterised by an absence of high -molecular-mass polysaccharides associated with the PSD gene cluster and an aberrant capsular phenotype in ~ 10% of LC cells (Patrick et al., 2009). Fibrinogen binding by the Δ BF1705 Δ BF1708 double gene deletion mutants in the present study indicates the participation of other adhesins, apart from the putative BF-FBP and MC polysaccharide layer in interactions with fibrinogen (Figure 3.15). However, since the Δ BF1708 strain still expresses large capsule, small capsule and might also harbour undetectable amounts of short-chain repeats of the MCs synthesised by other loci, the contribution of these polysaccharides to the observed fibrinogen binding cannot be ruled out. At the same time, the fibrinogen binding by the Δ BF1708 strain could be regarded as a consequence of the absent electron-dense layer which enhances the ability of other surface components involved in fibrinogen interaction. Exposure of other cell surface adhesins in the absence of capsular exopolymer expression has been reported to mediate interaction of *B. fragilis* with non-polarised Caco-2 cell surfaces (Ferreira et al., 2002). A similar mechanism was observed in the encapsulated Smith strain of *S. aureus* which prevented the interaction of bound coagulase with fibrinogen, when compared to the non-capsulated strain which

blocked uptake by phagocytes as a result of fibrinogen-mediated clumping (Kapral, 1966). In a similar manner, the inactivation of a gene encoding a fibronectin-binding protein in *B. fragilis* was observed to enhance its fibronectin-binding ability and was more adherent in biofilms when compared to the wild-type (Pauer et al., 2013). Therefore, the role of the remaining capsules as well as other surface adhesins in the observed fibrinogen binding by NCTC 9343 whole cells needs to be further investigated.

Nonetheless, the specific involvement of the BF1705 protein in binding fibrinogen was observed when cell-free culture supernatants of the wild-type and mutant strains were concentrated through membranes with a 100 kDa cut off and used in far-western analysis. Presence of BF1705 in the concentrated wild-type supernatant was evident from the detection of a band close to the position of the purified BF1705 protein on cross-reaction with anti-1705 polyclonal antibody in immunoblots (Figure 3.21). Absence of the band in concentrated culture supernatants of Δ BF1705 strain was consistent with the deletion of the protein-encoding gene in the same. These findings indicate the potential presence of BF1705 surface protein in OMV released by *B. fragilis*. A similar method of concentrated sample preparation and immuno analysis has previously suggested the presence of a ubiquitin homologue, BfUbb, in *B. fragilis* OMV (Patrick et al., 2011). OMVs are secretory vesicles predominant in Gram-negative bacteria and are conventionally surrounded by an outer membrane-derived bilayer membrane ranging between 25 and 200 nm in diameter (Kulp and Kuehn, 2010). These extracellular complexes function to deliver bacterial components to host cells and tissues as well as other bacteria (Kuehn and Kesty, 2005). Recent transmission electron microscopy (TEM) and mass-spectrometry (MS) studies on the structure and proteome of OMVs harvested from *B. fragilis* reveal the presence of outer membrane and periplasmic proteins in 30-80 nm sized vesicles. It was observed that proteins in OMV possess either a lipoprotein signal or a transmembrane domain which indicate that they are destined to the outer membrane before being included in OMVs (Elhenawy et al., 2014). Albeit the study did not detect the presence of BF1705 in the OMV proteome, sequence analysis of BF1705

revealed a signal sequence as well as a membrane lipoprotein lipid attachment site, thus indicating the expression of a putative cell-surface lipoprotein (Houston et al., 2010). These findings corroborate our assumptions that BF1705 is released into the intestinal milieu within secretory vesicles. Moreover, a recent unpublished study has identified BF1705 protein in the concentrated culture supernatant of wild-type 9343 by liquid chromatography-MS/MS technique (M. Kowal, pers. comm.). OMVs have been implicated in spread of infections owing to the transport of toxins and virulence factors produced by pathogenic bacteria (Bomberger et al., 2009). Therefore, the presence of BF1705 in OMV might enhance the virulence activity of *B. fragilis* by being able to interact with fibrinogen in extraintestinal abscesses.

A smeared band, with a distinct intensity close to the size of the BF1705 holoprotein as well as the matured protein confirmed binding of fibrinogen by concentrated wild-type 9343 supernatants (Figure 3.22, Figure 3.23 and Figure 3.24). The Δ BF1705 strain did not cross-react with anti-fibrinogen antibody in far-western analysis which confirms the specific role of BF1705 protein in fibrinogen binding. However, the presence of smeared multiple-intense bands for the wild-type far-immune blots raises questions on the mechanism of BF1705 secretion and fibrinogen binding. Smeared bands have often been associated with glycosylated proteins (Patrick et al., 2011). Since the triplet codons characteristic of consensus recognition site for *B. fragilis* O-glycosylation are absent in the BF1705 amino acid sequence, the possible involvement of this post-translation modification in the observed smearing could be ruled out (Fletcher et al., 2011). However, BF1705 possesses a lipid attachment site and the attached lipids might account for the observed smeared bands. There are two explanations plausible for the multiple-banding pattern in the wild-type concentrated supernatant. One of them would be with reference to the adopted sample preparation method. The conventional technique for extraction of pure OMVs involves further ultra centrifugation or chemical treatment of the concentrated cell-free supernatants (Bauman and Kuehn, 2006). The unprocessed cell-free supernatants might contain variable proportion of other components besides vesicles which might hinder the purity of OMV preparations. Concentrated culture supernatants represent the outer

membrane protein and secreted proteins in addition to OMV. Since the samples in the present study were restricted to crude concentrated supernatants, the BF1705 holoprotein might have been pulled down in the concentrated supernatant sample along with the processed protein, present on the cell surface. The detection of holoproteins that possess the signal peptide sequence in cell-free supernatant is rare since signal peptidases cleave the leader sequence once the protein is transported to its final destination. Thus, the holoprotein is generally detected only in the whole cell extract in most cases (Patrick et al., 2011). Proteins of the inner membrane and cytoplasmic components have not been reported in *B. fragilis* vesicles till date, although rare cases have been studied in an antarctic psychrotolerant bacterium, *Shewanella vesiculosa* M₇^T (Pérez-Cruz et al., 2013). Investigations by cryo-TEM identified a complex type of double bilayered OMVs characterised by the presence of plasma membrane, cytoplasmic proteins as well as entrapped DNA which constituted ~ 0.1% of the total vesicles released by *Shewanella vesiculosa* M₇^T. The presence of BF1705 full-length protein in the concentrated supernatant and its interaction with fibrinogen suggests that fibrinogen can bind to BF1705 both in the processed and unprocessed forms of the BF1705 protein. A similar double-band pattern was observed earlier in far-western analysis of collagen binding in OMP fraction of *B. fragilis*, wherein both glycosylated and non-glycosylated versions of the Cbp 1 protein bound collagen (Galvão et al., 2014). Another possibility is that the mature BF1705 protein might not be present in the pure OMV extracts after all, which might have prevented its detection from the reported OMV proteome studies. It might either be present as a surface exposed lipoprotein or follow a different secretory mechanism which allows the detection of full-length protein before signal cleavage. A recent analysis of outer membrane proteome and secretome of *B. fragilis* detected a multitude of secretion mechanisms operating in the bacterium. The study identified a large number of lipoproteins both on the cell surface as well as among the secreted proteins. Analysis of OMVs were excluded from the study which predicts an alternative fate for the BF1705 lipoprotein other than OMV-inclusion. Mass spectrometry predicted 108 out of 229 putative OMPs as lipoproteins. Most of

them were of unknown function and were suggested to be either peripherally associated with OM, bound to OMPs or embedded in OM by a novel mechanism independent of a β barrel structure. Apart from two predicted lipoproteins, of which one was a putative adherence protein and the other, a homologue of TonB-dependent transporter (TBDT)-associated heme-binding protein, HmuY, the remaining six identified in the secretome could not be assigned a specific function. Proteins of the secretome that contain a signal sequence are targeted to ABC transporter/TolC based export system following an initial periplasmic translocation mediated by Sec pathway (Wilson et al., 2015). These proteomic studies provide evidence of an OMV-independent existence of BF1705 protein that led to the detection of the full length protein in the sample preparation. The second explanation for the presence of multiple bands in the wild-type concentrated culture supernatant could be the presence of other fibrinogen-binding proteins in addition to BF1705 concurrent with the IFM results. However, this possibility could be ruled out since the mutant did not show binding to fibrinogen in far-western analysis. The lower molecular bands might suggest the presence of BF1705 degradation products that bind fibrinogen or even fibrinogen degradation products that cross-react with the anti-fibrinogen antibody. In the studies conducted by Elhenawy et al. (2014), a significant proportion of proteins that were exclusive to the OMV harvested from concentrated supernatants were identified as hydrolytic proteases. The retention of proteolytic activity by at least some of these proteases were detected by gelatin zymography. These intra-vesicular active proteases might have contributed to fibrinogen degradation in far-western analysis. Moreover, fibrinogenolytic activity has already been reported as an extracellular phenomenon in *B. fragiilis* using culture supernatants in fibrinogen zymography (Houston et al., 2010). Studies in *P. gingivalis* have identified a 150 kDa fibrinogen binding protein on its cell surface which is partially associated with temperature-mediated fibrinogen degradation in combination with another 120 kDa degrading component. Experimental evidence suggests that the binding of fibrinogen to the cell surface is followed by its proteolytic degradation into discrete peptides by *P. gingivalis*. It was observed that the degradation was initiated by fibrinogen

binding on the bacterial surface (Lantz et al., 1991a). The probable presence of a similar BF1705-mediated fibrinogenolytic activity in *B. fragilis* was investigated. Fibrinogen binding by BF1705 was independent of fibrinogen degradation since bioinformatic analysis of BF1705 mentions no obvious catalytic domain (Figure 3.27).

A relatively high concentration of purified BF1705 protein was required to detect fibrinogen binding in far-western blots. A dose-dependent manner of fibrinogen binding by BF1705 was also observed in the studies by Houston et al. (2010). The dependence on concentration suggests that higher amounts of BF1705 protein are required for effective binding to fibrinogen. A previous study had revealed that BF1705 interacted strongly with the B β -chain of fibrinogen in addition to weaker interactions with the A α - and γ -chains. Recent studies on the interaction of Efb protein harboured by *S. aureus* with fibrinogen have detailed the use of surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) techniques to confirm the specific interaction of N-terminal half of Efb protein with the D fragment of fibrinogen (Fg) (Ko et al., 2011). Characterisation of Efb-Fg interaction by using Biacore 3000 system with specific sensor chips derived with either Fg-D or with his-tagged Efb revealed full length Efb and Efb-N and not Efb-C bound to Fg-D. Efb-N bound to Fg-D in a dose-dependent manner with a K_D of 0.23 nM based on kinetic data. The stability of the complex formed was revealed by a reasonably fast on rate and slow off rate. Reversal of the system identified that Fg-D bound to Efb-N chip in a dose-dependent manner with a higher K_D of 2.67 nM due to a faster off rate. ITC examined the binding affinity of 2 GST tagged Fg-binding sites in Efb, namely Efb-A and Efb-O using plasmin-generated Fg-D fragments. Affinity of Efb-O for Fg-D was found to be 200 times higher than that of Efb-A with dissociation constants of 4.66 nM and 1.0 μ M, respectively. Hence, further work involving point mutation studies and surface plasmon resonance could be carried out to identify the domain-specific interaction of BF1705 with fibrinogen as well as determine the binding affinity.

Since fibrinogen is not present in the gastrointestinal tract, *B. fragilis* does not encounter this glycoprotein in its commensal niche. *B. fragilis* becomes exposed to fibrinogen only when this opportunistic pathogen is accidentally released from its normal habitat into the peritoneal cavity and encounters an infected site. Fibrinogen is recruited at the site of infection to initiate blood clotting and to form fibrin-mediated abscesses (Herrick et al., 1999). During this exposure, the binding of fibrinogen by *B. fragilis* might delay the process of defence and promote infection dissemination. However, the absence of fibrinogen in the bacterium's natural niche suggests one or more additional functions for the BF1705 protein in *B. fragilis*. Since the sequence analysis identified homology to cell surface proteins in *Tannerella forsythia* (32.88% identity in 295 aa), *B. thetaiotaomicron* (33.33% identity in 411 aa) and *Trichomonas vaginalis* (29.29% identity in 297 aa) as well as Pfam match entries to LRRs which facilitate protein-protein interactions, the BF1705 protein might perform an adherent function (Kobe and Kajava, 2001). Earlier studies have reported the increase in frequency of *B. fragilis* isolates from 4% in faeces to 42% in colonic mucosa cultures (Namavar et al., 1989). The mucosal layer comprises the viscous stratum that separates epithelial cells from the gut lumen and poses as a crucial barrier against pathogenic infection. Since only a small proportion of the endogenous intestinal microbiota can penetrate mucus and interact with the epithelial surface, the predominance of *B. fragilis* as a member of the adherent colonic microbiota when compared to the other *Bacteroides* spp. might have implications on the high prevalence of this species in gut associated infections (Swidsinski et al., 2005). Furthermore, invasion of mucus via mucin binding and degradation by *B. fragilis* to fulfil its nutritional purposes has been reported. Invasion of the mucosal layer confers epithelial attachment to *B. fragilis* (Huang et al., 2011). Frequent sloughing of epithelial cells might expose components of the extracellular matrix. Binding of *B. vulgatus*, the most common *Bacteroides* spp. of the faecal microbiota, to fibronectin has been regarded as a means of enhanced colonic establishment. The results presented here indicate a similar fibronectin-binding function in BF1705, albeit *B. fragilis* is frequently isolated from opportunistic

infections and binding to fibronectin might confer an invasive role to *B. fragilis* (Figure 3.28). It has been reported that fibrinogen expression is upregulated during an inflammatory response and is deposited in the ECM in a fibrillar form where fibrinogen colocalises with fibronectin fibrils (Pereira et al., 2002). Fibronectin possesses a fibrin-binding site and is codeposited at the site of injury to form a clot and protect the underlying tissue. Therefore it can be suggested that during opportunistic infections, the fibronectin-binding ability of *B. fragilis* undergoes transition to fibrinogen binding to persist in and spread infections. Binding and degradation of fibronectin by *P. gingivalis*, which is genetically close to *B. fragilis*, seemed to be mediated by the same surface protein that bound and degraded fibrinogen (Lantz et al., 1991b). A bifunctional protein, FnbpA, which binds fibrinogen in addition to fibronectin has also been identified in *Staphylococcus aureus* (Wann et al., 2000). Moreover, a TonB-dependent protein (encoded by *bf1991*) of 102 kDa in size had been identified as a fibronectin binding protein in *B. fragilis* strain 1405 (Pauer et al., 2009). However, this protein was not detected in the present study possibly due to use of a different strain. Also, BF1705 does not share significant homology to *S. aureus* FnbpA or the TonB-dependent protein encoded by *bf1991* (BF0466) in *B. fragilis*. Nonetheless, an ~32 kDa band was observed in the fibronectin far-western analysis in the present study which seems to be consistent with the band observed in the *bf1991* gene-inactivated strain (Pauer et al., 2013). Alternatively, it could be hypothesised that fibronectin possesses a greater binding affinity to a BF1705 degradation product at ~32 kDa. Binding affinity studies need to be conducted to investigate this hypothesis. The concentrated supernatant of Δ BF1705 failed to cross-react with anti-fibronectin antibody, therefore the fibronectin binding function of BF1705 within the gastrointestinal tract might be modified to fibrinogen binding in extraintestinal infections. The possibility of other outer membrane proteins adhering to fibrinogen, as observed in IFM results, and fibronectin cannot be ruled out since far-western blots can detect the functional role of only those proteins that can renature after SDS treatment (Lantz et al., 1991a). However, this study signifies the role of BF1705 protein both in the commensal as

well as pathogenic phase of *B. fragilis*. Future in vivo experiments using mouse models are essential to elucidate the intra- and extraintestinal functions of the BF1705 protein encoded by *B. fragilis*.

Chapter 4 Degradation of Fibrinogen by *Bacteroides fragilis*

4.1. Introduction

Adhesion of bacterial surface proteins to host epithelium and extracellular matrix, as reviewed in the previous chapter, might be succeeded by effective interactions to invade and colonise the host for the spread of opportunistic infections. Proteolytic degradation of host tissue is a common pathogenic interaction that can spread infections and trigger intense immune responses. The ability of the laminin-binding protein in *B. fragilis* to bind and activate plasminogen into a serine protease that degrades host matrix tissue has been documented (Ferreira et al., 2013). In a similar manner, pathogens can recruit as well as secrete a wide range of proteases and interfere with the host matrix. A brief outline of the effect of proteolytic interactions of pathogens on host is provided along with the proteolytic ability of *B. fragilis* and the consequences on host health.

4.1.1. Proteases

Proteases, also known as peptidases, are enzymes that catalyse the cleavage of peptide bonds in other proteins with which they interact. Proteases can act as positive and negative effectors of biological processes. The functions of proteases can be classified broadly as catalysts of protein degradation and specifically as selective agents that control physiological processes (Neurath, 1984). The significance of the different families of proteases is evident from the fact that 2-4% of genes in the human genome comprises the degradome (genes involved in protein degradation), whose gene products are vital for diverse functions (Puente et al., 2005). Enzymatic activity is also associated with the human microbiome, the microbes residing within humans, which secrete proteases involved in biological processes such as metabolism, development and virulence.

4.1.1.1. Classification

Proteases are multi-domain proteins with catalytic activity restricted to a single structural domain. The multi-domain nature results in diversity in action and

structure, despite having a specific enzymatic function (Carroll and Maharshak, 2013). A broad method of classification is where proteases are grouped as endopeptidases or exopeptidases depending on the cleavage location. Exopeptidases cleave peptide bond proximal to amino- or carboxy- termini of substrates. Endopeptidases cleave peptide bonds distant from termini of their substrates. Further classification outlines 5 distinct groups of proteases on the basis of the chemical nature of groups responsible for their catalytic activity namely aspartic, cysteine/thiol, metallo-, serine and unidentified proteases (Barrett et al., 2003). Additionally, Rawlings & Barrett (1993) developed a comprehensive method of classifying groups of enzymes on the basis of type of reaction they catalyse, chemical nature of catalytic site and evolutionary structure. This method of grouping is thus based on a hierarchical system of classification levels with 84 different proteases families being classified initially. The subsequent massive in-flow of information in the form of amino-acid sequence data and 3D pattern structures demanded an update in the classification system (Carroll and Maharshak, 2013). The MEROPS database thus came into existence and provides additional information regarding protein inhibitors of peptidases, small-molecule inhibitors and a collection of known protease cleavage sites and substrates (Rawlings & Barrett 1999; Rawlings 2004; Rawlings & Morton 2008; Rawlings 2009).

4.1.2. Microbial proteases

Microbes synthesise a vast array of aspartic, cysteine, metallo- and serine proteases which might be cell-associated or extracellular in action. Microbial aspartic proteases are specific for aromatic or bulky amino acid residues on both sides of a peptide bond and are broadly divided into pepsin- and rennin-like enzymes. Cysteine proteases are generally active in the presence of reducing agents and have been implicated in numerous virulence and inflammatory responses (Oido-Mori et al., 2001). Metallo-proteases require a divalent metal ion for activity and are classified into neutral and alkaline groups based on specificity of action (Rawlings and Barrett,

1995). Serine proteases are characterised by a serine group at the active site and possess a broad substrate specificity (Róka et al., 2007).

4.1.2.1. Proteases in the Intestinal Tract

Of all the organ systems in the human body, the gastrointestinal tract contains the highest levels of endogenous and exogenous proteases. Intestinal epithelium performs digestive, absorptive and secretory functions and transmits information to mucosal, immune, vascular and nervous system. Proteases in the gastrointestinal tract can be luminal, circulating, secreted, intracellular, intramembrane or pericellular enzymes (Antalis et al., 2007). The majority of proteases are synthesised as inactive precursors, known as zymogens, which are then activated through proteolytic cleavage. Linderstrom-Lang coined the term ‘limited proteolysis’ to recognise proteases elaborating intestinal epithelial cell (IEC) function through signalling processes (Neurath, 1999). IEC proteases regulate the intestinal environment via ECM remodelling (Medina and Radomski, 2006).

The gut microbiota synthesises substantial amounts of serine, cysteine and metallo-proteases. Macfarlane et al., (1986) examined the degradation of casein and BSA by human faecal-slurries to produce trichloroacetic acid (TCA)-soluble peptides, ammonia and volatile fatty acids over a 96h period and confirmed the prominence of proteolysis in the large intestine. *Bacteroides* spp., and *Propionibacterium* spp. were identified as the predominant proteolytic bacteria in faecal samples followed by lesser proportions belonging to the genera *Streptococcus*, *Clostridium*, *Bacillus* and *Staphylococcus*. Protein breakdown can be regarded as the first step in protein utilisation by gut commensals where large oligopeptides are progressively degraded into smaller peptides and amino acids by the combined activity of microbial and pancreatic proteolytic enzymes, which are then assimilated directly into microbial protein or fermented with the production of ammonia and volatile fatty acids.

Uncontrolled proteolysis, during events like premature activation of zymogens and intestinal dysbiosis by opportunistic infections, can lead to tissue invasion and destruction in the gastrointestinal tract. Tissue destruction causes disruption of the

intestinal barrier, thereby exposing the enteric immune system to luminal antigens (Carroll and Maharshak, 2013). Activation of intestinal macrophages which are the first line of defence, results in inflammatory responses in the gastrointestinal tract. Inflammatory bowel disease such as colitis is caused by dysregulated immune responses towards microbial antigens in genetically predisposed hosts and during conditions of intestinal barrier dysfunction. Adherent and invasive *E. coli* (AIEC), an enteric opportunistic pathogen has been implicated in gelatin degradation and biofilm formation, which has been observed in IBD. Biofilm formation in AIEC involves environmental stress-upregulation of genes encoding proteases by σ E-mediated upregulation of genes followed by proteolytic processing of bacterial aggregation proteins to promote intercellular aggregation (Martinez-Medina et al. 2009; Chassaing & Darfeuille-Michaud 2013).

4.1.2.2. Microbial proteases degrading fibrin and fibronectin

The ability of bacteria to degrade specific host proteins is a significant virulence factor, especially with regard to structural and defense proteins, such as fibrin and fibronectin. In the event of injury, accumulation of fibrin protects surrounding tissue from microbial invasion. High levels of fibrinogen, the precursor of fibrin, have been often detected in the blood of patients suffering from infectious diseases. Therefore, fibrinogen degradation products might delay fibrin polymerisation and prolong bleeding in infections (Wikstrom et al., 1983).

Oral bacterial species isolated from infections ranging from angular cheilitis and subgingival-related periodontal destructions were examined for fibrin and fibrinogen degrading ability. Degradation of fibrin and fibrinogen or one of them was detected in microbes isolated from all the sampling sources. Activity was predominant in strains belonging to *Actinomyces*, *Bacteroides/Porphyromonas*, *Fusobacterium*, *Peptococcus*, *Propionibacterium* and *Staphylococcus* spp. Enzyme specificity was observed and more strains were able to degrade fibrinogen than fibrin. Degradation of both the substrates were detected in *Clostridium*, *S. aureus* and *Streptococcus pyogenes* strains (Wikstrom et al., 1983). Further studies indicated that a

considerable proportion of these strains tested positive for fibronectinolytic activity too. Interaction with fibronectin is significant for initiation and progression of oral infectious processes. A few strains from the non-oral *Bacteroides* spp, *B. assacharolyticus* and *B. fragilis*, also revealed fibronectinolytic activity. Interestingly, fibronectin degradation was not observed in relation to cariogenic *Streptococci*, which exhibited fibronectin binding (Wikstrom and Linde, 1986).

Matsuka et al. (1999) studied *S. pyogenes*-associated fibrinogenolysis in detail and attributed the same to an extracellular cysteine protease known as streptopain or streptococcal pyrogenic exotoxin B (SpeB). The protease is expressed as an inactive 40 kDa precursor, which is converted into the 28kDa active enzyme by the autocatalytic removal of NH₂-terminal propeptide. SpeB-mediated fibrinogenolysis degraded the A α -chain of fibrinogen and the degradation was specific to the COOH-terminal region as analysed by western-blotting. In addition to fibrinogen, SpeB has been implicated in the cleavage of matrix proteins, such as fibronectin and vitronectin, resulting in loss of host tissue integrity. *S. pyogenes* expresses proteins that bind fibrinogen with fibrinogen-binding M1 protein being the most virulent (Esgleas et al., 2005; Macheboeuf et al., 2011). Therefore initial host surface attachment might be succeeded by the action of degrading enzymes that invade and colonise the host tissues, as seen in invasive diseases like STSS and necrotizing fasciitis caused by *S. pyogenes*. Furthermore, the contribution of SpeB protease to the bacterial virulence is evident by the production of defensive anti-streptococcal protease antibody in humans which readily inhibits SpeB mediated proteolysis.

Lantz et al. (1991) reported a coupled activity of fibrinogen binding, followed by degradation mediated by a 150kDa protein localised on the cell surface of *P. gingivalis* W12, a common periodontal pathogen. The cell-associated proteolytic activity was observed to occur at temperatures of 22°C and 37°C with the release of distinct fibrinogenolytic fragments of 97kDa and 50 kDa. Zymography analysis using substrate-containing gels identified two major fibrinogen degrading components at 120 kDa and 150 kDa, respectively. The dependence of

fibrinogenolysis on reducing agents and protease inhibitor studies indicated that the degrading components were thiol-dependent proteases that exhibited trypsin-like substrate specificity. Further studies suggested the involvement of the same surface component in similar interactions with fibronectin (Lantz et al., 1991b). Competitive inhibition studies identified unlabelled fibrinogen as a more potent inhibitor of labelled fibronectin binding, than unlabelled fibronectin. Fibrinogen was also observed to displace cell-associated fibronectin and its degradation products from *P.gingivalis*. However, the detection of fibronectinolytic activity at a relatively low temperature of 4°C, confirms fibronectin as a better substrate for *P. gingivalis* proteases than fibrinogen across varying temperatures. Periodontitis is associated with loss of connective tissue and bone from dental root areas, resulting in inflamed soft tissue pockets. Bacteria colonise these pockets and are submerged in gingival crevicular fluid containing plasma and tissue proteins. Thus, the complex adherence and degradation events displayed by the cell surface proteins allow *P. gingivalis* to attach and detach from inflamed tissue pockets and interfere with tissue repair. In addition to the mentioned surface component, *P. gingivalis* protease activity in cells, vesicles and spent medium have been shown to degrade collagen, C3, C5 and C5a receptors. 85-95% of the total proteolytic activity was attributed to three proteinases, PrtP, Rgp-1 and Rgp-2 of the cysteine protease family, encoded by 3 related genes (Barkocy-Gallagher et al., 1999).

4.1.3. Enzymatic activities of *Bacteroides* spp.

4.1.3.1. Glycosyl hydrolases

In a commensal interaction, *B. thetaiotaomicron* and *B. fragilis* have been identified to encode 172 and 89 glycosyl hydrolases, respectively. These enzymes in combination with starch binding proteins have been involved in binding and utilisation of dietary and host-derived polysaccharides (Xu and Gordon, 2003). However, the non-commensal proteolytic interactions of the genus *Bacteroides* have been of interest since the 1970s when the production of extracellular histolytic enzymes by its members were detected (Rudek and Haque, 1976). Hyaluronidase and

chondroitin sulphatase activities were detected in all the subspecies of *B. fragilis* tested. Elastase was associated with *B. coagulans*. It was hypothesised that these histolytic enzymes could have a potential invasive role in the infectious process of this genus through the degradation of the extracellular matrix.

Later on, Riepe et al. (1980) reported the destruction of human brush border sucrase and maltase activity on exposure to *B. fragilis* elastase-like serine proteases. Diminished disaccharidase activity caused brush border damage and altered the surface membrane of mucosal cells, as observed in blind loop syndrome associated with bacterial over growth in humans. The glycoprotein degrading ability of *B. fragilis* was examined to assess their role in mucin degradation (Macfarlane and Gibson, 1991). Mucins cover the mucosal surface of the large intestine and offer a slippery gel-like texture which prevent pathogenic attachment besides maintaining colonic health and integrity. The *B. fragilis* NCDO2217 strain produced cell-associated hydrolytic enzymes such as neuraminidase, fucosidase, N-acetylgalactosaminidase, β -galactosidase and N-acetylglucosaminidase which were able to degrade carbohydrate moieties of mucin glycoproteins. Due to the hydrolytic nature of the organism, they are able to grow on diverse substrates by inducing differential enzymatic production.

Guzmán et al. (1990) observed 24 out of 50 clinical samples of *B. fragilis* exhibiting a neuraminidase (NA) mediated attachment to mammalian epithelial cells. The adhesion was epithelial cell-specific and did not mediate attachment to PMNLs. The bacterium produced cell-bound neuraminidase which exposed a galactoside residue present on a mammalian cell receptor, thereby promoting adhesion of *B. fragilis*. Neuraminidase-dependent adhesion of epithelial cells to *B. fragilis* was competitively inhibited by galactosidase treatment which confirmed the significance of the residue in the process. A pili-independent increase in haemagglutination titre was also observed with RBCs pre-treated with NA, which coincided with an increased adherence efficiency to NA treated epithelial cells. Neuraminidase removes sialic acid present in the termini of cell membrane-associated glycoproteins

and glycolipids, thereby altering cellular integrity and inter-cellular relationships (Godoy et al., 1993). Bacterial neuraminidase activity might progressively strengthen a capsule/pili mediated bacterial adherence to host epithelial cells. As protease treatment was found to reduce adhesion, the adhesin was identified as an outer membrane protein that extends across the capsular layer, when present (Guzmán et al., 1990).

4.1.3.2. Protease activities

Protease production by *B. fragilis* has been compared in continuous and batch cultures in multiple studies to determine the effect of nutrient limitation. Gibson & Macfarlane (1988) studied the strain NCDO2217 of *B. fragilis* and reported cell-bound proteolytic activity which was eventually released from the bacterial cells during stationary phase. Therefore the bacterium might contribute to gut-associated extracellular proteolysis as well as that concerning self. Hydrolysis of casein, trypsin, chymotrypsin, azocasein and arylamidase activity against leucine p-nitroanilide (LPNA), leucine β -naphthylamide, glycyl-proline p-nitroanilide (GPRPNA) and valyl-alanine p-nitroanilide (VAPNA) were exhibited by the bacterium. There was no hydrolysis reported with respect to globular proteins. *B. fragilis* NCDO2217 could utilise an array of nitrogen sources for its growth, such as ammonia, casein and peptone. Growth on trypsin and chymotrypsin was also observed, which describe the bacterium's role in degrading pancreatic proteases. Collagenolytic activity was not detected in *B. fragilis* as the protein was degraded by pancreatic proteases, trypsin and chymotrypsin, leaving the substrate unavailable for proteases in the large intestine. The fastest growth was measured on ammonia and the usage of casein as substrate exhibited the highest protease activity. Cell-bound protease activity was observed to be growth rate dependent, especially in nitrogen-limited cultures. Slowing of growth in stationary phase marked the accumulation of protease activity in the medium, accompanied by a slight reduction in cell-bound protease activity. No such accumulation was observed in medium in continuous culture, although, ~ 25% of the total protease activity was detected to be extracellular at low growth rates in

continuous cultures. Thus, constitutive protease production was evidently influenced by nutrient availability and culture conditions. A similar influence was observed in protease synthesis and release in *B. fragilis* ATCC25285 grown in continuous and batch cultures (Macfarlane et al., 1992). Secretion of proteases in stationary phase was a distinct event and not associated with general lytic release of cytoplasmic contents. This was evident from the decline in the release of cell-bound marker enzymes β -galactosidase and alkaline phosphatase in stationary phase. In continuous culture, the ratio of intracellular: whole cell protease activity increased with growth rate and extracellular protease activity was detected in trace amounts at the lowest dilution rate, probably due to cell lysis. The secretory behaviour in continuous culture indicates that initiation of protease secretion requires a specific signal that is produced only at the end of active growth. The signal was suggested to be activated as a response to nitrogen depletion in stationary phase of batch cultures. The observed relationship between protease secretion to nutrient availability was correlated to in vivo conditions where *B. fragilis* synthesises high levels of cell-bound proteases during growth in the nutrient-rich cecum followed by release of enzymes in the nutrient limited distal colon. Thus it could be hypothesised that nitrogen limitation in the distal colon could signal the release of cell-bound proteases by *B. fragilis*. An increase in cell-associated protease production in exponential phase and subsequent release in stationary phase in batch cultures was also observed in *B. splanchnicus* NCTC10825, protease activities of which were similar to that of *B. fragilis* and differed only on the basis of inability to grow on sucrose and a few other disaccharides. Inhibition experiments in NCTC10825 indicated the presence of serine, thiol and possibly metallo-proteases, as was observed in *B. fragilis* (Macfarlane and Gibson, 1993). Further, the activity of three proteases, P1, P2 and P3 produced by *B. fragilis* NCDO2217 were examined in order to determine their substrate specificity, location and type. Cell fractionation demonstrated that P1 was located intracellularly and in the periplasm, whereas P2 and P3 were outer membrane-associated. Substrate specificities of the three enzymes were tested on azocasein and synthetic substrates, such as LPNA, VAPNA and GPRPNA. P1 was

characterised as an unusual thiol-group dependent serine exoprotease as it strongly hydrolysed VAPNA, GPRPNA and weakly hydrolysed azocasein, which was inhibited by phenylmethylsulfonyl fluoride (PMSF) and thimerosal a.k.a thiomersal in U.K. Since P2 and P3 hydrolysed azocasein and LPNA, they were regarded as endopeptidases, owing to the high protease : arylamidase ratios. Since P2 hydrolytic activity was inhibited by Ethylenediaminetetraacetic acid (EDTA) and P3 by thimerosal a.k.a thiomersal in U.K. and iodoacetate, they were recognised as metalloprotease and cysteine protease, respectively. Thus, the function of P2 and P3 proteases might be initiation of protein breakdown at the cell surface with P1 mediating further hydrolysis of peptides produced by outer membrane proteases (Gibson and Macfarlane, 1988b).

B. fragilis strains secreting an enterotoxin (BFT) belonging to the zinc metalloprotease family were implicated in diarrheal disease in animals. The secretion of enterotoxin is a unique feature of enterotoxigenic *B. fragilis* (ETBF) strains, which encode three isotypes of BFT on distinct *bft* loci present on a *B. fragilis* pathogenicity island (6kb) (Sears, 2001). Enterotoxin activity was first detected by Myers et al. (1984) using intestinal loop tests of *B. fragilis* strains isolated from faeces of 24-48h old lambs suffering from acute diarrhoea. Enterotoxigenic *B. fragilis* strains were first isolated from humans in 1987 and were later identified as a common causative agent of infant diarrhoea (Myers et al. 1987; Sack et al. 1994). Isolation of ETBF from sewage plants suggest that these strains could be members of the normal colonic microbiota (Shoop et al., 1990). *B. fragilis* strains lacking the enterotoxin gene (NTBF) failed to produce diarrhoea in rabbit models and there was no indication of bacteraemia, adherence or invasion of epithelial cells in the disease model (Myers et al., 1989). The extracellular enterotoxin was purified and characterised as a ~20 kDa single polypeptide (Van Tassell et al., 1992). Further analysis post cloning the gene encoding the polypeptide into *E. coli* identified a zinc binding motif which is a characteristic feature of the metzincin family of metalloproteases. Proteolytic activity at 37°C revealed autodigestion of enterotoxin and comigration of gelatinase activity, both of which were inhibited by metal

chelators thus confirming metalloprotease activity (Moncrief et al., 1995). In vivo, the secreted enterotoxin (fragilysin) induces the release of colonic epithelial cell surface proteins and cleavage of E-cadherin, the primary protein of zonula adherens epithelial tight junctions. Loss of intestinal tight junctions contributes to a reduction of the colonic barrier. Subsequent fluid accumulation results in ETBF-associated diarrhoea, loss of appetite and depression (Wu et al., 1998). However, more recent studies by the same group introduced a model by which the catalytic domain of BFT processes its receptor protein inducing cellular signal transduction that mediates the biological activity of BFT (Wu et al., 2006). Obiso et al. (1997) observed morphologic alterations and increased permeability of the paracellular barrier of HT-29 cells by enterotoxin. Enterotoxin-mediated intoxication of the paracellular barrier was independent of endocytosis and was also observed in kidney and lung epithelium by in vitro assays. The disruption of tight junctions on rat lung and canine kidney epithelium by enterotoxin suggest a possible role of the toxin in extraintestinal infections (Obiso et al., 1997). Studies have already identified the presence of ETBF in a larger proportion of *B. fragilis* isolated from human blood cultures suggesting a correlation to bacteraemia (Kato et al., 1996). Łuczak et al. (2001) detected the enterotoxin encoding gene in 12 out of 65 investigated *B. fragilis* strains isolated from extraintestinal clinical specimens from four countries.

4.1.3.3. Proteases induced by B. fragilis under different host conditions

B. fragilis was found to induce the release of host degrading proteases during its forage for nutrients, as observed in the bacterial requirement of heme for growth. Limitation of heme was a signal to induce protease expression for the release from heme-binding plasma proteins. Bacteriostatic conditions were observed immediately after the addition of hemopexin, a heme-binding plasma protein, which limited the availability of heme for *B. fragilis* growth. Eventually, production of hemopexin-degrading proteases by *B. fragilis* facilitated the release of heme from hemopexin. A heme-withholding mechanism contributes to host defence against infections by heme-requiring organisms. Pathogens circumvent this condition via heme-binding outer

membrane proteins. Black-pigmented members of the *B. fragilis* group degrade serum heme-carrier to provide free heme. Inhibition studies suggested that hemopexin degradation was caused by a serine protease in *B. fragilis* (Rocha et al., 2001).

Four homologues of gene encoding SpeB, a member of the C10 family of cysteine proteases have been identified in *B. fragilis* (*bfp1-bfp4*) and *B. thetaiotaomicron* (*btpA, btpB, btpC* and *btpZ*). *bfp1* and *bfp2* were present in all three strains studied whereas *bfp3* and *bfp4* were exclusive to the 638R strain and were present on mobile genetic elements (Thornton et al., 2010). C10 proteases are a group of papain-like cysteine proteases including Spe B and Interpain A produced by *Prevotella intermedia*. The contribution of these proteases to virulence has been studied extensively in SpeB-mediated cleavage of host matrix proteins, cytokines and release of kinin from kininogen. C3 degradation observed in Interpain A has also been associated with SpeB (Kagawa et al., 2000; Potempa et al., 2009). Therefore, the presence of C10 homologues in *Bacteroides* spp. could predict the ability of *B. fragilis* to disseminate infections. SpeB and Staphopain, a C47 cysteine protease expressed in staphylococci were transcriptionally coupled to genes encoding specific protease inhibitors, Spi and Staphostatin, respectively. A similar mechanism of coupling was observed in *bfp1* and *bfp4* encoded by *B. fragilis* and in *btpA, btpB* and *btpz* encoded by *B. thetaiotaomicron*. These inhibitory mechanisms could prevent bacterial cells from ectopic proteolytic damage. There was an increase in the expression levels of *bfp1* and *bfp4* encoding *B. fragilis* C10 proteases in response to oxygen. A similar inducible effect was observed in corresponding *B. thetaiotaomicron* C10 proteases as well, except *btpA* expression, which was stimulated in the presence of blood. The differential expression of these *Bacteroides* secreted proteases were observed only in extraintestinal habitats, such as blood and oxygen, whereas no induced expression was determined in caco-2 cell cultures. The observed differential expression with respect to the habitat contribute to its predicted role in spreading extraintestinal infections, as suggested in other members of C10 protease family (Thornton et al., 2012).

Hemolysins secreted by bacteria benefit pathogenesis by weakening the immune system and gaining access to nutrients. *B. fragilis* was considered as a non-hemolytic bacterium until Robertson et al. (2006) identified ten genes in *B. fragilis* which shared homology to other bacterial hemolysins. The identified hemolysins, HlyA - HlyI and HlyIII were subjected to functional analysis by cloning into a non-hemolytic *E. coli* strain. Subsequent analysis on blood agar plates detected hemolytic activity by the recombinant *E. coli*. HlyH was exclusive to the genome of *B. fragilis* NCTC 9343. A synergistic effect on hemolytic activity by HlyA and HlyB identified a two-component hemolysin. Differential regulation of HlyB and HlyA expression in the presence of oxygen and iron suggests that HlyBA is associated with *B. fragilis* pathogenicity in an oxygen- and iron- limited environment. This is in agreement with the observation that an oxygen limiting condition is required at the site of infection for the aetiology of extraintestinal anaerobic infection by *B. fragilis*. Further, in addition to a reported heme-degrading serine protease activity, *B. fragilis* might be able to overcome the host-imposed iron-limiting conditions through the regulation of HlyBA expression in iron depleted conditions. In vivo expression of hemolysins and role of HlyBA as a virulence factor was also addressed by Lobo et al. (2013) in an intra-abdominal infection model. Expression of eight hemolysin mRNAs were induced at different levels post-infection and there was a distinct reduction in the growth rate of a *hlyBA* mutant strain after 8 days following infection.

4.1.3.4. OMV-related enzymic activity

Macfarlane et al. (1992) had reported that ~10% of the extracellular proteases in cell-free culture supernatants of *B. fragilis* were associated with particulate fractions exhibiting high specific protease activities, reminiscent of *P. gingivalis* OMVs. Observations made in this study were strengthened by the work of Patrick et al. (1996) which provided evidence of haemagglutination and enzymic activity in relation to outer membrane vesicles, that were released from the *B. fragilis* outer membrane. OMV released by *P. gingivalis* had been implicated in virulence which includes enzyme production, cell toxicity, adherence and haemagglutination

(Mayrand and Grenier, 1989). Therefore, enzymic activity in terms of alkaline phosphatase, lipase, β galactosidase, glucosaminidase and glucosidase detected in OMV secreted by *B. fragilis* could possibly mediate tissue destruction by the degradation of host glycoproteins. Recently, a comparison profile of OM and OMV proteome in *B. fragilis* indicated a higher proportion of glycosidases and proteases in the OMV with more than 40 proteins being OMV-specific (Elhenawy et al., 2014). MS analysis identified a quarter of the proteins present in OMVs as glycosidases, peptidases, a phosphatase and a lipase. 80% of the OMV-specific proteins were acidic as indicated by isoelectric points. In contrast, OM proteome, which was more basic in nature, exhibited a higher proportion of TonB-dependent receptors and a lower abundance of hydrolases. The OMV proteome demonstrated proteolytic activity in zymograms against gelatin. Secretion of hydrolytic OMVs enhances the degradation of polysaccharides and glyco-conjugates which would contribute to microbial nutrition in the gut. The nutrients released by degradation could also help pathogenic gut-colonizers. A high level of digestive hydrolases detected in the OMV of bacterial predator, *Myxococcus xanthus*, were detected to have predatory effects and lytic activity against *E. coli*. Therefore, the *B. fragilis*-secreted hydrolytic OMVs could also operate against host tissue, such as mucus glycans, allowing unfavourable destruction and immune responses (Elhenawy et al., 2014).

4.1.3.5. Fibrinogen-degradation by *B. fragilis*

The ability of non-oral *B. fragilis* to degrade fibronectin was proposed in an earlier study (Wikstrom and Linde, 1986). Later, studies by Chen et al. (1995) purified and characterised a 100 kDa fibrinogen-degrading protease from *B. fragilis* strain YCH46. Inhibition studies revealed the purified enzyme to be a serine-thiol-like protease. In addition to fibrinogen, proteolytic activity was detected against casein, azocoll and gelatin. The protease was involved in the cleavage of A α -chain of fibrinogen, within a 10h time period, and repressed thrombin-mediated coagulation of fibrinogen. The crude extract of *B. fragilis* YCH46 was reported to cleave all three chains of fibrinogen within 24h. Since γ -chain hydrolysis was not found in the

purified protease, it was suggested that *B. fragilis* might encode other proteases with fibrinogenolytic activity. Further analysis by Houston et al. 2010 predicted the involvement of extracellular metalloproteases and a cysteine protease in fibrinogenolysis exhibited by strains NCTC 9343, 638R and YCH46, respectively. Analysis of exponential and stationary phase culture supernatants using zymograms identified fibrinogenolysis at ~ 45 and 50 kDa. Protease-specific activation buffers indicated metallo-proteolytic activity in fibrinogen degradation observed in NCTC 9343 and 638R. Comparison of fibrinogen degradation profiles of these strains and a further six clinical isolates showed that most of them initiated A α -chain degradation during exponential growth phase, followed by B β -chain hydrolysis. Hydrolysis of γ -chain hydrolysis was variable across the strains studied and commenced only in the stationary phase. However, MS analysis of proteins within the detected fibrinogenolytic range obtained from concentrated supernatants of the mid-exponential phase cultures of *B. fragilis* failed to identify any proteases predicted by the complete genome sequence annotation. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) activity, associated with virulence in certain pathogens, and the presence of lipoproteins was observed in MS analysis which might suggest the involvement of OMVs in the detected fibrinogenolysis (Houston et al., 2010).

B. fragilis, being the most common obligate anaerobe isolated from soft tissue abscesses and blood-related infections, might be able to prevent or delay the formation of the fibrin-mediated abscess wall, which normally curtails the spread of infections. The role of fibrinogenolytic proteases secreted by *B. fragilis* might accelerate infection dissemination by delaying the formation of fibrin walls, however, *B. fragilis* comes into contact with extraintestinal fibrinogen only when the intestinal barrier is disrupted. It has been suggested that the secretion of proteases might be triggered by the depletion of nutrients in the growth environment. Therefore, the secretory behaviour and regulation of proteases involved in fibrinogen degradation in nitrogen limiting conditions would open avenues to study the transition of this commensal into an opportunistic pathogen in extraintestinal nitrogen deprived conditions.

4.1.3 Aims

- To test fibrinogen degradation in *B. fragilis* NCTC 9343.
- To analyze the role of protease-encoding genes present in NCTC 9343 in fibrinogen degradation by zymography.
- To analyse the potential role of multiple proteases in fibrinogen degradation by generating multiple gene deletion mutants followed by their zymography.
- To express the extracellular NCTC 9343 proteases in *E. coli* and to investigate their specific role in fibrinogen degradation.

4.2. Fibrinogen degradation detected in *B. fragilis* strain NCTC 9343

Fibrinogen degradation was previously detected in *B. fragilis* NCTC 9343 grown at 37°C in supplemented Brain Heart Infusion medium containing 100 µg/ml plasminogen-free human fibrinogen. Detection of fibrinogenolysis by *B. fragilis*, subsequent to fibrinogen binding, might suggest a mechanism by which this opportunistic pathogen escapes from fibrin-mediated abscess walls in peritoneal infections, prevents clot formation and promotes blood-stream infections in humans (Houston et al., 2010; Patrick and Duerden, 2006; Patrick et al., 1995b). Partial A α -chain degradation was detected in 3h grown *B. fragilis* culture (Figure 4.1 A lane 1) followed by complete degradation of the A α -chain of fibrinogen after 48 h incubation, along with partial degradation of B β - and γ -chains (Figure 4.1 A lane 2). The A α -chain has been suggested to be an accessible target for proteolytic activity owing to its characteristic location in the fibrinogen molecule, where the chain protrude outwards via the D domain (Weisel, 2005). Fibrinogen was not degraded in BHI-S media even after a 48 h incubation period (Figure 4.1 A lane 4) indicating that the growth medium does not contribute to the observed degradation of fibrinogen. The ability of *B. fragilis* NCTC 9343 concentrated culture supernatant to degrade fibrinogen was studied using 0.1% fibrinogen SDS-zymography experiments. Supernatants of stationary phase cultures grown in BHI-S as well as glucose-DM were concentrated by centrifugation through a membrane with a 100 kDa molecular weight cut-off (MWCO) (2.7.2) to get a better representation of the OMV and associated proteins which might mediate interaction with the host proteins.

The concentrated culture supernatants were electrophoresed separately on SDS-PAGE gels containing 10% separating gel and on 0.1% fibrinogen SDS-zymogram as (2.7.3 and 2.7.5). Following protein refolding and incubation in serine protease and metalloprotease activation buffers, and visualisation, (2.7.5 and 2.7.4), fibrinogenolytic activity at ~45 and 50 kDa was detected in zymograms of BHI-S grown culture (Figure 4.1 B lane 1). However in glucose-DM grown culture, a zone of fibrinogenolysis was observed only at ~45 kDa (Figure 4.1 C lane 1). Activity was compared to negative control samples of concentrated BHI-S broth, in which no fibrinogenolytic bands were observed even after a 48h incubation in activation buffers (Figure 4.1 C lane 2). The results obtained were in agreement with fibrinogen degradation studies carried out by Houston et al. (2010) and confirmed that fibrinogen degradation is associated with non cell-related material since the culture supernatants exhibited fibrinogen degradation ability. However, a difference in fibrinogenolytic activity between the two media was not reported in the previous study. The observed difference suggests a variation in the mechanism of proteolysis adopted by *B. fragilis* grown in BHI-S and glucose-DM.

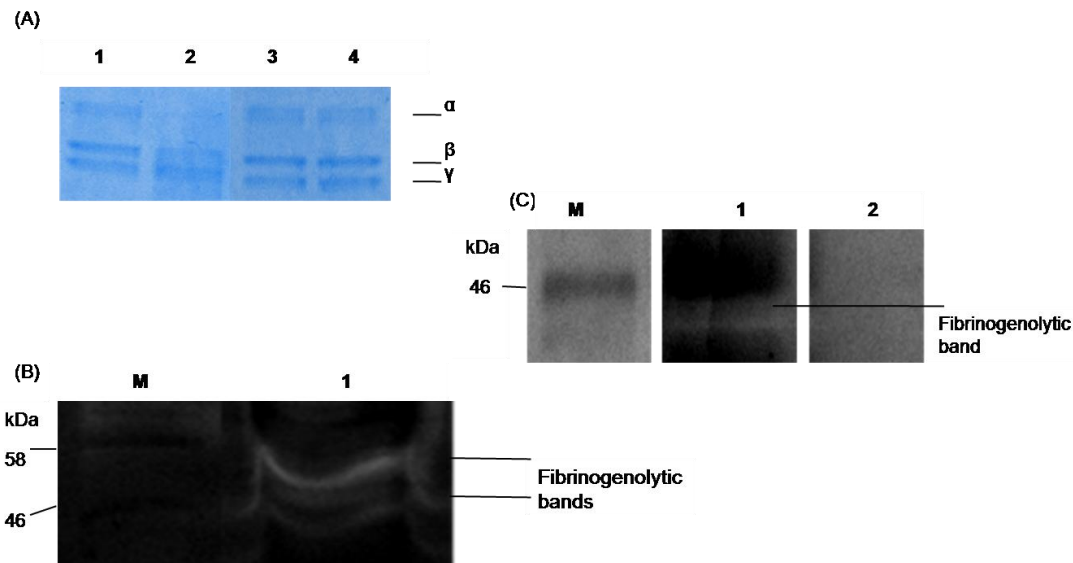


Figure 4.1: Detection of fibrinogen degradation by the wild-type 9343

(A) SDS-PAGE analysis of fibrinogen degradation by 3h (lane 1) and 48h (lane 2) grown cultures of wild-type 9343 in comparison with fibrinogen in BHI-S media as negative control at 3h (lane 3) and 48h (lane 4) incubation. Complete degradation of A α -chain and partial degradation of B β - and γ -chains were observed in 48h grown culture (lane 2). (B) 0.1% fibrinogen zymography of concentrated *B. fragilis* strain NCTC 9343 supernatant grown in BHI-S medium observed as two distinct bands of fibrinogenolysis (lane 1) (C) 0.1% fibrinogen zymography of concentrated *B. fragilis* strain NCTC 9343 supernatant grown in glucose-DM observed as a single band of fibrinogenolysis (lane 1) when compared to concentrated BHI-S medium as negative control (lane 2).

4.3. Generation of markerless protease-encoding gene deletion mutants in *B. fragilis*

Although fibrinogen zymography studies by Houston et al. (2010) suggested the involvement of metalloproteases in non-cell-associated fibrinogenolysis of *B. fragilis* strain NCTC 9343, MS analysis failed to reveal any specific proteases acting within the fibrinogenolytic range. However, studies conducted in other bacteria, such as *Pseudomonas aeruginosa*, *Streptococcus pyogenes*, *Staphylococcus aureus* and *Porphyromonas gingivalis* have confirmed fibrinogenolytic activity dependent on cysteine proteases and a metalloprotease (Ciborowski et al., 1994; Komori et al., 2001; Matsuka et al., 1999; Ohbayashi et al., 2011). Lantz, Allen, Vail, et al. (1991),

identified a 150 kDa cell surface component involved in fibrinogen binding and degradation by *P. gingivalis*, while Ciborowski et al. (1994) suggested that the two cell surface-associated fibrinolytic proteases at 150 and 120 kDa were conformational variants of a 180 kDa cysteine protease, porphypain.

To investigate the role of secreted proteases in fibrinogenolysis we used reverse genetics to delete genes encoding putative proteases. Four protease encoding genes were identified in the genome of *B. fragilis* NCTC 9343 using artemis genome browser, all of which encoded signal peptide sequences suggesting the expression of secreted proteases. The molecular-weight range of the identified extracellular proteases were ~ 45-50 kDa, which was within the range of observed fibrinolytic bands in the culture supernatant zymogram. The secreted proteases belonged to metallopeptidase and serine protease families and BLAST analyses revealed similarities to expressed proteases in other bacterial species (Table 4.1). Gene BF0275 encoded a metalloendopeptidase which was ~ 50 kDa in size and shared sequence homology with a putative peptidase BT3464 in *Bacteroides thetaiotaomicron* (80.77% identity in 437 aa) whereas BF0657 expressed a secreted metalloendopeptidase of ~ 37 kDa with sequence similarity to a putative aminopeptidase SCO4589 secreted by *Streptomyces coelicolor* (34.55% identity in 191 aa). Serine proteases of ~49 kDa encoded by BF1979 and BF3775 genes shared sequence homology with putative periplasmic protease BT0154 of *B. thetaiotaomicron* (43.86% identity in 440 aa) and putative secreted peptidase SCO7188 of *S. coelicolor* (28.53% identity in 431 aa). Expression studies conducted later have considered 27 other protease encoding genes containing signal peptide sequences in fibrinogen degradation (Table 4.2).

Table 4.1: Description of the selected protease encoding genes using artemis genome browser

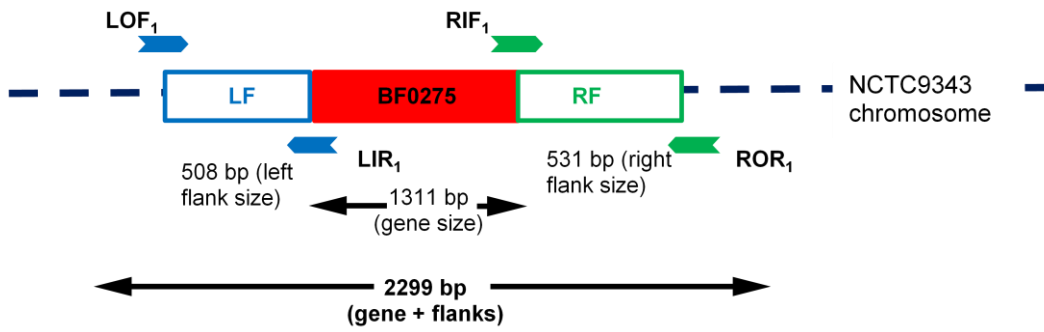
BF0275 (metalloendopeptidase), BF0657 (secreted metallopeptidase), BF1979 (periplasmic serine protease) and BF3775 (exported serine protease) present in NCTC 9343. The selected protein sequences shared identities with a putative peptidase from *B. thetaiotaomicron*, a putative aminopeptidase from *S. coelicolor*, a putative periplasmic protease from *B. thetaiotaomicron* and a putative secreted peptidase from *S. coelicolor*, respectively.

Feature Name	CDS	Size	Hit
BF0275 Metallo endopeptidase	304996...306306	1311 bp 437 aa 49.5947 kDa	<i>Bacteroides thetaiotaomicron</i> putative peptidase BT3464 (435 aa) E value : 1.4e-99 and 80.77% identity in 437 aa
BF0657 Secreted metalloendopeptidase	789188...790186	999 bp 333 aa 37.641 kDa	<i>Streptomyces coelicolor</i> putative aminopeptidase SCO4589 (324 aa) E value : 2.4e-08 and 34.55% identity in 191 aa
BF1979 Putative periplasmic serine protease	2312768...2314042	1275 bp 425 aa 47.6799 kDa	<i>Bacteroides thetaiotaomicron</i> putative periplasmic protease BT0154 (484 aa) E value : 1.1e-27 and 43.86% identity in 440 aa
BF3775 Exported serine protease	complement 4458153...4459505	1353 bp 451 aa 49.0703 kDa	<i>Streptomyces coelicolor</i> putative secreted peptidase SCO7188 (1239 aa) E value : 2.4e-07 and 28.53% identity in 431 aa

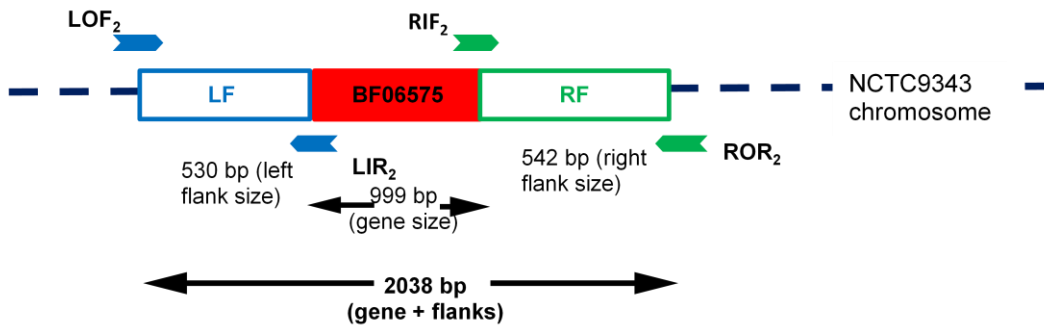
4.3.1. Primer design for amplifying the upstream and downstream regions of the selected protease-encoding genes

To amplify the upstream and downstream regions, four individual primer sets were designed for each gene using the Artemis genome browser and annotation tool and as explained in Chapter 3 (3.5.1). The left outside forward and left inside reverse primers were designed to amplify ~500 bp upstream of the desired gene whereas the right inside forward primer and right outside reverse primers were generated to amplify the region flanking the gene downstream (Figure 4.2 A, B, C and D). Primer sequences were in the range of 20-24 bases in length and melting temperature (T_m) values of the primer sequences were restricted to 60-66°C. Primers designed for the outside region of flanks (left outside and right outside primers) were attached with ~15 bp overlap from the vector sequence in order to ensure compatibility while cloning. To facilitate fusion PCR, the left inside and right inside primers were prefixed with reverse complements of primer sequences from each other (2.3).

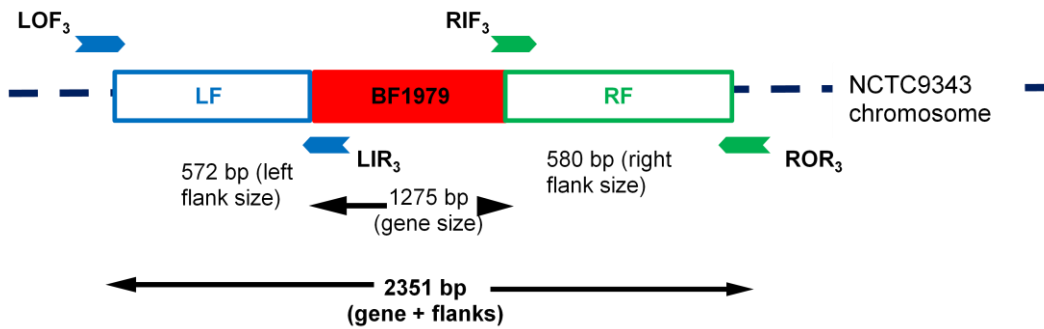
(A)



(B)



(C)



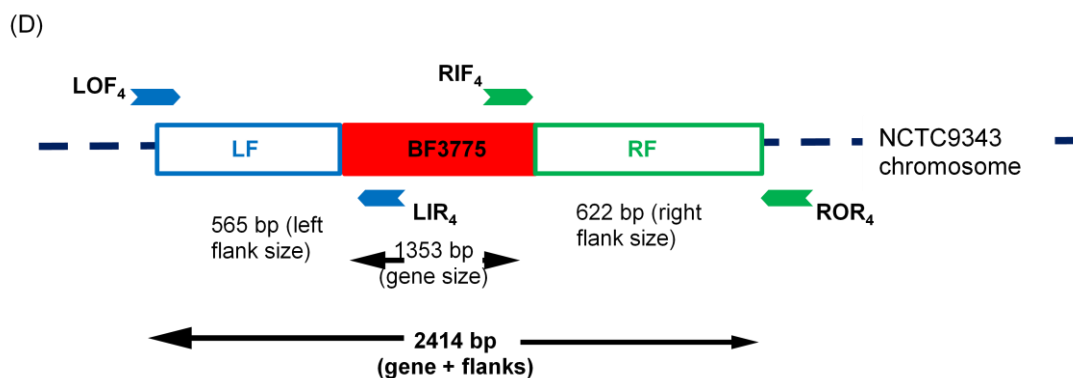


Figure 4.2: Graphic representation of the selected protease encoding genes and flanking regions

Graphic representation of (A) BF0275, (B) BF0657, (C) BF1979 and (D) BF3775 (red boxes) with left outside primers and left inside primers (LOF₁-LOF₄ and LIR₁-LIR₄) designed for amplifying the upstream regions of each gene (denoted in blue) and right inside and right outside primers (RIF₁-RIF₄ and ROR₁-ROR₄) designed for amplifying the downstream regions (denoted in green). Sizes of the genes, upstream and downstream amplicons are also indicated.

4.3.2. Generation of the deletion constructs

The deletion constructs for the protease-encoding genes were generated by amplifying the left upstream and right downstream flanks of each gene according to appropriate PCR conditions (2.5.3) using NCTC 9343 genomic DNA as the template. The amplified flanks of each gene were visualised after electrophoresis on 1% agarose gels and were in the range of 500-625 bp (Figure 4.3). The amplified flanks were then subjected to fusion PCR using left outside forward primer and right outside reverse primer to fuse the flanking sequences and generate 1000-1200 bp sized deletion constructs (Figure 4.4).

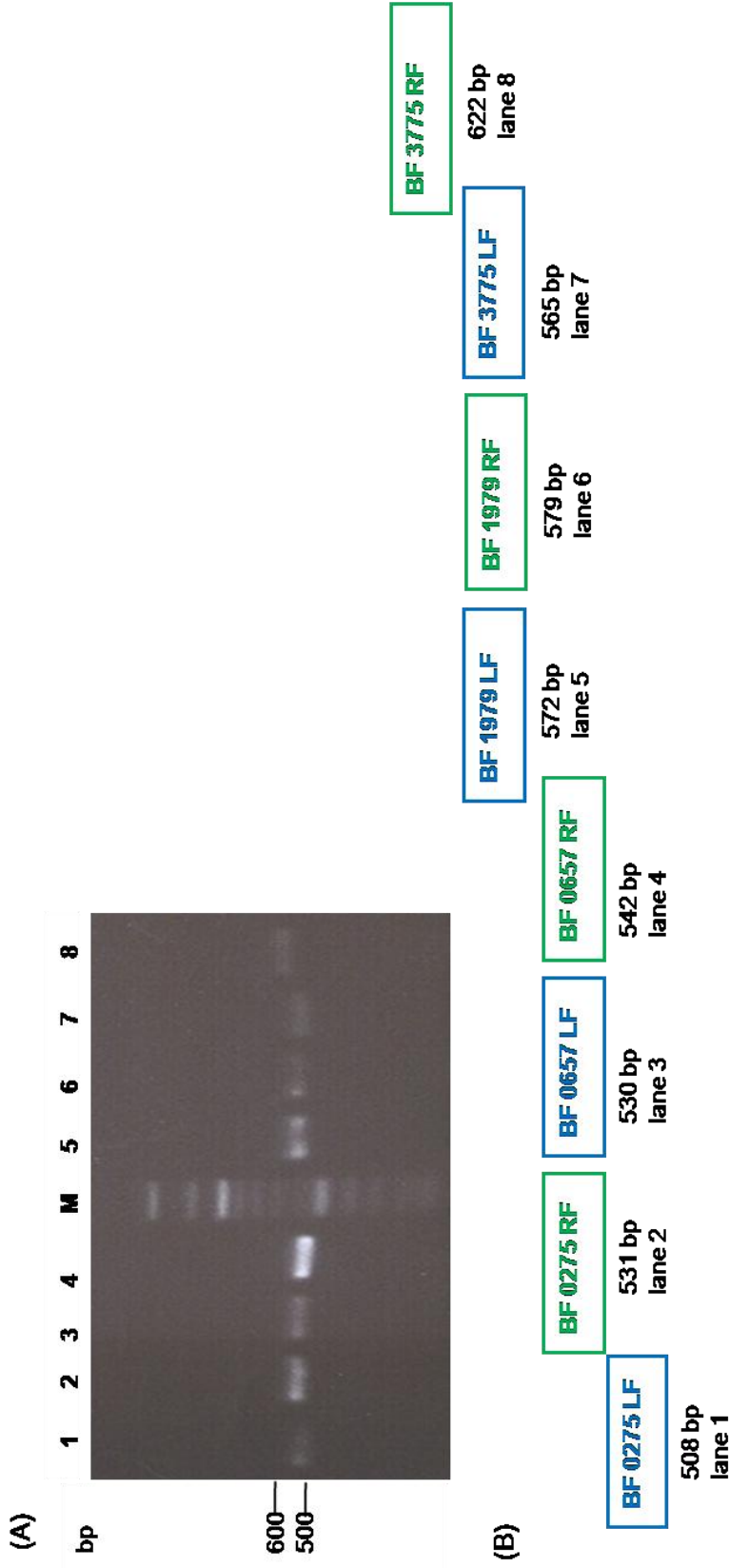


Figure 4.3: Amplification of flanking regions of the protease encoding genes

(A) Agarose gel analysis of PCR amplification of left and right flanks of BF0275 (lanes 1 and 2), BF0657 (lanes 3 and 4), BF1979 (lanes 5 and 6) and BF3775 (lanes 7 and 8) genes using LOF₁-LOF₄ and LIR₁-LIR₄ primers for left flanks and RIF₁-RIF₄ and ROR₁-ROR₄ primers for right flanks. Arrows and numbers on the left side indicate size of DNA in base pairs (bp) in a 100 bp ladder (M). The estimated sizes of the flanks were in the range of 500-625 bp in the gel. (B) Graphic representation of the expected sizes of the left flanks (denoted in blue) and right flanks (denoted in green) of the four genes.

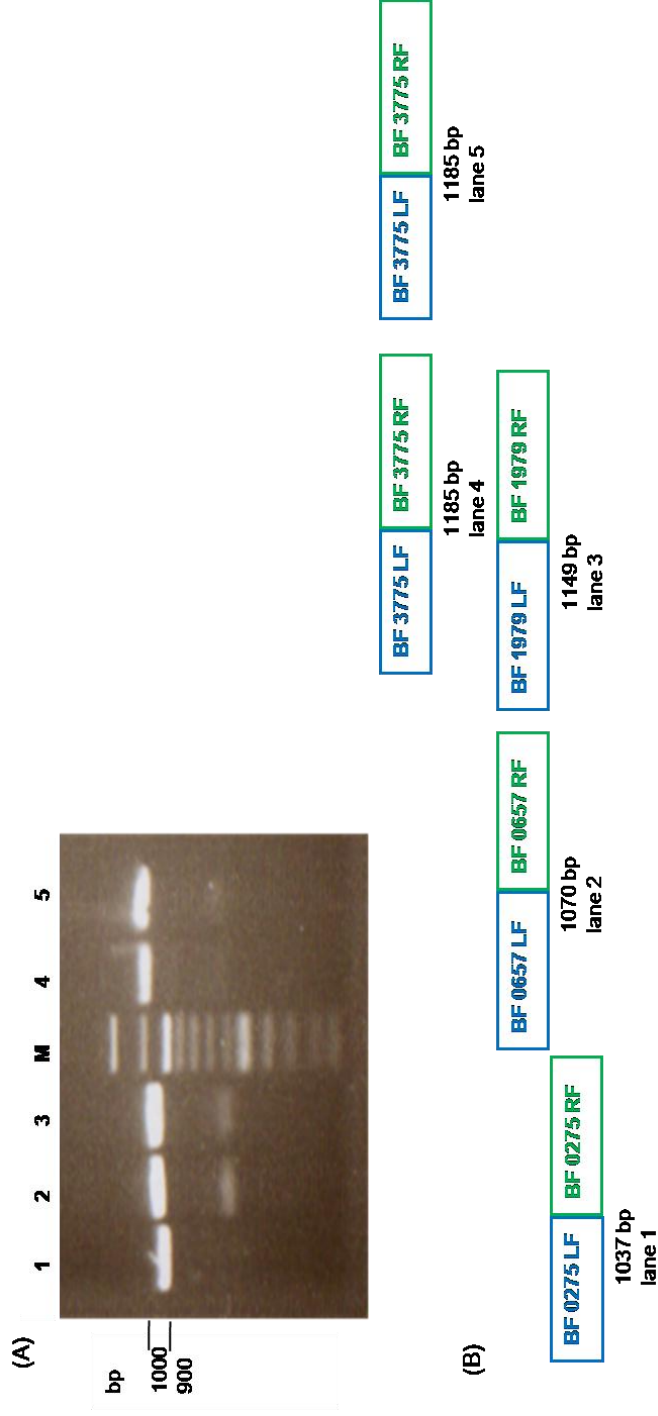
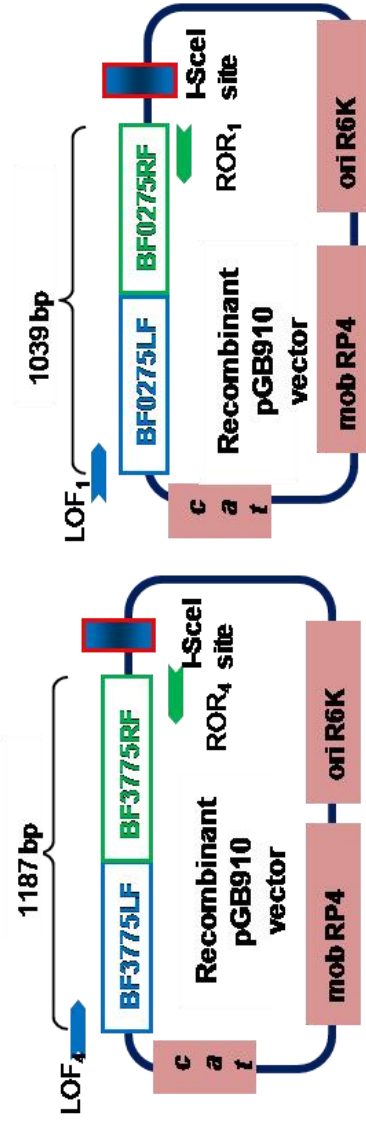
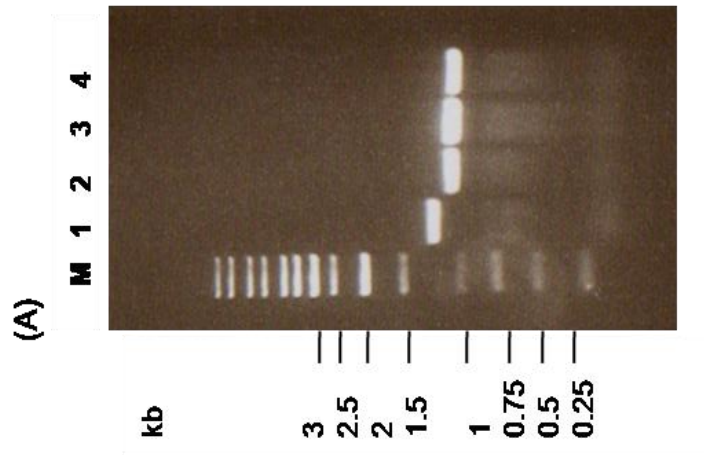


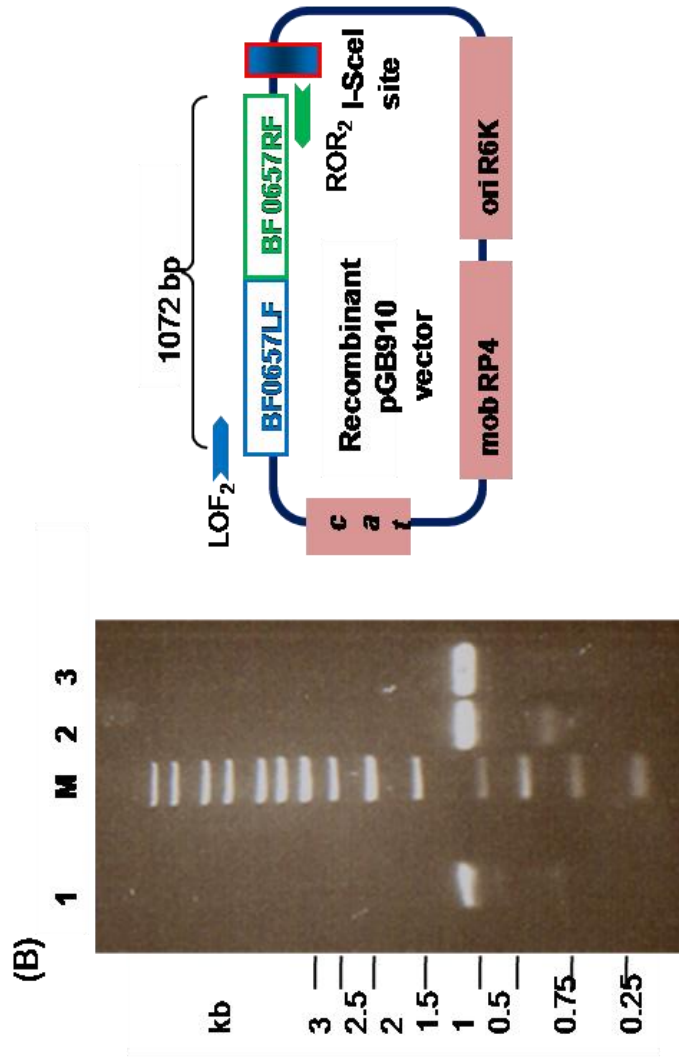
Figure 4.4: Amplification of deletion constructs of the protease encoding genes

(A) Agarose gel analysis of PCR amplification of deletion constructs generated by fusing together the left and right flanks of BF 0275 at ~1000 bp (lane 1), BF0657 at ~1000 bp (lane 2), BF1979 at ~1100 bp (lane 3) and BF3775 at ~1200 bp (lanes 3 and 4) using LOF₁-LOF₄ forward primers and RIF₁-RIF₄ reverse primers. Arrows and numbers on the left side indicate size of DNA in base pairs (bp) in a 100 bp ladder (M). The estimated sizes of the constructs were in the range of 1000-1200 bp. (B) Graphic representation of the deletion constructs of the four genes.

4.3.3. Cloning the deletion constructs into pGB910

The deletion constructs were cloned into the NotI restriction site in the multiple cloning site (MCS) region of linearised pGB910 by Infusion cloning, a ligation independent cloning method (2.5.8). The Infusion enzymes fuse DNA fragments, PCR generated sequences and linearised vectors, efficiently and precisely by recognising a 15 bp overlap at their ends. Infusion cloning reactions were set up by incubating each of the protease-encoding gene deletion constructs, linearised vector and the infusion enzymes (2.5.8). The pGB910 vector used for infusion cloning is an ~ 5300 bp RP4-based conjugative vector (3.5.3). In the present study, the vector was linearised by digesting with NotI (Figure 3.8). The infusion cloning mixture was used to transform *E.coli* S17-1 λ pir (2.4.3). The transformants were selected for chloramphenicol resistance on LB agar plates. Resistant colonies carrying the recombinant pGB910 vector were confirmed by PCR amplification of the deletion constructs (Figure 4.5 A, B and C) and Sanger-based sequencing (2.5.10).





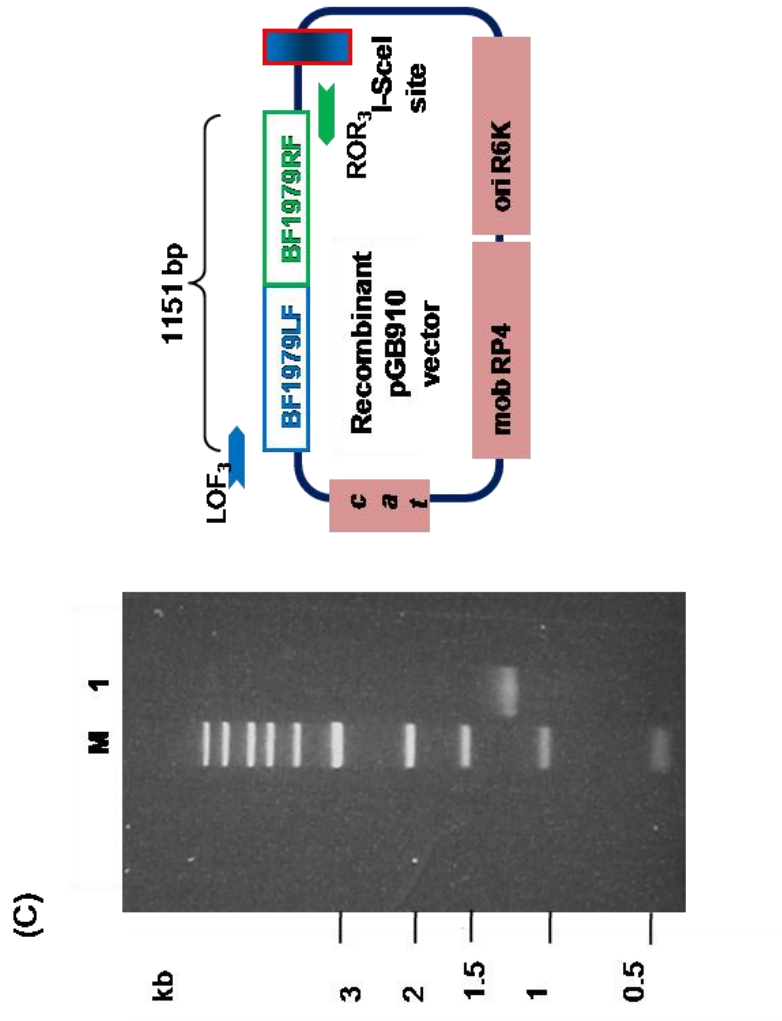
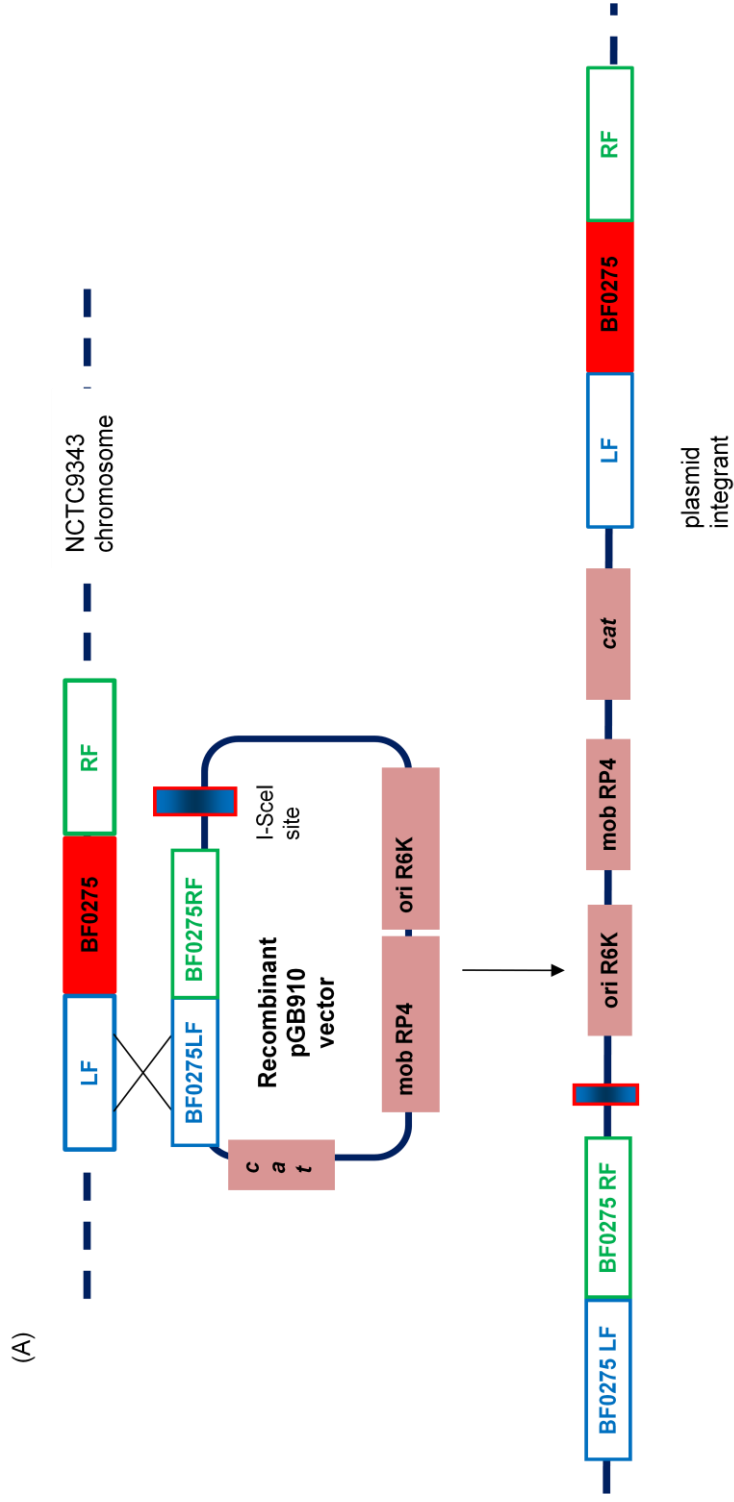


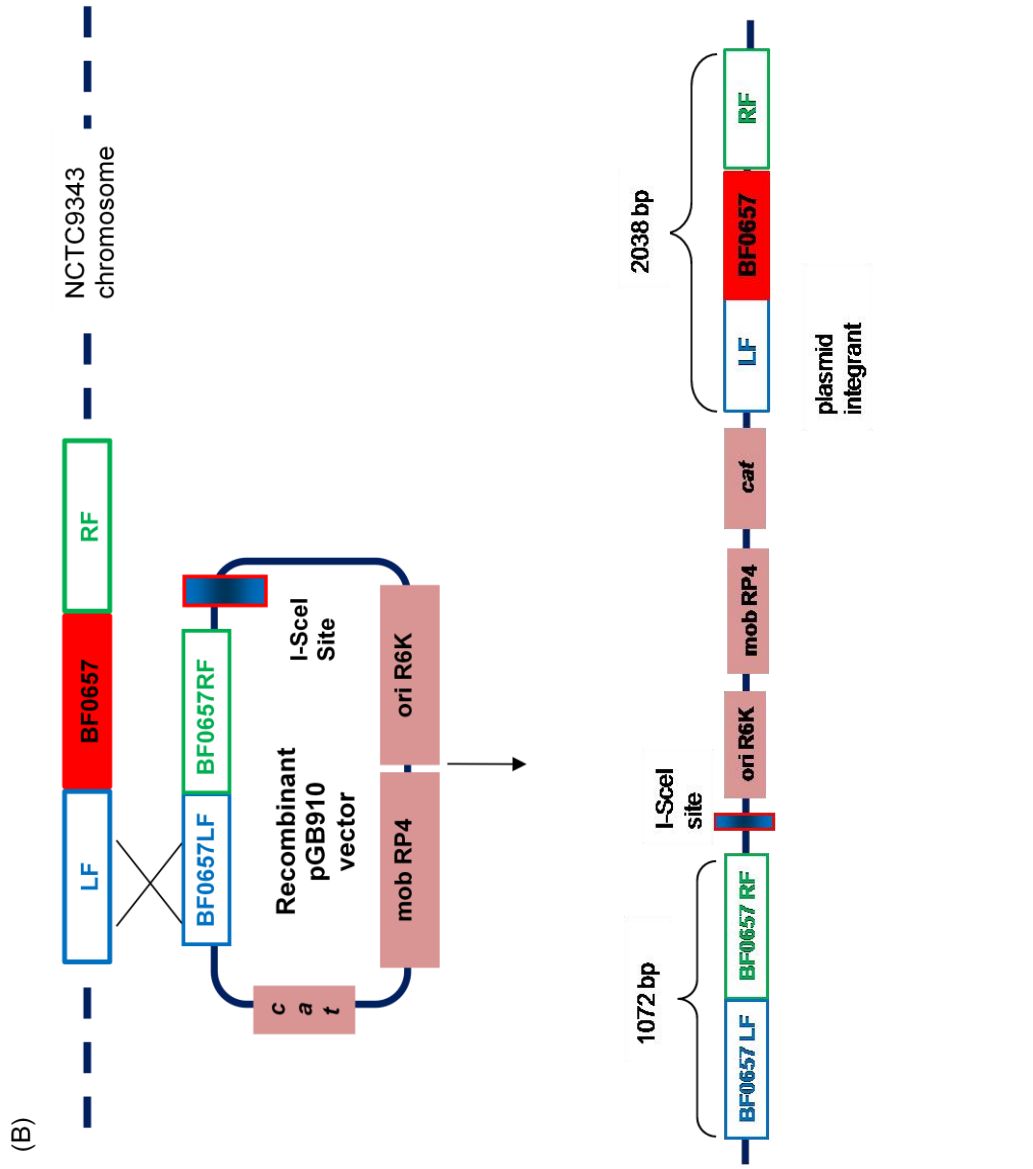
Figure 4.5: Confirmation of the deletion constructs present in the recombinant vector

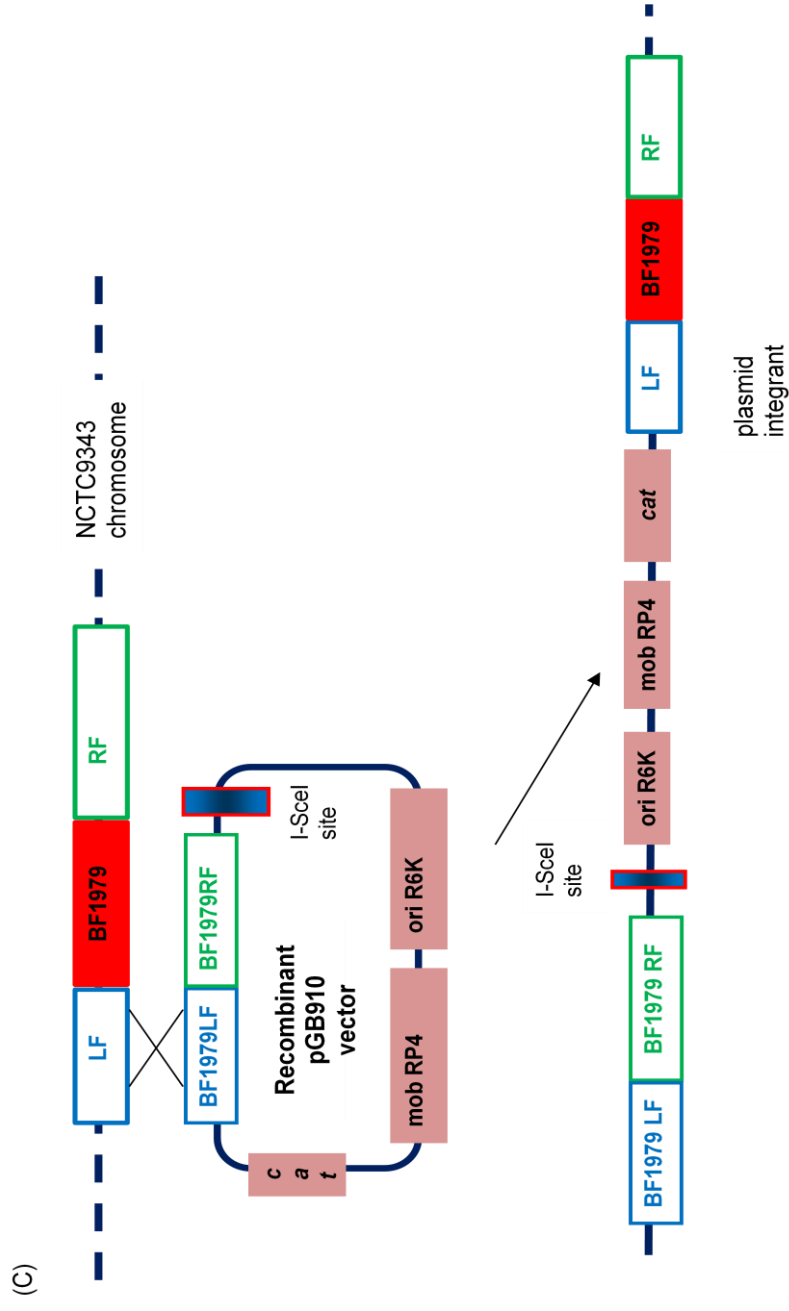
Agarose gel analysis of PCR amplifications of: (A) Δ BF3775 (lane 1) and Δ BF0275 (lanes 2, 3 and 4) (B) Δ BF0657 (lanes 1, 3 and 4) (C) Δ BF1979 (lane 1) present in recombinant pGB910 using specific left outside forward and right outside reverse primers (LOF₁-LOF₄ and ROR₁-ROR₄). Arrows and numbers on the left side indicate size of DNA in kilo bases (kb) in a 1 kb ladder (M). The estimated sizes of the constructs were in the range of 1000-1200 bp and matched with the expected sizes within the recombinant pGB910 shown in the graphic representation.

4.3.4. Conjugal transfer of gene deletion constructs into *B. fragilis* strain NCTC 9343

The recombinant vectors carrying the deletion constructs were mobilized from *E. coli* S17-1 λ pir into *B. fragilis* NCTC 9343 by conjugation (2.4.5). Post-conjugation, the strains were plated on BHI-S agar containing antibiotics, gentamicin sulphate to select against *E. coli* and erythromycin to select *B. fragilis* transconjugants carrying the recombinant pGB910 vector. The plated transconjugants were grown anaerobically for 2 days prior to confirming the presence of integrated plasmids in the erythromycin resistant colonies by PCR amplification of chromosomal DNA using left and right outside primers designed for the protease gene deletion constructs. Following conjugation, the *B. fragilis* strains carrying the integrated plasmid are partial diploids and therefore PCR amplification with left outside forward and right outside reverse primers of each protease encoding gene would result in amplicons corresponding to the size of the wild-type gene combined with the flanks as well as the size of the deletion construct (Figure 4.6 A, B, C and D). PCR amplification using left outside forward and right outside reverse primers, LOF₂ and ROR₂, specific to Δ BF0657 resulted in a band with an expected size of 2038 bp, corresponding to the size of the wild-type 0657 gene along with left and right flanks as well as a band at 1072 bp, which corresponded to the size of deletion construct (Figure 4.6 B lane 1). Likewise, the transconjugant partial diploid of BF3775 gene resulted in a band at 2414 bp and 1187 bp corresponding to the wild-type + flanks and the deletion construct, respectively on PCR amplification with LOF₄ and ROR₄ primers (Figure 4.6 D lane 1).







(D)

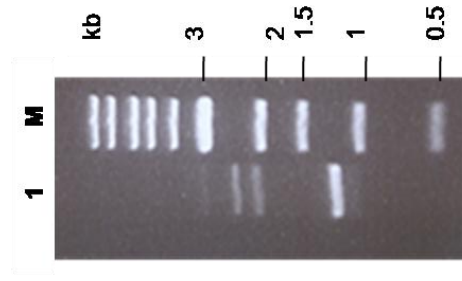
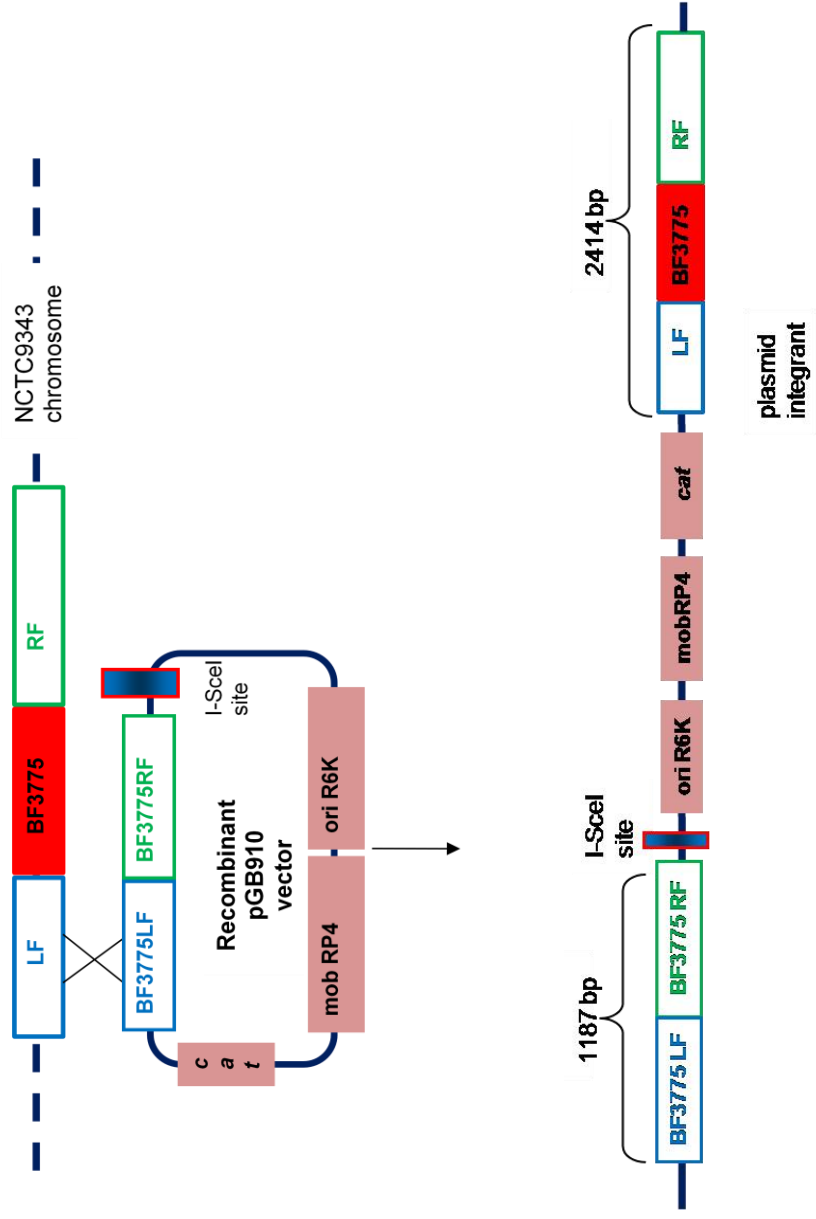
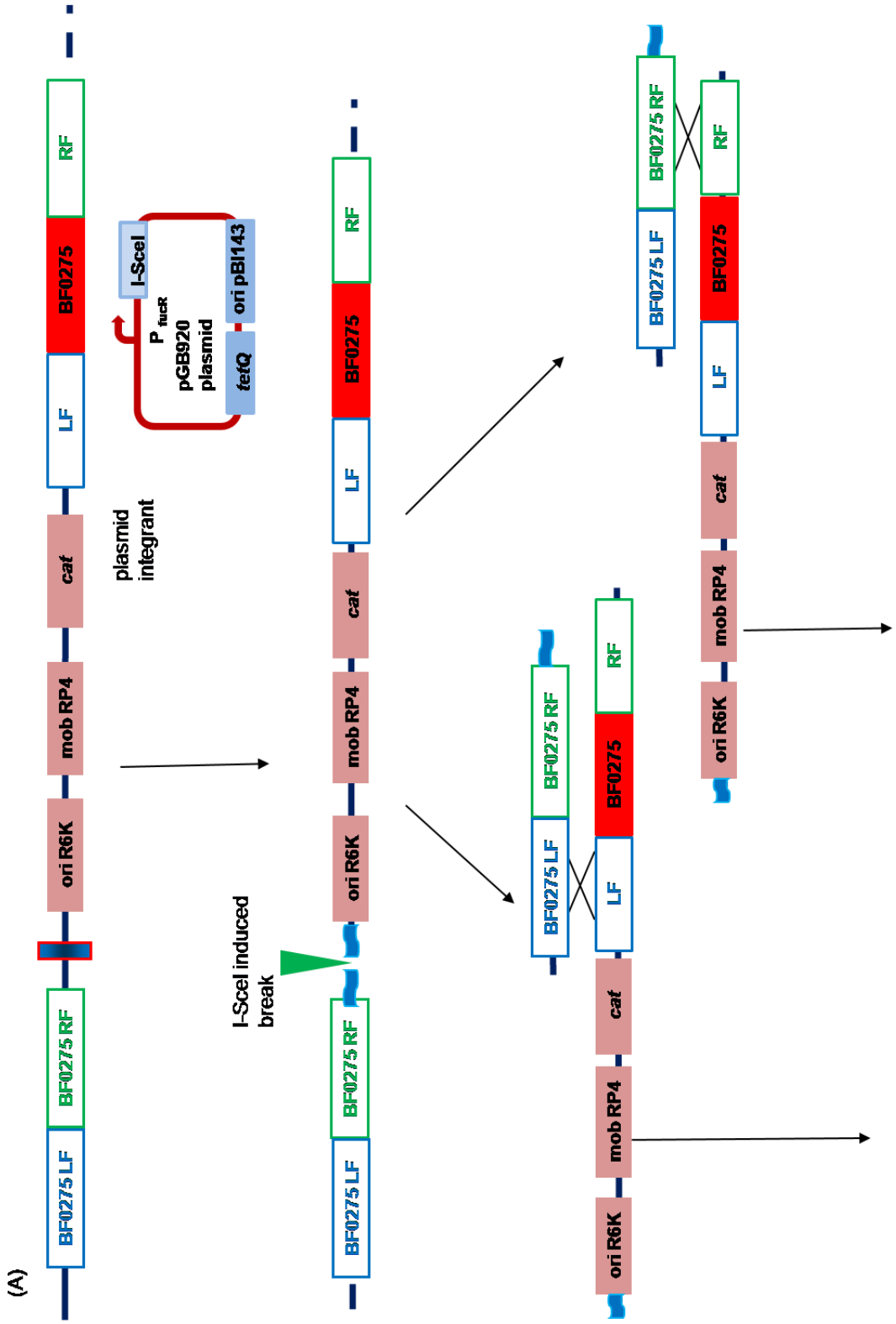


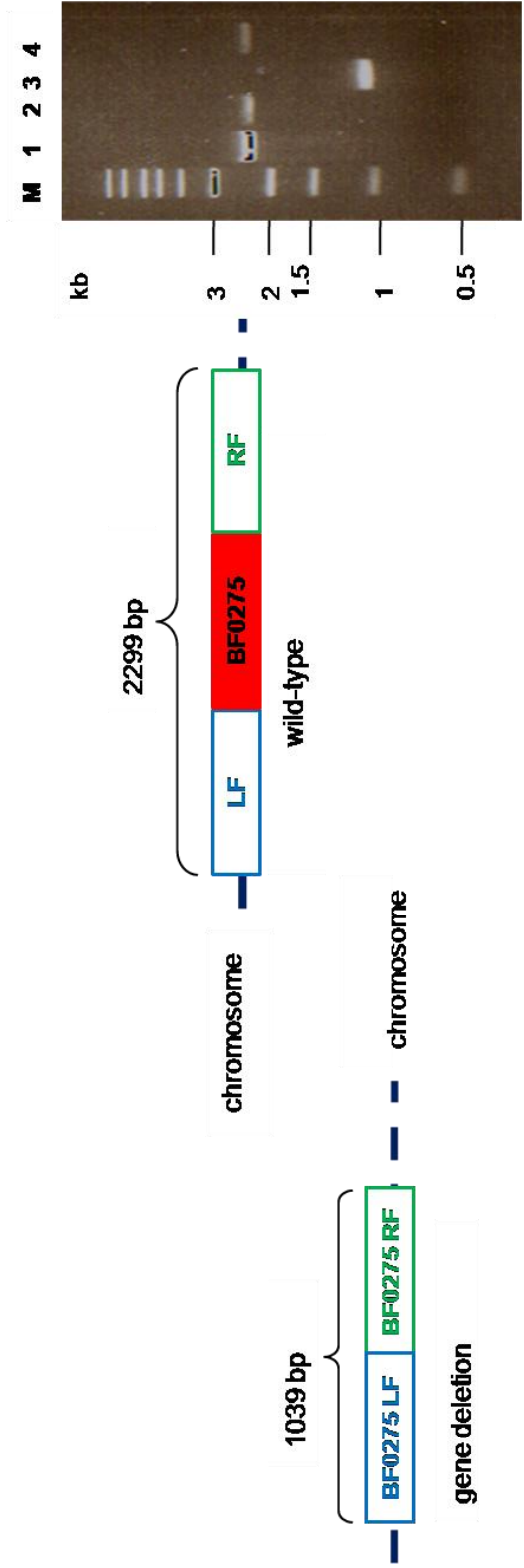
Figure 4.6: Amplification of protease encoding genes and flanking regions in the respective partial diploids

Arrows and numbers on the right side of the gel images indicate size of DNA in kilo bases (kb) in a 1 kb ladder (M). (A) Transfer of recombinant pGB910 carrying the BF0275 deletion construct into *B. fragilis*. (B) Transfer of recombinant pGB910 carrying the BF0657 deletion construct into *B. fragilis* and agarose gel analysis of PCR amplification of BF0657 partial diploid using primers LOF₂ and ROR₂. The amplicons at ~1000 bp and ~2000 bp (lane 1) correspond to the expected sizes of BF0657 deletion construct and wild-type mentioned in the graphic representation. The amplicon at ~3000 bp corresponds to the combined amplification as a result of non specific priming. (C) Transfer of recombinant pGB910 carrying the BF1979 deletion construct into *B. fragilis*. (D) Transfer of recombinant pGB910 carrying the BF3775 deletion construct into *B. fragilis* and agarose gel analysis of PCR amplification BF3775 partial diploid using LOF₄ and ROR₄. The amplicons at ~1100 bp and ~2400 bp (lane 1) correspond to the expected sizes of BF3775 deletion construct and wild-type mentioned in the graphic representation. The amplicon at~ 2500 bp might be a result of non specific priming.

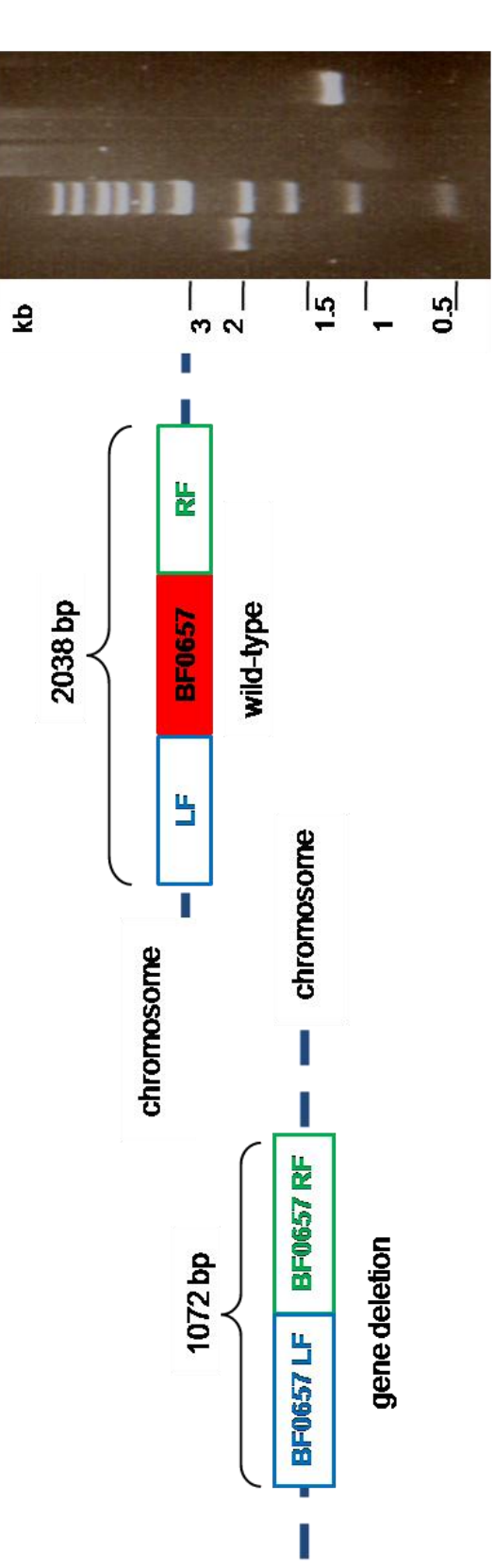
4.3.5. Resolution of partial diploids carrying the deletion constructs

The resulting partial diploid strains were transformed by electroporation with pGB920 (2.4.4). Tetracycline resistant colonies were selected and streaked on DM containing 2% fucose, which induced an I-SceI mediated double-strand break and allowing diploids to resolve to either wild-type or deletion genotypes. Colonies grown on fucose were patched onto plain BHI-S medium and BHI-S medium containing erythromycin to screen for loss of erythromycin resistance. Chromosomal DNA extracts of the erythromycin sensitive colonies were subjected to PCR amplification using left outside (LOF₁-LOF₄) and right outside (ROR₁-ROR₄) primers to confirm the presence of deletion constructs of BF 0275 at ~ 1000 bp (Figure 4.7 A lane 3), BF0657 at ~ 1000 bp (Figure 4.7 B lane 2), BF1979 at ~ 1100 bp (Figure 4.7 C lanes 3 and 4) and BF3775 at ~ 1200 bp (Figure 4.7 D lanes 1-3). However, the diploids that resolved into wild-type genotype resulted in a band corresponding to the size of the wild-type gene along with flanking regions (Figure 4.7 A lanes 1, 2, 4, B lane 1, C lanes 1, 2 and D lane 4).

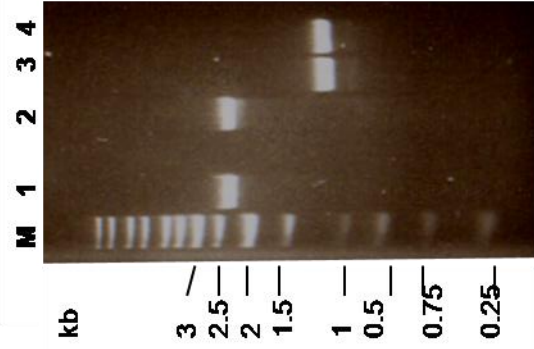
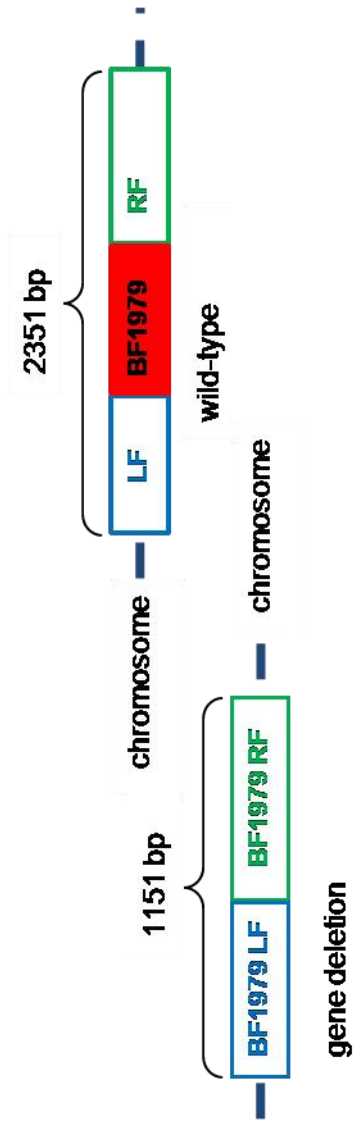




(B)



(C)



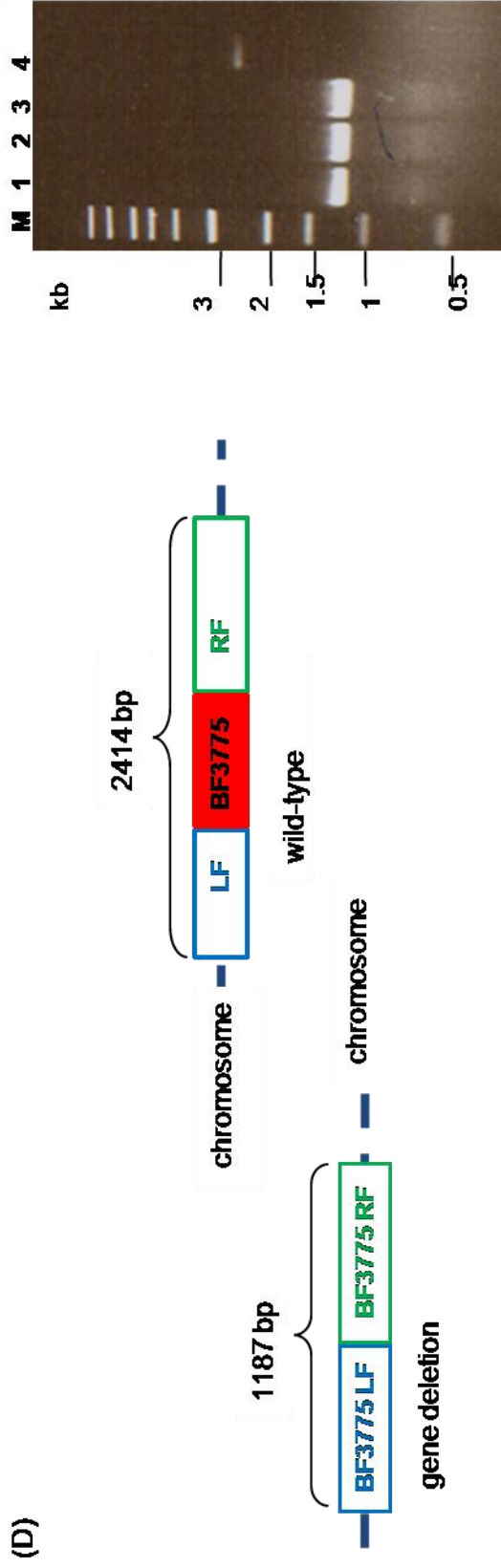


Figure 4.7: Generation of the protease encoding gene deletion mutants

Graphic representation of resolution of diploids by fucose-mediated I-SceI double-strand break into either wild-type or deletion genotype. Agarose gel analysis of PCR amplification using gene-specific primers of (A) Resolution of BF0275 partial diploid into wild-type (lanes 1, 2, 4) and deletion mutant (lane 3). (B) Resolution of BF0657 partial diploid into wild-type (lane 1) and deletion mutant (lane 2). (C) Resolution of BF1979 partial diploid into wild-type (lanes 1, 2) and deletion mutant (lanes 3, 4). (D) Resolution of BF3775 partial diploid into wild-type (lane 4) and deletion mutant (lanes 1-3). Arrows and numbers on the left side indicate size of DNA in kilo bases (kb) in a 1 kb ladder (M). The estimated sizes of wild-types and mutants in the gels were in agreement with the expected sizes indicated in the graphic representations.

4.4. Analysis of fibrinogen degradation by protease-encoding gene deletion mutants

Fibrinogen degradation by the deletion mutants was analysed by fibrinogen zymography in the present study. Mass-Spectrometry analysis by Elhenawy et al. (2014) identified 25% of the proteins present in OMVs released by *B. fragilis* strain NCTC 9343 to be homologous to hydrolases, of which 7 peptidases were exclusive to OMV and not detected in the outer membrane- proteome. Houston et al. (2010) have already indicated the role of non cell-associated proteases in fibrinogenolytic activity exhibited by *B. fragilis*. Therefore, supernatants of BHI-S grown stationary phase cultures of the wild-type strain NCTC 9343 and the four protease-encoding gene deletion mutants were concentrated by centrifuging in membranes containing a 100 kDa MWCO and analysed for their fibrinogenolytic ability in zymography experiments. Zymography is the electrophoretic separation of non-reduced proteins through a polyacrylamide gel containing a proteolytic substrate. The denaturing electrophoresis was followed by protein renaturation and incubation in appropriate buffers for activation of proteolytic activity. 0.1% fibrinogen was used as the proteolytic substrate in 10% SDS polyacrylamide gels. The SDS-fractionated protein samples were renatured and incubated in metalloprotease and serine protease activation buffers (2.7.5). After proteolytic activation, SDS-zymograms were checked for the presence of lytic bands in the form of clearance zones. Proteolytic bands were identified at ~50 kDa for the wild-type sample as well as for all the four single gene deletion mutants, Δ BF0275, Δ BF0657, Δ BF1979 and Δ BF3775 strains (Figure 4.8 B lanes 1,2,3,4 and 5). The presence of fibrinogenolytic bands confirmed that the deletion mutants retained proteolytic activity with respect to fibrinogen.

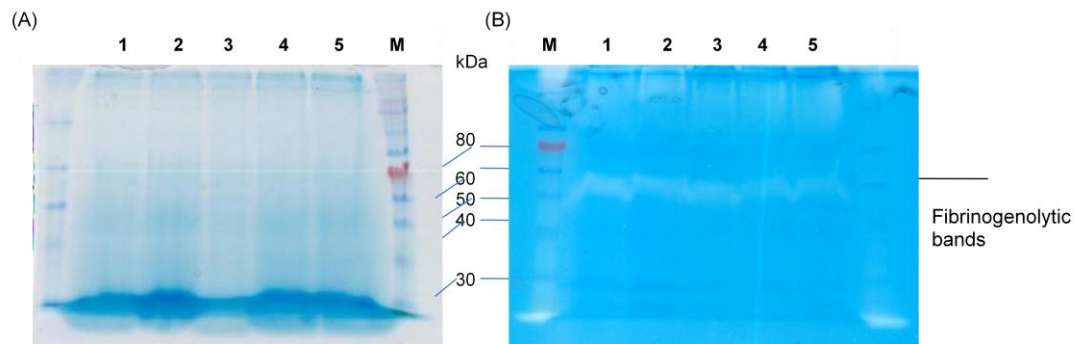


Figure 4.8: Analysis of fibrinogen degradation of the deletion mutants by zymography

(A) SDS-PAGE analysis of concentrated culture supernatants of BHI-S-grown wild-type 9343 (lane 1), Δ BF0275 (lane 2), Δ BF0657 (lane 3), Δ BF1979 (lane 4) and Δ BF3775 (lane 5) and (B) Fibrinogen zymography of the corresponding samples (lanes 1-5). The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. The observed clearance bands in the zymogram indicate the presence of a fibrinogen degrading component at ~50 kDa.

4.5. Analysis of fibrinogen degradation by the generation of multiple protease-encoding gene deletion mutants in *B. fragilis*

The study of virulence-associated proteases in the periodontal pathogen, *P. gingivalis* showed significant decreases in collagenolytic activity with respect to *prtP* single-knockout mutant as well as *rgp-1 prtP* double-knockout mutant. It was also hypothesised that a triple-knockout mutant involving the protease-encoding genes, *prtP*, *rgp-1* and *rgp-2* would block the *P.gingivalis*-mediated collagenolytic activity completely (Barkocy-Gallagher et al., 1999). The role of multiple proteases in fibrinogen degradation in the present study was investigated by performing step-wise deletions of all the four selected proteases in a single *B. fragilis* strain.

Since the functional analysis of single gene deletion mutants of the four selected protease-encoding genes in substrate zymography experiments still degraded fibrinogen, multiple gene deletions of the previously selected protease-encoding genes were generated. Combined effect of these proteases on fibrinogen degradation was analysed, starting from double deletion to quadruple deletion. Elimination of the

tetracycline resistance of single gene deletion mutants was performed prior to generating multiple markerless deletions to render stability to the mutant and for easier genetic manipulations of the mutant. The generation of a quadruple gene deletion mutant in any bacterial model is rare and was specifically tedious with respect to *B. fragilis* strain NCTC 9343 owing to the lack of antibiotic resistance markers. Δ BF3775 was made tetracycline sensitive by repeated sub-culturing in BHI-S without selection and was used as recipient for the generation of double deletion. Recombinant *E.coli* strain S17-1 λ pir carrying BF1979 gene deletion construct was mobilized into tetracycline sensitive BF3775 gene deletion mutant in *B. fragilis* by conjugation. As described earlier, erythromycin resistant transconjugants were electroporated with pGB920. Erythromycin sensitive colonies were confirmed for the presence of BF1979 gene deletion construct by PCR amplification using LOF₃ and ROR₃ primers. Amplicons of ~1100 bp were confirmed on 1% agarose gels (Figure 4.9 lanes A1-A3, A6, A7, A8, A11, A12, B1-B6, B7, B10). The Δ BF1979 Δ BF3775 mutants thus generated were the recipients for the mobilization of BF0275 gene deletion construct to generate the Δ BF1979 Δ BF3775 Δ BF0275 mutant with ~ 1000 bp amplicon corresponding to BF0275 deletion construct observed on agarose gel (Figure 4.10 lane 3). In a similar manner, a quadruple gene deletion mutant, Δ BF1979 Δ BF3775 Δ BF0275 Δ BF0657 was generated (Figure 4.11 lanes 1-4) using BF0657 gene deletion construct in *E. coli* strain S17-1 λ pir.

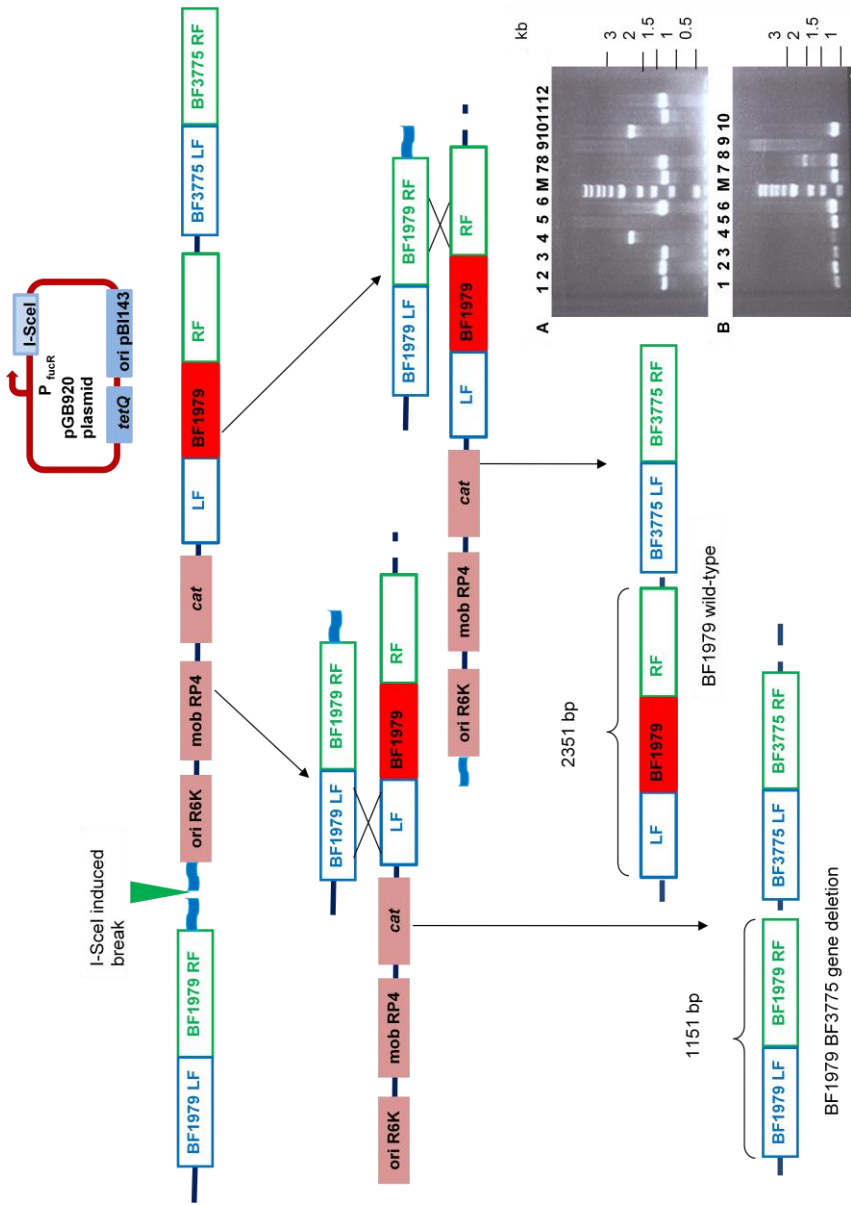


Figure 4.9: Generation of the Δ BF1979 Δ BF3775 double deletion mutant from the Δ BF3775

Graphic representation of introducing BF1979 deletion construct into Δ BF3775 strain and diploid resolution by fucose-mediated I-Sce double-strand break into either Δ BF1979 Δ BF3775 double gene deletion mutant or wild-type parent strain (Δ BF3775) and agarose gel analysis of BF1979 deletion construct in Δ BF1979 Δ BF 3775 (lanes A1-A3, A6, A7, A8, A11, A12, B1-B6, B7, B10) and BF1979 wild-type + flanks (lanes A4 and A10) by PCR amplification using primers LOF₃ and ROR₃. Arrows and numbers indicate size of DNA in kilo bases (kb) in a 1 kb ladder (M). Amplicon sizes observed in the gel are in agreement with the expected sizes mentioned in the graphic representation.

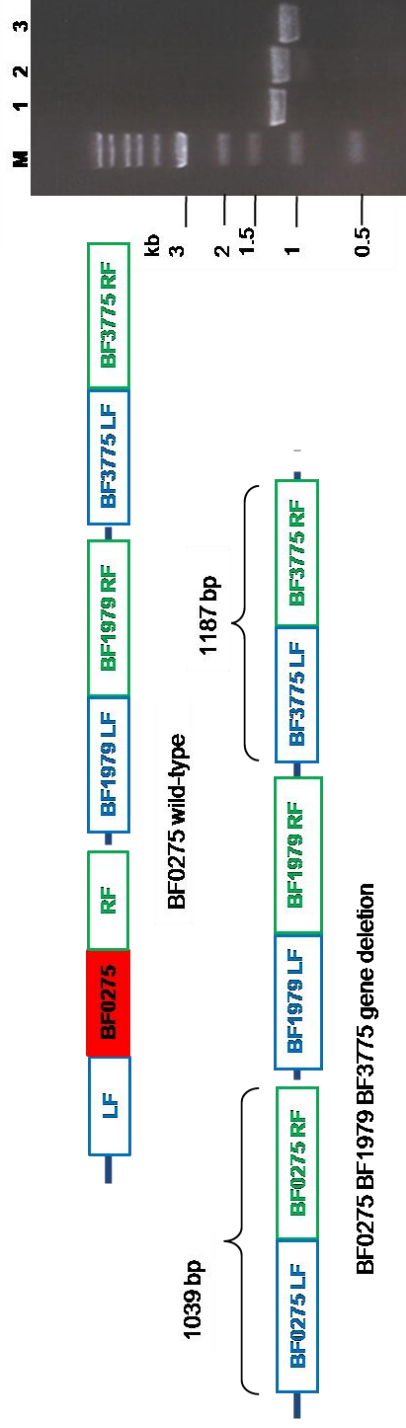


Figure 4.10: Generation of the Δ BF0275 Δ BF1979 Δ BF3775 triple deletion mutant from the Δ BF1979 Δ BF3775

Graphic representation of introducing BF0275 deletion construct into Δ BF1979 Δ BF3775 double gene deletion mutant strain followed by diploid resolution via fucose-mediated I-SceI double-strand break into either Δ BF0275 Δ BF1979 Δ BF3775 triple gene deletion mutant or wild-type parent strain (Δ BF1979 Δ BF3775 double gene deletion mutant) and agarose gel analysis of PCR amplification of BF0275 deletion construct in Δ BF0275 Δ BF1979 Δ BF3775 triple gene deletion mutant (lane 3) using primers LOF₁ and ROR₁. Arrows and numbers on the left side indicate size of DNA in kilo bases (kb) in a 1 kb ladder (M). Size of the BF0275 gene deletion construct observed in the gel corresponds to the expected size mentioned in the graphic representation.

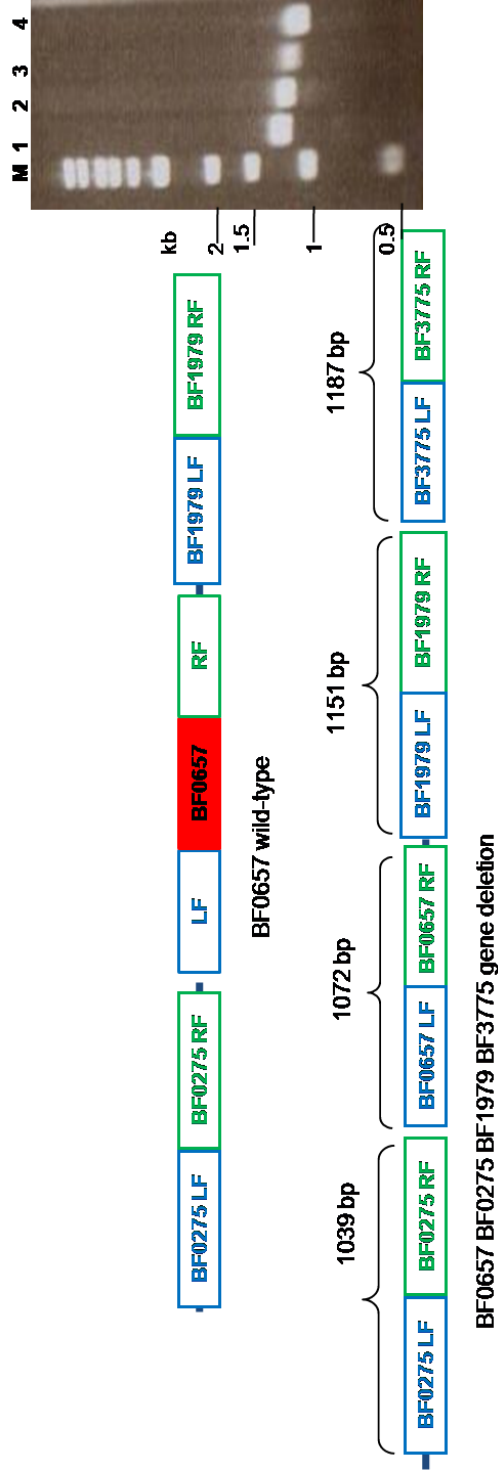


Figure 4.11: Generation of the Δ BF0657 Δ BF0275 Δ BF1979 Δ BF3775 quadruple gene deletion mutant from the Δ BF0275 Δ BF1979 Δ BF3775

Graphic representation of introducing BF0657 deletion construct into Δ BF0275 Δ BF1979 Δ BF3775 triple gene deletion mutant strain followed by fucose-mediated I-SceI double-strand break of the partial diploid into either Δ BF0657 Δ BF0275 Δ BF1979 Δ BF3775 quadruple gene deletion mutant or the wild-type parent strain (Δ BF0275 Δ BF1979 Δ BF3775 triple gene deletion mutant) and agarose gel analysis of PCR amplification of BF3775, BF1979, BF0657 and BF0275 deletion constructs (lanes 1-4) in the quadruple gene deletion mutant using primers LOF₄-LOF₁ and ROR₄-ROR₁. Arrows and numbers on the left side indicate size of DNA in kilo bases (kb) in a 1 kb ladder (M). Estimated sizes of all the four gene deletion constructs in the gel were in agreement with their expected sizes in the graphic representation.

4.5.1. Detection of fibrinogen degradation by the multiple deletion mutants

Fibrinogen zymography was repeated to analyse the ability of multiple deletion mutants to degrade fibrinogen. Concentrated culture supernatants of the double, triple and quadruple gene deletion mutants grown on BHI-S as well as glucose-DM were prepared. Cell-free supernatants concentrated from cultures grown in glucose DM produced a single proteolytic band at ~45 kDa on fibrinogen-SDS zymograms after reactivation of proteases and was detected in wild-type strain NCTC 9343 as well as the quadruple deletion mutant. Three replicates of the zymography experiments were performed and a representative example is shown (Figure 4.12). Concentrated BHI-S media was used as a control sample in which no proteolytic activity was detected (Figure 4.12 B lane 3). Two proteolytic bands at ~45 and 50 kDa were detected in SDS-zymograms of electrophoresed samples of multiple deletion mutant and wild-type 9343 concentrated supernatants grown in BHI-S media (Figure 4.13 lanes 1, 2 and 3). The observed proteolytic bands in zymography experiments suggest that fibrinogen degradation is independent of the combined proteolytic activity of all 4 selected proteases. Fibrinogenolytic bands were observed in all the samples after incubation in metalloprotease activation buffer alone and omitting serine protease activation buffer, thus suggesting that a metalloprotease is involved in the degradation of host fibrinogen by *B. fragilis* NCTC 9343.

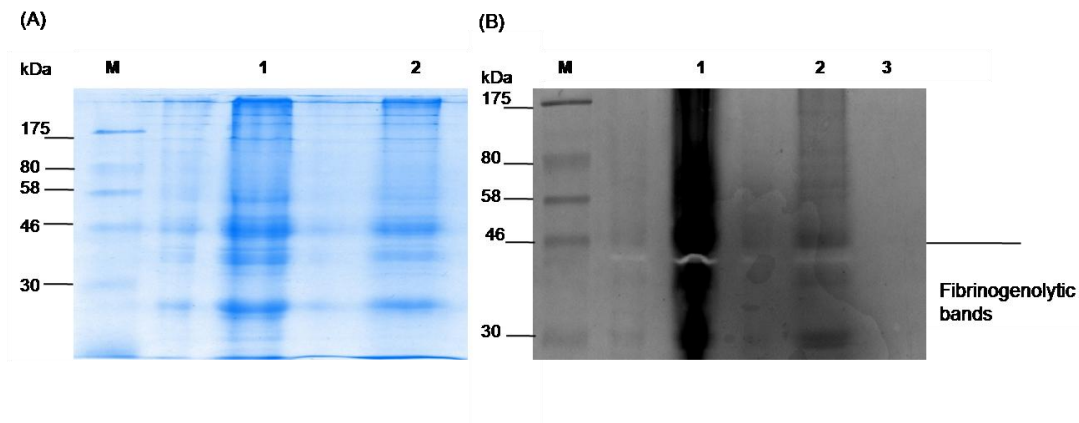


Figure 4.12: Analysis of fibrinogen degradation by the quadruple deletion mutant

(A) SDS gel analysis of concentrated culture supernatants of glucose-DM-grown $\Delta\text{BF0657}\Delta\text{BF0275}\Delta\text{BF1979}\Delta\text{BF3775}$ quadruple gene deletion mutant (lane 1) and wild-type 9343 (lane 2) and (B) Fibrinogen zymography analysis of the corresponding samples (lanes 1 and 2). The positions of molecular weight markers (M) are shown in kDa adjacent to the gel. The zone of fibrinogenolysis is observed at ~ 45 kDa in the zymogram of the two samples.

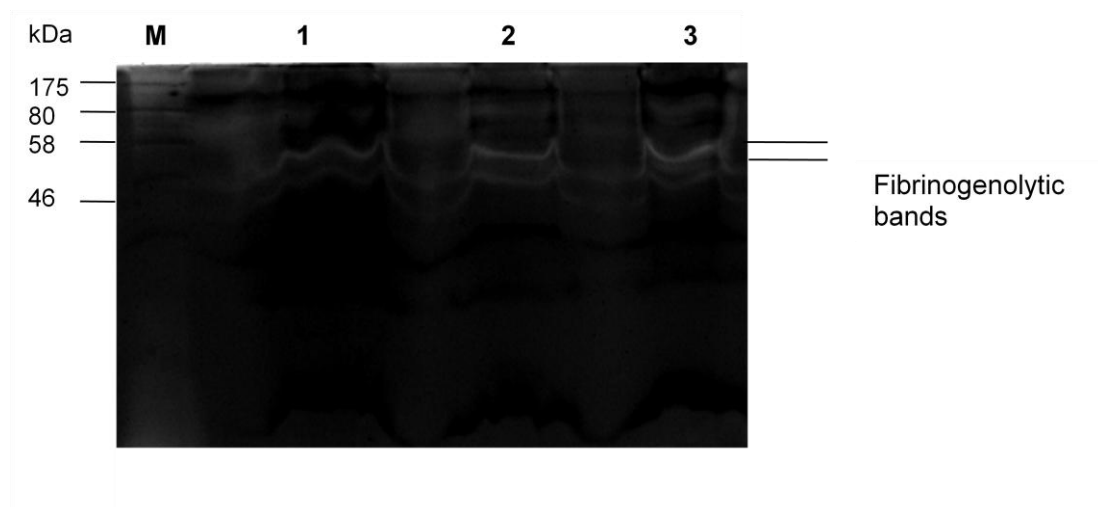


Figure 4.13: Analysis of fibrinogen degradation by the triple deletion mutant

Fibrinogen zymography analysis of concentrated culture supernatants of BHI-S-grown wild-type 9343 (lane 1), triple protease encoding gene deletion mutant (lane 2) and quadruple protease encoding gene deletion mutant (lane 3). The positions of molecular weight markers (M) are shown in kDa adjacent to the gel. Fibrinogenolytic bands are observed at ~ 45 kDa and 50 kDa in the three samples.

4.6. Identification, cloning and expression of metalloproteases secreted by *B. fragilis*

4.6.1. Identification of secreted proteases and peptidases in *B. fragilis*

Multiple deletion mutants generated in the present study retained the fibrinogenolytic activity displayed by the wild-type *B. fragilis* on zymograms. Therefore, other secreted protease-encoding genes within the molecular weight range (30-70 kDa) of the observed fibrinogenolytic activity (45-50 kDa) were investigated for their role in fibrinogen degradation. Additionally, Chen et al., (1995) detected fibrinogenolytic activity in relation to a putative 100 kDa serine-thiol-like protease from *B. fragilis* YCH46 albeit neither candidate gene nor an extracellular proteolytic activity has been conformed. Therefore, we included four protease-encoding genes with molecular weights > 75kDa in our investigation to detect a possible fibrinogenolysis at ~100 kDa on account of the described putative protease. Also, proteases belonging to the metalloprotease family were preferred in the present study since fibrinogen

degradation was detected in zymograms incubated with the metalloprotease activation buffer alone. Previous studies have detected fibrinolytic and fibrinogenolytic activities in relation to elastase secreted by *P. aeruginosa*. Elastase which was similar to *B. subtilis* thermolysin-like zinc-protease resulted in the complete hydrolysis of A α - and B β -chains of fibrinogen (Komori et al., 2001). Elastase-mediated fibrinogen degradation increased hemorrhagic tendency in animals suffering from acute pneumonia caused by *P. aeruginosa*. Albeit the present study deals with the non enterotoxigenic *B. fragilis* strain NCTC 9343, the enterotoxin associated with certain strains of *B. fragilis* implicated in diarrheal disease in animals and humans has been associated with the metzincin family of metalloproteases (Moncrief et al., 1995). Moreover, Houston et al. (2010) observed inhibition of fibrinogen degradation when zymograms containing concentrated culture supernatants of NCTC 9343 were pre-incubated with the metalloprotease inhibitor EDTA but not when pre-incubated with the serine protease inhibitor PMSF.

Other secreted protease and peptidase encoding genes in the *B. fragilis* NCTC 9343 genome were identified using Artemis genome browser. Apart from the previously identified four genes, 11 out of 35 protease encoding genes and 17 out of 49 peptidase encoding genes present in NCTC 9343 were annotated as containing signal peptide sequences (Table 4.2). 8 out of these genes were selected for expression analysis based on the expressed protein size and type of protease or peptidase encoded (metallo- and serine- catalytic type) (Table 4.2. highlights).

Table 4.2: List of the genes encoding putative secreted proteases and peptidases in NCTC 9343 genome

List of 27 genes encoding signal peptide-containing putative proteases and peptidases in NCTC 9343 genome. The highlighted 8 genes, 7 of which encode metallopeptidase and 1 encodes a serine peptidase, BF2517, were selected for expression and analysis. The selected peptidase-encoding genes expressed proteins >45 kDa in size, except BF1335 and BF1496 metallopeptidase encoding genes, and shared significant identities with protease-encoding genes in other bacterial spp.

Feature Name	CDS	Size	Hits
BF0899 Peptidase S9	1127056...1129263 c	2208 bp 736 aa 84.049 kDa	<i>P. gingivalis</i> dipeptidyl peptidase DPP IV (723 aa) E value : 1.6e-124 and 51.01% identity in 741 aa
BF0097 Peptidase S9	100099...102255 c	2157 bp 719 aa 81.303 kDa	<i>Xanthomonas maltophilia</i> dipeptidyl peptidase IV (741 aa) E value : 2.2e-56 and 32.47% identity in 622 aa
BF0145 Peptidase S26	151089...152021 c	933 bp 311 aa 35.795 kDa	Low similarity to <i>Bradyrhizobium japonicum</i> signal peptidase SipS (259 aa) E value : 0.00072 and 29.68% identity in 128 aa
BF0146 Peptidase S26	152032...153513 c	1482 bp 494 aa 57.3 kDa	C terminus of <i>E. coli</i> 06 signal peptidase I LepB or C3092 (324 aa) E value : 2.2e-06 and 44.3% identity in 79 aa
BF1408 Peptidase U34	1673656...1675290	1635 bp 545 aa 61.497 kDa	<i>Salmonella dublin</i> probable dipeptidase PipD (433 aa) E value : 8.5e-21 and 29.38% identity in 405 aa
BF1335	complement	798 bp 266 aa	<i>B. thetaiotaomicron</i>

Peptidase M48	1597400...1598197	28.786 kDa	peptidase BT0149 (260 aa) E value : 7e-71 and 78.86% identity in 265 aa zinc metallopeptidases
BF1496 Peptidase M23/37	1757378...1757926	549 bp 183 aa 20.251 kDa	C terminus of <i>Thermoanaerobacter tengcongensis</i> membrane proteins related to metallo endopeptidases NlpD4 and TTE1975 (389 aa) E value : 4.9e-19 and 44.02% identity in 159 aa
BF2064 Peptidase subfamily C1B	2415148...2416542	1395 bp 465 aa 53.132 kDa	<i>Lactobacillus delbrueckii</i> aminopeptidase C PepC (449 aa) E value : 1.4e-59 and 37.79% identity in 418 aa
BF2627 Peptidase S9	3056507...3058612	2106 bp 702 aa 79.461 kDa	<i>Caulobacter crescentus</i> prolyl oligopeptidase family protein CC1986 (683 aa) E value : 5e-71 and 32.12% identity in 663 aa
BF2764 Peptidase	3221523...3222722	1200 bp 400 aa 44.655 kDa	<i>Bacillus sphaericus</i> dipeptidyl-peptidase VI (271 aa) E value : 5e-10 and 25.09% identity in 267 aa
BF2989 Peptidase S26 Signal peptidase I	3479480...3480367	888 bp 296 aa 34.258 kDa	<i>Chlorobium tepidum</i> signal peptidase I LepB ot CT1450 (280 aa) E value : 1.1e-13 and 35.14% identity in 276 aa
BF3021 Peptidase M3	complement 3510933...3513032	2100 bp 700 aa 79.337 kDa	<i>Xylella fastidiosa</i> peptidyl-dipeptidase XF1944 (716 aa) E

			value : 6.9e-111 and 43.31% identity in 718 aa neutral zinc metallopeptidases
BF3195 Peptidase C39 like	3749935...3751116	1182 bp 394 aa 44.145 kDa	<i>B. thetaiotaomicron</i> aminopeptidase C BT1789 (394 aa) E value : 6.7e-136 and 89.59% identity in 394 aa
BF3237 Peptidase S9	3806942...3809476	2535 bp 845 aa 95.310 kDa	<i>B. thetaiotaomicron</i> putative alanyl dipeptidyl peptidase BT1838 (840 aa) E value : 0 and 82.02% identity in 840 aa
BF3247 Peptidase M49	complement 3820665...3822713	2049 bp 683 aa 77.070 kDa	<i>B. thetaiotaomicron</i> putative dipeptidyl-peptidase III BT1846 (675 aa) E value : 0 and 82.54% identity in 676 aa
BF3732 Peptidase M23/37	complement 4401251...4402129	879 bp 293 aa 32.884 kDa	<i>P. gingivalis</i> W83 peptidase M23/37 family PG2192 (337 aa) E value : 6.6e-43 and 46.34% identity in 287 aa
BF3918 Peptidase M13	complement 4615597...4617627	2031 bp 677 aa 76.629 kDa	<i>P. gingivalis</i> endothelin-like protein Pep0 (689 aa) E value : 8.5e-137 and 51.08% identity in 689 aa
BF0080 Peptidase S41B	76143...79412 c	3270 bp 1090 aa 122.3 kDa	<i>S. coelicolor</i> tricorin protease homolog 1 Tri1 sco2549 (1067 aa) E value : 4.7e-27 and 23.81% in 1121 aa
BF0506 Putative outer	586869...588197 c	1329 bp 443 aa	<i>Proteus mirabilis</i> metalloprotease

membrane transport protein		49.509 kDa	transporter component ZapD (449 aa) E value : 3.6e-12 and 24.57% identity in 407 aa
BF0935 Peptidase M 16	complement 1168153...1170969	2817 bp 939 aa 106.556 kDa	<i>E. coli</i> probable zinc protease PqqL (931 aa) E value : 1.5e-56 and 26.75% identity in 927 aa
BF1640 Peptidase C10	1914390...1915844	1455 bp 485 aa 56.677 kDa	Limited similarity to <i>B. thetaiotaomicron</i> putative pyrogenic exotoxin B BT2451 (426 aa) E value : 0.0088 and 27.53% identity in 483 aa
BF2517 Peptidase S41B	complement 2929009...2932239	3231 bp 1077 aa 120.152 kDa	<i>Streptomyces coelicolor</i> tricorin protease homolog 1 Tri 1 or SC02549 (1067 aa) E value : 1e-51 and 29.6% identity in 1091 aa
BF2762 Cysteine protease	3218992...3220323 c	1332 bp 444 aa 51.708 kDa	<i>Thermoanaerobacter tengcongensis</i> transglutaminase-like enzymes, putative cysteine proteases TTE0036 (440 aa) E value : 2.9e-27 and 31.81% identity in 418 aa
BF3752 Peptidase S41	4429629...4432913	3285 bp 1095 aa 123.499 kDa	<i>Thermoplasma volcanium</i> tricorin protease Tri or TV0915 (1030 aa) E value : 1.2e-16 and 21.73% identity in 1104 aa
BF3964 Transporter protein	4679352...4680668	1317 bp 439 aa 49.013 kDa	<i>Pseudomonas aeruginosa</i> alkaline protease secretion protein AprF or pal248 (481 aa) E

			value : 2.9e-10 and 20.58% identity in 413 aa
BF4172 Peptidase S41	4920643...4922322	1680 bp 560 aa 62.649 kDa	<i>Bartonella baciliformis</i> C terminal processing protease precursor CtpA (434 aa) E value : 2.5e-33 and 35.08% identity in 362 aa
BF4335 Peptidase S41	5157260...5158885	1626 bp 542 aa 60.946 kDa	<i>Bartonella baciliformis</i> C terminal processing protease precursor CtpA (434 aa) E value : 1.7e-33 and 36.33% identity in 355 aa

4.6.2. Choice of vector for the expression of the selected proteases in *E. coli*

To analyse their individual fibrinolytic potential, the selected *B. fragilis* protease-encoding genes were expressed in *E. coli*. The protease-/peptidase-encoding genes were expressed using pET-100/D-TOPO vector of 5764 bp size from the Champion pET Directional TOPO Expression kit (2.5.8 and 2.7.10). These expression kits utilise a highly efficient 5 minute TOPO cloning strategy to directionally clone a blunt-end PCR product into vector for high level T7 regulated expression in *E. coli*. Expression vectors available with the kit were originally developed by Studier & Moffatt (1986) and utilise the high activity and specificity of bacteriophage T7 RNA polymerase to facilitate regulated expression of heterologous genes in *E. coli* from T7 promoter. In the present study, pET-100/D-TOPO was the vector of choice which contains an N terminal expressed 6xHis fusion tag and ampicillin selection marker.

Features of the chosen pET-100 vector:

- i. T7 *lac* promoter for high level IPTG inducible expression of gene of interest in *E. coli*
- ii. Directional TOPO cloning site for rapid and efficient directional cloning of PCR products
- iii. N terminal fusion tags for detection and purification of recombinant fusion proteins
- iv. Protease recognition for cleavage of fusion tag from protein of interest
- v. *lacI* gene encoding lac repressor to reduce basal transcription from T7 *lac* promoter in pET-TOPO vector and from *lac UV5* promoter in *E. coli* host chromosome
- vi. Antibiotic resistance marker for selection in *E. coli*
- vii. pBR322 origin for low-copy replication and maintenance in *E. coli*

Topoisomerase I from *Vaccinia* virus binds to double-stranded DNA at specific sites and cleaves the phosphodiester backbone after 5'-CCCTT in one strand (Shuman, 1991). The energy produced is conserved by formation of a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue of topoisomerase I. The phospho-tyrosyl bond between the DNA and enzyme can subsequently be attacked by the 5' hydroxyl of the original cleaved strand thereby reversing the reaction and releasing topoisomerase (Shuman, 1994). TOPO cloning harnesses this reaction to clone blunt end PCR products into pET-TOPO vector.

4.6.3. Primer design and cloning of expression construct into pET-100 vector

Directional cloning was achieved by adding a 4 nucleotide (GTGG) overhang sequence to the TOPO-charged DNA coupled with the addition of a CACC overhang to the forward primer of the incoming DNA. GTGG overhang in the cloning vector invades the 5' end of the PCR product, anneals to the added bases and stabilises the

PCR product in the correct orientation. This technique was originally developed by Cheng & Shuman (2000). Primers were designed for the eight selected protease/peptidase encoding genes for strain NCTC 9343. CACC tag was added to the 5' end of all primers followed by the start codon and gene sequence. Signal peptide sequences were omitted as the gene was to be expressed in *E. coli*. Stop codons were included in the reverse primers (Figure 4.14). The primers were used to amplify selected protease encoding genes using appropriate PCR conditions (2.5.3). The sizes of the PCR amplified products were confirmed by 1% agarose gel electrophoresis (Figure 4.15 lanes 1-8).

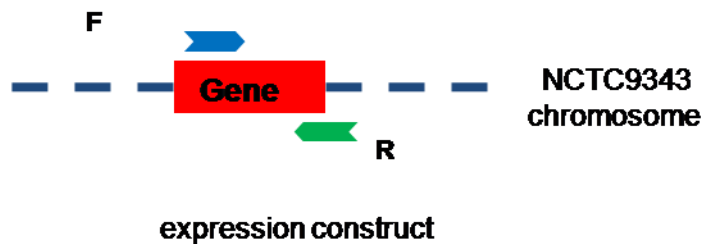


Figure 4.13: Graphic representation of the expression construct model

Graphic representation of the model of a gene (red box) along with forward primer (denoted in blue) and reverse primer (denoted in green) to amplify the expression construct.

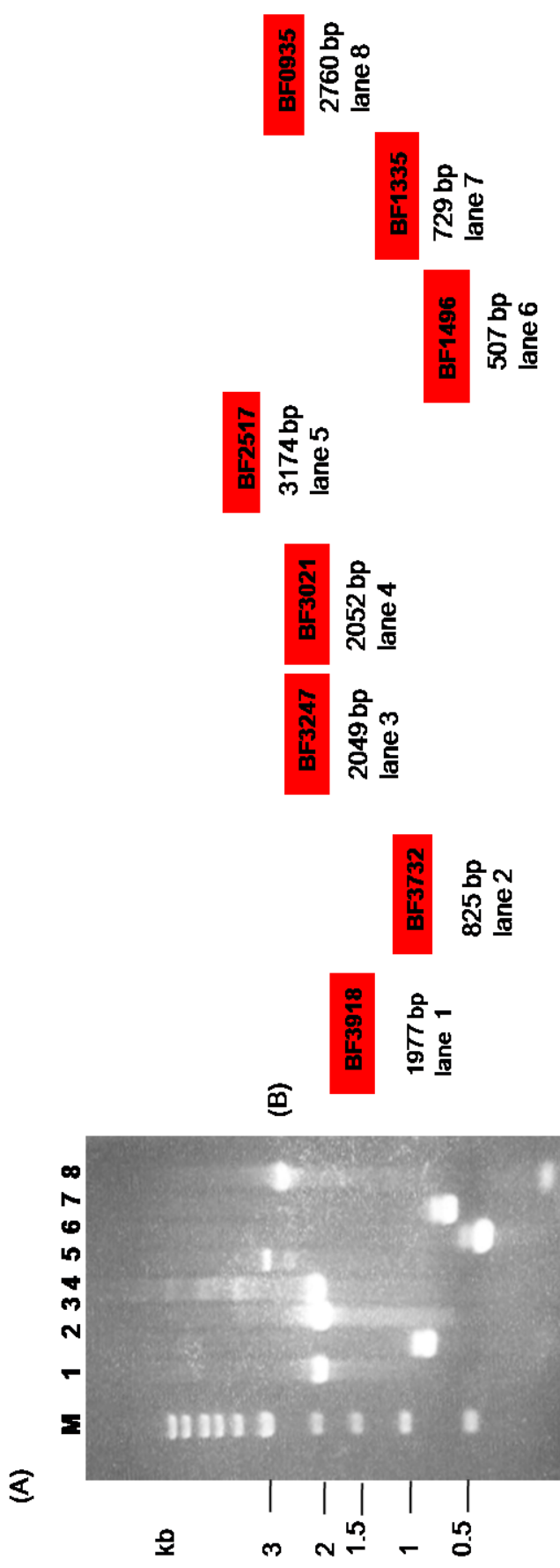
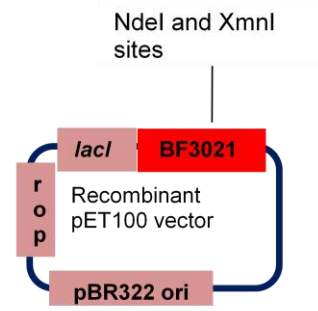
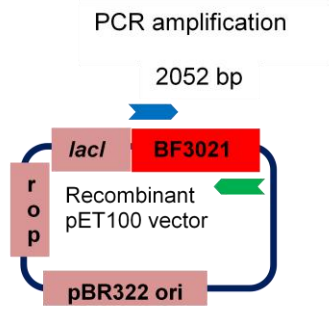
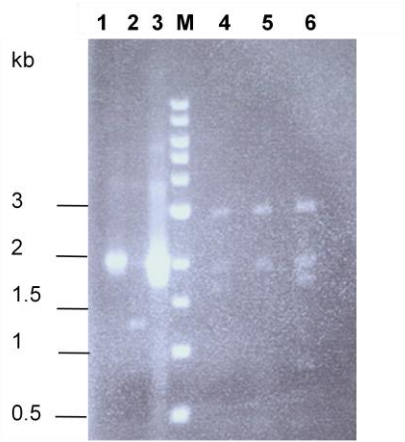


Figure 4.14: Amplification of the protease/peptidase encoding genes selected for expression

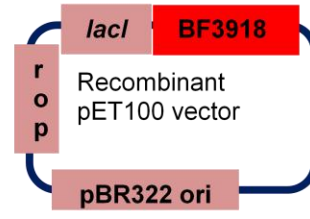
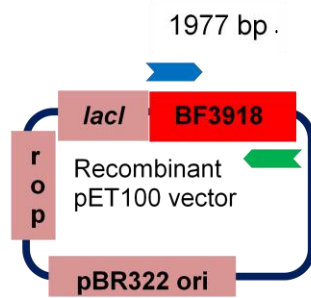
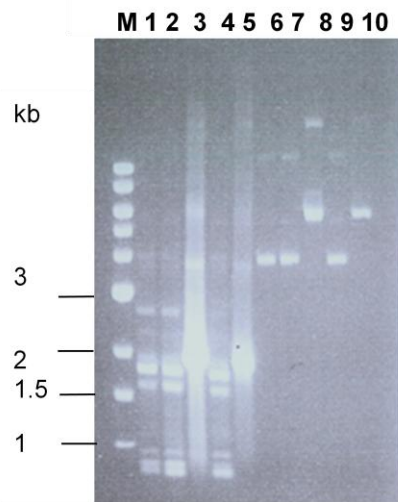
(A) Agarose gel analysis of PCR amplification of the eight selected genes (lanes 1-8) using the specific primers, F₁- R₁ to F₈-R₈. Arrows and numbers on the left side indicate size of DNA in kilo bases (kb) in a 1 kb ladder (M). The estimated sizes of the amplicons observed in the gel correspond to the expected sizes. (B) Graphic representation of the expected sizes of the protease/peptidase encoding genes.

TOPO cloning reaction products were recovered in *E. coli* Top10 for characterisation of the construct, propagation and maintenance. The presence of recombinant pET-TOPO in Top10 strain was confirmed by single colony analysis (Figure 4.16 C), PCR amplification of the recombinant pET-TOPO vector using specific primers (Figure 4.16 B and D) and restriction digestion analysis (Nde1 and Xmn1 double digestion for insert release) of the pET-TOPO clone carrying BF3021 (Figure 4.16 A). Recombinant pET-TOPO vectors were also subjected to Sanger sequencing and the sequences generated were compared to the nucleotide sequences of the respective expression constructs by nucleotide BLAST analysis. Sequence identities ranging from 92%-100% were observed for each forward and reverse reaction, except for the BF1496 clone which failed to exhibit any identity with the corresponding expression construct.

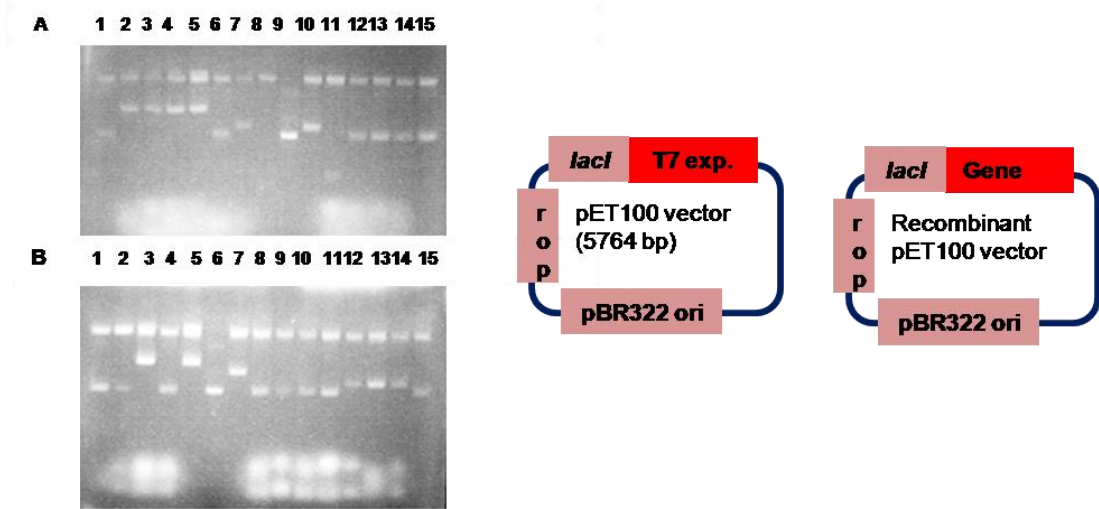
(A)



(B)



(C)



(D)

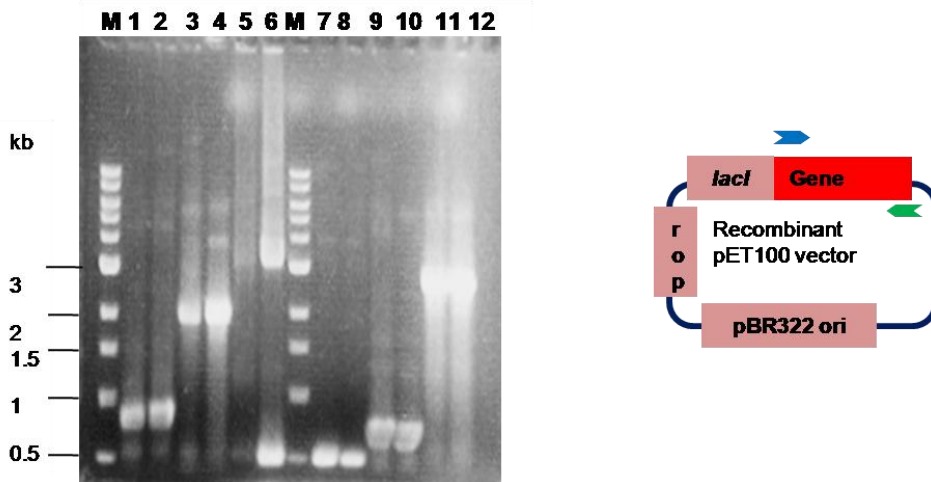


Figure 4.15: Confirmation of the clones positive for the protease-encoding gene expression constructs

Arrows and numbers on the left side indicate size of DNA in kilo bases (kb) in a 1 kb ladder (M). (A) Agarose gel analysis of PCR amplification of recombinant vector carrying BF3021 expression construct present in Top10 clones 1, 3 and 4 (lanes 1, 2 and 3) using BF3021-specific primers F₅ and R₅ and their respective NdeI-XmnI double restriction digestion to release BF3021 expression construct from the recombinant pET-100 vector (lanes 4, 5 and 6). Top10 clones 1 and 4 are positive for BF3021 expression construct as observed from the amplicon size (~2000 bp) and digestion pattern. (B) Agarose gel analysis of PCR

amplification of recombinant vector carrying BF3918 expression construct present in Top10 clones 1, 2, 3, 4 and 5 (lanes 1-5) using BF3918-specific primers F₈ and R₈ and their respective plasmid extracts (lanes 6-10). Top10 clones 3 and 5 are positive for BF 3918 expression construct as observed from the specific amplicon size (~2000 bp) and increase in plasmid size. (C) Agarose gel analysis of single colony assay of plasmid size in recombinant vector extracted from Top10 clones 1, 2, 5, 6 and 10 (lanes A1-A5) carrying BF0935 expression construct, clones 2, 4, 6, 7 and 9 (lanes A6-A8, A10 and A11) carrying BF1335 expression construct, clones 1, 3, 5 and 7 (lanes A12-A15) carrying BF1496 expression construct, clones 2, 4, 6, 9 and 10 (lanes B1-B5) carrying BF2517 expression construct, clones 1, 5, 8, 9 and 10 (lanes B7-B11) carrying BF3247 expression construct and clones 2, 4, 6 and 8 (lanes B12-B15) carrying BF3732 expression construct when compared to the size of the empty pET-100 vector backbone (lanes A9 and B6). An up-shift in size of the recombinant vectors carrying BF0935 (lanes A2-A5), BF1335 (lanes A7 and A10), BF1496 (lanes A12-A15), BF2517 (lanes B3 and B5), BF3247 (lane B7) and BF3732 (lanes B12-B14) when compared to the empty vector backbone (lanes A9 and B6) indicate the presence of the respective expression constructs in recombinant vector. (D) Agarose gel analysis of PCR amplification of BF3732 (lanes 1 and 2), BF3247 (lanes 3 and 4), BF 2517 (lanes 5 and 6), BF1496 (lanes 7 and 8), BF1335 (lanes 9 and 10) and BF0935 (lanes 11 and 12) expression constructs using respective gene-specific forward (F, denoted in green) and reverse primer (R, denoted in blue) sets. The amplicons were in agreement with the sizes of the expression constructs.

4.6.4. Expression of recombinant pET-TOPO vector

The selected protease-encoding genes were expressed in *E. coli* to determine their ability to degrade fibrinogen. The selected metalloprotease/peptidase- and serine peptidase- encoding genes in *B. fragilis* were expressed following IPTG induction in *E. coli* BL21 carrying the recombinant pET-TOPO vector (2.7.10). In pET-TOPO vectors, gene expression is controlled by a strong bacteriophage T7 promoter which is modified to contain a *lac* operator sequence. In TOPO system, T7 RNA polymerase is supplied by *E. coli* BL21 Star (DE3) in a regulated manner and on production of sufficient quantity of T7 RNA polymerase, it binds to T7 promoter and transcribes the gene of interest. BL21 carries the DE3 bacteriophage λ lysogen which contains a *lac* construct comprising the *lacI* gene encoding the *lac* repressor, the T7 RNA polymerase-encoding gene under the control of the *lacUV5* promoter and a small portion of the *lacZ* gene. Addition of IPTG induces expression of T7 RNA polymerase from the *lacUV5* promoter. The *lac* operator sequence placed downstream of the T7 promoter in the vector binds to the *lac* repressor and represses T7 polymerase-induced basal transcription of the gene of interest in BL21 host cells. The absence of lon protease and a mutation in the outer membrane protease, OmpT, reduces the degradation of heterologous proteins expressed in the chosen *E. coli* strain BL21 Star (DE3). Moreover, the mRNA stability of the expressed protein is ensured by a mutation in the *rne* resulting in a truncated Rnase E enzyme which lacks the ability to degrade mRNA.

The induced and uninduced cultures were electrophoresed through 12% SDS-polyacrylamide gels and the molecular weight of the expressed proteins were confirmed (Figure 4.17). Expressions of BF1335 and BF3732 were induced at ~ 30 kDa (Figure 4.17 lanes 3 and 14) when compared to their uninduced samples (Figure 4.17 lanes 6 and 13). Induction of BF3021, BF3247 and BF3918 were observed at ~ 75 kDa (Figure 4.17 lanes 10, 12 and 16) when compared to uninduced samples (Figure 4.17 lanes 9, 11 and 14). BF2517 and BF0935 were induced at sizes above 100 kDa (Figure 4.17 lanes 1 and 7) on comparison with the respective uninduced samples (Figure 4.17 lanes 2 and 8). The estimated sizes of the induced proteins from

the SDS-polyacrylamide gels were in agreement with their expected sizes. However an induced protein corresponding to ~20 kDa was not observed for BF1496 when compared to the respective uninduced sample (Figure 4.17 lanes 5 and 6).

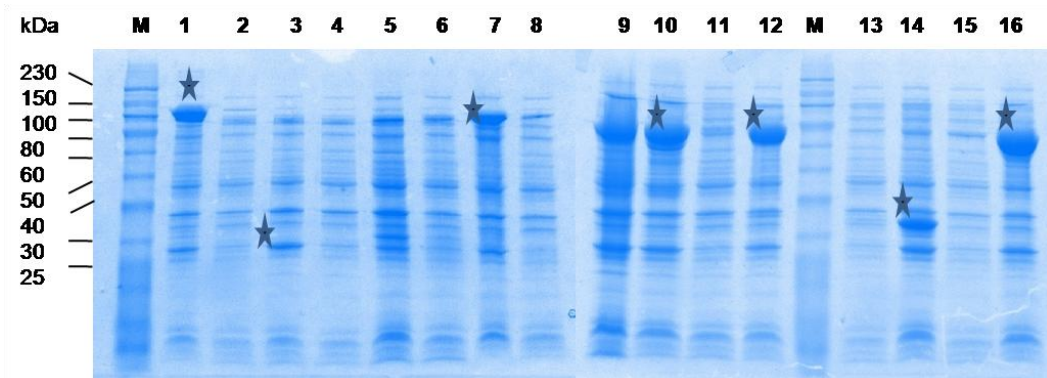


Figure 4.16: Expression of the selected protease-encoding genes in *E. coli*

SDS-PAGE analysis of IPTG-induced expression and uninduced samples of the selected *B. fragilis* protease-encoding genes in *E. coli*. The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. Induced expression of BF 0935 at ~100 kDa (lane 1), BF1335 at ~28 kDa (lane 3), BF2517 at ~120 kDa (lane 7), BF3021 at ~80 kDa (lane 10), BF3247 at ~77 kDa (lane 12), BF3732 at ~32 kDa (lane 14) and BF3918 at ~77 kDa (lane 16) when compared to the respective uninduced samples (lanes 2, 4, 6, 8, 9, 11, 13 and 15). The protein expression in each lane is marked by an★. No distinct band corresponding to BF1496 expression is observed (lane 5).

4.7. Analysis of fibrinogen degradation by the expressed proteases/peptidases in fibrinogen zymography

The ability of the induced strains to degrade human fibrinogen was examined by fibrinogen zymography using concentrated culture supernatant of wild-type NCTC 9343 as the positive control (Figure 4.18 lane 1). No zones of clearance at ~45 kDa were detected in the *B. fragilis* proteases/ peptidases expressed in *E. coli* on visualising the zymogram (Figure 4.18 lanes 2-8). Also, fibrinogenolysis was not detected at ~100 kDa confirming that the previously studied serine-thiol-like protease from *B. fragilis* YCH46 is either not expressed or functionally active in *B. fragilis*

NCTC 9343. Absence of fibrinolytic bands indicated that the proteases/peptidases analysed here are not capable of degrading fibrinogen when expressed in this host. Future work would therefore require the expression of other identified secreted metallo or serine proteases/peptidases have to be expressed and their fibrinolytic property analysed. There is also a possibility that the expressed proteases and peptidases might not be functionally active in *E. coli*, which might explain the inability to degrade fibrinogen. A similar difficulty was experienced in functional analysis of *P. gingivalis* encoded proteases in *E. coli* (Barkocy-Gallagher et al., 1999). The proteases could be functionally expressed when ecologically similar *Bacteroides* spp. was used as expression strain instead of *E. coli*.

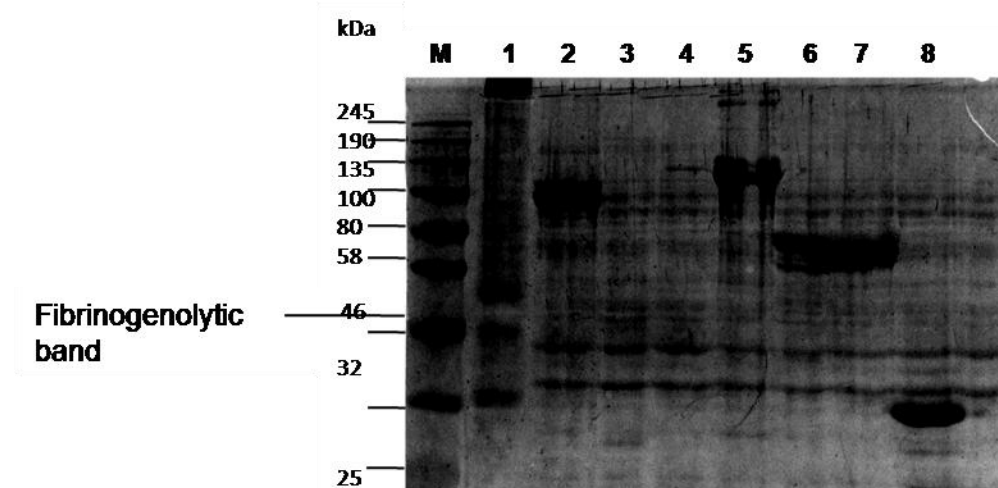


Figure 4.17: Analysis of fibrinogen degradation by the protease-encoding genes expressed in *E. coli*

Fibrinogen zymography analysis of concentrated culture supernatant of glucose-DM grown wild-type 9343 (lane 1) and *E. coli* expressed *B. fragilis* protease-encoding BF0935 (lane 2), BF1335 (lane 3), BF2517 (lane 5), BF3021 (lane 6), BF3247 (lane 7) and BF3732 (lane 8). The positions of molecular weight markers (M) in kDa are shown adjacent to the gel. Zone of clearance corresponding to fibrinogenolysis was observed at ~45 kDa in wild-type 9343 (lane 1).

4.8. Discussion

Degradation of fibrinogen by *B. fragilis* was first observed by Houston et al., (2010) and a possible role of secreted proteins/proteases in fibrinogenolysis was predicted. However, Mass-Spectrometry analysis of the region of proteolysis in SDS-fibrinogen zymograms failed to identify any proteases secreted by *B. fragilis*. The clinical significance of fibrinogen involves its proteolytic activation by thrombin to fibrin, the principal component of abscesses formed during infections. Abscess formation is a defence mechanism adopted by the host to sequester bacteria and other infectious agents, thereby curtailing the spread of infections. *B. fragilis* has been frequently isolated from soft tissue abscesses and peritoneal infections. Therefore, the ability to degrade fibrinogen might aid the bacterium in escaping from abscesses and spreading infections and bacteraemia in the human body.

It was observed in the present study that fibrinogen binding by BF1705 is not required for degradation since fibrinogen degradation was detected in both the wild-type NCTC 9343 and the Δ BF1705 concentrated culture supernatants (Figure 3.27). Wild-type cultures incubated with fibrinogen, and analysed by electrophoresis at different time points in growth, were observed to partially degrade the A α -chain of fibrinogen at 3h followed by complete degradation at 48h. On comparison, the B β -chain and γ -chain of fibrinogen were only partially degraded even after 48h (Figure 4.1 A). Previous studies have shown that the A α -, B β - and γ -chains of fibrinogen extend outwards from the central E domain, which contains their N-termini, to form the two globular outer D domains. The C-termini of the B β - and γ -chains are located within the D domains whereas the A α -chains extend outside the globular D domains and then reverse inwards towards the central E domain (Weisel, 2005; Weisel et al., 1985). Thus, the accessible location of A α -chains might play a role in its increased predisposition to proteolytic activation. Neerman-Arbez, (2001) identified that a reduction or absence of the A- α subunit alone, even in the presence of the other two subunits, was sufficient to adversely affect clot formation in a health condition known as afibrinogenaemia. This suggests that proteases secreted by *B. fragilis* can

recognise and cleave fibrinogen which subsequently prevents clot and abscess formation.

The ability of the Δ BF1705 to degrade fibrinogen led us to identify proteases secreted by *B. fragilis* that might be involved in fibrinogen degradation. Identification of a fibrinogen degrading protease might further help us in developing a small molecule to target the protein, and prevent escape of *B. fragilis* from abscesses. Since fibrinogen zymography using concentrated culture supernatants of wild-type *B. fragilis* have revealed lytic bands at ~ 45 kDa and 50 kDa, four genes encoding secreted proteases predicted to be within this range were selected as targets for deletion (4.3).

The role of the selected four secreted proteases in degrading fibrinogen was ruled out by the detection of fibrinogenolysis using single gene deletion mutants as well as multiple gene deletion mutants (Figure 4.8, Figure 4.12 and Figure 4.13). Medium contamination was not likely since the concentrated samples of plain BHI-S medium did not exhibit fibrinogenolytic bands. Although the construction of deletion mutants did not prevent fibrinogenolysis, technically, the present study was the first successful attempt at generating a quadruple gene deletion mutant in *B. fragilis* NCTC 9343. However, these proteases might still be involved in fibrinogen degradation, but the observed zone of lysis might represent more than one protease thereby suggesting redundancy of the secreted proteases. In addition to the four analysed proteases, 27 of the protease/peptidase-encoding genes which possessed a signal peptide were considered for their potential role in fibrinogen degradation. Eight of these 27 secreted peptidases/proteases were selected for expression studies. Selection of the protease/peptidase-encoding genes were based on their molecular weight in relation to fibrinogenolytic bands observed on fibrinogen-SDS zymograms in the present study as well as a previous study by Chen et al., (1995). Expression of these proteases in *E. coli* using pET-100/D-TOPO was followed by functional characterisation of the expressed proteases using fibrinogen zymography (Figure 4.17 and Figure 4.18). No lytic bands were observed and the electrophoresed lysates

were similar to the *E. coli* control culture. Either the selected proteases were not involved in fibrinogenolysis or they were unable to mediate the fibrinogenolytic function when expressed in *E. coli*. Purification of the expressed proteases using nickel columns or the adoption of shuttle vectors to express the proteases in a genetically closer organism might have to be adopted for further analysis.

An interesting observation was made while performing zymography experiments in the two growth media, BHI-S and glucose-DM. Two distinct fibrinogenolytic bands at ~45 kDa and 50 kDa were identified in cultures grown in BHI-S medium, whereas zymography using concentrated culture supernatants of *B. fragilis* NCTC 9343 grown in glucose-DM detected a single zone of lysis at ~ 45 kDa. Therefore, a response to a condition in the glucose-DM is leading to the repression of the gene expression of the larger protease. The difference in the number of fibrinogenolytic bands detected in *B. fragilis* cultures grown in the two different media led to the investigation of factors determining the repressive/activating component in glucose-DM and BHI-S growth media, respectively. Nutrient limitation-mediated protease secretion is already evident in *B. fragilis* (Gibson and Macfarlane, 1988a). Therefore, fibrinogenolysis, which could not be associated with a specific protease/proteases through deletion and expression studies in nutrient-rich growth media could be studied by analysing the transcriptional regulation of proteases under nitrogen-limiting conditions which might be potentially faced by the bacterium in vivo.

Chapter 5 Effect of Nitrogen Limitation on the transcriptome of *Bacteroides fragilis*

5.1. Introduction

This chapter deals with the effect of nitrogen limitation on global gene expression in *B. fragilis*, an opportunistic pathogen. In addition, a transcriptomic analysis would allow us to measure changes in expression of BF1705 and if so, might allow the identification of one or more proteases which were regulated in a similar manner. A difference in fibrinolytic profile was observed when *B. fragilis* was grown on BHI-S and glucose-DM as presented in the previous chapter (4.2). This proteolytic variation might reflect a difference in nitrogen utilisation of the bacterium between the two media. An examination of the transcriptome under nitrogen stress might be able to address the differential expression of proteases.

Previous studies conducted by Gibson and Macfarlane, (1988a, 1988b) observed that the secretion of proteases by *B. fragilis* into the growth medium at the beginning of the stationary phase was induced under nitrogen limiting conditions. This induction was proposed as a mechanism to target proteinaceous substances for replenishing the intracellular nitrogen content. For example, a mucin-degrading glycoprotease was identified in a strain of *B. fragilis* which might perform a similar scavenging role within the gastrointestinal tract (Macfarlane and Gibson, 1991). Since mucin is present on the mucosal surface that lines the intestinal epithelium, its degradation might contribute to the escape of *B. fragilis* into the peritoneal cavity following intestinal injury, thereby suggesting that protease production and pathogenicity are under nitrogen regulation. A similar example has been reported in *E. coli* where inhibitor studies have indicated the specific involvement of a serine protease in starvation-induced degradation. The serine protease inhibitor, Diisopropyl fluorophosphate (DFP), drastically reduced the rate of protein degradation in an *E. coli* K strain starved simultaneously for nitrogen and leucine (reviewed in Miller, 1975). An analogous behaviour was detected in *Bacillus* spp. by Adigüzel et al., (2009) where succedent secretions of collagenolytic, elastolytic and keratinolytic

proteases were detected in peptide-limited carbon-sufficient cultures of two strains of *Bacillus cereus*. Voigt et al., (2007) observed the induction of several transcripts encoding extracellular proteases and peptidases under nitrogen starved conditions in *Bacillus licheniformis*. Proteolytic induction was possibly brought about by deactivation of the global transcriptional repressor CodY due to lowering of the intracellular GTP levels under nutrient limitation and in media compositions that do not support maximal growth rates (Adigüzel et al., 2009; Molle et al., 2003). The intracellular GTP concentration was found to be decreased by a 5-fold when *B. subtilis* was transferred from medium containing casamino acids to a poorer glutamate-only medium (Lopez et al., 1981).

In vivo, the ingested dietary proteins mix with the endogenous luminal proteins derived from sources including gastric and pancreatic secretory products, desquamated intestinal epithelial cells and mucous proteins, in the small intestine lumen. The protein mixture is digested in the small intestine by proteases and peptidases originating from the pancreas and enterocytes (Macfarlane et al., 1986). The oligopeptides and amino acids thus formed are transported from the lumen into the bloodstream through a variety of transporters present in the epithelial membrane of the small intestine (Bröer, 2008). Amino acid metabolism is facilitated during this transport for use by the intestinal mucosa (Davila et al., 2013). Glutamine utilisation by enterocytes through mitochondrial glutaminase leads to the formation of ammonia and glutamate. Utilisation of glutamine in the intestine contributes significantly to the overall production of ammonia in the body (Duée et al., 1995). The microbiota inhabiting the small intestine utilize amino acids provided by dietary and endogenous proteins for protein synthesis, generation of metabolic energy and recycling of reduced cofactors. There is a progressive decrease in the luminal concentration of proteins, peptides and amino acids and increase in bacterial content from the duodenum to the ileum of the small intestine (Adibi and Mercer, 1973). The report by Millward, (2008) addresses the protein sufficiency in older people and describes that protein intakes above the Recommended Dietary Allowances (RDA) may have benefits for adults, especially the elderly population, since protein requirements

increase with age. Therefore, a high proportion of ileal protein digestion, either for mucosal absorption or by microbial utilisation, contributes to reducing the amount of endogenous and alimentary proteins and peptides entering the large intestine (Millward, 2008). A majority of the protein content required by the individual must be absorbed from the small intestine as the large intestine majorly accounts for only water absorption. The reduction in protein content entering the large intestine is also achieved by the presence of bacteria in the small intestine, such as *Lactobacillus johnsonii*, which lack relevant genes for amino acid biosynthesis and rely completely on the supply of exogenous amino acids and peptides for protein synthesis. The nitrogen acquiring functions in this bacterium require the expression of an extracellular protease, oligopeptide transporters, cytoplasmic peptidases and amino-acid permease type transporters (Pridmore et al., 2004). However, in spite of tightly regulated and efficient digestion in the small intestine, substantial amounts of nitrogenous compounds from dietary and endogenous origins pass through the ileo-caecal junction into large intestine. About 2-3 g of nitrogen content enter the large bowel every day, which comprises 10-15% urea, ammonia, and amino acids, 48-51% proteins and 34-42% peptides (Chacko and Cummings, 1988). The colonic microbiota along with residual pancreatic proteases degrade these proteins into peptides and amino acids followed by production of intermediary bacterial metabolites and end-products (Dai et al., 2011). Owing to the heavy bacterial load, transit of luminal material through the colon is considerably slower than the small intestine (Cummings et al., 1978). Due to low pH and high carbohydrate availability in the proximal colon, higher rates of protein degradation, fermentation and net ammonia production are confined to the distal part of the colon (Smith and Macfarlane, 1998). Thus the bacterial species residing in the proximal region of the colon are likely to experience a nitrogen limiting condition. A nitrogen-deprived environment in the colon might induce the secretion of nitrogen scavenging proteases. The specific requirement of *Bacteroides* spp. for ammonia as an effective nitrogen source, might impose a struggle among the gut microbiota for the limited ammonia in the colon (Varel and Bryant, 1974). However, *Bacteroides* spp. are not likely to

face a carbon limiting condition due to their ability to degrade host and plant polysaccharides (Macfarlane and Gibson, 1993). Thus, the proteases induced under nitrogen-limiting carbon-sufficient conditions target the host epithelial barrier thereby posing a threat to the stability of the intestinal microbiome. Species belonging to the *Bacteroides* genus have been reported to secrete proteases in the intestine with presumed elastase activity near the brush border of adsorptive cells (Riepe et al., 1980). Therefore it is of clinical significance to analyse the collagenolytic and fibrinogenolytic potential of *B. fragilis* encoded at the transcriptional level in nitrogen-sufficient and limiting conditions. Also, analysis at the transcriptional level would provide new insights into the regulation of host matrix adhesins such as the fibrinogen binding protein, BF1705. Potential binding of fibronectin in addition to fibrinogen has been discussed in chapter 3, Section 11(3.11). Since very few studies have dealt with the regulation of nitrogen utilisation in *Bacteroides* spp. at the genetic level, a nitrogen-limiting growth medium has been developed by comparing the enzymes involved in nitrogen metabolic pathway encoded in the *B. fragilis* genome to that of *E. coli*, a co-inhabitant of *B. fragilis* in the gastrointestinal tract as well as in polymicrobial infections (Farthmann and Schöffel, 1998). However, there are a few striking differences in the nutritional requirements and metabolism between the two which will be considered and discussed in this chapter. In addition to the regulation of proteases and enzymes involved in nitrogen assimilation, the nutrient limitation might also affect the regulation of stress and antibiotic resistance-related genes (Brown et al., 2014). A comparison of these genes under nutrient limiting and nutrient abundant conditions might provide a better understanding of the behaviour of this opportunistic pathogen within the gastrointestinal tract and also following its accidental release into the peritoneal cavity and bloodstream.

5.1.1. Metabolic regulation of *E. coli* in nitrogen-sufficient and nitrogen-limiting environments

Bacteria sense changes in its intracellular environment, brought about by variation in concentrations of carbon, nitrogen and phosphate nutrient sources and in growth conditions like pH level and oxygen availability. These sensory signals modulate the transcriptional regulatory system, thereby affecting the physiological and morphological changes which enable the organism to adapt effectively for survival.

Nitrogen is an essential element of many macromolecules in a bacterial cell, including proteins, nucleic acids and cell wall components. Although bacteria can assimilate a variety of nitrogen sources, ammonia/ammonium (NH_4^+) is preferred and supports the fastest growth rate (reviewed in Reitzer, 2003). In *E. coli*, ammonia is assimilated using α -ketoglutarate which results in the synthesis of glutamate and glutamine (reviewed in Shimizu, 2013). These two amino acids provide nitrogen for the synthesis of all other nitrogen-containing compounds. 88% of the cellular nitrogen comes from glutamate and the rest from glutamine. Glutamine synthetase, encoded by *glnA*, catalyses the only pathway for glutamine biosynthesis. Glutamate is synthesised by two pathways:

1. GS/GOGAT pathway: The combined action of glutamine synthetase (GS) and glutamate synthase (GOGAT), encoded by *gltBD* (Figure 5.1). GS/GOGAT possesses high affinity for NH_4^+ ($K_M < 0.2\text{mM}$) and is favoured in nitrogen-scarce, energy-sufficient environments since ATP is required for the reaction.

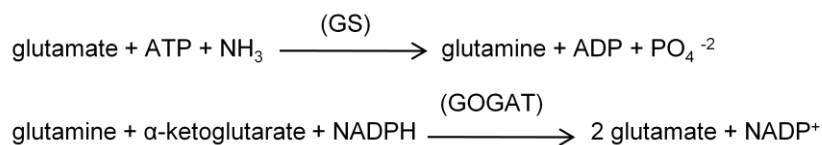


Figure 5.1: GS/GOGAT pathway

Synthesis of glutamine and glutamate using glutamine synthetase (GS) and glutamate synthase (GOGAT) by GS/GOGAT cycle. (Reitzer, 2003)

2. GDH pathway: Glutamate dehydrogenase (GDH) encoded by *gdhA* has low affinity for NH_4^+ ($K_m < 1\text{mM}$) and is dominant in nitrogen-sufficient, energy-limiting conditions (Yan, 2007) (Figure 5.2).

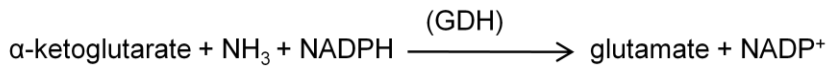


Figure 5.2: GDH pathway

Synthesis of glutamate using glutamate dehydrogenase in the presence of ammonia by GDH cycle. (Reitzer, 2003)

E. coli responds to nitrogen starvation by activating the σ^{54} (*rpoN*) dependent nitrogen regulated response (Ntr). Microarray-based as well as other in silico analyses of σ^{54} -dependent genes indicate that nitrogen limitation affects ~100 genes which facilitate nitrogen scavenging from alternate sources (Reitzer, 2003; Zimmer et al., 2000). It has been proposed that the absolute concentration of glutamine controls the Ntr regulatory cascade (Reitzer, 2003). The Ntr system is composed of three central functioning units:

1. Uridyl transferase or uridylyl-removing enzyme (UTase/UR) encoded by *glnD*
2. P_{II} trimeric nitrogen regulatory protein encoded by *glnB* and its paralogue encoded by *glnK*
3. NtrB/NtrC (NR_{II}/NR_I) two-component system encoded by *glnLG*.

GlnD controls GS activity by adenylation or deadenylation through the bifunctional *glnE* gene product, adenylyltransferase (ATase). Upon nitrogen limitation, UTase covalently modifies P_{II} (GlnB) by addition of a uridine monophosphate (UMP) group at a specific residue. The NtrB/NtrC two-component system and GlnE are the receptors of the modified GlnB signal transduction. NtrC (NR_I) transcriptionally activates σ^{54} -dependent promoters while NtrB (NR_{II}) is a bifunctional protein that can either transfer phosphate to NR_I for its activation or

control dephosphorylation of $\text{NR}_I\text{-P}$ (Shimizu, 2013). GlnB-UMP formed during nitrogen limitation does not complex with NtrB and the histidine kinase activity of NtrB dominates which results in NtrC activation by phosphoryl transfer. NtrC (NR_I) phosphorylation stimulates expression of the *glnALG* operon. The operon is controlled by tandem promoters *glnAP1*, a weak σ^{70} -dependent promoter which is transcriptionally activated by Crp and blocked by NtrC-P and *glnAP2* which is transcribed by RNA polymerase (E σ^{54}) and therefore activated by NtrC-P. Thus *glnAP2* is responsible for *glnA* transcription under nitrogen limitation. The other receptor, GlnE, regulates GS activity by reversible adenylylation. NtrC-P also activates *glnK* (the second P_{II} paralogue) and nitrogen assimilation control (*nac*) promoters under ammonia-depleting conditions. The ability of the P_{II} paralogues to regulate NR_{II} is also regulated by intracellular carbon and nitrogen availability via allosteric control. The *gltBDF* operon which encodes glutamate synthase is bound by the global regulators, Fumarate and nitrate reductase (Fnr) and cAMP reactive protein (Crp). Under nitrogen limitation, *fnr* transcript is upregulated and *crp* downregulated. The regulatory region of *gdhA*, encoding glutamate dehydrogenase, lacks $\text{NR}_I\text{-P}$ binding sites and is transcriptionally repressed by Nac under nitrogen limitation (Figure 5.3). Even though the P_{II} paralogue GlnB is constitutively expressed, it exists in the non-modified deuridylylated state which adenylylates GS (Maheswaran, 2003). Low intracellular glutamine levels signal P_{II} uridylylation and so ATase activity is switched to deadenylylation thereby activating GS (Shimizu, 2013).

Under nitrogen rich conditions, the uridylyl-removing activity of GlnD predominates and deuridylylated P_{II} promotes adenylylation of GS by ATase. Non-modified GlnB forms a complex with NtrB and dephosphorylates the NtrC response regulator thus downregulating *glnA* and *glnK* expression (Figure 5.3).

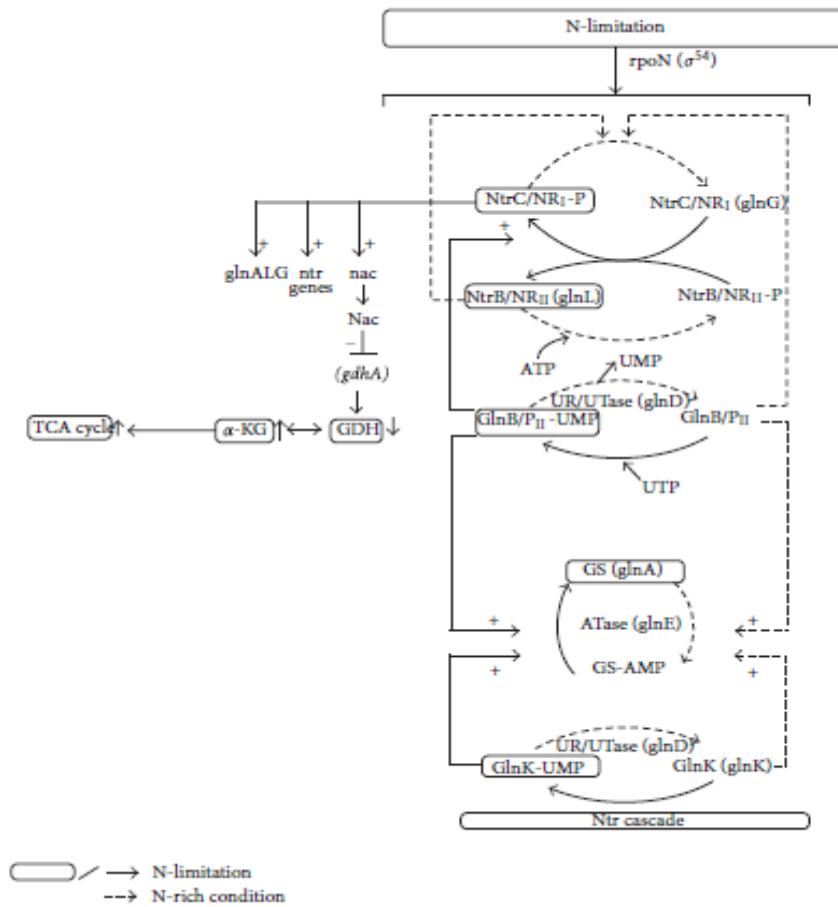


Figure 5.3: Nitrogen assimilation pathway in *E. coli*

Under nitrogen limitation, *E. coli* activates the *rpoN* (σ^{54})-dependent nitrogen regulated response (Ntr). UTase encoded by *glnD* covalently modifies P_{II} (GlnK and GlnB) by addition of a UMP group. GlnB-UMP activates NtrC by phosphoryl transfer which stimulates the *glnALG* operon (*glnA* transcription), other *ntr* genes and *nac*. *gdhA* is transcriptionally repressed by *Nac* under nitrogen limitation. The other receptor of GlnB-UMP, GlnE promotes GS deadenylation by ATase under nitrogen limitation. Under nitrogen rich condition, UR (uridylyl removing) activity of GlnD predominates and the deuridylylated P_{II} promotes adenylation of GS by ATase. UR action on P_{II} (UMP)³ occurs under higher nitrogen concentration (low C/N ratio). Phosphorylation of NtrB dephosphorylates NtrC which blocks *glnA* transcription. GlnB is also allosterically regulated by α -ketoglutarate (KG) thus integrating signals of C/N status. (Shimizu, 2013)

5.1.2. The link between nitrogen and carbon metabolism

The P_{II} regulatory protein and phosphotransferase system (PTS) are considered the central processing units of nitrogen and carbon metabolism, respectively. As mentioned in the previous section (5.1.1), P_{II} protein senses adenylate energy and regulates nitrogen assimilation in conjunction with the Ntr system. The focal point of the carbon and nitrogen pathway interdependence has been the conversion of α -ketoglutarate to glutamate. At high α -ketoglutarate concentration, GlnB binding sites are completely occupied which blocks complex formation of non-uridylylated GlnB with NtrB, thereby blocking phosphorylation and activation of NtrC. Thus GlnB integrates signals of nitrogen status with a signal of central carbon metabolism to coordinate expression of key enzymes of nitrogen assimilation in response to the carbon and nitrogen state. The tandem promoters that initiate *glnALG* transcription are also controlled by carbon and nitrogen states. The activity of the weak, σ^{70} -dependent *glnAP1* promoter requires binding of the cAMP receptor protein (Crp)-cAMP complex to an upstream activating sequence. The complex is formed when cells are grown in the presence of a poor carbon source that is not transported by the PTS system. The complex dissociates during growth on glucose leading to catabolite repression of a large number of genes. The σ^{54} -dependent strong *glnAP2* promoter is activated by NtrC-P which blocks transcription from the weak promoter under nitrogen limiting conditions resulting in high levels of *glnA* transcript from *glnAP2* (Maheswaran, 2003). Elevated levels of cyclic adenosine monophosphate (cAMP) through formation of the CRP-cAMP complex caused a 21-fold downregulation in transcription from *glnAP2* (Tian et al., 2001). All responses of *glnA* and *glnK* expression to carbon or cAMP signals were in agreement with the uridylylation status of P_{II}. The effect of carbon source on NtrC-dependent gene expression is mediated through cAMP-dependent glutamine uptake, affecting the glutamine sensitive UTase-P_{II} signalling system.

5.1.3. Aims

The main aims of this chapter are to identify the effect of nitrogen stress on the transcriptome of *B. fragilis* and to compare the differential expression of genes involved in nitrogen assimilation and scavenging, and stress responses under nitrogen limiting and abundant conditions.

- The first step was to choose a growth medium which limits nitrogen availability
- To determine the time point at which the growth slows, from the growth curve of *B. fragilis* in nitrogen limiting medium
- To determine the differential gene expression of *B. fragilis* under nitrogen replete and limiting conditions through total RNA sequencing

5.2. Choice of nitrogen-limiting medium for the growth of *B. fragilis*

Changes in medium composition were required to determine the amount of nitrogen required to reduce the growth rate, which we are defining as nitrogen limitation, when compared to the growth rate measured in the presence of excess ammonium. A number of previous studies have shown the inability of single free amino acids, nitrate, urea or a complex mixture of L-amino acids to effectively replace ammonia as the sole nitrogen source in *Bacteroides* spp. (Hullah and Blackburn, 1971; Pittman and Bryant, 1964; Stevenson, 1979; Varel and Bryant, 1974; Yamamoto et al., 1984). However, uptake of amino acids and peptides by *Bacteroides rumenicola* and *Bacteroides melaninogenicus* in the presence of ammonia and trypticase-yeast-extract-hemin medium has been observed (Miles et al., 1976; Pittman and Bryant, 1964; Stevenson, 1979). Varel and Bryant, (1974) specifically focused on the inability of *B. fragilis* to use organic sources such as amino acids as the sole nitrogen source in the absence of ammonia. Therefore, the influence of nitrogen content on the growth rate of *B. fragilis* was assessed in glucose-DM where ammonium sulphate ((NH₄)₂ SO₄) served as the source of nitrogen. The requirements for nitrogen-abundant, -sufficient and -limiting conditions were first tested by altering the concentration of ammonium sulphate in the glucose-DM. The amount of ammonium

sulphate present in the normal glucose-DM (0.02 M; ~ 0.58%) was considered as the nitrogen-abundant state. Nitrogen content was increased by incorporating 5 times the normal concentration of ammonium sulphate used (5x) to confirm that nitrogen was not limiting in the normal glucose-DM. Nitrogen-depleting conditions were brought about by reducing the amount of ammonium sulphate by 5 times (-5x) and 10 times (-10x) the normal concentration. Growth rates of *B. fragilis* in the nitrogen abundant and nitrogen limiting media and in the absence of L-cysteine were measured in comparison to the growth in normal glucose-DM, containing 0.02 M (NH₄)₂ SO₄, in a 15h growth curve experiment (2.4.1). No growth was observed in the absence of L-cysteine (data not shown). The inability of *B. fragilis* to grow in the absence of L-cysteine confirms the exclusive role of this amino acid in lowering the redox-potential of the medium and rules out a probable contribution to cellular nitrogen requirement. Analysis of the semi-log plot representing the growth curve of *B. fragilis* under varying concentrations of ammonium sulphate, obtained by plotting OD₆₀₀ values of biological triplicates of cultures corresponding to the three experimental conditions at 45 min intervals, revealed a difference in the growth rate of the bacterium when compared to the growth in normal glucose-DM (Figure 5.4). The doubling times of the populations were determined to be 84.99 min ± 6.56 min, 100 min ± 1.5 min, 80.26 min ± 4.75 min and 100.77 min ± 1.91 min, respectively, in the normal, 5x, -5x and -10x glucose-DM. A difference in biomass yield was also observed, with the lowest yield corresponding to -10x glucose-DM (Figure 5.4). In recent work by Atkinson et al., (2002) and Bren et al., (2013) which examined the growth curve of *E. coli* grown on varying amounts of ammonia, it was observed that the limiting level of nutrient merely limits the final OD₆₀₀ obtained rather than reducing the growth rate. The observation indicated the efficiency of *E. coli* in capturing ammonia, even at low concentrations. It was proposed that the substrate removed from the medium is incorporated into the bacterial biomass with a constant yield and that the total biomass produced is proportional to the OD₆₀₀. This was in agreement with the *B. fragilis* growth curve observed in the present study. However, the abrupt stop in growth from maximal to zero growth which was observed in *E.*

coli cells at low nitrogen levels in the former study was not detected under nitrogen limiting conditions in *B. fragilis* in the present study (Bren et al., 2013). Cells had gradually entered stationary phase from the exponential phase under all four conditions tested (Figure 5.4). A considerable reduction in growth rate was observed in cultures grown in 5x glucose-DM and -10x glucose-DM when compared to the other two conditions. Therefore the reduction in growth rate did not seem to be proportional to the decrease in the initial content of ammonia in the growth medium. However, since the reduction was pronounced as the cultures approach stationary phase, a marked distinction between the slowing of growth due to the onset of stationary phase and due to the nitrogen limitation could not be made. Therefore the medium and conditions for nitrogen limitation had to be further optimised to identify conditions for a reduction in growth rate.

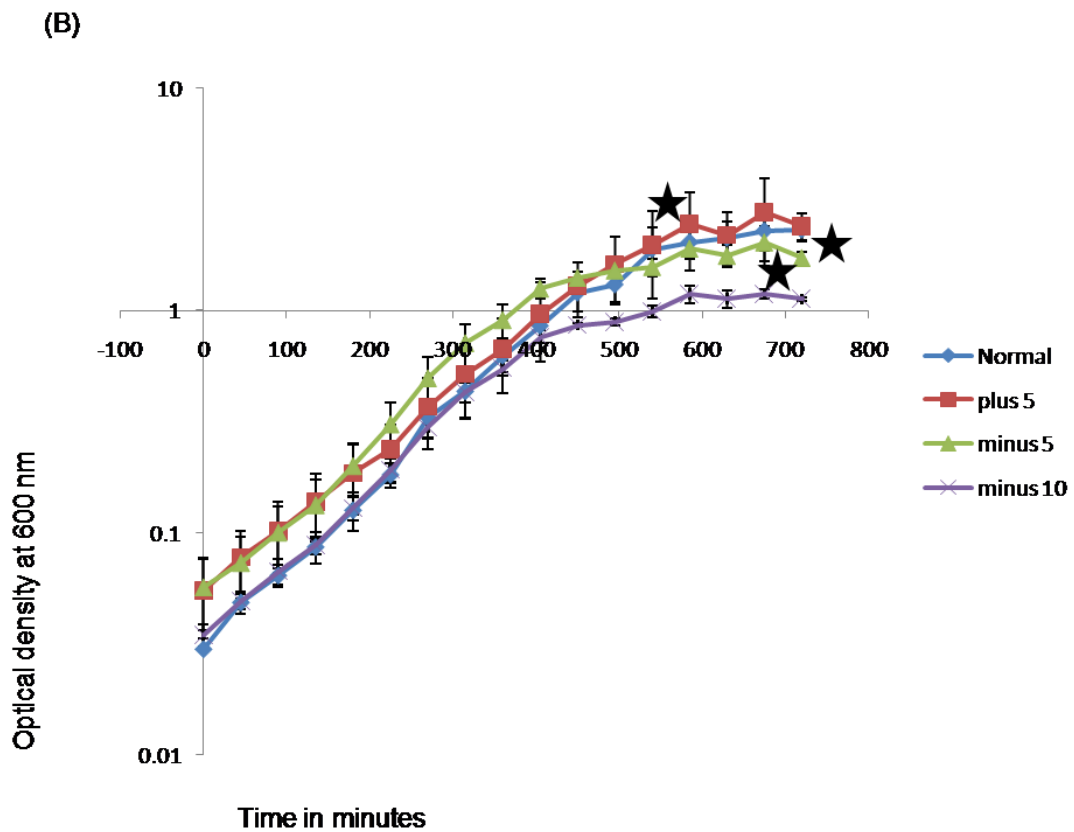


Figure 5.4: Effect of altered nitrogen (ammonium sulphate) content on *B. fragilis* growth

Semi-log plot of a 15h growth curve of *B. fragilis* (Log mean OD₆₀₀ vs time) on glucose-DM containing normal (denoted in blue), +5x (denoted in red), -5x (denoted in green) and -10x (denoted in purple) amounts of ammonium sulphate. Increase in bio-mass yield is observed upon increasing the concentration of ammonium sulphate from -10x to +5x, indicated by the asterisk mark. However, no significant difference in yield is observed between normal and +5x concentrations of ammonium sulphate.

5.2.1. Effect of glutamine as the sole nitrogen source on growth rate

The similarities between the growth curves of cultures grown in the defined media containing ammonia as the nitrogen source made defining nitrogen limitation problematic. Therefore, we considered alternative nitrogen sources that could substitute for ammonia in the glucose-DM. Although studies on *B. fragilis* have emphasised the requirement of ammonia as the major nitrogen source, a minimal medium with glutamine as the sole nitrogen source has been referred to as a nitrogen-limiting condition for *E. coli* growth (Maheswaran, 2003). Atkinson et al., (2002) also studied gene regulation in *E. coli* when grown in a medium with glutamine as the nitrogen source and observed a reduction in growth rate in mid-log phase in the glutamine-grown *E. coli* cultures. The nitrogen-limiting condition was consistent with the activation of NtrC-dependent gene expression involved in the assimilation of nitrogen in the cells grown in this medium. The studies on *E. coli* coupled with the fact that glutamine and glutamate form the intermediary products in the bacterial nitrogen assimilation pathway, and the P_{II} regulatory protein is signalled by the intracellular levels of glutamine, inspired us to assess the individual effect of these two amino acids as the sole nitrogen source on the growth of *B. fragilis*.

The ammonium sulphate in the normal glucose-DM was replaced with either 0.1% glutamine or 0.1% glutamic acid. Overnight cultures of NCTC 9343 in glutamic acid-containing glucose-DM (referred to hereafter as GCM) did not grow even after a 24h anaerobic incubation when compared to glutamine-containing glucose-DM (referred to hereafter as GNM) which had reached growth saturation in a 12h incubation period. The growth conditions were checked by inoculating normal glucose-DM as a control which exhibited saturated growth and a better biomass yield when compared to GNM, an indication of the nutrient-abundant state. Lack of growth observed on GCM indicates the inability of *B. fragilis* to utilise glutamic acid as the sole nitrogen source and the observation is consistent with previous studies carried out on this species which had confirmed that amino acids can only supplement the growth of *Bacteroides* in an ammonia-containing medium. The ability of *B. fragilis* to utilise glutamine as the sole nitrogen source in an ammonia-

deficient medium has not previously been investigated. However, increased uptake of amino acids and peptides by *Bacteroides melaninogenicus* in a glutamine-added trypticase medium has been studied (Lev, 1980). Therefore, to confirm the initial findings and to validate the observations on GCM, we compared the growth of 12h cultures of *B. fragilis* NCTC 9343 and *E. coli* MG1655 in normal glucose-DM, GNM and GCM in terms of OD₆₀₀. The ability of *E. coli* to utilise amino acids as the sole nitrogen source is already known from previous literature (Atkinson et al., 2002; Goux et al., 1996). The glutamic acid content in the glucose-DM (GCM) was doubled to 0.2% for both the strains. It was observed that the 12h aerobic and anaerobic cultures of MG1655 and NCTC 9343, respectively, on GNM attained final OD₆₀₀ values of 1.5 and 2.0 while growth of MG1655 on GCM reached a lower density with an OD₆₀₀ of 1.23 (Figure 5.5 A and B). There was no growth observed for 12h NCTC 9343 culture on GCM. Thus, a lower biomass yield was observed when *E. coli* was grown on a nutrient-limiting GCM. However, it is evident from the OD₆₀₀ values that *E. coli* can utilise glutamic acid whereas *B. fragilis* cannot. Even though, the genome of *B. fragilis* NCTC 9343 possesses *glnA* (glutamine synthetase) and *gdhB* (glutamate dehydrogenase) homologues for glutamine and glutamic acid utilisation, respectively, the bacterium can't use glutamic acid when supplied exogenously. The possible glutamine utilisation through the GS/GOGAT cycle for the production of glutamate in GNM and the failure to generate glutamine in the absence of ammonia in GCM by *B. fragilis* might be the reason for the observed differential utilisation. The inability of *B. fragilis* to grow on GCM even in the presence of L-cysteine indicates that this bacterium does not utilise cysteine as the sole nitrogen source in the absence of ammonia, unlike *B. thetaiotaomicron* (Atherly and Ziemer, 2014). Nevertheless, the observations made in the current study confirm that ammonia is the preferred source of nitrogen for bacterial growth (Figure 5.6 A and B), as demonstrated by the attainment of an OD₆₀₀ of 2.5 and 3.0 after a 12h growth of *E. coli* and *B. fragilis*, respectively, on normal glucose-DM containing ammonium sulphate.

To determine changes in gene expression that occur during growth under nitrogen limiting conditions, an RNA sequencing approach was used to examine the transcriptome. Before samples were taken for RNA preparation, we first measured the growth rate in GNM to establish the OD_{600} at which the culture became nitrogen-limited. The growth of NCTC 9343 was monitored in GNM containing 0.1% glutamine in comparison to that of normal glucose-DM containing ~ 0.58% ammonium sulphate in a 15h growth curve measured at hourly intervals. A marked difference in growth rate was observed by late-log phase between the two media. To confirm reproducibility, the experiment was repeated twice and the results showed good agreement. The final experiment was performed in six biological replicates and the mean OD_{600} values were plotted against time (Figure 5.7). In early-log phase (~ 0.4 OD_{600}), the mass doubling-times (T_d) in the normal glucose-DM and GNM were $78 \text{ min} \pm 4.2 \text{ min}$ and $90 \text{ min} \pm 6 \text{ min}$, respectively. The growth curve for GNM was observed to be 'biphasic', where the initial doubling time of $90 \text{ min} \pm 6 \text{ min}$ dropped by late-log phase (0.8 OD_{600}) to a T_d of $390 \text{ min} \pm 20 \text{ min}$ in GNM (Figure 5.7). The observed 5h decrease in T_d is indicative of a nitrogen limitation in the medium. For transcriptomic analysis, RNA samples were extracted from the 6 biological replicates of normal glucose-DM cultures grown to early-log phase ~ 0.4 OD_{600} (N set) and from replicates of GNM cultures grown to a similar ~0.4 OD_{600} (G1 set) as well as a time point in late-log phase, 0.8 OD_{600} , where the reduced growth rate was observed (GA set). Analysis of the whole transcriptome under the three experimental conditions would provide a detailed description of the regulation of genes encoding nitrogen scavenging proteins, nutrient transport systems as well as stress-related genes under normal, nitrogen-limiting and growth-limiting conditions.

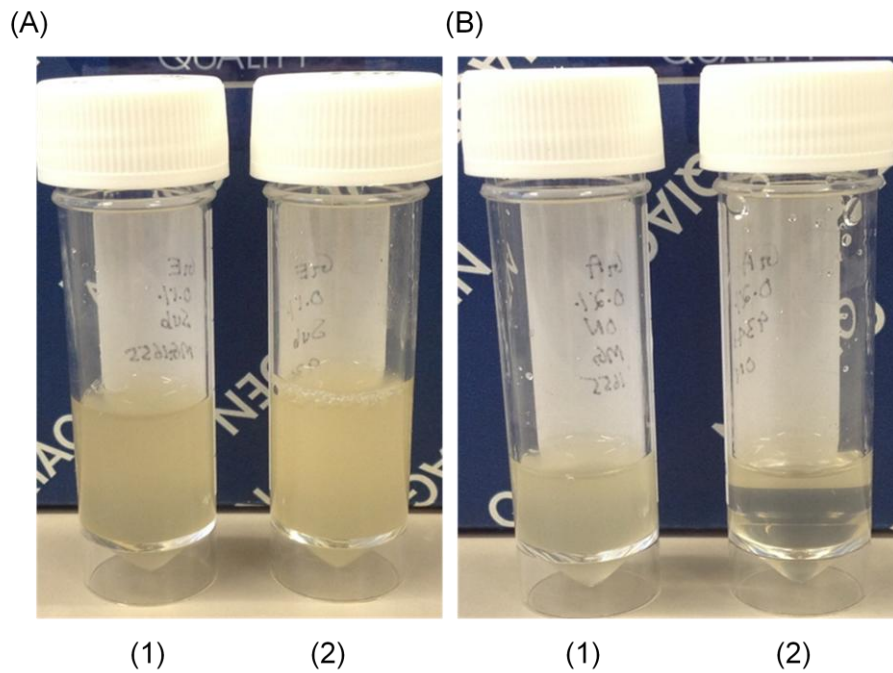


Figure 5.5: Comparison of bacterial growth in glutamine-containing glucose-DM and glutamic acid-containing glucose-DM

Overnight cultures of *E. coli* MG1655 and *B. fragilis* NCTC 9343 grown on 0.1% glutamine-containing glucose-DM (GNM) and 0.2% glutamic acid-containing glucose-DM (GCM). (A) MG1655 has attained OD_{600} 1.5 (1) and 9343 attained OD_{600} 2.0 (2), respectively on GNM. (B) MG1655 has attained sub-optimal growth of OD_{600} 1.23 (1) while no growth is observed for NCTC 9343 (2) on GCM.

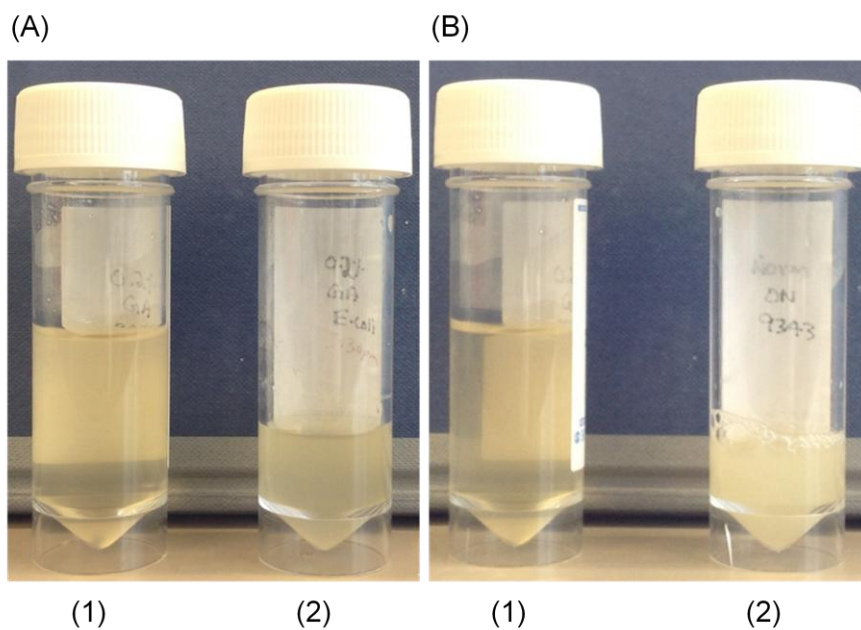


Figure 5.6: Comparison of *B. fragilis* growth in glutamic acid-containing glucose-DM and normal glucose-DM

Overnight cultures of *B. fragilis* NCTC 9343 grown on 0.2% glutamic acid-containing glucose-DM (GCM), *E. coli* grown on GCM and *B. fragilis* grown on normal glucose-DM. (A) No growth is observed for *B. fragilis* (1) and *E. coli* has attained a final OD_{600} of 1.19 (2) on GCM. (B) No growth is observed for *B. fragilis* on GCM (1) and has reached saturation at OD_{600} 2.21 on normal glucose-DM (2).

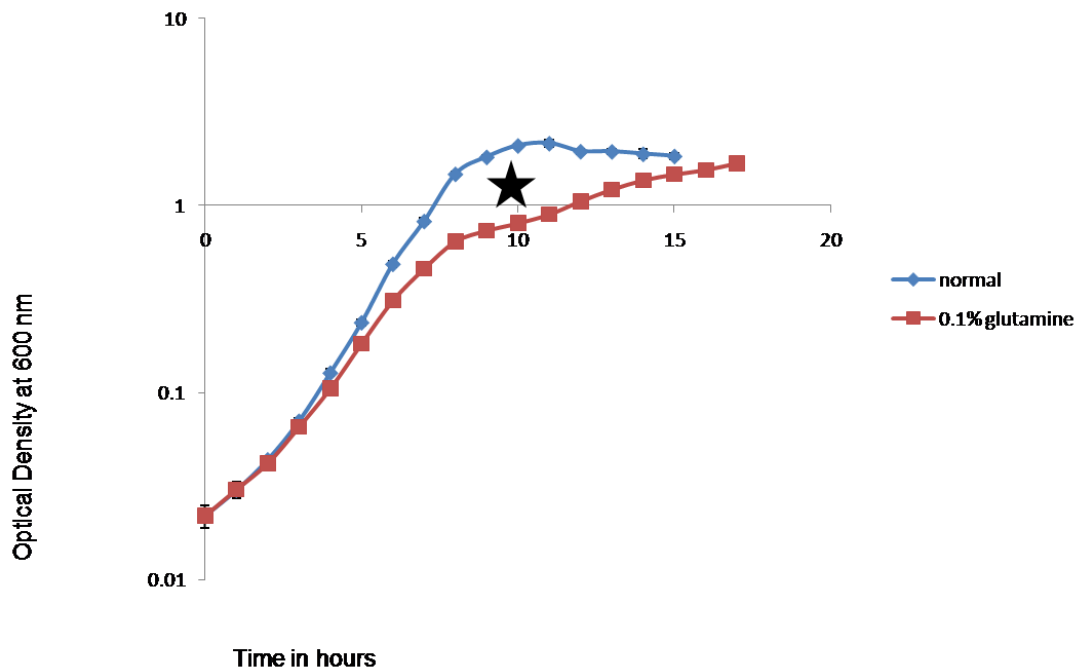


Figure 5.7: Effect of nitrogen limitation on *B. fragilis* growth

Semi-log plot of a growth curve of *B. fragilis* (mean OD₆₀₀ vs time) on glucose-DM containing the normal concentration of ammonium sulphate (~0.58% and 0.02 M) (denoted by blue diamond) and 0.1% glutamine (denoted by red box). Reduction in growth rate is observed at late-log phase (0.8 OD₆₀₀) in 0.1% glutamine-containing glucose-DM, indicated by the asterisk mark.

5.3. RNA sequencing

The most suitable method for analysing and comparing the transcriptional upregulation or downregulation of genes associated with nutrient metabolism, stress, drug resistance and virulence under nutrient-abundant and nutrient-limiting conditions was RNA sequencing. RNA-Seq, also known as whole transcriptome shotgun sequencing, depicts a snapshot of RNA presence and quantity from a genome at a given moment in time by using the capabilities of next-generation sequencing. Unlike microarray studies, RNA sequencing does not require prior knowledge about the genome. Also, an increased dynamic range is observed in RNA-seq reads with an overall higher sensitivity, reliability and reproducibility levels when compared to that of microarrays (Wang et al., 2009). Therefore in the present study, RNA samples derived from five biological replicates of cultures belonging to the nitrogen abundant glucose-DM at ~ 0.4 OD₆₀₀ (referred to hereafter as N set) and nitrogen limiting GNM at ~ 0.4 OD₆₀₀ (referred to hereafter as G1set) and 0.8 OD₆₀₀ (referred to hereafter as GA set) were subjected to RNA sequencing (2.6). The process of RNA-sequencing proceeds through a multitude of laboratory and programming steps starting from the raw RNA extract to the final transcript reads, of which the differential expression can be compared and analysed. Each step needs to be performed carefully and with accuracy in order to get final, good quality reads that have an optimum genome coverage. Therefore, the quality control steps are essential in determining the overall success of RNA-Seq (Figure 5.8).

Overview of RNA Sequencing

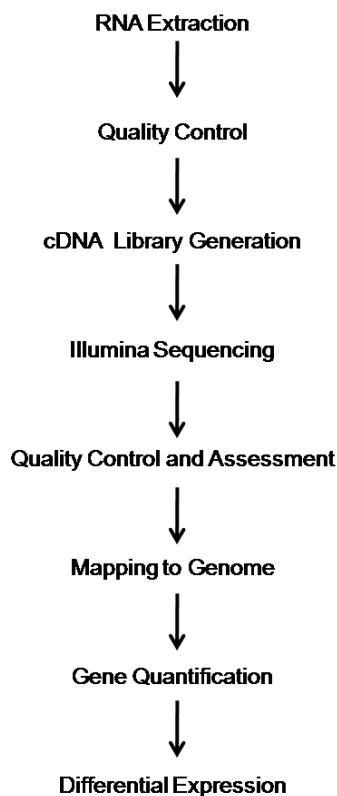


Figure 5.8: Schematic of RNA-Seq workflow

The RNA-Seq work flow initially involves extraction of the starting material, RNA, at the desired time points followed by its quality control to confirm its suitability for library generation and sequencing work. Stranded libraries are then constructed for each of the RNA samples. The pooled libraries are subjected to RNA sequencing using the HiSeq v4 (Illumina) instrument to generate 50 base paired-end reads. The read pairs corresponding to each sample are then examined by FastQC module to assess the quality of library construction and sequencing before processing and mapping the read pairs to the genome. Once mapped, the reads are assigned to features by gene quantification and finally the read-count data matrix is subjected to differential expression using the appropriate software tool and parameters.

5.3.1. Validation of RNA samples prior to sequencing

Since the integrity of the starting RNA material is significant in determining the quality and yield of subsequent steps such as library preparation and sequencing, the quality of the extracted RNA samples was first analysed. Nanodrop readings of all eighteen (3x6) RNA samples were in the range of 220-360 ng/ μ l, with corresponding 260/280 ratios of \sim 2. The best 5, in terms of concentration, from each experimental condition was selected for RNA sequencing (Table 5.1). Integrity of RNA samples were checked by agarose gel electrophoresis of formamide-treated aliquots (2.6.1.1). Two distinct bands corresponding to the 23S and 16S ribosomal RNA (rRNA) were observed (Figure 5.9). After examining the concentration, purity and integrity of the RNA samples and confirming that they met the specifications for sequencing, the samples were sent to the Edinburgh Genomics facility for whole transcriptome sequencing. A further quality control analysis was performed on the RNA samples at the facility using the RNA Integrity Number (RIN) software tool in combination with the 2100 Bioanalyzer (Agilent Technologies). On a scale of 1-10 RIN, '10' stands for a perfect RNA sample without any degradation products whereas '1' marks a completely degraded sample. All the 15 RNA samples were of good integrity since the RINs were between 9 and 10 (Figure 5.10 A and B).

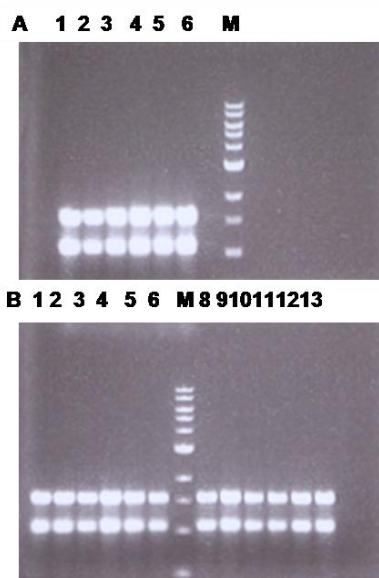


Figure 5.9: RNA extracted from *B. fragilis* grown on normal glucose-DM and GNM

Agarose gel analysis of RNA samples extracted from *B. fragilis* grown on (A) normal glucose-DM (lanes 1-6) to ~0.4 OD₆₀₀ and (B) glutamine-containing glucose-DM at early-log phase ~0.4 OD₆₀₀ (lanes 1-6) and late-log phase 0.8 OD₆₀₀ (lanes 8-13). RNA samples were intact and two distinct bands corresponding to 16S and 23S rRNAs were observed.

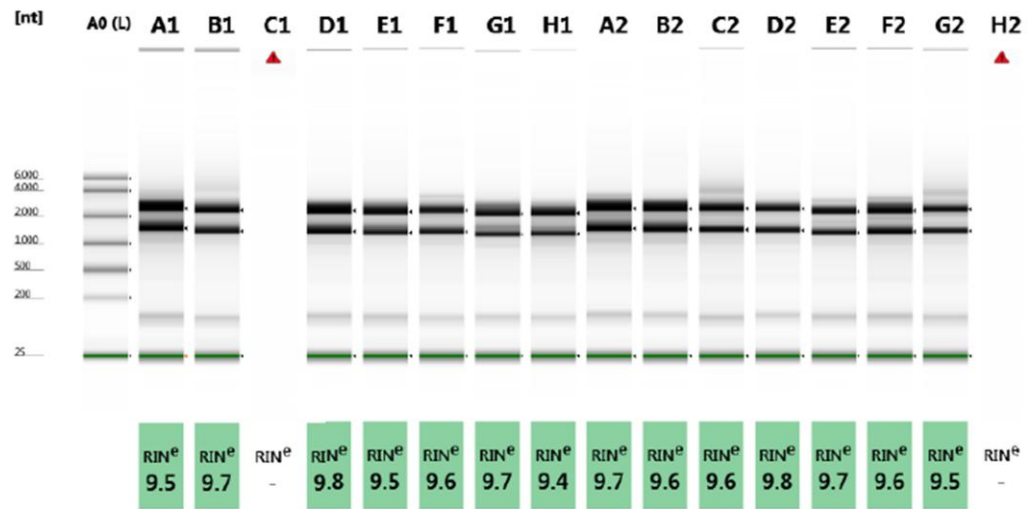
Table 5.1: Nanodrop-based quantification of RNA samples prior to sequencing

Nanodrop-based quantification of RNA samples selected for RNA sequencing. A 260/280 are above 2 and the 15 samples exhibited optimum concentration (> 150 ng/μl) for sequencing.

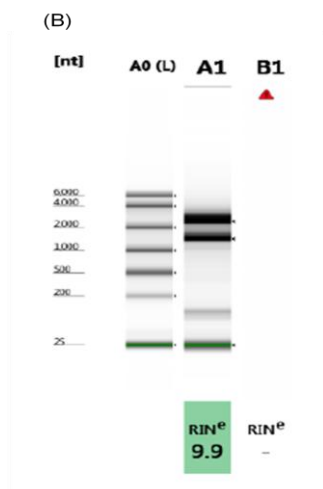
Samples (~0.4 OD normal media) N set	conc (ng/μl)	Absorbance 260/280	Absorbance 260/230	Total yield (μg)
N1	360.2	2.11	2.25	7.2
N3	214.6	2.12	2.23	4.3
N4	278.4	2.11	2.21	5.5
N5	316.7	2.1	2.07	6.3
N6	333.8	2.15	2.25	6.7
Samples (~0.4 OD glutamine media) G1 set				
Gu1	298.9	2.18	1.49	5.9

Gu2	278.2	2.17	2.14	5.5
Gu3	220.6	2.07	1.82	4.4
Gu4	309.4	2.15	1.81	6.2
Gu5	314.1	1.99	1.27	6.3
Samples (~0.8 OD glutamine media) GA set				
GuB	286.1	2.14	2.25	5.7
GuC	165.2	2.13	2.28	3.3
GuD	178.8	2.02	1.64	3.5
GuE	219.3	2.14	2.27	4.4
GuF	222	2.03	1.63	4.4

(A)



Well	RIN ^e	16S/23S (Area)	16S/23S (Height)	Conc. [ng/μl]	Sample Description	Alert	Observations
A0	-	-	-	116			Ladder
A1	9.5	0.9	1.2	282	GB9638		
B1	9.7	1.3	1.2	182	GB9639		
C1	-	-	-		GB9640		Marker(s) not detected
D1	9.8	1.4	1.2	212	GB9641		
E1	9.5	1.9	1.2	168	GB9642		
F1	9.6	1.2	1.0	143	GB9643		
G1	9.7	1.3	1.1	160	GB9644		
H1	9.4	1.8	1.2	163	GB9645		
A2	9.7	1.2	1.1	250	GB9646		
B2	9.6	0.9	1.2	214	GB9647		
C2	9.6	1.4	1.1	193	GB9648		
D2	9.8	1.3	1.1	138	GB9649		
E2	9.7	1.3	1.0	146	GB9650		
F2	9.6	0.9	1.0	188	GB9651		
G2	9.5	1.5	1.1	148	GB9652		
H2	-	-	-		Blank		Marker(s) not detected



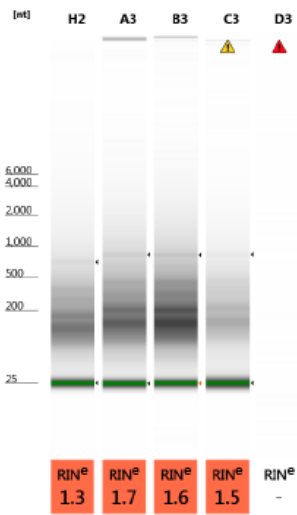
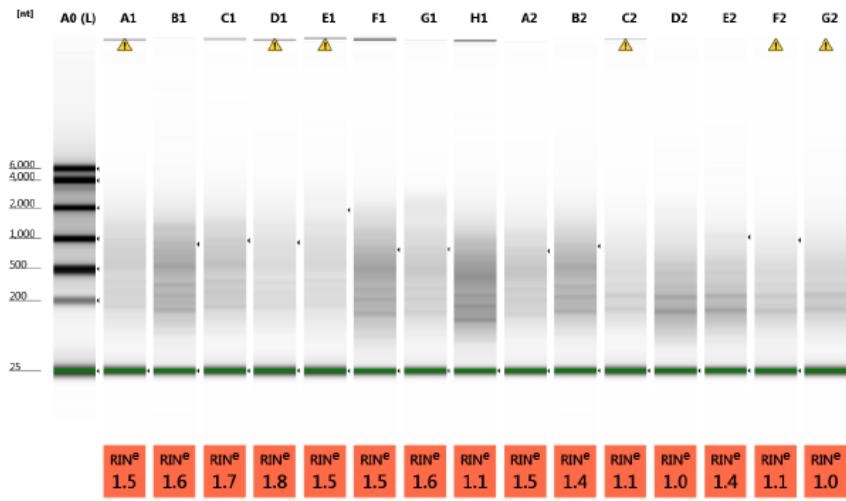
Well	RINe	16S/23S (Area)	16S/23S (Height)	Conc. [ng/μl]	Sample Description	Alert	Observations
A0	-	-	-	116			Ladder
A1	9.9	1.2	1.0	186	GB9640		
B1	-	-	-			⚠	Marker(s) not detected

Figure 5.10: RIN-based quantification of the extracted RNA samples

(A) and (B) RIN (RNA Integrity Number) values of all the RNA samples submitted for sequencing were above 9 and concentrations were above 100 ng/μl, respectively.

The majority, amounting to 50-80%, of the extracted and quantified total RNA content comprised of ribosomal RNAs, as was evident from the distinct bands observed by gel electrophoresis analysis (Figure 5.9). Therefore specialised methods were required to enrich the messenger RNA (mRNA) transcriptome for analysis. In the present study, the quantified RNA samples were subjected to Ribo-Zero (Illumina) treatment for the depletion of rRNA molecules through hybridization to magnetic bead-linked complementary oligonucleotides, thereby increasing the mRNA coverage for the subsequent complementary DNA (cDNA) library construction. Ribo-Zero kit performs well only on DNase-treated and purified RNA samples (2.6.1). The proficiency of the Ribo-Zero kit in rRNA depletion has been described in a study by Giannoukos et al., (2012) where different commercial rRNA removal kits were compared. Upto 99% removal of the large rRNA (23S and 16S)

as well as > 85% to 95% removal of 5S rRNA could be achieved through the use of Ribo-Zero kit. RIN software analysis was performed a second time to analyse the quality of the rRNA depleted samples. Since the RIN tool measures RNA integrity based on the presence of rRNA bands, poor RINs after Ribo-Zero treatment indicate successful removal of rRNAs. Consistently, RIN values of all 15 RNA samples were between 1.0 -1.8 after one round of Ribo-Zero treatment. Also, the distinct rRNA bands were missing from the electrophoresed agarose gel images of the samples post Ribo-Zero treatment (Figure 5.11).



Well	RIN ^e	23S/16S (Area)	Conc. [pg/ul]	Sample Description	Alert	Observations
A0	-	-	3850			Ladder
A1	1.5	-	874	GB9638_Ribo_Ampure	⚠	RNA concentration outside recommended range for RINe
B1	1.6	-	1800	GB9639_Ribo_Ampure		
C1	1.7	-	1070	GB9640_Ribo_Ampure		
D1	1.8	-	677	GB9641_Ribo_Ampure	⚠	RNA concentration outside recommended range for RINe
E1	1.5	-	722	GB9642_Ribo_Ampure	⚠	RNA concentration outside recommended range for RINe
F1	1.5	-	2040	GB9643_Ribo_Ampure		
G1	1.6	-	1040	GB9644_Ribo_Ampure		
H1	1.1	-	2400	GB9645_Ribo_Ampure		
A2	1.5	-	1030	GB9646_Ribo_Ampure		
B2	1.4	-	1600	GB9647_Ribo_Ampure		
C2	1.1	-	654	GB9648_Ribo_Ampure	⚠	RNA concentration outside recommended range for RINe





D2	1.0	-	1360	GB9649_Ribo_Ampure		
E2	1.4	-	1160	GB9650_Ribo_Ampure		
F2	1.1	-	755	GB9651_Ribo_Ampure		RNA concentration outside recommended range for RINe
G2	1.0	-	774	GB9652_Ribo_Ampure		RNA concentration outside recommended range for RINe
H2	1.3	-	1700	Ambion_KT_Ribo_Ampure		
A3	1.7	-	2200	Ambion_SW1_Ribo_Ampure		
B3	1.6	-	2770	Ambion_SW2_Ribo_Ampure		
C3	1.5	-	813	Ambion_SW3_Ribo_Ampure		RNA concentration outside recommended range for RINe
D3	-	-		Empty		Marker(s) not detected

Figure 5.11: RIN-based quantification of the RNA samples post rRNA depletion

Agarose gel analysis, RIN values and concentration ($\mu\text{g}/\mu\text{l}$) of the RNA samples following Ribo-Zero treatment to remove rRNAs. No distinct bands are observed on the agarose gel and the RIN values have reduced.

5.3.2. Library construction and Illumina-based Sequencing

The total RNA, following rRNA depletion, was converted into a library of cDNA molecules of known standard origin which facilitates downstream sequencing processes, such as cluster generation (TruSeq stranded total RNA-seq library kit, Illumina). Prior to cDNA library construction, the total RNA (1 µg/sample) from the 15 samples was fragmented into small pieces using divalent cations under elevated temperature. Cleaved RNA fragments were then copied into first strand cDNA using reverse transcriptase enzyme and random primers, followed by second strand cDNA synthesis using DNA polymerase I and RNase H. The strand specificity in cDNA library generation is achieved during second strand synthesis when the RNA template is removed and the second cDNA strand is constructed by incorporating dUTP in place of dTTP. Incorporation of dUTP quenches the second strand during amplification because the polymerase does not add past this nucleotide. Stranded libraries provide strand-specific information for each of the transcripts. The blunt ended double-stranded (ds) cDNA fragments are then prepared for hybridisation onto flow cell lanes of the sequencer by ligating multiple indexing adapters (2.6.2) to the adenylated 3' ends of the ds cDNA fragments. To enrich those DNA fragments that are adapter-ligated at both ends and to amplify the amount of DNA in the cDNA library, the fragments were subjected to 10 cycles of PCR (TruSeq stranded total RNA-seq library kit, Illumina). Finally, the cDNA libraries generated from the 15 RNA samples were quantified, normalised to 10 nM concentration and pooled in equal volumes for the cluster generation step in sequencing.

The pooled libraries were sequenced in a single lane of HiSeq v4 High Output (Illumina) using the 'sequence-by-synthesis' approach of the Solexa-Illumina sequencing platform. In brief, it involves the incorporation of novel reversible terminator nucleotides for the four bases, each labelled with a different fluorescent dye, to the cDNA strand followed by detection of the added bases at each step. Initially, the adapter-ligated DNA fragments in the cDNA library were denatured and one end was immobilised onto a solid support. The surface of the solid support is densely coated with adapters and complementary adapters. The complementary

adapters hybridise to the free end of the immobilised single stranded fragment leading to the formation of a bridge structure. Adapters on the surface act as primers for the subsequent bridge PCR amplification (cluster generation). The reaction mixes for sequencing reactions and DNA synthesis comprise the four reversible terminator nucleotides, each uniquely labelled with a fluorescent dye. On incorporation into the DNA strand, the terminator nucleotides as well as its position on the solid support is detected and identified via its fluorescent dye using charge-coupled device (CCD) camera. Subsequently, the terminator group at the 3' end of the base and fluorescent dye are removed from the base and the synthesis cycle is repeated (Ansorge, 2009) (Figure 5.12). Each cluster progressively leads to over a million amplified copies of the original single-stranded fragment, the sequencing reaction from which is sufficient for reporting incorporated bases at the required signal intensity for detection during sequencing (Mardis, 2008). In the present study, the sequence read length was specified at 50 nucleotides and paired-end reads were achieved by modifying the clusters in situ to regenerate the template for the paired read after the first read. The same clusters are then sequenced using a second sequencing primer to generate the second read (2.6.2). The amount of duplication in the libraries could be evaluated with paired-end reads, unlike single-end reads. The ability to sequence atleast 40 million clusters in parallel defines the high throughput feature of Illumina sequencers. A total of 220 million + 220 million reads was generated from a single lane of HiSeq v4 High Output (Illumina) in the present study. R1 and R2 read pairs from each of the 15 RNA samples were obtained as fastq files from the Edinburgh Genomics facility. Five files in fastq format belonging to reads from RNA samples of N set (taken from ~0.4 OD₆₀₀ cultures in normal glucose-DM) were referred to as N1_1, N1_2 and N3(_1 and _2) - N6(_1 and _2). Fastq files containing read pairs from RNA samples of G1 set (taken from ~0.4 OD₆₀₀ cultures in nitrogen-limiting glutamine medium) were referred as Gu1(_1 and _2) - Gu5(_1 and _2). In a similar manner read pair files from RNA samples of GA set (extracted from 0.8 OD₆₀₀ cultures in nitrogen-limiting glutamine medium) were given the nomenclature GuB(_1 and _2) - GuF(_1 and _2).

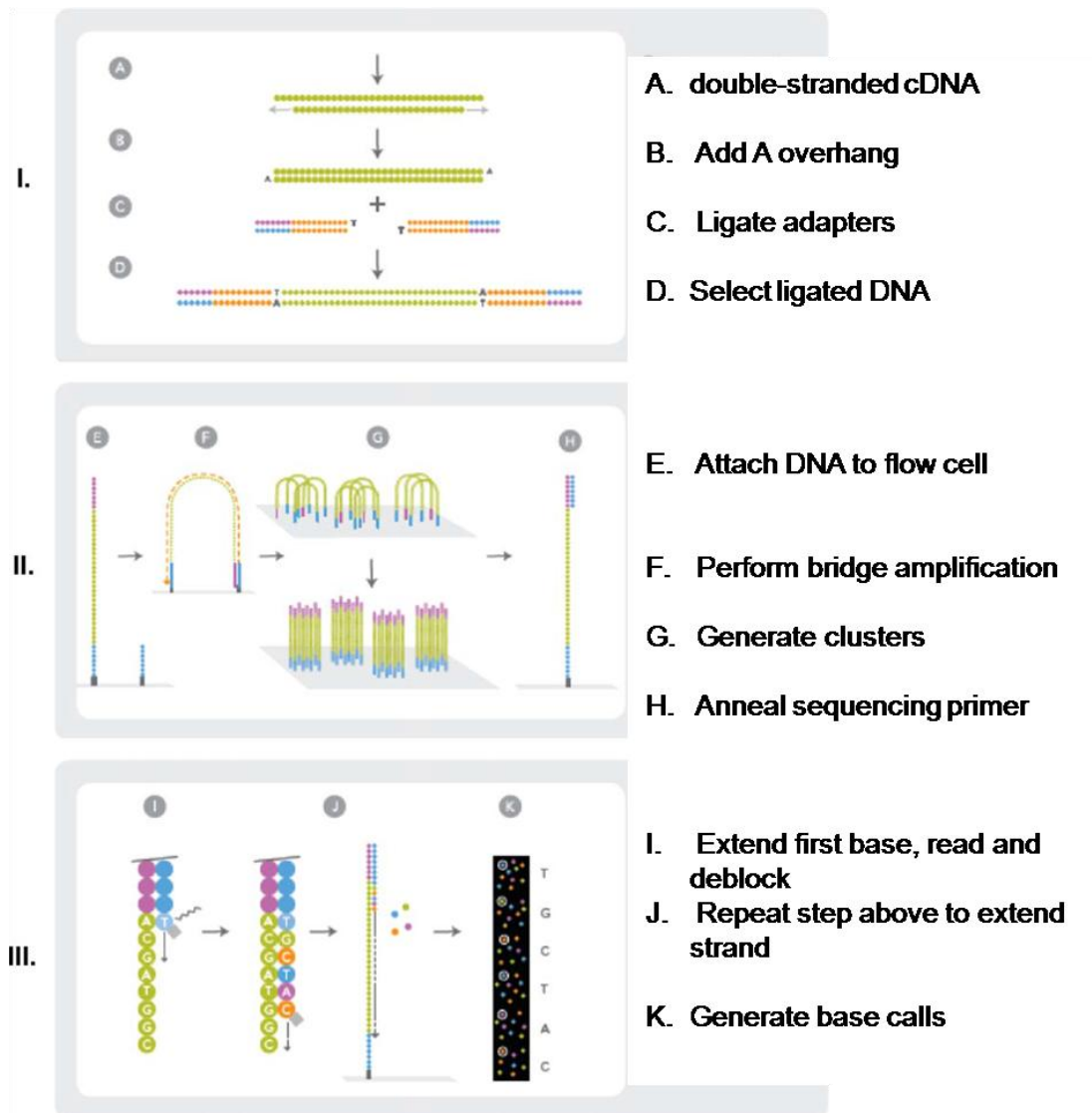


Figure 5.12: Outline of the Illumina (Solexa) sequencing platform

I. ds cDNA (A) synthesised from RNA fragments is adeylated at the 3' end (B) for adapter ligation (C). Adapter ligated DNA fragments are enriched by PCR (D). II. Denatured cDNA is hybridised to the flow cell (E) and bridge PCR is performed (F) to generate clusters (G). Sequencing primer is annealed for sequencing reactions (H). III. The first terminator nucleotide is incorporated into the base, read and removed (I) and the step is repeated to extend strand (J). Base calls are generated to the desired length (K). (Adapted from (Ansorge, 2009))

5.3.3. Validation of RNA-Seq reads post RNA sequencing

Quality of the generated read pairs was checked, processed and mapped against the reference *B. fragilis* NCTC 9343 genome using appropriate software tools and parameters for analysing regulation at the transcriptional level.

5.3.3.1. Quality Control and Assessment

To determine the quality of the read pairs in fastq files, FastQC v0.11.3 was performed initially (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The FastQC analysis is performed by a series of modules on each sequence read pair and the output is provided as an interactive display which summarises the modules that were run and provides a quick evaluation to check whether the results of the module seemed entirely normal (green tick), slightly abnormal (orange exclamation mark) or very unusual (red cross), an example of which is shown in Figure 5.13. Since the FastQC reports of all the 15 read pair files were similar, except for a few modules which have been discussed separately, the analysis of the FastQC read pair belonging to N set is used as an example (Figure 5.14 A-L).

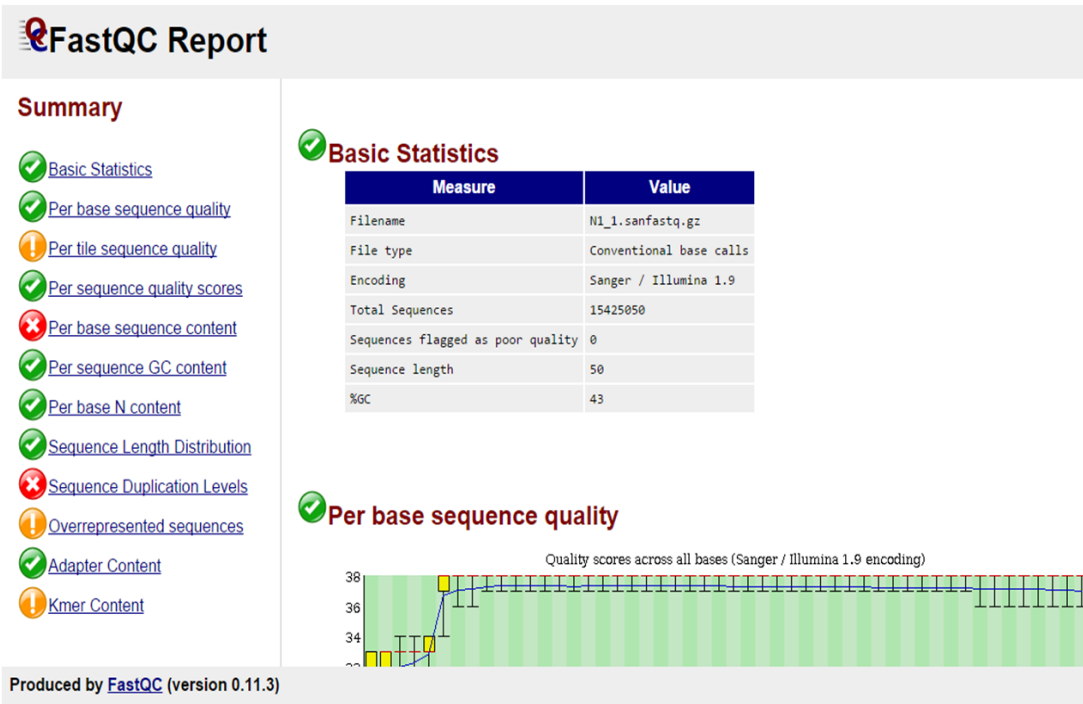


Figure 5.13: Screenshot providing the summary of the FastQC report of a read pair file

Summary of the N1_1 read pair file provides a quick evaluation of the results of the different modules in the FastQC analysis which identifies whether the results were normal (denoted by a green tick), slightly abnormal (denoted by an orange exclamation mark) or very unusual (denoted by a red cross).

i. Basic Statistics

The statistics table for each read pair file provides details of the filename, total sequence reads contained, sequences flagged as poor quality, sequence length and %GC content. None of the reads analysed in the present study contained sequences flagged as poor quality. The sequence length, 50, was identical for all read pairs as observed for N1_1 (Figure 5.14 A).

(A) Basic Statistics

Measure	Value
Filename	N1_1.sanfastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	15425050
Sequences flagged as poor quality	0
Sequence length	50
%GC	43

Figure 5.14: Display of results of the modules performed by the FastQC tool on N1_1 fastq file

(A) Basic statistics detailing the type and contents of the file. There are no sequences flagged as poor quality.

ii. Per Base Sequence Quality

This plot provides an overview of the range of quality values across all bases at each position in the FastQ file. For each position, a BoxWhisker type plot is drawn. The elements of the plot are as follows:

- The central red line is the median value
- The yellow box represents the inter-quartile range (25-75%)
- The upper and lower whiskers represent the 10% and 90% points
- The blue line represents the mean quality

The y-axis on the graph shows the quality scores. The higher the score the better the base call. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). Similar to N1_1, all the read pairs were confined within the green area which indicates a very good quality call in the present analysis (Figure 5.14 B). However, the base calls might fall into the orange area towards the end of the read owing to degradation as the run progresses.

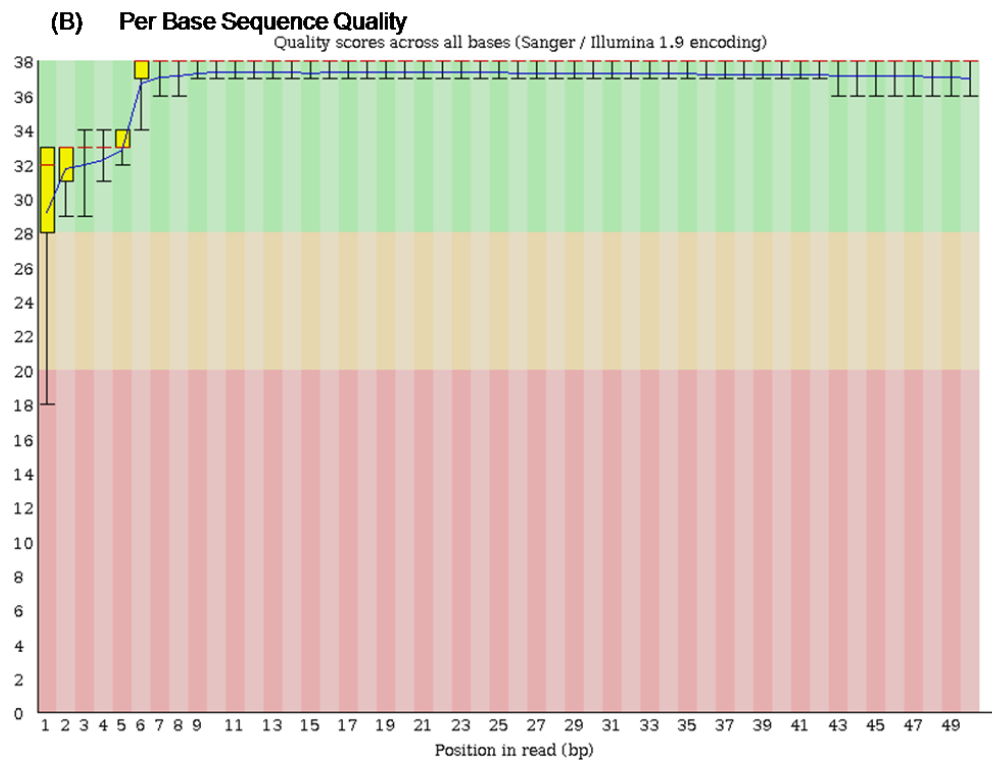


Figure 5.14 (B) Per base sequence quality plot of quality scores against position in N1_1 read. A good quality base call is observed with a slight fall at the start of the read, indicative of insertion deletion error which is dependent on the library preparation method.

iii. Per Tile Sequence Quality

The per tile sequence quality module is a distinct feature of the FastQC reports belonging to sequences generated from the Illumina library, which retains its original sequence identifiers. The flowcell tile coming from each read is encoded in this graph which also allows a check of quality scores from each tile across all bases to determine if a loss in quality was associated with only one part of the flowcell. The plot shows the deviation from the average quality for each tile. The colours are on a cold to hot scale, with cold colours (blues) being positions where the quality was at or below the average for that base in the run, and hotter colours (red) indicate that a tile had worse qualities than other tiles for that base. The reports from R1 reads across N, G1 and GA read sets in the present study issued a warning (orange triangle) for this plot. For instance, a few light blue and green spots are seen in N1_1 read pair (Figure 5.14 C). A warning is issued in this module if any tile shows a mean Phred score more than 2 less than the mean for that base across all tiles. Warnings or errors on this plot could be a result of transient problems such as bubbles going through the flowcell, or they could be more permanent problems such as smudges on the flowcell or debris inside the flowcell lane. Also, when a flowcell is generally overloaded, the variation in phred scores appears all over the flowcell rather than being specific to an area or range of cycles. Errors that mildly affect a small number of tiles for only 1 or 2 cycles are ignored whereas larger effects pertaining to high deviation in scores that persisted for several cycles are pursued. However, since the warnings posed by the module in this study were minimal and did not involve hotter (red) colours, it was not of major concern.

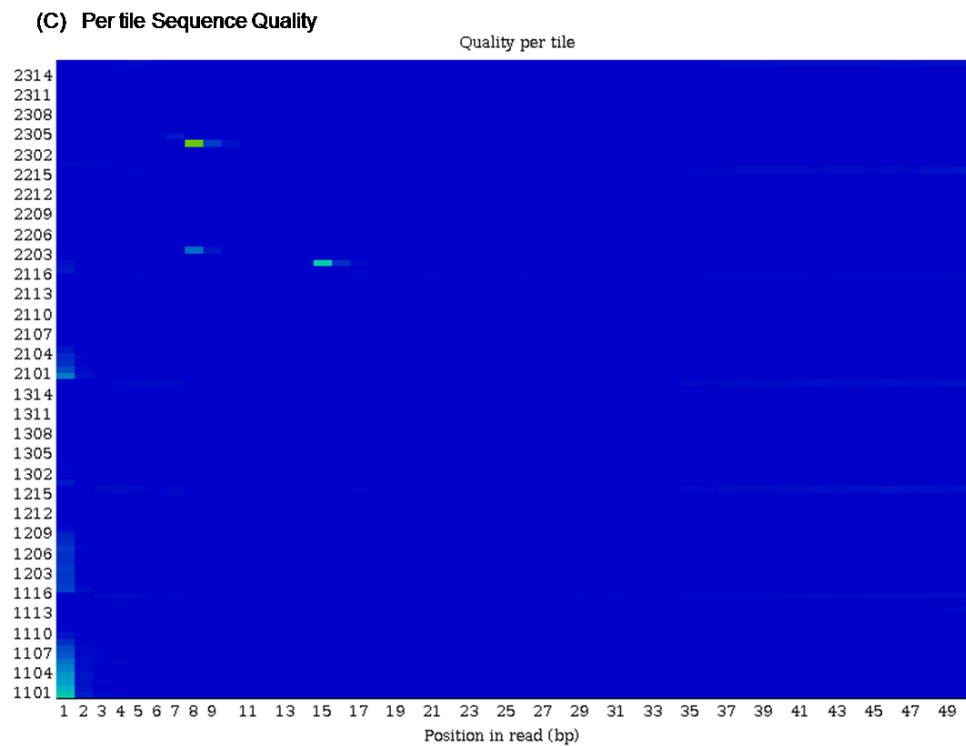


Figure 5.14 (C) Per tile sequence quality plot of the read file shows a deviation from the average quality in some tiles which indicate a bad quality (denoted in light blue and green) than other tiles (denoted in dark blue) for that base.

iv. Per Sequence Quality scores

The graph is generated by computing the average quality of a read and then plotting the distribution of this average quality. It thus determines if a subset of the sequences have universally low quality values and whether this subset is distributed across all reads. The y-axis depicts the number of reads and the x-axis, the mean quality score. It is often the case that a subset of sequences will have low mean quality scores. However, no low quality bases were observed across any reads in the present study as shown in the example of N1_1 (Figure 5.14 D). Also, no warnings were raised in this module since the highest peak in the distribution of all read pairs were mapped to a score higher than 20 on the x-axis (37).

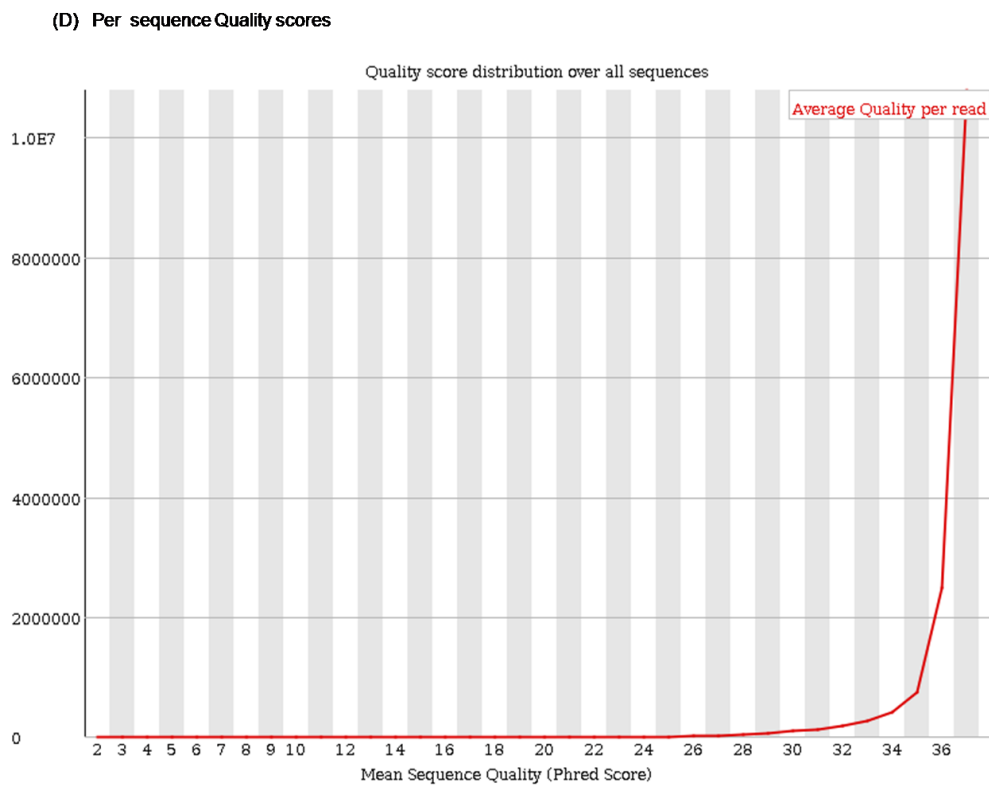


Figure 5.14 (D) The plot of per sequence quality scores representing the number of reads on y-axis and the mean quality score on x-axis. Highest peak in the distribution of reads in N1_1 file is mapped to a phred score of 37.

v. Per Base Sequence Content

This graph plots the proportion of each base position in a file for which each of the four normal DNA bases has been called. A random library is expected to have little or no difference between the different bases of a sequence run, therefore, the lines in this plot must run parallel with each other. However, all the read pairs generated by RNA-Seq in the present study raised an error (red cross) for this graph which is consistent with a >20% difference between A and T, or G and C in any position. It is common for some types of libraries, especially RNA-Seq libraries, to produce biased sequence composition, normally at the start of the read (first 12 bp of each run), as shown in N1_1 read pair (Figure 5.14 E). RNA-Seq libraries are produced by priming using random hexamers and those which were fragmented using transposases inherit an intrinsic bias in the positions at which the reads start. The bias is the result of an enrichment of a number of different K-mers at the 5' end of the reads. Nevertheless, this is not a problem which can be fixed by processing and the bias does not seem to adversely affect downstream analysis.

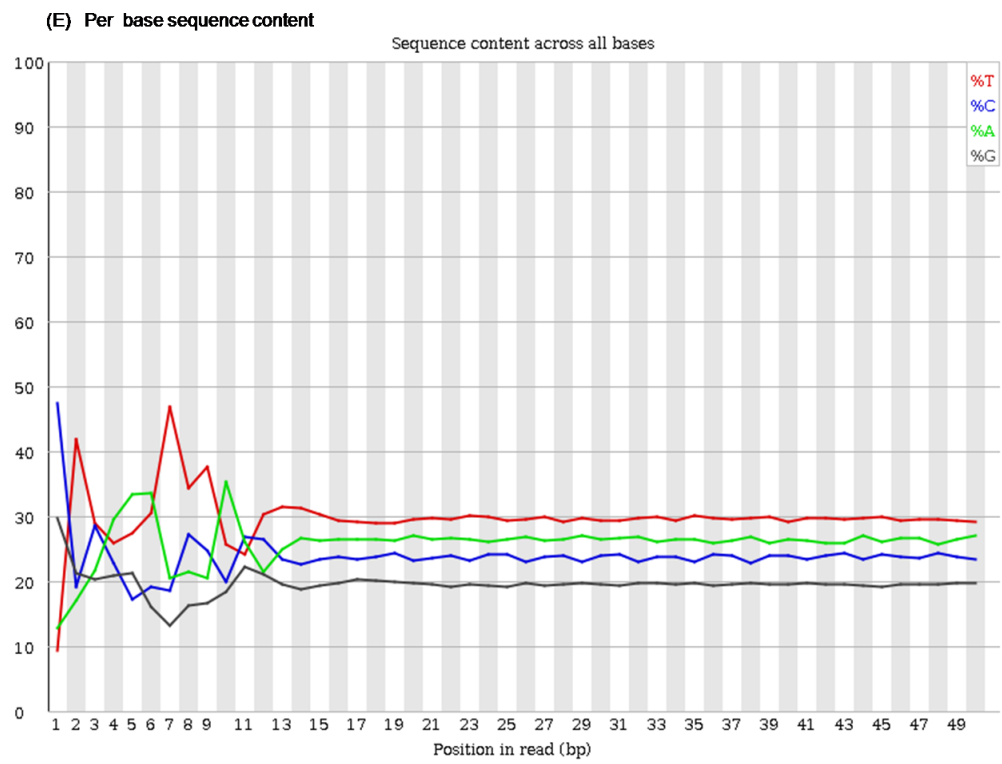


Figure 5.14 (E) Per base sequence content plot of each base position in a file for which each of bases has been called. ~ 50% difference between A and T or G and C is observed in the first 12 bp in N1_1 reads.

vi. Per Sequence GC Content

This module measures the GC content across the whole length of each sequence in a file and compares it to a modelled normal distribution of GC content. A roughly normal distribution of GC content where the central peak corresponds to the overall GC content of the underlying genome, as expected in a normal library, was observed in all the read pairs analysed (Figure 5.14 F).

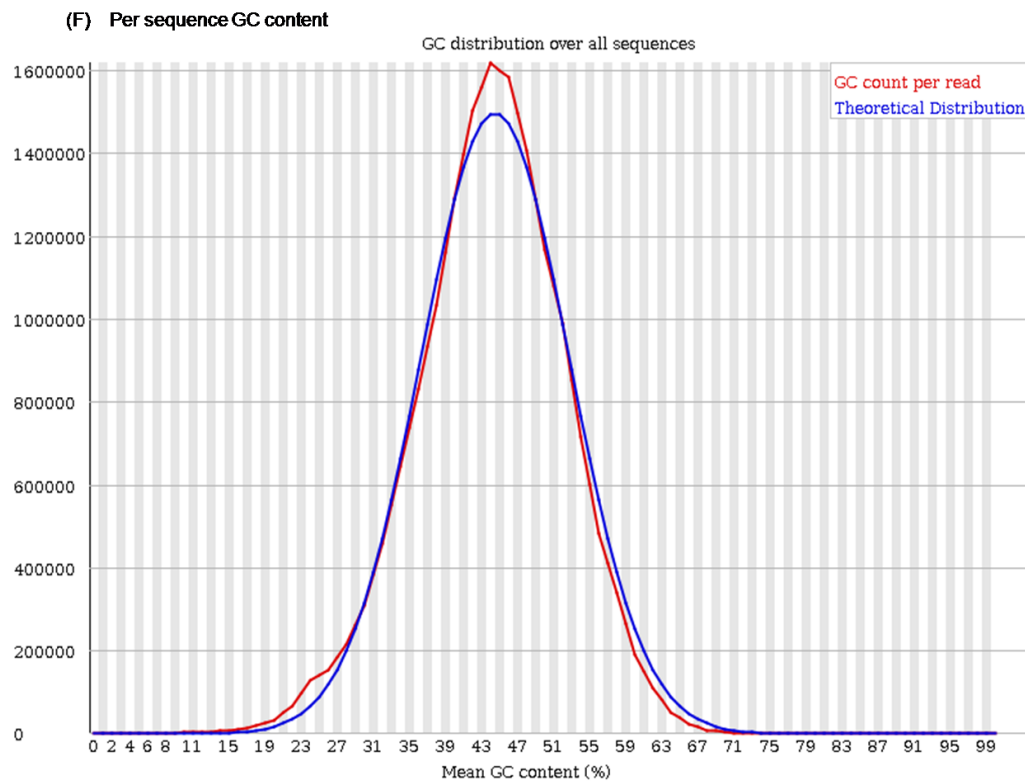


Figure 5.14 (F) Per sequence GC content depicts a roughly normal distribution of GC content across the whole length of the sequence in N1_1 (denoted in red) when compared to the overall GC content of the underlying genome (denoted in blue).

vii. Per Base N Content

The inability to make a base call with sufficient confidence will normally lead the sequencer to substitute an N in the place of a conventional base call. This module plots out the percentage of base calls at each position for which an N was called. However, the HiSeq v4 used in the present study has not made any such substitutions according to this plot (Figure 5.14 G).

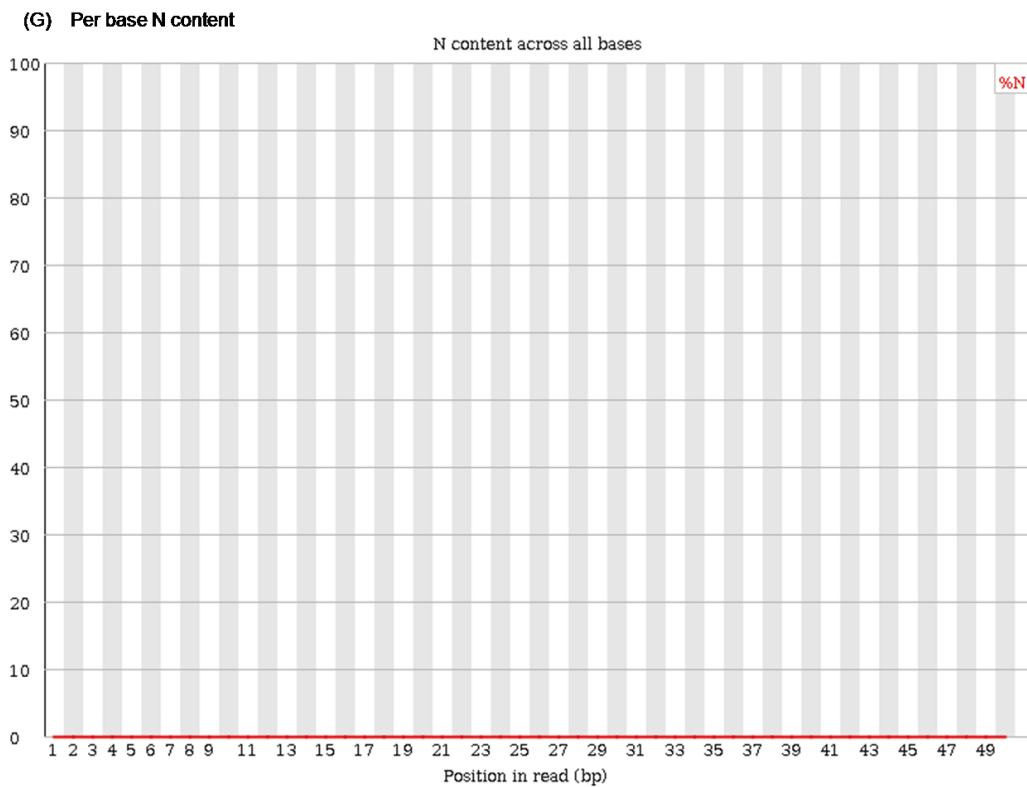


Figure 5.14 (G) Per base N content plot of N1_1 file does not detect any N substitutions in the place of bases.

viii. Sequence Length Distribution

This module generates a graph showing the distribution of fragment sizes in the file which was analysed. Since we generated 50 base reads using HiSeq, a graph showing a peak at 50 on x-axis was produced for all reads in the present study (Figure 5.14 H).

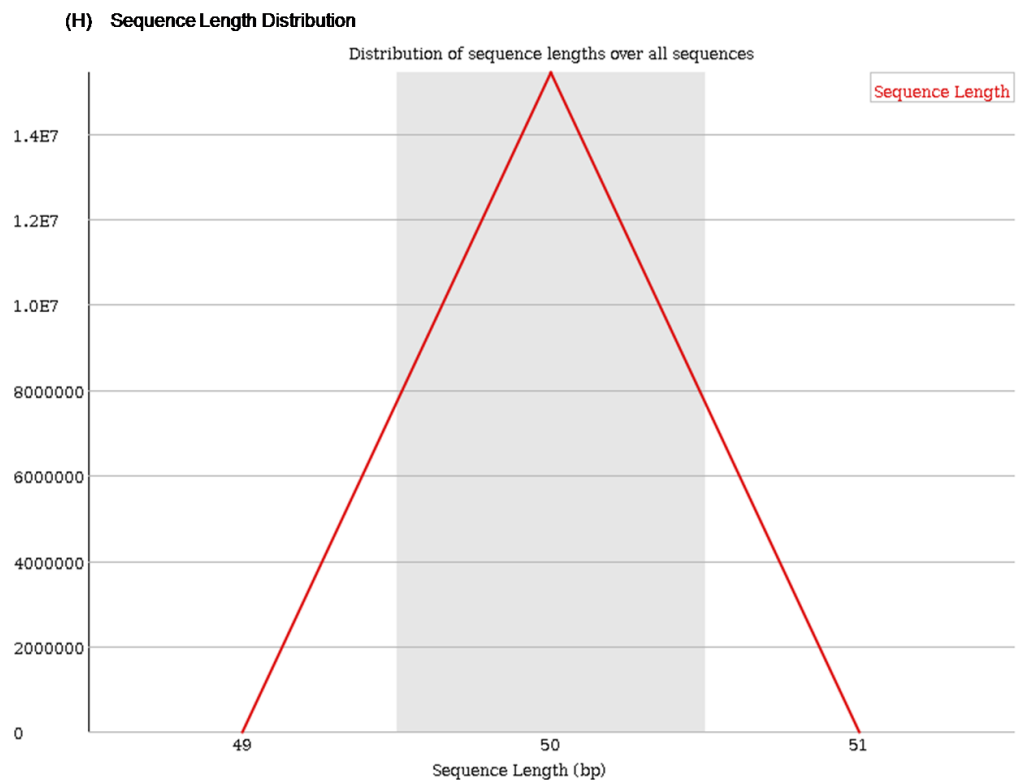


Figure 5.14 (H) The plot of sequence length distribution identifies a peak at 50 on x-axis which indicates that the length of all sequences in the file was 50.

ix. Sequence Duplication Levels

This graph shows the number of sequences with different degrees of duplication (indicated on the x-axis) relative to the number of unique sequences (which is set to 100%). In a diverse library most sequences will occur only once in the final set. A good coverage of the target sequence is indicated by a low level of duplication whereas a high level of duplication is more likely to indicate some kind of enrichment bias. To save memory, in this module, only sequences which first appear in the first 100 000 sequences in each file are analysed, which is adequate to get a good impression for the duplication levels in the whole file. A representative count of the overall duplication level is provided by tracking each sequence to the end of the file. To minimise the amount of information in the final plot any sequences with more than 10 duplicates are placed into grouped bins to give a clear impression of the overall duplication level without having to show each individual duplication value. The plot depicts the proportion of the library which is made up of sequences in each of the different duplication level bins. The blue line on the plot represents the full sequence set and shows how its duplication levels are distributed. Deduplicated sequences are indicated in the red plot and the proportions shown are the proportions of the deduplicated set which come from different duplication levels in the original data. The persistence of peaks even in the blue trace across all reads in the present analysis suggests a large number of different highly duplicated sequences in the read pairs, as seen in N1_1 (Figure 5.14 I). In addition, the module calculates an expected overall loss of sequence if the library were to be deduplicated. The calculated percentage value of sequences remaining after library deduplication varied across the read pairs analysed and fell within 13% - 20% range (Table 5.2). All the read pairs issued an error for this module since the non-unique sequences made up more than 50% of the total sequence reads. In RNA-Seq libraries, sequences from different transcripts occur at wildly different levels in the starting population. Therefore, it is common to greatly over-sequence high expressed transcripts for observing lowly expressed transcripts. The over-expression will potentially create a large set of duplicates. This duplication might explain the peaks produced in the higher

duplication bins in the present study. Examination of the distribution of duplicates in a specific genomic region will allow the distinction between over-sequencing and general technical duplication.

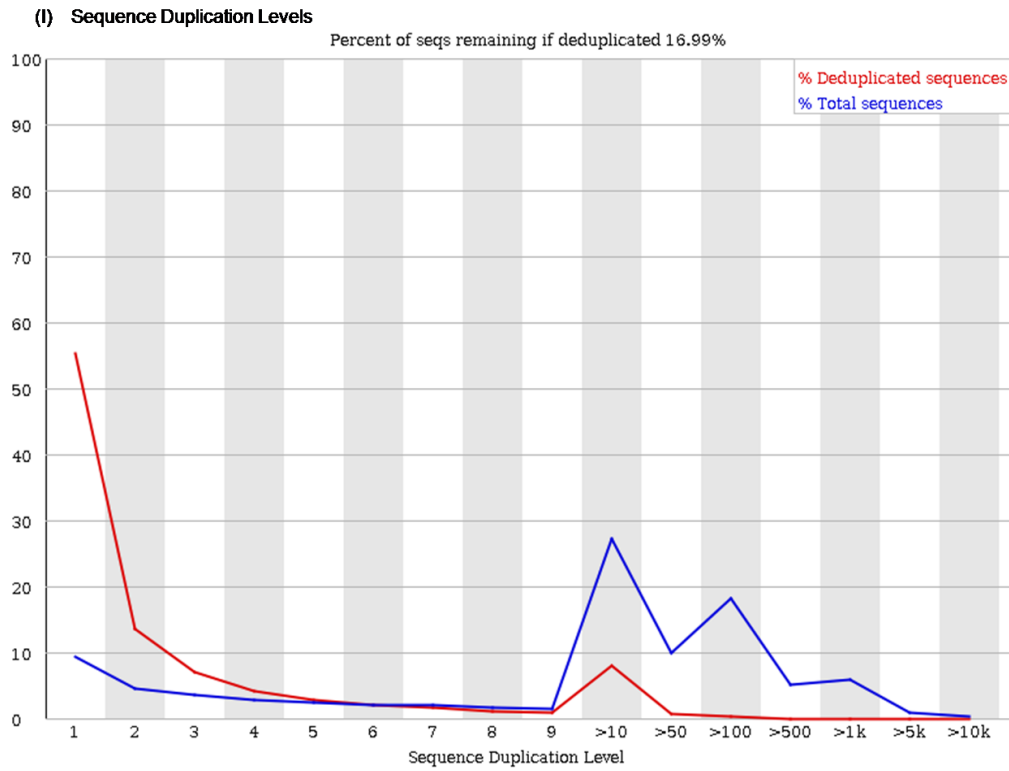


Figure 5.14 (I) Sequence duplication level graph shows persistence of peaks in blue plot in N1_1, indicative of different highly duplicated sequences.

x. Overrepresented Sequences

A normal high-throughput library will contain a diverse set of sequences, with no individual sequence making up a tiny fraction of the whole. Finding of specific overrepresented sequences in the set either means that it is highly biologically significant, or indicates that the library is contaminated, or not very diverse. This module in the present study listed all sequences that made up more than 0.1% of the total and thus raised a warning as observed in N1_1 (Figure 5.14 J). Eighteen overrepresented sequences were identified across the 15 read pair files (Table 5.3). To conserve memory, only sequences which appear in the first 200 000 sequences are tracked to the end of the file. Thus it is possible that a sequence which is overrepresented but does not appear at the start of the file could be missed by this module. For each overrepresented sequence the program will look for matches in a database of common contaminants and will report the best hit it finds. The 18 unique sequences in the present study were analysed using BLASTN against nr-databases and were confirmed to be part of NCTC 9343 genome rather than a result of contamination. Therefore, these 18 sequences were reported to be overrepresented in the genome according to the FastQC analysis and were not discarded.

(J) Overrepresented Sequences

Sequence	Count	Percentage	Possible Source
ATTATCTGAGTGGGGTGT CACCACACCACTTTAGCG GTCTACCCTCCGA	18606	0.12062197529343502	No Hit
CTAAGACGTACAGCCCGA TATGTCACCATACGGCTGG TGGGCTCTTACTC	16561	0.10736431972667837	No Hit
GTTTTTTTTTAAGGTA AAAA GATTGTTTTTCAGGATAACG ATGCAAATATA	16051	0.10405800953643587	No Hit

Figure 5.14 (J) List of three overrepresented sequences identified in N1_1 file which was present at more than 0.1% of the total sequences. No contamination hits were indicated.

xi. Adapter Content

This module does a specific search for adapter sequences and gives a view of the total proportion of the library which contain these separately defined Kmers. A result trace will always be generated for all of the sequences present in the adapter configuration file which enables the visualisation of the adapter content of the library even if its low. In the present FastQC report, the presence of the 4 analysed adapter sequences are almost close to zero across all reads suggesting that the adapter content is negligible (Figure 5.14 K).

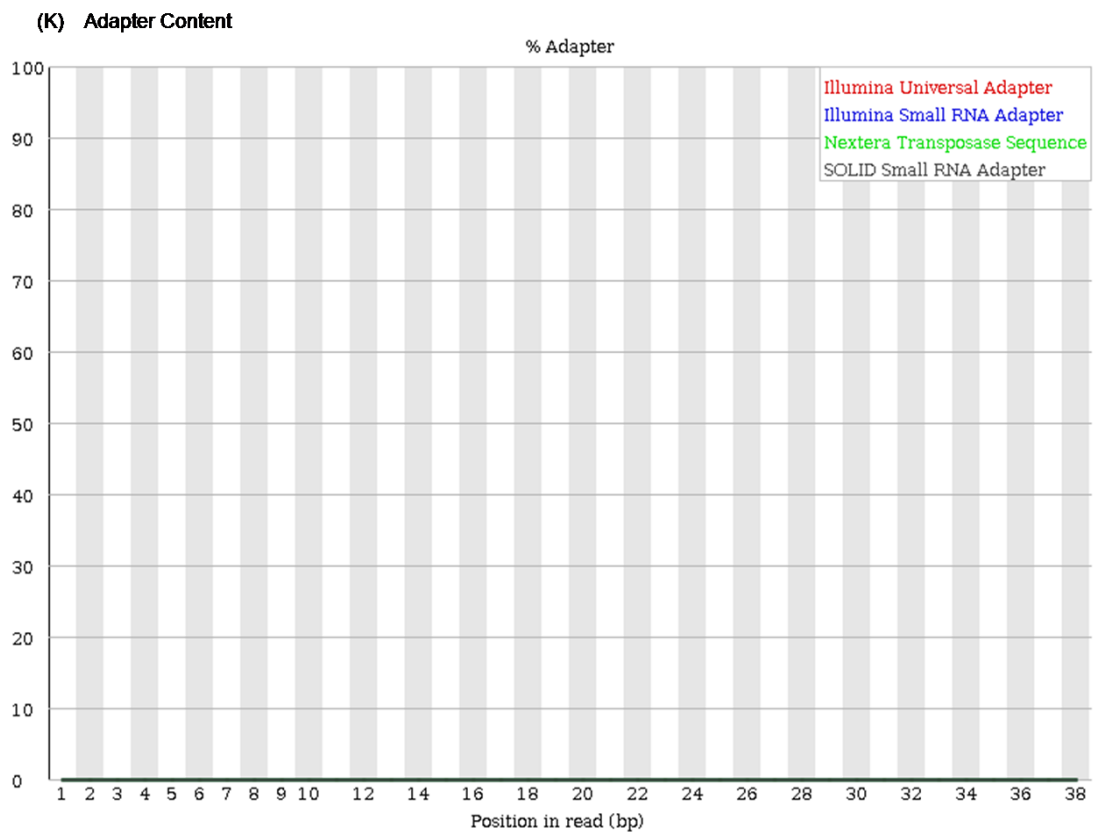
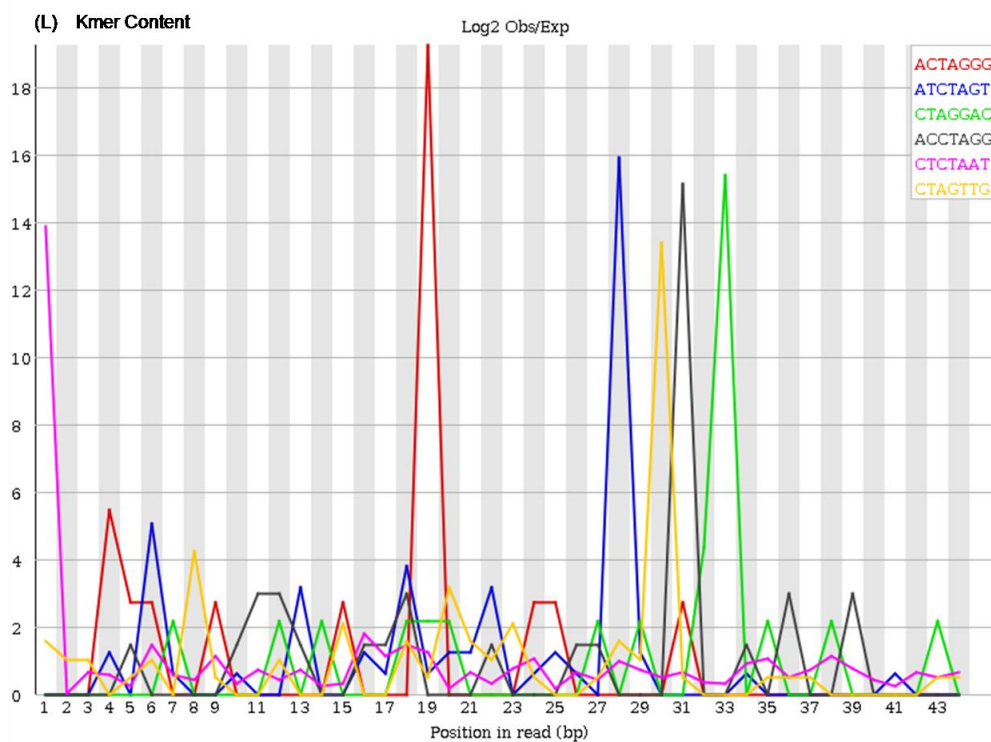


Figure 5.14 (K) The plot of adapter content in N1_1 file indicates negligible scores.

xii. Kmer Content

The Kmer module measures the number of each 7-mer at each position in the library and uses a binomial test to look for significant deviations from an even coverage at all positions. Any Kmers with positionally biased enrichment are reported. The top 6 most biased Kmers are additionally plotted to show their distribution. Only 2% of the whole library is analysed and the results are extrapolated to the rest of the library. Sequences longer than 500 bp are truncated to 500 bp for analysis. This module had issued an error for all reads in the present study due to a Kmer imbalance with a binomial p-value $< 10^{-5}$ (Figure 5.14 L). Individually overrepresented sequences that are not present at an adequate threshold to trigger the overrepresented sequences module will cause Kmers from those sequences to be highly enriched in this module, which appear as sharp spikes, as seen here, rather than a progressive or broad enrichment (Figure 5.14 L).



Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
ACTAGGG	80	9.001873E-5	19.248728	19
ATCTAGT	345	0.0	15.945006	28
CTAGGAC	100	4.9392943E-4	15.403727	33
ACCTAGG	145	2.2841705E-6	15.175938	31
CTCTAAT	3250	0.0	13.894159	1
CTAGTTG	410	0.0	13.417272	30
CTAGGGT	120	0.0019599372	12.836398	32
CTAATGT	3365	0.0	12.811049	3
CTCGGTT	2705	0.0	12.54051	1
GCCGAGC	1440	0.0	12.377312	28
CTCTAAG	1010	0.0	12.19496	6
CTAAGGA	795	0.0	12.179087	36
TACTAGC	145	5.855435E-4	12.140395	39
GTTTAAT	6280	0.0	11.960697	1
CTAGGAA	130	0.0035453169	11.860868	1
CTCGGAT	1460	0.0	11.768023	1
ACCCTCC	7505	0.0	11.752682	42
GTTTCGAG	2775	0.0	11.73039	11
TAAGGTA	6020	0.0	11.545276	10
ACATGGT	4035	0.0	11.446942	9

Figure 5.14 (L) Plot of distribution of the top 6 most biased Kmers in N1_1 file observed as sharp spikes and a list of 7-mers with positionally biased enrichment reported in N1_1.

Even though, the FastQC report issued a few errors and warnings, which are commonly observed in RNA-Seq libraries, the overall quality of each file was stated as 'Good' by the FastQC standards (Table 5.2). The unique set of 18 overrepresented sequences across the 15 read pair files were retained in downstream analyses since they were all part of the NCTC 9343 genome (Table 5.3).

Table 5.2: Fast QC report of the read pair files of the G1, GA and N sets

FastQC analysis indicating the number of reads, percentage of sequences remaining after deduplication and overall quality for the read pairs corresponding to each sample of G1, GA and N sets.

File Name	No. of Reads	Percent of sequences remaining after deduplication	Overall quality (inferred through the fastqc report)
Gu1_1	19,747,613	14.95%	Good
Gu1_2	19,747,613	13.97%	Good
Gu2_1	15,194,819	16.83%	Good
Gu2_2	15,194,819	16.18%	Good
Gu3_1	18,350,715	15.4%	Good
Gu3_2	18,350,715	14.34%	Good
Gu4_1	17,762,906	15.9%	Good
Gu4_2	17,762,906	15.23%	Good
Gu5_1	15,553,594	16.55%	Good
Gu5_2	15,553,594	16.52%	Good
GuB_1	18,321,211	15.2%	Good
GuB_2	18,321,211	14.56%	Good
GuC_1	17,509,499	14.88%	Good
GuC_2	17,509,499	14.72%	Good
GuD_1	16,127,604	14.66%	Good
GuD_2	16,127,604	15.8%	Good

GuE_1	14,148,510	16.14%	Good
GuE_2	14,148,510	16.49%	Good
GuF_1	10,563,179	18.61%	Good
GuF_2	10,563,179	19.06%	Good
N1_1	15,425,050	16.99%	Good
N1_2	15,425,050	16.06%	Good
N3_1	20,837,400	14.62%	Good
N3_2	20,837,400	14.1%	Good
N4_1	19,525,023	15.66%	Good
N4_2	19,525,023	14.32%	Good
N5_1	14,952,840	18.09%	Good
N5_2	14,952,840	16.54%	Good
N6_1	23,240,069	14.28%	Good
N6_2	23,240,069	13.58%	Good

Table 5.3: List of the unique overrepresented sequences

The 18 overrepresented sequences as reported by the FastQC across the 15 sample sets.

Sequence Name	Sequence
Seq1	ATTTATCTGAGTGGGGTGTCCACACACCACTTTAGCGGTCTACCC TCCGA
Seq2	CTAAGACGTACAGCCCGATATGTCACCATACGGCTGGTGGGCTCT TACTC
Seq3	CCGGTCTGTCGCGTGTACTTCGGTGCAGGAGGAAAGTCCGGGCA ACACAG
Seq4	CACAAACACAAAATAAAACACGACAAACTACTATGCAATCTTACC TGTC
Seq5	GTTTAATTACATGGTAACCTGCTTTGCAAGCAGTTTTTGAAGTTG CCAT
Seq6	CGCGTGTACTTCGGTGCAGGAGGAAAGTCCGGGCAACACAGAGC ATCCTA

Seq7	GGAAAGTCCGGGCAACACAGAGCATCCTACTTCCCTAACAGAAAGC TATCC
Seq8	GTTTTTTTTTAAGGTAAAAAGATTGTTTTTCAGGATAACGATGCAAAT ATA
Seq9	GTTCTTTTTTCGCAGTGGTGGCTCTTCCCATCTGTTTCGTCAAATAG TTGC
Seq10	CTTTTATTCATCAATCAAGTTCCTACTTTACTTTCGATGCCTTTGTTT TTC
Seq11	GCCAATTCTGTTCAAATTGTGCGCCACCAATCTTGATTACCTGAGC CGAT
Seq12	GTCGCGTGTACTTCGGTGCAGGAGGAAAGTCCGGGCAACACAGA GCATCC
Seq13	GTCATTTATCTGAGTGGGGTGCACCACACCACTTTAGCGGTCTA CCCTC
Seq14	GTCATGTTCTCCAGAATGGAAGTATCATCAGTCGATTCCACATGCA GACG
Seq15	GCCGAGTTCTGTACCCTGAATAAATCAAGGTGCCTGTCATTTATCT GAGT
Seq16	ATCTGAGTGGGGTGCACCACACCACTTTAGCGGTCTACCCTCCG ACATG
Seq17	GTTTTTCGGTAAATTCGGTAATGTATCCGTTTTCCAGGTAAGTTCC TTCA
Seq18	GTTTTTTTTGAAATGATAATAAATCCAAAGAAAAATCAAAAAGATGT TAT

5.3.4. Alignment against the reference genome

Adapter sequences were removed from raw fastq files and reads that were shorter than 20 nucleotides and longer than 50 nucleotides were discarded (2.8.2). The trimmed read pairs were aligned against the reference NCTC 9343 genome using Bowtie2-2.2.5 program. The genome sequence is available at:

ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/Bacteroides_fragilis_NCTC_9343_uid57639/NC_003228.fna.

The overall alignment rate was above 98% for all the read pairs processed which is indicative of a good genome coverage by the reads (Table 5.4). The number of unaligned reads in the alignment table are individual reads that could not be aligned to the genome anywhere under the parameters specified by Bowtie2 alignment. The failure to align these reads might be due to DNA/RNA contamination at the level of RNA extraction, library preparation or sequencing. A maximum of two mismatches were allowed while mapping reads to the genome. Alignment rate might lower if the sequence reads are significantly different at nucleotide levels, owing to events such as Single Nucleotide Polymorphisms. However, since the data presented here were of good quality with more than a 98% overall alignment rate, the number of unaligned reads was insignificant.

Table 5.4: List of read pairs aligned against the *B. fragilis* NCTC 9343 genome

Alignment was performed using Bowtie2 tool. The table indicates the total read pairs processed, number of paired reads aligned concordantly 1 time, >1 time and zero times, number of unaligned reads and the overall alignment rate.

Input files	Total read pairs processed	No. of paired reads aligned concordantly exactly 1 time	No. of paired reads aligned concordantly >1 times	No. of paired reads aligned concordantly 0 times	No. of unaligned reads	Overall alignment rate
Gu1_1.fastq.gz, Gu1_2.fastq.gz	19,710,239	19,221,607	229,873	258,759	167,514	99.58%
Gu2_1.fastq.gz, Gu2_2.fastq.gz	15,171,723	14,839,706	161,977	170,040	133,592	99.56%
Gu3_1.fastq.gz, Gu3_2.fastq.gz	18,327,330	17,891,795	222,077	213,458	220,292	99.40%
Gu4_1.fastq.gz, Gu4_2.fastq.gz	17,722,095	17,168,253	187,254	366,588	186,027	99.48%
Gu5_1.fastq.gz, Gu5_2.fastq.gz	15,502,292	14,883,445	182,571	436,276	149,867	99.52%

GuB_1.fastq.g z,GuB_2.fastq .gz	18,255,247	17,449,880	242,072	563,295	403,384	98.90%
GuC_1.fastq.g z,GuC_2.fastq .gz	17,451,374	16,662,667	241,870	546,837	273,130	99.22%
GuD_1.fastq.g z,GuD_2.fastq .gz	16,014,677	14,456,168	553,457	1,005,052	474,559	98.52%
GuE_1.fastq.g z,GuE_2.fastq .gz	14,067,743	13,342,879	235,810	489,054	220,583	99.22%
GuF_1.fastq.g z,GuF_2.fastq. gz	10,506,534	9,993,321	135,420	377,793	165,196	99.21%
N1_1.fastq.gz, N1_2.fastq.gz	15,396,878	14,947,375	206,633	242,870	150,908	99.51%
N3_1.fastq.gz, N3_2.fastq.gz,	20,798,225	20,215,676	257,031	325,518	212,001	99.49%
N4_1.fastq.gz, N4_2.fastq.gz	19,489,795	18,873,765	240,859	375,171	273,734	99.30%
N5_1.fastq.gz, N5_2.fastq.gz	14,923,486	14,367,930	249,442	306,114	340,337	98.86%
N6_1.fastq.gz, N6_2.fastq.gz	23,187,846	22,492,350	313,187	382,309	290,629	99.37%

5.3.5. Gene Quantification

Prior to quantitating the aligned reads, the reads were converted from Sequence Alignment/Map format (SAM) to its binary version, BAM (2.8.2). The alignment files in BAM format were then sorted by name, position and indexed (2.8.2). Finally, genes were quantified by counting the aligned reads assigned to features using the htseq count tool HTSeq-0.6.1 (Table 5.5). The HTSeq program goes through the sorted alignment file to assign a read pair to a feature (gene). Those read pairs where both the mates failed to align to the genome were ignored and are mentioned under 'Read pairs not aligned' column in the table. The read pairs that were mapped to the genome but could not be assigned to any feature due to inter genic alignment are counted in the 'Read pairs could not be assigned to any feature' category. However, a high proportion of read pairs were assigned to features, which is significant for the differential gene expression analysis.

Table 5.5: List of the number of read pairs assigned to genes

Quantification of read pairs was performed using HTSeq tool. The table indicates the number of read pairs assigned to a feature and those that could not be assigned to any feature. Read pairs that could have been assigned to more than one features and of low quality were discarded.

Input files	Total read pairs processed	Read pairs assigned to a feature	Read pairs could not be assigned to any feature	Read pairs could have been assigned to more than one features, hence discarded	Read pairs skipped due to low quality	Read pairs with more than one reported alignment	Read pairs not aligned
Gu1_nam esorted.bam	19,710,239	16,972,863	1,439,602	1,044,034	185,696	0	68,044
Gu2_nam esorted.bam	15,171,723	13,160,374	1,032,226	796,684	126,421	0	56,018
Gu3_nam esorted.bam	18,327,330	15,700,830	1,451,852	907,092	170,586	0	96,970
Gu4_nam esorted.bam	17,722,095	15,295,393	1,154,460	1,038,581	157,104	0	76,557
Gu5_nam esorted.bam	15,502,292	13,107,410	1,225,824	954,525	157,190	0	57,343

GuB_nam esorted.bam	18,255,247	15,263,001	1,668,959	925,025	225,155	0	173,107
GuC_nam esorted.bam	17,451,374	14,271,296	2,015,446	839,400	215,841	0	109,391
GuD_nam esorted.bam	16,014,677	12,418,607	1,815,201	975,545	610,015	0	195,309
GuE_nam esorted.bam	14,067,743	11,534,330	1,509,208	735,349	206,547	0	82,309
GuF_nam esorted.bam	10,506,534	8,697,974	1,042,340	575,896	125,152	0	65,172
N1_name sorted.bam	15,396,878	13,072,291	1,196,632	892,299	173,124	0	62,532
N3_name sorted.bam	20,798,225	17,684,203	1,634,015	1,176,848	214,539	0	88,620
N4_name sorted.bam	19,489,795	16,647,934	1,355,588	1,161,867	204,530	0	119,876
N5_name sorted.bam	14,923,486	12,552,478	1,209,988	785,039	220,111	0	155,870

m										
N6_name sorted.ba m	23,187,846	19,844,463	1,657,507	1,292,300	268,845	0				124,731

5.3.6. Differential Expression

The read pairs that were successfully aligned against the reference genome and assigned a feature were finally analysed using DESeq2 (DeSeq2_1.6.1) software to determine the differentially expressed genes. The DeSeq2 tool applies the Wald significance test to identify the differential expression in read pairs. The Wald test is used to test the true value of a parameter based on a sample estimate, wherever a relationship within or between data items can be expressed as a statistical model with parameters to be estimated from the sample.

5.3.6.1. Analysis of the read pairs belonging to N, G1 and GA sets post gene quantification

Prior to analysing the differential expression of genes, it is essential to validate the read pairs belonging to the three sets of experimental conditions using analysis plots. Firstly, the raw read-count data matrix ('genes' as rows and 'read pairs from the sample sets' as columns, as compiled from the HTSeq-count result) was transformed to log₂ scale and normalized with respect to library size using rlog transformation (<https://www.bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.pdf>). Qualitative analysis of the five read pair files belonging to each of the three sets was performed using Euclidean distance, Poisson distance and Principal Component Analysis (PCA) plots (Figure 5.15 A and B and Figure 5.16). For Euclidean distance plot, 'dist' function (<https://stat.ethz.ch/R-manual/R-patched/library/stats/html/dist.html>) in R programming language was used to compute Euclidean distance by the formula $\sqrt{\text{sum}((x_i - y_i)^2)}$ between the sample records from the complete rlog transformed data matrix. This dissimilarity information was provided to 'hclust' function (<https://stat.ethz.ch/R-manual/R-patched/library/stats/html/hclust.html>) which uses an agglomerative method, 'complete' for hierarchical cluster analysis (Figure 5.15 A). For Poisson distance plot, 'PoissonDistance' function from library 'PoiClaClu' (<https://cran.r-project.org/web/packages/PoiClaClu/PoiClaClu.pdf>) has been used on transpose of raw count data matrix. This function initially uses a median-of-ratio method (also

called *deseq* method) to normalize the count data and power transformation to transform the data so that the data fits the Poisson model more accurately and then, calculates Poisson distance matrix. The distance information was fed to 'hclust' for hierarchical cluster analysis (Figure 5.15 B). Hierarchical clustering using Euclidean and Poisson distance plots revealed a closer relationship between the N and G1 sets of read pairs when compared to the GA (Figure 5.15 A and B). This observation was in agreement with the growth rate of the culture samples used for RNA extraction, where N and G1 sets were extracted from cultures growing at similar growth rates and GA set was from cultures where the growth rate was reduced due to nitrogen limitation. PCA is a statistical procedure that uses an orthogonal information to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. In the present study, PCA plot was performed to analyse the top 500 genes, which were selected by highest variance across the 3 sample sets as per the *rlog* transformed matrix, using their first two principal components (Figure 5.16). The plot suggests that there was no effect of experimental covariates and batch effects as the samples belonging to the same set (N, G1 and GA) were clustered together in the plot.

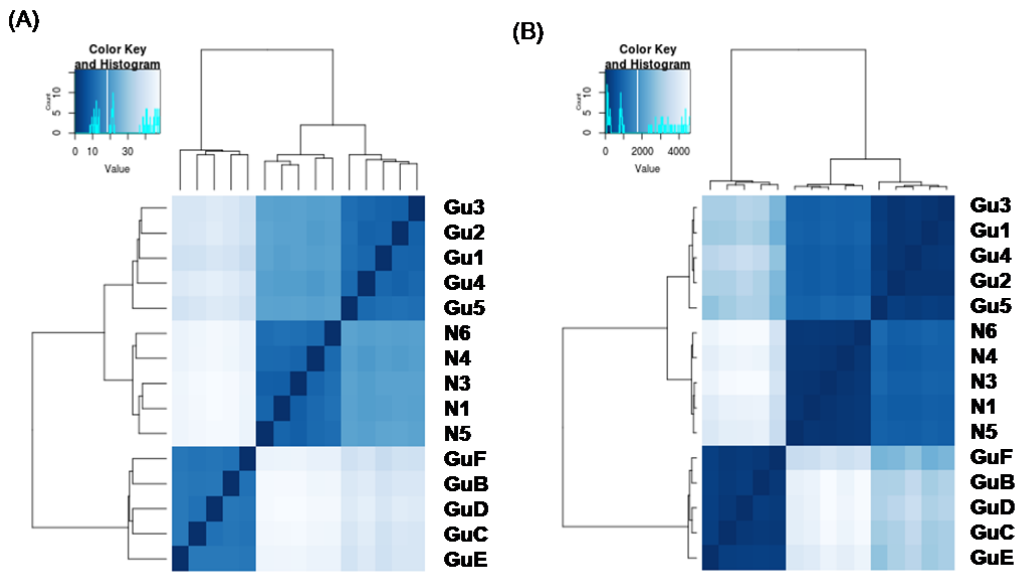


Figure 5.15: Distance plots of the normalized RNA-Seq reads

(A) Euclidean distance and (B) Poisson distance plots. The genes are represented in rows whereas columns represent sets of read pairs that have been transformed using rlog function. According to the colour key white-light blue shades depict genes that are separated by distance in each set and dark blue shades depict genes with zero distance. The clustering has grouped each read set together. N and G1 read pairs are more closely grouped than GA set according to heirarchial clustering.

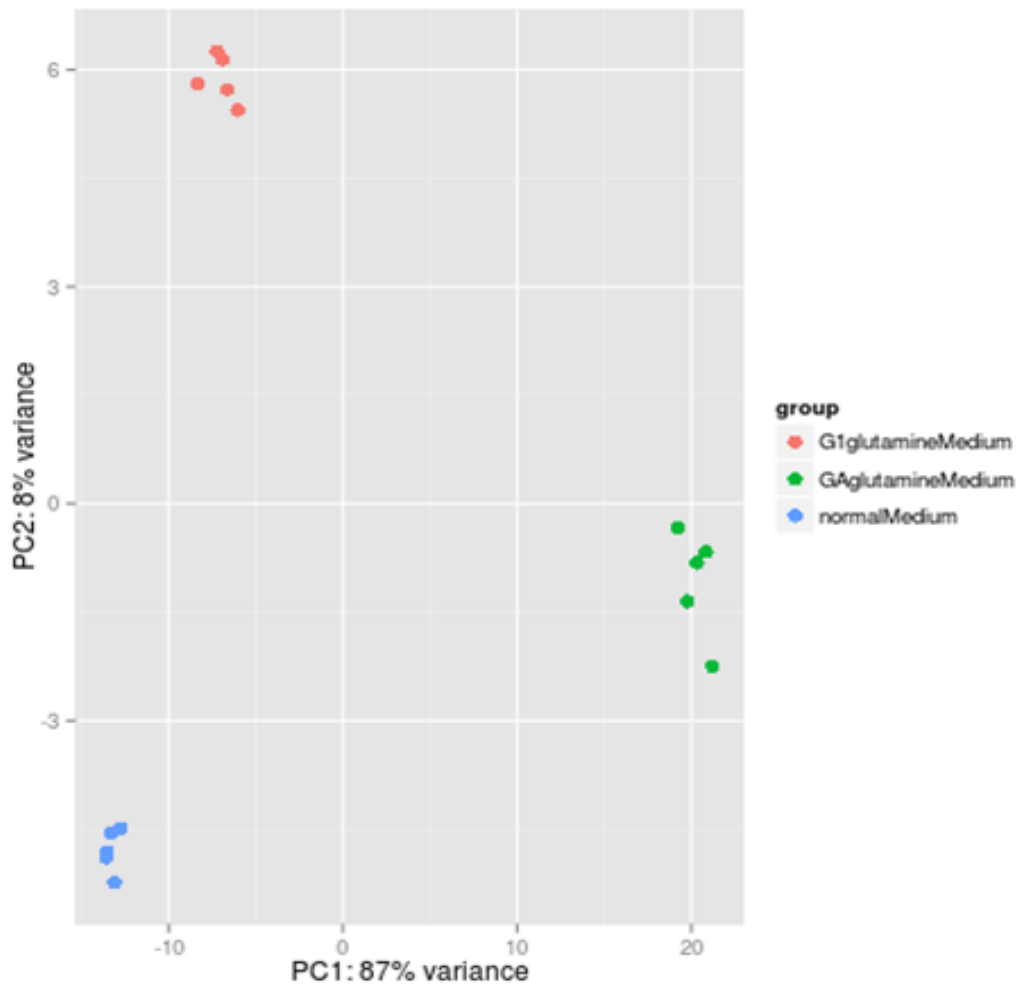


Figure 5.16: Principal component analysis of the normalized RNA-Seq reads

PCA plot representing samples belonging to N set (denoted in blue), GA set (denoted in green) and G1 set (denoted in red) spanned by their first two principal components. Samples of each specific set are clustered together.

5.3.6.2. Significance of normalization in differential expression

Having confirmed that the read pairs were clustered into three groups based on the sample set they originated from, we examined their normalized expression values and determined whether they could be compared across the three sample sets. Boxplot is a convenient way to compare the distribution of expression values among the different groups of interest. In RNA-Seq experiment, where we assume that the abundance of most genes will remain unchanged between groups, it is expected to have similar distribution for the normalized expression values for all the groups. Similar distribution indicates that the problem of having different sequencing depth in libraries in a next-generation sequencing (NGS) experiment has been minimized. Figure 5.17 suggests that the distribution of normalized expression values across all the three experimental sets were similar, and hence the normalized values of a gene can be compared across these sets of read pairs for identifying differential expression.

5.3.6.3. Assessing changes in gene expression across N, G1 and GA sets

The overall distribution of differentially expressed genes across the transcriptome from reads from the three sample sets were observed before analysing the regulation of genes of specific interest. Our primary aim was to compare the expression of genes in the GA set (from the transcriptome derived from glutamine-grown cultures at a T_d of 390 min \pm 20 min) with that in N set (from the transcriptome derived from normal glucose-DM cultures at a T_d of 78 min \pm 4.2 min) to identify the differentially expressed genes in the GA set as a result of a reduction in growth rate due to nitrogen limitation. Additionally, comparison of the G1 set (from the transcriptome derived from glutamine-grown cultures at a T_d of 90 min \pm 6min) with the reference N set was also performed to determine if the regulation of gene expression observed in the GA set was merely growth rate related or a combined effect of growth and nitrogen limitation. Differential gene expression of the GA set was also compared to G1 as reference set to validate the expression analysis.

Gene regulation was measured as a function of log₂ fold-change, such that, a log₂ FC (fold-change) value of 1 corresponds to the expression of a gene that has increased 2-fold (2^1) and a log₂ FC value of -1 corresponds to the expression of a gene that has decreased 2-fold (inverse of 2^{-1}) in the test sample when compared to the reference sample. For the ease of qualitative analysis and to provide an overview, genes that were assigned a positive log₂ fold-change value by DeSeq2 were considered as upregulated and those that were assigned a negative value, downregulated (no cut-off for log₂ fold-change value was assigned). The significance of the expression values was calculated by taking into account the false discovery rate (FDR). A FDR-adjusted p value (q value) of 0.05 implies 5% of the tests found to be statistically significant by p value could probably be false positives. Therefore, while performing multiple tests on the same sample, the use of q values instead of p values becomes advantageous. We have assigned a q value cut-off of < 0.05 all through our analyses. The q values (adjusted p values) that were above 0.05 were considered non-significant and omitted from the analyses.

Log₂ fold-change values of expression between the three groups were compared using scatter, MA and volcano plots by assigning a q value of 0.05 as cut-off. A three-way qualitative comparison based upon expression values, namely N (reference) vs GA (test), G1 (reference) vs GA (test) and N (reference) vs G1 (test) were performed using the plots to get an overview of the quantity and degree of the differentially expressed genes under the three experimental conditions. Scatter plots investigate the possible relationship between two groups. It highlights general similarities and specific outliers between two groups (or conditions). While comparing expression values of the GA set with G1 or N sets as references, a wide distribution of upregulated or downregulated genes was observed, which was indicative of high and low log₂ FC values (Figure 5.18 A, Figure 5.19 A). In these two plots, there was a remarkable decrease in the statistically non-significant values (at the specified cut-off), when compared to N vs G1 scatter plot, (Figure 5.20 A) and the number of not-applicable values (owing to the inability of DESeq2 to assign a q value) were below 15. A scatter plot comparing expression values of G1 set with

the reference N set displayed a significant proportion of genes as upregulated or downregulated. A slightly higher proportion of genes were either statistically non-significant at the specified cut-off or not applicable (Figure 5.20 A) when compared to GA plots with N or G1 as references. However, the majority of differentially regulated genes were close to the baseline indicating that the log₂ FC values of the upregulated or downregulated genes were not high and a stringent log₂ FC cut-off value might render the observed differential expression in the G1 set non-significant. These statistics suggest a higher magnitude of differentially expressed genes in the GA test set when compared to the G1 set as the test condition. Observations from the scatter plots agree with the experimental condition where the GA set pertained to RNA extracted from slow-growing cultures limited by nitrogen and therefore might possess a considerable difference in transcriptome when compared to the G1 and N sets.

In a two group-comparison problem, an MA plot depicts the trend of difference in expression over the average expression. The plot is a function of $M = \log_2(\text{Group}_2/\text{Group}_1)$ vs. $A = (\log_2\text{Group}_2 + \log_2\text{Group}_1)/2$ where group 2 is the test sample and group 1 is the reference sample. Under the assumption that most genes will have no change in their expression, most points are expected to lie around 'M'=0. A typical MA plot helps to spot the 'A' regions (such as low expressed, high expressed or moderately expressed) to which the differentially expressed genes belong at a user specified p.adj value and log₂ FC cut-off. In the present study differentially expressed genes were analysed across N vs GA, G1 vs GA and N vs G1 using MA plots at p.adj value (q value) < 0.05 and all log₂ FC values. In all the three plots, it was observed that a larger proportion of genes were significantly upregulated or downregulated when compared to the M0 line (Figure 5.18 B, Figure 5.19 B, Figure 5.20 B). MA plots do not indicate a considerable difference in differentially regulated genes between GA and G1 when using N as the reference. However, quantification of genes at stringent log₂ FC expression values would be required to obtain a clearer idea.

Qualitative analysis was also performed using volcano plots to spot the differentially expressed genes by considering both fold-change value and test statistic. It is used to display both the test statistic's p.adj value (signal-to-noise ratio) and fold-change (signal) using $-\log_{10}(\text{p.adj-value})$ vs. \log_2 fold-change, respectively. Consistent with the experimental observation, a larger proportion of differentially expressed genes at high or low \log_2 FC values was observed when GA was used as the test sample instead of G1 test with N as the reference set (Figure 5.18 C, Figure 5.19 C, Figure 5.20 C).

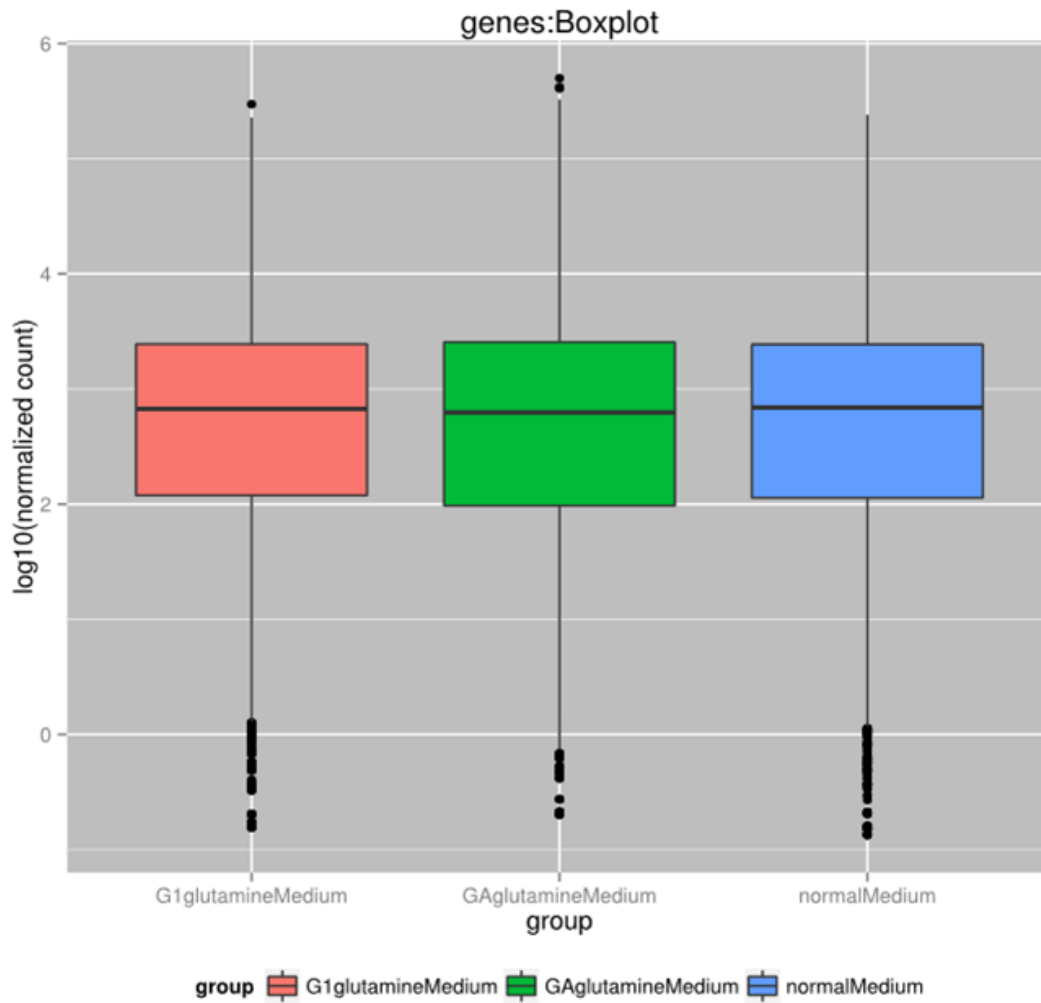
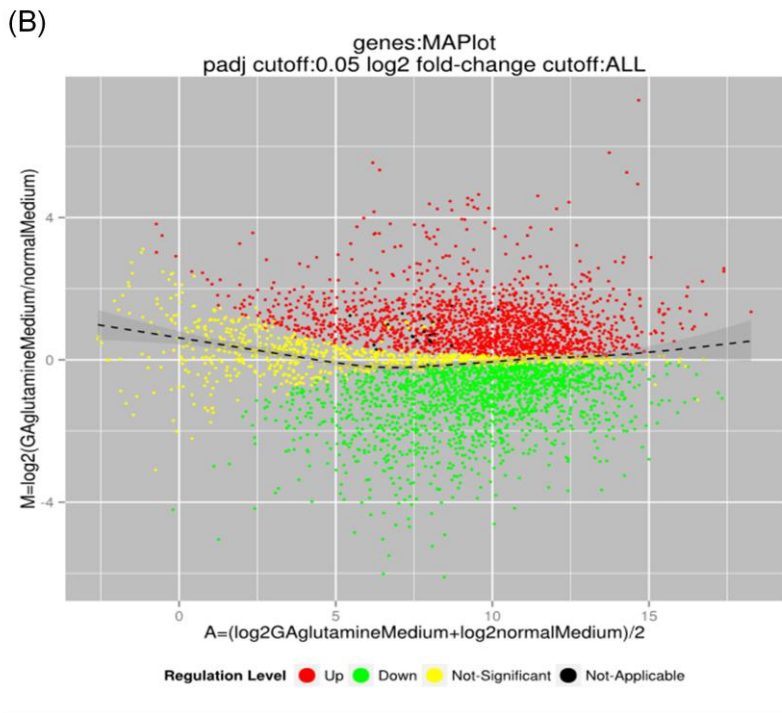
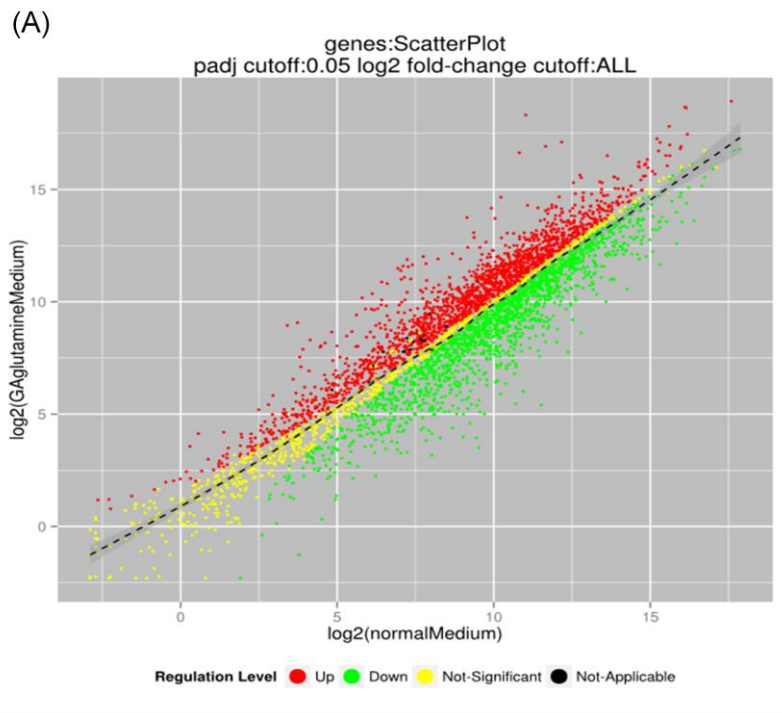


Figure 5.17: Representation of the normalized expression values

Box plot representing normalized expression values across G1 (denoted in pink), GA (denoted in green) and N (denoted in blue) read sets. The normalized values are similar and hence can be compared across the three sets for differential expression.



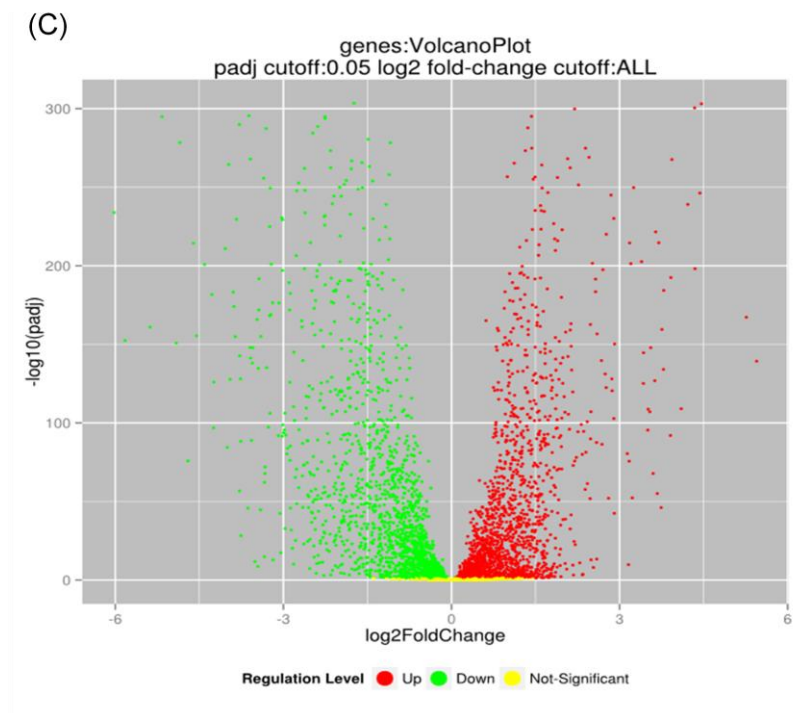
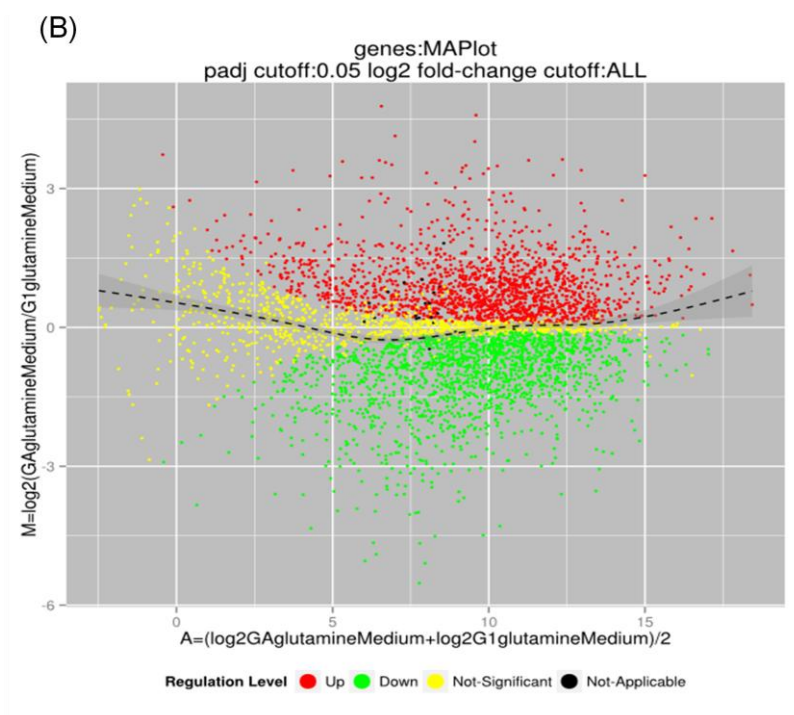
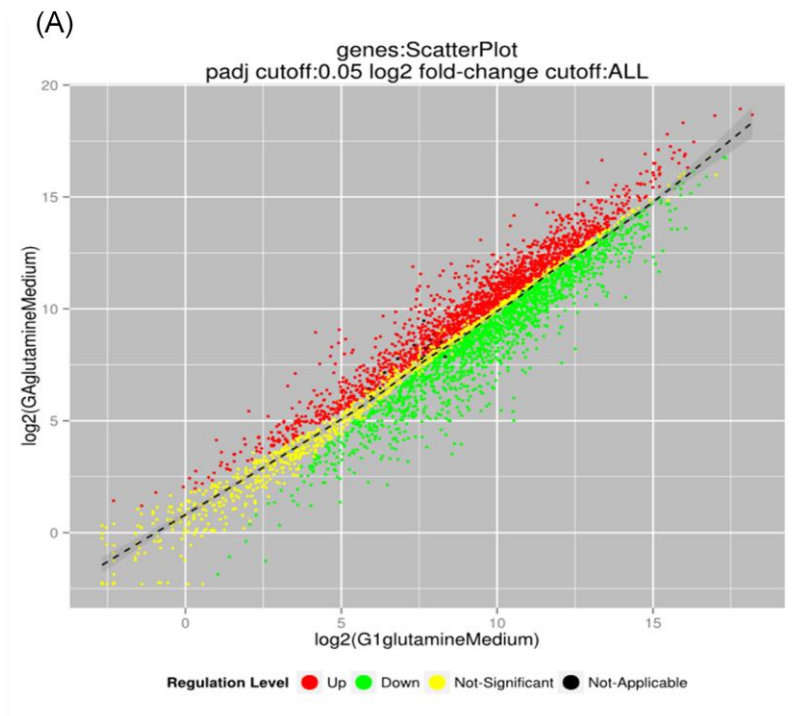


Figure 5.18: Qualitative analysis of the differentially expressed genes in the GA set with reference to the N set

Genes are represented as upregulated (denoted by red dots), downregulated (denoted by green dots), non-significant (denoted by yellow dots) and not-applicable (denoted by black dots) based on expression values at $p.adj$ value < 0.05 and all \log_2 FC values for GA when compared to reference N. (A) Scatter plot of \log_2 FC values of genes in N vs GA. Higher proportion of genes are either upregulated or downregulated with very few number of genes represented by non-significant or not-applicable $p.adj$ values. (B) MA plot as a function of $M = \log_2(\text{GroupGA}/\text{GroupN})$ vs. $A = (\log_2\text{GroupGA} + \log_2\text{GroupN})/2$. Dots farther from the M_0 line represent genes that are highly differentially expressed (upregulated or downregulated). (C) Volcano plot representing differentially expressed genes by considering both fold-change and test static. A greater number of genes attain higher values of differential expression.



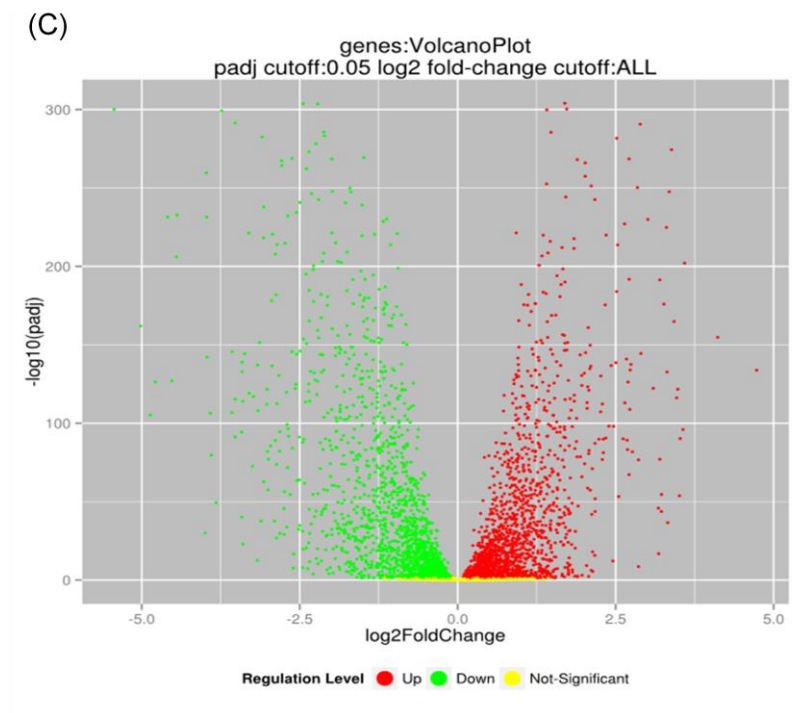
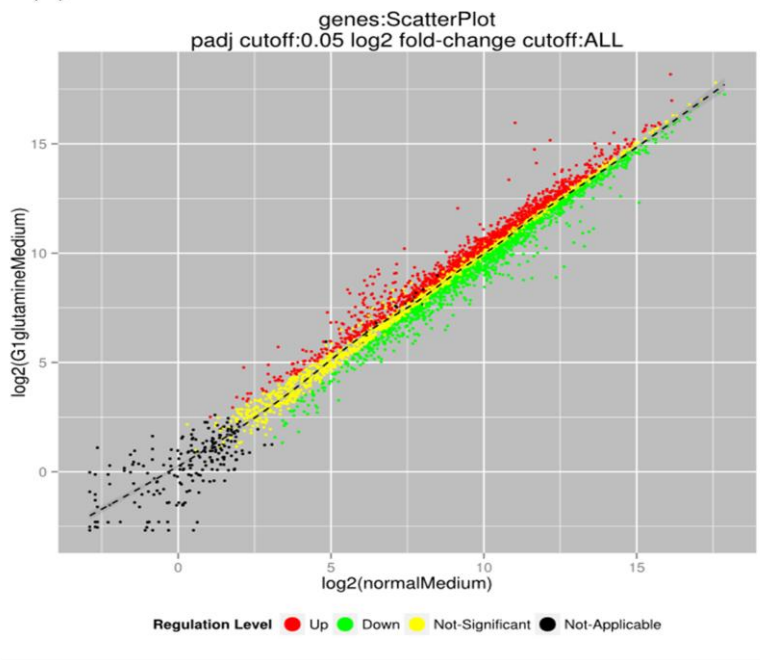


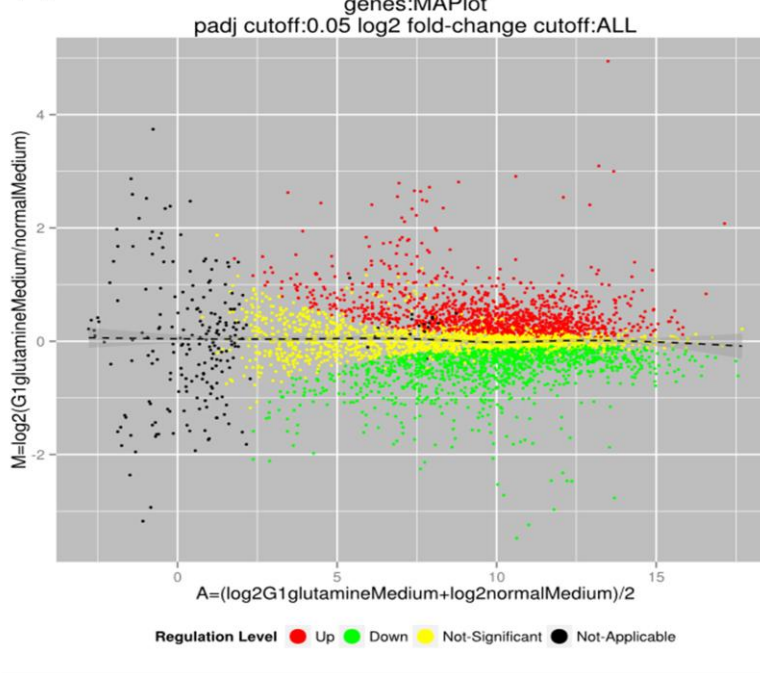
Figure 5.19: Qualitative analysis of the differentially expressed genes in the GA set with reference to the G1 set

Genes are represented as upregulated (denoted by red dots), downregulated (denoted by green dots), non-significant (denoted by yellow dots) and not-applicable (denoted by black dots) based on expression values at $p.adjust$ value < 0.05 for GA when compared to reference G1 (A) Scatter plot of \log_2 expression values of genes in G1 vs GA. Higher proportion of genes are either upregulated or downregulated with very few number of genes represented by non-significant or not-applicable $p.adjust$ values. (B) MA plot as a function of $M = \log_2(\text{GroupGA}/\text{GroupG1})$ vs. $A = (\log_2\text{GroupGA} + \log_2\text{GroupG1})/2$. Dots farther from the M_0 line represent genes that are highly expressed (upregulated or downregulated). (C) Volcano plot representing differentially expressed genes by considering both fold-change and test static. A greater number of genes attain higher values of differential expression.

(A)



(B)



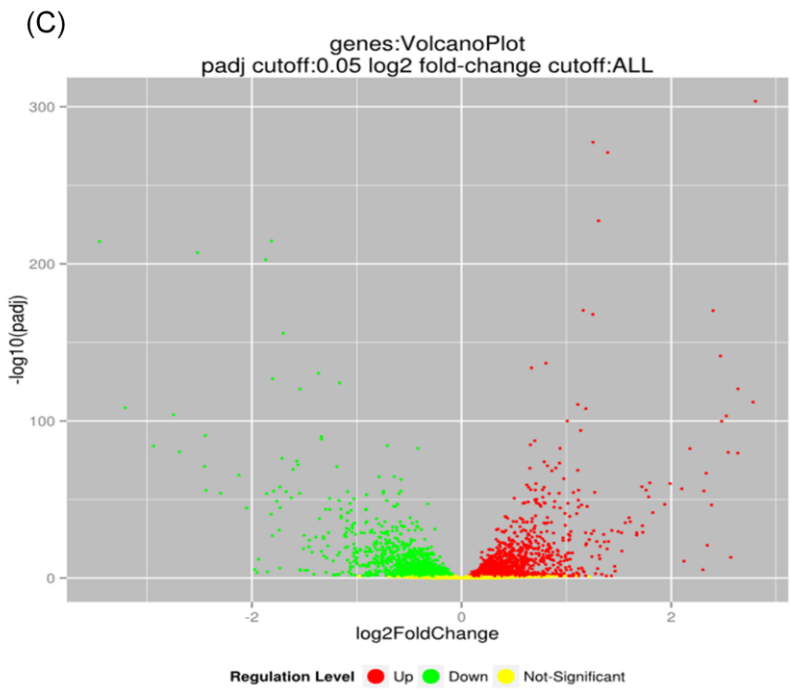


Figure 5.20: Qualitative analysis of the differentially expressed genes in the G1 set with reference to the N set

Genes are represented as upregulated (denoted by red dots), downregulated (denoted by green dots), non-significant (denoted by yellow dots) and not-applicable (denoted by black dots) based on expression values at p.adj value < 0.05 for G1 set when compared to reference N. (A) Scatter plot of log₂ expression values of genes in N vs G1. Higher proportion of genes are either upregulated or downregulated with a smaller proportion of genes represented by non-significant or not-applicable p.adj values. (B) MA plot as a function of $M = \log_2(\text{GroupG1}/\text{GroupN})$ vs. $A = (\log_2\text{GroupG1} + \log_2\text{GroupN})/2$. Dots farther from the M0 line represent genes that are differentially expressed (upregulated or downregulated). (C) Volcano plot representing differentially expressed genes by considering both fold-change and test static. A higher proportion of differentially regulated genes are closer to the baseline when compared to N vs GA and G1 vs GA volcano plots.

5.3.6.4. *Quantification of differentially expressed genes across N, G1 and GA*

To substantiate the qualitative analysis and to provide more information about the overall differential regulation, the DeSeq2 results of differential expression were quantified. At first, we determined the total number of differentially expressed genes at q value < 0.05 and no specified cut-off for \log_2 FC value (Table 5.6). All the values above 0 were considered as upregulated and below, downregulated. This was done to correlate the quantification results with the qualitative analysis plots which specified no \log_2 fold-change cut-offs. Of the 3433 differentially expressed genes in the N (reference) vs GA (test) sets, 1721 were upregulated and 1712 were downregulated. In the G1 (reference) vs GA (test) sets, 3312 were differentially expressed, of which an equal number of genes were determined as upregulated and downregulated. In N (reference) vs G1 (test) set, 2469 genes were differentially expressed (1249 upregulated and 1220 downregulated), which was surprising since the N and G1 sets belonged to reads from RNA extracts taken from cultures grown to similar OD_{600} values (0.4) with similar doubling times. Probably, more stringent \log_2 FC cut-offs have to be applied to observe its impact on the differential gene expression in the N vs G1 sets. In a microarray-based analysis of the global transcription response of *Mycobacterium smegmatis* to nitrogen stress, genes were considered to be significantly differentially expressed if their expression changed more than 2-fold at a q value < 0.05 for multiple analysis. A significant change in the differential expression of 16% of the genome was observed under these conditions (Williams et al., 2013). Therefore, we applied a \log_2 FC cut-off of 1 (2-fold increase) and -1 (2-fold decrease) at q value < 0.05 to the present analyses for determining the effect of higher stringent conditions on the number of genes differentially expressed (Table 5.7). There was an expected reduction in the number of differentially expressed genes, since genes with \log_2 FC values between -1 and +1 (< 2 -fold decrease and increase) were not considered significant under the revised stringency conditions. Appendices 8 and 9 provide the list of genes that were significantly upregulated and downregulated in the GA set when compared to the reference N set. In the N vs GA , 1590 genes ($\sim 40\%$ of the genome) were

differentially expressed, with 784 upregulated and 806 downregulated genes. 33% of the genome was differentially expressed in G1 vs GA. However, only 5.75% of the genome (230 genes) was differentially expressed in N vs G1 of which, 111 were upregulated and 119 were downregulated. Appendices 6 and 7 provide the list of genes that were significantly upregulated and downregulated in the G1 set with reference to the N set. Therefore, applying a log₂FC cut-off of 1 decreased the magnitude of differentially expressed genes in N vs G1 analysis. A difference of ~ 12 min was observed between the doubling times of cultures used for RNA extraction from N and G1 set of reads. This slight decrease in T_d might have an effect on the RNA derived from glutamine-grown cultures (G1 set) which might account for the observed differential expression. This difference in gene expression might indicate that the glutamine medium was posing a nitrogen limitation on bacterial growth even at an OD₆₀₀ of ~0.4 (at which RNA for the G1 set was extracted), commensurate with the slight decrease in T_d . To assess the influence of growth rate variation in the regulation of gene expression in the GA set of samples, the number of differentially regulated genes in both the G1 and GA set of samples were compared using the N set of samples as reference. Of the 111 significantly upregulated genes in G1 (p adj. value < 0.05 and log₂ FC cut-off 1), 35 were not upregulated and 5 were significantly downregulated in GA. On comparison of the 119 significantly downregulated genes in G1, 26 were not downregulated and 3 were upregulated in GA. This difference observed in the regulation of the transcriptome derived from two time points in the diauxic growth of cultures in glutamine medium might be an effect of nitrogen limitation as well as growth rate reduction. The GA sample set was derived from cultures with a considerably lower mass T_d when compared to that from the G1 sample set.

Table 5.6: List of the number of genes differentially expressed in N vs GA, G1 vs GA and N vs G1 analyses at q value < 0.05

Quantification of the differentially expressed genes in N vs GA, G1 vs GA and N vs G1 using DESeq2 software at q value < 0.05. A higher number of differentially expressed genes were upregulated than downregulated in GA and G1 when compared to N.

Reference Sample	Test Sample	Total number of differentially expressed genes at padj-value <0.05	No. of up-regulated genes	No. of down-regulated genes
Normal Medium	GAglutamine Medium	3433	1721	1712
G1glutamine Medium	GAglutamine Medium	3312	1656	1656
Normal Medium	G1glutamine Medium	2469	1249	1220

Table 5.7: List of the number of genes differentially expressed in N vs GA, G1 vs GA and N vs G1 analyses at q value < 0.05 and log2 FC cut-off of 1

Quantification of the differentially expressed genes in N vs GA, G1 vs GA and N vs G1 using DESeq2 software at q value < 0.05 and log2 fold-change cut-off of 1. A reduction in the number of differentially expressed genes was observed when stringency was applied.

Reference Sample	Test Sample	Total number of differentially expressed genes at padj-value <0.05 and log2FC cutoff 1	No. of up-regulated genes	No. of down-regulated genes
Normal Medium	GAglutamine Medium	1590	784	806
G1glutamine Medium	GAglutamine Medium	1312	600	712
Normal Medium	G1glutamine Medium	230	111	119

5.4. Analysis of differential gene expression with respect to nitrogen limitation

To examine the transcriptional regulation of the *B. fragilis* NCTC 9343 genome under nitrogen-limiting, ammonia-deficient conditions, the normalized data obtained from DESeq2 analytical tool was utilised to identify groups of genes that were significantly upregulated or downregulated in the GA set of reads (glutamine-grown cultures at 0.8 OD₆₀₀ with a calculated T_d of 390 min ± 20 min) and G1 set of reads (transcriptome derived from the glutamine-grown cultures at ~ 0.4 OD₆₀₀, non growth-limiting phase with a calculated T_d of 90 min ± 6 min) with reference to the N set (transcriptome derived from ammonia-grown cultures at ~0.4 OD₆₀₀ and T_d of 78 min ± 4.2 min). The results were considered reliable owing to the fact that each experimental set was obtained from the RNA extracted from samples from five culture replicates. Additionally, heat maps were obtained using heatmap.2 function from the R package "gplots" generated for each group of genes to visualize the pattern of differential expression. Regulation of genes in the N set were considered as reference and compared across the G1 and GA sets in heat maps. Being colour-coded, the heat maps help to illustrate the associated values in a graphic representation. Furthermore, the function clusters the genes by hierarchical clustering which indicates the evolutionary relationship of the genes involved. Although Williams et al., (2013) mentioned that a 2-fold cut-off for expression values was convenient in comparing significant changes in the differential expression of the *M. smegmatis* transcriptome under nitrogen limitation, they also acknowledged that the application of an arbitrary cut-off could mask less dramatic changes in gene expression and metabolism that may contribute to a co-ordinated response at the level of a metabolic pathway. Thus, in their study, the expression levels from the microarray data of all 6578 genes of the *M. smegmatis* genome were compared under varying nitrogen conditions using AMBIENT and KEGG analyses. Therefore, in the present analyses of differential expression of different groups of genes in *B. fragilis* under nitrogen limiting conditions, no log₂ FC cut-off was applied, albeit a log₂ FC value ≥ 1 (≥ 2-fold increase) and ≤ -1 (≥ 2-fold decrease) were considered significant while comparing a sub set of upregulated and downregulated genes.

5.4.1. Nitrogen assimilation and metabolism

Since biphasic growth was observed in a nitrogen-limiting glutamine medium (GA set), it is likely to pose a direct and immediate influence on the genes involved in nitrogen regulation. Also, the nitrogen metabolic pathway that facilitates the growth of *B. fragilis* on glutamine but not on glutamic acid has to be identified (5.2.1). Appendix 1 of the 'gene groups' folder provides the list of differentially expressed genes involved in overall nitrogen assimilation, glutamine and glutamate metabolism by annotation. The protein product encoded by BF9343_2369 was homologous to the *E. coli* nitrogen regulation protein NtrC (35.62% identity in 480 aa) and was found to be upregulated in GA with a log₂ fold-change value of 0.1572 when compared to N reference (Figure 5.21 A). This was concurrent with the increase in expression of *E. coli* NtrC observed in nitrogen-limiting condition in *E. coli* (Shimizu, 2013). Upregulation of NtrC transcription factor is essential for the activation of *glnA*, *glnK* and *nac* promoters for the conversion of glutamine into ammonia via the GS/GOGAT pathway (Atkinson et al., 2002). Examination of the glutamine-related reactions identified two genes, the raw annotations of which shared amino acid (aa) similarity with glutamine synthetase and were upregulated in GA reads in the present study. One homologue encoded by BF9343_2258, was similar to *B. thetaiotaomicron* glutamine synthetase I (88.2% identity in 500 aa) was upregulated to a log₂ fold-change value of 0.2330 in GA. BF9343_2936 expressing the other homologue was similar to a *P. gingivalis* W83 glutamine-dependent nicotinamide adenine dinucleotide (NAD⁺) synthetase NadE (55.43% identity in 644 aa) was upregulated to a log₂ fold-change of 0.1729 in GA. BF9343_0877 was found to encode a homologue of a Fnr-like (fumarate and nitrate reductase) protein in *Lactococcus lactis* (24.29% identity in 214 aa) and a transcription regulator in *B. thetaiotaomicron* (68.98% identity in 216 aa) and was upregulated to a value of 1.5748 log₂ fold-change, ie ~ 3-fold increase in GA (Figure 5.21 A and B). It was observed in *E. coli* nitrogen regulation studies that *fnr* transcript level increases under nitrogen limitation and the *gltBDF*, operon encoding glutamate synthase (GOGAT), had a high binding affinity for Fnr global regulator (Paul et al., 2007). The upregulation of NtrC and Fnr

transcription factor homologues in the GA samples coupled with the activation of genes encoding glutamine synthetase-I homologue suggests that the GS/GOGAT pathway is favoured in *B. fragilis* during conditions of nitrogen limitation. However, the log₂ FC values of expression were all lower than a significant 2-fold increase, except for the *fnr* transcript, which questions the reliability of the observed upregulation. Apart from these regulatory enzymes, considerable upregulation was also detected in gene products encoding glutamine amidotransferase enzyme homologues, two of them sharing identity with glutamine amidotransferase Class I and three with glutamine amidotransferase Class II family (Appendix 1 and Figure 5.21 B). Glutamine amidotransferase (GATase) enzymes catalyse removal of ammonia from a glutamine molecule and its transfer to a specific substrate thus creating a new C-N group on the substrate. On the basis of sequence similarities, GATases are classified into Class I (TrpG-type) and Class II (PurF-type). Class I GATase domains are defined by a conserved catalytic triad consisting of cysteine, histidine and glutamate. They have been found in enzymes such as the second component of anthranilate synthase and 4-amino 4-deoxychorismate synthase, CTP synthase, GMP synthase, glutamine-dependent carbamoyl phosphate synthase and histidine amidotransferase HisH. ClassII GATase domains have been identified in amidophosphoribosyl transferases, glucosamine fructose-6 phosphate aminotransferases, asparagine synthase and glutamate synthase (Weng and Zalkin, 1987). The upregulation of GATases in the GA set might account for the effective glutamine utilisation through ammonia generation and in the formation of a glutamate pool from glutamine via GOGAT in the absence of ammonia for *B. fragilis* growth. Albeit three out of the five GATases were not upregulated by a significant 2-fold in GA, they all posed a > 2-fold increase in G1, except *hisH* (0.7905 log₂ FC) when compared to the reference N. This observation suggests that the reduction to non-significant FC value in GA might be growth rate related (Appendix 1 and Figure 5.21 B). Also, two significantly downregulated GATases in GA were only downregulated by a < 2-fold decrease in G1.

However, BF9343_1251, a gene encoding a homologue of *Neisseria gonorrhoeae* glutamine synthase (23.17% identity in 410 aa) was downregulated in G1 and GA with reference to N, and a gene identical to *E. coli glnA* was downregulated in GA, but upregulated in G1 (Figure 5.21 B). The log2 fold-change value of *glnA* reduction was -1.1887 (2.27-fold decrease) in GA, whereas the gene was upregulated by a 5.3-fold (log2 FC of 2.3990) in G1. Upregulation of *glnA* in G1 is in agreement with the upregulation of this gene by sigma-54 in *E. coli* under nitrogen limitation (Shimizu, 2013). In addition, downregulation of BF9343_0887 gene product which was annotated as a nicotinamide adenine dinucleotide phosphate (NADPH)-dependent glutamate synthase small chain homologue (91.35% identity in 763 aa) in GA was also observed by a < 2-fold decrease (Figure 5.21 C). Two genes encoding homologues of N utilization substances A and B were downregulated in GA. Also, BF9343_4052, a gene annotated to be similar to both *Shewanella violacea* RNA polymerase sigma-54 factor RpoN (31.16% id in 507 aa) and *B. thetaiotaomicron* RNA polymerase σ^{54} BT2521 (86.84% id in 494 aa) was found to be downregulated in GA, indicative of slow growth. These findings suggest a different pathway than GS/GOGAT for nitrogen assimilation and glutamate generation in *B. fragilis* under nitrogen stress. Nevertheless, the fold-change values in expression of the downregulated genes was not greater than a 2-fold decrease except BF9343_1107 (N utilization substance B homologue) downregulation with a log2 fold-change value corresponding to -1.1685 (2.27-fold decrease).

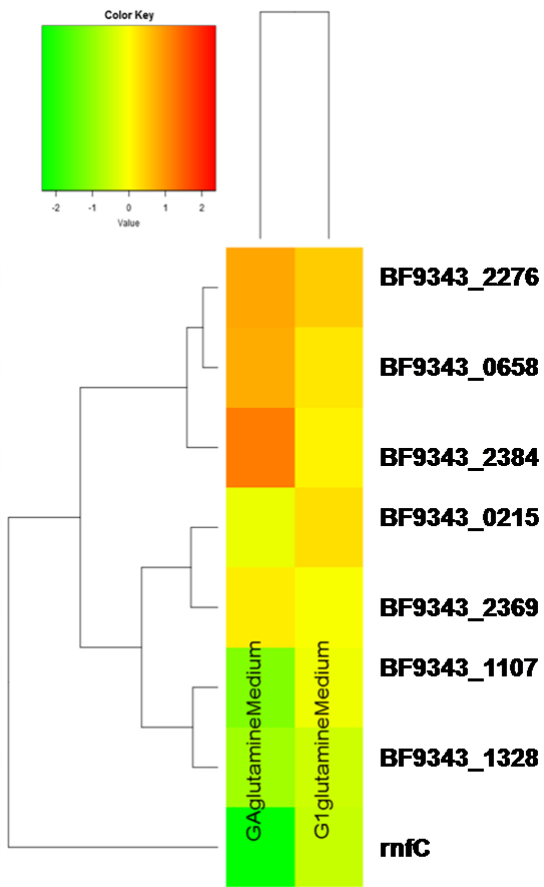
In contrast to *E. coli*, where the GDH pathway for glutamate synthesis is inactive during nitrogen limitation, the low ammonia concentration in batch cultures of *B. fragilis* has been shown to account for a higher cellular GDH activity. Low K_M values for ammonia and α -ketoglutarate in reactions with NADPH suggest that GDH encoded by *B. fragilis* functions in ammonia assimilation even under ammonia-limited conditions (Yamamoto et al., 1984). Although the medium of choice in the present study was ammonia-deficient, a significant upregulation (> 5-fold increase) was observed in GA and a 4-fold upregulation in G1 for *gdhB2*, the product of which is similar to *B. fragilis* NAD-specific glutamate dehydrogenase GdhB (69.21%

identity in 445 aa) and *Prevotella ruminicola* NADP-specific glutamate dehydrogenase GdhA (76.52% identity in 443 aa) (Figure 5.21 C). Nevertheless *gdhB*, which is identical to the gene encoding previously sequenced *B. fragilis* NAD-specific glutamate dehydrogenase GdhB and similar to *E. coli* NADP-specific glutamate dehydrogenase GdhA (56.98% id in 444 aa) was significantly downregulated in both the G1 and GA sets of reads when compared to the reference N (Figure 5.21 C). This difference in regulation observed in two homologues of glutamate dehydrogenase suggests a possible redistribution of resources as a consequence of reduced growth rate and nitrogen stress, similar to the differential regulation of four modules related to fatty acid metabolism via AMBIENT analysis of the differentially expressed genes in *M. smegmatis* (Williams et al., 2013).

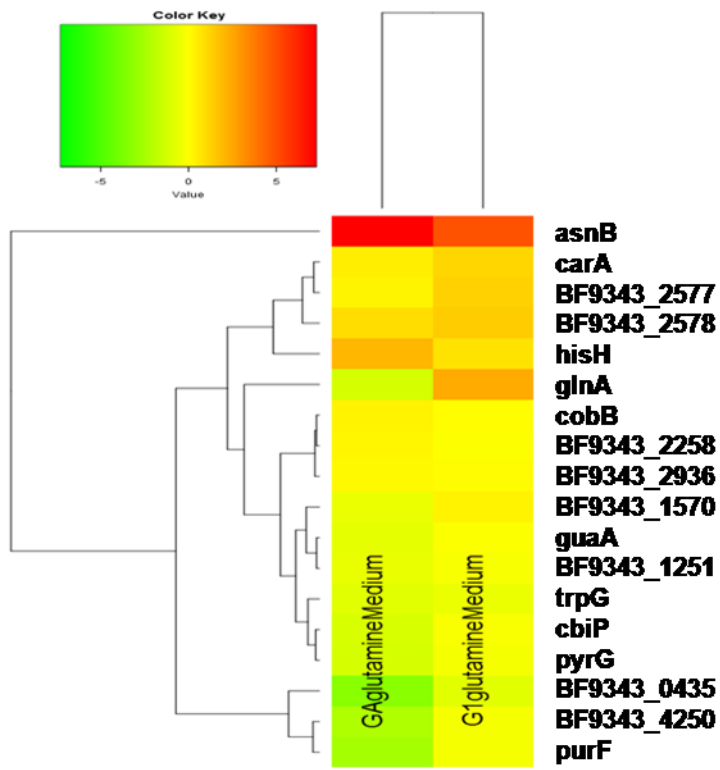
A downregulation of the genes encoding homologues of enzymes involved in the aa (proline and histidine) catabolic process to form glutamate was observed in GA. This was surprising since the bacterium was in need of ammonia under nitrogen stress. However, the observed downregulation in aa catabolism in GA could be related to the reduction in growth rate since the downregulation of these genes were identified by a < 2-fold decrease in G1 (Figure 5.21 C). Consistent with the downregulation of genes involved in aa catabolic process in the present study, Petridis et al., (2015) identified a 5-40-fold downregulation in a large cohort of genes involved in the catabolism of amino acids in the transcriptome derived from *M. smegmatis* continuous culture in response to nitrogen limitation, some of which (valine degradation) were attributed to a need to prevent unnecessary consumption of amino acids for CoA biosynthesis. A significant downregulation was also detected in *murF* encoding a homologue of *Bacteroides thetaiotaomicron* UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate-d-alanyl-d-alanyl ligase BT3644 (82.87% id in 432 aa) and a Pfam match to glutamate ligase domain (score 42.1, E-value 1.1e-09) in GA. Mur ligases are involved in the synthesis of peptidoglycan, specifically in the successive additions of L-alanine, D-glutamate, meso-diaminopimelate or L-lysine, and D-alanyl-D-alanine to UDP-N-acetylmuramic acid. Therefore it is probable that during the reduction in growth rate

under extreme nitrogen stress, as was observed in culture samples from which RNA was extracted for GA set, the cell size becomes smaller leading to a decrease in the amount of peptidoglycan required. The fold-change values observed for genes encoding Mur ligase homologues in G1 were not significant.

(A)



(B)



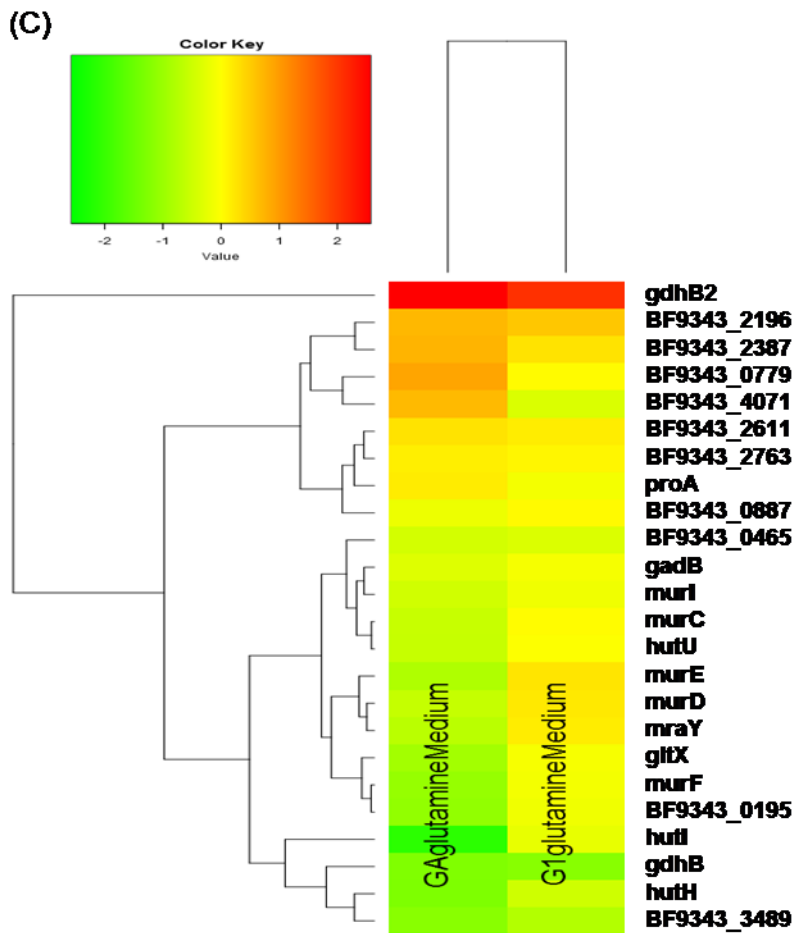


Figure 5.21: Qualitative analysis of the differentially expressed genes involved in nitrogen metabolism in the GA and G1 sets with reference to the N set

Heat map based analysis of differentially expressed genes using N as the reference set of reads and GA and G1 reads from RNA extracted from the biphasic growth in glutamine medium as the two tests (no cut-offs applied). Green-red colour key represents downregulated to upregulated genes based on \log_2 fold-change values with yellow representing unchanged gene expression. (A) Genes involved in overall nitrogen metabolism. Although a few genes are downregulated in GA, a larger number of genes are upregulated in G1 and GA when compared to N. BF9343_2384 encoding a carbon-nitrogen hydrolase homologue is highly expressed in GA (B) A majority of the genes involved in glutamine metabolism are upregulated in G1 and GA. Glutamine hydrolyzing asparagine synthase (*asnB*) is highly upregulated in both G1 and GA with reference to N. More number of genes are upregulated in G1. The differential regulation of *glnA* is observed when GA and G1 sets are compared to N (C) Differential expression of genes involved in glutamate metabolism in G1 and GA. Glutamate dehydrogenase2 (*gdhB2*) is markedly upregulated in

both G1 and GA while *gdhB* is significantly downregulated in both sets. Gene involved in aspartate/glutamate transport (BF9343_2196 and BF9343_2387) are also upregulated in GA.

As per the DeSeq analysis, if the GDH pathway is the favoured pathway for glutamate synthesis (upregulation of *gdhB2* in both the G1 and GA sets) in the nitrogen limiting phase of *B. fragilis* growth in glutamine-containing medium, the ammonia required for the synthesis needs to be generated by an alternative pathway. A striking upregulation corresponding to a log₂ fold-change value of 7.2795 (~ 155-fold increase) was identified for the *asnB* gene product which shares similarity to *E. coli* glutamine-dependent asparagine synthetase B (64.09% id in 557 aa) in GA (Figure 5.21 B). A 30-fold increase in *asnB* expression was also detected in G1 which suggests that the upregulation is not exclusively growth rate induced. Albeit non-significant, there was also a 1.688-fold increase in regulation (0.7557 log₂ FC) of a BF9343_2387 gene product, a homologue of *Escherichia coli* O157:H7 glutamate/aspartate transport ATP-binding protein GltL (28.5% id in 200 aa) (Figure 5.21 B). The upregulation of these two genes suggests their involvement in *B. fragilis* growth under nitrogen limiting condition. This observation led to the analysis of aspartate-related, amino-acid transport related and deaminase-encoding genes (Figure 5.25) in GA. Appendix 2 of the 'gene groups' folder provides the list of differentially expressed genes involved in general amino acid metabolism, aspartate and arginine metabolism, and deaminase activity by annotation. Of the identified 10 genes that encoded proteins predicted to be involved in amino acid metabolism two enzyme-encoding genes, apart from the previously mentioned glutamate/aspartate transport protein, were found to be upregulated in GA more than a 1.5-fold. These were involved in amino acid catabolic processes and branched chain amino acid biosynthetic process, respectively (Appendix 2 and Figure 5.25). The reasons for the upregulation of these two specific genes is unknown, despite a 4.3-fold increase in expression of the gene encoding a homologue involved in branched chain amino acid biosynthetic process. Downregulation of the other genes involved in aa transport in GA, might not be just a consequence of reduction in growth rate since none of the

genes were significantly upregulated in G1 too. The reduction in expression might either be a mechanism of conserving energy or a possible regulation at the level of transcription by non-coding regulatory RNAs. Kim et al., (2012) reported the presence of multiple transcription start sites (TSSs), in both *E. coli* and *Klebsiella pneumoniae* operons involved in amino acid biosynthesis, central metabolism and transport, which provided alternative transcripts and regulated gene expression. A > 2-fold decrease was detected in *opuAA* and *opuAB* genes encoding homologues of proteins involved in glycine betaine transport. Among aspartate-related metabolism group, the majority of gene products that were upregulated in *B. fragilis* were homologous to aspartate aminotransferases/transaminases which are involved in the conversion of glutamate to aspartate (Appendix 2). The synthesis of asparagine from glutamine and aspartate via asparagine synthetase (AsnB) involves ammonia generation (Figure 5.22). Therefore the small proportion of glutamate synthesised by the GOGAT pathway during growth in nitrogen limiting glutamine medium could be boosted via GDH pathway following ammonia generation, to be used in cellular biosynthetic process in the absence of exogenous ammonia. A study in *E. coli* using aspartate as the sole nitrogen source demonstrates ammonia generation as a product of arginine catabolism during aspartate degradation (Goux et al., 1996). Enzymes of the Krebs-Henseleit pathway incorporates the amino nitrogen of aspartate into the imino nitrogen of arginine (Figure 5.23). The ArgG catalyzes the condensation of aspartate and citrulline to form arginosuccinate, the immediate precursor of arginine (Figure 5.23). Consistent with this pathway, a significant upregulation of *argG* (1.5332 log₂ FC, ~ 3-fold upregulation), encoding a homologue of the *B. thetaiotaomicron* argininosuccinate synthase BT3760 (93.5% id in 400 aa) was observed in GA (Appendix 2). Arginine deiminase (ADI pathway), previously studied in *Streptococcus faecalis* and *Pseudomonas aeruginosa* catalyses arginine decomposition into ornithine, carbon dioxide and ammonia (Cunin et al., 1986) (Figure 5.24). The transcript from BF9343_2385 encoding a protein homologous to Porphyromonas-type peptidyl-arginine deiminase (score 440.8, E-value 1e-129) was upregulated to a log₂ fold-change value of 0.9783 in GA (Appendix 2). The

upregulation of these genes suggests a similar pathway for ammonia generation in *B. fragilis*. However three genes encoding homologues of arginine decarboxylase, arginine-ornithine carbamoyltransferase and arginine-ornithine antiporter were downregulated in GA and G1, but to values < a 2-fold decrease, which could be considered non-significant (Appendix 2).

The regulation of genes annotated as deaminases was also analysed since deamination of amino acids leads to ammonia generation. Seven out of 10 genes encoding enzymes similar to deaminases were upregulated in GA when compared to the reference N (Appendix 2). Two of the genes, BF9343_2059 and *cdd* homologous to the genes encoding *Rhizobium loti* probable riboflavin-specific deaminase mlr2151 (29.61% id in 233 aa) and *B. thetaiotaomicron* cytidine deaminase BT1539 (89.31% id in 131 aa) were upregulated to a significant log2 FC value of 3.16 and 1.22 (8.94- and 1.33-fold increase), respectively. This observation underlines the significance of the requirement of ammonia by the bacterium to replenish nitrogen under starvation.

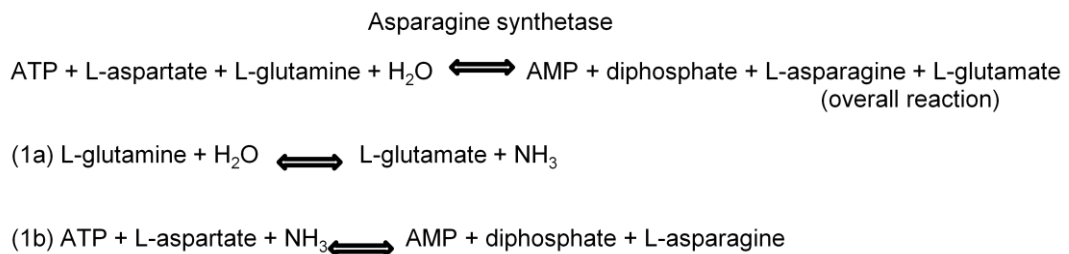


Figure 5.22: Synthesis of asparagine by glutamine hydrolysis

Reactions involved in the synthesis of asparagine by glutamine hydrolysis catalysed by asparagine synthetase. Ammonia is formed as a by-product during the reaction.

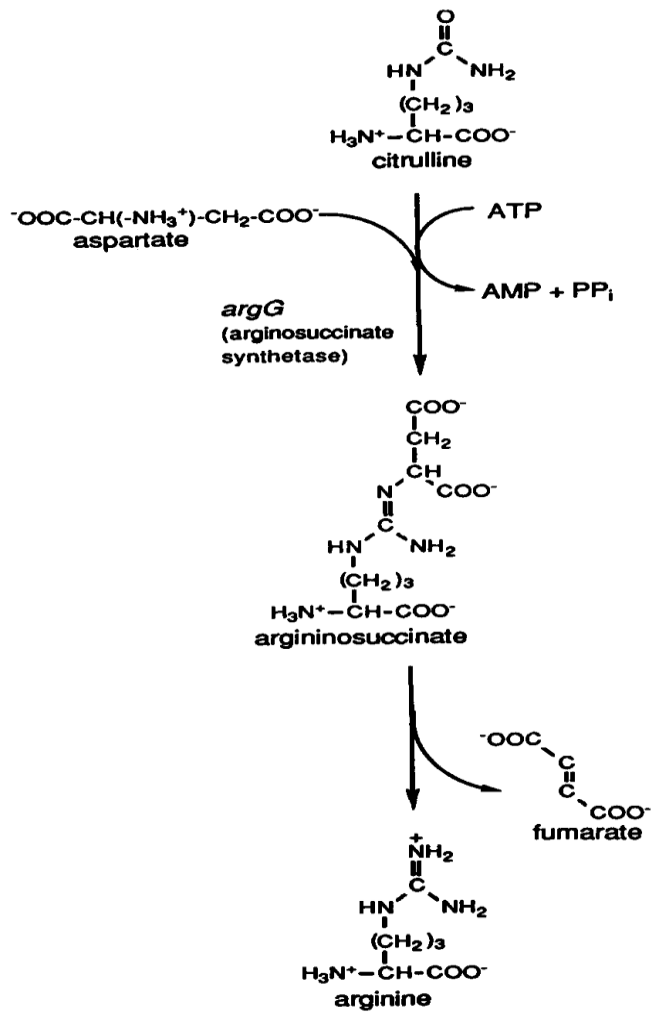


Figure 5.23: Krebs-Henseleit pathway-mediated formation of arginine as an intermediate in aspartate metabolism

Reactions involved in the condensation of aspartate and citrulline to form arginine with argininosuccinate as a precursor. (Goux et al., 1996)

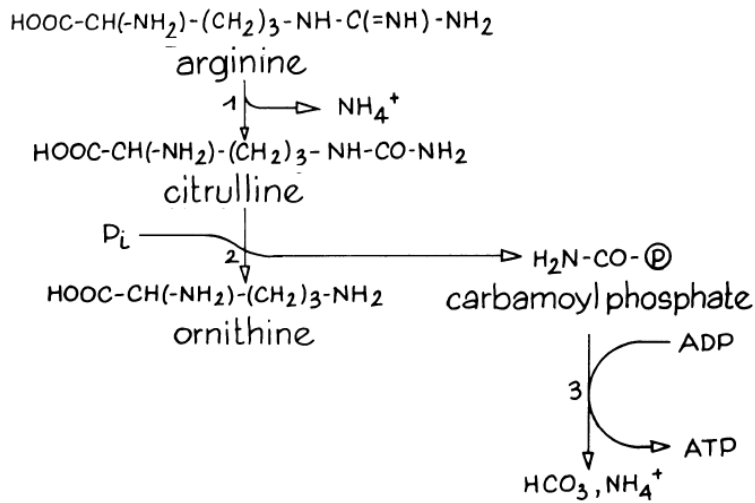


Figure 5.24: Arginine deiminase pathway for ammonia generation

Reactions involved in ammonia generation from arginine through the formation of ornithine by arginine deiminase (ADI) pathway. The first reaction involving conversion of arginine to citrulline is catalysed by arginine deiminase, the second reaction is catalysed by ornithine carbamoyltransferase whereas the the third reaction of the ADI pathway concerned with the generation of ammonia from carbamoyl phosphate is catalysed by carbamate kinase. (Cunin et al., 1986)

Additionally, the gene encoding a homologue of *B. subtilis* guanosine triphosphate (GTP) pyrophosphokinase RelA (36.04% identity in 763 aa) involved in guanosine tetraphosphate (ppGpp) metabolism, with a Pfam match to nitrogenases component 1 alpha and beta subunits, was found to be upregulated < 2-fold in both G1 and GA when compared to N (Figure 5.21 A). Albeit not significant in the present analysis, *relA*-induced ppGpp synthesis was activated by NtrC during nitrogen starvation in *E. coli*. ppGpp served as an effector alarmone of the bacterial stringent response and its level was modulated in *E. coli* by ppGpp synthetase RelA and ppGpp synthetase/hydrolase SpoT (Brown et al., 2014). Since RelA and SpoT contribute to stress adaptation, antibiotic resistance, expression of virulence traits and acquisition of persistent phenotype in pathogens, the induction of related genes were examined in *B. fragilis* during nitrogen depletion, which are discussed in the following sections. Genes encoding two enzymes belonging to the carbon-nitrogen hydrolase family were also found to be upregulated in the GA culture in the present study. Increased expression of these two genes might have an effect on the nitrogen scavenging capacity of this Gram-negative anaerobe (Figure 5.21 A).

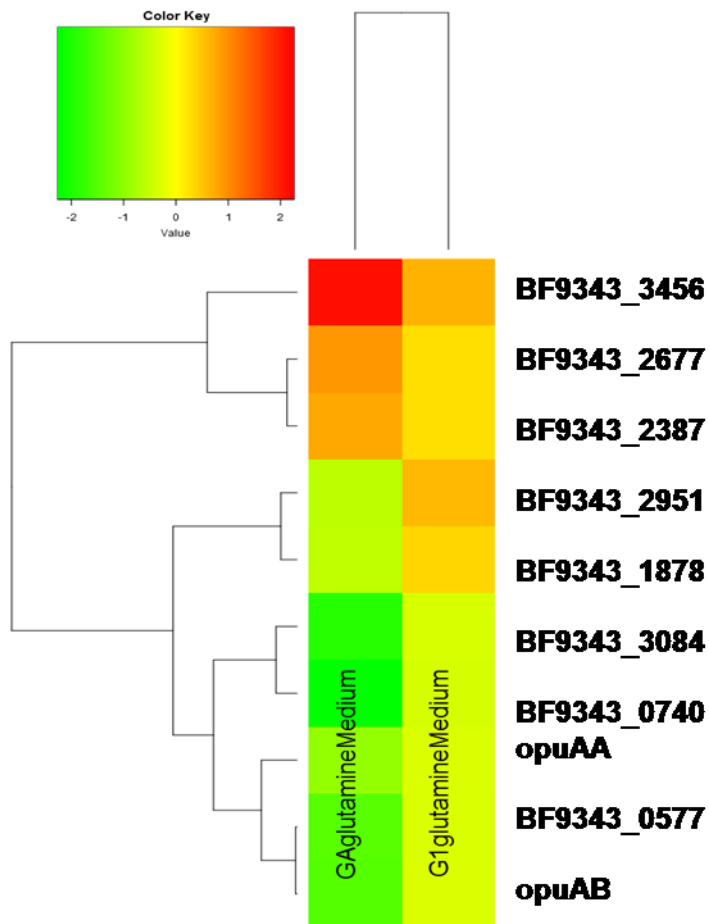


Figure 5.25: Qualitative analysis of the differentially expressed genes involved in general amino acid metabolism and transport in GA and G1 with reference to N

Heat map based visual analysis of differentially expressed genes involved in amino acid transport using N as the reference set and GA and G1 as the two tests (no cut-offs applied). Green-red colour key represents downregulated to upregulated genes based on log2 fold-change values with yellow representing unchanged gene expression. Although a higher number of genes are downregulated in GA than G1, a few genes are upregulated to a significantly higher fold in GA (BF9343_3456, BF9343_2677, BF9343_2387) when compared to G1. BF9343_1878 and BF9343_2951, involved in amino acid transport, are expressed to a higher level in G1 than GA.

5.4.2. Host Matrix adhesins and Secreted proteases

Analysis of the regulation of extracellular protease-encoding genes under nitrogen limiting conditions was performed with the intention of identifying proteases that might contribute to replenishing the nitrogen requirement of the bacteria via protein degradation, which might possibly include fibrinogen degradation. Studies related to *B. fragilis* growth in varying nutrient conditions have pointed out that secretion of proteases by *B. fragilis* is related to nitrogen starvation (Gibson and Macfarlane, 1988a). A similar correlation was also observed with an increase in the rate of protein degradation in *E. coli* and *S. typhimurium* (Miller, 1975). Protein degradation might help the bacterium scavenge for alternative nitrogen sources to satisfy the limitation and facilitate amino acid synthesis. Therefore, it is also possible that the expression of adhesin proteins, like MSCRAMMs, that recognise host matrix molecules might also be upregulated under nitrogen depleting conditions to facilitate nutrient binding for effective protein degradation by proteases. Appendix 3 provides the list of differentially expressed genes involved in host matrix adhesion and proteolytic activity by annotation.

The regulation of BF1705, which was found to bind both fibronectin and fibrinogen in the present study, was hence examined (3.9 and 3.11). Transcription of BF9343_1624 (BF1705 gene in artemis nomenclature) was upregulated by 0.4585 log₂ FC value (1.37-fold increase) in GA when compared to the reference N (Figure 5.26 A). However, the BF1705 differential expression was not significant when a 2-fold cut-off stringency is applied which suggests the function of BF1705 as an adhesin in the gastrointestinal tract. In contrast to BF1705, a > 2-fold upregulation of BF0586 (BF9343_0559), a hypothetical protein which was previously identified as a collagen binding protein (Cbp1) was also detected in GA (Galvão et al., 2014) (Appendix 3).

Sijbrandi et al., (2005) had identified a cytoplasmic α -enolase in *B. fragilis* strain BE1 that was upregulated under conditions of iron- or heme-starvation. This enzyme is a part of RNA-degradosomes that are involved in mRNA processing thereby

enabling bacteria to respond efficiently on the dynamic fluctuation of environmental signals. It was also proposed that the upregulation of α -enolase is essential under conditions of infection. The transcriptomic analysis of *B. fragilis* NCTC 9343 performed in the present study confirmed that a gene encoding enolase was upregulated to a log₂ FC value of 0.9351 (1.9-fold increase) in GA (Appendix 3). Therefore, in addition to iron-starvation, nitrogen limitation might also regulate *B. fragilis* α -enolase activity with potentially profound effects under conditions of infection.

Studies on *B. fragilis* strains BE1 and 638R had associated the laminin-1 binding and plasminogen activating functions of the bacterium with an outer membrane protein, Bfp60 (Ferreira et al., 2013; Sijbrandi et al., 2008). The activation of plasminogen to plasmin, a broad-spectrum serine protease is required for degradation of fibrin, collagen and other structural components of the matrix tissue. BF9343_4169 (BF4280) present in the NCTC 9343 genome encodes a protein (559 aa) that is 99% identical (558/559 aa; E-value:0) to the putative plasminogen binding protein Bfp60 of BE1. Differential gene expression analysis of BF9343_4169 detected an upregulation corresponding to a log₂ FC value of 0.86 (1.8-fold increase) in the GA set (Appendix 3).

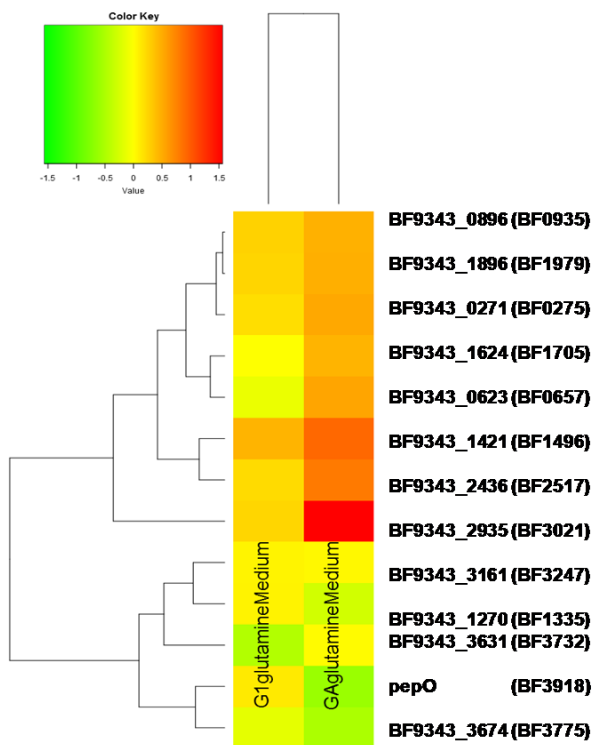
The observed changes in expression of genes putatively involved in host protein binding and degradation, under nitrogen depleting environment led us to analyse the regulation of NCTC 9343-secreted proteases under similar conditions. A specific protease involved in the observed fibrinogen degrading activity in *B. fragilis* was not detected among the 12 proteases studied using gene-deletion and gene-expression strategies coupled with functional analysis using zymography (Chapter 4). However, the medium used for of *B. fragilis* growth in the zymography experiments was not the nitrogen limiting glutamine medium. To determine if nitrogen limitation induced degradation capabilities in the *B. fragilis* secreted proteases to scavenge for nitrogen, the expression levels of protease-encoding genes in the GA set were analysed. Seven out of the 12 secreted proteases that were studied in the present work were

upregulated in GA with reference to N (Appendix 3 and Figure 5.26 A). Two of the proteases were assigned non-significant fold-change values at the assigned q value cut-off < 0.05 . Among the three that were downregulated when compared to reference N, one matched the serine protease family while the other two matched metallopeptidases (Appendix 3 and Figure 5.26 A). The observed pattern of regulation indicates that during nitrogen-limiting conditions, a major proportion of the secreted proteases are upregulated in *B. fragilis*. However, when the 2-fold stringency (\log_2 FC 1) cut-offs were applied, only one of the previously studied proteases, BF3021 (BF9343_2935) could be considered as upregulated in the GA set (\log_2 FC value of 1.5625 corresponding to a 3-fold increase in expression). BF3021 encodes an *Escherichia coli* peptidyl-dipeptidase Dcp homologue (42.03% id in 678 aa) with a Pfam match to metallopeptidase M3 family (score 466.1, E-value $2.5e-137$). These observations might suggest that the rest of the proteases studied are not differentially regulated under nitrogen limiting conditions which led us to extend our analysis to include the other known secreted proteases and peptidases that were not previously examined for their fibrinolytic potential in this study. Among the nineteen protease/peptidase-encoding genes that were analysed, eight were identified to be upregulated in GA when compared to N under no cut-off for \log_2 FC value (Appendix 3 and Figure 5.26 B). Three were assigned non-significant values for q value < 0.05 , however, two of these proteases were upregulated in the G1 set. Of the remaining eight, three were upregulated in G1 when compared to N indicating that the downregulation of these three peptidase-encoding genes in GA might pertain to reduced growth rate conditions. Of the nine upregulated genes in GA, two, namely, BF9343_0144 and BF9343_0145 matched *B. thetaiotaomicron* signal peptidase I BT3319 (25.64% id in 429 aa) and *B. thetaiotaomicron* signal peptidase I BT3319 (85.83% id in 494 aa), respectively. Another gene, BF9343_2903 exhibited similarities to *Chlorobium tepidum* signal peptidase I LepB (35.14% id in 276 aa) and was upregulated in G1 despite being non-significant at the q value cut-off in GA (Figure 5.26 B). Upregulation of signal peptidase-encoding genes is probably a reflection of an increase in the expression of secreted proteins possessing these signal

peptidases. The other GA-upregulated secreted peptidase/protease-encoding genes matched aminopeptidase C, dipeptidyl peptidase VI, alanyl dipeptidyl peptidase, tricorn protease, cysteine protease and an alkaline protease secretion protease, one or more of which might possess a non-specific fibrinogenolytic function under nutrient limiting conditions (Appendix 3 and Figure 5.26 B).

Under a log₂ FC cut-off of 1 (2-fold increase), three protease-encoding genes, BF9343_1980, BF9343_2676 and BF9343_2678 encoding putative homologues of *B. thetaiotaomicron* putative aminopeptidase C (81.54% id in 466 aa), *Thermoanaerobacter tengcongensis* transglutaminase-like enzymes, putative cysteine proteases (31.81% id in 418 aa) and *Bacillus sphaericus* dipeptidyl-peptidase VI (25.09% id in 267 aa), respectively were upregulated. BF9343_0483 and BF9343_3652, encoding homologues of *Proteus mirabilis* metalloprotease transporter component (24.57% id in 407 aa) and *Thermoplasma volcanium* tricorn protease (21.73% id in 1104 aa) with a Pfam match serine peptidase S41B family, respectively were downregulated > a 2-fold in GA.

(A)



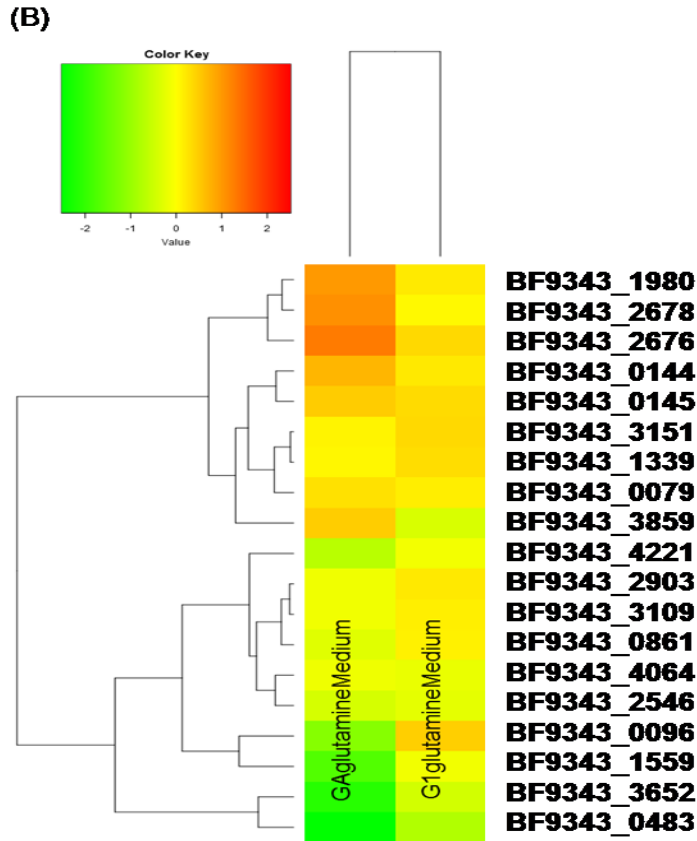


Figure 5.26: Qualitative analysis of the differentially expressed genes involved in host matrix adhesion and proteolytic activity in GA and G1 with reference to N

Heat map based visual analysis of differentially expressed genes involved in adhesion and proteolysis using N as the reference and GA and G1 as the two tests (no cut-offs applied). Green-red colour key represents downregulated to upregulated genes with yellow representing unchanged gene expression. (A) Genes putatively involved in host matrix adhesion and proteolytic activity as per previous studies. The gene nomenclature according to Artemis is given in brackets. A significant proportion of genes are highly upregulated in GA than in G1. (B) Genes annotated as proteases and peptidases. Although 4 genes are downregulated, the other protease/peptidase-encoding genes are considerably upregulated in GA.

5.4.3. Response to shock/stress

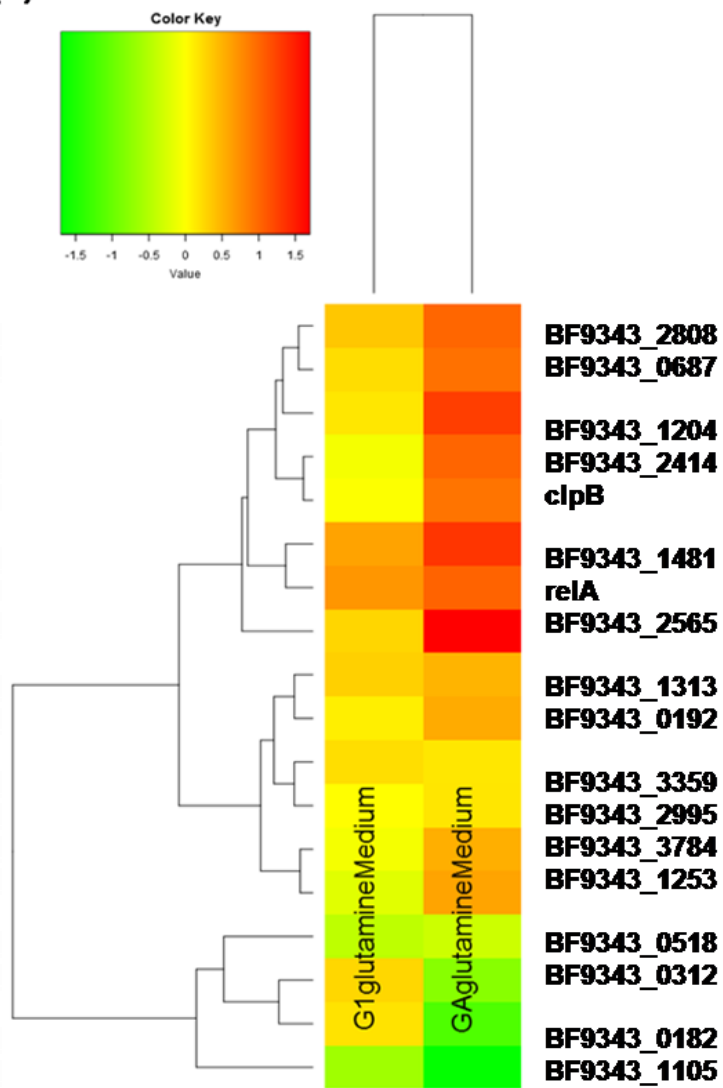
Owing to an upregulation detected in an open reading frame (ORF) homologous to a gene involved in ppGpp metabolism (5.4.1) under nitrogen limitation, the regulation of genes involved in stress/shock responses were analysed in detail in the GA set (Figure 5.27 A). Appendix 4 of the 'gene groups' folder provides the list of differentially expressed genes involved in shock/stress, oxidative stress and DNA repair by annotation. A higher proportion of these genes involved in general shock/stress were upregulated in both G1 and GA sets of reads when compared to the reference N, suggesting that the increase in transcript level was not a mere consequence of a sudden reduction in growth rate observed in the transcriptome derived from lower T_d phase cultures in glutamine medium. The observed upregulation might be the result of a prevalent nitrogen limitation in the glutamine growth medium, which is evident from the slight drop in T_d in cultures from which the transcriptome for G1 was taken, when compared to that of the reference N (Figure 5.27 A). Since nutrient limitation is sensed as a stress by the bacterium, the growth in nitrogen-limiting glutamine medium upregulates an array of genes in response to stress and shock which included genes encoding general stress proteins and a heat shock protein, in the present study. In agreement with the increase in ppGpp metabolism, a *relA* gene encoding a homologue of *B. thtaiotaomicron* GTP pyrophosphokinase BT0700 (82.52% id in 738 aa) was upregulated to a log2 FC value of 1.0394 in GA (Figure 5.27 A). The protein product of BF9343_3784 is annotated as being involved in the maintenance of stationary phase and shows homology to *Salmonella typhimurium* survival protein SurA (25.38% id in 461 aa) (Figure 5.27 A). This gene was upregulated in GA, which is consistent with the observation that the bacteria prepare to enter into stationary phase of growth in response to nutrient limiting conditions (Peterson et al., 2005). However, in the present study, culture replicates, from which RNA was extracted for the GA set of reads, were still growing, albeit at a slower rate, preparative of stationary phase entry. The upregulation of the SurA homologue was lower than 2-fold increase and therefore, not significant. A *clpB* gene product that matched to a *B. thtaiotaomicron*

endopeptidase Clp ATP-binding chain B BT4597 (93.85% id in 862 aa) and to a *Synechococcus* spp. heat shock ClpB protein (54.86% id in 873 aa) was also upregulated in GA. Of the four stress-related genes that were downregulated in GA, two were upregulated in G1 with reference to N (Appendix 4 and Figure 5.27 A).

Another interesting observation was the upregulation of seven genes, that were exclusively related to oxidative stress response, in GA. A similar upregulation of genes involved in peroxide metabolism was observed when the transcriptome of *M. smegmatis* was assessed in relation to nitrogen stress (Williams et al., 2013). *B. fragilis*, an obligate anaerobe requires protection against reactive oxygen species (ROS) when exposed to aerobic conditions following initial stages of infection (Rocha and Smith, 1997). However, the observed upregulation in genes was mostly not significant in G1, suggesting that the differential expression might be growth rate induced in GA. The *oxyR* gene product, redox-sensitive transcriptional activator OxyR, and *sodB* product, similar to *Bacteroides fragilis* superoxide dismutase [Fe] SodB (99.48% id in 193 aa) were upregulated to log₂ fold-change values of 1.26 (2.4-fold increase) and 1.36 (2.5-fold increase) in GA, respectively (Figure 5.27 B). SodB contains iron and functions to convert superoxide anions to hydrogen peroxide. A catalase encoding gene, *katA/katB* and three genes encoding enzymes similar to the Alkyl Hydroperoxide reductase/Thiol Specific Antioxidant (AhpC/TSA) family namely BF9343_2277, *bcp* and *tpx* were also upregulated in GA when compared to the N reference (Appendix 4). Catalases mediate the conversion of hydrogen peroxide into water and oxygen. In a previous study, it was observed that *B. fragilis katB* mRNA was strongly repressed by glucose under anaerobic condition (Rocha and Smith, 1997). However, no transcriptional regulation of *katB* mRNA has been reported under nitrogen limitation. Upregulation of genes encoding homologues of AhpC enzymes was also observed in the GA. AhpC enzymes function to reduce peroxides to alcohols with the aid of a reduced thiol donor and provide protection against oxidative burst generated by macrophages and neutrophils during host immune response to infection (Henry et al., 2012). The transcriptional regulation observed in stress and shock response, especially oxidative stress conveys the idea

that the ability to regulate gene expression in response to changes in nitrogen availability facilitates the survival of this obligate anaerobe in extraintestinal habitats by combating oxidative stress. In the study of *M. smegmatis* under nitrogen limitation, over a 500-fold upregulation in genes involved in ROS detoxification was explained with respect to the upregulation of a gene encoding amine oxidase, which released hydrogen peroxide during breakdown of primary amines into ammonia under nitrogen stress. Therefore, the upregulation in oxidative stress response might be a means by which the cell detoxifies ROS. However, since a significant upregulation of the genes described above were not observed in the G1 set, it implies that the increased expression in GA was related to reduction in growth rate in the present study. The regulation of a gene (BF9343_1121) encoding a protein similar to RecA (99.68% id in 318 aa) was also analysed since *recA* is co-transcribed as an operon alongwith two genes putatively involved in repairing oxygen damage (Steffens et al., 2010). RecA is a major DNA repair protein which performs strand exchange during homologous recombination in bacterial species through SOS response. However, the log₂ FC value of upregulation of RecA homologue was not significant at 0.2841 in GA (Appendix 4).

(A)



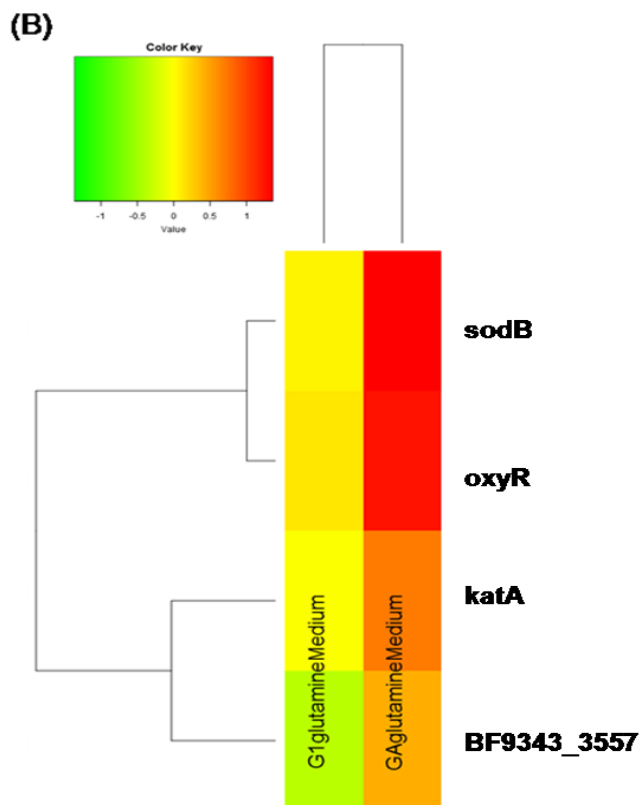


Figure 5.27: Qualitative analysis of the differentially expressed genes involved in response to shock and oxidative stress in GA and G1 with reference to N

Heat map based visual analysis of differentially expressed genes involved in response to shock/stress using N as the reference and GA and G1 as the two tests (no cut-offs applied). Green-red colour key represents downregulated to upregulated genes with yellow representing moderately regulated genes. (A) A major proportion of the genes involved in general response to shock/stress are significantly upregulated in GA. (B) All the four genes involved in oxidative stress response are highly upregulated in GA when compared to G1.

5.4.4. Drug/Antibiotic resistance

Concurrent with the upregulation of stress-response genes, transcriptional activation of a few genes encoding antibiotic resistance and drug transport activity were also identified in the transcriptome of *B. fragilis* under nitrogen-depleting conditions. Appendix 5 of the 'gene groups' folder provides the list of differentially expressed genes involved in drug/antibiotic resistance by annotation. Appendix 6 and Appendix 8 provides the list of genes that are significantly upregulated in the G1 and GA set, respectively, with the N set as the reference. Activation of the multi-drug resistance genes was reflected by an upregulation of multi-drug efflux pumps and drug transporters (Appendix 5 and Appendix 8) which operate to remove toxic drugs from the cytoplasm. Resistance to metronidazole, a 5-nitroimidazole pro-drug, is often attributed to a nitroimidazole resistance gene (*nim*) that encodes a nitroimidazole reductase which mediates the reduction of the nitro group of 5-nitroimidazole to an amino group resulting in the formation of an inactive 5-aminoimidazole (Husain et al., 2013). Two ORFs, BF9343_0350 and BF9343_4246, encoding homologues of *B. fragilis* nitroimidazole resistance protein NimE (38.41% id in 151 aa) and previously sequenced *B. fragilis* 5-nitroimidazole resistance determinant NimD protein (30.43% id in 161 aa), respectively were upregulated in GA by 1.5-fold compared to N (Figure 5.28). The regulation of these genes in G1 were not significant at the specified q value cut-off. Although downregulation was observed in a major proportion of resistance related genes, upregulation of *nimE* and *nimD* is significant owing to the fact that *B. fragilis* is primarily metronidazole-sensitive and the drug has been used in treating infections caused by this opportunistic pathogen (Schapiro et al., 2004). Five ORFs annotated to encode drug resistance proteins were significantly downregulated in GA, as shown by a > 2-fold decrease. These include genes encoding homologues of bicyclomycin resistance protein, NorM multidrug resistance protein and fosmidomycin resistance protein. Bicyclomycin selectively inhibits Rho and is used in the treatment of nonspecific diarrhoea. The drug exhibits activity against a broad spectrum of Gram-negative bacteria and against the Gram-positive bacterium *Micrococcus luteus* (Kohn and Widger, 2005). However, a

specific activity of bicyclomycin against *B. fragilis* has not been reported. Fosmidomycin is primarily an antimalarial drug and therefore does not play a role in the treatment of anaerobic infections (Jomaa, 1999). Nevertheless, activation of metronidazole resistance-related genes and a mechanism to combat oxidative stress observed under nitrogen-limitation-induced reduction in growth rate are reminiscent of the conditions which facilitate the survival of *B. fragilis* in the extraintestinal environment, following the accidental release from the gastrointestinal tract.

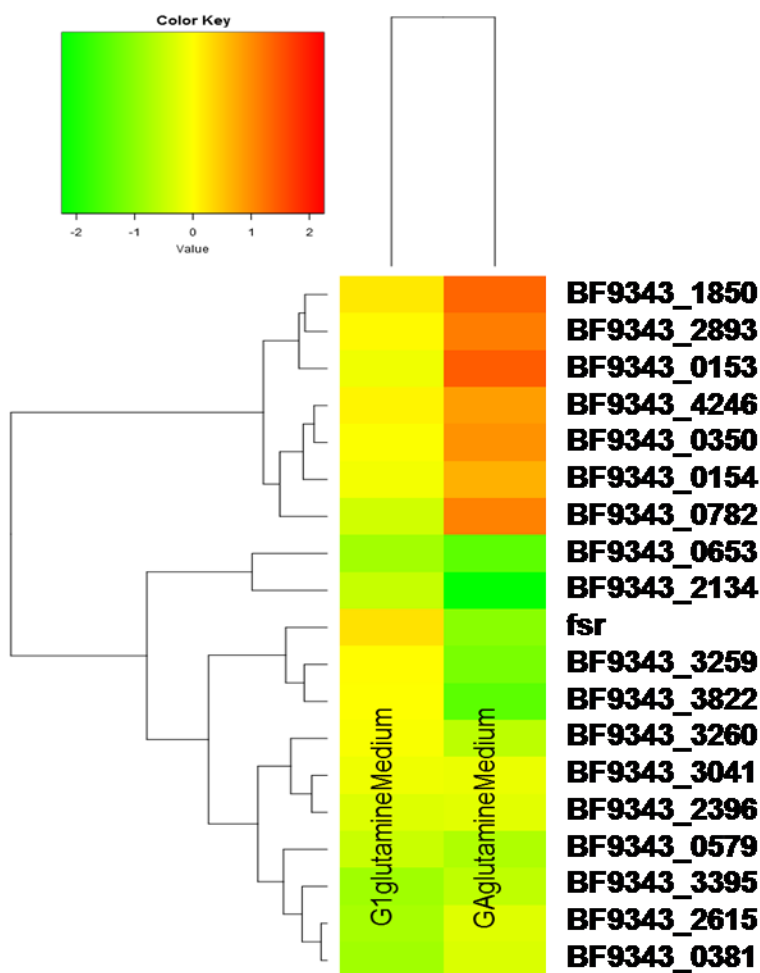


Figure 5.28: Qualitative analysis of the differentially expressed genes involved drug/antibiotic resistance in GA and G1 with reference to N

Heat map based visual analysis of differentially expressed genes involved in using N as the reference and GA and G1 as the two tests (no cut-offs applied). Green-red colour key represents downregulated to upregulated genes with yellow representing moderately regulated genes. Although most of the genes involved in drug/antibiotic resistance and transport are downregulated in GA when compared to G1, seven genes including two that are annotated as metronidazole-resistance genes (BF9343_4246 and BF9343_0350) are upregulated in GA.

5.5. Discussion

The effect of nitrogen on transcriptional regulation of the *B. fragilis* NCTC 9343 genome was studied by altering the nitrogen content of the growth medium from abundant to depleting conditions. A state of nitrogen limitation was achieved by substituting the ammonium sulphate (~0.58%, 0.02 M) present in the conventional defined medium (glucose-DM) with glutamine. The point of limitation of nitrogen was established by growth curve analysis (Figure 5.7). The stress posed by nitrogen limitation was utilised to examine the transcriptome for differential gene expression and address the mechanisms of nitrogen metabolism of this bacterium. The presence of glucose (10%) in both normal medium and nitrogen limiting medium ensured carbon sufficiency under the conditions tested.

5.5.1. Glutamine in the nitrogen-limiting medium

Based upon previous literature which emphasized the necessity of ammonia for the optimal growth of *B. fragilis*, the concentration of ammonium sulphate in the defined medium was altered initially to check for growth limitation (Smith and Macfarlane, 1998; Varel and Bryant, 1974). Studies as early as 1974 by Varel and Bryant, observed that this anaerobic bacterium grew well in defined medium containing NH_4Cl as the nitrogen source. A variant of this medium was used in our study where ammonium sulphate (0.02 M) was used in the glucose-DM instead of ammonium chloride. A decrease in the concentration of ammonia by 10x succeeded in reducing the biomass yield and the growth rate considerably in the present study (Figure 5.4). However, the inability to identify a difference between stationary phase-induced and nitrogen limitation-induced reduction in growth rate addressed the need for an alternative nitrogen source to study limitation. Attainment of similar growth rates and biomass yields by growth in normal glucose-DM and in a medium where the concentration of ammonia was increased by 5x indicated that the nitrogen content is not limiting in the normal glucose-DM. Neither single amino acids, nitrate, urea, nor a complex mix of L-amino acids or peptides were effective in replacing ammonia as the nitrogen source (Varel and Bryant, 1974). This led to the speculation that the

availability of organic nitrogen for *B. fragilis* growth in its natural colonic habitat is limited since the bacterium is unable to utilise nitrogen from organic compounds (Varel and Bryant, 1974). Ammonia, which is an inorganic nitrogen source, was more readily available since ~ 20-25% of a person's daily urea excretion is normally recycled to the intestine from blood and hydrolysed to NH_4^+ by ureolytic bacteria, thereby releasing 3-4 g ammonia. Based upon studies in *E. coli* by Atkinson et al., (2002); Goux et al., (1996); Liang and Houghton, (1981) and Maheswaran, (2003), where nitrogen-limiting growth conditions were attained by using glutamine or single free amino acids such as aspartate and arginine as the sole nitrogen source, growth of *B. fragilis* was examined in low concentrations of the two major intermediary amino acids of nitrogen metabolism, glutamine and glutamate (5.1.1). Additionally, studies have indicated that internal glutamine serves as a sensor of external nitrogen availability and its pool decreases upon nitrogen limitation (Maheswaran, 2003; Yan, 2007). Annotation of the NCTC 9343 genome indicates the presence of genes encoding enzymes involved in glutamine and glutamate utilisation which are required by the bacterium to utilise these metabolic products for nutrition and biosynthesis. Interestingly, no growth was observed on glutamic acid containing medium (GCM) whereas *B. fragilis* attained growth saturation with overnight OD_{600} 2.0 on glutamine containing medium (GNM). The limiting condition in late-log phase imposed a slow growth rate of the cell population on reaching the preparatory stationary phase in a growth curve. A decrease in growth rate is the most direct indication of nutrient limitation and therefore was of significance in the present work on transcriptome analysis which identified a change in growth rate in the log phase of *B. fragilis* growth in GNM. The inability to utilise glutamic acid by *B. fragilis* is consistent with previous reports (Varel and Bryant, 1974). However, the ability to grow on GNM with glutamine as the sole nitrogen source has not been previously documented to the best of our knowledge. Ammonia which is found at millimolar concentrations in the large intestinal lumen as a result of amino acid deamination and urea hydrolysis by intestinal microbes, increases proportionally to the concentration of alimentary proteins. The concentration of ammonia in the colon

represents a balance between deamination by amino acids, subsequent uptake by bacterial cells as nitrogen source for protein synthesis and colonic absorption (Davila et al., 2013). Also, the luminal concentration of ammonia increases progressively from ascending to descending colon owing to the low pH and high carbohydrate availability in the proximal colon. The reduced net production of ammonia in the proximal colon might exert a condition of nitrogen limitation on *B. fragilis*. In addition to the direct assimilation of ammonia by the intestinal microbiota, recent work by Eklou-Lawson et al., (2009) suggests that part of the ammonia produced is condensed with L-glutamate through the activity of glutamine synthetase for the synthesis of glutamine thereby ensuring that the intestinal ammonia concentration does not reach toxic levels. The activity of glutamine synthetase is relatively high in colonocytes in comparison with enterocytes. Therefore, a pathway has evolved in *B. fragilis* that enables utilisation of glutamine as the sole nitrogen source which is consistent with the observation of the bacterium growing in the glutamine medium.

5.5.2. Inability to grow on glutamic acid as the sole nitrogen source

The inability of *B. fragilis* to utilise glutamic acid in a similar manner to glutamine was confirmed by doubling the concentration of glutamic acid in GCM to 0.2% and examining growth in comparison with *E. coli* MG1655 strain (Figure 5.5 and Figure 5.6). GCM favoured *E. coli* growth when compared to *B. fragilis*. However, GNM was considered a more favourable growth medium even for *E. coli*, than GCM (Figure 5.5 and Figure 5.6). Studies have indicated that intestinal microbes prefer to assimilate nitrogen in the form of peptides or ammonia, rather than free amino acids. Eventhough *B. fragilis* and ruminobacter *B. amylophilus* failed to utilise amino acids as the sole energy source, amino acids are assimilated and incorporated by these organisms via six transport systems if ammonia is available, leading to the production of SCFA and ammonia, the major products of amino acid fermentation (Smith and Macfarlane, 1998). It was found that *B. amylophilus* proteases have a tryptic-type specificity when supplied with tryptic peptides and amino acids (Hullah and Blackburn, 1971). Assimilating peptides instead of free amino acids is

energetically advantageous and would be physiologically important in an energy deficient environment such as the large intestine (Pittman and Bryant, 1964).

Since adequate literature was not available on the uptake of glutamine and other single amino acids by *B. fragilis*, the search was extended to other species of the *Bacteroides* genera to address the rationale behind differential utilisation of glutamine and glutamate observed in *B. fragilis*. Miles et al., (1976) showed that the addition of individual amino acids to a trypticase-yeastextract-hemin medium affected growth rates and final yields of asaccharolytic and saccharolytic strains of *B. melaninogenicus*, an opportunistic pathogen inhabiting oral and intestinal niches, with maximum growth attained in aspartate. However, since fermentation of proteinaceous material was the main energy source of this bacterium it was more permeable to peptides than amino acids. The addition of 50 mM glutamine induced maximum uptake of glycine or alanine by *B. melaninogenicus* when added to a growth medium containing trypticase. Uptake of amino acids and peptides was stimulated by glutamine, and was concentration- and energy-dependent (Lev, 1980).

In a similar manner, the only compounds that replaced ammonia as the main nitrogen source in *Bacteroides ruminicola* were tryptic protein digests, peptide-rich fractions of tryptic digests of casein and octapeptides (Pittman and Bryant, 1964). It was found that ammonia served as the main source of nitrogen and that glutamine and potassium cyanide supported the growth of both the strains tested. However, it was doubtful whether glutamine and potassium cyanide were directly utilised since production of ammonia was detected from each compound in uninoculated media at a rate sufficient to account for the cellular growth observed in the inoculated media. Therefore, ammonia might be the only low molecular weight compound used efficiently as a nitrogen source for growth by *B. ruminicola*. The ability of the bacterium to grow on acid hydrolysate of casein during prolonged incubation was attributed to ammonia production, presumably due to the release of intracellular amino acid deaminases from old cultures (Pittman and Bryant, 1964). The reason for peptides to be a favoured source of nitrogen than its constituent amino acids was

either due to the difficulty in specific amino acid transport into the cell or due to the destruction of amino acids by the bacterium prior to its utilisation. Peptides containing more than two or three amino acids were observed to be most effective in bacterial nutrition. However the study by Stevenson, (1979) which focused on the uptake of amino acids in complete defined medium containing ammonia as nitrogen source identified six different amino acid uptake systems specific for L-isomers and barring proline transport in *B. ruminicola* confirming that the failure of free amino acids to act as sole nitrogen source is not due to the absence of transport systems. Centrifugation and resuspension of 18h culture pellets in fresh basal medium led to a 10-fold increase in amino acid uptake. Therefore it was hypothesised that the initial inhibition in aa uptake could be due to the depletion of a required substance or assimilation of an inhibitory substance resulting in bacterial growth. Omission of ammonium sulphate from the culture medium resulted in slowing and finally cessation of growth as well as amino acid uptake. Also, high levels of acetate, a volatile fatty acid, inhibited amino acid uptake. It was observed that the amino acid uptake required active cellular growth and effects of inhibitors of uptake were predicted to be either due to a direct effect on the energy mechanism of transport or due to indirect effects on cell energy levels. In conclusion, the uptake systems in *B. ruminicola* facilitate auxiliary supply of amino acids for immediate use in biosynthesis despite failing to fulfil the entire nitrogen requirement of the bacterium. These observations suggest that the ability of *B. fragilis* to utilise glutamine as the sole nitrogen source might be due to a mechanism by which it can generate ammonia, the principal nitrogen source for the bacterium, thereby initiating the conventional GS/GOGAT or GDH nitrogen assimilation pathways. The failure to utilise glutamic acid could be attributed to the inability to generate ammonia.

5.5.3. Choice of time points for RNA-Seq

Analysis of the NCTC 9343 transcriptome at the nutrient and growth limiting time point in GNM was performed to elucidate the regulatory mechanisms which control acquisition of nitrogen and the effects of limited nitrogen availability on this

bacterium. With the exception of studies dealing with nutritional requirements and regulation of enzymes involved in nitrogen metabolism, a whole transcriptome analysis of *B. fragilis* by RNA-Seq under nitrogen limiting conditions had not been previously reported. While examining the differential regulation of genes involved in *E. coli* nitrogen metabolism under varying nitrogen-limiting conditions, Atkinson et al., (2002) had suggested that expression analysis in glutamine-containing medium was risky since the bacterial growth rate dropped by mid-log phase (0.5 OD₆₀₀) which led to a sharp increase in the expression of genes involved in nitrogen assimilation. However, the growth rate of *B. fragilis* in GNM in the present study was similar to that in normal medium till late-log phase (0.8 OD₆₀₀). The drop in growth rate at this OD value, defined by a change in doubling time from 90 min ± 6 min to 390 min ± 20 min, thereby provided an opportunity to use two RNA samples taken from cultures with similar growth rate (78 min ± 4.2 min and 90 min ± 6 min), one from normal medium (N set) and the other from GNM (G1set) to compare with the test GA set (0.8 OD₆₀₀) in GNM (Figure 5.7) in RNA-Seq experiments. The reproducibility of the growth curve was confirmed by performing the experiment in five biological replicates.

5.5.4. Analysis of differentially regulated genes under nitrogen-limiting conditions

5.5.4.1. Nitrogen metabolism

RNA-Seq analysis of N vs GA differential gene expression indicated a significant downregulation of *glnA* (> 2-fold decrease) in *B. fragilis* (Figure 5.21 B). The differential regulation of *glnA* in G1 and GA with respect to N was most likely a reflection of growth rate since *glnA* expression had increased by over 2-fold in G1. Previous studies have confirmed the upregulation of GS, the *glnA* gene product, during conditions of nitrogen limitation thereby activating the GS/GOGAT pathway for nitrogen assimilation in *E. coli* (Shimizu, 2013; Yamamoto et al., 1984). GS expressed by *B. fragilis* was identified as a hexamer of ~490000 Da, with each identical subunit weighing 75000 Da (Southern et al., 1987). The enzyme was longer

than and exhibited no similarity to the previously characterised GSI and GSII enzymes, the holoenzymes of which were dodecamers and octamers, respectively, and therefore led to the introduction of the GSIII group of enzymes to categorize the GS expressed by *B. fragilis* (Hill et al., 1989). The inability of *B. thetaiotaomicron* to recognise the *B. fragilis glnA* promoter in a reporter gene assay suggested that species belonging to the colonic *Bacteroides* genera are genetically distinct from one another (Abratt et al., 1993). However, in spite of less than 50% homology between the *E. coli* and *B. fragilis* strain Bf-1 GS encoding genes, *E. coli glnA* deletion mutant when complemented with a recombinant plasmid carrying *B. fragilis glnA* was able to utilise ammonium sulphate as the sole nitrogen source. GS activity was repressed in the presence of glutamate and glutamine (Southern et al., 1986). The majority of previous studies on *B. fragilis* were focused on ammonia-limiting nitrogen regulation since glutamine utilisation by the bacterium had not been analysed in the past. The speculation that the observed *glnA* downregulation in GA was a function of growth rate reduction (Figure 5.21 B) was consistent with the study by Bren et al., (2013) on *E. coli* assimilation genes with respect to limiting amounts of nitrogen in M9 minimal growth medium. GFP reporter assays revealed moderate activity of the *glnA* during growth and, at about one generation preceding cessation of growth owing to nutrient limitation, the promoter activity rose by about a 4-6 fold. However, the activity dropped to low levels when growth stopped. In the present study, in G1 of GNM where the growth rate was similar to normal medium, the detected upregulation of *glnA* might allow the cells to maintain their growth rate for about a few more generations where they utilise the low glutamine levels. However, GA in the present study was not characterised by a complete cessation of growth although there was a decrease in growth rate when compared to the faster-growing phase in GNM. The observed decrease in growth rate might have contributed to *glnA* downregulation in GA. In a similar manner, *glnA* mutants complemented with *glnRAP-lacZ* fusion constructs in *B. subtilis* failed to induce GlnA expression in a slow-growing *glnA* mutant strain with a doubling time of around 150 min (Hu et al., 1999). Nevertheless, an ORF encoding a homologue of *B. subtilis* GS

(BF9343_2258) was slightly upregulated in GA, which might contribute to nitrogen assimilation during growth under nitrogen limitation (Figure 5.21 B). A gene that was 100% identical to the gene encoding GOGAT, the second enzyme of GS/GOGAT pathway for nitrogen assimilation and the major glutamine hydrolyzing enzyme, has not been annotated in the *B. fragilis* NCTC 9343 genome. However, an ORF, BF9343_0887, was observed to encode a product similar to *B. thetaiotaomicron* NADPH-dependent glutamate synthase small chain (91.35% id in 763 aa) on raw annotation (Figure 5.21 B). Although the expression level was unchanged in G1 and GA (log₂ FC cut-off 1 and q value <0.05), the enzyme might contribute to the initial glutamate synthesis from glutamine in the absence of ammonia. Previous studies have measured GOGAT activity as NADPH oxidation depending on glutamine and ketoglutarate. It was also observed that GOGAT activity was very low and might be associated with GS to produce glutamate from α -ketoglutarate and ammonia but not strong enough to play a main role in ammonia assimilation by *B. fragilis* (Yamamoto et al., 1984). Therefore, alternate methods of ammonia generation that supported the GOGAT activity were examined in the present study.

While considering the GDH pathway for nitrogen assimilation, a gene annotated as *gdhB* was downregulated in GNM, however another gene annotated as *gdhB2*, which shared similarity with *gdhB* (69.21% id in 445 aa) was upregulated in both G1 and GA (Figure 5.21 B). Studies have identified two electrophoretically distinct GDHs in *B. fragilis*: GDH with dual-specificity for pyridine nucleotides and NAD⁺-specific GDH. In the dual-specificity enzyme, NADPH-GDH has an anabolic role whereas NADH-GDH fulfils a catabolic function (Abrahams and Abratt, 1998). NADPH-GDH (*gdhA*) was found to be induced at low ammonia concentrations whereas high ammonia concentrations partially induced NADH-GDH. The highest NADH-sp (*gdhB*) activity was observed when *B. fragilis* was grown in the presence of organic nitrogen such as tryptone as the nitrogen source. A similar upregulation was observed under low concentration of glutamine, an organic nitrogen source, in the present study. The regulation of *gdhA* is not known since the gene has not been

identified in NCTC 9343 genome. Regulation of *gdhB* was suggested to be linked to proteolytic activity in vivo and hence to pathogenesis. This is consistent with the detection of reduced protease activity in *B. fragilis* under excess nitrogen (ammonia) conditions (Gibson and Macfarlane, 1988a). *gdhB* also displayed high amino acid identity and similarity to NADH-GDH of the oral pathogen *Porphyromonas gingivalis* which supported the predicted role of GDH in the pathogenic transition of *B. fragilis* (Abrahams and Abratt, 1998).

Yamamoto et al., (1984) analysed the activities of NADPH and NADH linked glutamate dehydrogenase and glutamine synthetase in *B. fragilis*. In their study, it was found that the cells exhibited higher activity of glutamate dehydrogenase at low ammonia concentrations, unlike *E. coli* which adopted the GDH pathway of ammonia assimilation under nitrogen excess conditions. However, GS was also observed to be higher in cultures with limited ammonia. In the rumen bacterium *B. amylophilus* GS was inactivated by addition of excess ammonia whereas GDH was active across all conditions tested. GOGAT activity was not detected in ammonia-limited continuous cultures. Nevertheless, activities of all three enzymes were detected in *B. fragilis* batch cultures harvested at early stationary phase of growth. In batch cultures, the highest NADPH-linked GDH activity was observed in cells grown in 0.5 mM ammonia. In NADPH-linked reactions with an optimal pH range (7.8-8.4) K_m values for ammonia, ketoglutarate and NADPH were 0.8 mM, 0.15 mM and 7 μ M, respectively. Low K_m values for ammonia and ketoglutarate in reactions with NADPH ensured that GDH functions in ammonia assimilation even at low ammonia conditions. When the ammonia concentrations were increased to 5 mM and 50 mM, GDH specific activities were 50% and 15%, respectively of that observed at 0.5 mM concentration. Moreover, NADPH- and NADH-linked GDH activities were not altered by the addition 20 mM L-glutamate to the growth medium (Yamamoto et al., 1984). Further studies on the dual coenzyme-specific GDH regulation by extracellular concentration of ammonia in the *B. fragilis* group identified that K_m values of GDH for NH_4Cl was in the range of 1-5 mM in NADPH-dependent amination and that specific activity of NADPH-GDH was 100 times more than GS

and GOGAT which was twice as high as that in other species under low ammonia conditions (Yamamoto et al., 1987).

In a study by Liang and Houghton, (1981) that analysed the coregulation of oxidised NADPH and GDH activities in enteric bacteria in differing concentrations and sources of nitrogen, it was found that at concentrations exceeding 20 mM ammonia, both GS and NADP-linked GDH activities declined in *E. coli* owing to the direct involvement of NH_4^+ in GDH for glutamate synthesis. In *Klebsiella aerogenes* GDH was regulated inversely to that of GS. Glutamine as the sole nitrogen source in the presence of glucose led to an increase in the levels of GS with repressed GDH and transhydrogenase. NAD(P)^+ transhydrogenase catalyses the reversible transfer of reducing equivalents between the pyridine nucleotides and therefore is of importance in reactions involving GDH. Incorporation of excess NH_4Cl restored high levels of GDH and transhydrogenase concurrent with repressed GS activity.

In nitrogen-fixing *Bacillus macerans* it was found that ammonia is assimilated predominantly by the GDH pathway during nitrogen fixation. GDH of *B. macerans* was found to be specific for NADPH and no NADH-GDH activity was detected. GOGAT activity in crude cell extracts was too low to be detected through measurement of the rate of oxidation of NADPH (Kanamori et al., 1987). Therefore it was concluded that GDH plays a significant role in ammonia assimilation in organisms that have a GDH with K_m for NH_4^+ in the range of 1-5 mM with very low levels of GOGAT. The lower energy requirement of the GDH pathway might be advantageous to nitrogen-fixing cells that require 4-29 ATP molecules to reduce 1 molecule of nitrogen. Ammonia assimilation by GDH, provided that it has moderate affinity for NH_4^+ , seemed feasible during nitrogen fixation.

The studies on GDH regulation described above, confirm the existence of a variation in GDH activities across species and from the existing literature, it could be predicted that *B. fragilis* grown at low concentrations of glutamine activates a homologue of catabolic NADH-GDH encoded by *gdhB2* which results in glutamate formation. However, GDH and GS enzymes are dependent on ammonia for their

activity. The requirement for ammonia could be partially satisfied by the deamination of GOGAT-formed glutamate and of the compounds formed from the activity of glutamine amidotransferases (GATases), which were identified to be upregulated in GA (Appendix 1 and Figure 5.21 B). The enzymatic mechanism of glutamine amidotransferases, which comprise enzymes such as carbamoyl phosphate synthase, involves binding of glutamine to the cysteine residue in the glutamine binding site (small subunit), release of ammonia from the amide group of glutamine followed by its transfer to an ammonia binding site often located on a separate subunit (large subunit) where it is used for an amination reaction. Therefore interactions between small and large subunits of the GATase enzyme mediates its catalytic function (Cunin et al., 1986).

Another interesting observation was the significant upregulation of glutamine-hydrolysing asparagine synthetase B (AsnB) in GNM. The upregulation was commensurate with that of aspartate amino transferases and an aspartate/glutamate transporter potentially involved in conversion of glutamate to aspartate (Figure 5.21 B). Synthesis of asparagine from glutamine and aspartate via AsnB generates ammonia which could support the requirements of *B. fragilis* for nitrogen under the limiting conditions that were observed in the biphasic growth in GNM. In a separate study on carbon-source-dependent nitrogen regulation in *E. coli*, it was found that the bacterium possesses a low affinity glutamine transport system mediated by periplasmic asparaginase (AsnB) which displays low glutaminase activity. The transport system was found to be activated either by CRP-cAMP complex or when *E. coli* was grown on poor carbon sources in the presence of glutamine (Maheswaran, 2003). Thus, a high concentration of exogenous glutamine was found to increase intracellular glutamine levels through an indirect mechanism which involved periplasmic glutaminase activity and uptake of hydrolysed ammonium followed by intracellular GS-mediated ammonium reaccumulation. However, in the present study, glucose, which is an efficient carbon source was used in conjunction with limiting glutamine concentration and hence the activation of a low affinity transport system is not likely. Goux et al., (1996) had observed the generation of ammonia by

E. coli through a previously uncharacterised degradation pathway of arginine, formed by aspartate catabolism when the bacterium was grown on aspartate as the sole nitrogen source. It was suggested that arginine might be an intermediate of aspartate metabolism since amino nitrogen of aspartate is incorporated into imino nitrogen of arginine via the Krebs-Henseleit pathway. Concurrent with this finding, in the present study, a gene annotated as *argG* (arginosuccinate synthase) was upregulated in GA (1.53 log₂ FC) (Appendix 2). ArgG catalyses the condensation of aspartate and citrulline to form arginosuccinate, the immediate precursor of arginine (Figure 5.23). An increase in expression of a homologue of arginine deiminase (BF9343_2385) was observed in GA (Appendix 2). Arginine deiminase, together with ornithine carbamoyltransferase and carbamate kinase enzymes comprise the arginine deiminase (ADI) pathway which catalyse the step-wise conversion of arginine into ornithine, ammonia and carbon dioxide with the formation of 1 mol of ATP/mol of arginine consumed (Cunin et al., 1986) (Figure 5.24). The ADI pathway has been identified in lactic bacteria, Bacilli, *Pseudomonas* spp., *Aeromonas* spp., Clostridia, *Mycoplasma* spp. and Cyanobacteria. Besides arginine utilisation for the generation of energy in *Streptococcus faecalis*, *Mycoplasma* spp., *Bacillus* spp., *P. aeruginosa* and Halobacteria, the ADI pathway might provide nitrogen when arginine is the only source for the bacterium. Utilisation of arginine via the ADI pathway as a nitrogen source has been studied in *Aeromonas formicans* and *Bacillus licheniformis*. In a *P. aeruginosa* strain, the three enzymes in the ADI pathway were induced strongly up to 50 or 100 fold by a shift from aerobic growth condition to a very low oxygen tension (reviewed in Mercenier et al., 1980). Arginine utilisation either via ADI pathway or decarboxylation is also of significance under conditions of nutrient depletion.

The present study suggests a similar mechanism of arginine utilisation as a source of ammonia in *B. fragilis* too. Nitrogen catabolite repression of ADI pathway by ammonia occurs in the presence of a good energy source such as glucose, consistent with the downregulation of arginine deiminase homologue observed in the normal defined medium. Although an ORF encoding a homologue of *P. aeruginosa*

ornithine carbamoyltransferase (OTCase) (BF9343_0422) was downregulated in GA, an increase in expression was detected in G1 (Appendix 2). Moreover, a gene annotated as *pyrB* with a Pfam match to aspartate/ornithine carbamoyltransferase was upregulated in G1 and GA (Appendix 2). Ornithine carbamoyltransferase serves two functions in arginine metabolism; arginine biosynthesis in which it catalyses the transfer of carbamoyl moiety of carbamoyl phosphate to 5-amino group of ornithine, forming citrulline and in the catabolic arginine deiminase pathway where it mediates the thermodynamically less favoured reverse reaction favouring citrulline phosphorolysis thereby yielding ornithine and carbamoyl phosphate. Organisms using both these functions possess distinct anabolic and catabolic OTCases. Repression of anabolic OTCase was detected in growing *B. subtilis* cells but was induced by arginine at the end of exponential growth (Cunin et al., 1986).

5.5.4.2. Secreted proteins

Consistent with the studies by Gibson and Macfarlane, (1988a, 1988b) which demonstrated that the nutrient limitation in stationary phase induced secretion of proteases by *B. fragilis*, a higher proportion of secreted proteases (possessing signal peptides) were identified to be upregulated in GA (Figure 5.26 A and B), albeit only four were upregulated to a significant > 2-fold increase. Although the culture had not completely attained stationary phase, there was a drop in growth rate owing to nitrogen limitation in the glutamine medium.

In *Salmonella typhimurium*, an enteric bacterium, nitrogen starvation was found to degrade over 40% of the bacterial protein during an 8h period. The observed degradation provided amino acids which were a source for synthesis of new proteins under shift-down conditions (Miller, 1975). Adigüzel et al., (2009) investigated the proteolytic activities of two *Bacillus cereus* strains grown in peptide-restricted, carbon sufficient medium and observed the sequential induction of different secreted proteases. Peptone was the preferred nitrogen source and its limitation led to the production of collagenolytic and elastolytic proteases which were later replaced by keratinolytic proteases. Thus it was concluded that *B. cereus* adjusted its proteolytic

affinity profile in response to organic nitrogen and sequentially secreted proteases with activities targeted against increasingly inaccessible proteinaceous substrates as nutritional availability in the environment deteriorated.

In addition to secreted proteases, three genes encoding signal peptidases were found to be upregulated in *B. fragilis* grown on nitrogen-limiting GNM (Figure 5.26 B). The direction of exported bacterial proteins, such as those involved in the usage of nutrient sources, in cell-cell communication, and proteins with antimicrobial activity and virulence factors, to transport machineries in the cytoplasmic membrane requires their synthesis as precursors with an N-terminal signal peptide (Voigt et al., 2006). Signal peptides are recognised in the cytoplasm and the proteins are targeted to the transport machinery in the membrane in an energy-dependent manner. Once the protein reaches the outer membrane, the signal peptide is cleaved by signal peptidases belonging to the serine protease type, to form the mature protein. Therefore, the activation of signal peptidases in *B. fragilis* suggest an upregulation of the proteins cleaved by them thereby enhancing protein transport and secretion. Different pathways have evolved in bacteria to secrete proteins, of which the major ones include the general secretory (Sec), twin arginine translocation pathway, ABC transporters and pseudopilin pathway. Of the 79 identified proteins containing N-terminal signal peptides in the *B. licheniformis* proteome under starved conditions, 18 were involved in transport processes and several among them were ABC transport binding proteins (Voigt et al., 2006). Consistent with this, an upregulation of proteins involved in ABC transport was observed in the present study (Appendix 6 and Appendix 8).

Apart from bacteria, serine proteases in pathogenic fungi, have been related to pathogenesis and nitrogen starvation. The *Paracoccidioides brasiliensis* Serine Protease (PbSp) transcript and the protein (serine protease) encoded by *Paracoccidioides brasiliensis* were induced during nitrogen starvation. PbSp, a serine thiol protease cleaved murine laminin, human fibronectin, type IV collagen and proteoglycans. Analysis of the amino acid promoter region revealed the presence

of a nitrogen metabolite repression (*nmr*) region binding protein responsible for positive regulation of genes in response to nitrogen metabolite presence. Therefore PbSp could be a molecule regulated by nitrogen metabolite presence (Parente et al., 2010). Since *B. fragilis* is an actively proteolytic organism, the proteases upregulated during nitrogen starvation need to be examined for host protein degradation.

5.5.4.3. Shock/Stress response

Brown et al., (2014) identified a link between the nitrogen stress response and stringent response in *E. coli*. Nitrogen starved *E. coli* cells synthesised guanosine tetraphosphate (ppGpp), an effector alarmone of bacterial stringent response. Levels of ppGpp in *E. coli* were modulated by ppGpp synthetase, RelA and ppGpp synthetase/hydrolase, SpoT. Phosphorylation of NtrC following nitrogen starvation was found to activate *relA* from a σ^{54} -dependent promoter in *E. coli* and other members of the Enterobacteriaceae family. RelA and SpoT contribute to stress adaptation, antibiotic resistance, expression of virulence traits and acquisition of persistent phenotype in pathogenesis. In agreement with this finding, an upregulation of an NtrC homologue (BF9343_2369) identified in GA in the present study was coupled with an upregulation of a ppGpp homologue (BF9343_0658) sharing similarities with *B. subtilis* RelA (Figure 5.21 A).

Reactive Oxygen Species (ROS) in the form of the superoxide anion radicals, hydrogen peroxide and hydroxyl radical can lead to oxidative damage of lipids, proteins and nucleic acids resulting in mutagenesis and cell death. Organisms evolve defence mechanisms to combat oxidative environment through genes that encode proteins involved in oxidative stress response. Rocha and Smith, (1997) examined the regulation of *B. fragilis katB* mRNA (catalase) by oxidative stress and carbon limitation and detected a 15- fold increase when anaerobic mid-log phase cultures were exposed to oxygen, oxygen paraquats or hydrogen peroxide and carbon limitation. Catalases prevent the accumulation of toxic hydrogen peroxide. However, under anaerobic conditions, levels of *katB* mRNA were found to increase in a growth-dependent manner up to the late-log phase or early-stationary phase followed

by a decrease in stationary phase. The *katA/katB* upregulation detected in GA (0.707 log₂ fold-change) is different from the conventional upregulation detected in anaerobic condition since the observed depression in growth curve, at which the GA set of samples were extracted, was induced by nitrogen limitation (Figure 5.27 B). Although earlier studies failed to identify a link between nitrogen limitation and *katB* mRNA regulation, total RNA-Seq analysis carried out in the present study detected an upregulation under nitrogen-deprived conditions albeit by < 2-fold. Induction of *katA/katB* catalase mRNA in the absence of oxygen under nitrogen limitation suggests that this enzyme may be important to protect cells against eventual oxygen exposure or hydrogen peroxide decomposition that might be formed from stationary phase or dormancy metabolism. Since carbon and nitrogen regulatory pathways are related, nitrogen limitation might also affect catalase expression as observed in the present study. In *E. coli*, the redox-sensitive regulator OxyR transcriptionally regulates *katG* and nine other genes induced in the presence of sublethal doses of hydrogen peroxide. OxyR acts as repressor under reduced condition and shifts to act as transcription activator under oxidative conditions (Rocha and Smith, 1997). The *oxyR* gene was upregulated both in the GA and G1 set of samples extracted from GNM (1.26 and 0.13 log₂ fold-changes, respectively), however the significant differential regulation in GA might be a consequence of the reduction in growth rate (Figure 5.27 B). OxyR and KatA/KatB upregulation might facilitate resistance to the oxidative burst of macrophages and phagocytic cells of the immune system encountered by *B. fragilis* on accidental escape from the gastrointestinal tract during transition to an opportunistic pathogen. Ohara et al., (2006) demonstrated that the superoxide dismutase-encoding gene of the obligate anaerobe *P. gingivalis* was regulated by the OxyR transcription activator. Investigation of oxidative-stress response proteins in *P. gingivalis* detected two proteins that were predominantly upregulated in oxidative conditions. N terminal amino acid sequencing revealed these two proteins to be superoxide dismutase and alkyl hydroperoxide reductase which were positively regulated by OxyR. OxyR functioned as an intracellular redox-sensor rather than peroxide sensor in *P. gingivalis*. The superoxide dismutase-encoding

gene (*sodB*) of *B. fragilis*, which is taxonomically related to *P. gingivalis*, was observed to be inducible by redox stresses but not controlled by *oxyR*. In addition to OxyR, SodB was also found to be upregulated in the present study (1.36 log₂ fold-change) (Figure 5.27 B). *E. coli* cells produce two types of superoxide dismutase (Sod), Mn-containing Sod (SodA) and Fe-containing Sod (SodB) which converts superoxide to hydrogen peroxide and *sodB* was found to be constitutively expressed independent of oxidative stress. An upregulation of genes annotated as *bcp*, *tpx* and a homologue encoding AhpC/TSA family were also identified in GA, potentially in relation with oxidative stress response (Appendix 4). Alkyl hydroperoxide reductase subunit C (AhpC), thiol-dependent peroxidase (Tpx) and bacterioferritin-comigratory protein (Bcp) are peroxiredoxins which confer resistance to hydrogen peroxide in other bacteria.

5.5.4.4. Antibiotic resistance

B. fragilis is inherently resistant to a wide range of drugs which makes the bacterium a challenging candidate for current therapeutic regimens. Resistant *B. fragilis* strains have been associated with adverse outcomes including increased morbidity and mortality. However, metronidazole has been regarded as one among a few drugs that are still reliable for treatment of infections involving *B. fragilis* and is the most commonly prescribed drug worldwide for anaerobic infections. Metronidazole belongs to the class of 5-nitroimidazole drugs which also include tinidazole and ornidazole, and exerts antimicrobial action via inhibition of DNA synthesis (Schapiro et al., 2004). The drug was first introduced against protozoal infections in the middle of the 20th century. Metronidazole-resistant *B. fragilis* strains have been isolated more frequently, resulting in adverse outcomes including death or amputation (Sherwood et al., 2011).

While determining the response of nitrogen limitation to the transcriptional regulation of genes involved in drug/antibiotic resistance in the present study, a > 1.5-fold increase in the expression of BF9343_4246 and BF9343_0350 gene products, which encode homologues of *B. fragilis* 5-nitroimidazole resistance

determinant proteins NimD and NimE, respectively, was detected in the GA set of samples (Figure 5.28). Activation of metronidazole from its prodrug state requires partial reduction of the nitro group to the toxic nitroso radical intermediate that then binds to DNA causing single and double-stranded breaks (Zahoor et al., 1987). Pathogens which do not possess this activation mechanism are intrinsically resistant to metronidazole. However, active metronidazole resistance in sensitive organisms is attributed to a nitroimidazole resistant gene (*nim*), which encodes a nitroimidazole reductase involved in the reduction of nitro group to amino group to make inactive 5-aminoimidazole as observed in *B. fragilis* (Carlier et al., 1997). Nim homologs have been identified in both Gram-positive and Gram-negative aerobic and anaerobic bacteria. Therefore the *nim* family is ancient, widespread and found on chromosomes or on mobilizable plasmids, thus posing a significant threat to the continuing utility of 5-nitroimidazole drugs (Husain et al., 2013). Molecular detection of *nim* genes in *Bacteroides* isolates was first described in 1996 using specific primers (Trinh and Reysset, 1996). To date, ten *nim* genes (*nimA-J*) have been described in *B. fragilis* (Alauzet et al., 2009; Gal and Brazier, 2004; Löfmark et al., 2005). Other mechanisms of metronidazole resistance in *B. fragilis* include upregulation of transcript levels of genes encoding proteins involved in drug efflux, alterations in DNA repair systems, metabolic changes and lack of activation of the metronidazole molecule. In agreement with this observation, a multitude of genes involved in drug transport and drug efflux systems were identified as upregulated in nitrogen-limiting GA in the present study (Appendix 5 and Appendix 8).

The upregulation of *nim* genes identified in GA in the present study suggests that nitrogen-limiting conditions in *B. fragilis* confers increased metronidazole resistance to this anaerobe. A similar dependence of antibiotic resistance on nitrogen conditions was previously reported in Staphylococcal and Streptococcal species. In Gram-positive bacteria nitrogen-dependent transcriptional regulation is primarily mediated by negative regulators GlnR and TnrA, unlike NtrB/C positive regulation in Gram-negative bacteria. It was observed that GlnR regulation and presence of glutamine were important for maintaining methicillin resistance in *Staphylococcus*

aureus. GlnR target genes, *glnA* and *glnP*, mediated colonization and survival of *Streptococcus pneumoniae* during infection. Additionally, the repression of *lctE* and *alsS* genes by NreC of NreABC system, similar to fumarate nitrate reductase regulator (Fnr), was necessary for biofilm formation and maturation in *S. aureus* (Somerville and Proctor, 2009). Nevertheless, nitrogen limitation might not be responsible for the observed metronidazole resistance in the present study since there was no significant increase in the G1 set. The upregulation of the two *nim* genes in GA might be a consequence of decrease in growth rate when compared to G1.

5.5.5.5. Conclusions

The study of transcriptional regulation under nitrogen limiting conditions is of significance in opportunistic pathogens to study the responses of the bacterium under nitrogen limitation in vivo (Parente et al., 2010). While studying the environmental regulation of glycosidase and peptidase production by *Streptococcus gordonii* involved in infective endocarditis, Mayo et al., (2000) observed that a shift from low oral pH to blood pH induced expression of several genes, including *msrA* (protection against oxidative stress) and enhanced bacterial growth. *S. gordonii* is a commensal organism of the oral cavity which becomes an opportunistic pathogen during colonisation of heart valves thereby contributing to infective endocarditis. The present work describes the upregulation of a group of oxidative-stress and virulence associated genes in nitrogen limiting conditions in *B. fragilis*. The study also emphasizes the requirement of ammonia for *B. fragilis* growth and presents possible mechanisms of ammonia generation from glutamine in an ammonia deficient medium. The differential regulation holds significance when the bacterium is accidentally released into the peritoneal cavity from its natural habitat, the GI tract. The changes in regulation at the transcriptomic level under nitrogen stress might aid the bacterium in its opportunistic pathogenic phase where it is exposed to oxygen and nitrogen-limiting conditions in the extraintestinal sites. Since the results of the RNA-Seq analysis discussed in this chapter are pioneering, they need to be validated with appropriate experiments for example, using reverse transcription polymerase chain

reaction (RT-PCR) on selected differentially expressed genes to provide additional information regarding nitrogen regulation by *B. fragilis*.

Chapter 6 General Discussion

The human gastrointestinal tract hosts the major proportion of the human microbiota, amounting to $\sim 10^{12}$ microbes per ml of luminal contents from 500-1000 diverse species (Savage, 1977; Xu and Gordon, 2003). The microorganisms that comprise the gut microbiome thrive in the gastrointestinal tract through specialised interactions with the intestinal epithelial mucosa as well as MSCRAMM-based interactions with the host extracellular matrix (Patti et al., 1994). *Bacteroides fragilis*, one among the commensals of the gastrointestinal tract, engage in various beneficial interactions with the host, encompassing host nutrition and immunomodulation activities (Troy and Kasper, 2010). However, it is also the most frequently isolated Gram-negative obligate anaerobe from peritoneal abscesses and systemic infections (Patrick and Duerden, 2006). Through the study presented here, we examined potential mechanisms involved in the transition of this commensal into an opportunistic pathogen. *B. fragilis* is known to interact with human fibrinogen, a protein which is involved in the formation of fibrin-containing abscess walls that curtail pathogens and other infectious agents. However, the ability to bind and degrade fibrinogen presents the bacterium with a possible opportunity to escape from abscesses and disseminate infection (Houston et al., 2010). Initially we focussed on identifying a specific protein involved in fibrinogen binding and on investigating its interaction with proteases. Reverse genetics revealed the involvement of the outer membrane protein, BF1705, in fibrinogen binding through the analysis of concentrated supernatant extracts of wild-type NCTC 9343 and the Δ BF1705 cultures. In addition to fibrinogen binding, a possible involvement of BF1705 in the previously established *B. fragilis* fibronectin binding activity was also confirmed. Fibronectin, an ECM component, recruits fibrinogen which colocalizes with the fibronectin at the site of injury, thereby functioning in wound healing (Pereira et al., 2002). However, a potential link between fibrinogen binding by 1705 and degradation was not detected in the present study. The genetic conservation of BF1705 suggests that it is also

possibly involved in survival within the gastrointestinal tract, where fibrinogen is not present.

Previous studies in *B. fragilis* indicated the induction of proteases to enhance nitrogen scavenging under conditions of nitrogen limitation in the growth environment (Gibson and Macfarlane, 1988a). Urease catalyses the hydrolysis of urea to carbamate and ammonia, thereby replenishing the nitrogen source necessary for bacterial growth. The Nac-dependent induction of urease enzyme has been identified in *Klebsiella pneumoniae* under nitrogen limiting conditions (Collins et al., 1993). Therefore we attempted to find a relation between nitrogen limitation and the regulation in the expression of binding proteins, including BF1705, and secreted proteases. Comparison of the transcriptomes derived from cultures grown in defined medium and glutamine-containing medium was performed. Culture mass doubling times were similar in defined medium (N) and glutamine medium (G1) till mid-log phase. A diauxic shift was apparent in the glutamine medium wherein the growth rate decreased when the culture reached an OD₆₀₀ of 0.8 (GA). Analysis of differential expression in the GA set with N set as the reference, demonstrated that the gene encoding the BF1705 binding protein was not significantly upregulated in the nitrogen limiting medium, which suggests its primary function as an adhesin rather than a role in nutrient binding. However, BF0586 (BF9343_0559), encoding a homologue of a previously identified collagen binding protein (Cbp1) and BF9343_4169 (BF4280) encoding a protein identical to a laminin-1 binding protein, were significantly upregulated by a ~2-fold at a q value < 0.05 (De et al., 2006; Galvão et al., 2014). The laminin-1 binding protein was identified to possess plasminogen activating function to plasmin, a serine protease known to degrade fibrin, collagen and other structural components of the matrix tissue (Ferreira et al., 2013; Sijbrandi et al., 2008). Moreover, among the four protease-encoding genes that were upregulated by 2-fold under nitrogen limiting conditions, three, namely BF9343_1980, BF9343_2676 and BF9343_2678 encoding putative homologues of aminopeptidase C, transglutaminase-like enzymes, cysteine proteases and dipeptidyl-peptidase VI, respectively, were predicted to be within the molecular size range of

45-55 kDa. The molecular sizes of these three proteases corresponded to the molecular size of the fibrinolytic zone of clearance (45-50 kDa) detected by fibrinogen zymography of wild-type *B. fragilis* protein extracts. However, the ability of *B. fragilis* to degrade fibrinogen had been detected in BHI-S and glucose-DM growth media but was not examined in the nitrogen limiting glutamine medium. The upregulation of the genes encoding these binding proteins and proteases suggest an increased proteolytic activity, including fibrinogen degradation, under nitrogen limiting conditions to assuage the cell's growing need for nitrogen.

While considering an overview of the number of differentially expressed genes, both in G1 and GA compared to the reference N at a 2-fold cut-off (q value < 0.05), ~40% of the genome was differentially expressed in GA whereas only 5.75% of the genome was differentially expressed in G1. The proportion of the genes differentially expressed in G1, albeit smaller, indicates that the bacterial growth was affected by nitrogen limitation in glutamine medium even at the time-point the G1 samples were extracted, despite similar doubling times of the cultures in N and G1 samples. The difference in the regulation of genes in G1 and GA sets when compared to the N reference was evident from the fact that 31% of the upregulated genes in G1 were not significantly upregulated in GA. Moreover, five genes that were upregulated in G1 were significantly downregulated in GA. A similar differential regulation was identified even in the proportion of downregulated genes in G1 wherein ~ 22% of the genes that were downregulated in G1 were not significantly regulated in GA and three were significantly upregulated in GA. The observed differential regulation was most likely attributed to the reduction in growth rate, observed after 0.4 OD₆₀₀ in glutamine medium, between the samples extracted for GA when compared to that from which the G1 samples were extracted. Therefore, the differential gene expression observed in GA might be an effect of both nitrogen limitation and growth rate reduction on the *B. fragilis* transcriptome. An example is the differential regulation of the glutamine synthetase encoding gene, *glnA*, between G1 and GA when compared to the N reference. Even though *glnA* was significantly upregulated by over 5-fold in G1, a > 2 -fold decrease in *glnA* expression was detected in GA.

However, previous studies using *E. coli* have confirmed the upregulation of *glnA* expression under nitrogen limiting conditions, owing to a reduction in intracellular glutamine concentration and activation of sigma 54 (Atkinson et al., 2002). Therefore, the downregulation observed in GA in the present study is most likely a consequence of the growth condition. A similar alteration in the number of differentially expressed genes was observed when the response of *Mycobacterium smegmatis* to nitrogen limitation was studied using microarray-based nitrogen run-out experiments in batch culture and RNA-Seq based study using continuous culture in a nitrogen-limited chemostat (Petridis et al., 2015; Williams et al., 2013). The former had identified a total of 1090 differentially expressed genes whereas the latter revealed significant changes in the expression of 357 genes. 52% of the 357 genes overlapped with the previous study using batch cultures. The study emphasized the difficulty in assigning transcriptional changes to a single stimulus in batch cultures due to the simultaneous changes in growth rate, nutrient depletion and end-product build up. However a comparative analysis of the individual gene expressions in G1 and GA with respect to the reference N helped us to focus on genes affected exclusively by nitrogen limitation in the present study.

The downregulation of genes involved in amino acid (aa) transport, including glycine betaine transport, both in G1 and GA were surprising since the bacterial cell would require an efficient transport of amino acids to adapt to the nitrogen limiting conditions and therefore should account for an upregulation of genes involved in aa transport. Petridis et al., (2015) detected the upregulation of eight GlnR-regulated amino acid transporters through the RNA-Seq analysis of the nitrogen limited *Mycobacterium smegmatis* transcriptome. Nevertheless, a previous study which had examined the interrelationships between diets, structure and operations of the human gut microbiota reported that total transcript levels corresponded to the protein intake (Faith et al., 2011). Microbial RNA-Seq on mice faecal samples, demonstrated that the number of *Bacteroides caccae* exhibited a positive correlation with casein (protein diet) with the transcripts remaining roughly constant per unit of casein. High levels of expression of genes involved in pathways using aa as substrates for nitrogen

or carbon energy sources were predicted for species that positively correlated with casein, which included the three *Bacteroides* spp. examined. Therefore, in the present study a reduction in the nitrogen content in the medium might correlate with the observed downregulation of *B. fragilis* genes involved in aa transport pathways.

Another possibility for the downregulation might be that expression of these genes are under the regulation of upstream non-coding small RNAs (sRNA), antisense RNAs or multiple transcription start sites (TSS) and 5' untranslated regions (UTR) of the coding mRNA. Contradictory to the previous conception that the prokaryotic transcriptome is not as complex as the eukaryotic transcriptome, bacteria have been shown to produce antisense RNA and previously unidentified non-coding RNAs (ncRNA) that regulate transcription and translation (Güell et al., 2011). Regulatory RNAs differ from protein-encoding regions in that they possess no distinct start and stop codons that denote their boundaries or occurrence in a genome. In enteric bacteria they are generally of 50-200 nucleotides in length and control a multitude of processes including stress responses, metabolic reactions and pathogenesis (Raghavan et al., 2011). Albeit in the present study we have determined differential gene expression only on the basis of reads that aligned against the genes (coding regions), it is essential to extend the analysis to detect fragments aligning to intergenic regions, owing to the rate at which the presence of ncRNAs are revolutionizing the field of bacterial gene expression studies. Despite a few studies suggesting that many fragments aligning to intergenic regions might define non-specific transcription initiation or leaky transcription termination of genes upstream, they might also account for the identification of novel ncRNAs (Livny et al., 2008; Raghavan et al., 2012). The presence and activity of ncRNAs has been analysed in many bacterial species. Thirteen previously unknown ncRNA species identified in the transcriptome of *Bacillus anthracis* were used to follow transcriptional changes during growth phases and sporulation. An RNA-Seq based study focussing on *S. enterica* serovar Typhi was used to correct annotation of genome sequence and also to identify transcriptionally active prophage genes and new members of the *ompR* regulon in addition to identifying four novel ncRNAs. Identification of a ncRNA

involved in the regulation of carbon metabolism was made possible through RNA-seq of *V. cholerae* transcriptome (van Vliet, 2010). sRNAs, especially antisense RNAs, work either in cis by targeting a gene within the same locus or in trans by targeting loci elsewhere in the genome. They regulate transcriptional interference, activation and control, and regulation of mRNA half life. The proportion of antisense RNA could be ~ 10-20% of the genes in bacteria (Güell et al., 2011). Sequencing of stranded libraries can help in locating trans-acting sRNAs opposite to coding genes. sRNAs can also have multiple functions through encoding functional peptides. RNAIII in *Staphylococcus aureus*, one of the first identified cis-acting RNAs is involved in quorum sensing (Novick and Geisinger, 2008). The SgrS sRNA causes translational repression when pre-annealed with the *ptsG* transcript in *E. coli*. The 5' region of SgrS contains an ORF termed SgrT which is translated during glucose-PO₄ stress. Cells expressing SgrT alone have a defective glucose uptake. Therefore this bifunctional sRNA encodes physiologically redundant but mechanistically distinct functions that contribute to the same stress response (Maki et al., 2010). Are similar actions of sRNAs involved in the observed downregulation of *B. fragilis* genes involved in amino acid transport under nitrogen limiting conditions?

In the present study, a > 2-fold increase in the expression of 17 and 13 consecutive genes beginning from BF9343_0698 to BF9343_0715 was observed in the G1 and GA samples, respectively. These genes are within the polysaccharide G (PS G or PS 7) locus of the 9343 genome. The detected upregulation suggest that these genes constitute an operon controlled by the same promoter. Previous studies analysing the NCTC 9343 genome revealed that eight PS loci (A-H) are related to micro-capsule (MC) expression, of which seven are reversibly switched on and off by the recombinase-mediated site-specific inversion of promoter sequences (Cerdeño-Tárraga et al., 2005). Although each PS locus encodes proteins that produce an antigenically distinct and phase-variable MC, they share a common genetic structure where two genes encoding putative transcriptional regulators, Up(a-h) Y homologue and Up(a-h) Z homologue, are located upstream of genes for polysaccharide biosynthesis and transport. UpxY and UpzY-mediated transcriptional regulation of

the phase-variable PS loci have been studied by Comstock, (2009). The cell population in a *B. fragilis* culture should exhibit random phase variation of the PS loci encoding each PS such that the expression of each locus is nearly uniform in the culture population. However, in the present study we detected an expression bias towards the genes of the PS7/PSG locus under nitrogen limiting condition. RNA-Seq data were obtained by pooling the reads from transcriptomes derived from biological replicates of cultures inoculated from six different colonies. Therefore, the probability of a 'founder effect' based on the inoculum in the observed expression bias seems unlikely. This raised the possibility that a capsule composed of polysaccharide G is providing an advantage to the cells under nitrogen limiting conditions.

Nevertheless, an increase in expression was observed only in a subset of genes belonging to the PSG locus which might suggest the presence of a suboperon within the putative operon. Mapping transcriptional initiation sites in *H. pylori* revealed 337 operons but internal TSSs added 192 additional cistrons (Sharma et al., 2010). In *M. pneumoniae* 341 reference operons could be divided into 447 smaller sub operons (Guell et al., 2009). Promoters could engage various transcription complexes that could produce different RNA outputs. By mapping TSSs to single base precision using RNA-Seq, it is possible to map all internal promoters within an operon. The analysis of the upstream promoter binding region and TSSs of these genes would provide insights into the possible presence of an operon (Güell et al., 2011). Mapping of TSSs define novel promoter regions and 5' UTRs upstream of coding genes. A number of TSSs have been reported with specific focus on genes involved in virulence and nitrogen metabolism in *Klebsiella pneumoniae*. A significant fraction of operons had multiple TSSs in both *E. coli* and *K. pneumoniae*, and encoded genes with essential functions including aa biosynthesis, central metabolism and transport (Kim et al., 2012). Also, multiple promoters upstream of a gene can be regulated by transcription factors in different ways. The *ure* operon in *K. pneumoniae* possesses two core promoters with distinct TSSs. One core promoter is Nac-dependent whereas the other is not. The 5' UTR can also contribute to the transcriptional regulation of a

downstream gene or operon by functioning as a transcription factor binding site. The ArgR transcription regulator is known to bind to the promoter region and the 5' UTR of the ArgC in *E. coli*. The conserved nature of these regions is evident from sequence analysis (Caldara et al., 2006; Charlier et al., 1992). OxyR regulator which autoregulates its transcription also exhibits similar 5' UTR binding activity (Toledano et al., 1994). ArgC, and ArgR and OxyR transcription regulators were observed to be significantly upregulated in the GA set in the present study. The upregulation might be facilitated through the 5'UTR binding of these transcription regulators under nitrogen limitation.

Another interesting observation from the transcriptomic analysis of *B. fragilis* under nitrogen limitation was the significant upregulation in the expression of eight genes encoding pseudogenes in the GA set. However, the expression of these genes were not significant in G1 and two separate genes encoding putative pseudogenes were downregulated in G1, suggesting that the upregulation of pseudogenes in GA might be a combined consequence of nitrogen limitation and growth reduction. Around 220 pseudogenes have been identified in *Salmonella enterica*. Even though the majority of them are inactivated in the serovar Typhi, they are intact in related host-promiscuous serovars such as typhimurium (Perkins et al., 2009). Certain pseudogenes are known to express functional truncated protein domains, such as the truncated cytotoxin encoded by a *Chlymadia trachomatis* pseudogene (Carlson et al., 2004). Genome degradation, as a consequence of pseudogene formation, is thought to restrict the host diversity of pathogens through a loss of pathways essential for survival in diverse hosts. However, a similar function is not likely in *B. fragilis*, which is not a host-specific pathogen. The pseudogenes in *B. fragilis* might correspond to incorrectly annotated genes or small functional polypeptides which might have a specific role under nitrogen limiting conditions. Alternatively, they might even be antisense RNAs regulating the expression of coding RNAs.

The microbiota and the residual pancreatic proteases present in the lower gastrointestinal tract, which is the natural habitat of the commensal *B. fragilis*, are

involved in the production of ammonia from semi-digested dietary and host-derived proteins (Dai et al., 2011). However, the production of ammonia progressively increases from the proximal to the distal part of the colon (Smith and Macfarlane, 1998). Therefore, it is likely that *B. fragilis* occupying certain niches in the proximal colon might face a nitrogen limitation. The accidental escape of *B. fragilis* from the gastrointestinal tract into the extraintestinal peritoneal cavity following intestinal injury would potentially leave the bacterium at a greater risk of being encountered by a shortage of ammonia. Although, no previous studies have reported the specific concentration of ammonia and other readily available nutrients in the peritoneal cavity, it is most likely that this site is limited in easily accessible sources of carbon and nitrogen especially when compared to the gastrointestinal tract where active digestion and absorption of diet-derived nutrients takes place. Moreover, extraintestinal sites represent an aerobic environment unlike the anaerobic gastrointestinal tract. Therefore, to persist in the peritoneal cavity following accidental release from the natural habitat, *B. fragilis* must regulate gene expression to adapt to carbon and nitrogen limitation, as well as oxygen exposure. Rocha and Smith, (1997) reported an increase in the expression of *B. fragilis katB* catalase gene when anaerobic cultures were exposed to oxygen. It was also shown that *katB* mRNA was strongly repressed by glucose. These findings suggest an increase in the expression of KatB as a means to adapt to the extraintestinal environment. An upregulation in *katB* transcription was also identified in the GA samples in the present study as a result of nitrogen and growth limitation. Therefore, it is possible that the transcriptional regulation of *B. fragilis* in the peritoneal cavity is similar to that analysed in the nitrogen limiting GA analysed in the present study. An increase in the expression of proteases and nutrient binding proteins observed in GA might therefore lead to the destruction of host tissues in the peritoneal cavity. Thus, in an attempt to scavenge for nitrogen in the extraintestinal environment, *B. fragilis* might promote tissue destruction and abscess formation. Prophylactic administration of metronidazole antibiotic (5-nitroimidazole) against anaerobic infections, especially against *Bacteroides* spp. that tend to be resistant against a wide range of

antimicrobial agents, have been reported in the past (Brazier et al., 1999). However, metronidazole-resistant *B. fragilis* strains are on the rise and have been associated with detrimental outcomes (Löfmark et al., 2005). The nitroimidazole resistance gene (nim)-based resistance is regarded as the primary mechanism of metronidazole resistance in *B. fragilis* (Trinh and Reysset, 1996). A 1.5-fold increase in two metronidazole resistance-encoding genes was also observed in GA which might facilitate persistence of *B. fragilis* in peritoneal infections. However, these assumptions are based on the potential similarities between transcription under the conditions encountered in GA and those reported following exposure to oxygen and carbon limitation. A conclusive study of the pathogenic determinants that promote the opportunistic phase of *B. fragilis* can only be performed by analysing the transcriptome of the bacterium isolated from soft tissue abscesses in conjunction with the infected host tissue transcriptome. Recent advancements in the field of RNA-Seq, such as Dual RNA-Seq which is capable of analysing changes in gene expression of both the pathogen and the host simultaneously would provide a timed monitoring of the infection process. However, the technique is cumbersome owing to optimising the RNA extraction, library construction and fixing the parameters for the effective analysis of both eukaryotic and prokaryotic transcriptomes simultaneously. Dual transcriptomics has been used recently to study *Candida albicans* and dendritic cells in mice (Westermann et al., 2012). The use of such techniques on *B. fragilis* clinical isolates and soft tissue abscesses would allow the quantification of transcriptomes of a small number of cells at the initial site of infection which would help us to understand changes during transition from commensal to pathogen and would also potentially allow development of novel methods for treatment of anaerobic infections.

Chapter 7 Bibliography

Abou-Zeid, C., Ratliff, T.L., Wiker, H.G., Harboe, M., Bennedsen, J., and Rook, G.A. (1988). Characterization of fibronectin-binding antigens released by *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG. *Infect. Immun.* *56*, 3046–3051.

Abrahams, G.L., and Abratt, V.R. (1998). The NADH-dependent glutamate dehydrogenase enzyme of *Bacteroides fragilis* Bf1 is induced by peptides in the growth medium. *Microbiology* *144* (Pt 6, 1659–1667.

Abratt, V.R., Zappe, H., and Woods, D.R. (1993). A reporter gene vector to investigate the regulation of glutamine synthetase in *Bacteroides fragilis* Bf1. *J. Gen. Microbiol.* *139*, 59–65.

Adibi, S.A., and Mercer, D.W. (1973). Protein Digestion in Human Intestine as Reflected in Luminal, Mucosal, and Plasma Amino Acid Concentrations after Meals. *J. Clin. Invest.* *52*, 1586–1594.

Adigüzel, A.C., Bitlisli, B.O., Yaşa, I., and Eriksen, N.T. (2009). Sequential secretion of collagenolytic, elastolytic, and keratinolytic proteases in peptide-limited cultures of two *Bacillus cereus* strains isolated from wool. *J. Appl. Microbiol.* *107*, 226–234.

Ahrné, S., Nobaek, S., Jeppsson, B., Adlerberth, I., Wold, A.E., and Molin, G. (1998). The normal *Lactobacillus* flora of healthy human rectal and oral mucosa. *J. Appl. Microbiol.* *85*, 88–94.

Alauzet, C., Mory, F., Teyssier, C., Hallage, H., Carlier, J.P., Grollier, G., and Lozniewski, A. (2009). Metronidazole Resistance in *Prevotella* spp. and Description of a New *nim* Gene in *Prevotella baroniae*. *Antimicrob. Agents Chemother.* *54*, 60–64.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *The Extracellular Matrix of Animals In Molecular Biology of the Cell* (New York: Garland Science).

Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. *N. Biotechnol.* *25*, 195–203.

Antalis, T.M., Shea-Donohue, T., Vogel, S.N., Sears, C., and Fasano, A. (2007). Mechanisms of disease: protease functions in intestinal mucosal pathobiology. *Nat. Clin. Pract. Gastroenterol. Hepatol.* *4*, 393–402.

Anthony, R.M., Rutitzky, L.I., Urban, J.F., Stadecker, M.J., and Gause, W.C. (2007). Protective immune mechanisms in helminth infection. *Nat. Rev. Immunol.* 7, 975–987.

Are, A., Aronsson, L., Wang, S., Greicius, G., Lee, Y.K., Gustafsson, J.-A., Pettersson, S., and Arulampalam, V. (2008). *Enterococcus faecalis* from newborn babies regulate endogenous PPAR γ activity and IL-10 levels in colonic epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* 105, 1943–1948.

Atherly, T., and Ziemer, C.J. (2014). *Bacteroides* isolated from four mammalian hosts lack host-specific 16S rRNA gene phylogeny and carbon and nitrogen utilization patterns. *Microbiologyopen* 3, 225–238.

Atkinson, M.R., Blauwkamp, T. a, Bondarenko, V., Studitsky, V., and Ninfa, A.J. (2002). Undergoes the Transition from Nitrogen Excess Growth to Nitrogen Starvation. *Society* 184, 5358–5363.

Aziz, R.K. (2009). A hundred-year-old insight into the gut microbiome! *Gut Pathog.* 1, 21.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915–1920.

Bakhtiar, S.M., LeBlanc, J.G., Salvucci, E., Ali, A., Martin, R., Langella, P., Chatel, J.-M., Miyoshi, A., Bermúdez-Humarán, L.G., and Azevedo, V. (2013). Implications of the human microbiome in inflammatory bowel diseases. *FEMS Microbiol. Lett.* 342, 10–17.

Barkocy-Gallagher, G. a., Foley, J.W., and Lantz, M.S. (1999). Activities of the *Porphyromonas gingivalis* PrtP proteinase determined by construction of prtP-deficient mutants and expression of the gene in *Bacteroides* species. *J. Bacteriol.* 181, 246–255.

Barrett, A.J., Tolle, D.P., and Rawlings, N.D. (2003). Managing peptidases in the genomic era. *Biol. Chem.* 384, 873–882.

Baughn, A.D., and Malamy, M.H. (2002). A mitochondrial-like aconitase in the bacterium *Bacteroides fragilis*: implications for the evolution of the mitochondrial Krebs cycle. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4662–4667.

Bauman, S.J., and Kuehn, M.J. (2006). Purification of outer membrane vesicles from *Pseudomonas aeruginosa* and their activation of an IL-8 response. *Microbes Infect.* 8, 2400–2408.

- Begg, K.J., and Donachie, W.D. (1978). Changes in cell size and shape in thymine-requiring *Escherichia coli* associated with growth in low concentrations of thymine. *J. Bacteriol.* *133*, 452–458.
- Biller, S.J., Schubotz, F., Roggensack, S.E., Thompson, A.W., Summons, R.E., and Chisholm, S.W. (2014). Bacterial vesicles in marine ecosystems. *Science* *343*, 183–186.
- Blairon, L., De Gheldre, Y., Delaere, B., Sonet, A., Bosly, A., and Glupczynski, Y. (2006). A 62-month retrospective epidemiological survey of anaerobic bacteraemia in a university hospital. *Clin. Microbiol. Infect.* *12*, 527–532.
- Bomberger, J.M., Maceachran, D.P., Coutermarsh, B.A., Ye, S., O'Toole, G.A., and Stanton, B.A. (2009). Long-distance delivery of bacterial virulence factors by *Pseudomonas aeruginosa* outer membrane vesicles. *PLoS Pathog.* *5*, e1000382.
- Brazier, J.S., Stubbs, S.L.J., and Duerden, B.I. (1999). Metronidazole resistance among clinical isolates belonging to the *Bacteroides fragilis* group: time to be concerned? *J. Antimicrob. Chemother.* *44*, 580–581.
- Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* *13*, 278–284.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., and Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* *185*, 6220–6223.
- Bren, A., Hart, Y., Dekel, E., Koster, D., and Alon, U. (2013). The last generation of bacterial growth in limiting nutrient. *BMC Syst. Biol.* *7*, 27.
- Brigham, C.J., and Malamy, M.H. (2005). Characterization of the RokA and HexA broad-substrate-specificity hexokinases from *Bacteroides fragilis* and their role in hexose and N-acetylglucosamine utilization. *J. Bacteriol.* *187*, 890–901.
- Bröer, S. (2008). Apical transporters for neutral amino acids: physiology and pathophysiology. *Physiology (Bethesda)*. *23*, 95–103.
- Brook, I. (1990). The clinical importance of all members of the *Bacteroides fragilis* group. *J. Antimicrob. Chemother.* *25*, 473–474.
- Brown, D.R., Barton, G., Pan, Z., Buck, M., and Wigneshweraraj, S. (2014). Nitrogen stress response and stringent response are coupled in *Escherichia coli*. *Nat. Commun.* *5*, 4115.

Caldara, M., Charlier, D., and Cunin, R. (2006). The arginine regulon of *Escherichia coli*: whole-system transcriptome analysis discovers new genes and provides an integrated view of arginine regulation. *Microbiology* *152*, 3343–3354.

Carlier, J.P., Sellier, N., Rager, M.N., and Reysset, G. (1997). Metabolism of a 5-nitroimidazole in susceptible and resistant isogenic strains of *Bacteroides fragilis*. *Antimicrob. Agents Chemother.* *41*, 1495–1499.

Carlson, J.H., Hughes, S., Hogan, D., Cieplak, G., Sturdevant, D.E., McClarty, G., Caldwell, H.D., and Belland, R.J. (2004). Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. *Infect. Immun.* *72*, 7063–7072.

Carroll, I.M., and Maharshak, N. (2013). Enteric bacterial proteases in inflammatory bowel disease- pathophysiology and clinical implications. *World J. Gastroenterol.* *19*, 7531–7543.

Carver, T., Berriman, M., Tivey, A., Patel, C., Bohme, U., Barrell, B.G., Parkhill, J., and Rajandream, M.-A. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* *24*, 2672–2676.

Cerdeño-Tárraga, A.M., Patrick, S., Crossman, L.C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M. a, et al. (2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* *307*, 1463–1465.

Chacko, A., and Cummings, J.H. (1988). Nitrogen losses from the human small bowel: obligatory losses and the effect of physical form of food. *Gut* *29*, 809–815.

Charlier, D., Roovers, M., Van Vliet, F., Boyen, A., Cunin, R., Nakamura, Y., Glansdorff, N., and Piérard, A. (1992). Arginine regulon of *Escherichia coli* K-12. A study of repressor-operator interactions and of in vitro binding affinities versus in vivo repression. *J. Mol. Biol.* *226*, 367–386.

Chassaing, B., and Darfeuille-Michaud, A. (2013). The σ E pathway is involved in biofilm formation by Crohn's disease-associated adherent-invasive *Escherichia coli*. *J. Bacteriol.* *195*, 76–84.

Chen, Z., and O'Shea, J.J. (2008). Th17 cells: a new fate for differentiating helper T cells. *Immunol. Res.* *41*, 87–102.

Chen, Y., Kinouchi, T., Kataoka, K., Akimoto, S., and Ohnishi, Y. (1995). Purification and characterization of a fibrinogen-degrading protease in *Bacteroides fragilis* strain YCH46. *Microbiol. Immunol.* *39*, 967–977.

- Cheng, C., and Shuman, S. (2000). Recombinogenic Flap Ligation Pathway for Intrinsic Repair of Topoisomerase IB-Induced Double-Strand Breaks. *Mol. Cell. Biol.* *20*, 8059–8068.
- Chi, B., Qi, M., and Kuramitsu, H.K. (2003). Role of dentilisin in *Treponema denticola* epithelial cell layer penetration. *Res. Microbiol.* *154*, 637–643.
- Ciborowski, P., Nishikata, M., Allen, R.D., and Lantz, M.S. (1994). Purification and characterization of two forms of a high-molecular-weight cysteine proteinase (porphypain) from *Porphyromonas gingivalis*. *J. Bacteriol.* *176*, 4549–4557.
- Cohen-Poradosu, R., McLoughlin, R.M., Lee, J.C., and Kasper, D.L. (2011). *Bacteroides fragilis* - Stimulated interleukin-10 contains expanding disease. *J. Infect. Dis.* *204*, 363–371.
- Collins, C.M., Gutman, D.M., and Laman, H. (1993). Identification of a nitrogen-regulated promoter controlling expression of *Klebsiella pneumoniae* urease genes. *Mol. Microbiol.* *8*, 187–198.
- Comstock, L.E. (2009). Importance of glycans to the host-bacteroides mutualism in the mammalian intestine. *Cell Host Microbe* *5*, 522–526.
- Coyne, M.J., Tzianabos, A.O., Mallory, B.C., Carey, V.J., Kasper, D.L., and Comstock, L.E. (2001). Polysaccharide Biosynthesis Locus Required for Virulence of *Bacteroides fragilis*. *Infect. Immun.* *69*, 4342–4350.
- Coyne, M.J., Weinacht, K.G., Krinos, C.M., and Comstock, L.E. (2003). Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 10446–10451.
- Coyne, M.J., Reinap, B., Lee, M.M., and Comstock, L.E. (2005). Human symbionts use a host-like pathway for surface fucosylation. *Science* *307*, 1778–1781.
- Coyne, M.J., Chatzidaki-Livanis, M., Paoletti, L.C., and Comstock, L.E. (2008). Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont *Bacteroides fragilis*. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 13099–13104.
- Cummings, J.H., Wiggins, H.S., Jenkins, D.J., Houston, H., Jivraj, T., Drasar, B.S., and Hill, M.J. (1978). Influence of diets high and low in animal fat on bowel habit, gastrointestinal transit time, fecal microflora, bile acid, and fat excretion. *J. Clin. Invest.* *61*, 953–963.
- Cunin, R., Glansdorff, N., Piérard, A., and Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. *Microbiol. Rev.* *50*, 314–352.

D'Elia, J.N., and Salyers, A.A. (1996). Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* *178*, 7180–7186.

Dai, Z.-L., Wu, G., and Zhu, W.-Y. (2011). Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. *Front. Biosci. (Landmark Ed.)* *16*, 1768–1786.

Davila, A.-M., Blachier, F., Gotteland, M., Andriamihaja, M., Benetti, P.-H., Sanz, Y., and Tomé, D. (2013). Intestinal luminal nitrogen metabolism: role of the gut microbiota and consequences for the host. *Pharmacol. Res.* *68*, 95–107.

De, E., Araújo Lobo, L., Barreiros Petrópolis, D., Dos, K.E., Ferreira, M.C., e Silva Filho, F.C., and Domingues, R.M.C.P. (2006). A *Bacteroides fragilis* surface glycoprotein mediates the interaction between the bacterium and the extracellular matrix component laminin-1. *Res. Microbiol.* *157*, 960–966.

Dethlefsen, L., McFall-Ngai, M., and Relman, D.A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* *449*, 811–818.

Dinkla, K., Rohde, M., Jansen, W.T.M., Carapetis, J.R., Chhatwal, G.S., and Talay, S.R. (2003). *Streptococcus pyogenes* recruits collagen via surface-bound fibronectin: A novel colonization and immune evasion mechanism. *Mol. Microbiol.* *47*, 861–869.

Dominiecki, M.E., and Weiss, J. (1999). Antibacterial action of extracellular mammalian group IIA phospholipase A2 against grossly clumped *Staphylococcus aureus*. *Infect. Immun.* *67*, 2299–2305.

Donskey, C.J. (2004). The role of the intestinal tract as a reservoir and source for transmission of nosocomial pathogens. *Clin. Infect. Dis.* *39*, 219–226.

Duée, P.H., Darcy-Vrillon, B., Blachier, F., and Morel, M.T. (1995). Fuel selection in intestinal cells. *Proc. Nutr. Soc.* *54*, 83–94.

Duncan, L., Yoshioka, M., Chandad, F., and Grenier, D. (2004). Loss of lipopolysaccharide receptor CD14 from the surface of human macrophage-like cells mediated by *Porphyromonas gingivalis* outer membrane vesicles. *Microb. Pathog.* *36*, 319–325.

Dziewanowska, K., Patti, J.M., Deobald, C.F., Bayles, K.W., Trumble, W.R., and Bohach, G.A. (1999). Fibronectin binding protein and host cell tyrosine kinase are required for internalization of *Staphylococcus aureus* by epithelial cells. *Infect. Immun.* *67*, 4673–4678.

- Dziewanowska, K., Carson, a. R., Patti, J.M., Deobald, C.F., Bayles, K.W., and Bohach, G. a. (2000). Staphylococcal fibronectin binding protein interacts with heat shock protein 60 and integrins: Role in internalization by epithelial cells. *Infect. Immun.* *68*, 6321–6328.
- Eckburg, P.B., Lepp, P.W., and Relman, D.A. (2003). Archaea and Their Potential Role in Human Disease. *Infect. Immun.* *71*, 591–596.
- Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the human intestinal microbial flora. *Science* *308*, 1635–1638.
- Eklou-Lawson, M., Bernard, F., Neveux, N., Chaumontet, C., Bos, C., Davila-Gay, A.-M., Tomé, D., Cynober, L., and Blachier, F. (2009). Colonic luminal ammonia and portal blood l-glutamine and l-arginine concentrations: a possible link between colon mucosa and liver ureagenesis. *Amino Acids* *37*, 751–760.
- Elhenawy, W., Debelyy, M.O., and Feldman, M.F. (2014). Preferential packing of acidic glycosidases and proteases into *Bacteroides* outer membrane vesicles. *MBio* *5*, 1–12.
- Esgleas, M., Lacouture, S., and Gottschalk, M. (2005). *Streptococcus suis* serotype 2 binding to extracellular matrix proteins. *FEMS Microbiol. Lett.* *244*, 33–40.
- Faith, J.J., McNulty, N.P., Rey, F.E., and Gordon, J.I. (2011). Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science* *333*, 101–104.
- Farthmann, E.H., and Schöffel, U. (1998). Epidemiology and pathophysiology of intraabdominal infections (IAI). *Infection* *26*, 329–334.
- Ferreira, E. de O., de Carvalho, J.B., Peixoto, R.J.M., Lobo, L.A., Zingalli, R.B., Smith, C.J., Rocha, E.R., and Domingues, R.M.C.P. (2009). The interaction of *Bacteroides fragilis* with components of the human fibrinolytic system. *FEMS Immunol. Med. Microbiol.* *56*, 48–55.
- Ferreira, E. de O., Teixeira, F.L., Cordeiro, F., Araujo Lobo, L., Rocha, E.R., Smith, J.C., and Domingues, R.M.C.P. (2013). The Bfp60 surface adhesin is an extracellular matrix and plasminogen protein interacting in *Bacteroides fragilis*. *Int. J. Med. Microbiol.* *303*, 492–497.
- Ferreira, E.D.O., Yates, E.A., Goldner, M., Vommaro, R.C., Silva Filho, F.C., Petrópolis, D.B., and Pilotto Domingues, R.M.C. (2008). The redox potential interferes with the expression of laminin binding molecules in *Bacteroides fragilis*. *Mem. Inst. Oswaldo Cruz* *103*, 683–689.

Ferreira, E.O., Falcão, L.S., Vallim, D.C., Santos, F.J., Andrade, J.R.C., Andrade, a. F.B., Vommaro, R.C., Ferreira, M.C.S., and Domingues, R.M.C.P. (2002). *Bacteroides fragilis* adherence to Caco-2 cells. *Anaerobe* 8, 307–314.

Fletcher, C.M., Coyne, M.J., and Comstock, L.E. (2011). Theoretical and experimental characterization of the scope of protein O-glycosylation in *Bacteroides fragilis*. *J. Biol. Chem.* 286, 3219–3226.

Flick, M.J., Du, X., Prasad, J.M., Raghu, H., Palumbo, J.S., Smeds, E., Hook, M., and Degen, J.L. (2013). Genetic elimination of the binding motif on fibrinogen for the *S. aureus* virulence factor ClfA improves host survival in septicemia. *Blood* 121, 1783–1794.

Flock, J.I., Fröman, G., Jönsson, K., Guss, B., Signäs, C., Nilsson, B., Raucchi, G., Höök, M., Wadström, T., and Lindberg, M. (1987). Cloning and expression of the gene for a fibronectin-binding protein from *Staphylococcus aureus*. *EMBO J.* 6, 2351–2357.

Franco, A. V, Liu, D., and Reeves, P.R. (1998). The wzz (cld) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. *J. Bacteriol.* 180, 2670–2675.

Furuse, K., Osawa, S., Kawashiro, J., Tanaka, R., Ozawa, A., Sawamura, S., Yanagawa, Y., Nagao, T., and Watanabe, I. (1983). Bacteriophage distribution in human faeces: continuous survey of healthy subjects and patients with internal and leukaemic diseases. *J. Gen. Virol.* 64 (Pt 9), 2039–2043.

Gal, M., and Brazier, J.S. (2004). Metronidazole resistance in *Bacteroides* spp. carrying nim genes and the selection of slow-growing metronidazole-resistant mutants. *J. Antimicrob. Chemother.* 54, 109–116.

Galvão, B.P.G. V., Weber, B.W., Rafudeen, M.S., Ferreira, E.O., Patrick, S., and Abratt, V.R. (2014). Identification of a Collagen Type I Adhesin of *Bacteroides fragilis*. *PLoS One* 9, e91141.

Ganesh, V.K., Rivera, J.J., Smeds, E., Ko, Y.P., Bowden, M.G., Wann, E.R., Gurusiddappa, S., Fitzgerald, J.R., and Höök, M. (2008). A structural model of the *Staphylococcus aureus* ClfA-fibrinogen interaction opens new avenues for the design of anti-staphylococcal therapeutics. *PLoS Pathog.* 4.

Giannoukos, G., Ciulla, D.M., Huang, K., Haas, B.J., IZard, J., Levin, J.Z., Livny, J., Earl, A.M., Gevers, D., Ward, D. V, et al. (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol.* 13, R23.

- Gibson, S. a, and Macfarlane, G.T. (1988a). Studies on the proteolytic activity of *Bacteroides fragilis*. *J. Gen. Microbiol.* *134*, 19–27.
- Gibson, S. a, and Macfarlane, G.T. (1988b). Characterization of proteases formed by *Bacteroides fragilis*. *J. Gen. Microbiol.* *134*, 2231–2240.
- Gibson, F.C., Onderdonk, a B., Kasper, D.L., and Tzianabos, a O. (1998). Cellular mechanism of intraabdominal abscess formation by *Bacteroides fragilis*. *J. Immunol.* *160*, 5000–5006.
- Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* *312*, 1355–1359.
- Godoy, V.G., Dallas, M.M., Russo, T. a., and Malamy, M.H. (1993). A role for *Bacteroides fragilis* neuraminidase in bacterial growth in two model systems. *Infect. Immun.* *61*, 4415–4426.
- Goldner, M., Coquis-Rondon, M., and Carlier, J.P. (1993). Effect of growth of *Bacteroides fragilis* at different redox levels on potential pathogenicity in a HeLa cell system: demonstration by confocal laser scanning microscopy. *Zentralbl. Bakteriol.* *278*, 529–540.
- Goux, W.J., Strong, A.A.D., Schneider, B.L., Lee, W.-N.P.L., and Reitzer, and L.J. (1996). Utilization of Aspartate as a Nitrogen Source in *Escherichia coli*. Analysis of nitrogen flow and characterization of the products of aspartate catabolism. *J. Biol. Chem.* *270*.
- Grangette, C., Müller-Alouf, H., Geoffroy, M., Goudercourt, D., Turneer, M., and Mercenier, A. (2002). Protection against tetanus toxin after intragastric administration of two recombinant lactic acid bacteria: impact of strain viability and in vivo persistence. *Vaccine* *20*, 3304–3309.
- Guan, J.L. (1990). Retroviral expression of alternatively spliced forms of rat fibronectin. *J. Cell Biol.* *110*, 833–847.
- Guell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kuhner, S., et al. (2009). Transcriptome Complexity in a Genome-Reduced Bacterium. *Science* (80-.). *326*, 1268–1271.
- Güell, M., Yus, E., Lluch-Senar, M., and Serrano, L. (2011). Bacterial transcriptomics: what is beyond the RNA horizo-me? *Nat. Rev. Microbiol.* *9*, 658–669.

- Gunasekaran, K., Tsai, C.-J., Kumar, S., Zanuy, D., and Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. *Trends Biochem. Sci.* 28, 81–85.
- Guzmán, C. a, Platé, M., and Pruzzo, C. (1990). Role of neuraminidase-dependent adherence in *Bacteroides fragilis* attachment to human epithelial cells. *FEMS Microbiol. Lett.* 59, 187–192.
- Hamer, H.M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F.J., and Brummer, R.-J. (2008). Review article: the role of butyrate on colonic function. *Aliment. Pharmacol. Ther.* 27, 104–119.
- Hattori, M., and Taylor, T.D. (2009). The human intestinal microbiome: a new frontier of human biology. *DNA Res.* 16, 1–12.
- Haurat, M.F., Aduse-Opoku, J., Rangarajan, M., Dorobantu, L., Gray, M.R., Curtis, M.A., and Feldman, M.F. (2011). Selective sorting of cargo proteins into bacterial membrane vesicles. *J. Biol. Chem.* 286, 1269–1276.
- Hay, E.D. (1991). *Cell Biology of Extracellular Matrix* (Springer Science & Business Media).
- Henry, L.G., McKenzie, R.M.E., Robles, A., and Fletcher, H.M. (2012). Oxidative stress resistance in *Porphyromonas gingivalis*. *Future Microbiol.* 7, 497–512.
- Herrick, S., Blanc-Brude, O., Gray, A., and Laurent, G. (1999). Fibrinogen. *Int. J. Biochem. Cell Biol.* 31, 741–746.
- Hill, R.T., Parker, J.R., Goodman, H.J.K., Jones, D.T., and Woods, D.R. (1989). Molecular Analysis of a Novel Glutamine Synthetase of the Anaerobe *Bacteroides fragilis*. *Microbiology* 135, 3271–3279.
- Hofstad, T. (1992). Virulence factors in anaerobic bacteria. *Eur. J. Clin. Microbiol. Infect. Dis.* 11, 1044–1048.
- Hooper, L. V, and Gordon, J.I. (2001). Commensal host-bacterial relationships in the gut. *Science* 292, 1115–1118.
- Hooper, L. V, Stappenbeck, T.S., Hong, C. V, and Gordon, J.I. (2003). Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat. Immunol.* 4, 269–273.

- Horstman, a L., and Kuehn, M.J. (2000). Enterotoxigenic *Escherichia coli* secretes active heat-labile enterotoxin via outer membrane vesicles. *J. Biol. Chem.* *275*, 12489–12496.
- Horstman, A.L., and Kuehn, M.J. (2002). Bacterial surface association of heat-labile enterotoxin through lipopolysaccharide after secretion via the general secretory pathway. *J. Biol. Chem.* *277*, 32538–32545.
- Houston, S., Blakely, G.W., McDowell, A., Martin, L., and Patrick, S. (2010). Binding and degradation of fibrinogen by *Bacteroides fragilis* and characterization of a 54 kDa fibrinogen-binding protein. *Microbiology* *156*, 2516–2526.
- Hu, P., Leighton, T., Ishkhanova, G., and Kustu, S. (1999). Sensing of nitrogen limitation by *Bacillus subtilis*: comparison to enteric bacteria. *J. Bacteriol.* *181*, 5042–5050.
- Huang, J.Y., Lee, S.M., and Mazmanian, S.K. (2011). The human commensal *Bacteroides fragilis* binds intestinal mucin. *Anaerobe* *17*, 137–141.
- Hullah, W.A., and Blackburn, T.H. (1971). Uptake and incorporation of amino acids and peptides by *Bacteroides amylophilus*. *Appl. Microbiol.* *21*, 187–191.
- Husain, F., Veeranagouda, Y., Hsi, J., Meggersee, R., Abratt, V., and Wexler, H.M. (2013). Two multidrug-resistant clinical isolates of *Bacteroides fragilis* carry a novel metronidazole resistance *nimJ* gene (*nimJ*). *Antimicrob. Agents Chemother.* *57*, 3767–3774.
- Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* *486*, 207–214.
- Ingham, H.R., Sisson, P.R., Middleton, R.L., Narang, H.K., Codd, A.A., and Selkon, J.B. (1981). Phagocytosis and killing of bacteria in aerobic and anaerobic conditions. *J. Med. Microbiol.* *14*, 391–399.
- Iozzo, R. V (1998). Matrix proteoglycans: from molecular design to cellular function. *Annu. Rev. Biochem.* *67*, 609–652.
- Jirmanova, L., Jankovic, D., Fornace, A.J., and Ashwell, J.D. (2007). Gadd45 α regulates p38-dependent dendritic cell cytokine production and Th1 differentiation. *J. Immunol.* *178*, 4153–4158.

Joh, D., Wann, E.R., Kreikemeyer, B., Speziale, P., and Höök, M. (1999). Role of fibronectin-binding MSCRAMMs in bacterial adherence and entry into mammalian cells. *Matrix Biol.* *18*, 211–223.

Jomaa, H. (1999). Inhibitors of the Nonmevalonate Pathway of Isoprenoid Biosynthesis as Antimalarial Drugs. *Science* (80-.). *285*, 1573–1576.

Jones, C.L., Napier, B. a., Sampson, T.R., Llewellyn, a. C., Schroeder, M.R., and Weiss, D.S. (2012). Subversion of Host Recognition and Defense Systems by *Francisella* spp. *Microbiol. Mol. Biol. Rev.* *76*, 383–404.

Kadurugamuwa, J.L., and Beveridge, T.J. (1995). Virulence factors are released from *Pseudomonas aeruginosa* in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion. *J. Bacteriol.* *177*, 3998–4008.

Kagawa, T.F., Cooney, J.C., Baker, H.M., McSweeney, S., Liu, M., Gubba, S., Musser, J.M., and Baker, E.N. (2000). Crystal structure of the zymogen form of the group A *Streptococcus* virulence factor SpeB: an integrin-binding cysteine protease. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 2235–2240.

Kahnt, J., Aguiluz, K., Koch, J., Treuner-Lange, A., Konovalova, A., Huntley, S., Hoppert, M., Søggaard-Andersen, L., and Hedderich, R. (2010). Profiling the outer membrane proteome during growth and development of the social bacterium *Myxococcus xanthus* by selective biotinylation and analyses of outer membrane vesicles. *J. Proteome Res.* *9*, 5197–5208.

Kanamori, K., Weiss, R.L., and Roberts, J.D. (1987). Role of glutamate dehydrogenase in ammonia assimilation in nitrogen-fixing *Bacillus macerans*. *J. Bacteriol.* *169*, 4692–4695.

Kapral, F. a (1966). Clumping of *Staphylococcus aureus* in the Peritoneal Cavity Clumping of *Staphylococcus aureus* in the Peritoneal Cavity of Mice. *J. Bacteriol.* *92*, 1188–1195.

Kasper, D.L. (1976). The Polysaccharide Capsule of *Bacteroides fragilis* Subspecies *fragilis*: Immunochemical and Morphologic Definition. *J. Infect. Dis.* *133*, 79–87.

Kato, N., Kato, H., Watanabe, K., and Ueno, K. (1996). Association of enterotoxigenic *Bacteroides fragilis* with bacteremia. *Clin. Infect. Dis.* *23 Suppl 1*, S83–S86.

Kelly, D., Campbell, J.I., King, T.P., Grant, G., Jansson, E.A., Coutts, A.G.P., Pettersson, S., and Conway, S. (2004). Commensal anaerobic gut bacteria attenuate

inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA. *Nat. Immunol.* 5, 104–112.

Kidd, P. (2003). Th1/Th2 balance: the hypothesis, its limitations, and implications for health and disease. *Altern. Med. Rev.* 8, 223–246.

Kim, D., Hong, J.S.-J., Qiu, Y., Nagarajan, H., Seo, J.-H., Cho, B.-K., Tsai, S.-F., and Palsson, B.Ø. (2012). Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. *PLoS Genet.* 8, e1002867.

Ko, Y.-P., Liang, X., Smith, C.W., Degen, J.L., and Höök, M. (2011). Binding of Efb from *Staphylococcus aureus* to fibrinogen blocks neutrophil adherence. *J. Biol. Chem.* 286, 9865–9874.

Kobe, B., and Kajava, A. V (2001). The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* 11, 725–732.

Kohn, H., and Widger, W. (2005). The molecular basis for the mode of action of bicyclomycin. *Curr. Drug Targets. Infect. Disord.* 5, 273–295.

Komori, Y., Nonogaki, T., and Nikai, T. (2001). Hemorrhagic activity and muscle damaging effect of *Pseudomonas aeruginosa* metalloproteinase (elastase). *Toxicon* 39, 1327–1332.

Kondo, K., Takade, A., and Amako, K. (1993). Release of the outer membrane vesicles from *Vibrio cholerae* and *Vibrio parahaemolyticus*. *Microbiol. Immunol.* 37, 149–152.

Krinos, C.M., Coyne, M.J., Weinacht, K.G., Tzianabos, A.O., Kasper, D.L., and Comstock, L.E. (2001). Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* 414, 555–558.

Kuehn, M.J., and Kesty, N.C. (2005). Bacterial outer membrane vesicles and the host-pathogen interaction. *Genes Dev.* 19, 2645–2655.

Kulkarni, H.M., and Jagannadham, M. V (2014). Biogenesis and multifaceted roles of outer membrane vesicles from Gram-negative bacteria. *Microbiology* 1–30.

Kulp, A., and Kuehn, M.J. (2010). Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu. Rev. Microbiol.* 64, 163–184.

Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., Kuhara, S., Hattori, M., Hayashi, T., and Ohnishi, Y. (2004). Genomic analysis of

Bacteroides fragilis reveals extensive DNA inversions regulating cell surface adaptation. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 14919–14924.

Lantz, M.S., Allen, R.D., Vail, T. a., Switalski, L.M., and Hook, M. (1991a). Specific cell components of *Bacteroides gingivalis* mediate binding and degradation of human fibrinogen. *J. Bacteriol.* *173*, 495–504.

Lantz, M.S., Allen, R.D., Duck, L.W., Blume, J.L., Switalski, L.M., and Hook, M. (1991b). Identification of *Porphyromonas gingivalis* components that mediate its interactions with fibronectin. *J. Bacteriol.* *173*, 4263–4270.

Lee, E.-Y., Choi, D.-Y., Kim, D.-K., Kim, J.-W., Park, J.O., Kim, S., Kim, S.-H., Desiderio, D.M., Kim, Y.-K., Kim, K.-P., et al. (2009). Gram-positive bacteria produce membrane vesicles: proteomics-based characterization of *Staphylococcus aureus*-derived membrane vesicles. *Proteomics* *9*, 5425–5436.

Lev, M. (1980). Glutamine-stimulated amino acid and peptide incorporation in *Bacteroides melaninogenicus*. *J. Bacteriol.* *143*, 753–760.

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 11070–11075.

Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* *124*, 837–848.

Liang, A., and Houghton, R.L. (1981). Coregulation of oxidized nicotinamide adenine dinucleotide (phosphate) transhydrogenase and glutamate dehydrogenase activities in enteric bacteria during nitrogen limitation. *J. Bacteriol.* *146*, 997–1002.

Liu, C.H., Lee, S.M., Vanlare, J.M., Kasper, D.L., and Mazmanian, S.K. (2008). Regulation of surface architecture by symbiotic bacteria mediates host colonization. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 3951–3956.

Livny, J., Teonadi, H., Livny, M., and Waldor, M.K. (2008). High-Throughput, Kingdom-Wide Prediction and Annotation of Bacterial Non-Coding RNAs. *PLoS One* *3*, e3197.

Lobo, L. a., Jenkins, A.L., Jeffrey Smith, C., and Rocha, E.R. (2013). Expression of *Bacteroides fragilis* hemolysins in vivo and role of HlyBA in an intra-abdominal infection model. *Microbiologyopen* *2*, 326–337.

- Löfmark, S., Fang, H., Hedberg, M., and Edlund, C. (2005). Inducible metronidazole resistance and nim genes in clinical *Bacteroides fragilis* group isolates. *Antimicrob. Agents Chemother.* *49*, 1253–1256.
- Lopez, J.M., Dromerick, A., and Freese, E. (1981). Response of guanosine 5'-triphosphate concentration to nutritional changes and its significance for *Bacillus subtilis* sporulation. *J. Bacteriol.* *146*, 605–613.
- Loughman, A., Fitzgerald, J.R., Brennan, M.P., Higgins, J., Downer, R., Cox, D., and Foster, T.J. (2005). Roles for fibrinogen, immunoglobulin and complement in platelet activation promoted by *Staphylococcus aureus* clumping factor A. *Mol. Microbiol.* *57*, 804–818.
- Łuczak, M., Obuch-Woszczatyński, P., Pituch, H., Leszczyński, P., Martirosian, G., Patrick, S., Poxton, I., Wintermans, R.G., Dubreuil, L., and Meisel-Mikołajczyk, F. (2001). Search for enterotoxin gene in *Bacteroides fragilis* strains isolated from clinical specimens in Poland, Great Britain, The Netherlands and France. *Med. Sci. Monit.* *7*, 222–225.
- Lutton, D.A., Patrick, S., Crockard, A.D., Stewart, L.D., Larkin, M.J., Dermott, E., and McNeill, T.A. (1991). Flow cytometric analysis of within-strain variation in polysaccharide expression by *Bacteroides fragilis* by use of murine monoclonal antibodies. *J. Med. Microbiol.* *35*, 229–237.
- Macatonia, S.E., Hosken, N.A., Litton, M., Vieira, P., Hsieh, C.S., Culpepper, J.A., Wysocka, M., Trinchieri, G., Murphy, K.M., and O'Garra, A. (1995). Dendritic cells produce IL-12 and direct the development of Th1 cells from naive CD4+ T cells. *J. Immunol.* *154*, 5071–5079.
- Macfarlane, G.T., and Gibson, G.R. (1991). Formation of glycoprotein degrading enzymes by *Bacteroides fragilis*. *FEMS Microbiol. Lett.* *61*, 289–293.
- Macfarlane, G.T., and Gibson, G.R. (1993). Characteristics of protease synthesis in *Bacteroides splanchnicus* NCTC 10825. *Appl. Microbiol. Biotechnol.* *39*, 506–511.
- Macfarlane, S., and Dillon, J.F. (2007). Microbial biofilms in the human gastrointestinal tract. *J. Appl. Microbiol.* *102*, 1187–1196.
- Macfarlane, G.T., Cummings, J.H., and Allison, C. (1986). Protein degradation by human intestinal bacteria. *J. Gen. Microbiol.* *132*, 1647–1656.
- Macfarlane, G.T., Macfarlane, S., and Gibson, G.R. (1992). Synthesis and release of proteases by *Bacteroides fragilis*. *Curr. Microbiol.* *24*, 55–59.

Macheboeuf, P., Buffalo, C., Fu, C., Zinkernagel, A.S., Cole, J.N., Johnson, J.E., Nizet, V., and Ghosh, P. (2011). Streptococcal M1 protein constructs a pathological host fibrinogen network. *Nature* 472, 64–68.

Maheswaran, M. (2003). Carbon-source-dependent nitrogen regulation in *Escherichia coli* is mediated through glutamine-dependent GlnB signalling. *Microbiology* 149, 2163–2172.

Maki, K., Morita, T., Otaka, H., and Aiba, H. (2010). A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA. *Mol. Microbiol.* 76, 782–792.

Manning, A.J., and Kuehn, M.J. (2011). Contribution of bacterial outer membrane vesicles to innate bacterial defense. *BMC Microbiol.* 11, 258.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9, 387–402.

Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 4, 447–457.

Martinez-Medina, M., Naves, P., Blanco, J., Aldeguer, X., Blanco, J.E., Blanco, M., Ponte, C., Soriano, F., Darfeuille-Michaud, A., and Garcia-Gil, L.J. (2009). Biofilm formation as a novel phenotypic feature of adherent-invasive *Escherichia coli* (AIEC). *BMC Microbiol.* 9, 202.

Matsuka, Y. V., Pillai, S., Gubba, S., Musser, J.M., and Olmsted, S.B. (1999). Fibrinogen cleavage by the *Streptococcus pyogenes* extracellular cysteine protease and generation of antibodies that inhibit enzyme proteolytic activity. *Infect. Immun.* 67, 4326–4333.

Matsuzaki, G., and Umemura, M. (2007). Interleukin-17 as an effector molecule of innate and acquired immunity against infections. *Microbiol. Immunol.* 51, 1139–1147.

Mayo, J.A., Harty, D.W.S., Cook, S.L., and Jacques, N.A. (2000). Environmental regulation of glycosidase and peptidase production by *Streptococcus gordonii* FSS2. *Microbiology* 146, 1923–1931.

Mayrand, D., and Grenier, D. (1989). Biological activities of outer membrane vesicles. *Can. J. Microbiol.* 35, 607–613.

- Mazmanian, S.K., Cui, H.L., Tzianabos, A.O., and Kasper, D.L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122, 107–118.
- Mazmanian, S.K., Round, J.L., and Kasper, D.L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* 453, 620–625.
- McBroom, A.J., and Kuehn, M.J. (2005). Outer Membrane Vesicles. *EcoSal Plus* 1.
- McBroom, A.J., Johnson, A.P., Vemulapalli, S., and Kuehn, M.J. (2006). Outer membrane vesicle production by *Escherichia coli* is independent of membrane instability. *J. Bacteriol.* 188, 5385–5392.
- McCarthy, K.J., Abrahamson, D.R., Bynum, K.R., St John, P.L., and Couchman, J.R. (1994). Basement membrane-specific chondroitin sulfate proteoglycan is abnormally associated with the glomerular capillary basement membrane of diabetic rats. *J. Histochem. Cytochem.* 42, 473–484.
- McDevitt, D., Nanavaty, T., House-Pompeo, K., Bell, E., Turner, N., McIntire, L., Foster, T., and Höök, M. (1997). Characterization of the interaction between the *Staphylococcus aureus* clumping factor (ClfA) and fibrinogen. *Eur. J. Biochem.* 247, 416–424.
- Medina, C., and Radomski, M.W. (2006). Role of matrix metalloproteinases in intestinal inflammation. *J. Pharmacol. Exp. Ther.* 318, 933–938.
- Mercenier, A., Simon, J.P., Vander Wauven, C., Haas, D., and Stalon, V. (1980). Regulation of enzyme synthesis in the arginine deiminase pathway of *Pseudomonas aeruginosa*. *J. Bacteriol.* 144, 159–163.
- Miles, D.O., Dyer, J.K., and Wong, J.C. (1976). Influence of amino acids on the growth of *Bacteroides melaninogenicus*. *J. Bacteriol.* 127, 899–903.
- Miller, C.G. (1975). *Peptidases and Proteases*.
- Miller, T.L., Wolin, M.J., Conway de Macario, E., and Macario, A.J. (1982). Isolation of *Methanobrevibacter smithii* from human feces. *Appl. Environ. Microbiol.* 43, 227–232.
- Millward, D.J. (2008). Sufficient protein for our elders? *Am. J. Clin. Nutr.* 88, 1187–1188.
- Molle, V., Nakaura, Y., Shivers, R.P., Yamaguchi, H., Losick, R., Fujita, Y., and Sonenshein, A.L. (2003). Additional targets of the *Bacillus subtilis* global regulator

CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J. Bacteriol.* *185*, 1911–1922.

Moncrief, J.S., Obiso, R., Barroso, L. a, Kling, J.J., Wright, R.L., Vantassell, R.L., Lyerly, D.M., and Wilkins, T.D. (1995). The Enterotoxin of *Bacteroides-Fragilis* Is a Metalloprotease. *Infec.Immunity Vol 63*, 175–181.

Muraca, M., Putignani, L., Fierabracci, A., Teti, A., and Perilongo, G. (2015). Gut microbiota-derived outer membrane vesicles: under-recognized major players in health and disease? *Discov. Med.* *19*, 343–348.

Murphy, K.M., and Reiner, S.L. (2002). The lineage decisions of helper T cells. *Nat. Rev. Immunol.* *2*, 933–944.

Myers, L.L., Firehammer, B.D., Shoop, D.S., and Border, M.M. (1984). *Bacteroides fragilis*: a possible cause of acute diarrheal disease in newborn lambs. *Infect. Immun.* *44*, 241–244.

Myers, L.L., Shoop, D.S., Stackhouse, L.L., Newman, F.S., Flaherty, R.J., Letson, G.W., and Sack, R.B. (1987). Isolation of enterotoxigenic *Bacteroides fragilis* from humans with diarrhea. *J. Clin. Microbiol.* *25*, 2330–2333.

Myers, L.L., Shoop, D.S., Collins, J.E., and Bradbury, W.C. (1989). Diarrheal disease caused by enterotoxigenic *Bacteroides fragilis* in infant rabbits. *J. Clin. Microbiol.* *27*, 2025–2030.

Nagy, E., Manncke, B., and Werner, H. (1994). Fibronectin and vitronectin binding of *Bacteroides fragilis* and eight other species of the genus. *Zentralbl. Bakteriologie.* *281*, 235–239.

Nakayama-Imahiji, H., Hirakawa, H., Ichimura, M., Wakimoto, S., Kuhara, S., Hayashi, T., and Kuwahara, T. (2009). Identification of the site-specific DNA invertase responsible for the phase variation of SusC/SusD family outer membrane proteins in *Bacteroides fragilis*. *J. Bacteriol.* *191*, 6003–6011.

Namavar, F., Theunissen, E.B.M., Verweij-Van Vught, a. M.J.J., Peerbooms, P.G.H., Bal, M., Hoitsma, H.F.W., and Maclaren, D.M. (1989). Epidemiology of the *Bacteroides fragilis* group in the colonic flora in 10 patients with colonic cancer. *J. Med. Microbiol.* *29*, 171–176.

Neerman-Arbez, M. (2001). The molecular basis of inherited afibrinogenaemia. *Thromb. Haemost.* *86*, 154–163.

Neurath, H. (1984). Evolution of proteolytic enzymes. *Science* *224*, 350–357.

Neurath, H. (1999). Proteolytic enzymes, past and future. *Proc. Natl. Acad. Sci.* *96*, 10962–10963.

Ng, K.M., Ferreyra, J.A., Higginbottom, S.K., Lynch, J.B., Kashyap, P.C., Gopinath, S., Naidu, N., Choudhury, B., Weimer, B.C., Monack, D.M., et al. (2013). Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* *502*, 96–99.

Nikitina, A.S., Kharlampieva, D.D., Babenko, V. V, Shirokov, D.A., Vakhitova, M.T., Manolov, A.I., Shkoporov, A.N., Taraskina, A.E., Manuvera, V.A., Lazarev, V.N., et al. (2015). Complete Genome Sequence of an Enterotoxigenic *Bacteroides fragilis*. *3*, 4–5.

Novick, R.P., and Geisinger, E. (2008). Quorum sensing in staphylococci. *Annu. Rev. Genet.* *42*, 541–564.

Obiso, R.J., Azghani, A.O., and Wilkins, T.D. (1997). The *Bacteroides fragilis* toxin fragilysin disrupts the paracellular barrier of epithelial cells. *Infect. Immun.* *65*, 1431–1439.

Ohara, N., Kikuchi, Y., Shoji, M., Naito, M., and Nakayama, K. (2006). Superoxide dismutase-encoding gene of the obligate anaerobe *Porphyromonas gingivalis* is regulated by the redox-sensing transcription activator OxyR. *Microbiology* *152*, 955–966.

Ohbayashi, T., Irie, A., Murakami, Y., Nowak, M., Potempa, J., Nishimura, Y., Shinohara, M., and Imamura, T. (2011). Degradation of fibrinogen and collagen by staphopains, cysteine proteases released from *Staphylococcus aureus*. *Microbiology* *157*, 786–792.

Oido-Mori, M., Rezzonico, R., Wang, P.L., Kowashi, Y., Dayer, J.M., Baehni, P.C., and Chizzolini, C. (2001). *Porphyromonas gingivalis* gingipain-R enhances interleukin-8 but decreases gamma interferon-inducible protein 10 production by human gingival fibroblasts in response to T-cell contact. *Infect. Immun.* *69*, 4493–4501.

Onderdonk, a B., Kasper, D.L., Cisneros, R.L., and Bartlett, J.G. (1977). The capsular polysaccharide of *Bacteroides fragilis* as a virulence factor: comparison of the pathogenic potential of encapsulated and unencapsulated strains. *J. Infect. Dis.* *136*, 82–89.

Ozeri, V., Rosenshine, I., Mosher, D.F., Fässler, R., and Hanski, E. (1998). Roles of integrins and fibronectin in the entry of *Streptococcus pyogenes* into cells via protein F1. *Mol. Microbiol.* *30*, 625–637.

- Pamer, E.G. (2007). Immune responses to commensal and environmental microbes. *Nat. Immunol.* 8, 1173–1178.
- Parente, J.A., Salem-Izacc, S.M., Santana, J.M., Pereira, M., Borges, C.L., Bailão, A.M., and Soares, C.M.A. (2010). A secreted serine protease of *Paracoccidioides brasiliensis* and its interactions with fungal proteins. *BMC Microbiol.* 10, 292.
- Pasare, C., and Medzhitov, R. (2004). Toll-like receptors: linking innate and adaptive immunity. *Microbes Infect.* 6, 1382–1387.
- Patrick, S. (2002). Bacteroides. In *Molecular Medical Microbiology*, M. Sussman, ed. (London: Academic Press), pp. 1921–1948.
- Patrick, S., and Duerden, B.I. (2006). Gram-negative non-spore forming obligate anaerobes. In *Principles and Practice of Clinical Bacteriology*, S.H. Gillespie, and P. Hawkey, eds. (London: Wiley), pp. 541–556.
- Patrick, S., and Lutton, D. (1990). Bacteroides fragilis surface structure expression in relation to virulence. *Med. Mal. Infect.* 20, 19–25.
- Patrick, S., Reid, J.H., and Coffey, a (1986). Capsulation of in vitro and in vivo grown Bacteroides species. *J. Gen. Microbiol.* 132, 1099–1109.
- Patrick, S., Lutton, D. a., and Crockard, a. D. (1995a). Immune reactions to Bacteroides fragilis populations with three different types of capsule in a model of infection. *Microbiology* 141, 1969–1976.
- Patrick, S., Stewart, L.D., Damani, N., Wilson, K.G., Lutton, D. a, Larkin, M.J., Poxtons, I., and Browns, R. (1995b). Immunological detection of Bacteroides fragilis in clinical Samples. *J. Med. Microbiol.* 43, 99–109.
- Patrick, S., McKenna, J.P., O’Hagan, S., and Dermott, E. (1996). A comparison of the haemagglutinating and enzymic activities of Bacteroides fragilis whole cells and outer membrane vesicles. *Microb. Pathog.* 20, 191–202.
- Patrick, S., Parkhill, J., McCoy, L.J., Lennard, N., Larkin, M.J., Collins, M., Sczaniecka, M., and Blakely, G. (2003). Multiple inverted DNA repeats of Bacteroides fragilis that control polysaccharide antigenic variation are similar to the hin region inverted repeats of Salmonella typhimurium. *Microbiology* 149, 915–924.
- Patrick, S., Houston, S., Thacker, Z., and Blakely, G.W. (2009). Mutational analysis of genes implicated in LPS and capsular polysaccharide biosynthesis in the opportunistic pathogen Bacteroides fragilis. *Microbiology* 155, 1039–1049.

- Patrick, S., Blakely, G.W., Houston, S., Moore, J., Abratt, V.R., Bertalan, M., Cerdeño-Tárraga, A.M., Quail, M. a., Corton, N., Corton, C., et al. (2010). Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiology* *156*, 3255–3269.
- Patrick, S., Jobling, K.L., O'Connor, D., Thacker, Z., Dryden, D.T.F., and Blakely, G.W. (2011). A unique homologue of the eukaryotic protein-modifier ubiquitin present in the bacterium *Bacteroides fragilis*, a predominant resident of the human gastrointestinal tract. *Microbiology* *157*, 3071–3078.
- Patti, J.M., Allen, B.L., McGavin, M.J., and Höök, M. (1994). MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu. Rev. Microbiol.* *48*, 585–617.
- Pauer, H., Ferreira, E.D.O., Dos Santos-Filho, J., Portela, M.B., Zingali, R.B., Soares, R.M.A., and Domingues, R.M.C.P. (2009). A TonB-dependent outer membrane protein as a *Bacteroides fragilis* fibronectin-binding molecule. *FEMS Immunol. Med. Microbiol.* *55*, 388–395.
- Pauer, H., Cavalcanti, S.N. V, Teixeira, F.L., Santos-Filho, J., Vommaro, R.C., Oliveira, A.C.S.C., Ferreira, E.O., and Domingues, R.R.M.C.P. (2013). Inactivation of a fibronectin-binding TonB-dependent protein increases adhesion properties of *Bacteroides fragilis*. *J. Med. Microbiol.* *62*, 1524–1530.
- Paul, L., Mishra, P.K., Blumenthal, R.M., and Matthews, R.G. (2007). Integration of regulatory signals through involvement of multiple global regulators: control of the *Escherichia coli* gltBDF operon by Lrp, IHF, Crp, and ArgR. *BMC Microbiol.* *7*, 2.
- Penkett, C.J., Dobson, C.M., Smith, L.J., Bright, J.R., Pickford, A.R., Campbell, I.D., and Potts, J.R. (2000). Identification of residues involved in the interaction of *Staphylococcus aureus* fibronectin-binding protein with the (4)F1(5)F1 module pair of human fibronectin using heteronuclear NMR spectroscopy. *Biochemistry* *39*, 2887–2893.
- Pereira, M., Rybarczyk, B.J., Odrliin, T.M., Hocking, D.C., Sottile, J., and Simpson-Haidaris, P.J. (2002). The incorporation of fibrinogen into extracellular matrix is dependent on active assembly of a fibronectin matrix. *J. Cell Sci.* *115*, 609–617.
- Pérez-Cruz, C., Carrión, O., Delgado, L., Martinez, G., López-Iglesias, C., and Mercade, E. (2013). New type of outer membrane vesicle produced by the gram-negative bacterium *Shewanella vesiculosa* M7T: Implications for DNA content. *Appl. Environ. Microbiol.* *79*, 1874–1881.
- Perkins, T.T., Kingsley, R.A., Fookes, M.C., Gardner, P.P., James, K.D., Yu, L., Assefa, S.A., He, M., Croucher, N.J., Pickard, D.J., et al. (2009). A strand-specific

RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet.* 5, e1000569.

Peterson, C.N., Mandel, M.J., and Silhavy, T.J. (2005). *Escherichia coli* starvation diets: essential nutrients weigh in distinctly. *J. Bacteriol.* 187, 7549–7553.

Petridis, M., Benjak, A., and Cook, G.M. (2015). Defining the nitrogen regulated transcriptome of *Mycobacterium smegmatis* using continuous culture. *BMC Genomics* 16, 821.

Piroth, L., Que, Y.A., Widmer, E., Panchaud, A., Piu, S., Entenza, J.M., and Moreillon, P. (2008). The fibrinogen- and fibronectin-binding domains of *Staphylococcus aureus* fibronectin-binding protein A synergistically promote endothelial invasion and experimental endocarditis. *Infect. Immun.* 76, 3824–3831.

Pittman, K.A., and Bryant, M.P. (1964). Peptides and other nitrogen sources for growth of *Bacteroides ruminicola*. *J. Bacteriol.* 88, 401–410.

Potempa, M., Potempa, J., Kantyka, T., Nguyen, K.-A., Wawrzonek, K., Manandhar, S.P., Popadiak, K., Riesbeck, K., Eick, S., and Blom, A.M. (2009). Interpain A, a cysteine proteinase from *Prevotella intermedia*, inhibits complement by degrading complement factor C3. *PLoS Pathog.* 5, e1000316.

Pridmore, R.D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A.-C., Zwahlen, M.-C., Rouvet, M., Altermann, E., Barrangou, R., et al. (2004). The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2512–2517.

Proctor, L.M. (2014). *The Human Microbiome: A True Story about You and Trillions of Your Closest (Microscopic) Friends.* 4–7.

Puente, X.S., Sánchez, L.M., Gutiérrez-Fernández, A., Velasco, G., and López-Otín, C. (2005). A genomic view of the complexity of mammalian proteolytic systems. *Biochem. Soc. Trans.* 33, 331–334.

Pumbwe, L., Skilbeck, C.A., and Wexler, H.M. (2006). The *Bacteroides fragilis* cell envelope: quarterback, linebacker, coach-or all three? *Anaerobe* 12, 211–220.

Raghavan, R., Groisman, E.A., and Ochman, H. (2011). Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res.* 21, 1487–1497.

Raghavan, R., Sloan, D.B., and Ochman, H. (2012). Antisense Transcription Is Pervasive but Rarely Conserved in Enteric Bacteria. *MBio* 3, e00156–12 – e00156–12.

- Rakoff-Nahoum, S., Coyne, M.J., and Comstock, L.E. (2014). An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr. Biol.* *24*, 40–49.
- Rakowski, S.A., and Filutowicz, M. (2013). Plasmid R6K replication control. *Plasmid* *69*, 231–242.
- Rao, S.P., Gehlsen, K.R., and Catanzaro, A. (1992). Identification of a beta 1 integrin on *Mycobacterium avium*-*Mycobacterium intracellulare*. *Infect. Immun.* *60*, 3652–3657.
- Rawlings, N.D. (2004). MEROPS: the peptidase database. *Nucleic Acids Res.* *32*, 160D – 164.
- Rawlings, N.D. (2009). A large and accurate collection of peptidase cleavages in the MEROPS database. *Database 2009*, bap015–bap015.
- Rawlings, N.D., and Barrett, A.J. (1993). Evolutionary families of peptidases. *Biochem. J.* *290* (Pt 1), 205–218.
- Rawlings, N.D., and Barrett, A.J. (1995). Evolutionary families of metallopeptidases. *Methods Enzymol.* *248*, 183–228.
- Rawlings, N.D., and Barrett, A.J. (1999). MEROPS: the peptidase database. *Nucleic Acids Res.* *27*, 325–331.
- Rawlings, N.D., and Morton, F.R. (2008). The MEROPS batch BLAST: a tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie* *90*, 243–259.
- Reid, J.H., and Patrick, S. (1984). Phagocytic and serum killing of capsulate and non-capsulate *Bacteroides fragilis*. *J. Med. Microbiol.* *17*, 247–258.
- Reitzer, L. (2003). Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu. Rev. Microbiol.* *57*, 155–176.
- Riepe, S.P., Goldstein, J., and Alpers, D.H. (1980). Effect of secreted *Bacteroides* proteases on human intestinal brush border hydrolases. *J. Clin. Invest.* *66*, 314–322.
- Robertson, K.P., Smith, C.J., Gough, A.M., and Rocha, E.R. (2006). Characterization of *Bacteroides fragilis* hemolysins and regulation and synergistic interactions of HlyA and HlyB. *Infect. Immun.* *74*, 2304–2316.

- Rocha, E.R., and Smith, C.J. (1997). Regulation of *Bacteriodes fragilis* katB mRNA by oxidative stress and carbon limitation. *J. Bacteriol.* 179, 7033–7039.
- Rocha, E.R., Smith, A., Smith, C.J., and Brock, J.H. (2001). Growth inhibition of *Bacteroides fragilis* by hemopexin: Proteolytic degradation of hemopexin to overcome heme limitation. *FEMS Microbiol. Lett.* 199, 73–78.
- Róka, R., Rosztóczy, A., Leveque, M., Izbéki, F., Nagy, F., Molnár, T., Lonovics, J., Garcia-Villar, R., Fioramonti, J., Wittmann, T., et al. (2007). A pilot study of fecal serine-protease activity: a pathophysiologic factor in diarrhea-predominant irritable bowel syndrome. *Clin. Gastroenterol. Hepatol.* 5, 550–555.
- Romberger, D.J. (1997). Fibronectin. *Int. J. Biochem. Cell Biol.* 29, 939–943.
- Rotimi, V.O., and Eke, P.I. (1984). The Bactericidal Action of Human Serum on *Bacteroides* Species. *J. Med. Microbiol.* 18, 355–363.
- Rotstein, O.D., Nasmith, P.E., and Grinstein, S. (1987). The *Bacteroides* by-product succinic acid inhibits neutrophil respiratory burst by reducing intracellular pH. *Infect. Immun.* 55, 864–870.
- Round, J.L., Lee, S.M., Li, J., Tran, G., Jabri, B., A. Chatila, T., and Mazmanian, S.K. (2011). The Toll-like receptor pathway establishes commensal gut colonization. *Science* (80-.). 332, 974–977.
- Rudek, W., and Haque, R.U. (1976). Extracellular enzymes of the genus *bacteroides*. *J. Clin. Microbiol.* 4, 458–460.
- Sack, R.B., Albert, M.J., Alam, K., Neogi, P.K., and Akbar, M.S. (1994). Isolation of enterotoxigenic *Bacteroides fragilis* from Bangladeshi children with diarrhea: a controlled study. *J. Clin. Microbiol.* 32, 960–963.
- Salyers, A.A. (1984). *Bacteroides* of the human lower intestinal tract. *Annu. Rev. Microbiol.* 38, 293–313.
- Savage, D.C. (1977). *Microbial Ecology of the Gastrointestinal Tract.*
- Schaar, V., Nordström, T., Mörgelin, M., and Riesbeck, K. (2011). *Moraxella catarrhalis* outer membrane vesicles carry β -lactamase and promote survival of *Streptococcus pneumoniae* and *Haemophilus influenzae* by inactivating amoxicillin. *Antimicrob. Agents Chemother.* 55, 3845–3853.
- Schapiro, J.M., Gupta, R., Stefansson, E., Fang, F.C., and Limaye, A.P. (2004). Isolation of metronidazole-resistant *Bacteroides fragilis* carrying the nimA

- nitroreductase gene from a patient in Washington State. *J. Clin. Microbiol.* *42*, 4127–4129.
- Schwarzbauer, J.E. (1991). Fibronectin: from gene to protein. *Curr. Opin. Cell Biol.* *3*, 786–791.
- Schwarz-Linek, U., Höök, M., and Potts, J.R. (2004). The molecular basis of fibronectin-mediated bacterial adherence to host cells. *Mol. Microbiol.* *52*, 631–641.
- Sears, C.L. (2001). The toxins of *Bacteroides fragilis*. *Toxicon* *39*, 1737–1746.
- Seksik, P., Rigottier-Gois, L., Gramet, G., Sutren, M., Pochart, P., Marteau, P., Jian, R., and Doré, J. (2003). Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* *52*, 237–242.
- Sharma, A., Sojar, H.T., Glurich, I., Honma, K., Kuramitsu, H.K., and Genco, R.J. (1998). Cloning, expression, and sequencing of a cell surface antigen containing a leucine-rich repeat motif from *Bacteroides forsythus* ATCC 43037. *Infect. Immun.* *66*, 5703–5710.
- Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., et al. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* *464*, 250–255.
- Shen, W., Steinrück, H., and Ljungh, a (1995). Expression of binding of plasminogen, thrombospondin, vitronectin, and fibrinogen, and adhesive properties by *Escherichia coli* strains isolated from patients with colonic diseases. *Gut* *36*, 401–406.
- Shen, Y., Giardino Torchia, M.L., Lawson, G.W., Karp, C.L., Ashwell, J.D., and Mazmanian, S.K. (2012). Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. *Cell Host Microbe* *12*, 509–520.
- Sherwood, J.E., Fraser, S., Citron, D.M., Wexler, H., Blakely, G., Jobling, K., and Patrick, S. (2011). Multi-drug resistant *Bacteroides fragilis* recovered from blood and severe leg wounds caused by an improvised explosive device (IED) in Afghanistan. *Anaerobe* *17*, 152–155.
- Shimizu, K. (2013). *Metabolic Regulation of a Bacterial Cell System with Emphasis on. 2013.*
- Shoop, D.S., Myers, L.L., and LeFever, J.B. (1990). Enumeration of enterotoxigenic *Bacteroides fragilis* in municipal sewage. *Appl. Environ. Microbiol.* *56*, 2243–2244.

Shuman, S. (1991). Recombination mediated by vaccinia virus DNA topoisomerase I in *Escherichia coli* is sequence specific. *Proc. Natl. Acad. Sci. U. S. A.* 88, 10104–10108.

Shuman, S. (1994). Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *J. Biol. Chem.* 269, 32678–32684.

Sijbrandi, R., Den Blaauwen, T., Tame, J.R.H., Oudega, B., Luirink, J., and Otto, B.R. (2005). Characterization of an iron-regulated alpha-enolase of *Bacteroides fragilis*. *Microbes Infect.* 7, 9–18.

Sijbrandi, R., Stork, M., Luirink, J., and Otto, B.R. (2008). Pbp, a cell-surface exposed plasminogen binding protein of *Bacteroides fragilis*. *Microbes Infect.* 10, 514–521.

Simon, G.L., Klemper, M.S., Kasper, D.L., and Gorbach, S.L. (1982). Alterations in Opsonophagocytic Killing by Neutrophils of *Bacteroides fragilis* Associated with Animal and Laboratory Passage: Effect of Capsular Polysaccharide. *J. Infect. Dis.* 145, 72–77.

Smith, C.J. (1995a). Genetic transformation of *Bacteroides* spp. using electroporation. *Methods Mol. Biol.* 47, 161–169.

Smith, H. (1995b). The revival of interest in mechanisms of bacterial pathogenicity. *Biol. Rev.* 70, 277–316.

Smith, E., and Macfarlane, G. (1998). Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. *FEMS Microbiol. Ecol.* 25, 355–368.

Somerville, G.A., and Proctor, R.A. (2009). At the crossroads of bacterial metabolism and virulence factor synthesis in *Staphylococci*. *Microbiol. Mol. Biol. Rev.* 73, 233–248.

Sonnenburg, J.L., Angenent, L.T., and Gordon, J.I. (2004). Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? *Nat. Immunol.* 5, 569–573.

Southern, J.A., Parker, J.R., and Woods, D.R. (1986). Expression and purification of glutamine synthetase cloned from *Bacteroides fragilis*. *J. Gen. Microbiol.* 132, 2827–2835.

- Southern, J.A., Parker, J.R., and Woods, D.R. (1987). Novel Structure, Properties and Inactivation of Glutamine Synthetase Cloned from *Bacteroides fragilis*. *Microbiology* *133*, 2437–2446.
- Stappenbeck, T.S., Hooper, L. V, and Gordon, J.I. (2002). Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 15451–15455.
- Steffens, L.S., Nicholson, S., Paul, L. V., Nord, C.E., Patrick, S., and Abratt, V.R. (2010). *Bacteroides fragilis* RecA protein overexpression causes resistance to metronidazole. *Res. Microbiol.* *161*, 346–354.
- Steidler, L., Hans, W., Schotte, L., Neiryneck, S., Obermeier, F., Falk, W., Fiers, W., and Remaut, E. (2000). Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. *Science* *289*, 1352–1355.
- Stentz, R., Horn, N., Cross, K., Salt, L., Brearley, C., Livermore, D.M., and Carding, S.R. (2015). Cephalosporinases associated with outer membrane vesicles released by *Bacteroides* spp. protect gut pathogens and commensals against β -lactam antibiotics. *J. Antimicrob. Chemother.* *70*, 701–709.
- Stevenson, R.M.W. (1979). Amino acid uptake systems in *Bacteroides ruminicola*. *Can. J. Microbiol.* *25*, 1161–1168.
- Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* *189*, 113–130.
- Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L.P., and Lochs, H. (2005). Spatial Organization and Composition of the Mucosal Flora in Patients with Inflammatory Bowel Disease. *J. Clin. Microbiol.* *43*, 3380–3389.
- Szöke, I., Pascu, C., Nagy, E., Ljung, Å., and Wadström, T. (1996). Binding of extracellular matrix proteins to the surface of anaerobic bacteria. *J. Med. Microbiol.* *45*, 338–343.
- El Tahir, Y., and Skurnik, M. (2001). YadA, the multifaceted *Yersinia* adhesin. *Int. J. Med. Microbiol.* *291*, 209–218.
- Tally, F.P., and Ho, J.L. (1987). Management of patients with intraabdominal infection due to colonic perforation. *Curr. Clin. Top. Infect. Dis.* *8*, 266–295.
- Tanaka, H., Ito, F., and Iwasaki, T. (1992). Purification and characterization of a sialidase from *Bacteroides fragilis* SBT3182. *Biochem. Biophys. Res. Commun.* *189*, 524–529.

- Van Tassell, R.L., Lyerly, D.M., and Wilkins, T.D. (1992). Purification and characterization of an enterotoxin from *Bacteroides fragilis*. *Infect. Immun.* *60*, 1343–1350.
- Thornton, R.F., Kagawa, T.F., O’Toole, P.W., and Cooney, J.C. (2010). The dissemination of C10 cysteine protease genes in *Bacteroides fragilis* by mobile genetic elements. *BMC Microbiol.* *10*, 122.
- Thornton, R.F., Murphy, E.C., Kagawa, T.F., O’Toole, P.W., and Cooney, J.C. (2012). The effect of environmental conditions on expression of *Bacteroides fragilis* and *Bacteroides thetaiotaomicron* C10 protease genes. *BMC Microbiol.* *12*, 190.
- Tian, Z.X., Li, Q.S., Buck, M., Kolb, A., and Wang, Y.P. (2001). The CRP-cAMP complex and downregulation of the *glnAp2* promoter provides a novel regulatory linkage between carbon metabolism and nitrogen assimilation in *Escherichia coli*. *Mol. Microbiol.* *41*, 911–924.
- Toledano, M.B., Kullik, I., Trinh, F., Baird, P.T., Schneider, T.D., and Storz, G. (1994). Redox-dependent shift of OxyR-DNA contacts along an extended DNA-binding site: A mechanism for differential promoter selection. *Cell* *78*, 897–909.
- Trinh, S., and Reyssset, G. (1996). Detection by PCR of the *nim* genes encoding 5-nitroimidazole resistance in *Bacteroides* spp. *J. Clin. Microbiol.* *34*, 2078–2084.
- Troy, E.B., and Kasper, D.L. (2010). Beneficial effects of *Bacteroides fragilis* polysaccharides on the immune system. *Front. Biosci. (Landmark Ed.)* *15*, 25–34.
- Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* *444*, 1027–1031.
- Tzianabos, A.O., Onderdonk, A.B., Rosner, B., Cisneros, R.L., and Kasper, D.L. (1993). Structural features of polysaccharides that induce intra-abdominal abscesses. *Science* *262*, 416–419.
- Tzianabos, A.O., Wang, J.Y., and Lee, J.C. (2001). Structural rationale for the modulation of abscess formation by *Staphylococcus aureus* capsular polysaccharides. *Proc. Natl. Acad. Sci. U. S. A.* *98*, 9365–9370.
- Valentine, P.J., Shoemaker, N.B., and Salyers, a. a. (1988). Mobilization of *Bacteroides* plasmids by bacteroides conjugal elements. *J. Bacteriol.* *170*, 1319–1324.

- Varel, V.H., and Bryant, M.P. (1974). Nutritional features of *Bacteroides fragilis* subsp. *fragilis*. *Appl. Microbiol.* 28, 251–257.
- Vieira, J.M.B.D., Vallim, D.C., Ferreira, E.O., Seabra, S.H., Vommaro, R.C., Avelar, K.E.S., De Souza, W., Ferreira, M.C.S., and Domingues, R.M.C.P. (2005). *Bacteroides fragilis* interferes with iNOS activity and leads to pore formation in macrophage surface. *Biochem. Biophys. Res. Commun.* 326, 607–613.
- Vieira, J.M.B.D., Seabra, S.H., Vallim, D.C., Américo, M. a., Fracallanza, S.E.L., Vommaro, R.C., and Domingues, R.M.C.P. (2009). *Bacteroides fragilis* induce necrosis on mice peritoneal macrophages: In vitro and in vivo assays. *Biochem. Biophys. Res. Commun.* 387, 627–632.
- Van Vliet, A.H.M. (2010). Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiol. Lett.* 302, 1–7.
- Voigt, B., Schweder, T., Sibbald, M.J.J.B., Albrecht, D., Ehrenreich, A., Bernhardt, J., Feesche, J., Maurer, K.-H., Gottschalk, G., van Dijl, J.M., et al. (2006). The extracellular proteome of *Bacillus licheniformis* grown in different media and under different nutrient starvation conditions. *Proteomics* 6, 268–281.
- Voigt, B., Hoi, L.T., Jürgen, B., Albrecht, D., Ehrenreich, A., Veith, B., Evers, S., Maurer, K.-H., Hecker, M., and Schweder, T. (2007). The glucose and nitrogen starvation response of *Bacillus licheniformis*. *Proteomics* 7, 413–423.
- Vrhovski, B., and Weiss, A.S. (1998). Biochemistry of tropoelastin. *Eur. J. Biochem.* 258, 1–18.
- Wang, Q., McLoughlin, R.M., Cobb, B.A., Charrel-Dennis, M., Zaleski, K.J., Golenbock, D., Tzianabos, A.O., and Kasper, D.L. (2006). A bacterial carbohydrate links innate and adaptive responses through Toll-like receptor 2. *J. Exp. Med.* 203, 2853–2863.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63.
- Wann, E.R., Gurusiddappa, S., and Höök, M. (2000). The fibronectin-binding MSCRAMM FnbpA of *Staphylococcus aureus* is a bifunctional protein that also binds to fibrinogen. *J. Biol. Chem.* 275, 13863–13871.
- Watanabe, K. (2004). Collagenolytic proteases from bacteria. *Appl. Microbiol. Biotechnol.* 63, 520–526.

- Watts, C., and Powis, S. (1999). Pathways of antigen processing and presentation. *Rev. Immunogenet.* *1*, 60–74.
- Weisel, J.W. (2005). Fibrinogen and fibrin. *Adv. Protein Chem.* *70*, 247–299.
- Weisel, J.W., Stauffacher, C. V, Bullitt, E., and Cohen, C. (1985). A model for fibrinogen: domains and sequence. *Science* *230*, 1388–1391.
- Weng, M.L., and Zalkin, H. (1987). Structural role for a conserved region in the CTP synthetase glutamine amide transfer domain. *J. Bacteriol.* *169*, 3023–3028.
- Westerlund, B., and Korhonen, T.K. (1993). Bacterial proteins binding to the mammalian extracellular matrix. *Mol. Microbiol.* *9*, 687–694.
- Westermann, A.J., Gorski, S.A., and Vogel, J. (2012). Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* *10*, 618–630.
- Wexler, H.M. (2007). Bacteroides: The good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* *20*, 593–621.
- Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* *75*, 39–68.
- Whitworth, D.E. (2011). Myxobacterial vesicles death at a distance? *Adv. Appl. Microbiol.* *75*, 1–31.
- Wikstrom, M., and Linde, a. (1986). Ability of oral bacteria to degrade fibronectin. *Infect. Immun.* *51*, 707–711.
- Wikstrom, M.B., Dahltn, G., and Linde, A. (1983). activity in oral microorganisms . Fibrinolytic and Fibrinolytic Activity in Oral Microorganisms. *17*, 759–767.
- Williams, K.J., Bryant, W.A., Jenkins, V.A., Barton, G.R., Witney, A.A., Pinney, J.W., and Robertson, B.D. (2013). Deciphering the response of *Mycobacterium smegmatis* to nitrogen stress using bipartite active modules. *BMC Genomics* *14*, 436.
- Willis, A.T. (1991). Abdominal sepsis. In *Anaerobes in Human Disease*, B.I. Duerden, and D. B.S., eds. pp. 197–223.
- Wills-Karp, M., Santeliz, J., and Karp, C.L. (2001). The germless theory of allergic disease: revisiting the hygiene hypothesis. *Nat. Rev. Immunol.* *1*, 69–75.

- Wilson, M.M., Anderson, D.E., and Bernstein, H.D. (2015). Analysis of the Outer Membrane Proteome and Secretome of *Bacteroides fragilis* Reveals a Multiplicity of Secretion Mechanisms. *PLoS One* *10*, e0117732.
- Winkler, J.R., John, S.R., Kramer, R.H., Hoover, C.I., and Murray, P. a. (1987). Attachment of oral bacteria to a basement-membrane-like matrix and to purified matrix proteins. *Infect. Immun.* *55*, 2721–2726.
- Wu, C., Keivens, V.M., O'Toole, T.E., McDonald, J.A., and Ginsberg, M.H. (1995). Integrin activation and cytoskeletal interaction are essential for the assembly of a fibronectin matrix. *Cell* *83*, 715–724.
- Wu, S., Lim, K.C., Huang, J., Saidi, R.F., and Sears, C.L. (1998). *Bacteroides fragilis* enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 14979–14984.
- Wu, S., Shin, J., Zhang, G., Cohen, M., Franco, A., and Sears, C.L. (2006). The *Bacteroides fragilis* Toxin Binds to a Specific Intestinal Epithelial Cell Receptor. *Infect. Immun.* *74*, 5382–5390.
- Xu, J., and Gordon, J.I. (2003). Honor thy symbionts. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 10452–10459.
- Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., et al. (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol.* *5*, e156.
- Yamamoto, I., Abe, A., Saito, H., and Ishimoto, M. (1984). The pathway of ammonia assimilation in *Bacteroides fragilis*. *J. Gen. Appl. Microbiol.* *30*, 499–508.
- Yamamoto, I., Saito, H., and Ishimoto, M. (1987). Comparison of properties of glutamate dehydrogenases in members of the *Bacteroides fragilis* group. *J. Gen. Appl. Microbiol.* *33*, 429–436.
- Yan, D. (2007). Protection of the glutamate pool concentration in enteric bacteria. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 9475–9480.
- Zahoor, A., Lafleur, M. V, Knight, R.C., Loman, H., and Edwards, D.I. (1987). DNA damage induced by reduced nitroimidazole drugs. *Biochem. Pharmacol.* *36*, 3299–3304.
- Zimmer, D.P., Soupene, E., Lee, H.L., Wendisch, V.F., Khodursky, A.B., Peter, B.J., Bender, R.A., and Kustu, S. (2000). Nitrogen regulatory protein C-controlled genes

of *Escherichia coli*: Scavenging as a defense against nitrogen limitation. *Proc. Natl. Acad. Sci.* 97, 14674–14679.

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

www.studyblue.com