THE UNIVERSITY *of* EDINBURGH

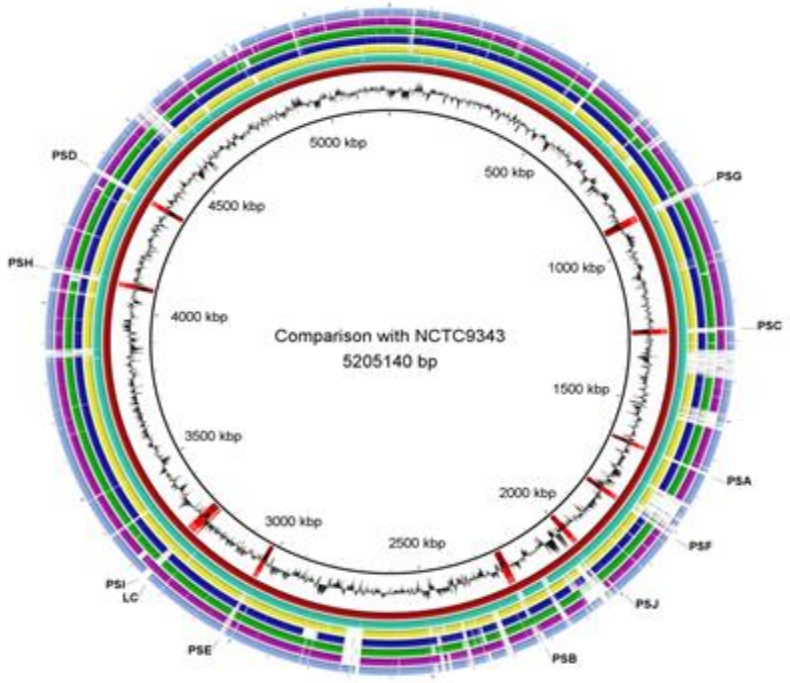
This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- 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.

Horizontal gene transfer in

Bacteroides fragilis

Kelly L. Jobling



Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2013

Declaration

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

Acknowledgements

I am truly grateful to everyone who has supported me through this PhD and made it an unforgettable experience. I would like to begin by thanking my supervisor, Dr Garry Blakely, for his support and guidance throughout this process. Most of all, I would like to thank Garry for inspiring me with his enthusiasm for this project. I would also like to thank my second supervisor, Dr Maurice Gallagher, and our collaborator, Prof. Sheila Patrick, for their valuable input. I am grateful to the BBSRC which provided funding for this work.

I would like to say a huge thank you to all of the Blakely lab members, past and present, who have not only put up with my bossiness in the lab, but have also made it a fantastic place to work. I would particularly like to thank Dr Prerna Vohra for all of her invaluable help and support, and David Roberts for always managing to put a smile on my face regardless of how terrible a science day it had been. Thanks to all of my friends in the Darwin Building who have celebrated and commiserated with me throughout the past four years. An extra special thank you must go to Dr Nicola McNiven and Dr Charlie Cockram, who have provided me with an endless supply of encouragement, support and friendship, it has been greatly appreciated.

Finally, I would like to thank my parents, my brother Ben and all of my family, for their love and support. I would especially like to thank my Mam for teaching me that I can achieve anything I put my mind to, none of this would have been possible without her.

Publications

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(4): 152-155.

Patrick, S., Jobling K.L., O'Connor D., Thacker Z., Dryden D.T. 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(11): 3071-3078.

Table of contents

Declaration i

Acknowledgements ii

Publications iii

List of Tables x

Abbreviations xi

Abstract 1

Lay Abstract 3

Chapter 1. Introduction 4

1.1. Horizontal Gene Transfer 4

1.2. Mechanisms of HGT 5

1.2.1. Conjugation 5

1.2.2. Bacteriophages and Transduction 6

1.2.3. Transformation 9

1.3. The microbiota of the human gastrointestinal tract 12

1.3.1. Resistance to pathogens 12

1.3.2. Development of the host immune system 14

1.4. Bacteroides species 15

1.4.1. *Bacteroides fragilis* Polysaccharide A 16

1.5. B. fragilis: An opportunistic pathogen 17

1.5.1. Virulence factors 17

1.5.1.1. Capsular polysaccharides 17

1.5.1.2. Enzymes 21

1.5.1.3. Enterotoxin 21

1.6. Bacterial outer membrane vesicles 22

1.7. Aims 25

2. Materials and Methods 26

2.1. Bacterial culture conditions 26

2.2. Antibiotics 26

2.3. Strains 27

2.4. Oligonucleotides 28

2.5. Plasmids 29

2.6. Bacterial techniques 29

2.6.1. *B. fragilis* growth curves 29

2.6.2. Preparation of chemically competent cells 29
2.6.3. Transformation – chemically competent cells 30
2.6.4. Transformation - Electroporation 30
2.6.5. Conjugation of *E. coli* S-17 with *B. fragilis* 30

2.7. DNA techniques *31*
2.7.1. Genomic DNA extraction 31
2.7.2. Plasmid DNA purification 31
2.7.3. Polymerase chain reaction (PCR) 31
2.7.4. Agarose gel electrophoresis 31
2.7.5. DNA purification 32
2.7.6. DNA quantification 32
2.7.7. Gel Extraction 32
2.7.8. Restriction endonuclease digestion 32
2.7.9. Dephosphorylation of DNA 32
2.7.10. Ligation of DNA 33
2.7.11. Plasmid construction using In-fusion HD Cloning kit 33
2.7.12. Whole genome sequencing 33
2.7.12.1. Sample preparation 33
2.7.12.2. Library preparation 34

2.8. Protein techniques *34*
2.8.1. Solutions 34
2.8.2. Preparation of whole cell protein extracts 35
2.8.3. Preparation of periplasmic protein extracts 35
2.8.4. Sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 35
2.8.5. Western blot analysis 36
2.8.6. Bradford assay 36
2.8.7. Immunofluorescence Microscopy 36

2.9. Bacteroides fragilis OMV *37*
2.9.1. Preparation of *B. fragilis* concentrated supernatants 37
2.9.2. DNA analysis of OMV 37
2.9.3. Transformation of *E. coli* XL1-Blue MRF' and *B. fragilis* Δ*ubb* with *B. fragilis* OMV 38
2.9.4. Transmission Electron Microscopy 38

2.10. Cell culture techniques *39*
2.10.1. Culture and passaging of cells 39
2.10.2. Cell counts 39
2.10.3. Mycoplasma detection 39
2.10.4. Incubation of *B. fragilis* concentrated supernatants with Caco-2 cells 40
2.10.5. Confocal Microscopy 40

2.11. Bacteriophage techniques *40*
2.11.1. Isolation of Bacteriophages 40
2.11.2. Plaque assay for phage titre 40

2.11.3. Transduction by bacteriophages *41*
2.11.4. Preparing bacteriophage DNA from lysates 41
2.11.4.1. PEG precipitation 41
2.11.4.2. Caesium Chloride gradients 41

2.11.4.3. Formamide DNA extraction 42

2.11. Bioinformatics *43*
2.11.1. Sequencing data, assembly and annotation 43
2.11.2. Genome comparisons 44

Chapter 3. Inter-strain Horizontal Gene Transfer **45**

3.1. Introduction *45*

3.2. Results *45*
3.2.1. Multi-drug resistant *B. fragilis* isolate, WAL610 45
3.2.2. Illumina MiSeq sequencing 47
3.2.2. Assembly and Annotation 48
3.2.3. Analysis of the polysaccharide biosynthesis loci 49
3.2.3.1 Strain RD48 52
3.2.3.2. Strain GNAB92 56
3.2.3.3. Strain BE1 59
3.2.3.4. Strain LS66 61
3.2.3.5. Large capsule assembly 62
3.2.3.6. Identifying Wzx and Wzy homologues 65
67
3.2.4. Whole genome comparison 69
3.2.5. DNA Restriction and Modification systems 70
3.2.6. *Bacteroides fragilis* toxin (BFT) 72

3.3 Conclusion *74*

Chapter 4. Inter-kingdom horizontal gene transfer **76**

4.1. Introduction *76*
4.1.1. Ubiquitin 76
4.1.2. Ubiquitin modifications 77
4.1.3. Ubiquitin and the immune system 78
4.1.4. Ubiquitin Binding Domains 79
4.1.5. Pathogens and the host ubiquitylation pathway 80

4.2. Results *82*
4.2.1. Comparison of protein sequence of BfUbb and eukaryotic ubiquitin 82
4.2.2. Ubiquitin and the *ubb* containing region 83
4.2.2.1. The presence of *ubb* in clinical isolates 83
4.2.2.2 Analysis of the low GC region of *B. fragilis* isolates by PCR 86
4.2.2.3. Analysis by whole genome sequencing. 88
4.2.3 Growth rate of Δ*ubb* 91
4.2.5 Expression of BfUbb 92
4.2.6 BfUbb forms dimers under reducing and non-reducing conditions 93
4.2.7 Generation and analysis of a *ubb* signal sequence deletion construct. 94
4.2.8 *Bacteroides fragilis* OMV 97
4.2.8.1 Isolation of OMV 97
4.2.8.2 Generation and analysis of a GFP-BfUbb fusion protein. 99
4.2.8.3 Detection of GFP-BfUbb fluorescence 103

4.2.8.5 *B. fragilis* OMV contain plasmid and chromosomal DNA 105
4.2.8.6 *B. fragilis* OMV-mediated transformation 106

4.3 Conclusion *108*

Chapter 5. Bacteriophages as mediators of horizontal gene transfer **111**

5.1. Introduction *111*

5.2. Results *111*
5.2.1. Isolation and identification of *B. fragilis*-specific bacteriophage 111
5.2.2. Restriction endonuclease digestion of bacteriophage DNA 118
5.2.3. Transduction of *ermF* cassette 124
5.2.4. Transmission electron microscopy 124
5.2.5. OMV interact with bacteriophage to decrease efficiency of infection 127

5.2.6. Development of a recombineering system in B. fragilis *129*
5.6.1.1. Construction of a PhiRecET expression vector 130

5.5. Conclusion *134*

6. General Discussion **136**

Bibliography **144**

List of Figures

Figure 1.1: Mechanisms of horizontal gene transfer in bacteria **10**

Figure 1.2: The lytic and lysogenic lifecycles of bacteriophages **11**

Figure 1.3*: B. fragilis* capsules **19**

Figure 1.4: Wzy-dependent polysaccharide biosynthesis **20**

Figure 2.1: Caesium chloride gradient set-up for bacteriophage isolation **43**

Figure 3.1: Artemis DNA plotter image of pWAL610 **47**

Figure 3.2: Comparison of RD48 polysaccharide biosynthesis loci **54**

Figure 3.3: RD48-Immunofluorescence Microscopy with PSB-specific antibody **55**

Figure 3.4: Comparison of GNAB92 polysaccharide biosynthesis loci **58**

Figure 3.5: ACT comparison of GNAB92 PSC and YCH46 PSA **59**

Figure 3.6: Comparison of BE1 polysaccharide biosynthesis loci **61**

Figure 3.7: Comparison of LS66 polysaccharide biosynthesis loci **64**

Figure 3.8: Comparison of the LC assembly genes **65**

Figure 3.9: Whole genome comparison with NCTC9343 **70**

Figure 3.10: Restriction modification related region **71**

Figure 3.11: Deletions in the type I R-M systems of RD48 **72**

Figure 3.12: ACT comparison of BfPAI pathogenicity island in GNAB92 and NCTC9343 **73**

Figure 4.1: The ubiquitylation pathway. **77**

Figure 4.2: Surfaces which interact with ubiquitin binding domains **80**

Figure 4.3: Alignment of the predicted protein sequence of BfUbb and human ubiquitin. **82**

Figure 4.4: Tertiary structure comparison of Ubb and BfUbb **83**

Figure 4.5: *ubb* in clinical isolates of *B. fragilis* **86**

Figure 4.6: Analysis of the low GC region of *B. fragilis* clinical isolates by PCR **87**

Figure 4.7: Sequence analysis of the low GC region of *B. fragilis* **89**

Figure 4.8: ACT comparison of the *ompA* to *mutL* region **90**

Figure 4.9: Comparison of the growth rate of NCTC9343 and Δ*ubb* **91**

Figure 4.10: Western blot analysis of anti-bovine ubiquitin antibody specificity **92**

Figure 4.11: Expression of BfUbb in whole cell extracts of *B. fragilis* **93**

Figure 4.12: BfUbb in reducing and non-reducing conditions **94**

Figure 4.13 : Construction and expression of a *ubb* signal sequence deletion plasmid **96**

Figure 4.14 : *B. fragilis* OMV **98**

Figure 4.15: Detection of BfUbb in concentrated supernatants. **99**

Figure 4.16: The generation of a GFP-BfUbb fusion construct **102**

Figure 4.17: Caco-2 cells incubated with concentrated supernatants containing GFP-BfUbb **104**

Figure 4.18: Colonic carcinoma cells incubated with *B. fragilis* OMV **105**

Figure 4.19: PCR detection of DNA isolated from *B. fragilis* OMV **106**

Figure 5.1: Bacteriophage titres on NCTC9343 and 638R **117**

Figure 5.2: Restriction endonuclease cleavage profiles of bacteriophages similar to ΦED01 **122**

Figure 5.3: Bacteriophages with restriction endonuclease cleavage profiles different to ΦED01 **123**

Figure 5.5: *B. fragilis* OMV interact with bacteriophage. **128**

Figure 5.6: *phirecE* and *phirecT* genes of *B. fragilis*-specific bacteriophage ΦED01 **132**

List of Tables

Table 2.1: List of Antibiotics **26**

Table 2.2: List of strains **27**

Table 2.3: List of oligonucleotides **28**

Table 2.4: List of plasmids **29**

Table3.2: Genome assembly data **50**

Table 3.3: Polysaccharide associated biosynthesis loci in *B. fragilis* BE1, GNAB92, RD48 and LS66. **51**

Table 3.4: List of potential Wzx and Wzy proteins **67**

Table 4.1: *B. fragilis* clinical isolates **85**

Table 5.1: Bacteriophage titres **112**

Abbreviations

ACT	Artemis Comparison Tool
APS	Ammonium persulphate
BFT	*Bacteroides fragilis* toxin
BfUbb	*Bacteroides fragilis* ubiquitin
BHI	Brain heart infusion
BLAST	Basic Local Alignment Search Tool
BRIG	BLAST Ring Image Generator
BSA	Bovine serum albumin
CDS	Coding sequence
CFTR	Cystic fibrosis transmembrane conductance regulator
CsCl	Caesium chloride
CSP	Competence-stimulating peptide
DAPI	Diamidino-2-phenylindole
DC	Dendritic cell
DLS	Dynamic Light Scattering
DM	Defined medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double-stranded deoxyribonucleic acid
DTT	Dithiothreitol
DUB	Deubiquitinating enzymes
ETBF	Enterotoxigenic *Bacteroides fragilis*
EDTA	Ethylenediaminetetraacetic acid
EHEC	Enterohaemorrhagic *Escherichia coli*
EMEM	Minimum essential medium eagle
EPEC	Enteropathogenic *Escherichia coli*
ETEC	Enterotoxigenic *Escherichia coli*
FBS	Foetal bovine serum
FITC	Fluorescein isothiocyante
GFP	Green fluorescent protein
GI	Gastrointestinal
HECT	Homologous to the E6-AP Carboxyl Terminus
HGT	Horizontal gene transfer
HMMPS	High molecular mass polysaccharide
IBD	Inflammatory bowel disease
ICE	Integrative and conjugative elements
KCl	Potassium chloride
LC	Large capsule
LT	Heat labile enterotoxin
LPS	Lipopolysaccharide
MC	Micro-capsule
MDR	Multi-drug resistant
MHC	Major histocompatibility complex
NOD2	Nucleotide-binding oligomerization domain-containing protein 2
NF-κB	Nuclear factor kappa B
OD	Optical density

OMV	Outer membrane vesicles
ORF	Open reading frame
oriT	Origin of transfer
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular pattern
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PRR	Pathogen recognition receptor
PSA	Polysaccharide A
PSB	Polysaccharide B
PSC	Polysaccharide C
PSD	Polysaccharide D
PSE	Polysaccharide E
PSF	Polysaccharide F
PSG	Polysaccharide G
PSH	Polysaccharide H
PSI	Polysaccharide I
PSJ	Polysaccharide J
PUL	Polysaccharide utilisation loci
RAST	Rapid Annotation using Subsystems Technology
RIP1	Receptor interacting protein 1
R-M	Restriction-modification
RNA	Ribonucleic acid
SC	Small capsule
SCFA	Short chain fatty acids
SDS	Sodium dodecyl sulphate
ssDNA	Single-stranded deoxyribonucleic acid
TAE	Tris-acetate EDTA
TCP	Toxin co-regulated pilus
TEM	Transmission electron microscopy
TEMED	Tetramethylethylenediamine
TLR	Toll-like receptor
TMHMM	Transmembrane Hidden Markov Model
TNF-α	Tumor necrosis factor-α
TRAF	TNF receptor associated factor
Treg	T regulatory cell
UBD	Ubiquitin-binding domains
und-P	Undecaprenyl phosphate
VPI	*Vibrio* pathogenicity island
WebACT	Web Artemis Comparison Tool

Abstract

Horizontal gene transfer (HGT) is one of the man driving forces of evolution in prokaryotes, and can also promote within-strain variation of bacterial species. The genomes of three previously sequenced *Bacteroides fragilis* strains, NCTC9343, 638R and YCH46 displayed evidence of extensive HGT, demonstrated by the presence of 28 divergent capsular polysaccharide-associated biosynthesis loci. The genomes of a further four *B. fragilis* strains, LS66, GNAB92, RD48 and BE1 were sequenced and analysed. Genomic comparisons of BE1 and GNAB92 with NCTC9343, 638R and YCH46 identified ten new divergent polysaccharide biosynthesis loci. There is consequently, the potential to express 38 different polysaccharides amongst these five strains. Such a high level of variation in capsular polysaccharides, in so few strains has not been previously observed. HGT has occurred in *B. fragilis* despite the presence of diverse Restriction-Modification systems.

The genome sequences of NCTC9343 and 638R contained a gene, *ubb*, the product of which, BfUbb, has 63% identity to human ubiquitin. The closest DNA sequence homology is to a migratory grasshopper entomopox virus, suggesting acquisition of this gene was via inter-kingdom HGT. The *ubb* gene was also identified in the newly sequenced genomes of *B. fragilis* strains LS66 and RD48. BfUbb had a predicted signal sequence; both full-length and processed forms were detected in whole-cell extracts by Western blot analysis. The inability to detect BfUbb in periplasmic extracts isolated from a *B. fragilis* strain containing an *ubb* signal sequence deletion construct, supported the periplasmic location of the processed form of the protein and the requirement for the signal peptide for transport from the cytoplasm. BfUbb was also detected in concentrated supernatants containing outer membrane vesicles, suggesting a mechanism by which the protein may be delivered to the host. This is the first example of ubiquitin being produced by a prokaryote.

Transduction by bacteriophages is one mechanism by which horizontal gene transfer can occur and can also be a useful tool for genetic manipulation. Fifteen potentially new *B. fragilis*-specific bacteriophages were isolated from filtered sewage and characterised by phage titres and restriction endonuclease cleavage profiles. Of the fifteen, seven phages appeared to be different to the previously identified phage ΦED01. None of the bacteriophages were capable of transduction.

B. fragilis is a predominant member of the gastrointestinal microbiota. To survive within this specific niche, bacteria must successfully compete with other organisms for nutrients and space, and withstand attacks from bacteriophages. HGT may aid in the survival of *B. fragilis* as a commensal.

Lay Abstract

Horizontal gene transfer (HGT) is a method by which pieces of DNA can be exchanged between bacteria. These pieces of DNA can contain genes which are advantageous to an organism, for example, antibiotic resistance genes which can aid in survival. HGT is the major driving force of bacterial evolution and can lead to variation between strains of the same species.

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. Three previously analysed strains of *Bacteroides fragilis*, NCTC9343, YCH46 and 638R displayed an extensive amount of HGT, shown by a high variation of polysaccharides. The three strains have the potential to produce 28 different polysaccharides. The DNA sequence of four further *B. fragilis* strains, BE1, GNAB92, RD48 and LS66, were determined and analysed. Comparisons were made between strains BE1 and GNAB92 and the three previously analysed strains. Ten new different polysaccharides were identified within BE1 and GNAB92. There is consequently, the potential to produce 38 different polysaccharides amongst these five strains. Such a high level of variation in polysaccharides, in so few strains has not been previously observed.

A further occurrence of HGT in *B. fragilis* is the presence of a gene, ubb, which produces a protein, BfUbb that is similar to a protein, ubiquitin, found in abundance in human cells. Ubiquitin is essential for a wide range of cellular functions, including the hosts' immune response to bacterial infection. BfUbb is exported from the bacterial cell, potentially via outer membrane vesicles (OMV), which are small spherical vesicles released by the bacterium. The association of BfUbb with OMV suggests a way in which BfUbb may be delivered to the host cell and therefore potentially interfere with the host ubiquitin pathway and affect the normal functioning of the cell.

B. fragilis is a main member of the human commensal microbiota. To survive in the gut, bacteria must compete with each other for food and space. HGT may help *B. fragilis* survive and thrive as a commensal organism.

Chapter 1. Introduction

1.1. Horizontal Gene Transfer

Horizontal gene transfer (HGT) is one of the main driving forces of evolution in prokaryotes and the mechanism by which bacteria acquire new genetic material from other species, genera or kingdoms. Some horizontally acquired genes will have a detrimental effect on the bacterial recipient, therefore these bacteria will be lost from the population over time. However, horizontally acquired DNA which provides a selective advantage to the organism, such as antibiotic resistance, can spread rapidly through the population (Thomas & Nielsen, 2005). A genomic view of a bacterial species now suggests that bacteria have a core genome, which is conserved within a species, and a larger pan-genome consisting of the core genome and an accessory genome, which is highly variable (Tettelin *et al.*, 2008). The first pan-genome analysis was carried out on eight pathogenic isolates of *Streptococcus agalactiae,* which is a leading cause of death in new-born babies. The authors determined that ~80% of each genome represented the core genome, which was shared by all eight isolates, and ~20% of genes were variable between strains (Tettelin *et al.*, 2005). Similarly the core genome of *Pseudomonas aeruginosa* is highly conserved and accounts for ~90% of the overall genome, with the remaining 10% representing genes absent from some isolates (Kung *et al.*, 2010). Pan-genome studies which have included a larger number of genomes have indicated that the core genome of some species may be much smaller. For example, following a comparison of 61 sequenced *Escherichia coli* genomes it was shown that only 20% of genes belonged to the core genome of the organism, with the other 80% of genes not being found in other *E. coli* isolates analysed and therefore being part of the variable accessory genome (Lukjancenko *et al.*, 2010). The large amount of variation between *E. coli* strains was originally suggested by a comparison of the genome sequence of the enterohaemorrhagic *E. coli* O157:H7 strain, isolated in Sakai, Japan, with that of the laboratory strain K-12. The two strains share a conserved 4.1 Mb backbone sequence but *E. coli* O157:H7 Sakai also contains 1.3 Mb of DNA, ~24% of the total genome, which is not found in *E. coli* K-12 (Ohnishi *et al.*, 2001).

1.2. Mechanisms of HGT

In addition to mutations and deletions within the core genome, HGT is a main contributor to the constituents of bacterial accessory genomes, which contain a number of mobile genetic elements including plasmids, genomic islands and bacteriophages (Jackson *et al.*, 2011). There are three mechanisms of HGT: conjugation, transduction and transformation, which facilitate the generation diversity in bacterial genomes.

1.2.1. Conjugation

The transfer of DNA by conjugation requires direct cell-cell contact between the donor and recipient. Transfer is mediated by the formation of a pilus, which connects the two cells (Figure 1.1a) (Sorensen *et al.*, 2005). It has been well established that plasmids, which are transferred by conjugation, carry a number of genes which can confer a selective advantage to the recipient cell, such as virulence and antibiotic resistance genes (Rankin *et al.*, 2011). Conjugation of plasmids can facilitate the horizontal transfer of large amounts of DNA, which can result in an evolutionary 'jump' in the recipient organism (Jackson *et al.*, 2011). Plasmids are circular, self-replicating, extra-chromosomal elements which usually consist of an essential 'backbone' of genes encoding the replicative functions and conjugative genes and 'accessory' genes which contribute adaptive traits to the host (Norman *et al.*, 2009). Most conjugative plasmids encode a relaxase which nicks the double-stranded DNA at a specific site, known as the origin of transfer (*oriT*) to produce a single-stranded substrate for transfer to the recipient cell (Frost *et al.*, 2005). All known conjugative plasmids from Gram-negative bacteria with the exception of the *Bacteroides* species, encode type IV secretion systems which are associated with the assembly of extracellular filaments, such as pili, which transport DNA from the donor to recipient cell (Sorensen *et al.*, 2005). Conjugative plasmids tend to be larger in size than non-conjugative plasmids due to the extra genes encoding the conjugation machinery. They are also usually found in low copy numbers to reduce the metabolic burden on the host (Norman *et al.*, 2009).

Integrative and conjugative elements (ICEs), which are chromosome-borne mobile genetic elements such as conjugative transposons, are also transferred by conjugation. Under certain conditions ICEs are excised from the host chromosome, following which they circularise, are replicated, and transferred to the recipient cell via conjugation. The ICE in the recipient cell is integrated into the chromosome and the remaining copy of the ICE in

the donor cell reintegrates into the donor cell chromosome. Unlike plasmids, ICEs are incapable of independent replication and therefore cannot exist outside of the chromosome (Wozniak & Waldor, 2010). ICEs are self-transmissible mobile genetic elements and encode their own conjugation machinery in addition to proteins which control their excision from and integration into the host chromosome (Osborn & Boltner, 2002). ICEs encode an integrase (usually designated Int) which enables their integration into the host chromosome and determines the site of insertion. These integrases, which are usually tyrosine recombinases, though serine recombinases have also been identified, promote site-specific recombination between the ICE and a target sequence on the host chromosome (Burrus & Waldor, 2004). ICEs also encode excisionases (Xis), which facilitate the excision of the ICE from the host genome, allowing it to circularise in preparation for transfer to the recipient cell. The conjugative transfer of ICEs is believed to be similar to that of plasmids, with many ICEs identified in Gram-negative bacteria encoding homologues of type IV secretion system proteins, suggesting transfer by a secretion channel such as a pilus. Many identified ICEs also encode putative relaxases and *oriT* sequences, suggesting processing of the DNA for transfer is also similar to conjugative transfer of plasmids (Wozniak & Waldor, 2010). Initial ICEs were identified because they conferred easily identifiable phenotypes within an organism, such as antibiotic resistance. For example, *V. cholera* strains containing the SXT element are resistant to the antibiotics sulphamethoxazole, trimethoprim, chloramphenicol and streptomycin (Beaber *et al.*, 2002). It has also been shown that ICEs can encode a number of other genes which give rise to a selective advantage within the host, such as virulence factors or pathways for the degradation of toxic compounds (Burrus & Waldor, 2004). ICEs can contribute to a large proportion of the variable accessory genome of some bacterial species. For example, two-thirds of the regions of diversity identified by Tettelin *et al.* (2008) in the genomes of eight *S. agalactiae* isolates are ICEs or genomic islands, believed to be defective ICEs which are no longer mobile (Brochet *et al.*, 2008).

1.2.2. Bacteriophages and Transduction

The second mechanism of HGT is transduction (Figure 1.1b), which is mediated by bacteriophages. Phages are host-specific, genomic viruses estimated to outnumber their bacterial hosts ten-fold (Brüssow & Hendrix, 2002). Infection of bacterial cells by bacteriophages can be classed as lytic or lysogenic (temperate). Lytic infection results in the

multiplication of the bacteriophage and lysis of the host cell, whereas lysogenic infection results in the insertion of the phage genome into the host chromosome as a prophage which will be replicated along with the bacterial genome. The bacteriophage lifecycle involves a number of steps (Figure 1.2): adsorption, release of DNA from the protein coat, insertion and replication of phage DNA within the host chromosome, new phage assembly, release from the host cell and transmission (Weinbauer, 2004). The first step of bacteriophage infection is the adsorption of the virion to the surface of the host cell by the recognition of a host-specific receptor by phage receptor binding proteins (Shao & Wang, 2008). A number of bacterial surface components have been identified as receptors for phage adsorption, including outer membrane proteins, flagella, LPS and capsular polysaccharides. LPS in particular are a common phage receptor in Gram-negative bacteria, for example T phage adsorb to LPS on the surface of *Shigella* and *E. coli* (Xu *et al.*, 2013). Following adsorption, the host cell membrane is penetrated by the bacteriophage tail and phage DNA is injected into the cell. Mechanisms of penetration can vary. In the case of T7, proteins ejected from the phage form a channel which crosses the cell membrane allowing DNA to be internalised, however, the channel formed to transfer the lambda genome is formed by host-derived proteins (Molineux, 2001). Following the injection of DNA, the bacteriophage genome is transcribed and translated in two phases, early and late. The early phase involves the expression of proteins to repair the bacterial cell membrane and a phage-specific polymerase which is required for the replication of the bacteriophage DNA. In the late phase structural proteins such as those involved in capsid and tail assembly are expressed, to enable new virions to form. This is followed by lysis of the host cell to release the bacteriophages which disseminate to neighbouring cells and infect (Weinbauer, 2004). Transduction occurs when host cell DNA is accidentally packaged into the new virions instead of bacteriophage DNA and later injected into a new cell (Figure 1.1b). The transduced DNA must then recombine with the host chromosome of the recipient cell. Due to this and the host-specific nature of bacteriophages, transduction usually occurs between bacteria of the same or closely related species (Frost *et al.*, 2005).

Temperate bacteriophages undergo a lysogenic lifecycle in which the phage genome is integrated into the host chromosome as a prophage and is replicated with the host genome (Figure 1.2), with the exception of bacteriophage P1, which exists as a plasmid when undergoing lysogeny. Temperate phage use the same unique integration site on the phage chromosome but do not necessarily integrate their DNA at the same site on the bacterial

chromosome. In *E. coli* for example, lambda phage normally integrates at one chromosomal site but bacteriophage Mu DNA integrates into the host chromosome at random (Casjens, 2003). A stimulus, such as DNA damage can cause the prophage to be excised from the host chromosome. This excised can now act as a lytic phage, leading to the synthesis of new virions and host cell lysis (Canchaya *et al.*, 2003a). However, not all prophage elements encode functional bacteriophages. These defective prophages, or cryptic prophages may still encode some functional genes but are unable to initiate replication of the phage genome, therefore cannot undergo lysis and remain within the host chromosome (Casjens, 2003).

Prophages can contribute to a large amount of the mobile DNA of a host genome and many sequenced bacterial genomes contain multiple prophages, which can represent >10% of the total bacterial DNA in some organisms (Brüssow & Hendrix, 2002). An example of this is the genome of the *E. coli* O157:H7 strain Sakai which contains 18 prophages accounting for 16% of the total genome (Ohnishi *et al.*, 2001). Sequenced *S. pyogenes* strains contain between four and six prophage regions, which amount to 12% of the total genome content (Canchaya *et al.*, 2003b). The acquisition of prophages plays a role in bacterial evolution through the transfer of useful genes to the host, which can increase the fitness of an organism (Brüssow *et al.*, 2004). A major driving force in the emergence of pathogenic isolates is the horizontal transfer and acquisition of virulence factors. It has been demonstrated that prophages can encode virulence factors including toxins, adhesins for host attachment and effector proteins involved in invasion (Boyd & Brüssow, 2002). Since the discovery that the diphtheria toxin of *Corynebacterium diphtheriae* is encoded on the β-phage genome (Freeman, 1951), phage encoded toxins have been identified in both Gram-positive and Gram-negative bacteria, including *Vibrio cholerae*, *Clostridium botulinum*, *S. pyogenes*, *Staphylococcus aureus* and *E. coli* (Eklund *et al.*, 1972; Weeks & Ferretti, 1984; Betley & Mekalanos, 1985; Huang *et al.*, 1987). The expression of prophage-encoded virulence factors is usually regulated by the host chromosome rather than phage-encoded factors. One exception to this is the Shiga toxins (Stx1 and Stx2) of enterohaemorrhagic *E. coli* (EHEC), which are encoded by *stx* genes located on λ-related prophages and expression of which is controlled by transcriptional regulators encoded by the prophages (Wagner *et al.*, 2002). Some bacterial adhesins involved in attachment to host cells have been shown to be prophage-encoded, for example, the toxin co-regulated pilus (TCP) of *V. cholerae*, which is important for colonisation of the host is encoded by the genome of bacteriophage VPIΦ.

This prophage is located adjacent to the *Vibrio* pathogenicity island (VPI), which encodes the cholera toxin, in the bacterial chromosome (Karaolis *et al.*, 1999). Bacteriophages are important mediators of HGT between bacterial species and the presence of prophages can confer inter-strain genetic variability within bacteria of the same species (Brüssow *et al.*, 2004).

1.2.3. Transformation

The third mechanism of HGT is transformation (Figure 1.1c), which is the natural uptake, integration and expression of extracellular DNA by a recipient cell. Unlike conjugation and transduction, acquisition of DNA by natural transformation is initiated by the recipient rather than the donor cell. Cells can be exposed to extracellular DNA that has been passively released by dead bacterial cells which have undergone lysis or via the active release of DNA by organisms such as *Acinetobacter*, *Bacillus* and *Pseudomonas* (Thomas & Nielsen, 2005). The ability of bacteria to bind and take up extracellular DNA is dependent on the cell being naturally competent, a state which involves the expression of specialised competence proteins. With the exception of *Neisseria gonorrhoeae* and *N. meningitidis,* which constitutively express their competence proteins, competence in naturally transformable bacteria is dependent on specific environmental or cellular signals (Solomon & Grossman, 1996). In *Streptococcus pneumoniae* natural competence occurs in early exponential phase of growth and is induced by the binding of the secreted competence-stimulating peptide (CSP) to its membrane-embedded receptor ComD (Johnsborg *et al.*, 2007). Conversely, the competence of *Bacillus subtilis* occurs at the beginning of stationary phase and involves a number of different stimuli, including cell density and nutritional stress (Hamoen *et al.*, 2003). Extracellular DNA binds to the cell surface of competent bacteria and in most Gram-negative bacteria, transport of DNA through the outer membrane involves type IV pili (Chen & Dubnau, 2004). Double-stranded DNA is converted to single-stranded DNA during translocation across the inner membrane to the cytoplasm and can integrate into the host chromosome via homologous recombination. This requires the incoming DNA to contain small regions of homology to the recipient genome (Thomas & Nielsen, 2005).

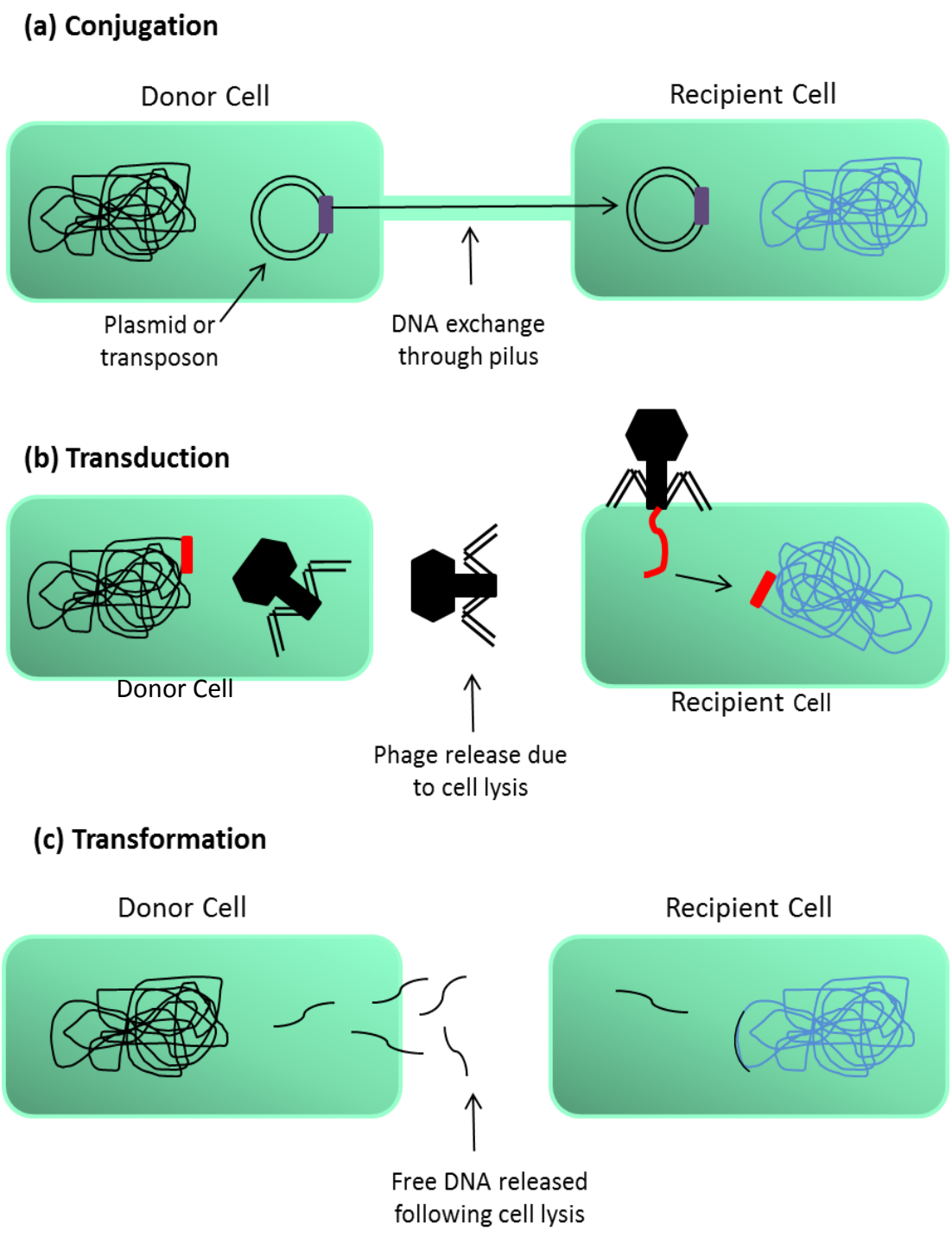


Figure 1.1: Mechanisms of horizontal gene transfer in bacteria

(a) Conjugation occurs by direct contact between the donor and recipient cells via a pilus. Plasmid or ICE DNA is exchanged which can result in acquisition of new genes by the recipient. (b) In transduction, the transfer of genes from donor to recipient cells is mediated by bacteriophages. Transferred genes can integrate into the recipient cell chromosome. (c) Transformation occurs when naked DNA is released following the lysis of the donor cell and can be taken up by a naturally competent bacterial cell.

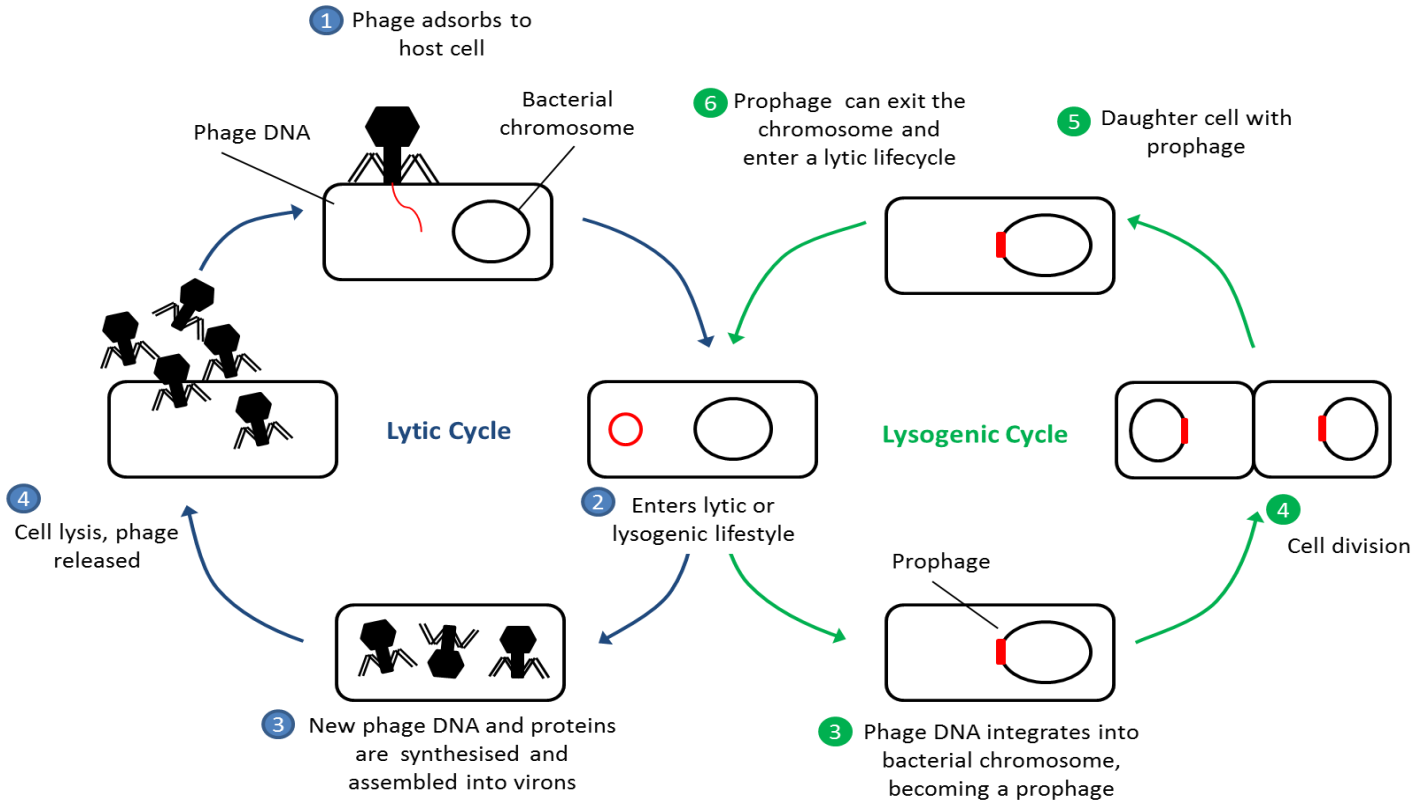


Figure 1.2: The lytic and lysogenic lifecycles of bacteriophages

Bacteriophages adsorb to the surface of the host cell via the recognition of host-specific surface receptors and phage DNA is injected into the cell (1). Lytic bacteriophages hijack the host cell machinery to synthesise and assemble new bacteriophage particles (2, 3). The host cell is lysed and the bacteriophages are released, these new phages disseminate to infect new cells (4). The DNA of lysogenic bacteriophages is integrated into the host chromosome and forms a prophage (3). The phage DNA is replicated along with the host chromosome and is present in each new daughter cell (4). An external stimulus can lead to the excision of the prophage from the host chromosome, at which time it enters a lytic lifecycle (5).

1.3. The microbiota of the human gastrointestinal tract

The dominant members of the 'normal' human gastrointestinal microbiota are the anaerobic bacteria. These organisms outnumber aerobes and facultative anaerobes by 100-1000 fold (Sommer & Backhed, 2013). The GI-tract is colonised by >1000 different bacterial species, but these represent only a few of the known bacterial phyla. The two most abundant of these are Bacteroidetes and Firmicutes, but members of other phyla, such as Proteobacteria and Actinobacteria, are also present (Backhed *et al.*, 2005). Colonisation of the GI-tract begins at birth with organisms from the mother's vagina and skin, and during breast feeding with organisms from milk. The composition of the intestinal microbiota changes with early host development but stabilises between the ages of 1-2 and begins to resemble the adult GI-tract (Sekirov *et al.*, 2010). The 'normal' adult intestine contains ~10^{14} bacterial cells, which is ten times the number of human cells in the body. Their combined bacterial genomes contain more than 5 million genes which provides the intestinal microbiota with a diverse range of gene products to facilitate their interaction with the host (Xu & Gordon, 2003). These commensal organisms perform several essential roles, including the metabolism of indigestible plant polysaccharides and the production of key vitamins, which are not synthesised by the host. Members of intestinal microbiota are also required for the development and regulation of the host immune system, to maintain homeostasis within the GI-tract and to provide protection against colonisation by opportunistic pathogens (Sommer & Backhed, 2013). In turn, the microbiota are provided with a protected, nutrient-rich environment to inhabit, which is maintained at a constant temperature (Troy & Kasper, 2010).

1.3.1. Resistance to pathogens

The intestinal microbiota provide protection against infection by pathogens, as demonstrated in germ-free mice which are more susceptible to infection than conventionally reared mice (Round & Mazmanian, 2009). For example, germ-free animals challenged with *Shigella flexneri* displayed decreased resistance to infection and increased mortality compared to conventionally reared animals (Sprinz *et al.*, 1961). Germ-free mice also demonstrate a reduced clearance of the intracellular pathogen *Listeria monocytogenes*, which has been attributed to a T-cell trafficking defect which normally directs lymphocytes to the site of infection (Zachar & Savage, 1979; Inagaki *et al.*, 1996). Finally, infection with *Salmonella* Typhimurium caused more acute gastroenteritis in germ-

free mice (Round & Mazmanian, 2009). Furthermore, treatment with antibiotics, which disrupts the intestinal microbiota, has been associated with increased colonisation by pathogens in both humans and mice.

A number of mechanisms by which the microbiota can inhibit colonisation by pathogens have been identified. The first is by direct competition for nutrients (Kamada *et al.*, 2013). Several studies have shown that commensal bacteria can successfully out-compete pathogens for nutrients in the GI-tract. For example, *E. coli* out-competes EHEC for amino acids, organic acids and other nutrients, suggesting that commensal *E. coli* is better adapted to utilise nutrients in the gut than the pathogenic strain. *E.coli* can also out-compete the pathogen *Citrobacter rodentium,* which, like *E. coli*, uses monosaccharides as their nutrient source. In contrast, *Bacteroides thetaiotaomicron* and *Bacteroides vulgatus*, which utilise mono and polysaccharides, cannot out-compete *C. rodentium*, suggesting commensals are more successful at out-competing pathogens that are metabolically related and therefore utilise the same nutrient sources (Kamada *et al.*, 2012). This was also demonstrated by the inability of *E. coli* MG1655 to out compete pathogenic *E. coli* strain EDL933. The two strains metabolise a number of the same sugars, however, EDL933 can also utilise sugars which MG1655 cannot. This suggests a mechanism by which pathogenic strains can gain a colonisation advantage in the GI-tract, by consuming sugars which are not primarily metabolised by the commensal microbiota (Fabich *et al.*, 2008).

Secondly, the intestinal microbiota promotes the function of the mucosal barrier, which prevents the attachment of pathogens to the surface of the intestinal epithelium. The epithelium of the colon produces mucus which covers the epithelial cells, facilitates gastrointestinal transport and forms a physical barrier against colonisation by both commensal and pathogenic organisms (Sommer & Backhed, 2013). The intestinal microbiota plays a role in mucus production, as shown by germ-free mice which have a thinner mucus layer and fewer mucus-producing goblet cells than conventionally reared mice. This defect can be corrected by stimulation with bacterial products such as lipopolysaccharide (LPS) and peptidoglycan (Petersson *et al.*, 2011). The gut microbiota also induces the production of antimicrobial peptides by the epithelial cells to protect against invading pathogens (Kamada *et al.*, 2013). These antimicrobial peptides can kill bacteria through enzymatic activity on the cell wall or by disrupting the bacterial cell inner

membrane of Gram-negative bacteria. Several antimicrobial peptides, such as α- and β-defensins, are constitutively expressed by the epithelium. However, the expression of others is induced following the recognition of bacterial signals by the host cell pattern recognition receptors (Hooper & Macpherson, 2010). For example, antimicrobial peptide RegIIIγ, which preferentially acts upon the peptidoglycan of Gram-positive bacteria, is released from Paneth cells in the lumen of the GI-tract following recognition of the intestinal microbiota by Toll-like receptors (TLRs) (Cash *et al.*, 2006). Germ-free mice display minimal levels of REGIIIγ expression and mice with REGIIIγ deficiencies have increased susceptibility to infection by enteric pathogens including *L. monocytogenes* (Brandl *et al.*, 2007). In mono-colonised mice, *B. thetaiotaomicron* can stimulate Paneth cells to produce the antimicrobial peptide angiogenin 4, which is bactericidal to *L. monocytogenes* and *Enterococcus faecalis,* but not to the commensal organisms *E. coli* K12 and *B. thetaiotaomicron* (Hooper *et al.*, 2003).

1.3.2. Development of the host immune system

Germ-free mice studies revealed that the intestinal microbiota is required for the normal development of the host gut-associated lymphoid tissues. These play a role in antigen detection and presentation, lymphocyte functions and the activation of the host immune response (Kamada *et al.*, 2013). Germ-free mice display a number of intestinal immunological defects including reduced antibody production, smaller and fewer Peyer's patches and mesenteric lymph nodes, and fewer $CD4^+$ and $CD8^+$ T-cells, in comparison to conventionally reared mice (Round & Mazmanian, 2009). Additionally, the intestinal epithelial cells, which have a number of immunological functions including cytokine secretion and expression of pattern recognition receptors, display reduced cell turnover and reduced expression of TLRs and class II major histocompatibility complex (MHC II) molecules in germ-free animals (Hooper *et al.*, 2012). Intraepithelial lymphocytes, which are specialised T-cells located between epithelial cells that release cytokines to kill invading cells, are also reduced in number and have a decreased cytotoxicity (Lee & Mazmanian, 2010). Germ-free mice also display extra-intestinal defects in the spleen including reduced numbers of $CD4^+$ T cells, reduced antibody levels and fewer and smaller germinal centres, which are the location of B cell differentiation and proliferation (Mazmanian *et al.*, 2005). When germ-free mice are colonised with members of the intestinal microbiota, these

defects can be corrected, demonstrating that commensal organisms modulate the development of the host immune system (Reading & Kasper, 2011).

1.4. Bacteroides species

Bacteroides species are non-spore-forming, Gram-negative rods and are the predominant anaerobic organisms in the GI-tract, representing ~25% of the organisms identified in faeces (Wexler, 2007). The *Bacteroides* spp. have established themselves as commensal organisms by playing an important role in host nutrition and in the development of the host immune system. The impact of the intestinal microbiota, including *Bacteroides* spp. on nutrition was demonstrated in germ-free mice, which required 30% more calorific intake a day to maintain their body weight than conventionally reared mice (Wostmann *et al.*, 1983). *Bacteroides* spp. can degrade a wide range of ingested plant polysaccharides which are otherwise indigestible by the host. These polysaccharides are broken down and fermented into short chain fatty acids (SCFA), such as acetate, propionate and butyrate which can then be utilised by the host. In humans, approximately 10% of their daily absorbed calories are obtained from these bacterial fermentation products (Hooper *et al.*, 2002). In addition to their nutritional value, SCFA have also been implicated in epithelial cell proliferation and differentiation and stimulating intestinal blood flow (Scheppach, 1994). The best understood mechanism by which commensals degrade polysaccharides is the starch utilisation system (*sus*) of *B. thetaiotaomicron.* Starch is bound to a protein complex present on the bacterial cell surface where it is hydrolysed into smaller oligosaccharides by α-amylases. These smaller molecules are transported via a porin across the outer membrane and into the periplasm, where they are broken down to glucose monomers. The *sus* genes are under the control of a transcriptional activator which responds to the presence of larger oligosaccharides so that they are only expressed when the appropriate nutrients are available. This ensures that nutrient metabolism is an energy efficient process (Flint *et al.*, 2008). In addition to *sus,* the genome of *B. thetaiotaomicron* contains 87 other polysaccharide utilisation loci (PUL) which encode proteins involved in sensing and degrading specific glycans in the GI-tract. This allows the organism to degrade a number of host derived glycans, such as mucin. The ability to utilise these glycans provides the organism with a readily available nutrient source, which is constantly replenished due to epithelial cell turnover. It also confers a selective advantage as the competition for these glycans will be minimal as they are not degraded by all bacterial species (Comstock, 2009).

1.4.1. *Bacteroides fragilis* Polysaccharide A

In addition to polysaccharide metabolism, *Bacteroides* spp., in particular *Bacteroides fragilis* are involved in the maintenance of the host immune system. Germ-free mice display a number of immunological defects which can be corrected following colonisation with the intestinal microbiota (1.3.2), *B. fragilis* polysaccharide A (PSA) has been shown to play an important role in the development of the host immune system. PSA is a zwitterionic polysaccharide, a class of polysaccharides which contain both a positive and negative charge in each of their repeating units, and are presented on MHC class II molecules to antigen-presenting cells (Wexler, 2007; Reading & Kasper, 2011). *B. fragilis* is also able to prevent the pathology of experimental colitis in mice, which has been attributed to PSA. Oral treatment of mice with PSA can protect against symptoms of colitis including weight loss and pro-inflammatory cytokine expression within the gut (Mazmanian *et al.*, 2008). It has been shown that colonisation of germ-free mice with wild type *B. fragilis* resulted in increased numbers of $CD4^+$ T cells in comparison to mice colonised with *B. fragilis* strains which cannot produce PSA. Mono-colonisation of mice with wild type *B. fragilis* has also been shown to mediate the differentiation of $CD4^+$ T cells into Foxp3+ T regulatory (Treg) cells. This process was defective in mice that were mono-colonised with *B. fragilis* strains that did not produce PSA. Furthermore, the defect in $CD4^+$ T cell differentiation could be corrected by the administration of purified PSA alone (Mazmanian *et al.*, 2005; Round & Mazmanian, 2010). The protective effects of PSA in experimental colitis models are attributed to IL-10 production by Treg cells which prevents inflammation. IL-10 deficient Treg cells were unable to confer the protective effects of PSA. Moreover, *B. fragilis* strains lacking PSA failed to induce IL-10 and instead had a pro-inflammatory effect on the host immune system, suggesting that PSA is an essential component in success of *B. fragilis* as a commensal organism (Lee & Mazmanian, 2010). PSA is also involved in splenic structural development. The spleens of mice mono-colonised with wild type *B. fragilis* display normal splenic development consistent with conventionally reared mice. However, mice colonised with a ΔPSA strain of *B. fragilis* showed defects similar to those observed in germ-free mice, including smaller and less defined follicles and depletion of lymphocyte zones (Mazmanian *et al.*, 2005). Shen *et al* (2013) determined that *B. fragilis* outer membrane vesicles (OMV) (see section 1.6) are the delivery mechanism by which PSA interacts with the host. *B. fragilis* OMV are internalised by dendritic cells (DCs) which sense OMV-associated PSA

through TLR2. The authors also demonstrated that treatment of DCs with PSA containing OMV prevented experimental colitis.

1.5. *B. fragilis*: An opportunistic pathogen

B. fragilis accounts for only 1-2% of the human colonic microbiota, however, it is the most commonly isolated Gram-negative bacterium from clinical anaerobic infections (Patrick & Duerden, 2006). *B. fragilis* is an opportunistic pathogen; infection usually occurs when the wall of the gastrointestinal tract has been disrupted or perforated, usually following surgery, and the contents of the GI-tract can enter the sterile peritoneal cavity. Early colonisation is by facultative anaerobes such as *E. coli*, which cause the primary tissue damage and cause a reduction in oxygen tension. This allows obligate anaerobic bacteria to colonise (Wexler, 2007). The range of *B. fragilis* infection includes abscess formation, intra-abdominal sepsis, gynaecological sepsis, peritonitis, soft tissue infection and bacteraemia (Patrick, 2002). During peritonitis, intra-abdominal abscesses are formed, which sequester bacterial cells along with cellular debris and dead polymorphonuclear leukocytes within a fibrous membrane. If untreated, abscesses can expand and cause intestinal obstructions and their rupture can result in dissemination of infection and bacteraemia. *B. fragilis* is the most common cause of anaerobic bacteraemia, with a potential mortality rate of up to ~30% (Cheng *et al.*, 2009).

1.5.1. Virulence factors

1.5.1.1. Capsular polysaccharides

One characteristic of *B. fragilis* which may play a role in its success as an opportunistic pathogen is the expression of within-strain variable surface polysaccharides. Three *B. fragilis* capsular types have been observed (Figure 1.3) (Patrick *et al.*, 1986): the within-strain variable large capsule (LC), which confers resistance to phagocytic uptake and killing by polymorphonuclear leukocytes *in vitro* (Reid & Patrick, 1984); the small capsule (SC), and an electron dense layer or micro capsule (MC) composed of within-strain and phase variable polysaccharides, which have been associated with resistance to complement mediated killing (Reid & Patrick, 1984) and abscess formation (Tzianabos *et al.*, 1993). The genomes of three sequenced *B. fragilis* strains, NCTC9343, 638R and YCH46 each encode ten regions (PSA-J) associated with extracellular polysaccharide biosynthesis, eight of which

(PSA-H), have been implicated in the expression of the phase variable MC (Cerdeno-Tarraga *et al.*, 2005; Patrick *et al.*, 2010).

Extracellular polysaccharides such as lipopolysaccharide (LPS) and capsular polysaccharides have diverse roles in mediating interactions between the bacterium and the environment, in protecting the cell from bacteriophage adsorption and the host immune response and can also be involved in colonisation. LPS is found on the outer membrane of most Gram-negative bacteria. It is comprised of three regions, the lipid anchor (lipid A), a core oligosaccharide and the O-antigen, which is a chain of polysaccharide repeat units of variable length. The nature of LPS in *B. fragilis* has been the subject of much debate, but it has been hypothesised that the polysaccharides of the MC are equivalent to the O-antigen of enteric LPS (Patrick *et al.*, 2009). However, unlike *E. coli* LPS which contains short chain length O-antigen polysaccharides, the *B. fragilis* MC is composed of high molecular mass polysaccharides (Lutton *et al.*, 1991; Patrick *et al.*, 2003). The mechanism by which the *B. fragilis* LC is synthesised is believed to be equivalent to that of *E.coli* group 1 and 4 capsules (Patrick *et al.*, 2009).

In *E. coli*, the polysaccharides required for the formation of the LPS O-antigen and group 1 and 4 capsules are synthesised by a Wzy-dependent pathway (Figure 1.4). On the cytoplasmic face of the inner membrane, sugar phosphates are bound to the carrier lipid undecaprenyl phosphate (und-P) by a glycosyltransferase complex containing WabP (Figure 1.4 A). The und-PP-linked repeat units are transferred across the inner membrane to the periplasmic face by the O-antigen flippase Wzx, and then the polymerase, Wzy assembles the und-PP-linked repeats into polymers (Figure 1.4 B). In the case of capsular polysaccharide biosynthesis, the activity of Wzy is modulated by Wzc, which also interacts with the export channel Wza to transport the polysaccharide to the outer membrane (Figure 1.4 C). In O-antigen biosynthesis, the chain length of the polymer is determined by the O-antigen co-polymerase, Wzz and the polysaccharide is ligated to the lipid A-core by WaaL (Whitfield & Paiment, 2003; Whitfield, 2006b).

In *B. fragilis* NCTC9343, three genes have been implicated in LC production, a putative glycosyltransferase WabP homologue, a putative export protein Wza homologue and a putative co-polymerase Wzc homologue. This suggested that the biosynthesis of *B. fragilis* LC was similar to that of *E. coli* group 1 and 4 capsules (Patrick *et al.*, 2009). The MC-associated polysaccharide biosynthesis loci of NCTC9343, YCH46 and 638R all contain genes

predicted to encode Wzx and Wzy proteins, suggesting that these polysaccharides may be exported to the outer membrane or linked to the lipid A-core (Patrick *et al.*, 2010). A single homologue of the Wzz co-polymerase has been identified in the genome of NCTC9343 and deletion of *wzz* resulted in the absence of MC, suggesting that the MC-associated polysaccharides may be linked to the lipid A-core. However, no homologues of WaaL have been identified in NCTC9343, therefore the mechanism by which polysaccharides could be attached to the lipid A-core is unknown (Patrick *et al.*, 2009).

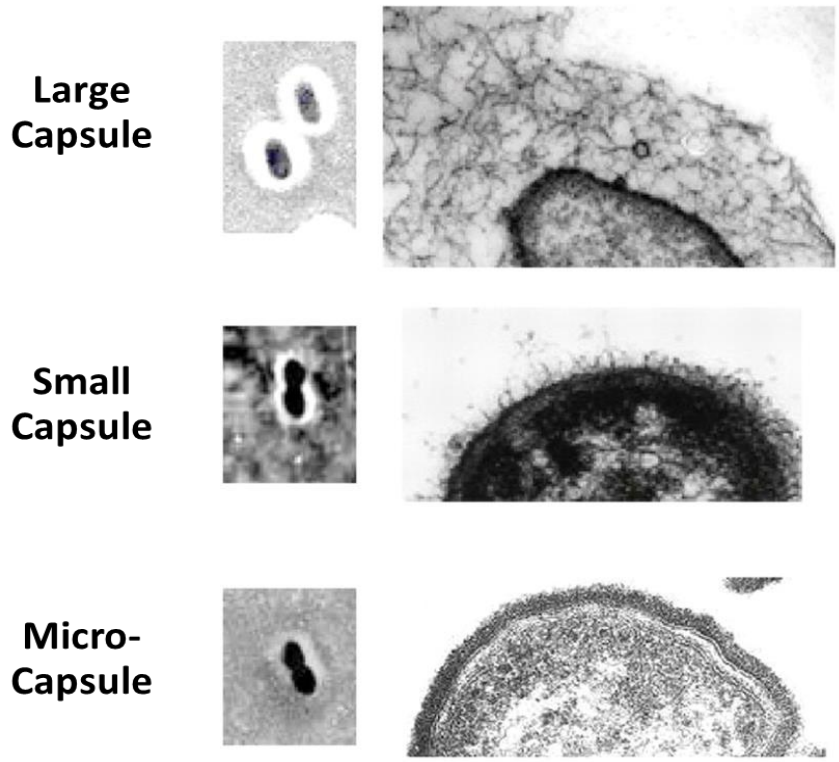


Figure 1.3*: B. fragilis* capsules (taken from Patrick *et al* 2009)

Three *B. fragilis* capsular types have been identified: A within-strain, antigenically variable large capsule (LC), a small capsule (SC) and a variable micro-capsule (MC). *B. fragilis* capsular polysaccharides have been associated with resistance to phagocytic uptake and killing, complement resistance and abscess formation.

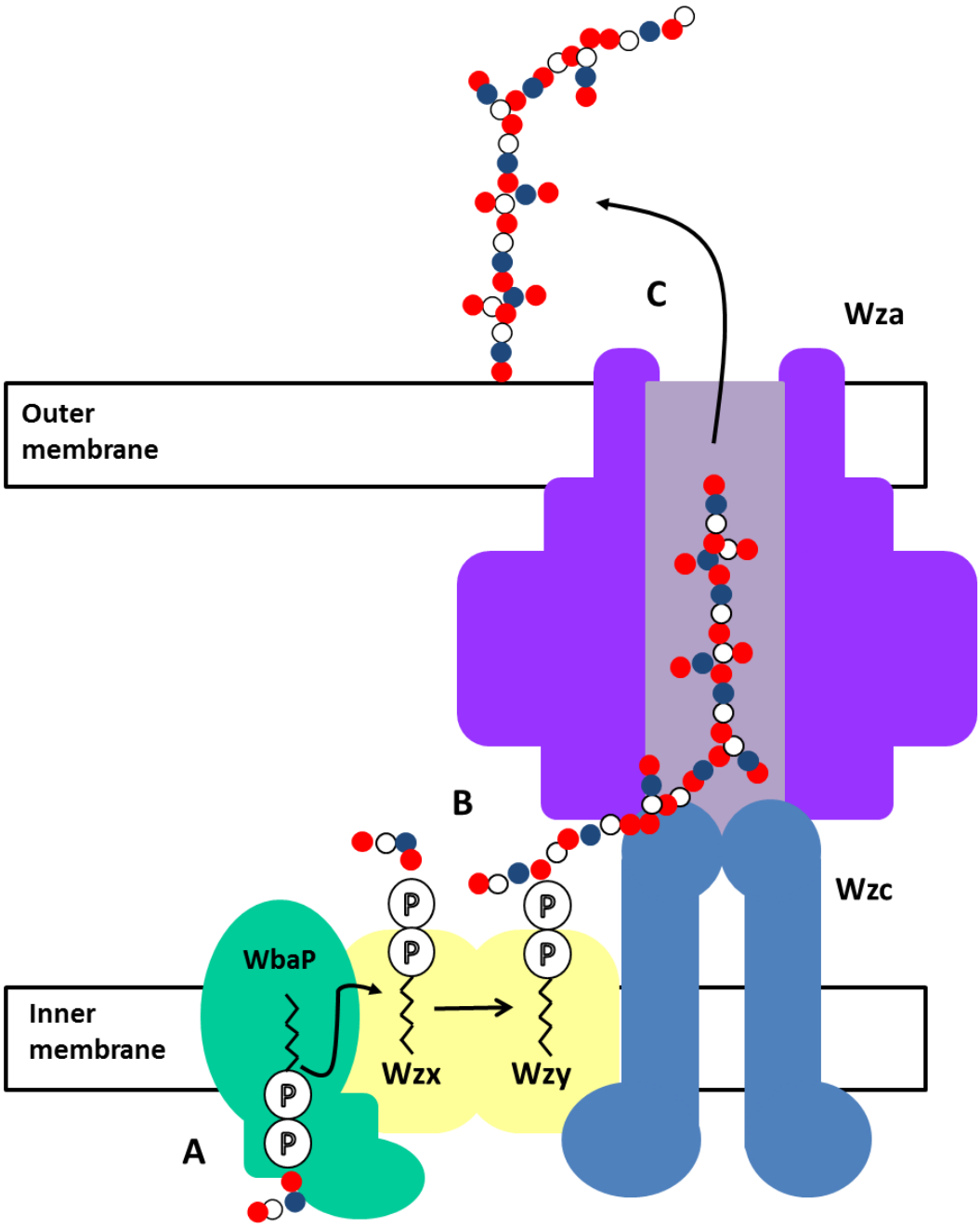


Figure 1.4: Wzy-dependent polysaccharide biosynthesis (adapted from Whitfield, 2006)

E. coli O-antigen polysaccharides and group 1 and 4 capsules are synthesised via a Wzy-dependent pathway. A) First, sugar phosphates are bound to und-PP on the cytoplasmic face of the inner membrane by a glycosyltransferase complex containing WbaP. B) The repeat units are flipped across the inner membrane by the O-antigen flippase, Wzx and the O-antigen polymerase, Wzy assembles the units into polymers. C) In the formation of capsules, polysaccharides are transported beyond the outer membrane via the export channel Wza. O-antigen polysaccharides are attached to the lipid A-core before export (not shown).

1.5.1.2. Enzymes

The destruction of host cell tissues is important during *B. fragilis* infection as these can be utilised by the bacterium as a nutrient source. Several proteases have been implicated in tissue destruction including hyaluronidase, chondroitin sulfatase and hemolysins. These proteases destroy the host extracellular matrix (Wexler, 2007). Hemolysins, which have been identified as virulence factors in a number of Gram-positive and Gram-negative organisms, are cytotoxic proteins which have a haemolytic effect on host cells by causing damage to membranes. The lysis of neutrophils and other host cells by hemolysins can affect the host immune response and also provide a source of nutrients for invading organisms. Two hemolysins, HlyA and HlyB have been characterised in *B. fragilis*. These proteins act in conjunction and have been shown to cause the hemolysis of erythrocytes (Robertson *et al.*, 2006). *Bacteroides* species also produce neuraminidase, encoded by the *nanH* gene, which cleaves mucin polysaccharides to produce glucose which can be utilised by the bacterium (Berg *et al.*, 1983). *B. fragilis* strains NCTC9343, 638R and YCH46 have been shown to bind human fibrinogen, a major component in abscess formation. These strain also display fibrinogenolytic activity, which might allow *B. fragilis* to slow down or prevent abscess formation or facilitate abscess degradation, allowing infection to disseminate, potentially resulting in bacteraemia (Houston *et al.*, 2010).

1.5.1.3. Enterotoxin

Enterotoxigenic *B. fragilis* (ETBF) strains encode a ~20 kDa zinc metalloprotease protein toxin termed *B. fragilis* enterotoxin (BFT). These strains have been associated with an acute form of disease, characterised by diarrhoea and loss of appetite in children, adults and livestock (Franco *et al.*, 1999). Studies indicate that BFT may destroy zonula adherens tight junctions, which are cell-cell junctions in the intestinal epithelium linked to the actin cytoskeleton. BFT cleaves zonula adherens protein E-cadherin which results in the rearrangement of the actin cytoskeleton and loss of tight junctions. This in turn results in leakage in the epithelium and diarrhoea (Wu *et al.*, 1998). The enterotoxin gene has been identified in strains isolated in a number of different countries, including USA, Japan and Italy. BFT was also detected in 18% of clinical isolates tested from the UK, Poland, France and Holland (Luczak *et al.*, 2001). In one study, ETBF strains were associated with patients suffering from inflammatory bowel disease (Prindiville *et al.*, 2000).

1.6. Bacterial outer membrane vesicles

Outer membrane vesicles (OMV) are naturally secreted from Gram-negative bacteria during all stages of growth. OMV form when a portion of the bacterial outer membrane encapsulates periplasmic content and 'blebs' from the cell, however, the mechanism by which this occurs is still unknown. These vesicles are ~50-250 nm in diameter, depending on the strain, and act as a secretion and delivery system which can disseminate bacterial products and are a mechanism by which bacteria can interact with their environment and the host (Kulp & Kuehn, 2010).

The rate of vesicle production is not consistent and can be altered by environmental factors and cellular stress. External conditions which upregulate the σ^E envelope stress response have been shown to increase vesiculation in *E. coli* and appear to play a role in cell survival during envelope stress. *E. coli* mutants that over-produce vesicles exhibit enhanced survival following extracellular insults in comparison to OMV under-producing and wild type strains. OMV allow bacterial cells to dispose of toxic products such as mis-folded proteins, therefore reducing membrane stress (McBroom & Kuehn, 2007). Antibiotics can also influence vesicle production. Treatment of *Shigella dysenteriae* with mitomycin C induces an increase in vesicle production, an increase in OMV size and in OMV toxicity (Dutta *et al.*, 2004). Treatment of *P. aeruginosa* with gentamicin also increases vesiculation and OMV size, while changes in the composition of the OMV were also observed. These antibiotic induced vesicles contained inner membrane and cytosolic components in addition to outer membrane and periplasmic proteins (Kadurugamuwa & Beveridge, 1995). OMV produced by some organisms can confer a survival advantage by having a bactericidal effect on other bacteria, therefore reducing competition. This was demonstrated by protease- and toxin-containing OMV from *E.coli*, *Shigella* and *Pseudomonas* which lyse other Gram-positive and Gram-negative bacteria (Kuehn & Kesty, 2005). OMV may also act as a defence mechanism to bacteriophage infection. Manning and Kuehn (2011) demonstrated that bacteriophage T4 can adsorb to OMV isolated from *E. coli* and that T4 infectivity was reduced following incubation with vesicles. This suggests OMV may act as a decoy system to prevent the adsorption of bacteriophages to bacterial cells.

OMV play a well-established role in the delivery of bacterial products, such as toxins and other virulence-associated bacterial products to the host. Experimental studies have indicated that the interaction of OMV with host cells can occur in one of two ways. The first is via fusion of the OMV with the eukaryotic cell plasma membrane and delivery of their contents directly into the cytosol. This has been demonstrated by the detection of fluorescently labelled components of the outer membrane of OMV on the surface of host cell membrane (Galka *et al.*, 2008). Bomberger *et al* (2009) proposed that OMV deliver their bacterial proteins by fusing with host cell lipid raft microdomains. The authors showed that disruption of lipid rafts with the cholesterol binding agent Filipin III eliminated the fusion of *P. aeruginosa* OMV with airway epithelial cells. A similar study by Kesty *et al* (2004) determined that Filipin III also interfered with the ability of *E. coli* OMV to interact with host cells. The second mechanism by which OMV interact with the host cells is by adherence and internalisation via a receptor-mediated endocytic pathway (Ellis & Kuehn, 2010). Toxins can act as adhesins for OMV, for example enterotoxigenic *E. coli* (ETEC) produces heat labile enterotoxin (LT) which is associated with the bacterial LPS and has been identified as a surface component of OMV. ETEC OMV interact with epithelial cell membranes and binding occurs between LT and its lipid raft-associated receptor. OMV are then internalised via cholesterol-mediated endocytosis. Following entry into the cell, LT is trafficked to the endoplasmic reticulum and the cytoplasm, resulting in enterotoxicity (Kesty *et al.*, 2004). Similarly, uptake of *Helicobacter pylori* OMV by gastric epithelial cells also involves interactions with the lipid rafts of host cells, however, cholesterol is not required for this process, and instead internalisation is via clathrin-mediated endocytosis. Vesicle-associated *H. pylori* vacuolating toxin (VacA) binds to a number of host cell receptors, including lipid rafts and has been shown to increase the uptake of OMV by epithelial cells, suggesting it plays a role in vesicle internalisation (Parker *et al.*, 2010).

The delivery of virulence factors to the host via OMV can play an important role in bacterial pathogenesis. For example, mucociliary clearance in the lung is driven by the cystic fibrosis transmembrane conductance regulator (CFTR) secretory chloride channels located in the plasma membranes of airway epithelial cells (Swiatecka-Urban *et al.*, 2006). These channels are constitutively cycled in and out of the cell membrane through endosomal compartments. CFTR is ubiquitylated and incorporated into an early endosome, deubiquitylating enzyme USP10 deubiquitylates CFTR and the channel is returned to the membrane. However, *P. aeruginosa* toxin Cif, which is delivered to host cells via OMV

inhibits the deubiquitylating activity of USP10 which leads to the lysosomal degradation of CFTR. This results in a decreased number of CFTR channels in the plasma membrane, which in turn leads to a reduced mucociliary clearance of bacterial infection and compromises the immune defences of the lung (MacEachran *et al.*, 2007; Bomberger *et al.*, 2009; Bomberger *et al.*, 2011b).

As OMV contain many of the surface antigens present on bacteria, they have the potential to stimulate or modulate the host cell response to bacterial infection. OMV have been shown to activate the host inflammatory response, for example, OMV produced by *Salmonella* Typhimurium induce dendritic cell maturation, activate macrophages and increase the expression of MHC class II molecules, and they also stimulate pro-inflammatory cytokine secretion and $CD4^+$ T cell activation (Alaniz *et al.*, 2007). A pro-inflammatory response to OMV has also been observed for other pathogens, including *H. pylori* and *P. aeruginosa,* which both induce the expression of IL-8 (Ellis & Kuehn, 2010). Vesicles may also have an anti-inflammatory effect on the host immune system to allow the organism to escape detection during colonisation. Gingipain proteases on the surface of *Porphyromonas gingivalis* degrade the LPS receptor, CD14, of human macrophages. This could result in a reduced immune reaction in response to *P. gingivalis* colonisation (Duncan *et al.*, 2004). *Moraxella catarrhalis* OMV contain the surface protein UspA1 which has been shown to have an inhibitory effect on the inflammatory response of epithelial cells (Schaar *et al.*, 2011).

B. fragilis OMV play an important role in the mutualistic relationship between the organism and the host because OMV-mediated delivery of *B. fragilis* PSA dendritic cells promotes host immune development and protects against experimental colitis (see section 1.4.1) (Shen *et al.*, 2012). However, *B. fragilis* OMV have also been shown to contain degradative enzymes and can agglutinate erythrocytes, indicating that vesicles may also be involved in virulence (Patrick *et al.*, 1996).

1.7. Aims

The main aim of this thesis was to identify instances of horizontal gene transfer within the genome of *B. fragilis* and to consider their potential contribution to the success of the organism as both a commensal and an opportunistic pathogen.

- The genomes of three previously sequenced *B. fragilis* strains, NCTC9343, YCH46 and 638R displayed extensive horizontal gene transfer. Whole genome sequencing of four further *B. fragilis* strains was performed and comparisons were made with YCH46, 638R and NCTC9343 to determine the within-strain variation of polysaccharide biosynthesis loci and diversity of R-M systems.

- The genomes of NCTC9343 and 638R contained a gene, *ubb*, which had 63% identity to human ubiquitin. The closest DNA sequence homology was to a Migratory Grasshopper entomopox virus, suggesting acquisition by inter-kingdom HGT. The aim of this study was to determine if BfUbb was expressed by *B. fragilis* and to explore the potential of the protein to interact with the eukaryotic host.

- Finally, *B. fragilis*-specific bacteriophages were isolated, characterised and assessed for their ability to mediate transduction, with the aim of identify new tools to aid in the genetic manipulation of *B. fragilis.*

2. Materials and Methods

2.1. Bacterial culture conditions

B. fragilis strains were grown in a Don Whitley Scientific (UK) MiniMacs anaerobic work station 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). *E. coli* strains were grown in Luria-Bertani Broth (LB) (Difco, USA) shaken at 200 rpm, or on LB agar at 37 °C. Media were supplemented with appropriate antibiotics (Table 2.1) where stated. Strains were stored in a final concentration of 25% glycerol at -80°C.

2.2. Antibiotics

The concentration and preparation medium of antibiotics which were used in this study are shown in Table 2.1.

Table 2.1: List of Antibiotics

Antibiotic	Stock concentration	Working concentration	Preparation
Ampicillin	100mg ml^{-1}	100μg ml^{-1}	In sterile H_2O
Erythromycin	10mg ml^{-1}	10μg ml^{-1}	In ethanol
Gentamicin sulphate	50mg ml^{-1}	50μg ml^{-1}	In sterile H_2O
Metronidazole	5mg ml^{-1}	5μg ml^{-1}	In sterile H_2O
Rifampicin	50mg ml^{-1}	50μg ml^{-1}	In methanol
Tetracycline hydrochloride	10mg ml^{-1}	10μg ml^{-1}	In 50% ethanol

2.3. Strains

The stains used in this study and their source are shown in Table 2.2.

Table 2.2: List of strains

Strain	Genotype	Source
DH5α	F⁻ ϕ80*lacZ*ΔM15Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ⁻	Invitrogen
S17-1 (λpir)	*rec*A *thi pro hsd*RM+ RP4::2-Tc::mu::Km Tn 7 λ pir	Simon *et al* 1983
XL-1 Blue MRF'	Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr* 173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F´ *proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^r)	Stratagene, Aligent Technologies
Δ*ubb*	NCTC9343 *Δubb::ermF*	Patrick *et al* 2011
Δ*wzz*	NCTC9343 *Δwzz::ermF*	Patrick *et al* 2009
Δ*upcy*	NCTC9343 *Δupcy*::*ermF*	Patrick *et al* 2009
638R	*B. fragilis* clinical isolate	C.J. Smith
3259	*B. fragilis* clinical isolate	Sheila Patrick
BCH1	*B. fragilis* clinical isolate	Sheila Patrick
BE1	*B. fragilis* clinical isolate	Sheila Patrick
BE3	*B. fragilis* clinical isolate	Sheila Patrick
GNAB4	*B. fragilis* clinical isolate	Sheila Patrick
GNAB92	*B. fragilis* clinical isolate	Sheila Patrick
JC6	*B. fragilis* clinical isolate	Sheila Patrick
JC15	*B. fragilis* clinical isolate	Sheila Patrick
JC17	*B. fragilis* clinical isolate	Sheila Patrick
JC19	*B. fragilis* clinical isolate	Sheila Patrick
LS27	*B. fragilis* clinical isolate	Sheila Patrick
LS54	*B. fragilis* clinical isolate	Sheila Patrick
LS66	*B. fragilis* clinical isolate	Sheila Patrick
MPRL3499VI	*B. fragilis* clinical isolate	Ian R. Poxton
MPRL3535V	*B. fragilis* clinical isolate	Ian R. Poxton
MPRL3536X	*B. fragilis* clinical isolate	Ian R. Poxton
NCTC9343	*B. fragilis* clinical isolate	Sheila Patrick
NCTC9344	*B. fragilis* clinical isolate	Sheila Patrick
NCTC10584	*B. fragilis* clinical isolate	Sheila Patrick
RD47	*B. fragilis* clinical isolate	Sheila Patrick
RD48	*B. fragilis* clinical isolate	Sheila Patrick
VH	*B. fragilis* clinical isolate	Sheila Patrick
WAL610	*B. fragilis* clinical isolate	Hannah Wexler
YCH46	*B. fragilis* clinical isolate	Sheila Patrick
LS84	*B. thetaiotaomicron* clinical isolate	Sheila Patrick

2.4. Oligonucleotides

PCR was carried out as described in 2.7.3 and the oligonucleotides used in this study are hown in Table 2.3

Table 2.3: List of oligonucleotides

Oligonucleotide	Length (bp)	Sequence (5'-3')
GyrB_full_FOR	21	TTGGTGGGATTCTGTTCTTCG
GyrB_full_REV	21	TAGCTATAAACGTATGCAACG
GyrB3_FOR	19	CCACACAGGAACCAAACAG
GyrB3_REV	20	TATTGTACATGATAGCCACC
ubb_FOR	21	CACACCTTTTTCAACAGACCC
ubb_REV	22	CTTAGTGTTAACAGTCGACATC
Pubb_FOR_1	36	ATTCGAGCTCGGTACCACACCTTTTTCAACAGACCC
ubb_del-sig_FOR	27	CAAGTTTTTATAAAAAACAGATATGGC
ubb_puc19_REV	37	CTAGAGGATCCCCGGCTTAGTGTTAACAGTCGACATC
Pubb_REV2	50	CTGTTTTTTATAAAAACTTGCATTTGTTTTTATATTTAAAAATAA TATGG
GFP_FOR	38	CTACTGTAAATGCAATGCGTAAAGGAGAAGAACTTTTC
GFP_REV	42	GTTTTTTATAAAAACTTGCATTTTGTATAGTTCATCCATGCC
ss_REV	42	GTTCTTCTCCTTTACGCATTGCATTTACAGTAGATGGCAATG
ubb_FOR	38	GATGAACTATACAAAATGCAAGTTTTTATAAAAAACAG
tetQ_FOR	21	GAATTATCTCCTTAACGTACG
tetQ_REV	21	CGGTTCAAGATTACGAAGATA
BF3878_FOR	20	ATACGCAAATCAACAGAACG
mutL_REV	20	GCCTTAACTTTACATGGAAT
Phi_RecET_FOR	52	TAATTTATTTATTTAATTAAAACGAGTATGATTAAAAATTTGCC TAACATTC
Phi_RecET_REV3	36	GCTTTTAAGCTTTTTGCTAAATCTTCATTTATAACC
PrpsF_FOR2	32	GCTTTTGGTACCCAACTTTGCAAAGATAGATG
PrpsF_REV2	27	ACTCGTTTTAATTAAATAAATAAATTA
finB_FOR	27	GGCATATGGAGATAATAGGCTACGCTC
finB_REV	27	AAGGATCCGACTTGAAAGGTAACGCCC

2.5. Plasmids

The plasmids used in this study are shown in Table 2.4.

Table 2.4: List of plasmids

Plasmid	Description	Source
pLYL01	Shuttle vector, Amp^R Mob+ (Tet^R only in *Bacteroides*)	Li *et al*., 1995
pGFPmut3.1	Shuttle vector, AmpR *gfpmut3.1*	Clontech
pKJ01	pLYL01 with *ubb* signal sequence deletion cassette in KpnI	This study
pKJ02	pLYL01 with *ubb* in KpnI	This study
pKJ03	pLYL01 with *gfpmut3.1-ubb* fusion cassette in KpnI	This study
pKJ04	pLYL01 with P*rpsf, phirecE* and *phirecT* in KpnI/HindIII	This study

2.6. Bacterial techniques

2.6.1. *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 1:40 dilution was made into fresh pre-reduced BHI-S broth or DM without antibiotic selection and incubated anaerobically at 37°C. Optical density readings were taken at 30 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 OD_{600} reading 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 triplicate.

2.6.2. Preparation of chemically competent cells

An overnight culture of *E. coli* was diluted 1:100 into 100 ml, pre-warmed LB and grown until an OD_{600} of 0.4 was reached. Cells were incubated on ice for 20 min prior to centrifugation at 6000 x g at 4°C for 10 min. The supernatant was discarded and cells were re-suspended in 10 ml ice-cold 100 mM $CaCl_2$ and incubated on ice for 2 h. Cells were harvested by centrifugation as previously described and re-suspended in 2 ml ice-cold 100 mM $CaCl_2$ with 20% (v/v) glycerol. 100 µl aliquots were stored at -80°C.

2.6.3. Transformation – chemically competent cells

For transformation, 5 – 10 µl of DNA were added to 50 µl of competent cells and incubated on ice for 30 min. Cells were heat shocked at 42°C for 90 seconds followed by incubation on ice for 5 min. 500 µl pre-warmed LB was added and cells were incubated at 37°C for 2 h, with the exception of transformations for which the antibiotic used for selection was ampicillin. These reactions required no further incubation time. Cells were spread onto LB plates containing the antibiotic of selection and incubated overnight at 37°C.

2.6.4. Transformation - Electroporation

A 1:40 dilution of *B. fragilis* overnight culture was made into 10 ml BHI-S and grown to an OD_{600} of 0.4. Cells were harvested by centrifugation at 4000 x g for 15 min at 4°C, then washed (5 x 15 min) in 10 ml pre-reduced ice cold dH_2O by centrifugation as previously described. Cells were re-suspended in 100 µl of ice cold dH_2O, 40µl of cells were placed into a pre-chilled electroporation cuvette (Bio-Rad Gene Pulser® Cuvette 0.1cm gap electrode, Bio-Rad, UK), to which 0.5-1.5 µl of DNA were added. Cells were electroporated in a Bio-Rad MicroPulser™ with a voltage of 2 kV. 1 ml pre-reduced, pre-warmed BHI-S was added to cells followed by anaerobic incubation for 1-3 h, depending on the antibiotic used for selection and 100 µl of cells were spread-plated onto BHI-S plates containing the appropriate antibiotics and incubated anaerobically (Patrick *et al.*, 2009).

2.6.5. Conjugation of *E. coli* S-17 with *B. fragilis*

Overnight cultures of *E. coli* and *B. fragilis* were grown with the appropriate antibiotic in LB and BHI-S, respectively. A 1:100 dilution of *E. coli* was made into LB and a 1:40 dilution of *B. fragilis* was made into BHI-S without antibiotic selection. Both strains were grown to an OD_{600} of 0.6 and mixed donor to recipient in a ratio of 1:10. 200 µl were pipetted onto a BHI-S plate and incubated aerobically at 37°C for 2 h, followed by anaerobic incubation overnight. Cells were washed from the plate using 1 ml BHI-S and pelleted by centrifugation at 16000 x g for 5 min. Pellets were re-suspended in 200 µl BHI-S, 100 µl aliquots were plated onto BHI-S containing gentamicin and the appropriate antibiotic and incubated anaerobically.

2.7. DNA techniques

2.7.1. Genomic DNA extraction

Genomic DNA was extracted from 1 ml of overnight culture using the Wizard Genomic DNA Purification Kit (Promega, UK) according to manufacturer's instructions. Pellets were re-suspended in 100 µl nuclease free water, overnight at 4°C and DNA was stored at -20°C.

2.7.2. Plasmid DNA purification

Plasmid DNA was extracted using the QIAprep Spin Mini prep Kit or the QIAfilter Plasmid Midi Kit (Qiagen, UK), according to manufacturer's instructions. Using the Mini prep kit, 1ml of overnight culture was used and DNA was eluted with 30-50 µl of elution buffer (10 mM Tris-Cl, pH 8.5). If the Midi kit was used, 50 ml of overnight culture was used and DNA was eluted with 500 µl of elution buffer (10 mM Tris-Cl, pH 8.5), overnight at 4°C. DNA was stored at -20°C.

2.7.3. Polymerase chain reaction (PCR)

Oligonucleotides were synthesised by Eurofins MWG Operon (Germany) and the sequences are shown (Table 2.4). PCR reactions were performed in a Techne Progene Thermal Cycler (Bibby Scientific Limited, UK) or an Eppendorf Mastercycler gradient cycler. 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 (NEB, UK), 1 µl ~0.1 ng/µl template DNA and 35.5 µl dH_2O. Unless otherwise stated, the thermal program was 1 cycle at 98°C for 1 min, 30 cycles at 98°C for 10 seconds, x°C for 20 seconds, 72°C for 30 seconds per kb, followed by a final extension of 72°C for 10 min, where x represents the annealing temperature of each primer pair.

2.7.4. Agarose gel electrophoresis

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 dH_2O for 10 min. Gels were visualised using a UVP VisiDoc-It Imaging system (UVP LLC, USA).

1xTAE buffer: 40 mM Tris, 1 mM EDTA pH8, 19 mM Glacial acetic acid

2.7.5. DNA purification

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

2.7.6. 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 or by a Qubit® 2.0 Fluorometer (Invitrogen, USA) according to manufacturer's instructions.

2.7.7. Gel Extraction

Gel extraction was performed using the QIAquick gel extraction kit (Qiagen, UK) following manufacturer's instructions.

2.7.8. Restriction endonuclease digestion

DNA was digested with restriction enzymes and their appropriate buffer as outlined in the manufacturer's instructions (New England Biolabs, UK). A standard 20 µl digestion contained: 1 µl restriction enzyme, 2 µl 10x reaction buffer, 5 µl DNA and 12 µl dH_2O. BSA was added where appropriate at a final concentration of 100 µg/ml. Digestions were incubated for 1 h at 37°C unless otherwise stated.

2.7.9. Dephosphorylation of DNA

Following restriction endonuclease digestion, DNA was dephosphorylated to prevent re-ligation. To the 20 µl digestion reaction, 1 µl of Antarctic Phosphatase (NEB, UK) and 5 µl Antarctic Phosphatase 1x buffer were added and the sample incubated at 37°C for 15 min. The enzyme was heat inactivated by incubation at 65°C for 5 min and enzymes and buffers were removed using the Qiagen PCR purification kit following manufacturer's instructions.

2.7.10. Ligation of DNA

A standard DNA ligation was carried out in a 10 µl reaction volume containing, 1 µl 10x reaction buffer, 0.5 µl T4 DNA ligase (Promega, UK), Varying ratios of insert to vector were used (1:1, 1:2, 1:3, 2:1, 2:2 and 2:3) and dH_2O. Reactions were incubated at 16°C for 2 h or overnight.

2.7.11. Plasmid construction using In-fusion HD Cloning kit

Plasmids were constructed using the In-Fusion HD Cloning kit (Clontech), in which PCR-generated DNA fragments were fused to linearised plasmid DNA by homologous recombination. In all In-Fusion reactions, PCR amplicons were inserted into the KpnI site of plasmid pLYL01. Oligonucleotides were designed to generate PCR products with 15 bp of homology to the KpnI site of pLYL01 and the plasmid backbone. DNA insert components were fused and amplified via crossover PCR before addition to the In-fusion cloning reaction.

Plasmid DNA was digested in a 50 µl reaction containing: 2.5 µl restriction enzyme, 6 µl 10x reaction buffer, 30 µl DNA and 11.5 µl dH_2O. Digests were incubated at 37°C, overnight. Plasmid DNA and PCR products were purified using the Wizard SV PCR Clean-up System (Promega, UK) following manufacturer's instructions and stored at -20°C.

PCR amplicons and linearised plasmid were fused in the In-Fusion cloning reaction. A standard 10 µl reaction contained: 2 µl 5X In-Fusion HD Enzyme Premix, 100 ng linearised vector, 100 ng insert DNA and x µl dH_2O, where x was the volume of dH_2O required to adjust the total reaction volume to 10 µl. Reactions were incubated at 50°C for 15 min, then placed on ice.

Chemically competent *E. coli* S-17 cells (see section 2.6.2) were transformed using 2.5 µl of cloning reaction (see section 2.6.3). Cells were spread onto LB plates containing ampicillin and incubated aerobically overnight.

2.7.12. Whole genome sequencing

2.7.12.1. Sample preparation

B. fragilis genomic DNA was isolated (see section 2.7.1) and the DNA pellet was rehydrated in 100 µl nuclease free water for 30 min at 65°C. RNA was removed by the addition of 2 µl

Riboshredder (Illumina, USA) for 2 h at 37°C. DNA was cleaned and purified using the Genomic DNA Clean and Concentrator kit (Zymo Research, USA), according to manufacturer's instructions. DNA was quantified using the Qubit (see section 2.7.7) and diluted in ddH_2O to a concentration of 0.2ng/µl.

2.7.12.2. Library preparation

DNA sequencing libraries were prepared using the NextERA Tagmentation Illumina sequencing library kit (Illumina, USA), according to manufacturer's instructions. Illumina sequencing barcodes were added to each sample using PCR, 5 µl of barcode primer 1 and 5 µl of barcode primer 2 were added to the Tagmentation product. The thermal program was 1 cycle at 72°C for 3 min, 1 cycle at 95°C for 30 seconds, 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds followed by a final extension of 72°C for 5 min. Illumina Miseq whole genome sequencing was carried out by GenePool Sequencing Service (Edinburgh, UK).

2.8. Protein techniques

2.8.1. Solutions

Tris-glycine SDS-PAGE

4x Resolving Buffer: 1.5 M Tris-HCl, pH 8.8
4x Stacking Buffer: 0.5 M Tris-HCl, pH 6.8

10 x Running Buffer: 0.25 M Tris base, 1.9 M glycine

1x Running Buffer: 1/10 dilution of 10 x Running buffer with 0.1% SDS

5% Stacking gel: 1.5 ml 30% bis-acrylamide, 2.25 ml 4x stacking buffer, 5 ml dH_2O, 90 µl 10% SDS, 100 µl ammonium persulphate (APS), 8 µl Tetramethylethylenediamine (TEMED).

12% Resolving gel: 4.8 ml 30% bis-acrylamide, 3 ml 4x resolving buffer, 4 ml dH_2O, 120 µl 10% SDS, 100 µl APS, 8 µl TEMED.

Tris-tricine SDS-PAGE

Gel Buffer: 3 M Tris-Cl, 0.3% SDS, pH 8.45

1x running buffer: 0.1 M Tris, 0.1 M Tricine, 0.1% SDS

Stacking gel: 0.8 ml 30% bis-acrylamide, 1.55 ml gel buffer, 3.9 ml dH_2O, 100 µl APS, 8 µl TEMED.

Resolving gel: 4.9 ml 30% bis-acrylamide, 5 ml gel buffer, 3.5 ml dH_2O, 1.6 ml glycerol, 100 µl APS, 8 µl TEMED.

1x Transfer buffer: 25 mM Tris, 192 mM Glycine

PBS Tween (PBS-T): PBS containing 0.05% (w/v) Tween 20

Coomassie Stain: 45% (v/v) dH_2O, 45% (v/v) Methanol, 10% (v/v) Glacial acetic acid, 0.125% (w/v) Coomassie brilliant blue

Coomassie de-stain: 45% (v/v) dH_2O, 45% (v/v) Methanol, 10% (v/v) Glacial acetic acid

2.8.2. Preparation of whole cell protein extracts

B. fragilis strains were grown overnight in glucose defined medium. Cells were pelleted by centrifugation at 16 000 x g for 5 min and re-suspended in 500 µl 3x SDS Sample Buffer (NEB). Samples were stored at -20°C.

2.8.3. Preparation of periplasmic protein extracts

B. fragilis strains were grown overnight in glucose defined medium. Cells were pelleted by centrifugation at 16 000 x g for 2 min at 4°C and the pellet was washed by centrifugation as above in 1 ml PBS. Pellets were re-suspended in 150 µl cold 20% sucrose/10 mM Tris-Cl (pH 7.5) and vortexed before the addition of 5µl 0.5 M EDTA (pH 8). Samples were incubated on ice for 10 min and centrifuged at 16 000 x g for 5 min at 4°C, pellets were re-suspended in 100 µl ice-cold water and incubated on ice for 10 min (Quan *et al.*, 2013). The supernatant containing the periplasmic fraction was removed and stored at -20°C.

2.8.4. Sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were resolved by SDS-PAGE using Tris-glycine or Tris-tricine polyacrylamide gels, made following the recipes shown in 2.8.1. Resolving protein gels were prepared using a mini protean gel casting apparatus (Bio-Rad, UK) and sealed using 250 µl isopropanol (Sigma-Aldrich, UK). Stacking gels were prepared and added to the polymerised resolving gel and a comb was inserted. Once polymerised, the protein gel was assembled in a gel tank and covered with 1x running buffer. Protein samples were boiled in 1x Protein sample buffer (New England Biolabs, UK) for 10 min at 98°C before loading onto the gel. Where indicated, the reducing agent dithiothreitol (DTT) (New England Biolabs, UK) was added to protein samples prior to boiling at a final concentration of 50 mM. ColorPlus Prestained

Protein Marker (New England Biolabs, UK) was loaded alongside samples which had a molecular weight range of 7-175 kDa. Gels were electrophoresed at 180 V for 45 min, or until the dye front had just run off the bottom of the resolving gel. Gels were either Coomassie Blue stained (see section 2.7.1 for stain and de-stain buffers) or transferred to PVDF membrane (Amersham Hybond™, GE Healthcare, UK) for Western blot analysis (as described in section 2.8.5 below).

2.8.5. Western blot analysis

Following SDS-PAGE separation, proteins were transferred from the gel to PVDF membrane at 100 V, 350 mA for 60 min with a Mini-Trans blot cell (Bio-Rad, UK) in the presence of transfer buffer (see section 2.8.1). After transfer, the membrane was incubated in PBS containing 5% (w/v) powdered milk and 0.05% (w/v) Tween 20 (Sigma-Aldrich, UK). The membrane was then incubated for 1 h at room temperature with primary antibody at the dilutions stated in each chapter. The membrane was washed 3 times with PBS-T for 5 min per wash, and then incubated in the corresponding secondary antibody conjugated to horse radish peroxidase for 1 h at room temperature. Following secondary antibody incubation, membranes were again washed in PBS-T (3 x 5 min) followed by a 10 min wash. Finally, the membrane was incubated with the enhanced chemiluminescence (ECL Plus) system (Amersham, GE Healthcare, UK) for 5 min and exposed to autoradiographic film (Amersham Hyperfilm ECL, GE Healthcare, UK) for varying periods of time. Films were developed using a Konica SRX-101A X-ray film processor.

2.8.6. Bradford assay

Protein concentration was determined according to the Bradford method. Samples were made up to 50 µl in dH_2O and 200 µl of Bradford dye reagent (Bio-Rad, UK) was added to each sample to give a 1:4 ratio of sample to reagent. Samples were incubated at room temperature for 5 min and the absorbance measured at 595 nm. Samples were analysed in triplicate and compared with a standard curve constructed using serial dilutions of bovine serum albumin (BSA).

2.8.7. Immunofluorescence Microscopy

Immunoblots were prepared as previously described (Patrick *et al.*, 2003). *B. fragilis* strains were grown overnight in BHI-S broth or DM as previously described. Cultures were diluted in sterile PBS to an OD_{600} of 0.3, and 20 µl were pipetted onto the wells of an 8-well Teflon

coated slide (ICN). Slides were air dried at 37°C and fixed in 100% methanol (Sigma-Aldrich, UK) at -20°C for 20 min. The slides were allowed to air dry and 20 µl of primary antibody were applied to each well. Slides were then incubated in a humidified box at 37°C for 1 h. Slides were washed, first by the application of sterile PBS down the centre of the slide and secondly in a shaking PBS bath for 20 min. The area surrounding the wells was dried using filter paper and 20 µl secondary antibody (IgG anti-mouse/anti-rabbit fluorescein isothiocyanate conjugate (FITC), Sigma-Aldrich, UK) was applied. The slides were incubated in a humidified box and washed as before, and then mounted in glycerol-PBS antifade reagent (Invitrogen). A coverslip was applied and the edges of the slide were sealed by nail varnish. Slides were stored in foil at -20°C.

2.9. *Bacteroides fragilis* OMV

2.9.1. Preparation of *B. fragilis* concentrated supernatants

Outer membrane vesicles were prepared based on a previously described method (Horstman & Kuehn, 2000) with modifications for *B. fragilis*. *B. fragilis* strains were grown overnight in 100 ml DM. Cells were removed by centrifugation at 6000 x g for 30 min, followed by centrifugation of the supernatant at 6000 x g for 15 min. The supernatant was filtered through a 0.45 µm PVDF filter (Millipore) to remove any remaining bacterial cells. Supernatants were concentrated using a 70 ml Vivaspin centrifugal concentrator (Sartorius Stedim Biotech, Germany) with a MWCO (molecular weight cut-off) of 100 kDa, by centrifugation at 1000 x g for 3 h or until a final volume of approximately 1.5 ml was achieved. If necessary, OMV were concentrated further by ultracentrifugation (Sorvall OTD ultracentrifuge, TH-641 swinging bucket rotor) at 100 000 x g for 1 h at 4°C, pellets were then washed twice in PBS by ultracentrifugation as previously described and re-suspended in 500 µl PBS. Protein concentration was determined using Bradford assay (see 2.8.6) and samples were stored at -20°C. All concentrates were checked to ensure there were no viable cells. This was done by plating 10 µl of concentrate onto two BHI-S plates, one was incubated anaerobically and one aerobically.

2.9.2. DNA analysis of OMV

DNA was isolated from *B. fragilis* concentrated supernatants based on previously described methods (Yaron *et al.*, 2000; Renelli *et al.*, 2004; Rumbo *et al.*, 2011). *B. fragilis* concentrated supernatants (100 µg/ml) were treated with RQ1 RNase-Free DNase

(Promega, UK), at a final concentration of 50 µg/ml. A standard reaction contained 200 µl concentrated supernatant, 10 µl RQ1 DNase, 25 µl RNase-Free 10 x reaction buffer and 15 µl dH_2O. Reactions were incubated at 37°C for 2 h before the enzyme was inactivated by the addition of 10 µl DNase Stop Solution and incubation at 65°C for 10 min. OMVs were pelleted by ultracentrifugation at 60 000 x g for 1 h at 4°C. Pellets were washed three times in 1 ml PBS, by ultracentrifugation as previously described. Following the final wash, the supernatant was collected and the pellet re-suspended in 100 µl PBS, overnight at 4°C. DNA was extracted by Mini-prep (see section 2.7.2) and PCR was carried out (see section 2.7.3) using oligonucleotides specific for *finB*, *ubb* and BF*4298* (Table 2.4).

2.9.3. Transformation of *E. coli* XL1-Blue MRF' and *B. fragilis* Δ*ubb* with *B. fragilis* OMV

Transformations were carried out based on a previously described method (Rumbo *et al.*, 2011). For *E. coli*, an overnight culture of *E. coli* XL1-Blue MRF' grown in LB tetracycline (10 µg/ml) was diluted 1:100 into fresh LB without antibiotic selection. Cultures were grown to an OD_{600} of 0.4 and cells were pelleted by centrifugation at 16 000 x g for 10 min. Pellets were re-suspended in 500 µl LB. Transformation reactions containing 800 µl LB, 100 µl *E. coli* XL1-Blue MRF' cell suspension, 100 µg/ml concentrated supernatant isolated from *B. fragilis* NCTC9343 containing pLYL01 and 50 µg/ml DNase were incubated at 37° C for 1 h without shaking. Reactions were then incubated for 2 h at 37°C, shaken at 200 rpm. 10 ml LB was added and cultures were incubated at 37°C overnight. Cells were pelleted by centrifugation at 6 000 x g for 15 min. Pellets were re-suspended in 500 µl LB and 5 x 100 µl aliquots were plated onto LB ampicillin (100 µg/ml) and tetracycline (10 µg/ml) and incubated at 37°C overnight. For *B. fragilis* Δ*ubb*, the method was as above with the following modifications: B. fragilis cells were grown in BHI-S and following transformation, reactions were incubated anaerobically without agitation. Samples were plated onto BHI-S erythromycin (10 µg/ml) and tetracycline (10 µg/ml) and incubated anaerobically overnight.

2.9.4. Transmission Electron Microscopy

OMV samples were diluted to 10 µg/ml and 1 µg/ml in sterile PBS, negatively stained with 2% uranyl acetate and visualized by transmission electron microscopy (TEM).

2.10. Cell culture techniques

2.10.1. Culture and passaging of cells

Cell culture experiments were carried out with human epithelial colorectal adenocarcinoma cell line, Caco-2 (Sigma Aldrich, UK). Cells were cultured in Eagle's minimum essential medium with Earle's salts (EMEM, Sigma-Aldrich, UK) supplemented with 10% foetal bovine serum (FBS, Sigma-Aldrich, UK), 1% non-essential amino acids (Sigma-Aldrich, UK), 5 mM L-glutamine (Sigma-Aldrich, UK and Pen-Strep (100 U/ml penicillin, 10 µg/ml streptomycin, Sigma-Aldrich, UK). The cells were recovered from liquid nitrogen stocks by thawing for 1 min at 37°C before the addition of 15 ml of pre-warmed medium and centrifugation at 1000 x g for 5 min. Cells were re-suspended in fresh medium, diluted to an appropriate dilution and incubated in a humid incubator at 37°C with 5% CO_2. The cells were checked regularly by microscopic examination and when monolayers reached ~90% confluence, they were passaged by washing 3 times with 10 ml pre-warmed PBS followed by addition of 5ml 0.05% trypsin-EDTA (Sigma-Aldrich, UK) and incubation at 37°C for 5 min. 5 ml warm EMEM was added to cells to inactivate the trypsin, and cells were scraped off and collected by centrifugation at 1000 x g for 5 min. The cells were re-suspended at an appropriate concentration in fresh EMEM and incubated at 37°C with 5% CO_2. Frozen stocks were prepared by freezing $5x10^5$ cells/ml in EMEM supplemented with 20% FBS and 10% DMSO (Sigma-Aldrich, UK). Cells were incubated at -20°C for 1 h, -80°C overnight and finally stored in liquid nitrogen.

2.10.2. Cell counts

Cells were counted using The Countess™ automated cell counter (Invitrogen, USA) following manufacturer's instructions.

2.10.3. Mycoplasma detection

The LookOut™ Mycoplasma PCR Detection Kit (Sigma-Aldrich) was used to detect Mycoplasma infections in the Caco-2 cell line when media appeared contaminated. Cells were cultured in the absence of Pen-Strep until 100% confluent and 100 µl of supernatant was removed. The supernatant was boiled for 5 min at 95°C and PCRs were prepared and performed following manufacturer's instructions. The products were visualised by agarose gel electrophoresis as described in 2.7.4.

2.10.4. Incubation of *B. fragilis* concentrated supernatants with Caco-2 cells

Caco-2 cells were seeded onto poly-L-lysine (Sigma-Aldrich) treated glass coverslips and grown until 90% confluent. Cells were incubated with 100 µg/ml *B. fragilis* concentrated supernatants containing OMV for 2 h. The cells were then washed three times in PBS, fixed in 4% paraformaldehyde and stained by incubation with 100 µl, 300mM DAPI (Invitrogen, USA) for 5 min. Cells were visualised using confocal microscopy.

2.10.5. Confocal Microscopy

Confocal fluorescent images were obtained using a Leica SP5 confocal microscope with a x63 objective. Images were analysed by ImageJ software.

2.11. Bacteriophage techniques

2.11.1. Isolation of Bacteriophages

Bacteriophages were isolated from 20L of raw sewage that was coarse filtered and fine filtered, then concentrated by tangential flow filtration. Ten-fold serial dilutions of sewage were plated onto 0.7% BHI-S top agar containing 100 µl of stationary phase *B. fragilis* culture and plates were incubated anaerobically, overnight. Individual plaques were extracted using a sterile glass pipette and used to inoculate 10 ml BHI-S containing 100 µl *B. fragilis* cells, cultures were incubated anaerobically overnight. Cells were then removed by centrifugation at 6,000 x g for 10 min and supernatants were filter sterilised through a 0.45 µm filter (Millipore) and stored at 4°C.

2.11.2. Plaque assay for phage titre

Ten-fold serial dilutions of bacteriophage lysates were made in phage buffer and 10 µl aliquots were spotted onto 0.7% (w/v) agar BHI-S top agar containing 100 µl *B. fragilis* cells. Plates were incubated anaerobically overnight and plaques were counted for each dilution. Plaque forming units (PFU) were calculated as follows:

$$\text{PFU/ml} = \frac{\text{Number of plaques}}{d \times V}$$

d = dilution
V = volume added to the plate

Phage buffer: 50mM Tris-HCl pH 7.5, 150mM NaCl, 10mM $MgSO_4$, 2mM $CaCl_2$

2.11.3. Transduction by bacteriophages

Bacteriophage lysates (100 µl) were plated onto BHI-S top agars containing 100 µl *B. fragilis* NCTC9343 *ΔupcY*::*ermF* and incubated anaerobically, overnight. The top agar was washed with 2 ml bacteriophage buffer, and scraped into a glass universal. This was incubated at room temperature for 20 min before the addition of 500 µl of chloroform and incubation for a further 15 min. Samples were centrifuged at 4000 x g for 20 min and the supernatant filter sterilized using a 0.45 µm filter (Millipore). Neat, 1:10 and 1:100 dilutions of sterilized supernatants were used to inoculate 100 µl NCTC9343 or 638R, these cultures were incubated at 37°C for 15 min, followed by the addition of 500 µl BHI-S and incubation anaerobically for 30 min. Cultures were spread onto BHI-S erythromycin (10 µg/ml) plates and incubated anaerobically overnight.

2.11.4. Preparing bacteriophage DNA from lysates

2.11.4.1. PEG precipitation

Large bacteriophage lysates were prepared in 50 ml BHI-S by inoculating 100 µl bacteriophage and 500 µl *B. fragilis* overnight culture, followed by anaerobic incubation for 16 h. Cells were removed by centrifugation at 6 000 x g for 20 min at 4°C and supernatants were filter sterilized through a 0.45 µm filter (Millipore). The filtered lysate was transferred to a 100 ml graduated cylinder and 5x PEG solution was added to a final concentration of 1x. Tubes were inverted to mix and incubated at 4°C for 48 h. The lysate was centrifuged at 3000 x g for 10 min at 4°C, the supernatant discarded and the pelleted white precipitate was re-suspended in 100 µl of bacteriophage buffer. The amount of KCl (Sigma-Aldrich) required to make a 1M solution was determined and this was added to the precipitate in three aliquots of approximately equal size, vortexing after each addition. The samples were then incubated on ice for 30 min before centrifugation at 12 000 x g for 10 min at 4°C to pellet the PEG. The supernatant, containing the bacteriophage was decanted and stored at 4°C.

5x PEG Solution: 20% (w/v) PEG-8000 (Sigma-Aldrich), 2.5M NaCl (Sigma-Aldrich)

2.11.4.2. Caesium Chloride gradients

Caesium chloride (CsCl) step gradients were used to isolate the bacteriophage particles. CsCl solutions were prepared in dH_2O, filter sterilized and stored at 4°C. Three layers of CsCl

were added to a centrifuge tube as follows: first layer, 3 ml CsCl solution, *d* = 1.7 g/ml; second layer, 2 ml CsCl solution, *d* = 1.5 g/ml; third layer, 2 ml CsCl solution, *d* = 1.3 g/ml. Layers were added slowly using a syringe to avoid mixing. 2 ml of bacteriophage lysate isolated from the PEG precipitation in 2.11.4.1 was carefully layered on top of the gradient (Figure 2.1a). Gradients were ultracentrifuged at 104 000 x g for 2 h at 4°C. Following centrifugation, a blue band was visible in the *d* = 1.5 g/ml layer containing bacteriophages, two further bands were sometimes visible, a second blue band in the *d* = 1.3 g/ml layer containing empty phage heads and a white band above the *d* = 1.3 g/ml layer containing cell debris (Figure 2.1b). The band containing the bacteriophages was extracted using a 10 ml syringe with a 25-G needle and stored at 4°C.

2.11.4.3. Formamide DNA extraction

The volume of the bacteriophage band isolated in 2.11.4.2 was weighed and a 0.1 vol of 2 M Tris-Cl (pH 8.5)/0.2 M EDTA was added. Tubes were inverted to mix. An equal volume of formamide (Sigma-Aldrich) was added and the sample was incubated at room temperature for 30 min. Two volumes, each equal to the volume of the original phage band, of room temperature, 100% ethanol were added, the sample was vortexed and centrifuged at 13 000 rpm (16, 100 x g) for 2 min. The pellet was rinsed in 70 % ethanol and re-suspended in 100 µl 1x TE (pH 8) by incubation at 37°C for 30 min. DNA concentration was measured by nanodrop as described in 2.7.6 and samples were stored at -20°C.

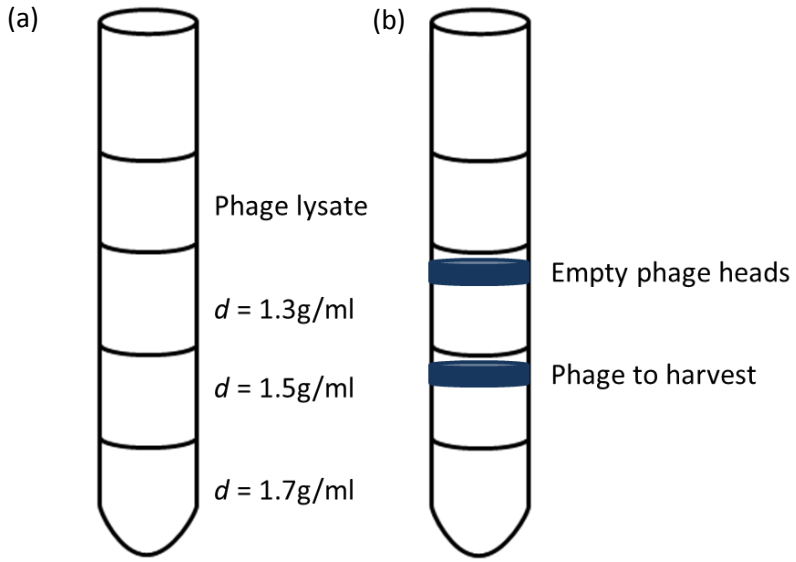


Figure 2.1: Caesium chloride gradient set-up for bacteriophage isolation
(a) Representation of the CsCl gradient prior to centrifugation, a gradient consisting of three densities of CsCl was added to the centrifuge tube. The phage lysate isolated from the PEG precipitation was layered on top. (b) Following ultracentrifugation, a blue-ish band containg bacteriophage was visible within the 1.5g/ml layer. A second band containing empty phage heads was also visible within the 1.3g/ml layer.

2.11. Bioinformatics

2.11.1. Sequencing data, assembly and annotation

The quality of the whole genome sequencing data was assessed using FastQC (Andrews) and contig assemblies were performed using CLC Genomics workbench v 6.4. When the FASTA files containing the sequencing reads were imported, the data were identified as Illumina, paired-end (reverse-forward) reads under the general options section. The minimum and maximum read distance was set at 100 and 250 bp, respectively. Contigs were assembled using the De Novo assembly tool with the following parameters: word size, 45; bubble size, 98; minimum contig length, 1000bp.

Contigs were re-ordered against the genome of NCTC9343 using the progressive Mauve program of the Mauve Genome Alignment Software (Darling *et al.*, 2004; Darling *et al.*, 2010). Re-ordering was carried out using default settings, with the exception of the seed weight which was set at 11. Genomes were annotated using xbase (Altschul *et al.*, 1997; Lowe & Eddy, 1997; Kurtz *et al.*, 2004; Delcher *et al.*, 2007; Lagesen *et al.*, 2007; Chaudhuri

et al., 2008) with *B. fragilis* NCTC9343 selected as the reference genome, or using the Rapid Annotation using Subsystem Technology (RAST) server (Aziz *et al.*, 2008).

2.11.2. Genome comparisons

Genomes were viewed using the Artemis genome browser (Rutherford *et al.*, 2000) and genome comparisons were carried out with the Artemis Comparison Tool (ACT), WebACT (Carver *et al.*, 2005) or the BLAST Ring Image Generator (BRIG).

Chapter 3. Inter-strain Horizontal Gene Transfer

3.1. Introduction

The acquisition of genes from other strains and species through HGT is a main contributor to bacterial evolution. The products of genes gained through HGT can confer a selective advantage to the organism, such as the acquisition of virulence factors. HGT can also mediate the rapid dissemination of antibiotic resistance determinants throughout bacterial species, usually via the conjugative transfer of plasmids (see section 1.2.1). The acquisition of antibiotic resistance genes can have a profound impact on the ability to treat bacterial infections; many of the bacterial pathogens associated with epidemics of disease have evolved into multi-drug resistant (MDR) forms (Davies & Davies, 2010).

HGT also influences inter-strain variation, demonstrated by the presence of accessory genomes which are highly variable between bacteria of the same species (See section 1.1). Genomic comparisons of the complete genome sequences of *B. fragilis* strains NCTC9343, 638R and YCH46 originally isolated in the UK, USA and Japan, respectively, revealed that extensive horizontal gene transfer has occurred within these strains. Patrick *et al* (2010) demonstrated that each strain contained ten divergent polysaccharide biosynthesis loci, none of which were similar between NCTC9343 and 638R. YCH46 had one common locus with 638R and one with NCTC9343, therefore a total of twenty eight divergent polysaccharide biosynthesis loci were identified amongst three strains. The authors also observed the presence of other horizontally acquired elements within the three genomes, such as prophages and conjugative transposons. To further examine the variation in *B. fragilis*, the genomes of seven *B. fragilis* clinical isolates were sequenced using the Illumina MiSeq sequencing platform and comparisons of four of these genomes, BE1, GNAB92, RD48 and LS66 were made with *B. fragilis* strains NCTC9343, 638R and YCH46.

3.2. Results

3.2.1. Multi-drug resistant *B. fragilis* isolate, WAL610

B. fragilis is the most frequently isolated Gram-negative pathogen from anaerobic infection and can cause severe intra-abdominal infections, postoperative wound infections, skin and soft-tissue infections and bacteraemia (see section 1.5). Metronidazole is the most effective antibiotic for the treatment of *B. fragilis* infections, however, there have been reports of

MDR strains. The emergence of metronidazole resistance in *B. fragilis* is a worrying development as these strains can prove difficult to treat and could result in increased mortality (Wareham *et al.*, 2005). A MDR strain of *B. fragilis,* WAL610, was isolated from the leg wounds of an American soldier who sustained multiple injuries in an improvised explosive device blast in Afghanistan. The otherwise healthy male soldier sustained open fibula and tibia fractures to both legs in the explosion and the wounds were exposed to water contaminated with raw sewage. The wounds became infected and despite treatment with broad-spectrum antibiotics, wound irrigation and the removal of infected tissue, repeated wound cultures tested positive for *B. fragilis*. Upon returning to the USA, intravenous metronidazole was added to the antibiotic therapy, but despite this intervention the patient required amputation of his right leg below the knee. Subsequent antibiotic susceptibility testing revealed that the *B. fragilis* strain was resistant to penicillin, clindamycin, metronidazole, cefoxitin, meropenem, imipenem, piperacillin/tazobactam, and tigecycline (Sherwood *et al.*, 2011). We were contacted by Hannah Wexler from the West Los Angeles Veterans Medical Centre and asked to determine if resident plasmids were involved in this strain's resistance to metronidazole.

Plasmid DNA was extracted and purified from a 50ml overnight culture of *B. fragilis* WAL610, grown in BHI-S (see section 2.7.2). Two plasmids of 5.5 kb and 8.3kb were identified and subsequently sequenced using the Roche 454 platform. The 5.5kb plasmid was identical in sequence to pHAG1, isolated from a multidrug resistant strain of *B. fragilis* in the UK (Pumbwe *et al.*, 2007), and to pBFB35, frequently isolated from *B. fragilis* clinical isolates in Hungary (Sóki *et al.*, 2010). The second plasmid, pWAL610 was annotated by the Rapid Annotation using Subsystems Technology (RAST) server (Aziz *et al.*, 2008) and analysed using BLAST and Artemis. A circular plasmid map was generated using DNAplotter (Carver *et al.*, 2009) (Figure 3.1). pWAL610 contained the nitromidazole resistance gene, *nimE,* which shared 100% DNA sequence identity to the *nimE* gene of *B. fragilis* plasmid pBF388c, from a metronidazole resistant strain of *B. fragilis* isolated in Kuwait (Sóki *et al.*, 2004). A transposase containing a putative insertion sequence was also identified, with 99% identity to transposase Bf6 of pBF388c. Whole genome sequencing of *B. fragilis* WAL610 is currently being completed by the Broad Institute.

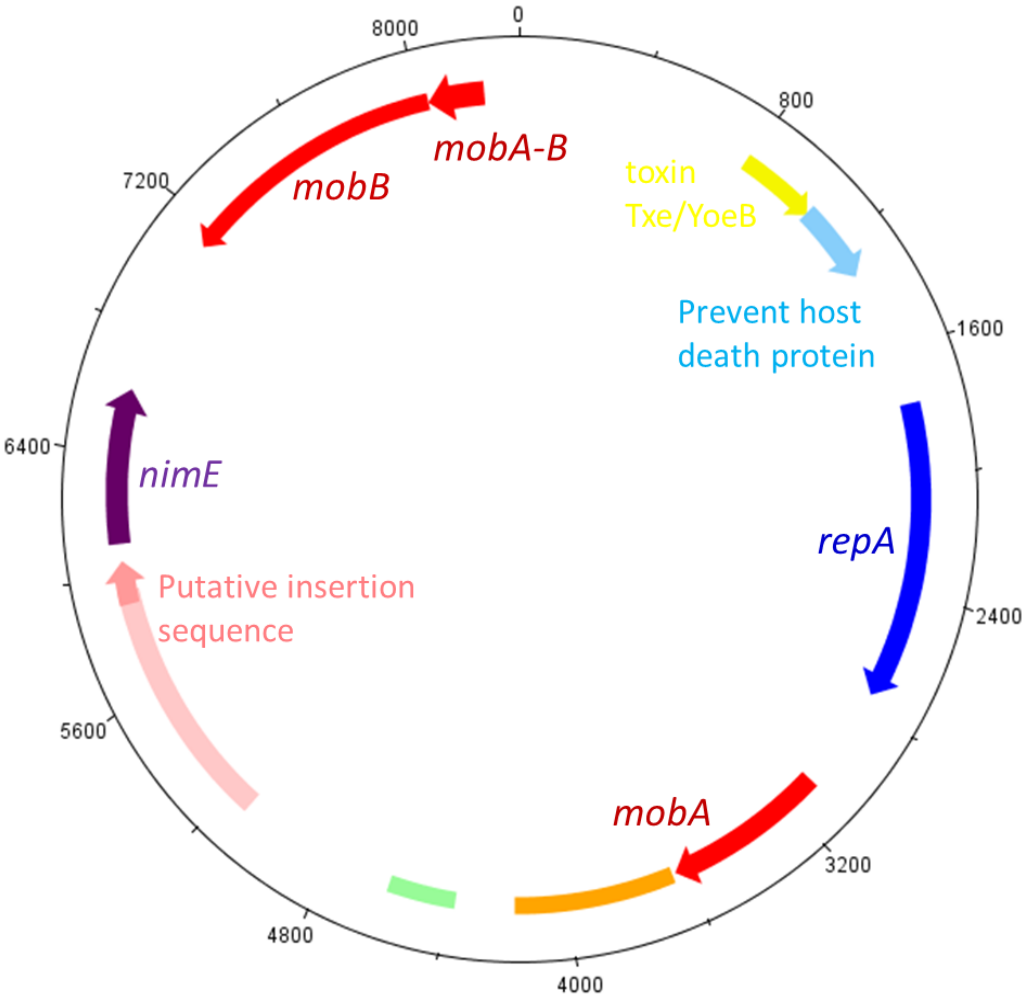


Figure 3.1: Artemis DNA plotter image of pWAL610

8.3kb pWAL610 isolated and sequenced from *B. fragilis* WAL610. Circular tracks from outside to inside: 1, DNA co-ordinates; 2, forward strand coding sequenced; 3, reverse strand coding sequences. The nitromidazole resistance gene, *nimE* was identified, labelled and shown in purple, along with a transposase containing a putative insertion sequence, represented in pink.

3.2.2. Illumina MiSeq sequencing

The genomes of seven *B. fragilis* clinical isolates, BE1, RD48, GNAB92, LS66, MPRL3499VI, MPRL3535V and MPRL3536X were sequenced using the Illumina MiSeq platform to further investigate inter-strain variation present in *B. fragilis.* These strains were isolated from

patients in Edinburgh, Amsterdam and Belfast (Table 4.). Sample preparation and genome sequencing was carried out as described in 2.7.12.1. Bacterial chromosomal DNA was isolated from overnight cultures, grown in BHI-S, and any RNA was removed using Riboshredder. The RNase-treated DNA was purified, concentrated. A sequencing library was then generated using a different barcoded adaptor for each isolate. Samples were pooled and sequenced using the Illumina MiSeq platform, which generated 250 bp paired-end reads for all samples, with the exception of GNAB92, which generated 150 bp paired-end reads.

3.2.2. Assembly and Annotation

The raw data were filtered to remove barcoded adaptors and the quality of the data was assessed using FastQC (Andrews). The basic statistics are summarised in Table 3.1. All seven genomes had a GC content between 42-45%, consistent with the three previously sequenced strains, which have an average GC content of 43%. The estimated coverage of the genomes ranged from 33-fold (MPRL 3536X) to 389-fold (MPRL 3535V).

Contig assemblies were performed on CLC Genomics Workbench v 6.4 using default parameters for De Novo Assembly. A minimum contig length was set at 1000 bp to eliminate smaller contigs, which are difficult to map. The assembly quality data are shown in Table 3.2. The quality of the genomes was assessed using the N50 value, the number of contigs, the average contig length and the overall size of the genome. N50 is a statistical measure of the data, in which 50% of the entire assembly is contained in contigs equal to or larger than this value. For example, 50% of the reads of GNAB92 were contained in contigs of 102122 bp or larger. A large number of contigs and a small average contig size indicate a poor genome assembly as it is difficult to align a high number of small fragments. The overall size of the genome should be similar to that of previously sequenced strains; a difference in size indicates there may not be complete coverage of the genome. Based on these data, it was decided that three genomes, MPRL 3536X, MPRL 3499VI and MPRL 3535V had not assembled well enough to be analysed further. The three genomes had N50 < 25 kb and an average contig length < 12kb. All three assemblies contained > 400 contigs.

Table 3.1: FastQC summary statistics of genomes

Strain	Total reads	Sequence length (bp)	GC content (%)	Estimated coverage
GNAB92	2276671	21-150	44	65x
BE1	898420	148-251	43	87x
RD48	1373685	169-251	43	134x
LS66	1094161	158-251	45	214x
MPRL 3499VI	474823	154-251	45	49x
MPRL 3536X	333633	155-251	44	33x
MPRL 3535V	1335241	155-251	42	389x

The contigs of the remaining four strains, GNAB92, BE1, RD48 and LS66 were aligned against NCTC9343 using Mauve Genome Alignment Software (Darling *et al.*, 2004; Darling *et al.*, 2010). Genomes were then annotated using xBASE (Altschul *et al.*, 1997; Lowe & Eddy, 1997; Kurtz *et al.*, 2004; Delcher *et al.*, 2007; Lagesen *et al.*, 2007; Chaudhuri *et al.*, 2008). Genomes were viewed using the Artemis genome browser (Rutherford *et al.*, 2000), and comparisons were carried out with the Artemis Comparison Tool (ACT) and the online version, WebACT (Carver *et al.*, 2005).

3.2.3. Analysis of the polysaccharide biosynthesis loci

B. fragilis NCTC9343, 638R and YCH46 each contain ten annotated regions (PSA-J) which have been implicated in polysaccharide production. Genetic comparisons of these loci revealed a high level of diversity, across the three strains, with twenty-eight different polysaccharide biosynthesis loci being identified (Patrick *et al.*, 2010). Eight of the ten loci, A-H, were implicated in the expression of polysaccharides which form the within-strain variable MC. Seven of these loci are switched on and off by the site-specific inversion of promoter sequences, which is catalysed by the invertase, *finA* (Patrick *et al.*, 2003). *finA* was identified in GNAB92 (CDS 2980), BE1 (CDS 2671), RD48 (CDS 2882) and LS66 (CDS 2768), each gene had 99-100% DNA sequence identity to NCTC9343, YCH46 and 638R. This suggested that some of the polysaccharide biosynthesis loci of BE1, RD48, GNAB92 and LS66 may also be switched on and off by site-specific inversion.

Table3.2: Genome assembly data

	Strains						
Feature	**GNAB92**	**BE1**	**RD48**	**LS66**	**MPRL 3499VI**	**MPRL 3536X**	**MPRL 3535V**
N75	56 843	45 651	53 326	27 602	5120	11 783	13 227
N50	102 122	79 420	85 038	49 700	9834	22 618	24 418
N25	183 866	144 275	143 311	84 597	17 416	38 914	39 212
Min. contig length	1048	1070	1085	1051	1006	1006	1009
Max. contig length	439 311	245 186	223 301	224 573	42 524	87098	94 041
Average contig length	49 281	43 032	50 689	25 630	6183	11954	10 224
Number of contigs	108	120	102	203	800	417	478
Size of genome (bp)	5 322 301	5 163 884	5 170 323	5 202 944	4 946 621	4 985 020	4 022 826

Transcription antitermination factors

Although divergent, the polysaccharide biosynthesis loci A-H of NCTC9343, YCH46 and 638R, share at least one common genetic feature. The first two genes in each of the loci encode UpxY and UpxZ proteins, where x represents the polysaccharide locus the proteins are associated with (UpaY to UphY and UpaZ to UphZ). These genes are required for polysaccharide synthesis because the encoded proteins act as transcription antitermination factors, preventing the premature termination of transcription of the associated polysaccharide operon (Chatzidaki-Livanis *et al.*, 2009).

Eight potential polysaccharide biosynthesis loci were identified in GNAB92, BE1, RD48 and LS66, all of which contained genes which encoded transcription antitermination proteins, with the exception of RD48 which did not encode homologues of *upbY* and *upbZ.* The DNA sequences of the *upxY* and *upxZ* genes identified in these strains were aligned to those of the reference genomes using ClustalW (Larkin *et al.*, 2007). The DNA sequence identity of the genes was 92-100%, the lowest being the *upcZ* gene of LS66 which had 92% identity to the *upcZ* genes of the other strains.

Polysaccharides I and J

The PSI locus in NCTC9343 has yet to be associated with a capsular type, however it has been identified in all three reference genomes, with 638R and YCH46 sharing the same 12 genes in this region. NCTC9343 encodes a different set of genes suggesting the production of a different polysaccharide (Patrick *et al.*, 2010). PSI appears to lack an invertible promoter, but does contain a transcription antitermination gene, *upiY*. PSJ in NCTC9343

contains the only homologue of the O-antigen chain length determinant protein, Wzz, identified in the genome of NCTC9343 (CDS 1708). Deletion of this gene resulted in the absence of HMMPS, which form the MC (Patrick *et al.*, 2009). This locus is shared by YCH46 and 638R. The composition of these loci were determined in the four new genomes.

Alignment with the reference genomes

Polysaccharide biosynthesis loci A - J of each of the four new genomes were compared to those of NCTC9343, 638R and YCH46 in a five-way ACT comparison, apart from LS66, which was compared to 638R and YCH46 only. The comparisons were performed to identify any divergent loci present in the sequenced strains. As the transcriptional regulators are highly conserved, the loci were aligned using these genes as a fixed point. The locations of each of these loci are shown in Table 3.3.

Table 3.3: Polysaccharide associated biosynthesis loci in *B. fragilis* BE1, GNAB92, RD48 and LS66.

Locus	CDSs *				Capsule Association
	BE1	GNAB92	RD48	LS66 [3]	
A	1350 - 1366	1396 - 1401	1418 - 1428	1367 - 1368	MC
B	1739 - 1757	1922 - 1938	1951 - 1957 [2]	1832 - 1854	MC
C	1045 - 1052	1078 - 1099	1075 - 1092	1062 - 1086	MC
D	3752 - 3777	4006 - 4007 [1]	3855 - 3871	3742 - 3748	MC
E	2475 - 2489	2780 - 2796	2678 - 2697	2567 - 2577	MC
F	1448 - 1472	1551 - 1595	1597 - 1614	1521 - 1529	MC
G	0756 - 0777	0767 - 0796	0788 - 0811	0733 - 0746	MC
H	3417 - 3434	3681 - 3690	3609 - 3624	3486 - 3498	MC
I	2681 - 2708	2990 - 3019	2892 - 2920	2777 - 2803	Unknown
J	1613 - 1617	1738 - 1742	1757 - 1761	1672 - 1676	MC
LC assembly	2673 - 2675	2982 - 2984	2884 - 2886	2770 - 2772	LC

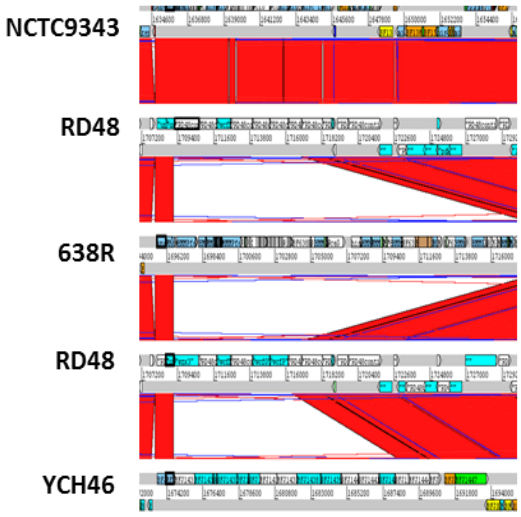
*Coding sequences
[1] location of *updY* and *updZ*
[2] apparent deletion, CDSs of remaining genes
[3] some PS loci poorly aligned, CDSs of genes which could be located

3.2.3.1 Strain RD48

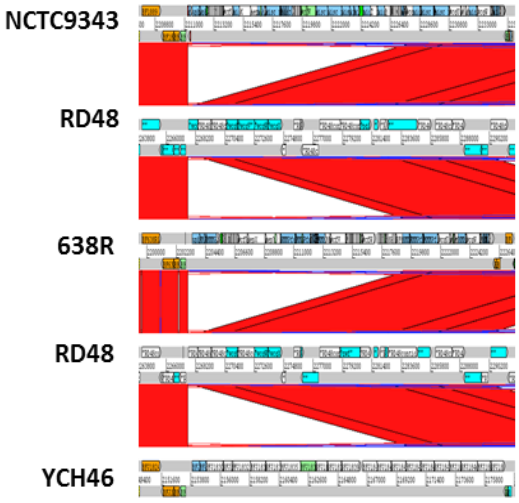
Genomic comparisons of RD48 with YCH46, 638R and NCTC9343 revealed that the strain shared (97-100% DNA sequence identity) nine of the polysaccharide biosynthesis loci with NCTC9343 (Figure 3.2). Eight of the nine common loci contained genes encoding a Wzx and Wzy homologue. These proteins shared 99-100% amino acid sequence identity to the Wzx and Wzy homologues encoded by the corresponding NCTC9343 locus. The PSB locus of RD48 appeared to contain a deletion of ~ 16.5 kb which included the transcriptional regulators *upbY* and *upbZ*, *wzx*, *wzy* and a number of glycosyltransferase genes (Figure 3.2 PS-B). The transcription antitermination factors have been shown to be essential for polysaccharide production (Chatzidaki-Livanis *et al.*, 2009). Therefore the absence of *upbY* and *upbZ* should render RD48 unable to express PSB. The ability of RD48 to express PSB was determined using IFM, with the PSB-specific monoclonal antibody, QUBf25 (see section 2.8.7). Δ*wzz* cells, which do not produce MC polysaccharides, were used as a negative control and NCTC9343 cells which have been shown to cross-react with QUBf25 (Patrick *et al.*, 2009), as a positive control.

Fluorescence was detected in a subset of RD48 cells (Figure 3.3), demonstrating that PSB was expressed by this strain and indicating that polysaccharide produced from this locus shares epitopes with PSB from NCTC9343. This suggested the deletion observed in the genome is likely to be an assembly associated error. MC expression is driven by invertible promoters, meaning that in a population there will be cells which have promoters on or off for multiple polysaccharide loci. This explains why only a subset of NCTC9343 cells cross-reacted with the PSB-specific monoclonal antibody. As expected, no fluorescence was detected in the Δ*wzz* control.

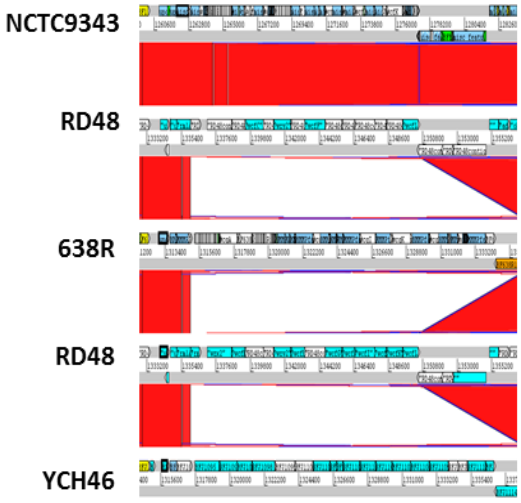
PS-A



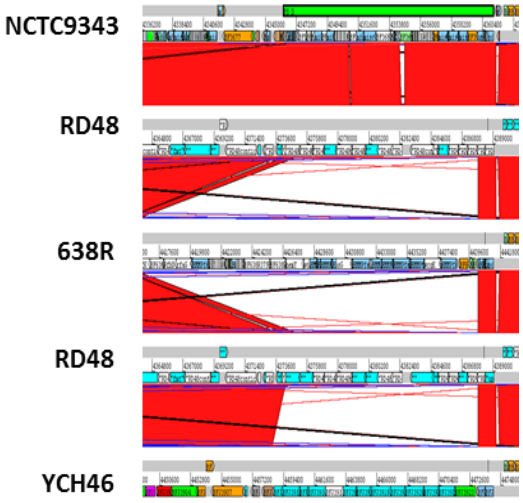
PS-B



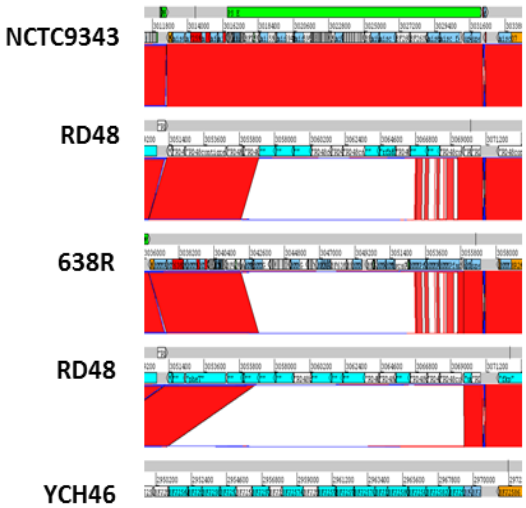
PS-C



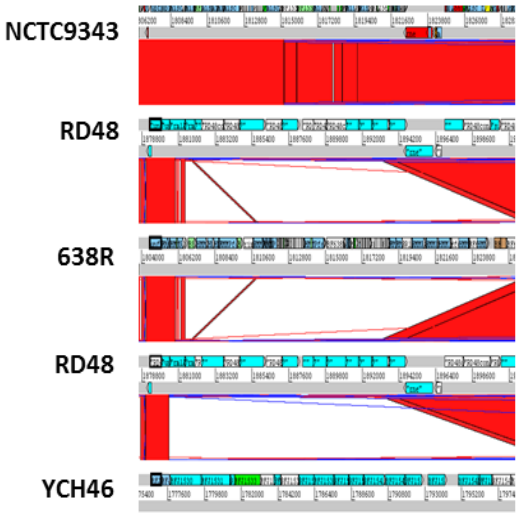
PS-D



PS-E



PS-F



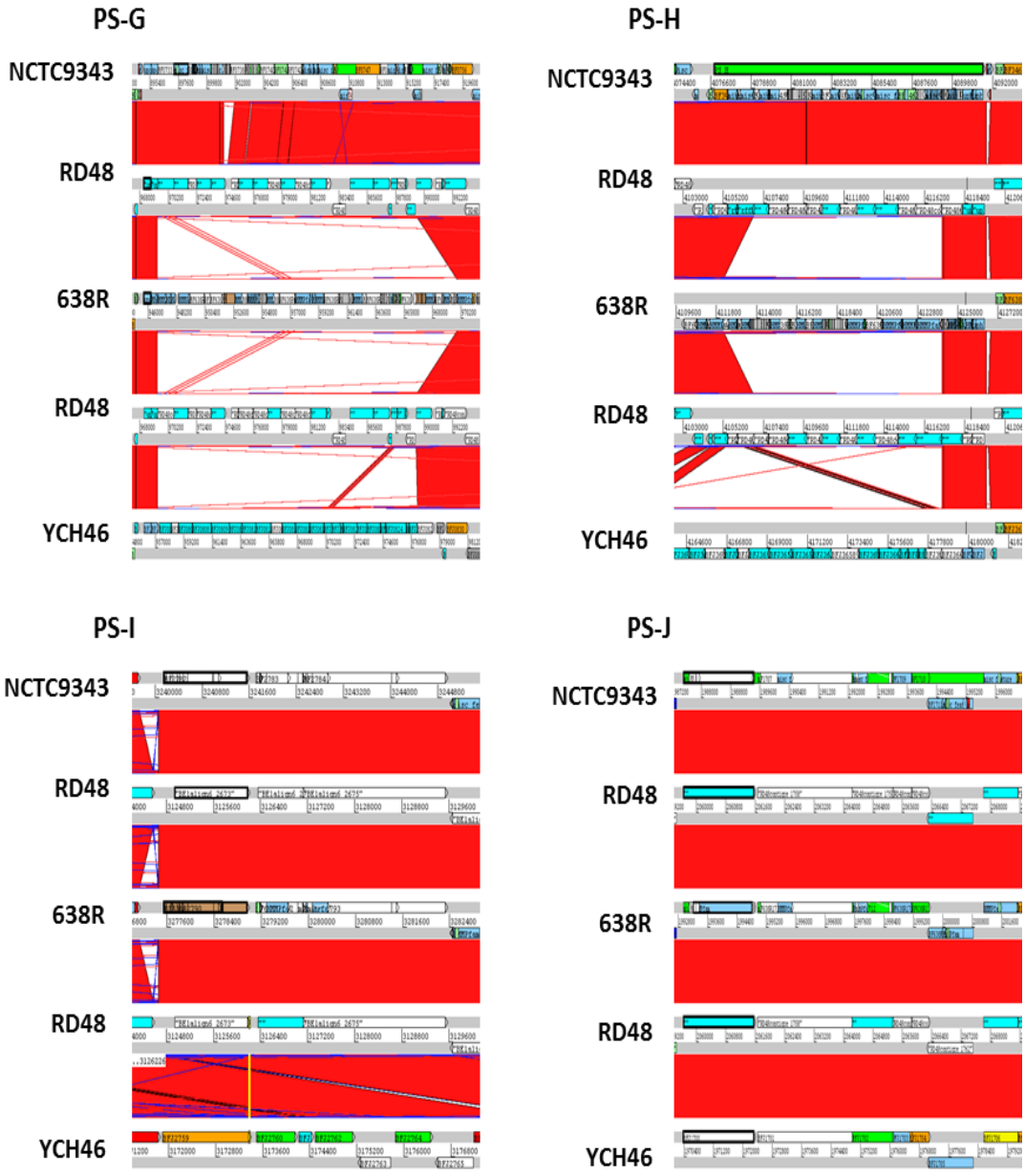


Figure 3.2: Comparison of RD48 polysaccharide biosynthesis loci

Five-way ACT genome comparison of the polysaccharide biosynthesis loci of RD48 with *B. fragilis* strains NCTC9343, 638R and YCH36. Each polysaccharide is labelled and strain names are shown to the left of the ACT comparison. The presence of red colour between sequences indicates sequence identity, no colour indicates divergent sequences and dark blue indicates a DNA inversion. Sequence identity was evident between RD48 and NCTC9343 at all loci apart from PSB. There is sequence identity between all four strains at locus PSJ.

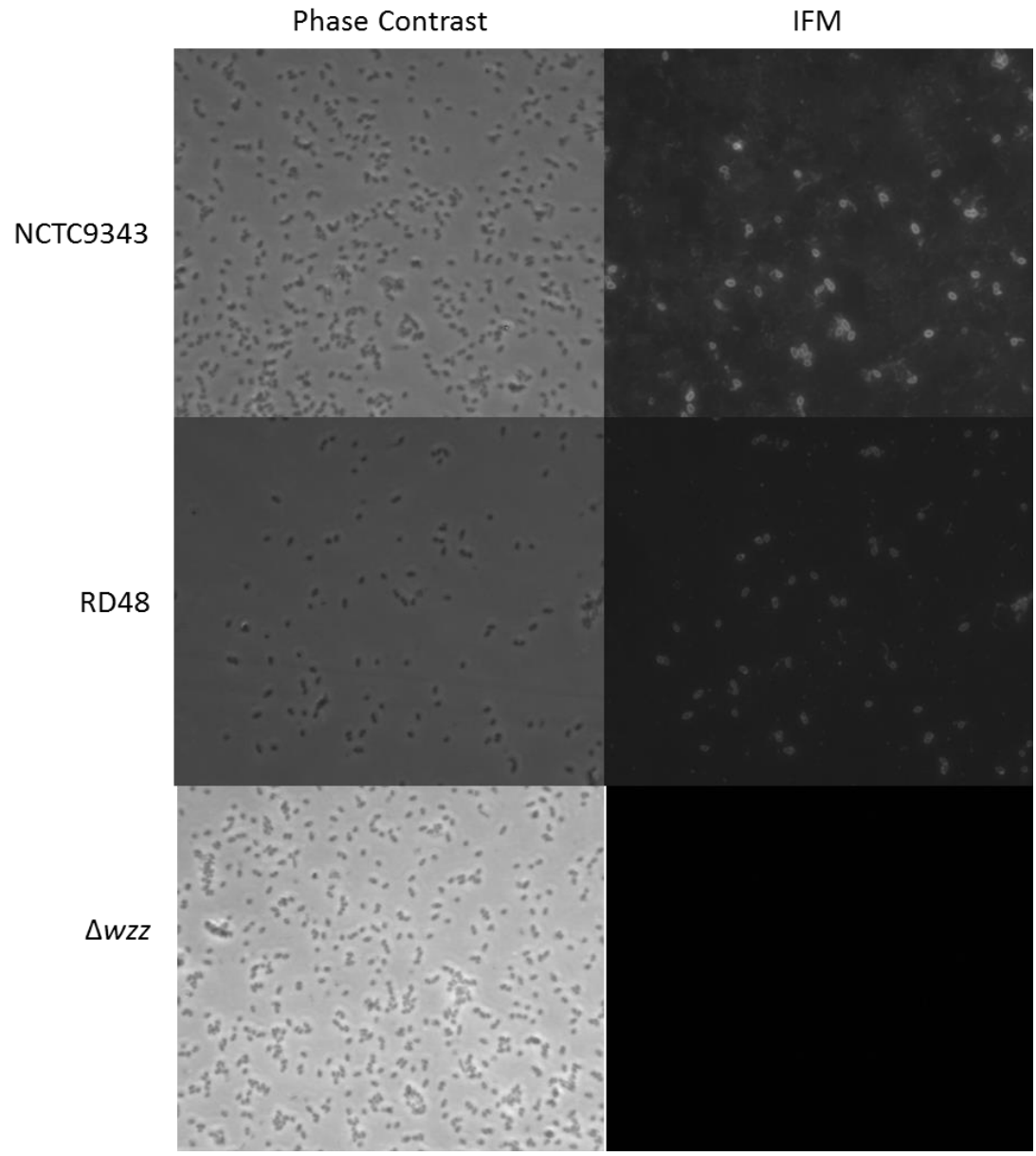


Figure 3.3: RD48-Immunofluorescence Microscopy with PSB-specific antibody
IFM using PSB-specific monoclonal antibody QUBf25 to probe cells isolated from *B. fragilis* strains NCTC9343, RD48 and Δ*wzz.* Fluorescence was detected in a subset of NCTC9343 and RD48 cells, indicating PSB production. No fluorescence was detected in Δ*wzz* cells, which do not produce MC polysaccharides

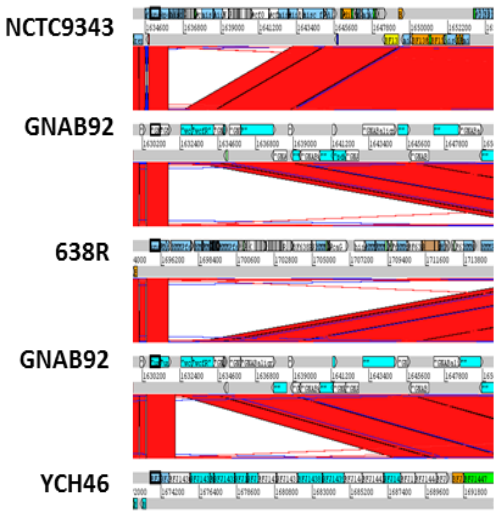
3.2.3.2. Strain GNAB92

Genomic comparison of GNAB92 with NCTC9343, 638R and YCH46 (Figure 3.4) showed that the strain contained eight complete polysaccharide biosynthesis loci, however, there appeared to be a deletion in both PSA and PSD (Figure 3.3a and d). The PSA locus contained five of the genes found in the corresponding locus of NCTC9343 and YCH46; the transcription antitermination genes *upaY* and *upaZ*, along with the final three genes of the operon encoding two glycosyltransferases and an aminotransferase were present. The genes encoding Wzx and Wzy were absent, suggesting this strain is unable to produce PSA. However, it was observed that PSC, which had some identity to both YCH46 and 638R, contained eleven extra genes, two of which encoded a second predicted Wzx and Wzy. These genes had homology to the YCH46 PSA locus (Figure 3.5), suggesting that instead of a deletion in PSA, there was an error during assembly of the genome and that this locus was actually present.

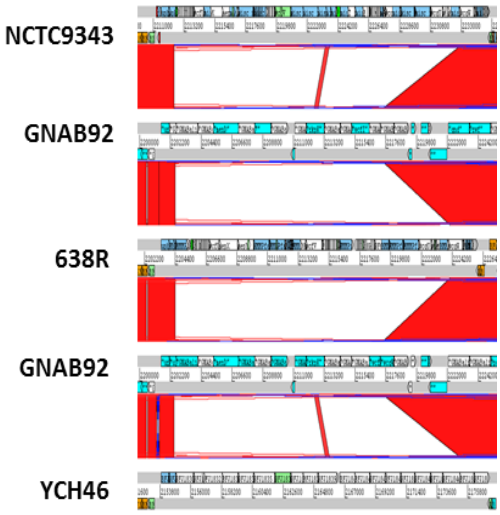
The PSD locus contained only *updY*, *updZ* and the final gene of the 18 gene operon, suggesting GNAB92 might be unable to express PSD. PSE encoded 17 genes, 8 of which had identity to the PSE locus of 638R. The predicted homologues of *wzx* and *wzy* however, were divergent therefore this locus was classified as different.

In comparison to the three reference genomes, GNAB92 contained five divergent polysaccharide biosynthesis loci, PSB, PSE, PSG, PSH and PSI. GNAB92 PSA might have been a misassembly and therefore represent a common locus with YCH46. PSC was common with 638R and YCH46, PSF was similar with 638R and PSJ was conserved in all four strains (between 99-100% DNA sequence identities). A potential Wzx flippase and Wzy polymerase was encoded within each locus, shown in Table 3.4 and discussed in Section 3.2.3.6.

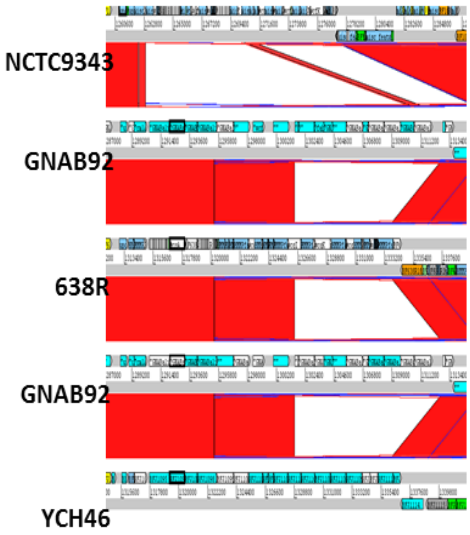
PS-A



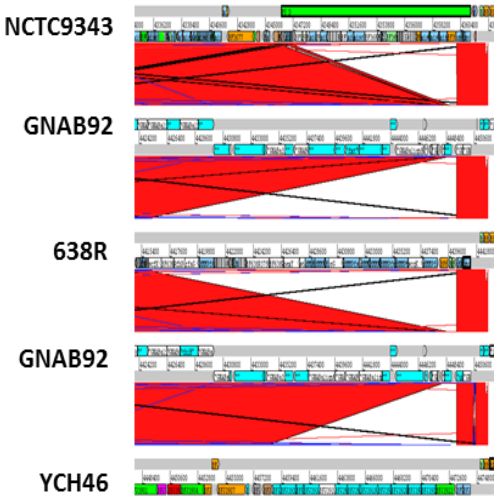
PS-B



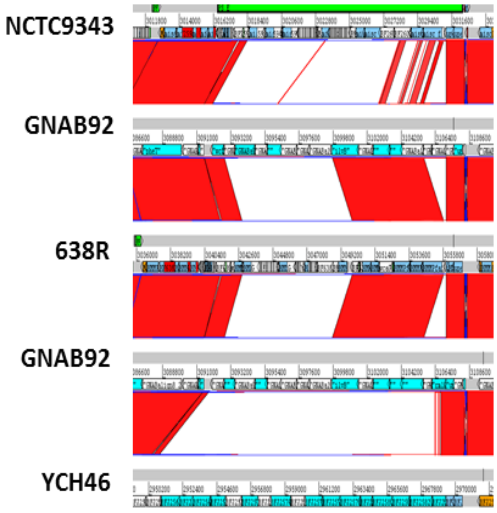
PS-C



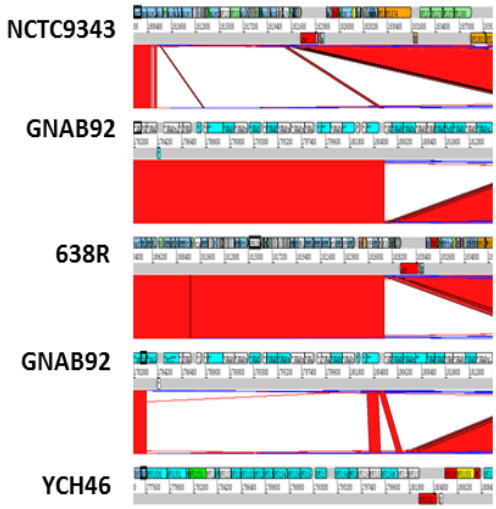
PS-D



PS-E



PS-F



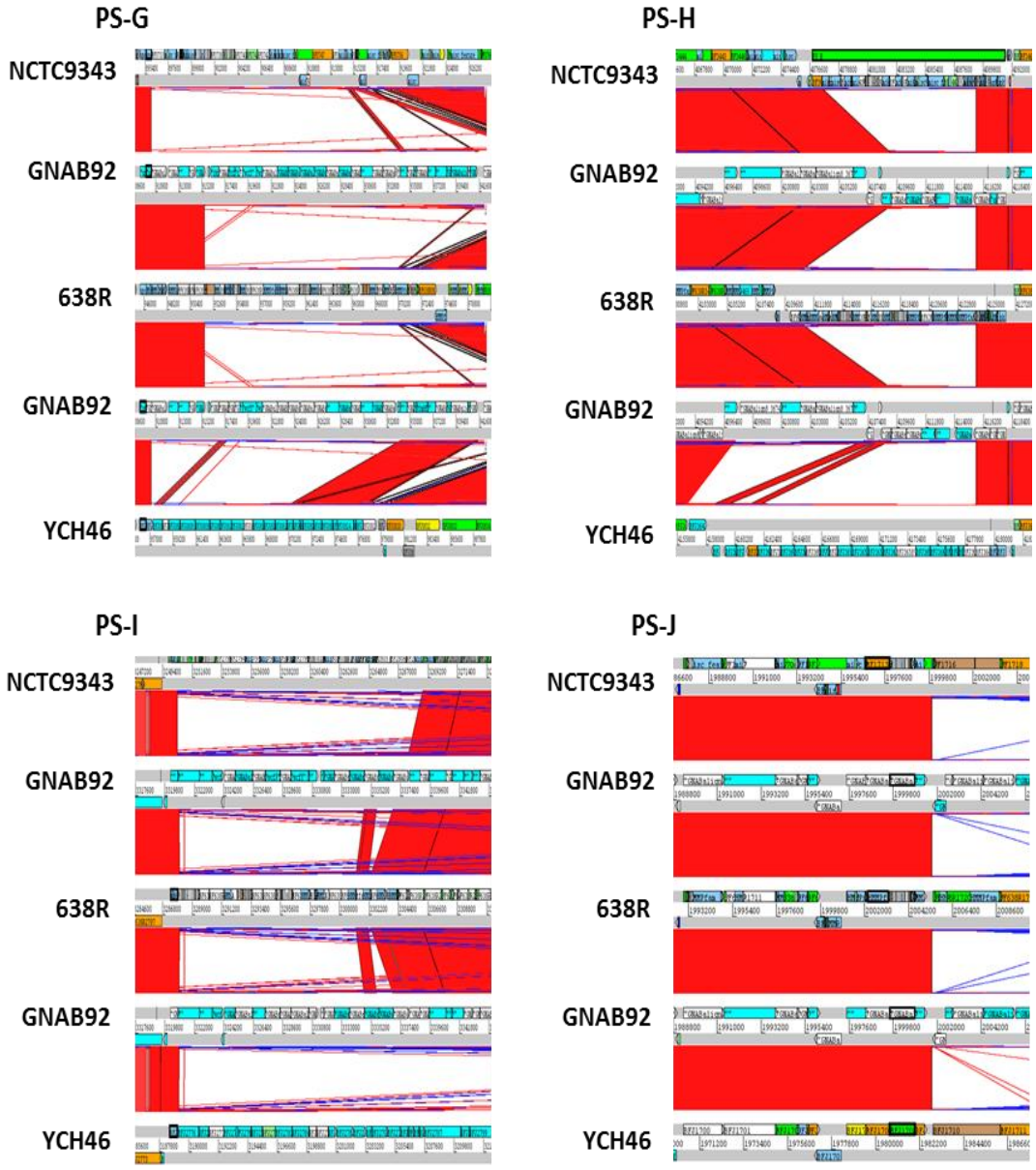


Figure 3.4: Comparison of GNAB92 polysaccharide biosynthesis loci

Five-way ACT genome comparison of the polysaccharide biosynthesis loci of GNAB92 with *B. fragilis* strains NCTC9343, 638R and YCH36. Each polysaccharide is labelled and strain names are shown to the left of the ACT comparison. Sequence identity was evident between GNAB92 and 638R at locus PSF. There is sequence identity between all four strains at locus PSJ.

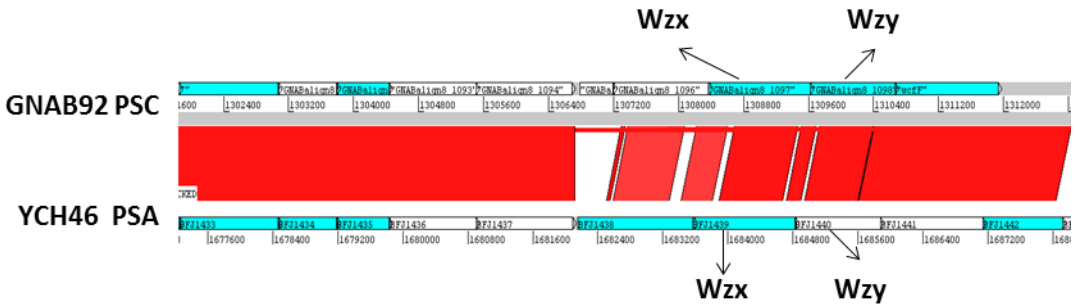


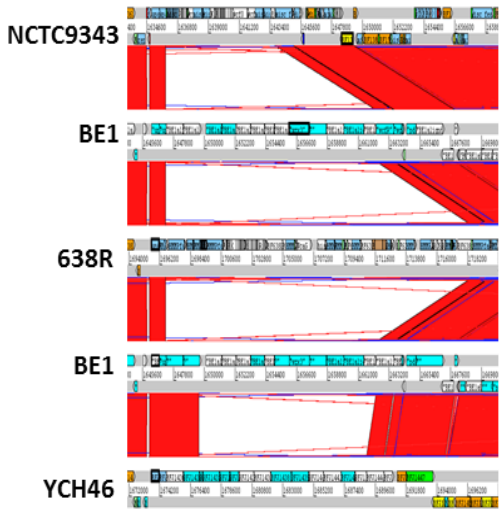
Figure 3.5: ACT comparison of GNAB92 PSC and YCH46 PSA
Two-way ACT comparison of eleven genes identified in GNAB92 PSC, which had homology to YCH46 PSA, including genes encoding Wzx (Similar to *Bacteroides fragilis* YCH46 oligosaccharide repeat unit transporter, E-value = 0.0, 95% identity in 418aa) and Wzy (Similar to *Bacteroides fragilis* YCH46 LPS related polymerase, E-value = 0.0 99% identity in 315aa), suggesting a misassembly in the GNAB92.

3.2.3.3. Strain BE1

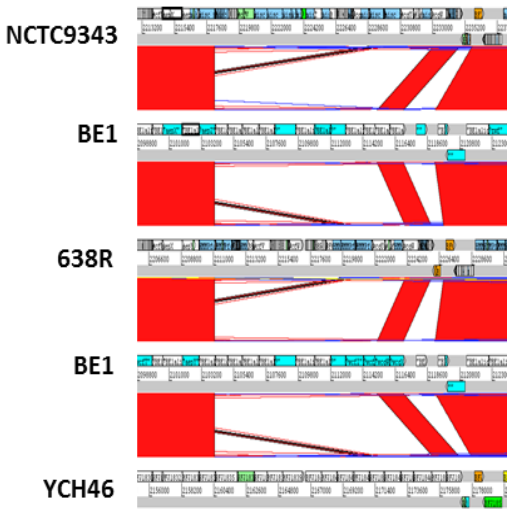
An ACT comparison of BE1 with the three reference genomes indicated that the strain contained ten polysaccharide biosynthesis loci. In comparison to the reference genomes, five of these, PSA, PSB, PSD, PSE and PSG were divergent (Figure 3.6). A sixth locus, PSH contained 99% DNA sequence identity to 638R in eleven of the eighteen genes present in the operon. Due to the number of similar genes, this locus was classified as common with 638R. There was, however, no homology between the *wzx* and *wzy* genes.

BE1 PSF and PSI were common with YCH46 and NCTC9343 respectively, PSJ was similar to all three reference genomes. BE1 PSC had some homology to 638R and YCH46, 99% DNA sequence identity to both strains in the first seven genes of the locus, however, it appeared to contain a deletion in the second half of the locus. The remaining genes of the operon were located elsewhere in the BE1 genome, suggesting a misassembly rather than a deletion of this locus. This indicated that BE1 PSC was a common locus with both YCH46 and 638R. A potential Wzx flippase was encoded in all loci, a potential Wzy polymerase was encoded in eight loci, with the exception of PSD and PSE (Table 3.4).

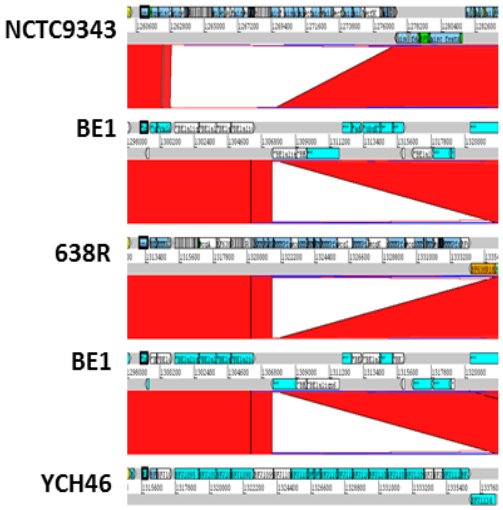
PS-A



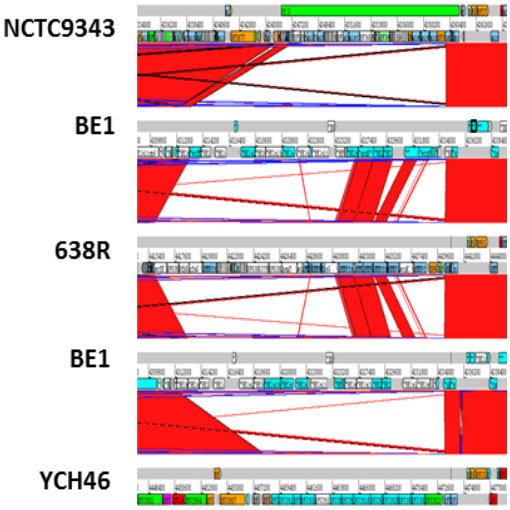
PS-B



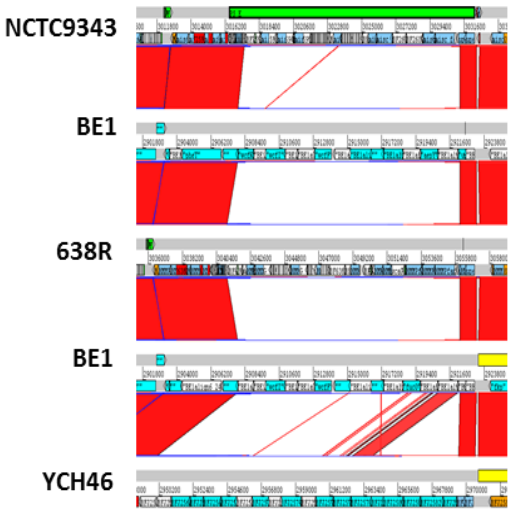
PS-C



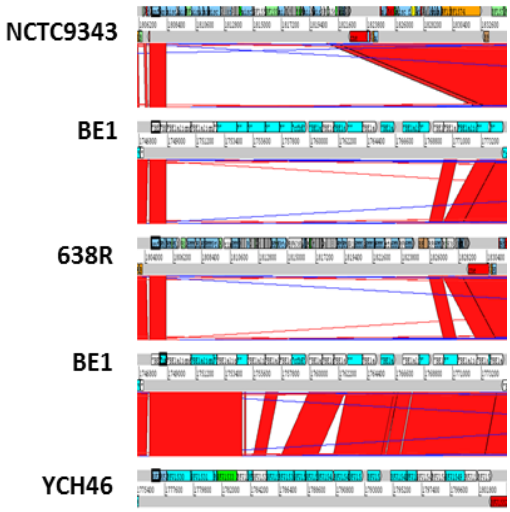
PS-D



PS-E



PS-F



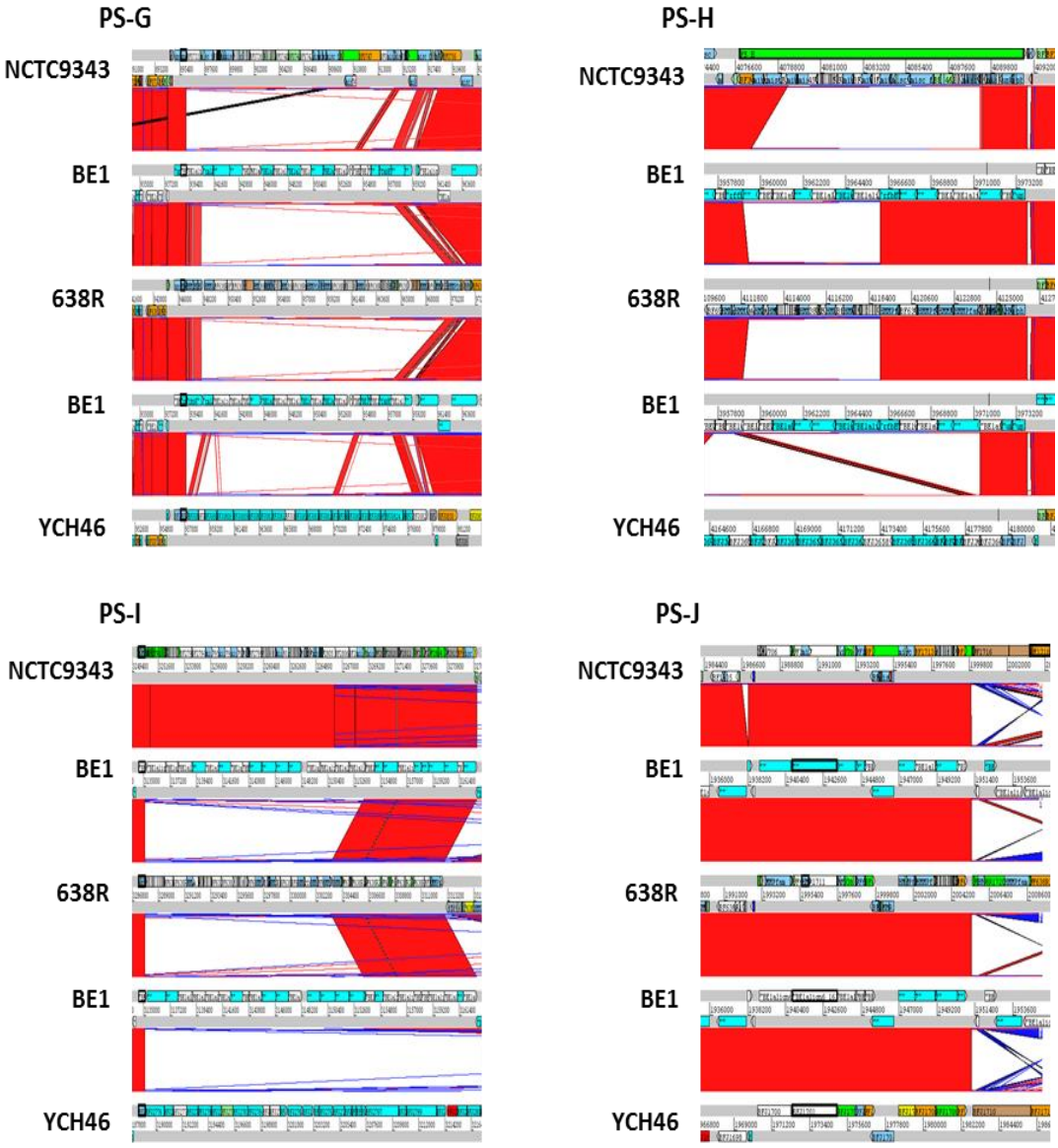


Figure 3.6: Comparison of BE1 polysaccharide biosynthesis loci

Five-way ACT genome comparison of the polysaccharide biosynthesis loci of BE1 with *B. fragilis* strains NCTC9343, 638R and YCH36. Each polysaccharide is labelled and strain names are shown to the left of the ACT comparison. The presence of red colour between sequences indicates sequence identity, no colour indicates divergent sequences and dark blue indicates a DNA inversion. Sequence identity was evident between BE1 and NCTC9343 at locus PSI and with YCH46 at locus PSF. There is sequence identity between all four strains at locus PSJ.

3.2.3.4. Strain LS66

The alignment of LS66 with the reference genomes was more problematic than the other three strains. This was probably due to the assembly of the genome, which contained more contigs and had a smaller N50 value than the other genomes, which made it more difficult

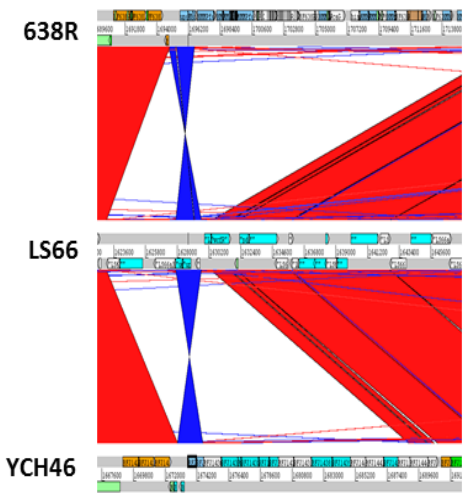
for Mauve to rearrange them against the reference genome. Following alignment of the contigs against NCTC9343 a number of inverted contigs were observed; it is possible that some of these were aligned incorrectly. It was particularly difficult to compare the genome of LS66 to NCTC9343 using the ACT comparison tool, therefore three-way comparisons were carried out with YCH46 and 638R (Figure 3.7).

Five of the LS66 polysaccharide biosynthesis loci, PSA, PSD, PSE, PSF and PSH appeared to contain deletions. Based on the comparisons alone it could not be determined if these were true deletions or a result of improper genome assembly.

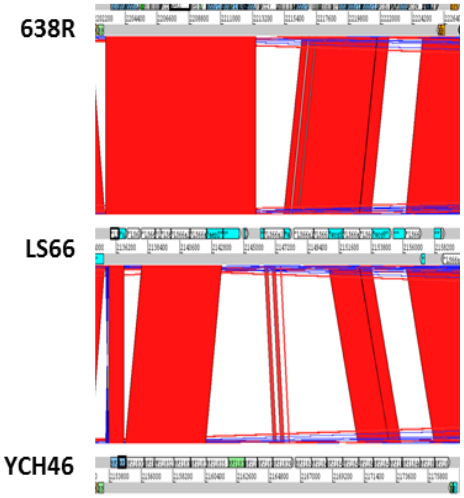
3.2.3.5. Large capsule assembly

Three genes encoding proteins have been implicated in the assembly of the within-strain variable LC of *B. fragilis*: a glycosyltransferase WbaP homologue, a polysaccharide export/assembly Wza homologue and a polysaccharide co-polymerase Wzc homologue (Patrick *et al.*, 2010). The genes encoding these proteins were conserved in RD48, BE1, GNAB92 and LS66, with 99-100% DNA sequence identity to the reference genomes (Figure 3.8).

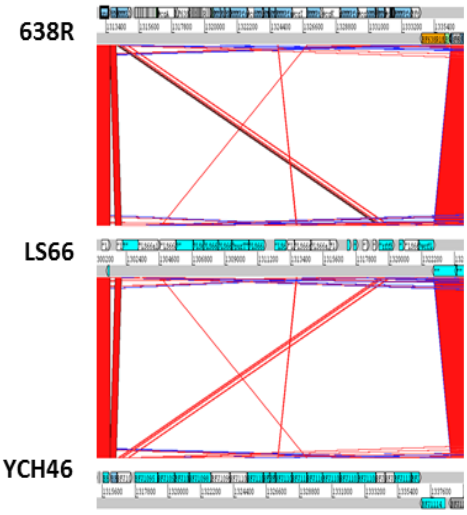
PS-A



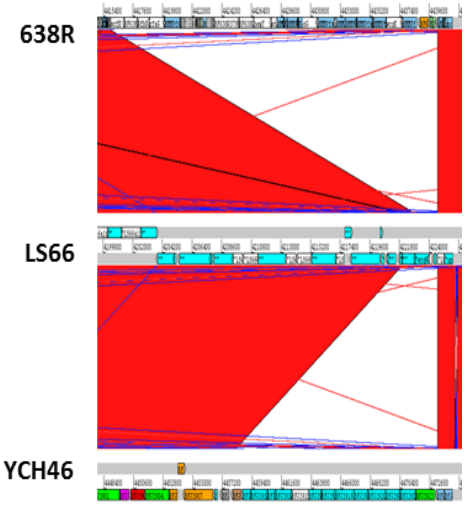
PS-B



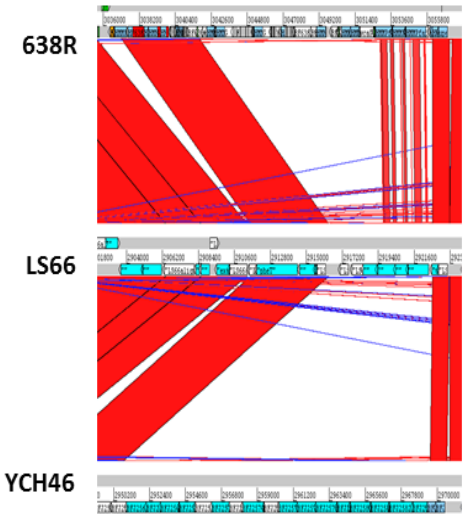
PS-C



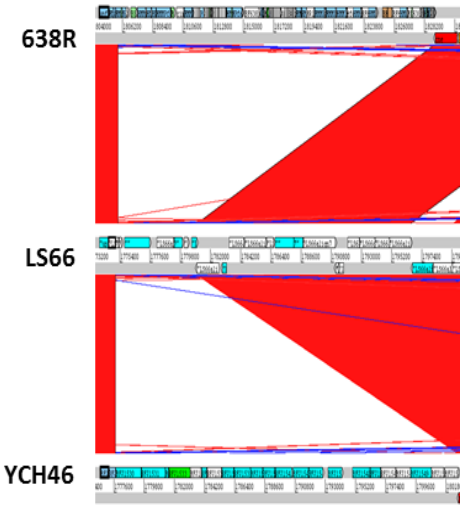
PS-D



PS-E



PS-F



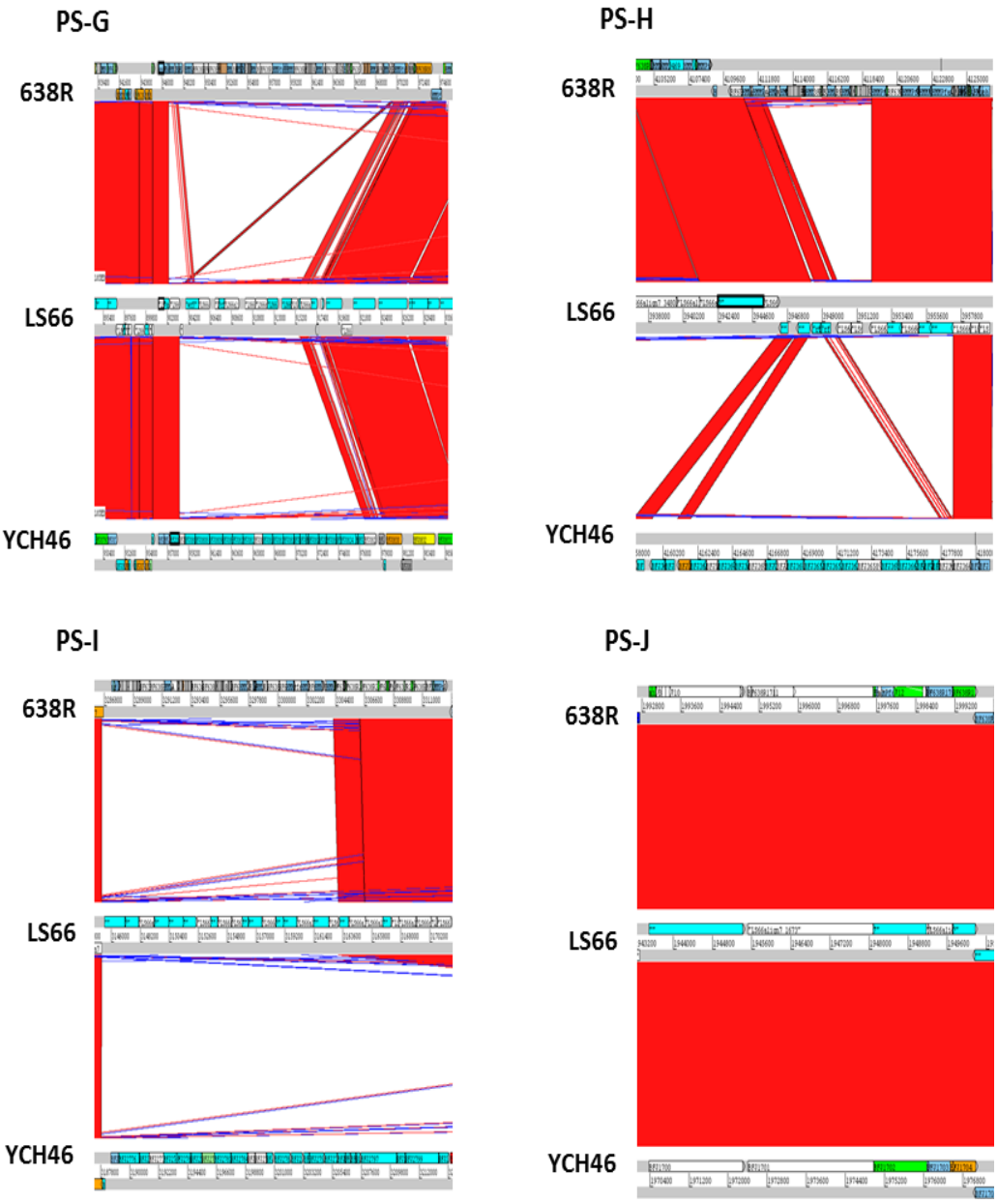


Figure 3.7: Comparison of LS66 polysaccharide biosynthesis loci

Three-way ACT genome comparison of the polysaccharide biosynthesis loci of LS66 with *B. fragilis* strains 638R and YCH36. Each polysaccharide is labelled and strain names are shown to the left of the ACT comparison. The presence of red colour between sequences indicates sequence identity, no colour indicates divergent sequences and dark blue indicates a DNA inversion. There was some sequence identity between the three strains at PSB. Sequence identity was evident between all three strains at locus PSJ.

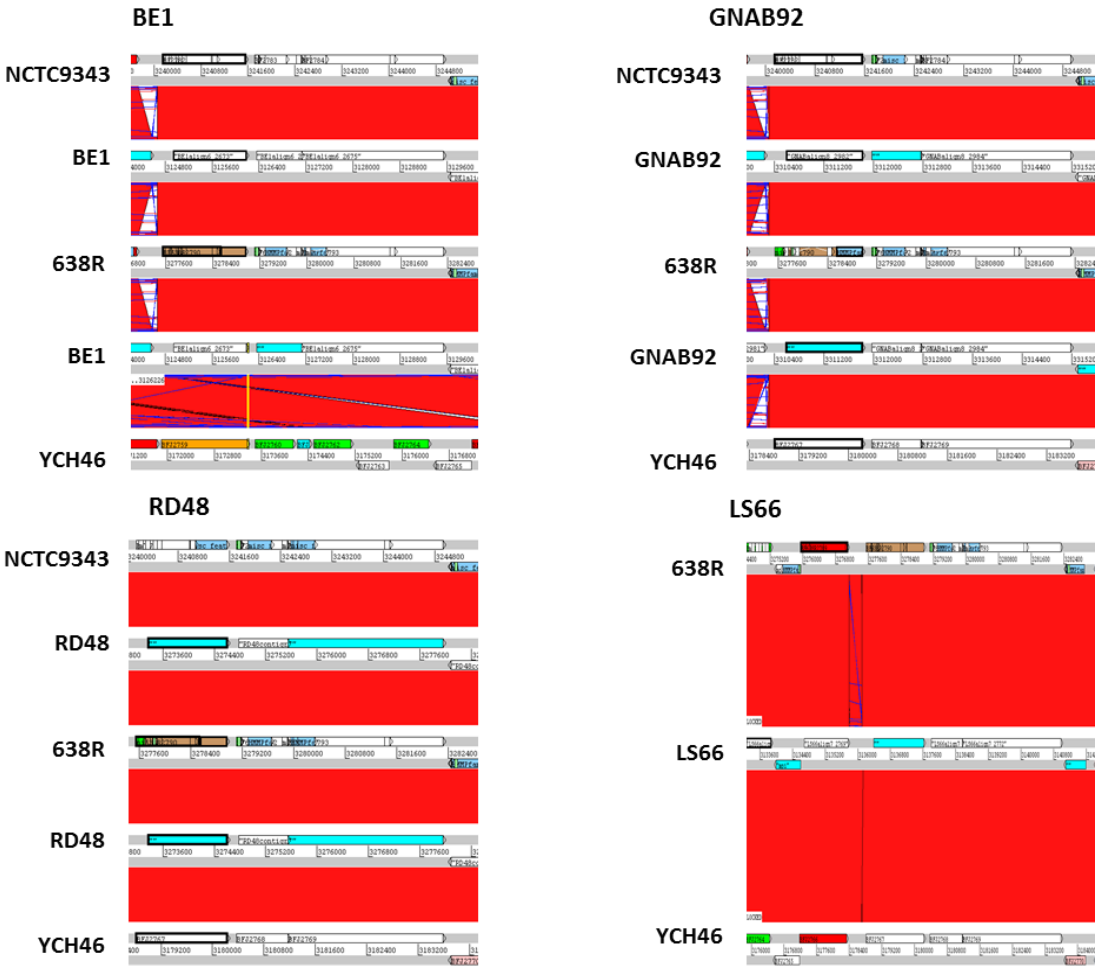


Figure 3.8: Comparison of the LC assembly genes

Five-way ACT comparison of the large capsule assembly loci of GNAB92, BE1 and RD48 with *B. fragilis* NCTC9343, 638R and YCH46. Three-way comparison of LS66 with 638R and YCH46. All seven genomes share homology within the locus.

3.2.3.6. Identifying Wzx and Wzy homologues

The sixteen polysaccharide biosynthesis loci which are implicated in MC formation in NCTC9343 and 638R were characterised by the presence of O-antigen flippase encoding genes (*wzx*) and polysaccharide polymerase encoding genes (*wzy*), which indicate that export of *B. fragilis* polysaccharides is by a Wzy-dependent pathway (Whitfield, 2006a; Patrick *et al.*, 2010). Although Wzx and Wzy proteins are found in a wide range of bacteria, there is very little conservation of amino acid sequences. Wzx proteins are predicted to contain 10-14 transmembrane helices and a loosely conserved polysaccharide biosynthesis protein domain (Polysacc_synt domain Protein families database of alignments and HMMs) (Marolda *et al.*, 2004; Islam & Lam, 2013). Wzy proteins are predicted to contain 8-12

transmembrane helices and some contain the O-antigen polymerase conserved domain Wzy_C (Daniels *et al.*, 1998; Samuel & Reeves, 2003).

For the fifteen potentially new polysaccharide biosynthesis loci, all genes encoding proteins without an annotated function were examined to identify potential Wzx and Wzy homologues. Amino acid sequence comparisons were carried out using the Basic Local Alignment Search Tool (BLAST). Conserved domains was identified using the Pfam protein families database, and transmembrane helices were predicted using the TMHMM (Transmembrane Hidden Markov Model) v 2.0 Server (Sonnhammer *et al.*, 1998; Krogh *et al.*, 2001). These results are summarised in Table 3.4. The Wzx and Wzy proteins of RD48 were not included as all polysaccharide biosynthesis loci present in this strain were homologous to NCTC9343.

Genes encoding potential Wzx flippases were identified in all loci present in BE1, and potential Wzy polymerases were identified in all loci with the exception of PSD and PSE. All of the putative Wzx proteins were predicted to contain 10-14 transmembrane helices, with seven of the nine proteins containing homology to the Polysacc_synt family. Six of these proteins, located in PSA, PSB, PSC, PSF, PSG and PSI had 32-100% amino acid identity to previously identified *B. fragilis* flippases. The seven potential Wzy proteins of BE1 were predicted to contain 8-14 transmembrane helices. This is consistent with published data in which identified Wzy proteins contain 8-14 transmembrane domains (Daniels *et al.*, 1998).

All of the polysaccharide biosynthesis loci of GNAB92 encoded proteins with identity to Wzx flippases and Wzy polymerases. However, the putative polymerase of PSB (CDS 1931) had very low homology to any other polymerases and only contained 4 predicted transmembrane regions suggesting it was not a Wzy homologue. There were no other genes within this locus with homology to an O-antigen polymerase and the locus contained no genes which encoded a hypothetical protein.

Within the five intact polysaccharide biosynthesis loci of LS66, only one potential Wzx protein was identified. This protein had 11 predicted transmembrane helices, and 99% identity to a previously characterised *B. fragilis* flippase in 638R. The lack of Wzx and Wzy homologues identified in LS66 may be due to the misassemblies present in the genome or to the poor amino acid sequence conservation within these protein families.

Table 3.4: List of potential Wzx and Wzy proteins

Strain and Locus	Wzx/Wzy	CDS	Homology	Identity	E value	Protein length (aa)	TMH*	Pfam domain search
GNAB92 PSA	Wzx	1097	oligosaccharide repeat unit transporter [Bacteroides fragilis YCH46]	398/418(95%)	0.0	418	12	Polysacc_synt
GNAB92 PSA	Wzy	1098	polysaccharide polymerase [Clostridium perfringens ATCC 13124]	84/336(25%)	3e-12	345	10	EspG
GNAB92 PSB	Wzx	1924	flippase [Bacteroides fragilis YCH46]	86/364(24%)	6e-16	506	14	Polysacc_synt
GNAB92 PSB	Wzy	1931	polysaccharide polymerase [Oceanicola sp. S124]	38/129(29%)	0.036	257	4	No hits
GNAB92 PSC	Wzx	1081	flippase [Bacteroides fragilis YCH46]	507/509(99%)	0.0	509	14	Polysacc_synt_3
GNAB92 PSC	Wzy	1084	polymerase [Bacteroides fragilis YCH46]	477/478(99%)	0.0	478	12	O-ag_pol_Wzy
GNAB92 PSE	Wzx	2787	O-antigen flippase [Yersinia similis]	127/368(35%)	5e-56	447	12	Polysacc_synt
GNAB92 PSE	Wzy	2782	O-antigen polymerase [Ignavibacterium album JCM 16511]	40/131(31%)	1.8	431	10	O-ag_pol_Wzy
GNAB92 PSF	Wzx	1560	putative transporter/flippase [Bacteroides fragilis 638R]	475/475(100%)	0.0	475	11	Polysacc_synt
GNAB92 PSF	Wzy	1566	polysaccharide polymerase Eps11O [Lactobacillus paraplantarum]	63/235(27%)	0.027	382	9	EspG
GNAB92 PSG	Wzx	782	O-antigen flippase [Escherichia coli]	125/439(28%)	1e-42	471	11	Polysacc_synt
GNAB92 PSG	Wzy	788	lipid A core-O-antigen ligase-like enyme [Alistipes sp. CAG:831]	69/250(28%)	4e-08	394	8	No hits
GNAB92 PSH	Wzx	3687	O-antigen flippase [Escherichia coli]	97/357(27%)	3e-27	375	8	No hits
GNAB92 PSH	Wzy	3683	O-antigen polymerase like protein [Yersinia bercovieri]	89/334(27%)	2e-14	369	8	O-antigen_lig
GNAB92 PSI	Wzx	2991	flippase [Bacteroides fragilis YCH46]	278/512(54%)	0.0	513	12	Polysacc_synt
GNAB92 PSI	Wzy	2996	O-antigen polymerase [Bacillus megaterium WSH-002]	48/184(26%)	0.16	425	10	Wzy_C

Strain and Locus	Wzx/Wzy	CDS	Homology	Identity	E value	Protein length (aa)	TMH*	Pfam domain search
BE1 PSA	Wzx	1360	LPS biosynthesis flippase [Bacteroides fragilis NCTC 9343]	135/426(32%)	6e-58	474	14	Polysacc_synt
BE1 PSA	Wzy	1363	Wzy [Escherichia coli]	72/253(28%)	5e-07	464	14	O-ag_pol_Wzy
BE1 PSB	Wzx	1750	capsular polysaccharide transporter [Bacteroides fragilis YCH46]	181/482(38%)	3e-93	481	11	Polysacc_synt
BE1 PSB	Wzy	1751	polysaccharide polymerase [Bacillus cereus]	48/190(25%)	0.80	428	12	O-antigen_lig
BE1 PSC	Wzx	1049	flippase [Bacteroides fragilis YCH46]	509/509(100%)	0.0	509	14	Polysacc_synt_3
BE1 PSC	Wzy	1052	Wzy [Escherichia coli]	101/402(25%)	3e-06	478	12	O-ag_pol_Wzy
BE1 PSD	Wzx	3759	O-antigen export [Clostridium botulinum H04402 065]	63/271(23%)	2e-04	480	13	Polysacc_synt
BE1 PSE	Wzx	2485	O-antigen transporter [Parabacteroides distasonis ATCC 8503]	132/394(34%)	2e-53	414	10	Polysacc_synt
BE1 PSF	Wzx	1453	hypothetical protein BF1533 [Bacteroides fragilis YCH46]	507/509(99%)	0.0	509	12	MatE
BE1 PSF	Wzy	1465	O-antigen polymerase [Paenibacillus sp. JC66]	77/293(26%)	3e-07	419	9	O-antigen_lig
BE1 PSG	Wzx	0760	capsular polysaccharide transporter [Bacteroides fragilis YCH46]	477/481(99%)	0.0	481	12	Polysacc_synt
BE1 PSG	Wzy	0766	polymerase [Clostridium thermocellum]	50/202(25%)	0.11	394	8	O-antigen_lig
BE1 PSH	Wzx	3426	O-antigen flippase [Yersinia similis]	149/432(34%)	7e-72	454	12	Polysacc_synt
BE1 PSH	Wzy	3422	O-antigen polymerase [Pectobacterium atrosepticum SCRI1043]	87/323(27%)	5e-08	369	8	No hits
BE1 PSI	Wzx	2682	hypothetical protein BF2791 [Bacteroides fragilis NCTC 9343]	514/514(100%)	0.0	514	12	MatE
BE1 PSI	Wzy	2691	Wzy [Escherichia coli]	52/203(26%)	0.12	368	8	No hits
LS66 PSG	Wzx	0741	polysaccharide transporter/flippase [Bacteroides fragilis 638R]	472/475(99%)	0.0	475	11	Polysacc_synt

*Transmembrane helices

3.2.4. Whole genome comparison

RD48, BE1, GNAB92, BE1, YCH46 and 638R were aligned to NCTC9343 using the BLAST Ring Image Generator (BRIG) (Figure 3.9). The locations of the polysaccharide biosynthesis loci of NCTC9343 are indicated on the outer ring of the comparison. The absence of colour within the rings representing the genomes of RD48, BE1, GNAB2, 638R and YCH46 indicated regions which were divergent from NCTC9343. For example, there was an absence of colour in the 638R (dark blue), YCH46 (green) and GNAB92 (light blue) rings in the region which encoded PSI, as the equivalent polysaccharide biosynthesis locus in these three genomes had little sequence identity to NCTC9343. However, there was an unbroken ring of colour in the same region of BE1 (purple), LS66 (yellow) and RD48 (light green) as these strains had 99-100% DNA sequence identity with NCTC9343 PSI.

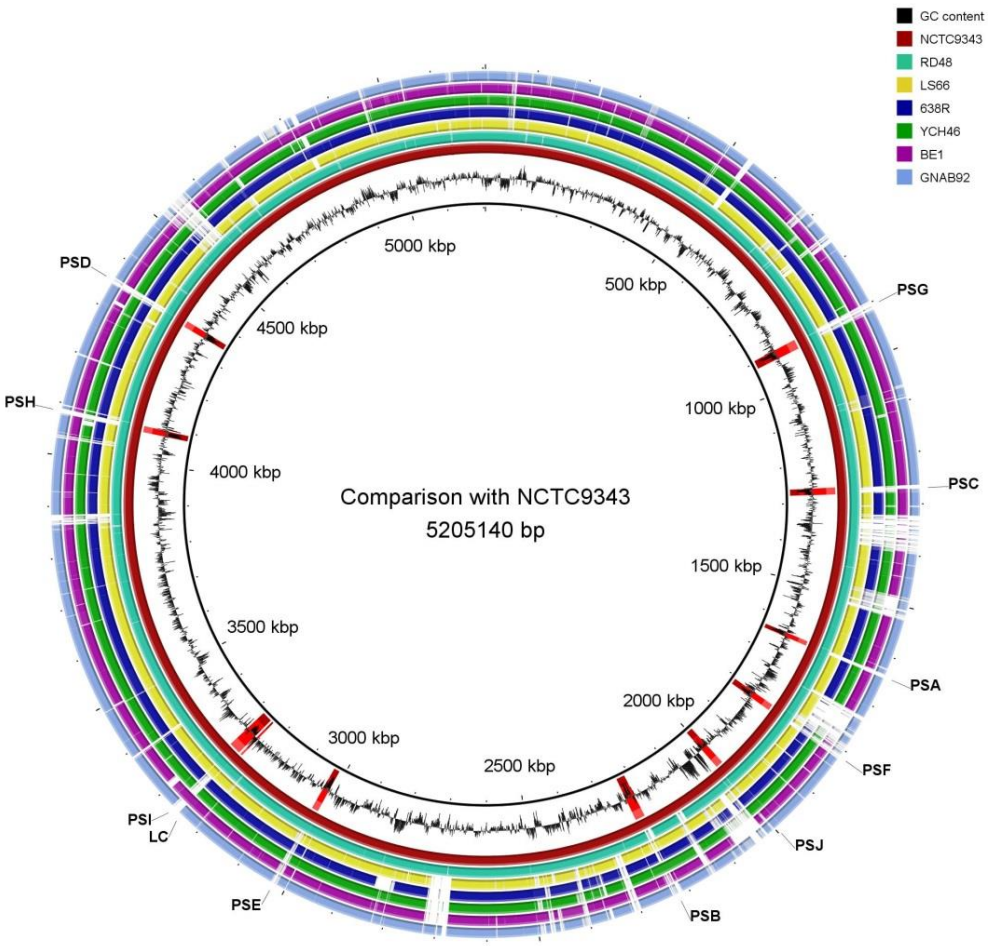


Figure 3.9: Whole genome comparison with NCTC9343
The genomes of RD48, LS66, 638R, YCH46, BE1 and GNB92 were compared to NCTC9343 using BRIG. The inner ring represents the GC content of NCTC9343, each consecutive ring represents an individual genome identified by the colour key. The loci of polysaccharides A-I and the LC assembly genes are indicated in red in the inner ring and are labelled on the outer ring. The presence of colour indicates homology with NCTC9343, no colour indicates divergent sequences.

3.2.5. DNA Restriction and Modification systems

B. fragilis contains R-M systems, which protect the cell from incoming foreign DNA. NCTC9343 encodes three type I and two type III R-M systems; in comparison, 638R encodes two type I, one type IIS and one type III R-M systems (Cerdeño-Tárraga *et al.*, 2005; Patrick *et al.*, 2010). The genome of YCH46 encodes one type IIS and one type III R-M systems. The presence of R-M systems in BE1, RD48 and GNAB92 were determined and comparisons were made to NCTC9343, 638R and YCH46.

Patrick *et al* (2010) demonstrated that both type I R-M systems present in 638R were contained within a horizontally acquired island. A region, displaying little sequence identity between strains, was present in all five genomes and in a similar chromosomal location (Figure 3.10a). This region ranged in size from ~24 kb in YCH46 to ~ 74 kb in NCTC9343, and was flanked by bacteriophage related genes, including a gene encoding a putative phage integrase. Interestingly, this region encoded a type I R-M system in all strains with the exception of YCH46. The 638R *hsdS* shufflon (BF 638R 1146-1149) was located within this region, a similar type I shufflon was also encoded in the same location in GNAB92 (1165-1168) (Figure 3.10b). In NCTC9343 the region also encoded one of the type III R-M systems.

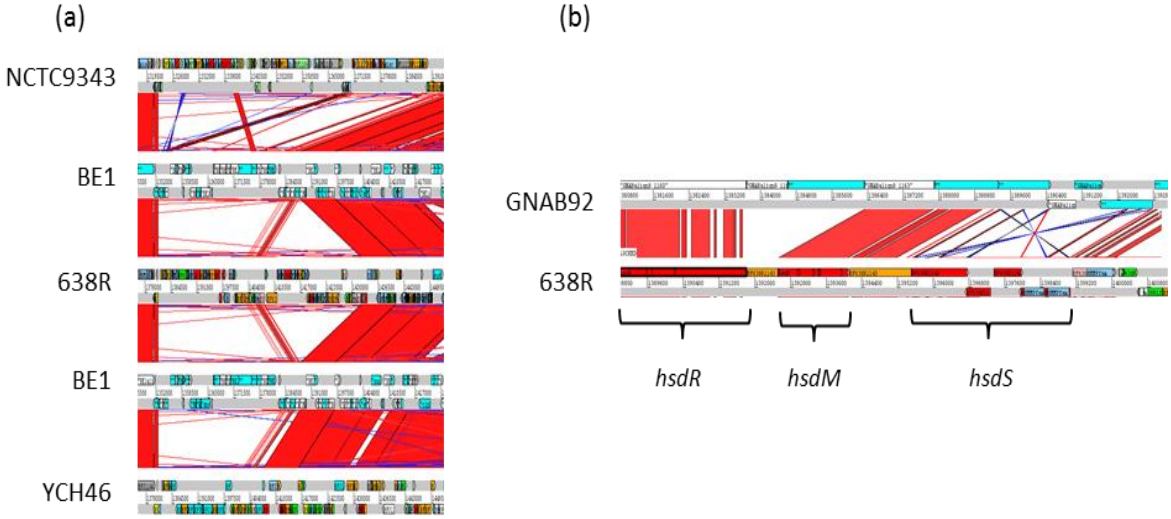


Figure 3.10: Restriction modification related region
(a) Five-way ACT comparison between NCTC9343, BE1, 638R and YCH46 showing a phage-related, divergent region present in all genomes containing R-M systems. (b) Two-way ACT comparison of the type I R-M shufflon present in GNAB92 and 638R. The red colour indicates homology.

In addition to the type I shufflon encoded by GNAB92 there was a second type I R-M system (4017-4019), which was also associated with a phage related integrase encoding gene (4014), and a type III R-M system (1132-1133). RD48 contained the same R-M systems present in NCTC9343, however, the type I shufflon of RD48 appeared to contain a deletion of both the HsdM and HsdR subunits (Figure 3.11a), with only the specificity subunits remaining (RD48 1891-1893), suggesting this was a non-functional R-M system. A second type I system in RD48 (RD48 1510-1511), appeared to contain a deletion in the HsdS subunit (Figure 3.11b). A type IIS enzyme (BE1 4482), related to those encoded by 638R (1811) and YCH46 (3008) was identified in BE1, along with a type III R-M system. This type

III locus was the only R-M system common to all six of the *B. fragilis* genomes. This suggests that diversity observed in the genes encoding the other R-M systems was generated by horizontal gene transfer.

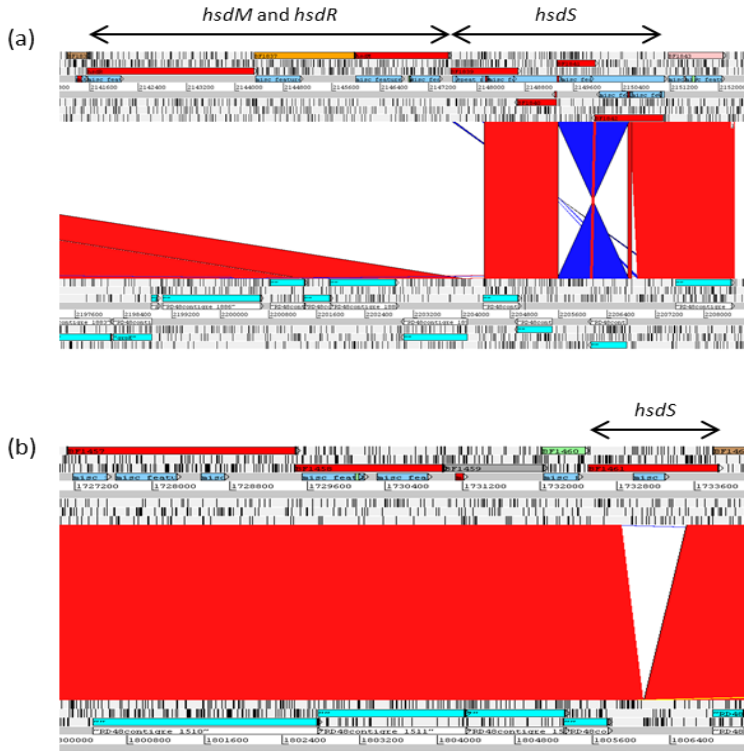


Figure 3.11: Deletions in the type I R-M systems of RD48
ACT comparisons of NCTC9343 (top) and RD48 (bottom) showing (a) a deletion of the *hsdR* and *hsdM* genes of the type I R-M shufflon. (b) a deletion of the hsdS gene of the type I R-M system equivalent to NCTC9343 1077-1079. The red colour indicates homology and the blue colour indicates a sequence inversion.

3.2.6. *Bacteroides fragilis* toxin (BFT)

A subset of *B. fragilis* strains, termed enterotoxigenic *B. fragilis*, contain the virulence factor BFT (Franco *et al.*, 1997). The *bft* gene is contained within an ~ 6kb pathogenicity island, BfPAI, which is inserted at a specific chromosomal location in EBFT strains (Moncrief *et al.*, 1998). BfPAI has ~12 kb flanking regions which contains genes encoding mobilisation proteins and this region has been identified in approximately half of the non-enterotoxigenic *B. fragilis* strains tested including NCTC9343 (Franco *et al.*, 1999). Franco (2004) determined that the BfPAI flanking region of NCTC9343 is contained within a ~65 kb

conjugative transposon (CTn9343). A similar transposon, designated CTn86, was identified in the ETBF 86-5443-2-2 strain which contains BfPAI.

The presence of BfPAI and the conjugative transposon in GNAB92, BE1, RD48, YCH46 and 638R were determined. GNAB92 encoded the BFT pathogenicity locus (Figure 3.12), with 98% DNA sequence identity to BfPAI isolated from ETBF VPI13784 (GenBank accession AF038459). GNAB92 also contained a ~ 61.5 kb region with 85% identity to CTn9343. RD48 encoded the complete CTn9343 transposon but did not contain the BfPAI. None of the other genomes encoded BfPAI or the conjugative transposon.

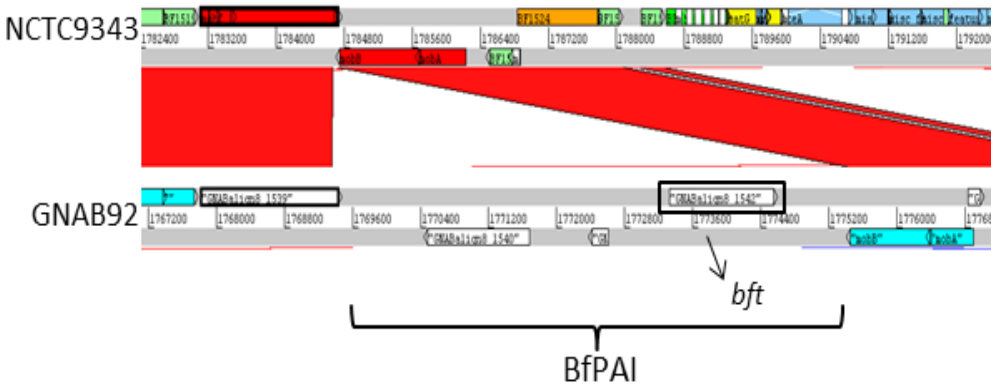


Figure 3.12: ACT comparison of BfPAI pathogenicity island in GNAB92 and NCTC9343
GNAB92 encoded the BfPAI pathogenicity island which contained *bft*, highlighted by a black box. The locus was flanked by mobilisation genes, conserved in NCTC9343.

3.3 Conclusion

The genome of *B. fragilis* shows evidence of extensive horizontal gene transfer. The genomes of seven *B. fragilis* clinical isolates were sequenced to further examine the role of HGT in genetic diversity within this organism. Following sequencing, the quality of the data was assessed and three of the genomes were deemed unsuitable for further analysis. The data of the four remaining strains, RD48, LS66, BE1 and GNAB92 were assembled and annotated. Analysis of these genomes focused on three previously identified examples of HGT within *B. fragilis*: polysaccharide diversity, R-M systems and *B. fragilis t*oxin.

Patrick *et al* (2010) demonstrated that the genomes of *B. fragilis* strains NCTC9343, 638R and YCH46 encoded 28 divergent capsular polysaccharide-associated biosynthesis loci. Genomic comparisons were performed between these strains and RD48, LS66, BE1 and GNAB92. The genome sequences of GNAB92 and BE1 contained eleven polysaccharide biosynthesis loci which were divergent from those previously identified in NCTC9343, YCH36 and 638R. One of these loci, PSG, was a common locus. Therefore within the two strains, ten different polysaccharide biosynthesis loci were found. There is, consequently, the potential to express 38 different polysaccharides amongst these five strains. A further three possible divergent polysaccharides were identified in LS66. However, due to problems with the assembly of the genome this could not be confirmed. RD48 encoded nine polysaccharide biosynthesis loci identical to those previously identified in NCTC9343. The diversity of polysaccharides, observed in so few strains of *B. fragilis* is unprecedented. For example, in *Streptococcus pneumoniae*, 90 different polysaccharide loci have been identified, however, each single locus is associated with a single serotype (Bentley *et al.*, 2006). Similarly in *Acinetobacter baumanii*, 25 polysaccharide loci have been identified in 27 serotypes, but each serotype contains only one locus (Hu *et al.*, 2013).

A number of divergent R-M systems were identified within the analysed genomes. GNAB92 contained two type I R-M systems, one of which contained a shufflon with homology to a type I shufflon encoded by 638R, and a type III R-M. A type IIS and a type III R-M system were identified in BE1. This type III locus was the only conserved R-M system present in all strains analysed. The presence of only one common system suggests that the diversity of the R-M systems within *B. fragilis* strains was the result of horizontal gene transfer.

During analysis of the four newly sequenced genomes, a number of apparent deletions in polysaccharide loci were observed, one of which was a deletion of RD48 PSB. Immunofluorescence microscopy revealed that PSB was expressed by this strain, indicating the observed deletion was an artefact of the genome assembly. The other putative deletions should be investigated to determine whether they are true deletions. This could be achieved through the use of a sequencing platform which produces longer read lengths, which would result in fewer contigs and reduce the chance of a misassembly.

In conclusion, four new *B. fragilis* genomes were sequenced. The genomes displayed a high level of HGT that had occurred despite the presence of multiple, diverse restriction-modification systems.

Chapter 4. Inter-kingdom horizontal gene transfer

4.1. Introduction

An unusual example of HGT in *Bacteroides fragilis* was observed when the complete genome of *B. fragilis* NCTC9343 was sequenced and analysed. An open reading frame was identified which encodes a 76 amino acid protein with 63% identity to human ubiquitin. The protein-modifier is conserved within eukaryotes but, until now, was believed to be absent from prokaryotes. The closest DNA sequence homology to this ORF is found in a migratory grasshopper entomopox virus, which suggests the gene was acquired by inter-kingdom HGT (Patrick *et al.*, 2011).

4.1.1. Ubiquitin

Ubiquitin is a 76 amino acid, highly conserved, eukaryotic protein-modifier. Tagging of target proteins with ubiquitin (ubiquitylation) is essential to a wide range of cellular functions, including DNA repair, endocytosis, transcription, apoptosis and immune response. The most studied role of ubiquitylation is in the regulation of proteolysis, in which a target protein is marked for degradation by the 26S proteasome**.** This process is facilitated by three enzymes, E1, E2 and E3. Ubiquitin is activated by the formation of a thioester bond between the C-terminal glycine residue and the cysteine active site of the ubiquitin-activating enzyme, E1. The bound ubiquitin is subsequently transferred to the cysteine residue in the active site of the ubiquitin-conjugating enzyme, E2. An isopeptide bond is formed between ubiquitin and a lysine residue of the target protein by the ubiquitin ligase, E3. Polyubiquitin chains can form on the target protein. In the case of proteolysis, Lys48-linked chains are recognised and the protein is degraded by the 26S proteasome (Figure 4.1) (Komander & Rape, 2012).

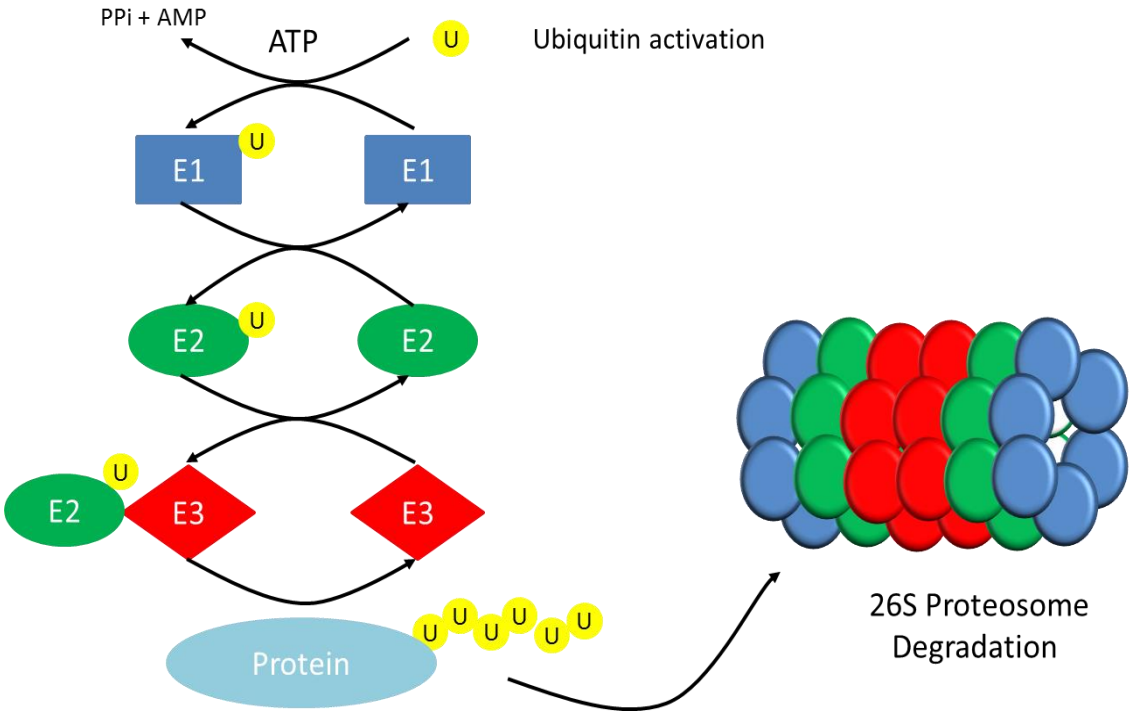


Figure 4.1: The ubiquitylation pathway.
The activation of ubiquitin occurs by the formation of a thioester bond between the C-terminal glycine residue and the cysteine active site of the ubiquitin-activating enzyme. Ubiquitin is then transferred to the active site of the E2, conjugating enzyme before E3 facilitates binding to the target protein. Lys48 linked polyubiquitin chains are recognised by the proteasome and the protein is degraded.

4.1.2. Ubiquitin modifications

Modifications of proteins can occur with one ubiquitin molecule (monoubiquitylation), multiple ubiquitin monomers attached at different lysine residues on the target protein (multiubiquitylation) or as a chain (polyubiquitylation) (Sadowski *et al.*, 2012). Polyubiquitin chains are linked through one of the seven lysine residues; Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 and Lys63. The formation of polyubiquitin chains can be homogeneous where all linkages are formed through the same lysine residue, or chains can contain a mixture of linkage types (Peng *et al.*, 2003). The type of ubiquitin modification of the target protein determines its fate; for example, monoubiquitylation is involved in regulating endocytosis, lysosomal targeting and meiosis, while polyubiquitylation is implicated in protein degradation, DNA repair and immune regulation (Haglund *et al.*, 2003; Hicke & Dunn, 2003).

4.1.3. Ubiquitin and the immune system

Ubiquitylation plays a major role in the regulation of the host pro-inflammatory response through the activation of the transcription factor, nuclear factor kappa B (NF-κB), which occurs in response to a variety of stimuli including invading pathogens. In the absence of a stimulus such as cytokines, NF-κB is present in the cytoplasm of the cell and is inactive, due to its association with inhibitor IκBα. During infection, pathogen associated molecular patterns (PAMPs) of extracellular pathogens are recognised by pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs) and NOD-like receptors, which initiates the NF-κB signalling cascade (Oeckinghaus *et al.*, 2011). This cascade results in the ubiquitylation and subsequent degradation of IκBα, which releases NF-κB, allowing it to translocate to the nucleus and initiate transcription of pro-inflammatory response genes (Schmukle & Walczak, 2012). Traditionally Lys48 and Lys63 linked polyubiquitin chains have been implicated in the regulation of the pathway. Recently however, Lys11 linked chains have been identified as playing a role (Dynek *et al.*, 2010; Bremm & Komander, 2011). Lys48 linked chains mediate the essential degradation of a number of components in this pathway. For example, following the phosphorylation of IκBα, these proteins are ubiquitylated with Lys48 chains and are targeted for proteolytic destruction. Lys63 chains are involved in the activation of NF-κB by modifying multiple components of the pathway upstream of IκBα phosphorylation, for example TRAF2 and NEMO, two proteins which are required for the recruitment of IKK (Chen & Chen, 2013). Homogeneous Lys11 linked chains are involved in the degradation of a number of cell cycle regulators, however, Lys11 linkages present in mixed polyubiquitin chains have been identified in many other cell processes including the activation of NF-κB via the tumor necrosis factor-α (TNF-α) pathway. Receptor interacting protein 1 (RIP1) is ubiquitylated by polyubiquitin chains containing Lys11 linkages after stimulation of the cell by a cytokine such as TNF-α (Bremm & Komander, 2011; Schmukle & Walczak, 2012).

Lys11 linkages have also been identified in the endocytosis and trafficking of transmembrane receptors, for example Kaposi's sarcoma associated herpesvirus produces two immunomodulatory proteins, K3 and K5, which can down-regulate MHC class 1 molecules on the host cell. These molecules are responsible for presenting peptides to cytotoxic T-cells which alert the immune system to the presence of intracellular pathogens. K3 causes Lys63-linked polyubiquitylation of MCH 1, however K5 mediated endocytosis of

MCH I requires polyubiquitin chains consisting of both Lys63 and Lys11 linkages. K3 and K5 act as ubiquitin E3 ligases and facilitate the polyubiquitylation of MHC 1 which targets it for degradation (Boname *et al.*, 2010).

All of the other possible lysine linkages, Lys6, Lys27, Lys29 and Lys33 have been detected in cells but the processes in which they are involved are less well defined (Komander & Rape, 2012)

4.1.4. Ubiquitin Binding Domains

Ubiquitin acts as a signalling molecule which initiates cellular events often by forming non-covalent interactions with proteins containing ubiquitin-binding domains (UBDs), of which there are estimated to be more than 150 different types (Hurley *et al.*, 2006; Dikic *et al.*, 2009). UBDs interact with different ubiquitin surface residues (Figure 4.2), the most common of which is the hydrophobic Ile44 patch consisting of the residues Ile44, Leu8, Val70 and His68. This patch binds most UBDs and is bound by the proteasome during degradation of substrates. Another hydrophobic patch involved in UBD interactions includes the residues Ile36, Leu71 and Leu73, which are recognised by deubiquitylating enzymes (DUBs) and HECT E3s (Komander & Rape, 2012). The Phe4 patch, consisting of residues Gln2, Phe4 and Thr12, is involved in cell division in yeast but is not required for ubiquitin conjugation or proteasome degradation (Sloper-Mould *et al.*, 2001). The TEK-box of ubiquitin, which consists of the residues Thr12, Thr14, Glu34, Lys6 and Lys11, is required for mitotic degradation and UBE2C- and UBE2S-mediated assembly of Lys11-linked chains (Jin *et al.*, 2008; Kulathu & Komander, 2012).

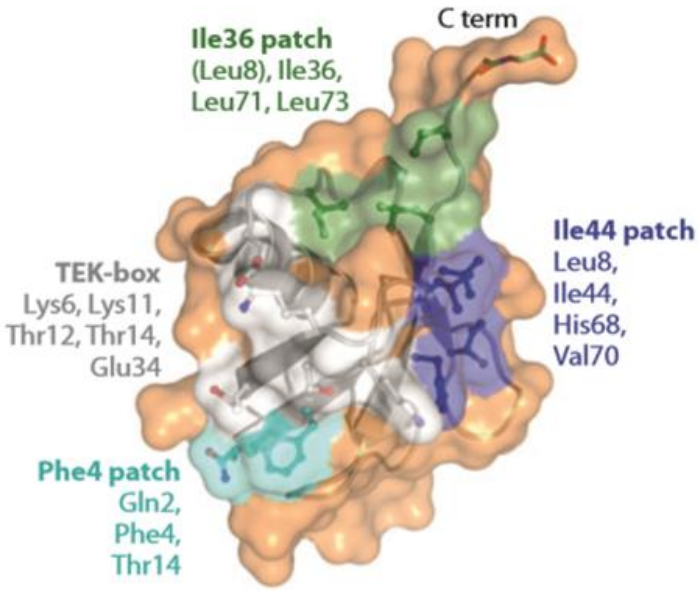


Figure 4.2: Surfaces which interact with ubiquitin binding domains
(taken from Kommander & Rape., 2012)

The four main binding surfaces of human ubiquitin, Ile36 patch, Ile44 patch, Phe4 patch and TEK-box. Each surface is colour-coded and the residues involved are shown.

Since ubiquitin plays a major part in the development and function of the immune system, mis-regulation of the ubiquitin pathway has been implicated in a number of diseases. These include arthritis, cancer and type 2 diabetes (Hoeller *et al.*, 2006; Wang & Maldonado, 2006; Schwartz & Ciechanover, 2009; Donath & Shoelson, 2011).

4.1.5. Pathogens and the host ubiquitylation pathway

The role of ubiquitin in such a vast number of cellular processes has made it an ideal target for microbial manipulation. A number of pathogens have evolved mechanisms to subvert and manipulate the host immune response through the ubiquitin pathway to aid in pathogenesis. This is accomplished by the delivery of bacterial effector proteins to the host cell via Type III and Type IV secretion systems, or OMV. Once inside the host cell, effector proteins can act upon the ubiquitin pathway in a number of different ways (Anderson & Frank, 2012).

One mechanism by which bacterial effector proteins can subvert the host ubiquitylation pathway is to mimic host DUB, for example *Yersinia* species Type III secretion effector, YopJ

plays a role in inhibiting the inflammatory response and inducing apoptosis in macrophages. YopJ interferes with the NF-κB pathway by acting as a DUB for ubiquitylated IκBα, preventing its degradation, therefore inhibiting the translocation of NF-κB to the nucleus and the transcription of the pro-inflammatory response genes (Rytkönen & Holden, 2007).

The intracellular pathogen*, Shigella flexneri* injects an effector, OspG, via its Type III secretion system into epithelial cells which interferes with the NF-κB pathway. OspG molecules interact with E2 ubiquitin-conjugating enzymes, specifically UbcH5 which are necessary for the transfer of ubiquitin to the E3 ligase involved in the degradation of IκBα. Blocking this interaction results in inhibition of the NF-κB inflammatory response, which facilitates colonisation by the organism (Angot *et al.*, 2007). *S. flexneri* also produces OspI that has deamidase activity and modifies the E2, UBC13, which prevents recognition of the enzyme by the TRAF6 E3 ligase, preventing polyubiquitylation and downstream activation of the immune response (Sanada *et al.*, 2012).

In addition to delivery by secretion systems, bacterial effector proteins can enter host cells via OMV. *Pseudomonas aeruginosa* uses OMV to export the protein Cif, which interferes with the recycling of cystic fibrosis transmembrane conductance regulator (CFTR) chloride channels in the lung. Mutations in these channels have been associated with cystic fibrosis and result in a reduced clearance of mucus in the airways, therefore a reduced clearance of any pathogens present. Cif stabilises a complex containing the DUB, USP10, preventing it from deubiquitylating CFTR and results in its degradation rather than recycling of the receptor back into the plasma membrane. The reduced number of CFTR channels in the membrane leads to reduced mucociliary clearance and compromises the immune defences of the lung (Bomberger *et al.*, 2011a).

A number of other organisms, such as *Listeria monocytogenes*, *E. coli* and *Salmonella* Typhimurium have also developed mechanisms to interfere with host signalling pathways to aid pathogenesis (Veiga & Cossart, 2005; Rytkönen & Holden, 2007; Patel *et al.*, 2009). Although the organisms outlined above can act upon the eukaryotic ubiquitin pathway, none produce ubiquitin.

4.2. Results

4.2.1. Comparison of protein sequence of BfUbb and eukaryotic ubiquitin

The sequencing of *B. fragilis* NCTC9343 identified a gene, *ubb* which encodes a protein with 63% identity (48 out of 76 amino acids) to human ubiquitin (Figure 4.3a). A sequence comparison showed that *B. fragilis* ubiquitin (BfUbb) and human ubiquitin have two main differences. The first 28 amino acids of BfUbb represent a signal peptide, indicating mobilisation of the protein to the periplasm of the cell. Secondly, the two terminal glycine residues have been replaced with an aspartic acid and a cysteine residue in BfUbb. These glycine residues in ubiquitin are required for covalent interaction between ubiquitin and the E1 and E2 enzymes in the ubiquitylation pathway and also for the formation of polyubiquitin chains.

Many of the amino acids important for the function of eukaryotic ubiquitin are conserved in BfUbb, including six of the seven lysine residues known to be involved in polyubiquitylation (Figure 4.3). Lys11 which is involved in cell cycle regulation and NF-κB activation through RIP1 has been replaced by a tryptophan residue. BfUbb has a seventh lysine residue at position 40, which replaces a glutamine present in human ubiquitin. BfUbb also contains a second cysteine residue at position 70. A di-sulphide bridge could potentially form between the C-terminal cysteine of one BfUbb molecule and Cys70 of another BfUbb molecule, suggesting a mechanism by which *B. fragilis* ubiquitin may form polyubiquitin chains. Due to the positioning of Cys70, these chains might mimic Lys63-linkages in eukaryotic ubiquitin.

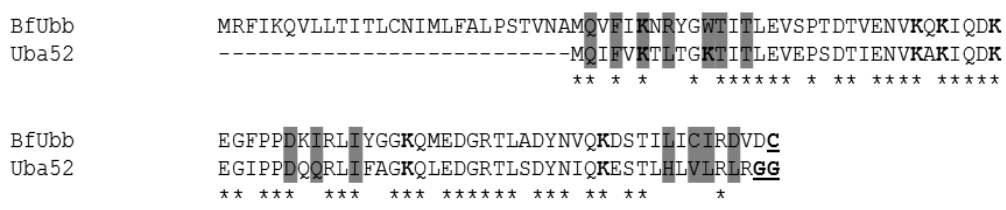


Figure 4.3: Alignment of the predicted protein sequence of BfUbb and human ubiquitin.
The first 28 residues of BfUbb represent a signal sequence. The lysine residues necessary for polyubiquitin chain formation are highlighted in bold and the C-terminal residues are underlined. The residues involved in the surfaces which interact with UBDs are highlighted in grey.

The tertiary structure of BfUbb was predicted using the PHYRE2 Protein Fold Recognition Server (Kelley & Sternberg, 2009) (Figure 4.4), which predicted it was similar to the tertiary structure of human ubiquitin, however, the residues involved in binding to UBDs, are less well conserved (Figure 4.3). BfUbb does not contain any of the residues required to form the Ile36 patch and only contains one residue, Ile44 of the Ile44 patch, which is required for proteosomal degradation in eukaryotic ubiquitin. All of the residues which form the Phe4 patch and all but one residue, Lys11 of the TEK-box are conserved in BfUbb. This suggests that BfUbb may have evolved specificity to interact with a specific component of the eukaryotic ubiquitylation pathway.

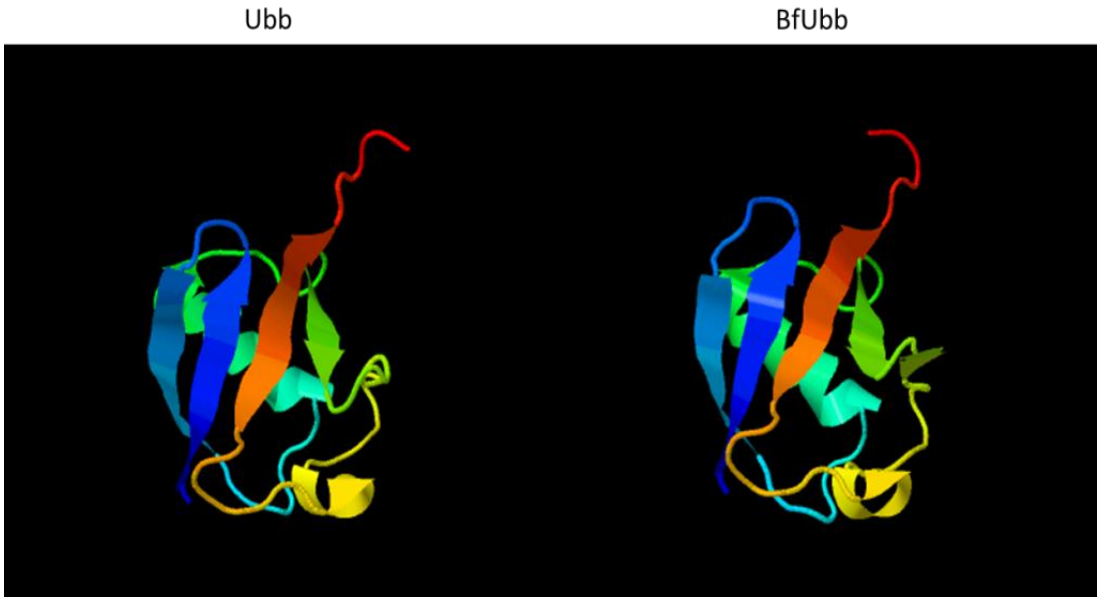


Figure 4.4: Tertiary structure comparison of Ubb and BfUbb
The tertiary structure of Ubb and BfUbb, predicted using Phyre2 with 99.8% confidence. The predicted structures of both Ubb and BfUbb contain 5 β sheets and 3 α helices.

4.2.2. Ubiquitin and the *ubb* containing region

4.2.2.1. The presence of *ubb* in clinical isolates

The *ubb* gene is contained within an 11kb, low GC content region of the NCTC9343 genome. *B. fragilis* 638R also contains a similar low GC region which includes *ubb,* however, *B. fragilis* YCH46 contains only three of the same genes in the equivalent region but does not contain *ubb*. The presence of *ubb* and the composition of the ubiquitin-containing region in *B. fragilis* clinical isolates were analysed by PCR and whole genome sequence comparison.

To determine the prevalence of *ubb,* 22 *B. fragilis* clinical isolates, collected from patients in Germany, Northern Ireland, UK, Ireland and The Netherlands (Table 4.1), were analysed using primers specific for *ubb* (Figure 4.5). The identity of the *B. fragilis* isolates was confirmed by PCR using primers specific for the gene encoding DNA gyrase B. NCTC9343 chromosomal DNA was used as a positive control and produced an amplicon of 450 bp (Figure 4.5, Lane NCTC9343). Chromosomal DNA isolated from LS84, which was previously identified as *B. thetaiotaomicron*, was a negative control, and no PCR product was visible (Lane 11). Of the 22 isolates tested, 5 produced amplicons consistent with the presence of *ubb,* LS66 (Lane 3), LS67 (Lane 13), RD48 (Lane 15), BE3 (Lane17) and GNAB4 (Lane 18). These samples were isolated in Northern Ireland (LS66 and LS67), Ireland (RD48), The Netherlands (BE3) and the UK (GNAB4). There is no apparent correlation between the source of the sample and the presence or absence of the *ubb* gene.

Table 4.1: *B. fragilis* clinical isolates

Number	Strain name	Place of Origin	Strain isolated from
1	NCTC 10584	Germany	Pus sample
2	3529	Unknown	Clinical isolate
3	LS66	Craigavon, Northern Ireland	Pus from abdominal wound; male
4	JC15	Northern Ireland	Pus sample
5	GNAB92	Edinburgh, UK	Clinical isolate
6	LS54	Craigavon, Northern Ireland	Pus from perianal abscess
7	JC15	Northern Ireland	Clinical isolate
8	NCTC9344	London, UK	Septic, post-operative infection
9	JC19	Northern Ireland	Pus sample
10	JC6	Northern Ireland	Pus sample
11	LS84*	Craigavon, Northern Ireland	Pus sample
12	JC17	Northern Ireland	Pus sample
13	LS67	Craigavon, Northern Ireland	Pus sample
14	RD47	Donegal, Ireland	Healthy faeces
15	RD48	Donegal, Ireland	Healthy faeces
16	BE1	Amsterdam, The Netherlands	Pus from wound
17	BE3	Amsterdam, The Netherlands	Clinical isolate
18	GNAB4	Edinburgh, UK	Pus sample
19	BCH1	Belfast, Northern Ireland	Post-operative wound abscess
20	VH	Northern Ireland	Clinical isolate, abscess
21	MPRL3499VI	Edinburgh, UK	IBS: Rectal Biopsy
22	MPRL3535V	Edinburgh, UK	Quiescent UC (extensive): Rectal Biopsy
23	MPRL3536X	Edinburgh, UK	Active UC (pancolitis): Biopsy ascending colon

IBS = Irritable Bowel Syndrome

UC = Ulcerative colitis

* This isolate was identified as *B. thetaiotaomicron* by whole genome sequencing

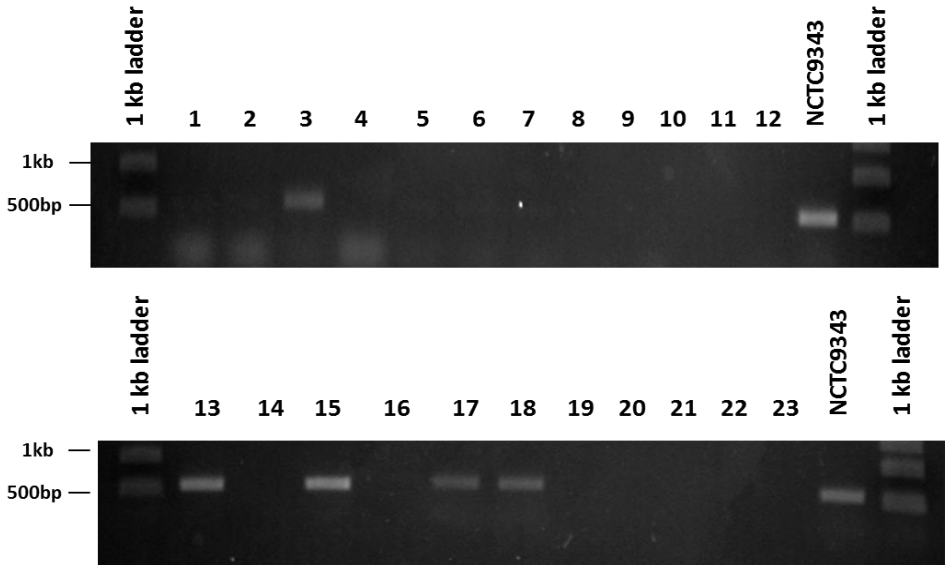


Figure 4.5: *ubb* in clinical isolates of *B. fragilis*

Agarose gels showing detection of the *ubb* gene by PCR in clinical isolates of *B. fragilis* from Ireland, Northern Ireland, Germany, The Netherlands and the UK. NCTC9343 and *B. thetaiotaomicron* (Lane 11) DNA was used as a positive and negative control, respectively. Five clinical isolates, LS66 (3), LS67 (13), RD48 (15), BE3 (17) and GNAB4 (18), produced amplicons consistent with the presence of *ubb.* The first and last lanes of each gel contain DNA molecular size markers.

4.2.2.2 Analysis of the low GC region of *B. fragilis* isolates by PCR

A comparison of the three previously sequenced *B. fragilis* genomes, NCTC9343, 638R and YCH46 indicated that the ubiquitin containing region displayed variation beyond the presence or absence of *ubb.* To determine the variation in size of the low GC region in the 22 clinical isolates, primers were designed in the DNA mismatch repair protein gene *mutL,* downstream of *ubb* and a hypothetical protein (*BF3878* in NCTC9343) upstream of the ubiquitin gene (Figure 4.6a). These genes were chosen as they were conserved in NCTC9343, 638R and YCH46 and the four newly sequenced genomes described in Chapter 3. The size of the low GC region in each the sequenced genomes was: NCTC9343, 4298 bp; 638R, 3121 bp; YCH46, 1608 bp; BE1, 2898 bp; GNAB92, 1607 bp; LS66, 3199 bp; RD48, 3598 bp. NCTC9343 chromosomal DNA was used as a positive control to generate an amplicon of 4298 bp. PCR analysis of DNA from the 23 clinical isolates (Table 4.1) produced amplicons ranging in size from 1.4-4kb (Figure 4.5b), confirming variation in this region between strains. The low GC region of *ubb* containing isolates, including NCTC9343, produced amplicons of 4 different sizes, which ranged from ~3.1 - 4 kb, LS66 (Lane 3, 3199 bp), LS67 (Lane 13, ~3 kb), RD48 (Lane 15, 3598 bp), BE3 (Lane 17, ~3.5 kb) and GNAB4

(Lane 18, ~3 kb). Those isolates which did not contain *ubb* produced amplicons of 5 different sizes, which ranged from ~ 1.4- 2.9 kb, the smallest of which were *B. fragilis* strains JC15 (Lane 6) and VH (Lane 20). Strains NCTC 10584 (Lane 1), 3529 (Lane 2) and MPRL3535V (Lane 22), which did not encode *ubb*, produced amplicons of ~2.9 kb.

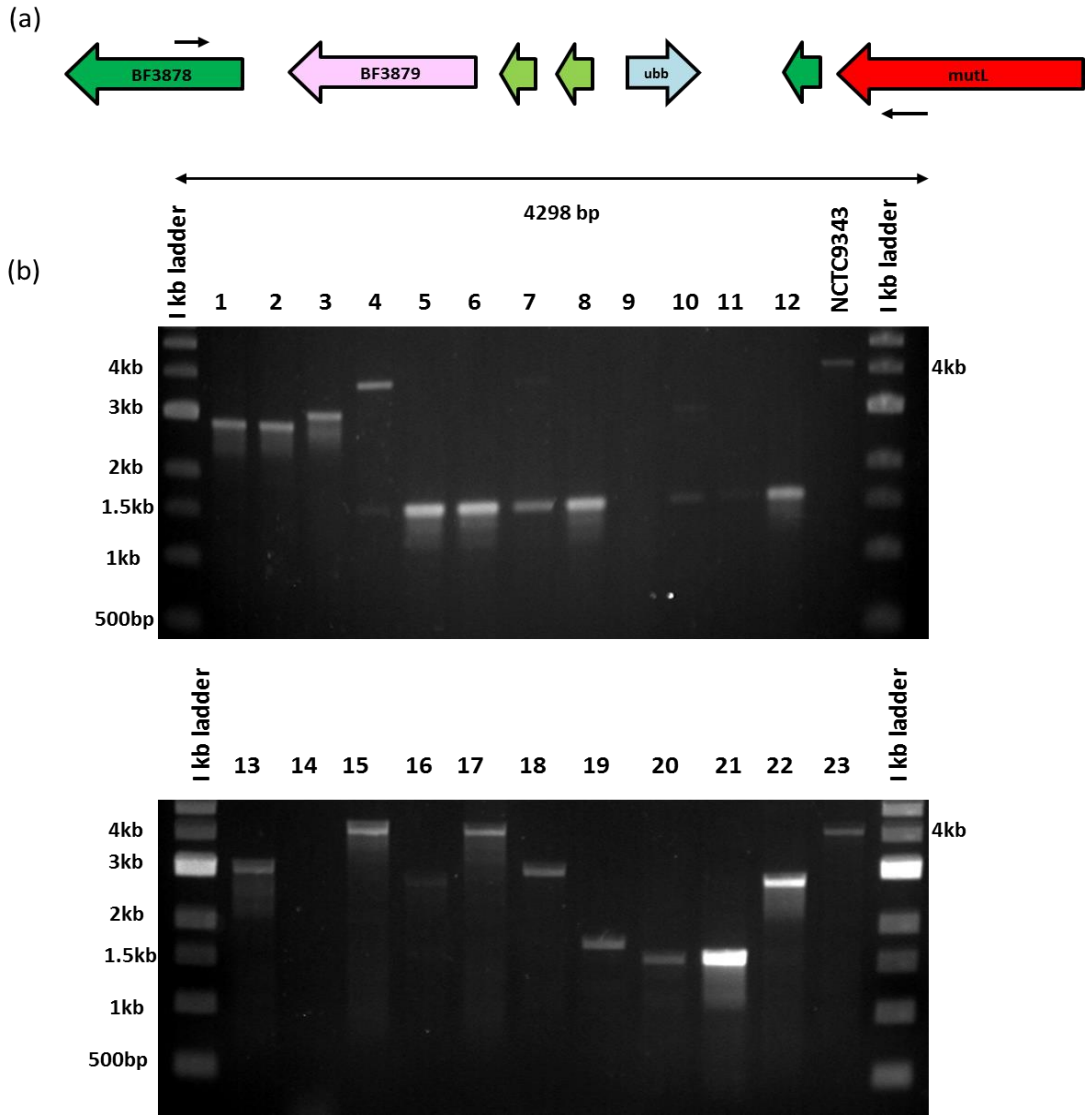


Figure 4.6: Analysis of the low GC region of *B. fragilis* clinical isolates by PCR

(a) Schematic diagram showing the low GC region of NCTC9343 flanked by *mutL* and *BF3878*. The two black arrows indicate the approximate location of the forward and reverse primers which produced an amplicon size of 4298 bp from DNA isolated from the positive control, NCTC9343. (b) Agarose gels showing the variation in size of the low GC regions of 23 clinical isolates and NCTC9343. The first and last lanes of each gel contain molecular size markers.

4.2.2.3. Analysis by whole genome sequencing.

Comparisons of the low GC region of NCTC9343, YCH46 and 638R were made with the four newly sequenced genomes RD48, BE1, GNAB92 and LS66 using the Artemis comparison tool (Figure 4.7b) to examine the variation present within this area of the genome. Two of the genomes, LS66 and RD48 contained *ubb,* which was confirmed by PCR (Figure 4.5, lanes 3 and 15, respectively). Genome sequencing also confirmed the variations in size of the low GC region as shown by PCR in Figure 4.6. The region was 1607 bp in GNAB92, 3598 bp in RD48, 2898 bp in BE1 and 3119 bp in LS66. Although there was variation within the region, there were some common genes present in strains which contained *ubb* and similarly in those in which ubiquitin was absent. The low GC region between *BF3878* and *mutL* of NCTC9343 contained six ORFs: *BF3879* (840 bp), encoding for a transposase and IS element; *BF388*0 (168 bp), *BF3881* (258 bp) and *BF3882* (279 bp), each encoding a hypothetical protein; *ubb* (228 bp) encoding BfUbb and *BF3884* encoding a putative membrane protein (231 bp) (Figure 4.7a). *ubb* positive strains, 638R, LS66 and RD48 all contained ORFs with 100% DNA sequence identity to BFBF3881, BF3882, ubb and BF3884. The strains in which *ubb* was absent, YCH46, BE1 and GNAB92 all contained a gene encoding a hypothetical protein (288 bp) and gene encoding a protein with 88% identity to an XRE transcriptional family regulator. In addition, BE1 contained an ORF (1134 bp) encoding a hypothetical protein (BE1 3921) not present in the other sequenced genomes but which has been identified in a number of other *Bacteroides* spp.

During analysis, it was observed that preceding the low GC region there was another area of variation. An ACT alignment (Figure 4.8a) of this region from *ompA* to *mutL* shows an approximately 38kb insertion present in NCTC9343 and RD48 which is absent from the five other genomes. This insertion, shown in NCTC9343 in Figure 4.8b is flanked by a transposase, BF3879 and a tyrosine recombinase, BF3839 and contains 40 genes which are predicted to encode a number of hypothetical proteins, outer membrane proteins and a homologue of the ferric iron transporter FeoB. In the genomes of YCH46, 638R, BE1, LS66 and GNAB92, the region from *ompA* to *mutL* is approximately 25kb in size. Although there was some homology between strains, this region also displayed a high level of variation, suggesting this area of the genome has potentially undergone numerous instances of HGT.

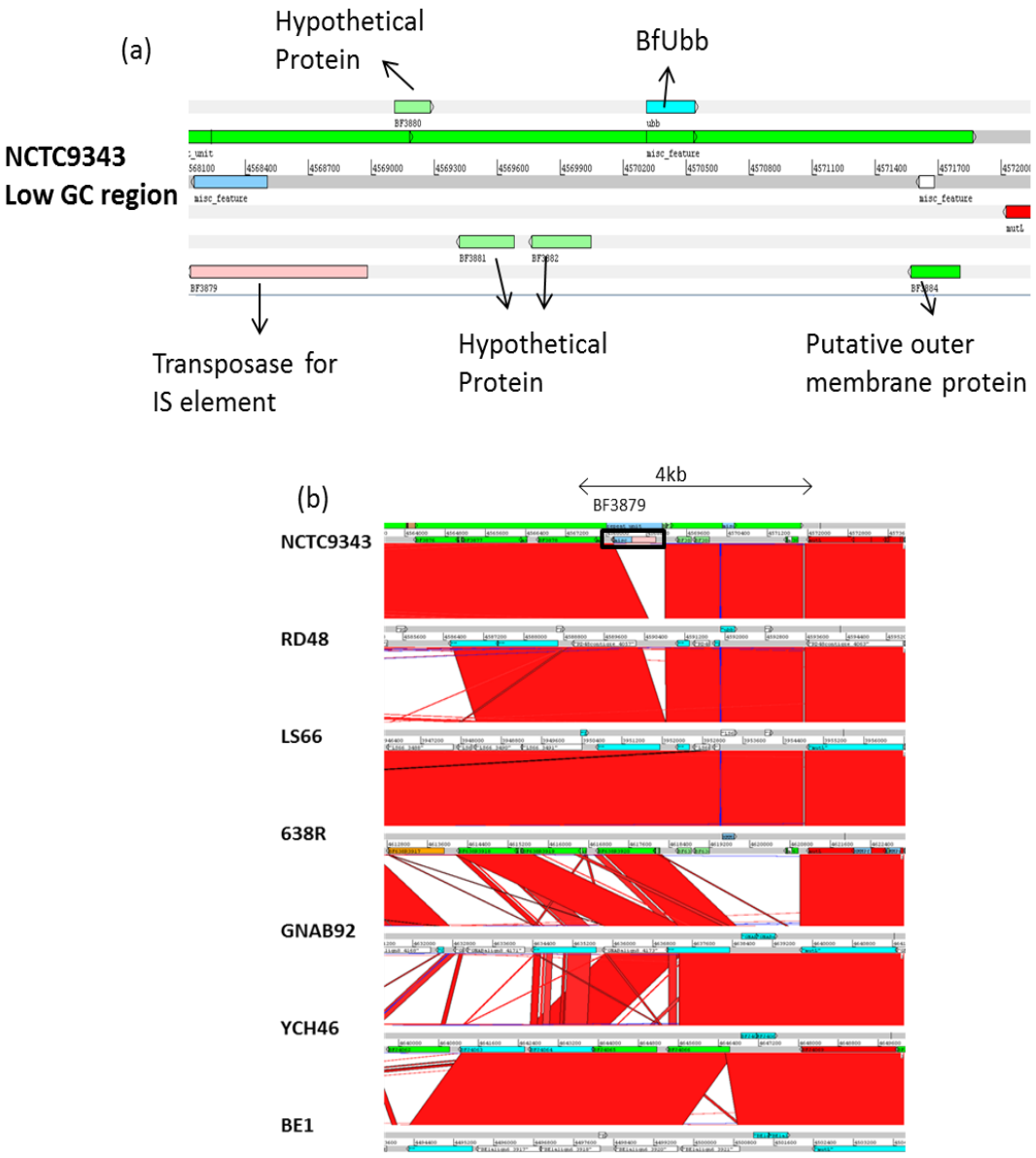


Figure 4.7: Sequence analysis of the low GC region of *B. fragilis*

Comparison of the low GC region of the four sequenced *B. fragilis* genomes with the reference genomes (NCTC9343, YCH46 and 638R) using ACT. (a) Alignment of genomes showing the 4kb region between *mutL* and *BF3878*. The transposase present in NCTC9343 is highlighted. (b) Schematic showing the similarities between the strains containing *ubb* and those in which it is absent.

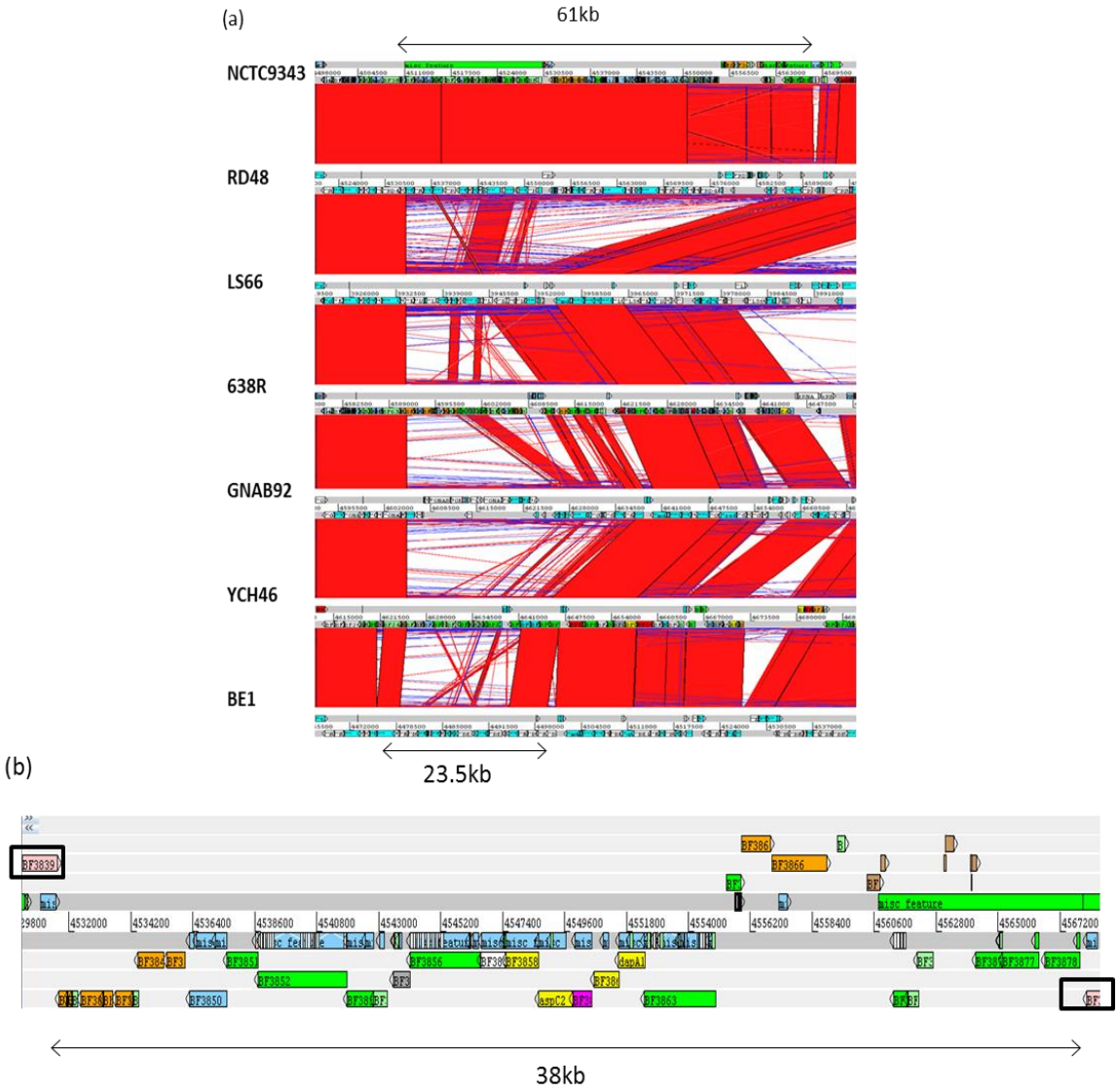


Figure 4.8: ACT comparison of the *ompA* to *mutL* region

(a) ACT comparison of the region between ompA and mutL of the reference sequences and the four sequenced *B. fragilis* genomes highlighting the difference in size of the region in NCTC9343 and RD48 and the other genomes (b) Artemis view of the insert region in NCTC9343 flanked by the transposase BF3829 and the transposase BF3879.

4.2.3 Growth rate of Δ*ubb*

As YCH46, which lacks *ubb* has a slower growth rate and attains a lower cell density in stationary phase than NCTC9343 (Patrick *et al.*, 2010), the growth rate of *B. fragilis* strain Δ*ubb*. The *ubb* deletion strain was generated by Dr Zubin Thacker, using the method described by Patrick *et al* (2009,) in which *ubb* was replaced with an erythromycin resistance cassette (Δ*ubb*::*ermF*). The growth rate of this strain was compared to that of NCTC9343 in both BHI-S and glucose defined medium (DM). Δ*ubb* had a similar growth rate to NCTC9343 (Figure 4.9), suggesting *ubb* is not essential for *B. fragilis* growth in laboratory medium. Growth experiments were performed in triplicate and an average was calculated.

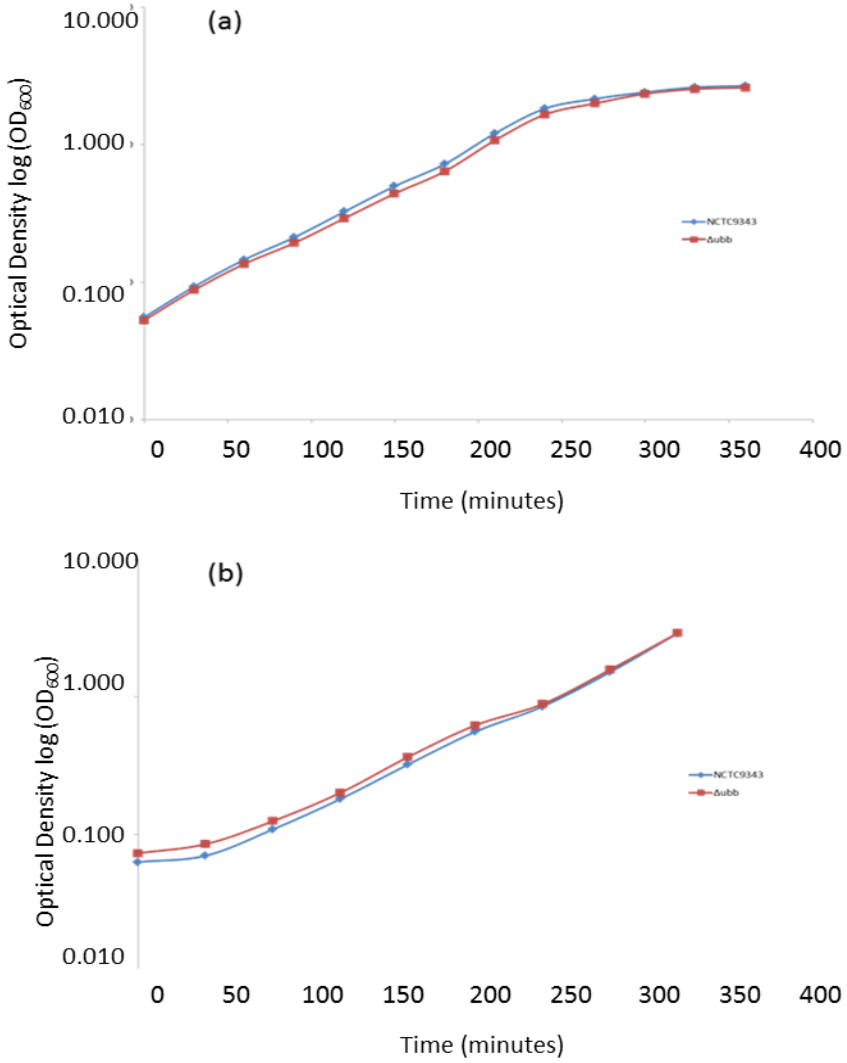


Figure 4.9: Comparison of the growth rate of NCTC9343 and Δ*ubb*

Growth of *B. fragilis* NCTC9343 and Δ*ubb* in (a) BHI-S and (b) glucose DM at 37°C under anaerobic conditions.

4.2.5 Expression of BfUbb

Following identification of the ubiquitin gene, the next step was to determine if the protein was expressed by the bacterium. Initial experiments were carried out using anti-bovine ubiquitin antibodies, which were shown to detect purified BfUbb (100 µg/ml) at a concentration of 5 µg/ml primary antibody (Figure 4.10). NCTC9343, 638R and Δ*ubb* were grown in different media, including BHI-S, DM and glucose DM, and cells were harvested during both stationary and exponential phases of growth. Whole cell protein extracts were prepared (see section 2.8.2). As BHI-S contains eukaryotic ubiquitin, all samples grown in this medium were extensively washed in PBS. The strains were also grown in the presence of polymyxin B (10 mg/ml), H_2O_2 (50 µM for 15 minutes) or ethanol to induce bacterial cell envelope stress. Whole cell extracts were analysed by SDS-PAGE using Tris-glycine polyacrylamide gels and Western blot. The purified 6xHis-BfUbb control (purified by the Dryden lab, Chemistry, University of Edinburgh) was detected using 0.4 µg/ml HRP-conjugated anti-rabbit secondary antibody, but there was no detectable level of BfUbb present in any of the whole cell extracts, suggesting either the antibody avidity was not high enough for detection or ubiquitin was not present in any of the strains tested or the concentration was too low in comparison to the purified protein.

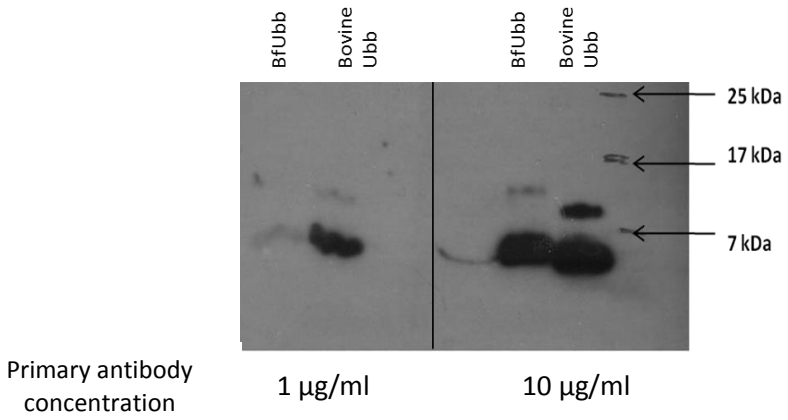


Figure 4.10: Western blot analysis of anti-bovine ubiquitin antibody specificity
Anti-bovine ubiquitin antibodies cross-reacted with BfUbb with a different binding avidity to bovine ubiquitin. Blot shows reaction of antibodies with BfUbb and bovine ubiquitin at two different concentrations, 1 µg/m and 10 µg/ml.

Using purified 6xHis-tagged BfUbb, rabbit polyclonal antiserum was generated and the whole cell extracts were re-probed for the presence of BfUbb. At this point SDS-PAGE was carried out using Tris-tricine polyacrylamide gels as they provided a better separation of proteins <15 kDa. BfUbb was detected in whole cell extracts of NCTC9343 grown in both BHI-S and DM. As shown in Figure 4.11, Western blot analysis of whole cell extracts from NCTC9343 grown in glucose DM revealed three bands. BfUbb is represented as a band of ~9 kDa, consistent with the size of the protein following cleavage of the signal peptide, a second band of ~12 kDa, consistent with the size of BfUbb before processing and a lower molecular mass band which might represent a degradation product.

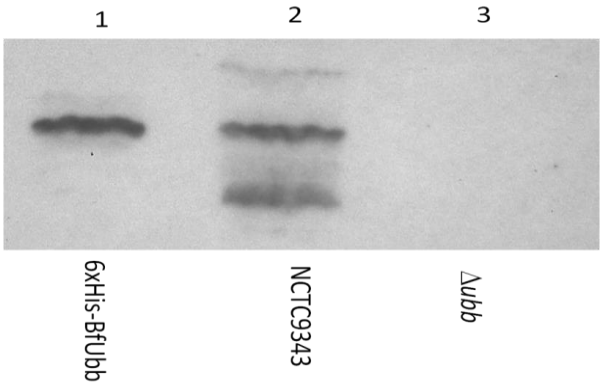


Figure 4.11: Detection of BfUbb in whole cell extracts of *B. fragilis*
Western blot using rabbit anti-BfUbb polyclonal antiserum. Lanes: 1, purified 6xHis-BfUbb; 2, whole cell extract of *B. fragilis* NCTC9343; 3, whole cell extract of *B. fragilis* Δ*ubb*. The three bands in lane two represent BfUbb before signal peptide cleavage (12kDa), BfUbb (8kDa) and the lowest band represents a degradation product of BfUbb

4.2.6 BfUbb forms dimers under reducing and non-reducing conditions

Eukaryotic ubiquitin forms polyubiquitin chains via the interaction of the C-terminal glycine residue of one ubiquitin molecule with one of seven lysine residues on the subsequent ubiquitin molecule (see section 4.1.2). BfUbb contained a C-terminal cysteine residue, suggesting that BfUbb molecules may form poly-BfUbb chains via a di-sulphide bridge (see section 4.2.1). Protein di-sulphide bonds are formed by the oxidation of thiol groups of two cysteine residues in a non-reducing environment (Gilbert, 1995). In prokaryotic and eukaryotic cells di-sulphide bonds form in extra-cytoplasmic compartments, for example, in *E. coli*, these bonds form in the periplasm of the cell (Shouldice *et al.*, 2010). Therefore the periplasm of *B. fragilis* was a potential location for poly-BfUbb chain formation. The ability of BfUbb to form poly-BfUbb chains was determined.

Samples of 6xHis-tagged BfUbb were prepared in the presence and absence of the reducing agent DTT before SDS-PAGE and Western blot analysis. In both the reducing and non-reducing sample, two bands (Figure 4.12) were detected at ~9 kDa and ~18 kDa corresponding to the size of a BfUbb monomer and dimer, respectively. The difference in intensity of the two bands suggests BfUbb was primarily present in its monomeric form. The ability of BfUbb to form a dimer in the presence of DTT is unusual, although, there are examples of di-sulphide bond formation in reducing environments. Proteins of the vaccinia virus have been shown to form di-sulphide bonds in the eukaryotic cytosol and were resistant to reduction by DTT, SDS, urea and β-mercaptoethanol (Locker & Griffiths, 1999).

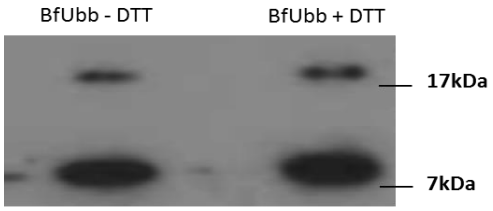


Figure 4.12: BfUbb in reducing and non-reducing conditions
Western blot analysis of BfUbb under reducing and non-reducing conditions. BfUbb appears to form dimers under both reducing and non-reducing conditions.

4.2.7 Generation and analysis of a *ubb* signal sequence deletion construct.

The presence of a signal sequence and its apparent cleavage is consistent with BfUbb being transported across the inner membrane and into the periplasm of the cell. To confirm this, a plasmid in which the BfUbb signal sequence had been deleted was constructed (pKJ01) using the Clontech In-Fusion®Cloning System (2.7.11). Primers were designed to amplify both the ubiquitin promoter (P*ubb*) and *ubb* without the signal sequence. The two components were isolated and then fused via crossover PCR and amplified using the two outside primers. These primers contained 15bp of homology to the KpnI restriction site present in the multiple cloning site of the shuttle vector pLYL01 (Figure 4.13a). A plasmid containing ubiquitin with the signal sequence still attached was also generated (pKJ02). The *E.coli* S-17 strain was transformed with the two plasmids and then used for conjugation with *B. fragilis* Δ*ubb*. Tetracycline resistant *B. fragilis* colonies were selected and plasmids were extracted for analysis. The presence of *ubb* was confirmed using PCR and Sanger sequencing. PCR was performed using the two outside primers (Figure 4.13b and c). The

PCR product of pKJ01 (Figure 4.13b lanes 3-5) was 84 bp smaller than that of the NCTC9343 chromosomal DNA control (Figure 4.13b lane 2) which is consistent with the deletion of the signal sequence. As expected, the PCR product of pKJ02 (Figure 4.13c lanes 3-5) was the same size as the NCTC9343 control (~500 bp) (Figure 4.13c lane 2).

Whole cell and periplasmic extracts of *B. fragilis* strains Δ*ubb*, Δ*ubb* containing pKJ01 and Δ*ubb* containing pKJ02 grown in glucose DM were prepared (2.8.3). Western blot analysis using anti-BfUbb polyclonal antiserum detected BfUbb in whole cell and periplasmic extracts of the Δ*ubb* strain containing pKJ02, demonstrated by the presence of an 8 kb band (Figure 4.13d lanes 3 and 4). In contrast, BfUbb was detected in whole cell extracts of the Δ*ubb* strain containing pKJ01 (lane 5) but not in the periplasmic extract (lane 6). These data support the periplasmic location of the processed form of BfUbb and the requirement of the signal peptide for transport from the cytoplasm. It should be noted however, that this is a negative result.

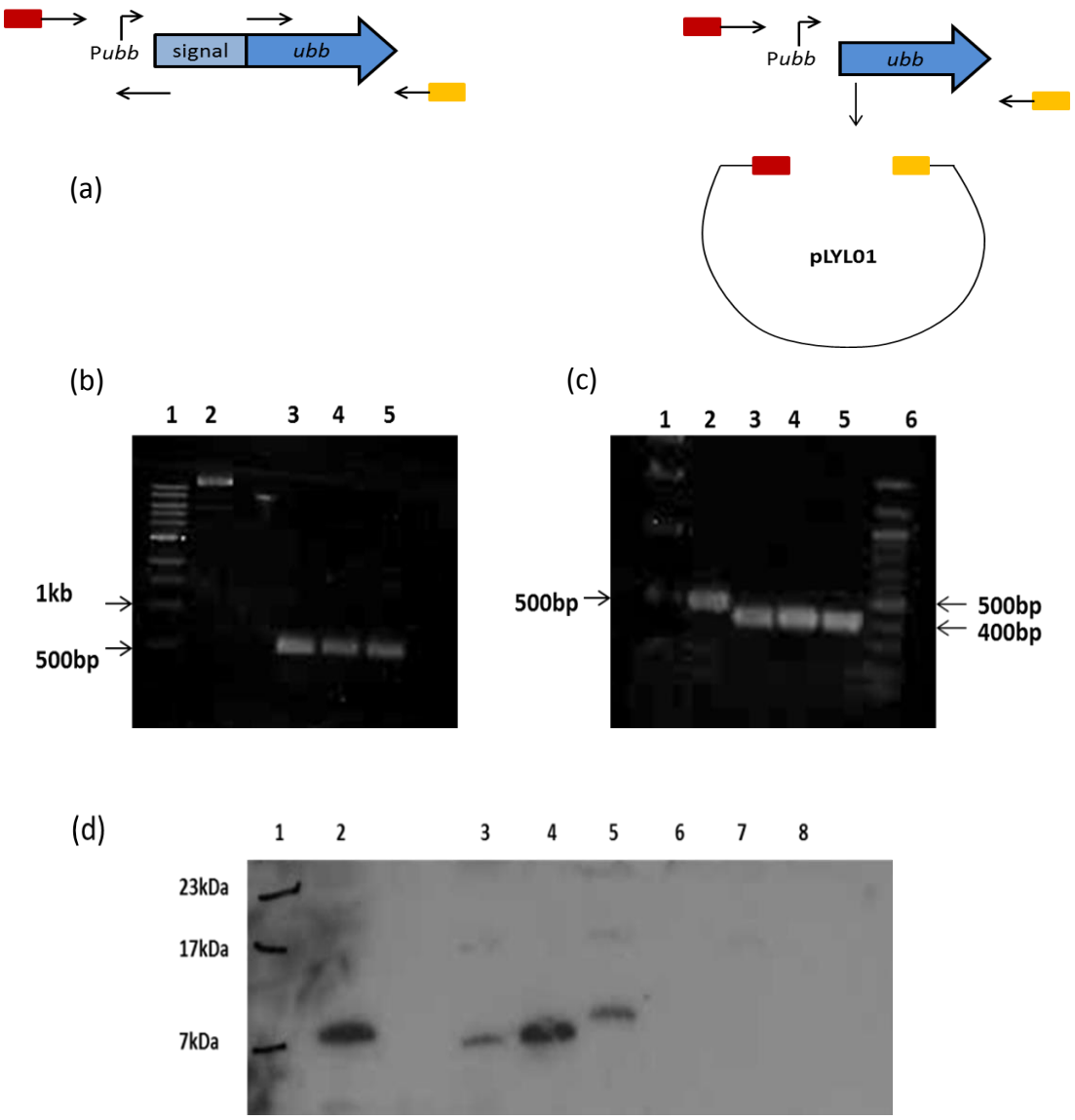


Figure 4.13 : Construction and expression of a *ubb* signal sequence deletion plasmid

(a) Schematic of the construction of a signal sequence deletion. The black arrows indicate the location of the primers. The two outside primers had homology to the KpnI restriction site present in the multiple cloning site of pLYL01. (b) Agarose gel showing pKJ02 in *Δubb* by PCR for *ubb*. Lanes: 1, I kb DNA size marker; 2, uncut pLYL01; 3,NCTC9343 chromosomal DNA; 4, pKJ02 clone 1 mini-prep; 5, pKJ02 clone 2 mini-prep. (c) Agarose gel showing pKJ01 in *Δubb* by PCR for *ubb*. Lanes: 1, I kb DNA size marker; 2,NCTC9343 chromosomal DNA; 3, pKJ01 clone 1 mini-prep; 4, pKJ01 clone 2 mini-prep; 5, pKJ01 clone 3 mini-prep; 6, 100 bp DNA size marker. (d) Immunoblot using rabbit anti-BfUbb polyclonal antiserum Lanes: 1, Prestained protein size marker; 2, purified 6xHis-tagged BfUbb; 3, whole-cell extract *Δubb* containing pKJ02; 4, periplasmic extract *Δubb* containing pKJ02*;* 5, whole-cell extract *Δubb* containing pKJ01; 6, periplasmic extract *Δubb* containingpKJ01; 7, whole-cell extract *Δubb; 8,* periplasmic extract *Δubb.*

4.2.8 *Bacteroides fragilis* OMV

Our data support the notion that BfUbb is present in the periplasm of the cell, however, to interact with the eukaryotic host and modulate ubiquitylation, the protein would need to be secreted from the bacterium. *B. fragilis* produces OMV which consist of outer membrane and periplasmic contents (see section 1.6). These vesicles have been shown to be phagocytosed by dendritic cells (Shen *et al.*, 2012). Concentrated supernatants containing OMV were isolated and the presence of BfUbb was detected by Western blot analysis.

4.2.8.1 Isolation of OMV

B. fragilis concentrated supernatants containing OMV were isolated (see section 2.9.1). Cell free supernatants from overnight cultures of *B. fragilis* grown in DM were filter sterilised and concentrated with a 100 kDa cut-off membrane. OMV preparations were then pelleted by ultracentrifugation and washed in PBS. Aliquots were plated and incubated anaerobically at 37°C overnight to check for bacterial cell contamination. The protein concentrations of OMV samples were measured using a Bradford assay (see section 2.8.6).

OMV were visualised using Transmission Electron Microscopy (TEM) (Figure 4.14) and the size distribution of OMV was measured using Dynamic Light Scattering (DLS). DLS indicated *B. fragilis* OMV range in size from 100-250nm which is consistent with the size of OMV produced by other Gram-negative bacteria. Negatively stained spherical structures were visible by TEM. Outer membrane, shown by a lighter ring surrounding the darker centre was observed in some vesicles (Figure 4.14).

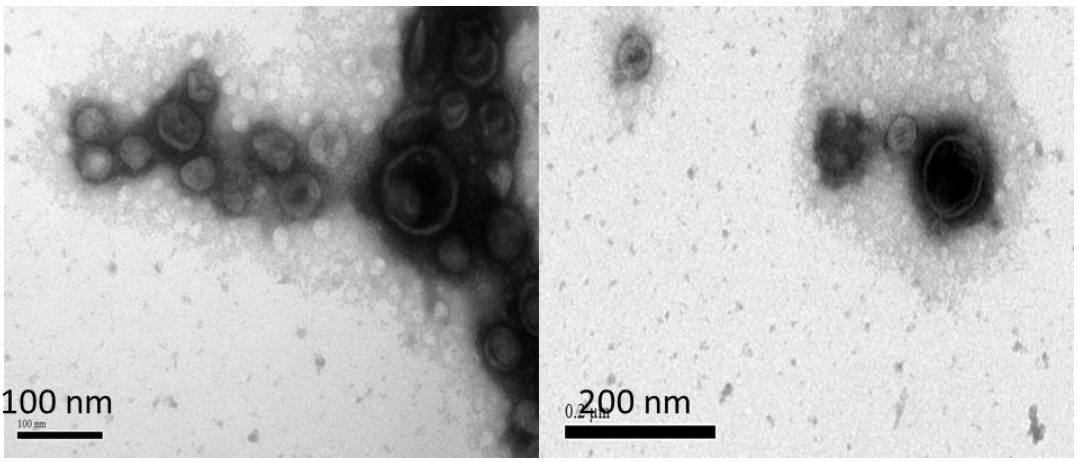


Figure 4.14 : *B. fragilis* OMV

Transmission electron micrographs of 1% (w/v) uranyl acetate stained *B. fragilis* OMV present in concentrated supernatants. OMV represented as negatively stained spherical structures which ranged in size from 100-250 nm. The outer membrane was visible in some OMV, , encompassing a darker stained contents, presumably representing the periplasmic component of the vesicles

To determine if BfUbb is associated with *B. fragilis* OMV, thereby suggesting a mechanism by which BfUbb may be delivered to the host cell, concentrated supernatants were prepared from overnight cultures of NCTC9343 and Δ*ubb* strains and analysed by Western blot (Figure 4.15). BfUbb was detected using rabbit anti-BfUbb polyclonal antiserum in both whole cell extracts and concentrated supernatants of NCTC9343, as indicated by a band at ~ 8.5 kDa. A second band of ~ 17 kDa was also detetcted in the concentrated supernatants and whole cell extracts suggesting both BfUbb monomers and multimers may be associated with OMV. BfUbb was not detected in samples prepared from strain Δ*ubb.*

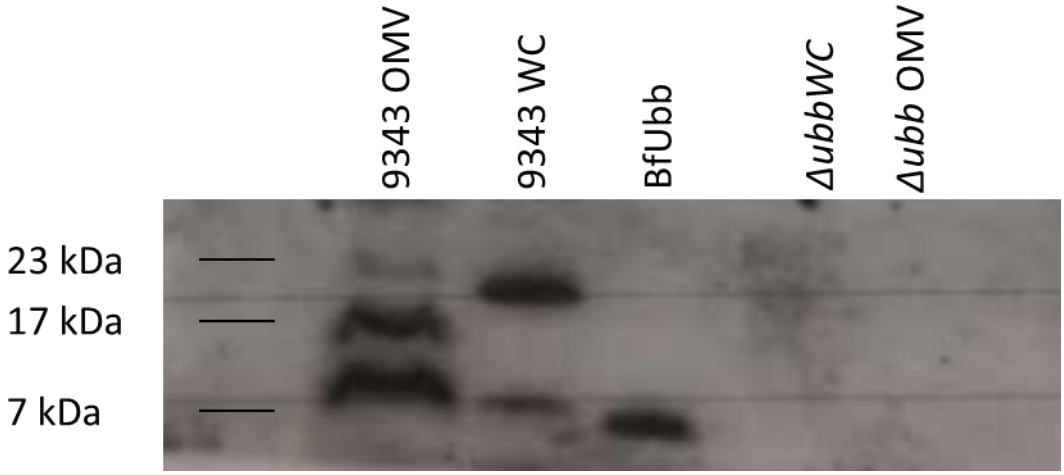


Figure 4.15: Detection of BfUbb in concentrated supernatants.
Western blot using rabbit anti-BfUbb polyclonal antiserum. BfUbb was present in whole cell extracts and concentrated supernatant of NCTC9343 but not Δ*ubb* under reducing conditions. A band of ~17 kDa was detected in the concentrated supernatant and whole cell extract of NCTC9343, possibly indicating a BfUbb dimer. Some non-specific bands were also visible.

4.2.8.2 Generation and analysis of a GFP-BfUbb fusion protein.

As a tool to aid investigation of BfUbb within eukaryotic cells, a plasmid expressing a GFP-BfUbb fusion protein was generated. Green fluorescent protein (GFP) is a 23 kDa reporter protein isolated from the genus *Aequorea* (Chalfie *et al., 1994*). It is commonly used due to its autofluorescence and ability to be viewed under a range of conditions. GFP was chosen as it has been shown that eukaryotic ubiquitin tagged with GFP (GFP-Ub) retained the functionality of the untagged protein (Qian *et al.*, 2002). Since the formation of the final GFP fluorophore only occurs following correct folding of the protein, which is dependent on the presence of oxygen, it is usually an unsuitable reporter in anaerobes. However, it has been shown that GFP expressed in organisms grown under anaerobic conditions can flouresce after oxygen has been introduced (Cubitt *et al.*, 1995; Scott *et al.*, 1998; Zhang *et al.*, 2005).

The GFP-BfUbb expressing fusion plasmid, pKJ03, was generated using the Clontech In-Fusion®Cloning System (see section 2.7.11). The *gfp* gene was isolated from the plasmid

pFU31 which encodes a derivative of GFP, GFPmut 3.1 which has a 35- fold increase in fluorescence compared to wild-type GFP. *ubb* and its promoter, P*ubb* were isolated from NCTC9343 chromosomal DNA. Primers to generate the insert used in the cloning reaction were designed to isolate P*ubb*, *gfp and ubb*. Primers were designed with approximately 20bp of homology to the next component of the construct, for example, SS_REV contained homology to *gfp.* This was to facilitate the assembly of the insert by crossover PCR. To ensure transport of the protein into the periplasm, *gfp* was inserted following the signal sequence. The reverse primer to isolate *gfp* (GFP_REV) was designed with the stop codon omitted so that translation was not terminated before *ubb*. The two outside primers (Pubb_puc19_FOR and ubb_puc19_REV) contained 15bp homology to the KpnI restriction site of the multiple cloning site of pLYL01.

E.coli S17 cells were transformed with pKJ03 and ampicillin resistant colonies were selected and screened for the presence of the plasmid. pKJ03 was then transferred into NCTC9343 by conjugation. Plasmid DNA was extracted from tetracycline resistant clones and the presence of the insert was confirmed by PCR (Figure 4.16b) and Sanger sequencing. The presence of the whole DNA insert confirmed using the two outside primers produced an amplicon of ~1.4 kb (lane 5), which was consistent with the size of the insert used for cloning (lane 2). Amplicons of ~250 bp (lane 3) and ~1.1 kb were generated following PCR with primers for *ubb* and *gfp*, respectively.

Expression of the GFP-BfUbb fusion protein was confirmed by Western blot analysis under reducing conditions. Whole cell extracts and concentrated supernatants were prepared from NCTC9343 containing pKJ03 and NCTC9343 containing pLYL01 as the control, grown in DM. Blots were probed with rabbit anti-BfUbb (Figure 4.16ci) and rabbit anti-GFP polyclonal antisera (Figure 4.16cii) to detect the presence of BfUbb and GFP, respectively. Purified BfUbb (Figure 4.16ci lane 2) and samples prepared form NCTC9343 containing pLYL01 (Figure 4.16ci lanes 3 and 4) cross reacted with anti-BfUbb antiserum, indicated by the presence of a band of ~8.5 kDa. Purified BfUbb and whole cell extract of NCTC9343 containing pLYL01 also showed the presence of a second band of ~17 kDa which is the same size as the BfUbb dimer, shown previously to form even in the presence of DTT (Figure 4.12). Purified BfUbb and proteins from the whole cell extract and concentrated supernatant of NCTC9343 containing pLYL01 did not cross-react with anti-GFP anti-serum (Figure 4.16cii lanes 2-4).

Proteins from whole cell extracts and concentrated supernatants isolated from NCTC9343 containing pKJ03 cross-reacted with both anti-BfUbb and anti-GFP antiserum. Bands were detected at the same molecular mass on both blots, indicating the two proteins, BfUbb and GFP were migrating to the same location on the gel. Two main bands were detected. The first of ~30-35 kDa corresponded to the size of a GFP-BfUbb monomer (~31kDa) confirming the generation and expression of a fusion protein. The second band was ~60-75 kDa; an exact size was difficult to determine as there was less separation of the protein size marker near the top of the gel (Figure 4.16cii lanes 5 and 6). The size of a dimer of the fusion protein would be ~62 kDa. The higher band observed may indicate that GFP-BfUbb can form multimers, however, since this was a denaturing gel, the mechanism by which multimers are formed is unclear.

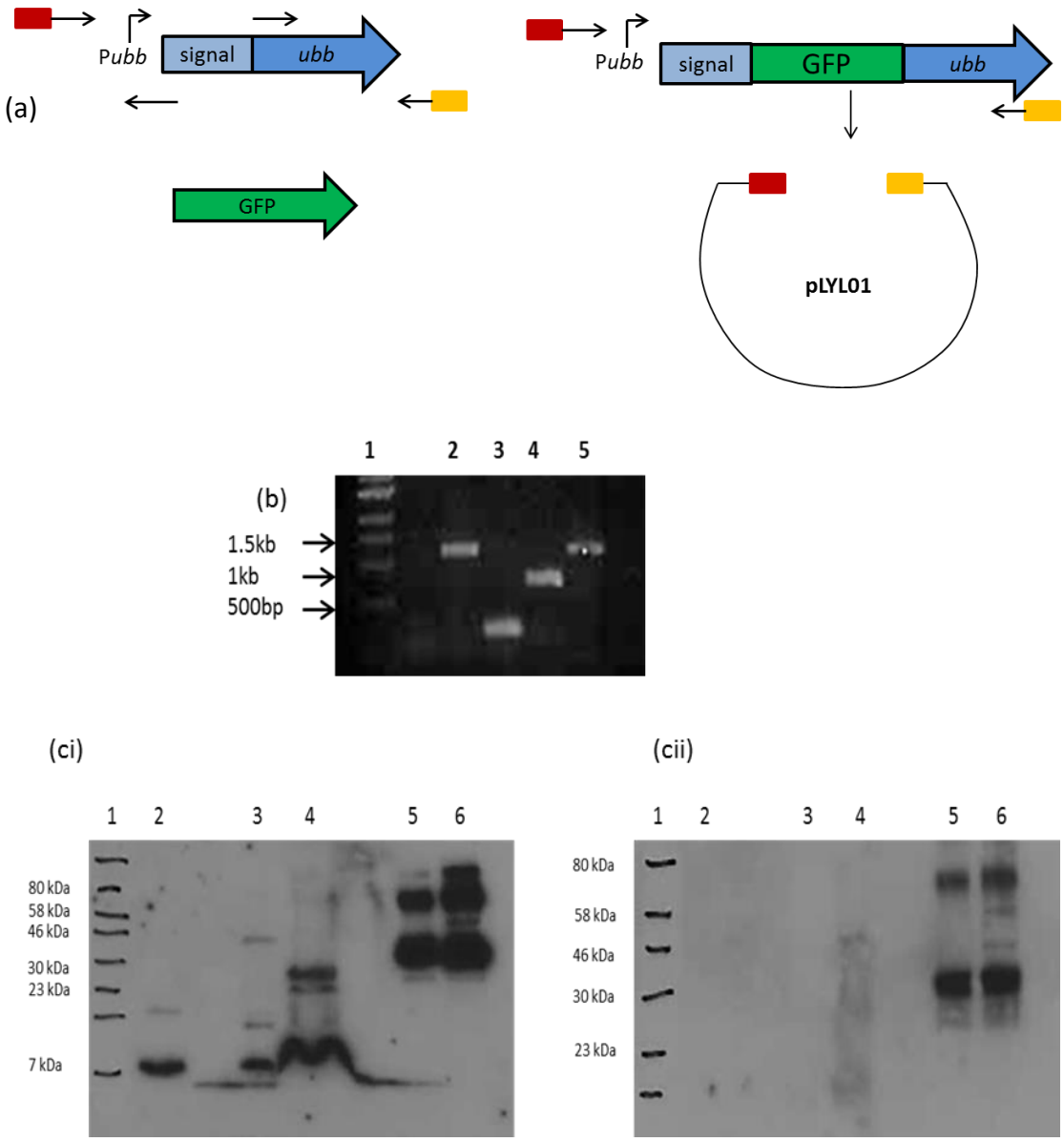


Figure 4.16: The generation of a GFP-BfUbb fusion construct

(a) Schematic representation of the cloning of the fusion construct indicating where the primers were located. Each component was isolated separately before fusion using crossover PCR. (b) Agarose gel of PCR products from pKJ03 plasmid DNA extracted from *B. fragilis* NCTC 9343::pKJ03. Lanes: 1, I kb molecular size marker; 2, GFP-*ubb* insert control; 3, *ubb*; 4, *gfp*; 5, whole construct. (c) Western blot analysis of whole cell extracts and OMV, grown in glucose DM of NCTC9343::pLYL01 and NCTC9343::pKJ03. Lanes: 1, ColorPlus PreStained protein size marker; 2, BfUbb; 3, NCTC9343::pLYL01 whole cell; 4, NCTC9343::pLYL01 OMV; 5, NCTC9343::pKJ03 whole cell; 6, NCTC9343::pKJ03 OMV. (ci) was probed using rabbit anti-BfUbb polyclonal antiserum. (cii) was probed using rabbit anti-GFP polyclonal antiserum.

4.2.8.3 Detection of GFP-BfUbb fluorescence

GFP fluorescence is dependent on the correct folding of the fluorophore, which only occurs in the presence of oxygen (see section 4.2.8.2). To determine the ability of the GFP-BfUbb fusion protein to fluoresce, concentrated supernatants, containing OMV, isolated from *B. fragilis* strain NCTC9343 containing pKJ03 were incubated with Caco-2 cells for 2 hours and fluorescence was detected using confocal microscopy (see section 2.10.5). Caco-2 cells were grown and maintained (see section 2.10.1) before they were seeded onto glass coverslips and grown until confluent. Cells were then incubated with OMV (100ug/ml protein concentration as measure by a Bradford assay). After 2 hours, cells were washed in PBS and fixed in 4% paraformaldehyde. Cells were visualised using confocal microscopy (Figure 4.17). It was observed that following incubation with the OMV, cells appeared to be undergoing cell death, characterised by rounding up, disruption of cell membranes and release of cytoplasmic content. Death was not observed in cells which were incubated without *B. fragilis* OMV (Figure 4.17). The experiment was carried out in triplicate, but due to the cell death there were very few intact cells and GFP fluorescence was only detected in one cell in two of the three replicates. However, detection of fluorescence indicated that the GFP-BfUbb fusion protein was delivered to the eukaryotic cell via OMV. It also demonstrated the correct folding of GFP after its introduction to an aerobic environment, which suggested that GFP-BfUbb could be used as a reporter to investigate the function of BfUbb within the host cell.

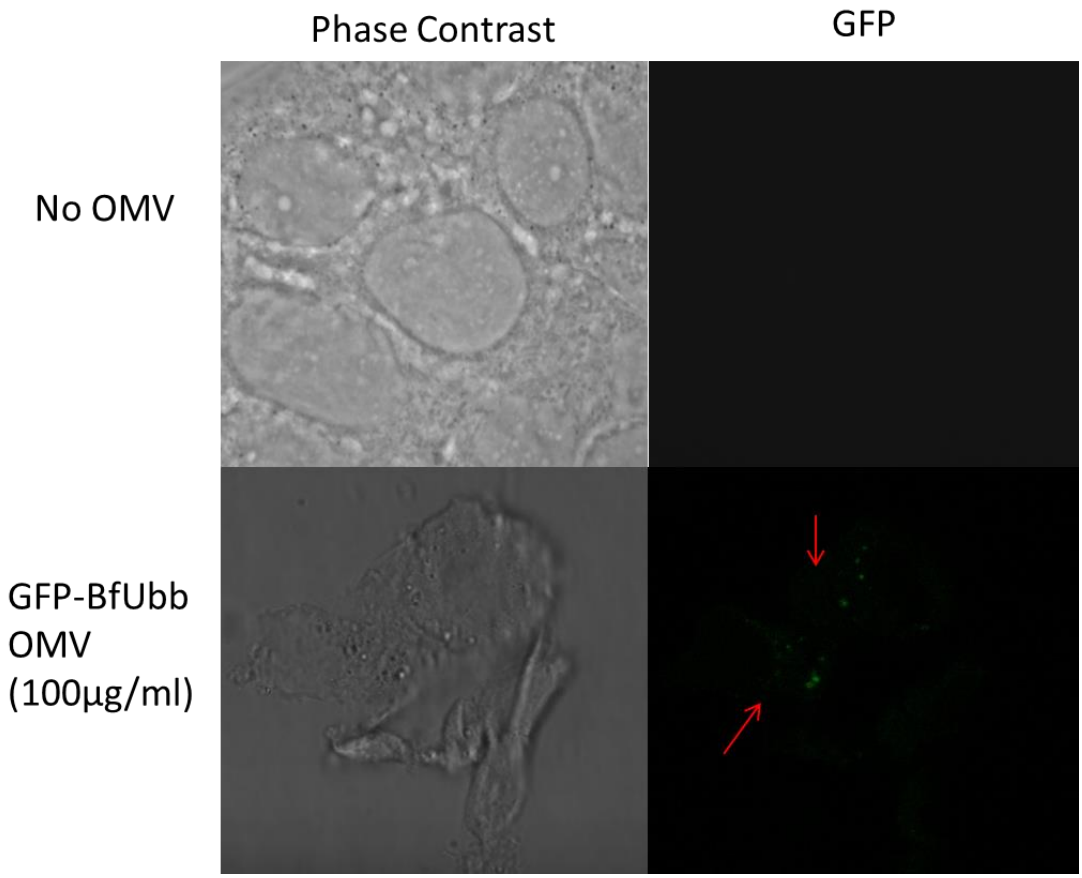


Figure 4.17: Caco-2 cells incubated with concentrated supernatants containing GFP-BfUbb GFP fluorescence detected in Caco-2 cells following incubation with OMV containing GFP-BfUbb for 2 h. The bottom right-hand panel shows fluorescence within the cell, indicated by the red arrows. Death was observed in cells which had been incubated with OMV. This was not seen in cells which had been incubated in the absence of OMV (top-left panel). This suggests *B. fragilis* OMV had a toxic effect on Caco-2 cells.

Cell death was also observed after the incubation of Caco-2 cells with 100 µg/ml (protein concentration) NCTC9343 OMV. Cells were incubated with vesicles for 2 hours before they were washed, fixed in 4% paraformaldyhyde and stained with DAPI. Cells were visualised using confocal microscopy (Figure 4.18). In comparison to the OMV free control, cells incubated with OMV appeared to be rounded and undergoing cell death. In the control cells DAPI was confined to the nucleus, however, cells which had been incubated with OMV showed DAPI in the cytoplasm, which indicated the breakdown of the nucleus and dissemination of the nuclear material throughout the cell. The observed cell death may be

due the contents of the OMV being released into the cell, however, it may also be due to *B. fragilis* LPS which is a component of the OMV.

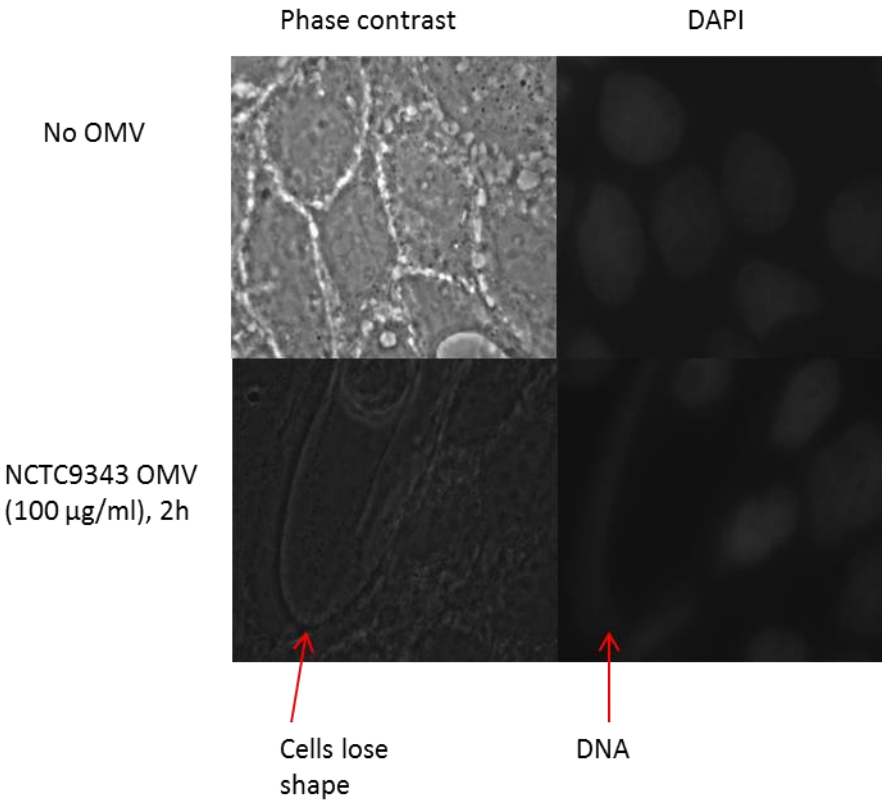


Figure 4.18: Colonic carcinoma cells incubated with *B. fragilis* OMV
Confocal images of Caco-2 cells following incubation with NCTC9343 concentrated supernatants containing OMV (bottom panels). Cells were incubated with OMV for 2 hours prior to washing and fixation with 4% paraformaldehyde. Cellular DNA was stained with DAPI. Cells which were incubated with OMV underwent cell death and dissemination of the nuclear material into the cytoplasm. Cell death was not observed in the OMV free control (top panels).

4.2.8.5 *B. fragilis* OMV contain plasmid and chromosomal DNA

A number have studies have identified DNA as component of OMV. To date, chromosomal and plasmid DNA have been isolated from OMV produced by *Haemophilus, Neisseria, E. coli*, *P. aeruginosa* and *Acinetobacter* (Kahn *et al.*, 1982; Dorward *et al.*, 1989; Yaron *et al.*, 2000; Renelli *et al.*, 2004; Rumbo *et al.*, 2011). To investigate the presence of DNA in *B. fragilis* OMV, vesicles were treated with DNase (see section 2.9.2) to remove exogenous DNA present in the supernatant. OMV were isolated from NCTC9343 which contains a

36.5kb plasmid, pBF9343. DNA was extracted from samples taken from the cell-free supernatant before any filtration or concentration, the supernatant of the final PBS wash and the final OMV sample. This was done to confirm that any DNA detected was isolated from the OMV and not simply present in the supernatant as a result of bacterial cell lysis. PCR assays were carried out using primers specific for the pBF9343 encoded Hin invertase homologue *finB*, and the chromosomally encoded genes *ubb* and *4298* (Figure 4.19).

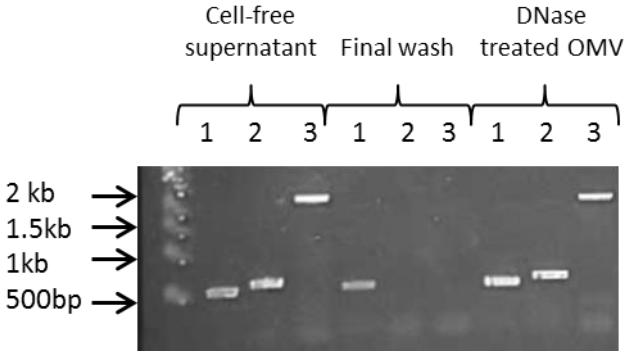


Figure 4.19: PCR detection of DNA isolated from *B. fragilis* OMV
Agarose gel showing detection of (1) *ubb* (2) *finB* (3) *4298* by PCR in mini-preps from untreated cell-free supernatant, supernatant of the final PBS wash post-DNase treatment and OMV post-DNase treatment.

Amplicons for all three genes were detected in the cell free supernatant which had not undergone concentration or treatment with DNase (Figure 4.19 cell-free supernatant 1-3). All three genes were also detected in DNA extracted from the OMV, however, *ubb* was also detected in the final wash sample indicating it was present in the supernatant rather than the OMV and that DNase treatment was not sufficient to remove all DNA. *finB* and BF*4298* were not detected in the final wash by PCR, suggesting both plasmid and chromosomal DNA can be packaged into *B. fragilis* OMV.

4.2.8.6 *B. fragilis* OMV-mediated transformation

The packaging of DNA into OMV may be a mechanism by plasmids can be transferred between bacteria. Rumbo *et al* (2011) demonstrated that *Acinetobacter bumannii* can transfer carbapenem resistance plasmids between strains via OMV. The ability of *B. fragilis* OMV to mediate transformation was investigated (see section 2.9.3). NCTC9343 was

transformed with pLYL01, which contains an *amp*^{R} gene and OMV were isolated from this strain. The presence of the plasmid in OMV was confirmed by mini-prep and agarose gel electrophoresis. OMV isolated from NCTC9343 containing pLYL01 (concentrations of 200, 100 and 50 μg/ml were used) were incubated with *E. coli* XL-1 Blue MRF' cells (~10^5), which are deficient in Type I and Type IV restriction systems and tetracycline resistant, in the presence of 50 μg/ml DNase. Cultures were grown aerobically at 37°C, samples were taken at 2h, 4h and 12h and plated onto LB plates containing ampicillin and tetracycline. The ability of OMV to mediate transformation between two *B. fragilis* strains was also investigated. OMV isolated from NCTC9343 containing pLYL01 (concentrations of 200. 100 and 50 μg/ml) were incubated with *B. fragilis Δubb* cells (~10^5), which are erythromycin resistant, in the presence of 50 μg/ml DNase. All stages of the experiment were carried out anaerobically at 37°C. Samples of culture were taken at 2h, 4h, 12h and 24 h and plated onto BHI-S plates containing erythromycin and tetracycline. These experiments were carried out in triplicate and repeated three times, in all instances, no resistant colonies were identified. This indicated that under these conditions *B. fragilis* OMV were unable to mediate transfer of plasmid DNA into *E. coli* or other trains of *B. fragilis*.

4.3 Conclusion

Bacteroides fragilis encodes and expresses a unique homologue of the eukaryotic protein-modifier, ubiquitin. This is the first example of ubiquitin being produced by a prokaryote. The number of *Bacteroides* present in the GI-tract is estimated between 10^9 to 10^{14} cells per gram of faeces, approximately 4-14% of which are *B. fragilis* (Patrick *et al.*, 2009). This means that some people could have a large amount of BfUbb present in their GI-tract. BfUbb may interact with the host ubiquitylation pathway to aid the survival of the organism as a member of the resident microbiota or as an opportunistic pathogen.

The genome of *B. fragilis* does not appear to encode the activating or conjugating enzymes of the ubiquitylation pathway, nor does it encode any components of the proteasome. The changes in the amino acid sequence and the presence of a signal sequence, which facilitates transport of the protein into the periplasm, suggests that BfUbb has evolved to perform a unique role.

The detection of BfUbb associated with OMVs suggests a mechanism by which the protein is packaged and exported from the bacterial cell. OMVs are secretory vesicles which have a well-established role in the delivery of bacterial effector proteins to host cells. *B. fragilis* OMV have been shown to be phagocytosed by dendritic cells (Shen *et al.*, 2012), therefore the target for BfUbb may be the host ubiquitylation pathway in this cell type.

The residues important for the interaction of ubiquitin with the E1 activating and E2 conjugating enzymes have been conserved in BfUbb. It has also been shown that BfUbb can covalently bind to E1 and inhibit eukaryotic ubiquitylation *in vitro* (Patrick *et al.*, 2011). This suggests that the interaction between BfUbb and these enzymes could be important. However, BfUbb could also potentially be interacting with other eukaryotic proteins to modify their function.

In eukaryotic ubiquitylation, the binding of ubiquitin to a target protein involves the formation of an isopeptide bond between the C-terminal glycine residue and a lysine residue of the target protein. Although BfUbb has a cysteine residue at the C-terminus it is possible that it could form a thiol-ester bond with other amino acids on the target protein. Eukaryotic ubiquitin has been shown to bind to residues other than lysine on a target substrate. For example, the viral E3 ligase, mK3, which targets the MHC I heavy chain for degradation can catalyse the binding of ubiquitin to lysine, serine and threonine residues of

the target protein (Wang *et al.*, 2007). The authors demonstrated that polyubiquitylation still occurred after the mutation of all of the lysine residues present in the substrate. Vosper *et al* (2009) also showed that ubiquitylation of the transcription factor Neurogenin can occur via a cysteine residue, after the mutation of lysine residues, and the protein was still degraded by the proteasome.

The identification of BfUbb dimers (Figure 4.12) suggests the protein may be able to form multimers via a di-sulphide bridge between C-terminal cysteine residues or Cys70 giving the potential for poly-BfUbb chain formation. Traditionally di-sulphide bridges cannot form in the reducing environment of the eukaryotic cytosol, however there is some evidence to suggest that intracellular organelles such as endosomes and lysosomes contain a non-reducing environment (Austin *et al.*, 2005; Yang *et al.*, 2006). As BfUbb contains six of the seven lysine residues involved in polyubiquitin chain formation, could BfUbb form mixed chains with the native ubiquitin?

Although the predicted tertiary structures of BfUbb and eukaryotic ubiquitin are similar, the residues important for interactions with UBDs are less well conserved. BfUbb is missing residues belonging to the two hydrophobic patches involved in the majority of UBD interactions. BfUbb has retained all of the residues involved in the formation of the Phe4 patch. The change in the C-terminal residue of *B. fragilis* ubiquitin and the loss of binding surfaces which are important in the function of ubiquitin, suggests BfUbb has evolved specificity. Unlike eukaryotic ubiquitin which is involved in a vast array of cellular processes, BfUbb may interact with a specific part of the host ubiquitylation pathway.

B. fragilis is a commensal organism, and BfUbb may aid in the organism's ability to establish itself as a member of the resident microbiota of the GI-tract. A number of bacterial pathogens have developed mechanisms to subvert the host ubiquitin pathway to aid pathogenesis. The target of many of these mechanisms is the regulation of the pro-inflammatory response, in which ubiquitin plays a major role. Commensal bacteria have been shown to interfere with this pathway to maintain homeostasis in the gut. Excessive NF-κB activation in intestinal epithelial cells may lead to colitis. However, commensals can cause an anti-inflammatory effect by interfering with NF-κB signalling (Goto & Ivanov, 2013). *Bacteroides thetaiotaomicron* has an anti-inflammatory effect by enhancing the export of the NF-κB subunit RelA from the nucleus back into the cytoplasm. This reduces the amount of time NF-κB is active, therefore limiting the inflammatory response (Kelly *et*

al., 2004). Could BfUbb play a role in inhibiting NF-κB activation and dampening the inflammatory response?

It has been established that the resident microbiota of the GI-tract have a role in the development of the host immune system, of which ubiquitin is also an integral component. Inflammatory bowel disease (IBD) is caused by a combination of genetic susceptibility and an abnormal mucosal immune response (Xavier & Podolsky, 2007). It has been demonstrated that *B. fragilis* has an intimate association with the mucosa of the GI-tract (Lee *et al.*, 2013) and that in patients with IBD it was the predominant member of the mucosal biofilm, accounting for >60% of the biofilm mass (Swidsinski *et al.*, 2005).

Crohn's disease, a type of IBD has been associated with dysregulation of the NOD2 signalling pathway. Nucleotide binding and oligomerization domain 2 (NOD2) is an intracellular PRR which recognises a component of bacterial peptidoglycan and initiates a cascade which results in the activation of a number of signalling pathways, including the NF-κB and autophagy pathways (Abbott *et al.*, 2004). Both activation of NF-κB by NOD2 and downregulation of this signalling involves ubiquitin, DUBs and a number of ubiquitin E3 ligases (Tigno-Aranjuez & Abbott, 2012). Interference in this signalling pathway could contribute to the inflammation observed in Crohn's disease.

The mis-regulation of ubiquitin has been implicated in a number of diseases, including, neurodegenerative disorders, cancer and cardiac disease. The causes of diseases such as IBD are not currently understood, however, a number of cellular pathways in which ubiquitin is an essential component have been implicated. Could BfUbb play a role in the development of these diseases?

Chapter 5. Bacteriophages as mediators of horizontal gene transfer

5.1. Introduction

Transduction by bacteriophages is one mechanism by which horizontal gene transfer can occur and might contribute to the high level of variation observed between the *B. fragilis* genomes (see section 1.2.2 and chapter 3). Bacteriophage transduction is also a useful tool for bacterial genetic manipulation, for example, cloning using P1 transduction in *E. coli*. Genetic manipulation of *B. fragilis* is difficult, probably in part due to the presence of variable R-M systems (see section 3.2.5). Transduction by bacteriophages may be a mechanism by which the host R-M systems could be subverted to improve genetic manipulation of *B. fragilis*. The non-transducing, *B. fragilis*-specific bacteriophage ΦED01 was previously isolated from sewage collected in the Edinburgh area, its host-range was identified and its genome sequenced (G.W Blakely, personal communication). The aim of this chapter was to identify further *B. fragilis*-specific phages, different to ΦED01 and determine their transducing capability.

5.2. Results

5.2.1. Isolation and identification of *B. fragilis*-specific bacteriophage

20L of raw sewage from the Edinburgh area (Veolia Seafield sewage treatment plant) were coarse filtered, fine filtered and then concentrated by tangential flow filtration. Samples of treated sewage were plated onto soft agar overlays containing *B. fragilis* 638R and NCTC9343. Plaques were extracted and amplified by incubation with the appropriate *B. fragilis* strain (see section 2.11.1). Initial isolations were performed by students in an undergraduate practical class.

In total, fifteen potentially new *B. fragilis*-specific bacteriophages were isolated (Table 5.1), twelve of which were isolated by propagation on NCTC9343 and three by propagation on 638R. To determine the titre and host range of each phage, 10-fold serial dilutions were carried out, and 10 µl of each dilution was spotted onto soft agar overlays containing NCTC9343 or 638R (Figure 5.1). Spot plating was carried out in triplicate and average plaque counts were used to determine bacteriophage titres on each *B. fragilis* strain (Table 5.1). Nine isolated bacteriophages were able to infect NCTC9343, one infected 638R and five infected both strains, including previously identified phage, ΦED01. Four

bacteriophages, ΦED01, ΦED09, ΦED13 and ΦED15 which infected both NCTC9343 and 638R did not form clear plaques when spotted onto 638R. Instead, hazy zones were formed, suggesting these phage may have been entered a lysogenic lifecycle or the host cells were partially resistant to phage infection. Resistance of a bacterial strain to phage infection could occur if the receptor to which the phage binds had been lost or expression turned off. Those phage which infected both strains of *B. fragilis*, did so at different titres, for example the titre of ΦED08 on NCTC9343 was $2x10^9$ PFU/ml, but on 638R it was $8x10^5$ PFU/ml (Table 5.1). This indicated host-mediated restriction of bacteriophage infection which is consistent with the different R-M systems identified in the genomes of these strains. ΦED14 was unable to infect either *B. fragilis* strain after isolation, suggesting this phage was lost during the amplification and purification process.

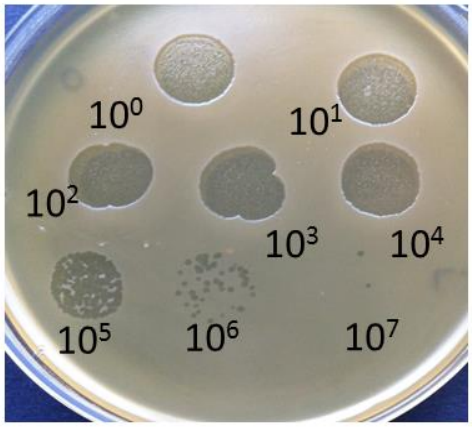
Table 5.1: Bacteriophage titres

Bacteriophage	Original strain of isolation	Titre on NCTC9343 (PFU/ml)	Titre on 638R (PFU/ml)
ΦED01	NCTC9343	$3.6x10^{10}$	neat*
ΦED02	NCTC9343	$1.1x10^5$	X
ΦED03	NCTC9343	$1.6x10^3$	X
ΦED04	NCTC9343	$2x10^4$	X
ΦED05	NCTC9343	$3x10^3$	X
ΦED06	NCTC9343	$1x10^4$	X
ΦED07	NCTC9343	$4x10^2$	X
ΦED08	NCTC9343	$2x10^9$	$8x10^5$
ΦED09	NCTC9343	$3x10^7$	neat*
ΦED10	NCTC9343	$2.4x10^5$	X
ΦED11	638R	X	neat
ΦED12	638R	$8x10^2$	X
ΦED13	638R	$4x10^4$	10^4*
ΦED14	NCTC9343	x	X
ΦED15	NCTC9343	$3.6x10^9$	neat*
ΦED16	NCTC9343	$5x10^6$	X

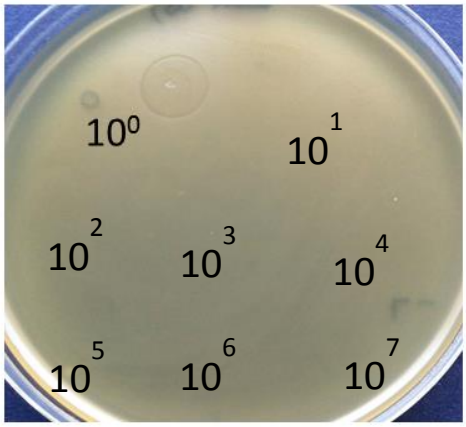
x No infection

* Hazy zone of infection rather than a zone of clearance

ΦED01

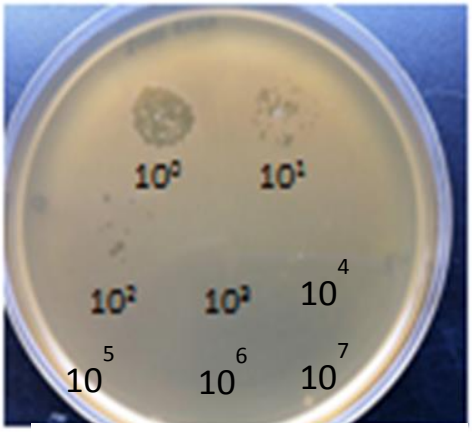


NCTC9343

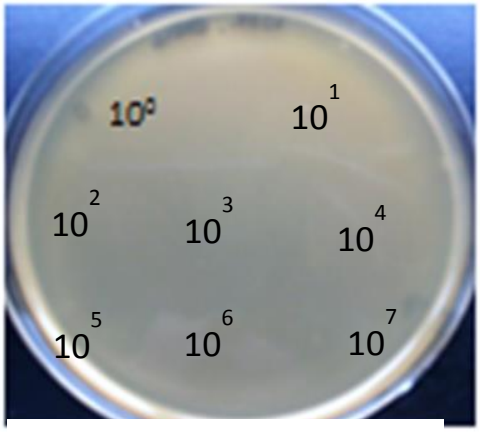


638R

ΦED02

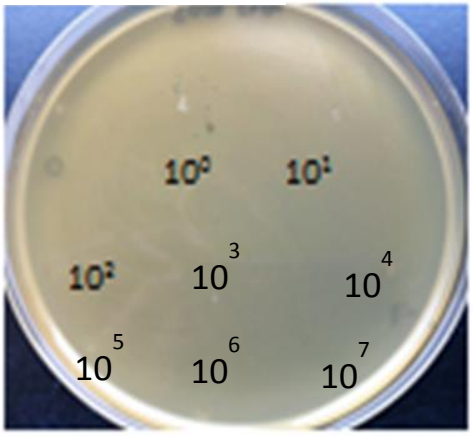


NCTC9343

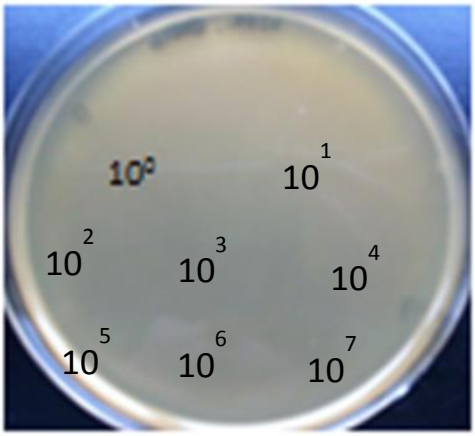


638R

ΦED03

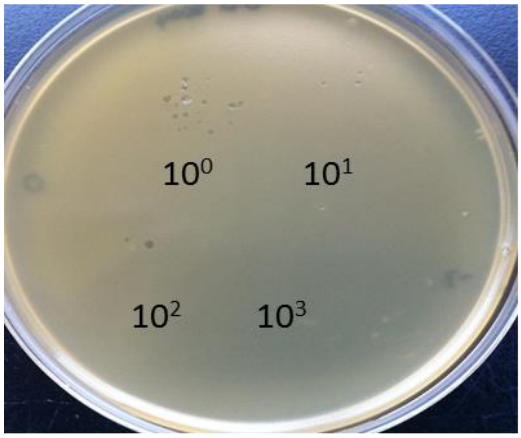


NCTC9343

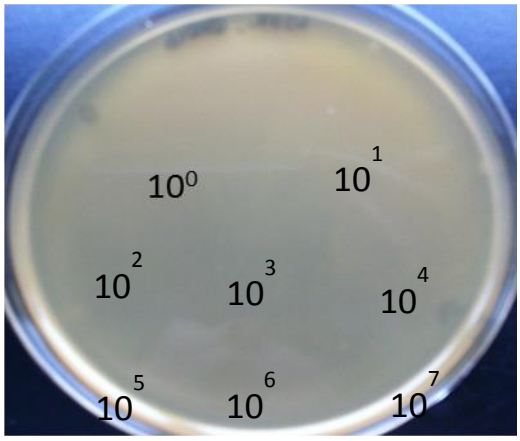


638R

ΦED04

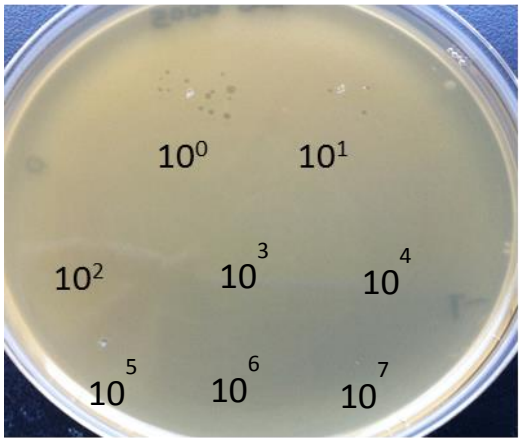


NCTC9343

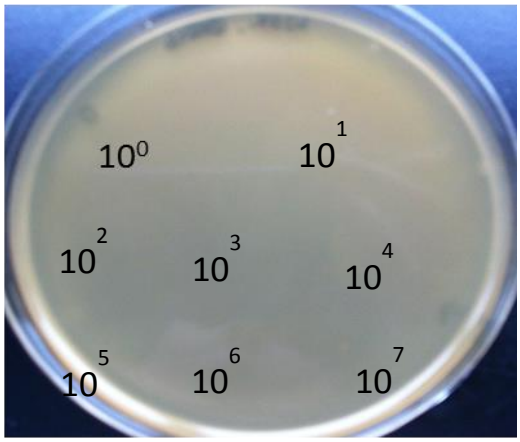


638R

ΦED05

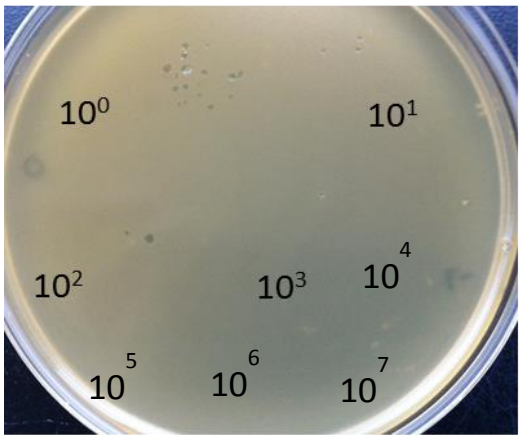


NCTC9343

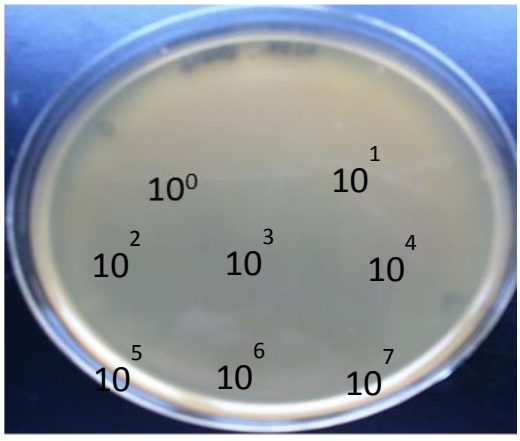


638R

ΦED06

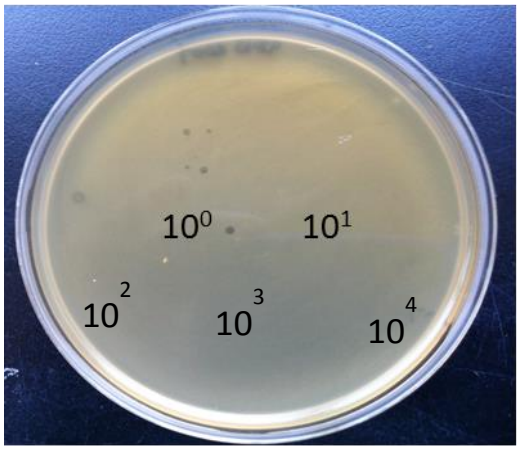


NCTC9343

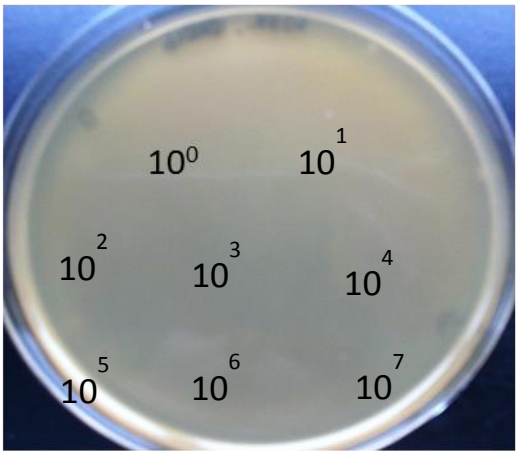


638R

ΦED07

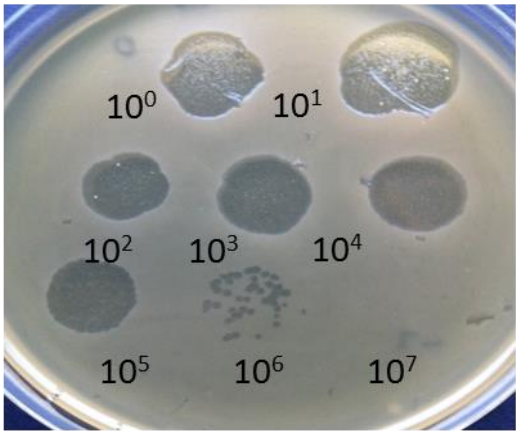


NCTC9343

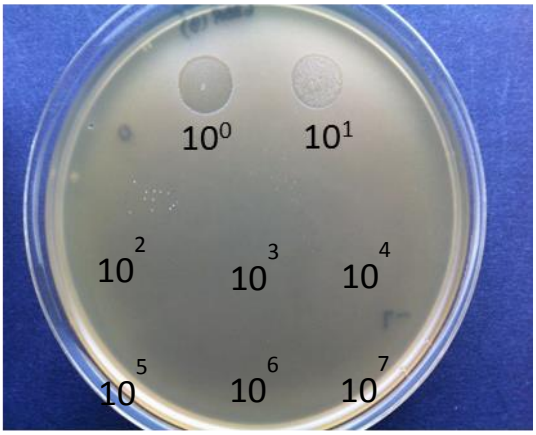


638R

ΦED08

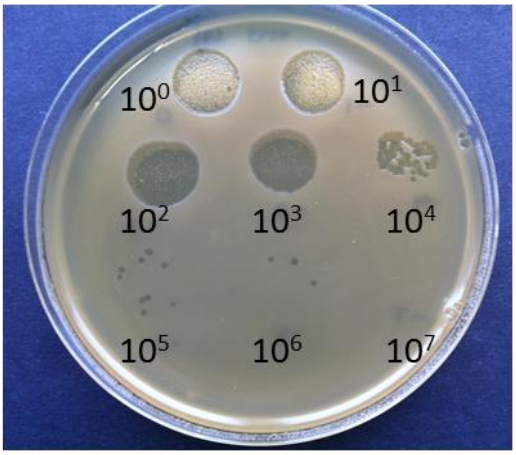


NCTC9343

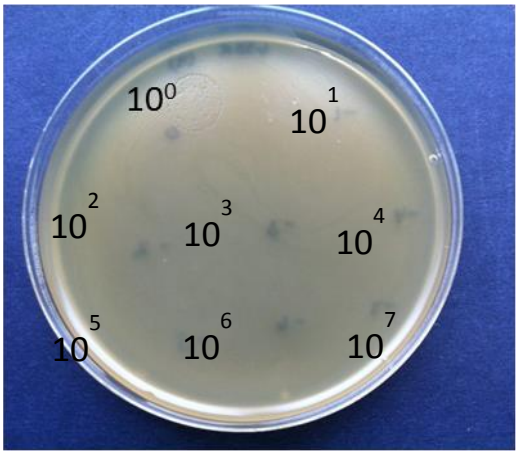


638R

ΦED09

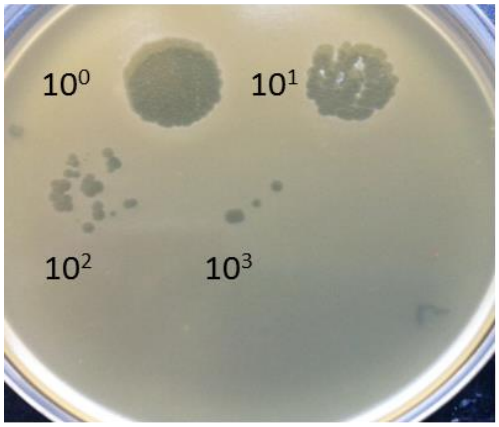


NCTC9343

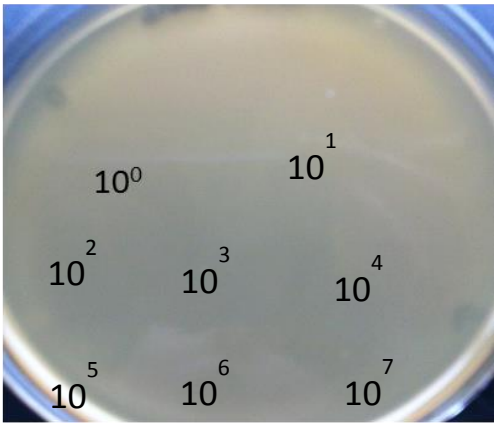


638R

ΦED10

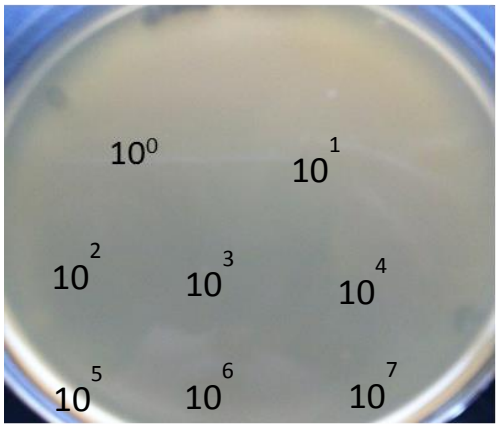


NCTC9343

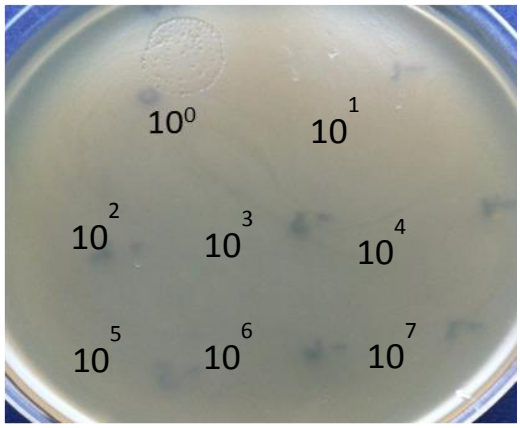


638R

ΦED11

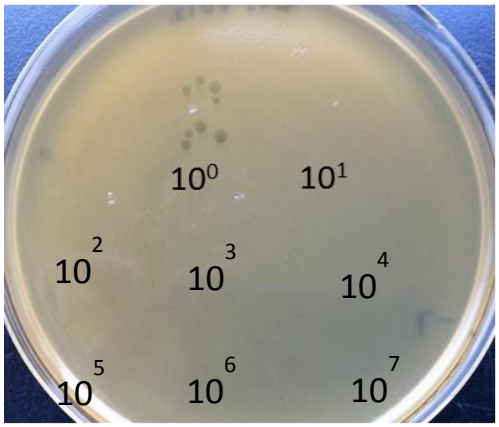


NCTC9343

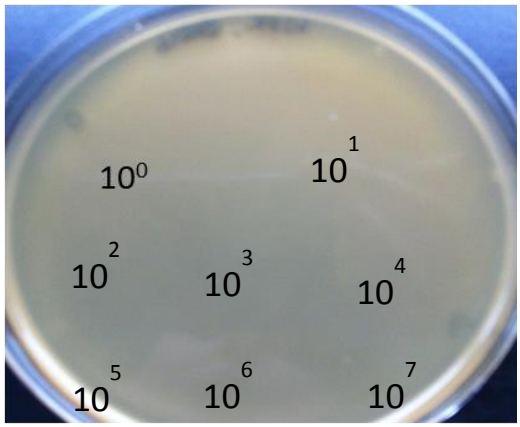


638R

ΦED12



NCTC9343



638R

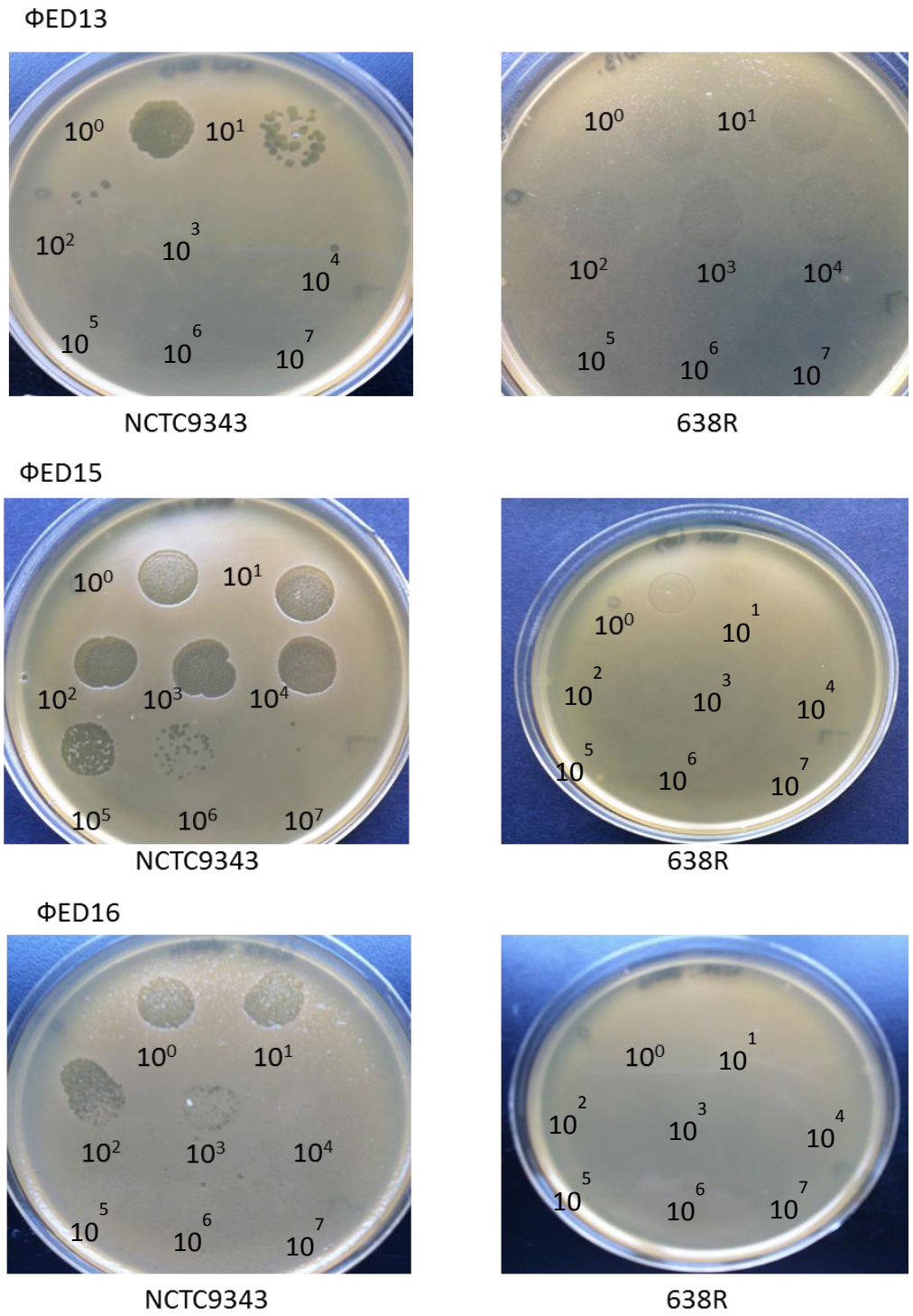


Figure 5.1: Bacteriophage titres on NCTC9343 and 638R

10µl of 10-fold dilutions of the indicated phage were spotted on soft agar overlay seeded lawns of NCTC9343 and 638R. Nine phages were able to infect NCTC9343, one infected 638R and five infected both strains. Four phages, ΦED01, ΦED09, ΦED13 and ΦED15 appeared to infect 638R with only a partial zone of clearance, indicating either a lysogenic lifestyle or host resistance to infection.

5.2.2. Restriction endonuclease digestion of bacteriophage DNA

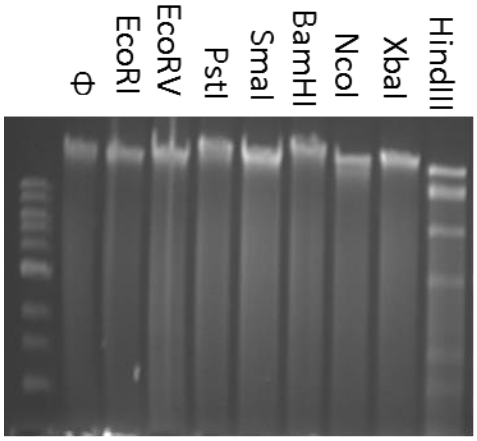
Restriction endonuclease mapping of phage DNA was carried out with the aim of identifying which of the isolated bacteriophages were related to ΦED01 or which differ. Phage DNA was extracted from CsCl purified lysates using formamide (see section 2.11.4). Purified DNA was digested individually with EcoRI, EcoRV, PstI, SmaI, BamHI, NcoI, XbaI and HindIII, with the exception of ΦED02 and ΦED13; DNA isolated from these phage was digested with all enzymes apart from XbaI (Figure 5.3 ΦED02 and ΦED13). The previously sequenced phage, ΦED01, was only sensitive to cleavage by HindIII, and six DNA fragments were observed when the digest was analysed by agarose gel electrophoresis (Figure 5.2 ΦED01). The genome sequence of ΦED01 contained eight target sequences for HindIII, and also contained three target sequences recognised by SmaI, one recognised by NcoI and one recognised by XbaI, as determined by NEBcutter v2.0 (Vincze *et al.*, 2003), however, cleavage by these enzymes was not observed suggesting DNA modification. Restriction endonuclease profiles revealed 6 other phages, ΦED03, ΦED04, ΦED06, ΦED07, ΦED10 and ΦED15 that were only digested by HindIII (Figure 5.2). As with ΦED01, multiple DNA fragments, between four and seven, were visible by gel electrophoresis following incubation with HindIII, indicating that multiple recognition sites were present within the genomes of these phages. All of these phage were isolated from NCTC9343, and with the exception of ΦED01 and ΦED15, only infected NCTC9343 (Table 5.1). ΦED01 and ΦED15 appeared to infect 638R, but with a hazy zone of clearance when an undiluted sample was spotted onto a soft agar overlay containing 638R (Figure 5.1). This suggested these six bacteriophages may be identical to the previously sequenced phage, ΦED01.

Seven further bacteriophages, ΦED02, ΦED05, ΦED08, ΦED09, ΦED11, ΦED13 and ΦED16 had restriction endonuclease cleavage profiles which differed from ΦED01 (Figure 5.3). ΦED02 was isolated from NCTC9343 and was not capable of infecting 638R; this DNA was cleaved by PstI, SmaI and BamHI (Figure 5.3 ΦED02). ΦED05 was isolated from NCTC9343 and was cleaved by both HindIII, which produced six DNA fragments and NcoI, which produced two DNA fragments (Figure 5.3 ΦED05). Cleavage of ΦED05 DNA with HindIII produced a digestion pattern similar to that of ΦED01, which also contains a target sequence recognised by NcoI, suggesting that this phage could be the same as ΦED01 however, it may contain modified DNA. ΦED08 and ΦED11 were isolated from NCTC9343 and 638R respectively. DNA from these phage was resistant to digestion by all of the

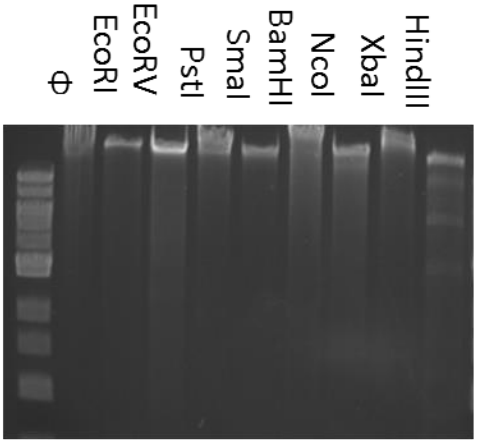
endonucleases tested (Figure 5.3 ΦED08 and ΦED11). Since ΦED11 was only able to infect 638R and ΦED08 was able to infect both NCTC9343 and 638R (Figure 5.1), the difference in host-range suggested they may be different. ΦED09 was isolated from NCTC9343 and infected 638R with a hazy zone of lysis rather than clear plaques. DNA isolated from this phage was susceptible to cleavage by EcoRI, SmaI and HindIII (Figure 5.3 ΦED09). Digestion of ΦED09 by HindIII resulted in three DNA fragments, which differs from ΦED01, in addition the genome of ΦED01 does not contain any target sequenced for EcoRI, suggesting this bacteriophage may be different from ΦED01. ΦED13 was originally isolated from a plate containing 638R but was also able to infect NCTC9343. This phage DNA was cleaved by HindIII and SmaI to generate five and three DNA fragments, respectively (Figure 5.3 ΦED13). The genome of ΦED01 contained three target sequences recognised by SmaI, suggesting that ΦED13 may be the same bacteriophage but contain polymorphisms. Finally, ΦED16 also appeared to have a different restriction endonuclease cleavage profile to the other bacteriophages. DNA isolated from ΦED16 was susceptible to cleavage by HindIII, NcoI and SmaI. Six DNA fragments were visible following digestion with HindIII, two with SmaI and two with NcoI (Figure 5.3 ΦED16). Since the genome of ΦED01 contained target sequences which are recognised by HindIII, SmaI and NcoI, it is possible that ΦED16 may be the same phage as ΦED01.

All of the isolated bacteriophages were susceptible to cleavage by HindIII, with the exception of ΦED08, ΦED11 and ΦED02. DNA from seven of the phages was only cleaved by this enzyme, suggesting they may be the same as bacteriophage ΦED01, however since the number of DNA fragments differed, some polymorphisms may be present. The remaining seven phages had unique restriction endonuclease cleavage profiles which differed from ΦED01, however, DNA from ΦED05, ΦED13 and ΦED16 was susceptible to cleavage by HindIII and at least one other restriction enzyme, the target sequence of which was also present in the genome of ΦED01. This suggested that these bacteriophages may be the same as ΦED01, but may contain DNA modifications. DNA isolated from ΦED09 was also susceptible to cleavage by HindIII, but was also cleaved by EcoRI. The genome of ΦED01 did not contain a target sequence recognised by this enzyme, suggesting this bacteriophage may be different. ΦED02 DNA was cleaved by BamHI and PstI, the target sequences of which were not present within the ΦED01 genome. DNA from ΦED08 and ΦED11 was resistant to cleavage from all restriction endonucleases tested, indicating they may also be different from ΦED01. Restriction endonuclease cleavage profiles suggested that of the

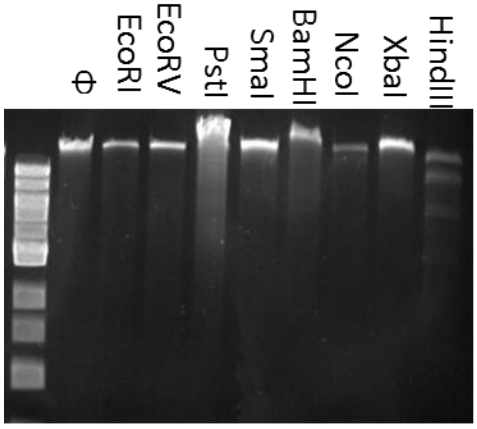
fourteen isolated bacteriophages, four new phages; ΦED02, ΦED08, ΦED09 and ΦED11 may have been identified. In addition, three phages, ΦED05, ΦED13 and ΦED16 had cleavage profiles which differed from ΦED01, therefore are potentially new bacteriophages. Sequencing of the bacteriophage genomes would confirm this and also determine if the bacteriophages which were susceptible to cleavage by HindIII alone were ΦED01.



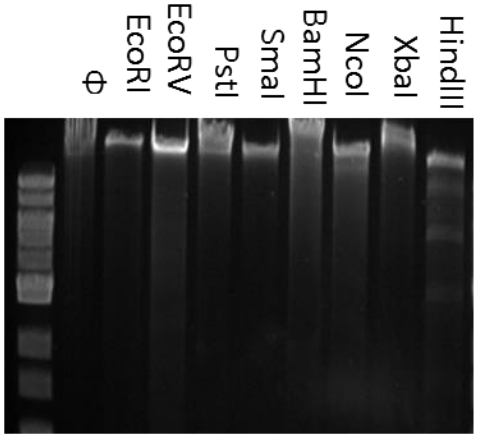
ΦED01



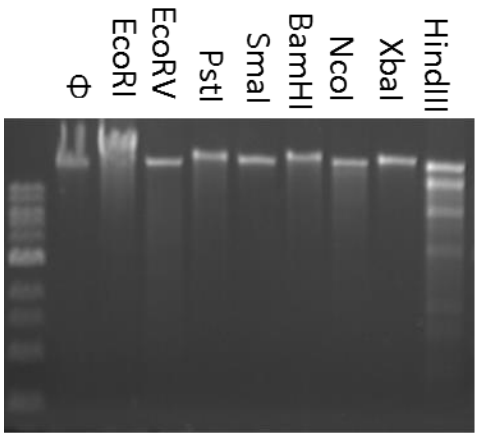
ΦED03



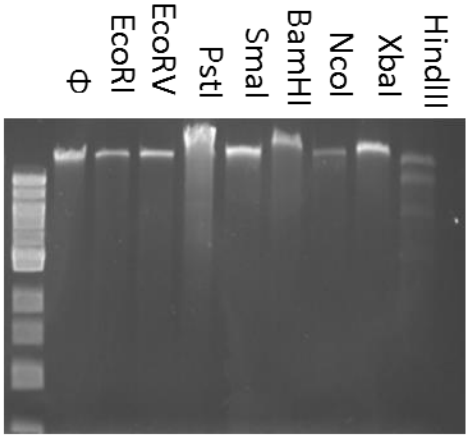
ΦED04



ΦED06



ΦED07



ΦED10

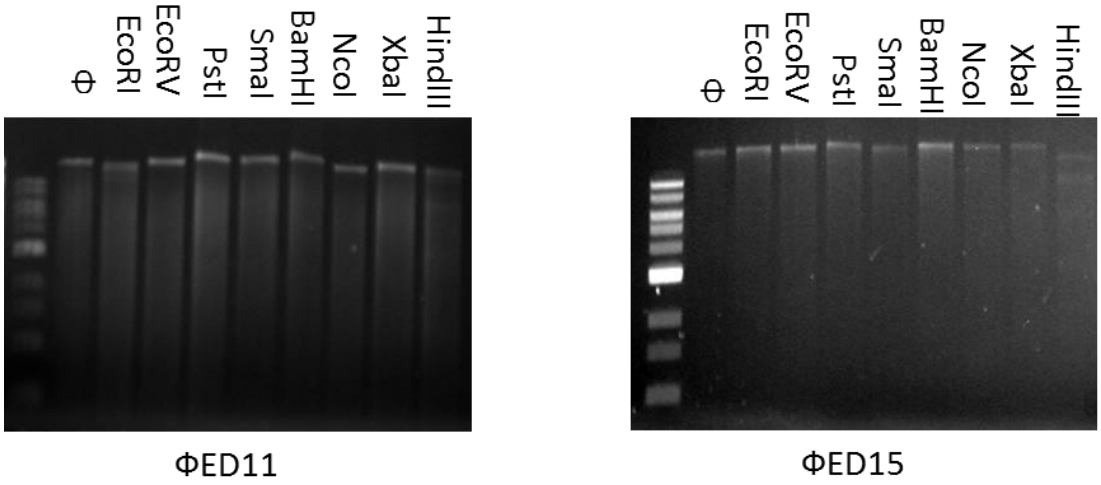
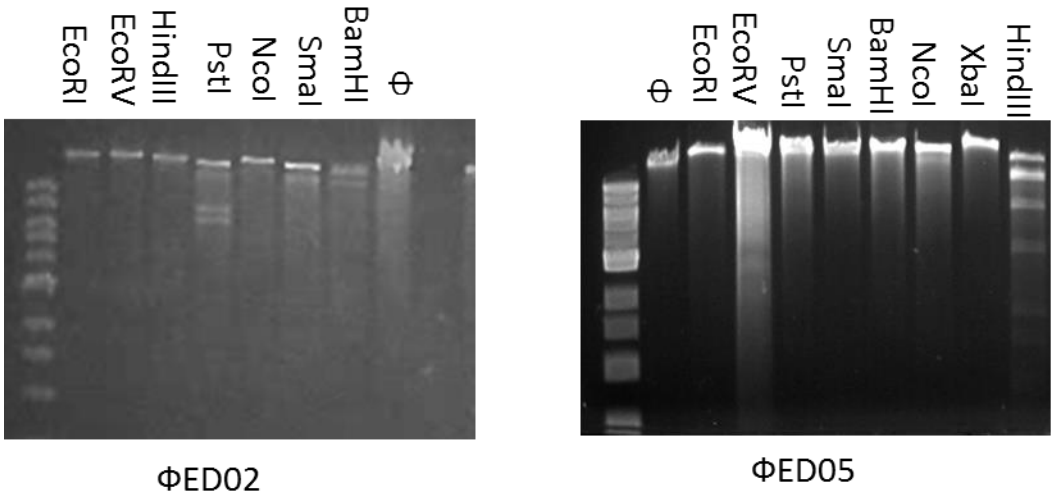


Figure 5.2: Restriction endonuclease cleavage profiles of bacteriophages similar to ΦED01

Agarose gels showing *B. fragilis*-specific bacteriophage DNA digested with a range of restriction enzymes for 1 hour at 37°C, with the exception of SmaI which was incubated at 25°C. The first lane of each gel contains molecular size markers. Phage DNA was digested with the restriction enzyme indicated. DNA from the seven isolated bacteriophages was only susceptible to cleavage by HindIII, suggesting they may be the same as bacteriophage ΦED01.



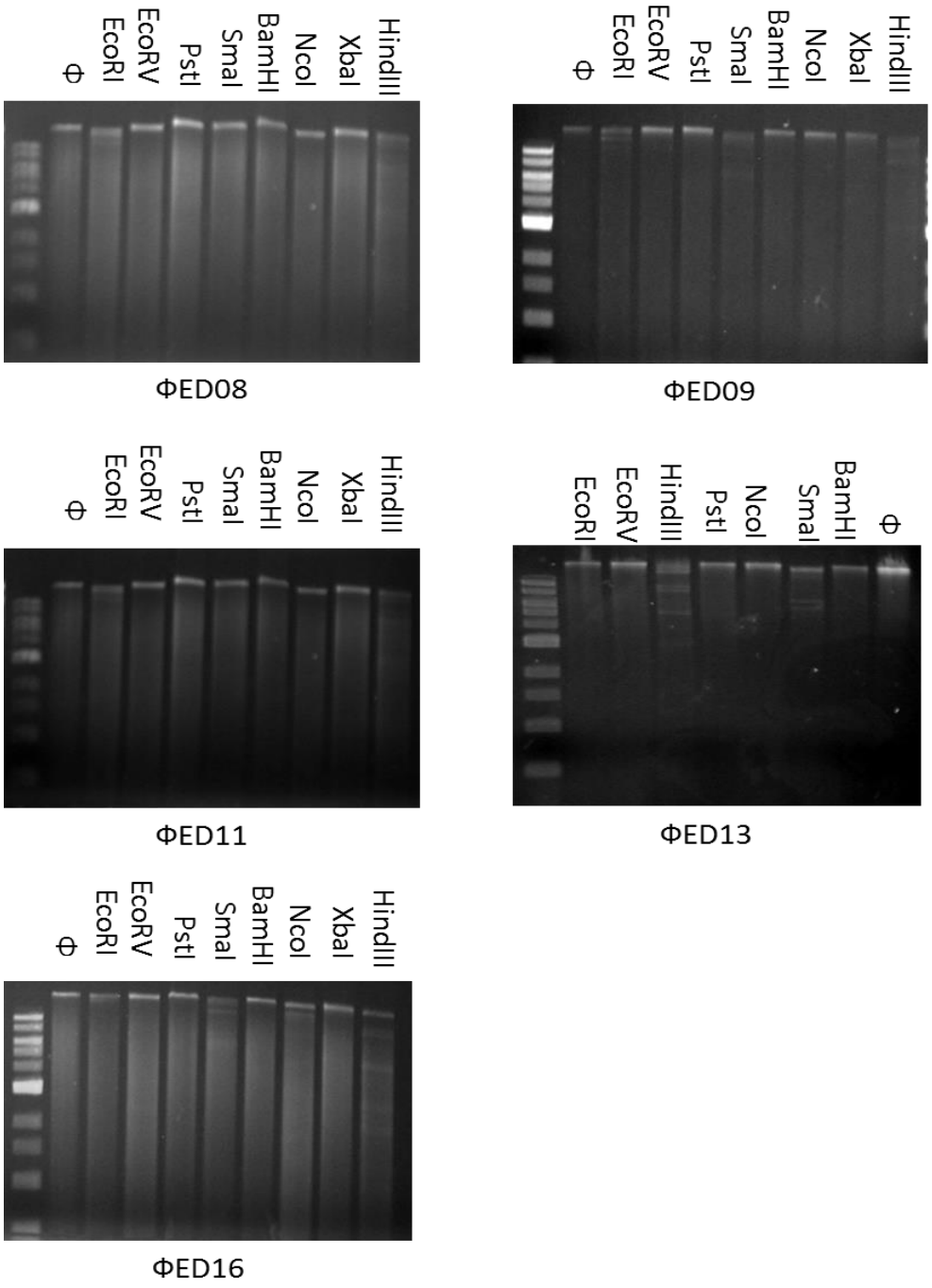


Figure 5.3: Bacteriophages with restriction endonuclease cleavage profiles different to ΦED01

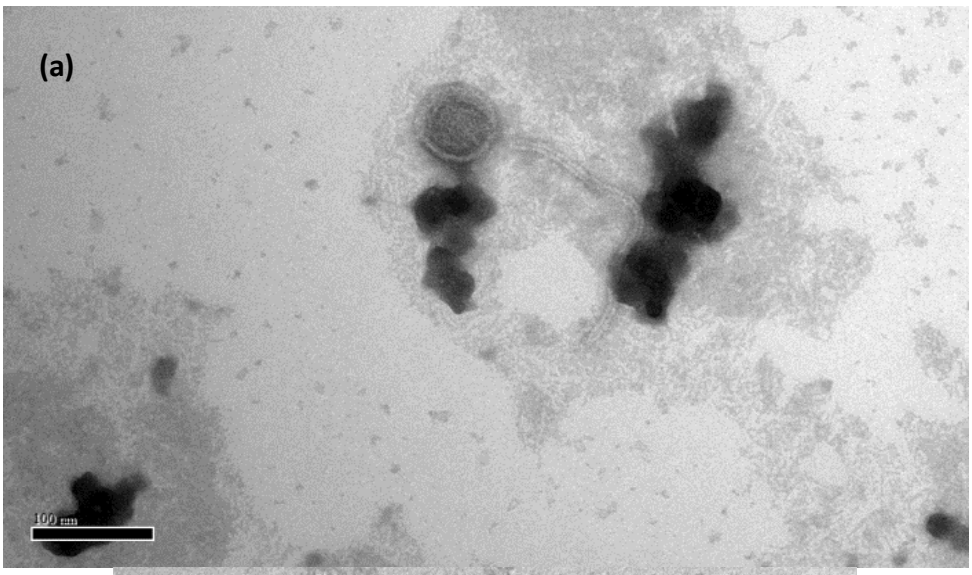
Agarose gels showing *B. fragilis*-specific bacteriophage DNA digested with a range of restriction enzymes for 1 hour at 37°C, with the exception of SmaI which was incubated at 25°C. The first lane of each gel contains molecular size markers. DNA was digested with the restriction enzyme indicated. DNA from the seven bacteriophages displayed a restriction endonuclease profile which differed from that of ΦED01. DNA from two phages, ΦED08 and ΦED11 was resistant to cleavage by all of the restriction endonucleases tested.

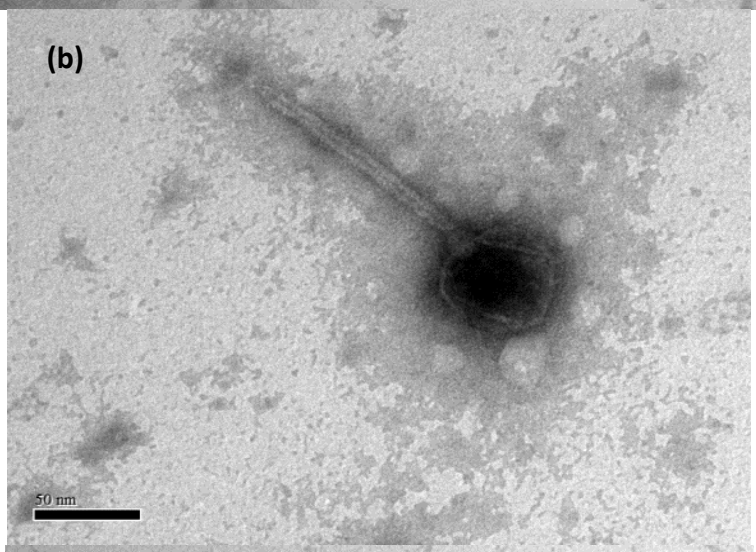
5.2.3. Transduction of *ermF* cassette

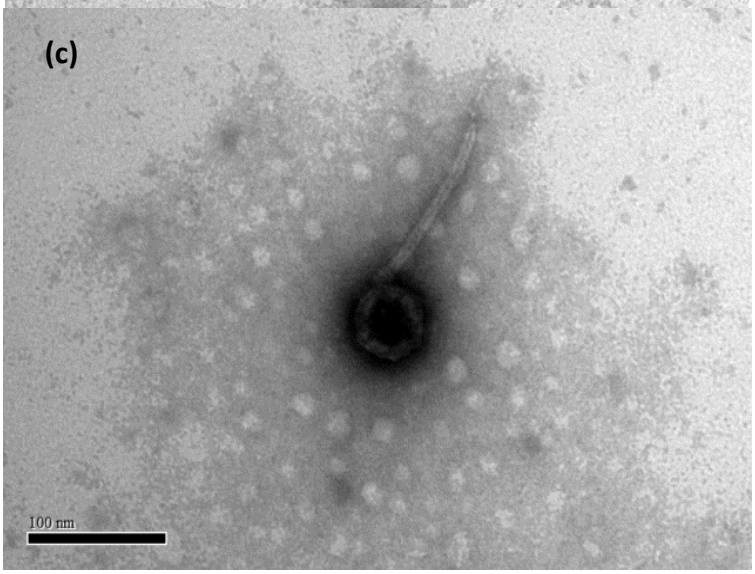
The ability of the isolated bacteriophages to transduce DNA between *B. fragilis* strains was tested by plating bacteriophage lysates onto soft agar overlays inoculated with a strain containing an erythromycin resistance gene and incubated anaerobically overnight before harvesting (see section 2.11.3). Dilutions of the bacteriophages (0.5µl neat, 1:5, 1:10 and 1:100), made in phage buffer, were incubated with 100 µl *B. fragilis* NCTC9343 or 638R, grown to OD_{600} 0.6 in BHI-S. Cells were incubated anaerobically for 15 minutes before the addition of 500µl BHI-S and incubation for a further 30 minutes. Cells were pelleted and re-suspended in 200µl BHI-S, and two 100 µl aliquots of each dilution were spread plated onto BHI-S erythromycin (10 µg/ml) plates and incubated for up to three days. Transduction experiments were repeated three times for each bacteriophage, but no erythromycin resistant colonies were identified. This indicated under these conditions, none of the identified bacteriophages were capable of transducing DNA containing the *ermF* gene.

5.2.4. Transmission electron microscopy

Transmission electron microscopy was performed to determine phage morphology and allow classification (Figure 5.4) (see section 2.9.4). CsCl purified phage lysate was diluted 1:100 in PBS, a drop was added to a carbon/formvar-coated grid, and stained with 2% uranyl acetate for 30 seconds. Negatively stained preparations of ΦED13 (Figure 5.4a), ΦED02 (Figure 5.4b), ΦED15 (Figure 5.4c), ΦED07 (Figure 5.4d), ΦED04 (Figure 5.4e) and ΦED08 (Figure 5.4f) clearly showed phage heads and tails of varying length. ΦED04 appeared to have the shortest tail, an average of ~96nm, taken from ten phages, over four fields of view. The tail lengths of ΦED02 and ΦED15 were similar (~120-130nm), an average measured from 5 and 10 bacteriophage respectively. The tails of ΦED13 appeared to be longer, however length was difficult to estimate as the tails were curled rather than straight (Figure 5.4a). ΦED07 and ΦED08 were always seen clumped together (Figure 5.4d) or in pairs (Figure 5.4f), therefore tail length was not determined. Tail striations were observed on all of the bacteriophages imaged.


(a)
100 nm


(b)
50 nm


(c)
100 nm

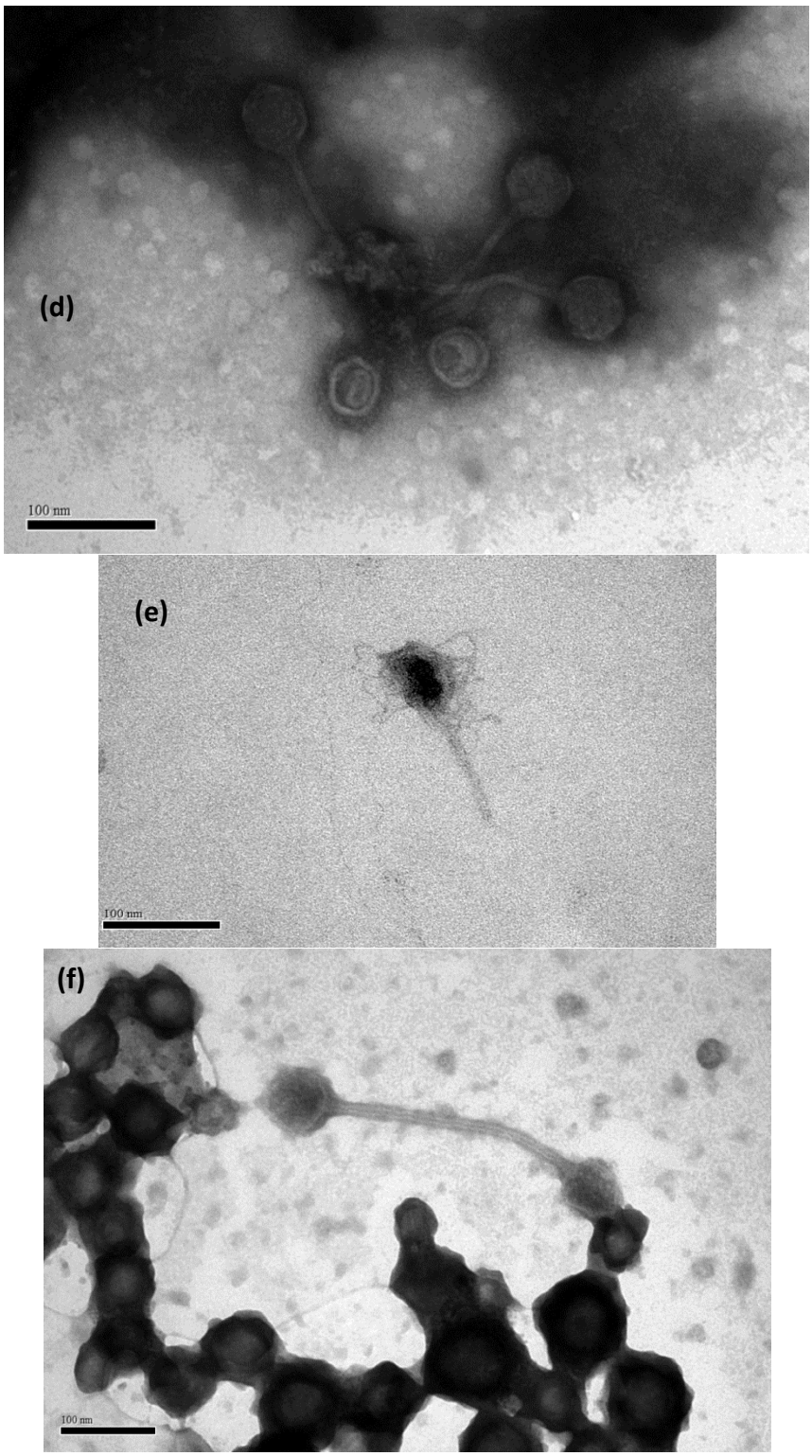


Figure 5.4: Transmission electron microscopy of *B. fragilis*-specific bacteriophage

Electron micrographs of bacteriophage negatively stained in 2% uranyl acetate (a) ΦED13, (b) ΦED02, (c) ΦED15, (d) ΦED07, (e) ΦED04 and (f) ΦED08. Tail length ranged from 95nm-130nm and striations were visible the tails of all phage imaged.

5.2.5. OMV interact with bacteriophage to decrease efficiency of infection

Manning & Kuehn (2011) hypothesised that bacterial OMV may act as a defence mechanism in response to external stressors, including bacteriophage infection. The authors demonstrated that the infectivity of T4 bacteriophage reduced following incubation with *E.coli* OMV. The ability of *B. fragilis* OMV to decrease the efficiency of infection of ΦED01 and ΦED13 was determined. 10^5 ΦED01 were co-incubated with 1µg NCTC9343 OMV and ΦED13 with both NCTC9343 OMV and 638R OMV, for 2 hours at 37°C. Ten-fold serial dilutions were subsequently spotted onto soft agar overlays containing 100µl of the appropriate *B. fragilis* strain and plates were incubated overnight under anaerobic conditions. As controls, ΦED01 and ΦED13 were incubated with 1xPBS under the same conditions before dilutions were plated onto NCTC9343 and 638R (Figure 5.5). The experiment was carried out in triplicate, plaques were counted and an average titre was calculated. Following incubation with PBS, the average titre of ΦED01 on NCTC9343 was $6x10^{11}$ PFU/ml; there was a ten-fold decrease in infection after pre-incubation of ΦED01 with NCTC9343 OMV which had a post-infection titre of $4x10^{10}$ PFU/ml (Figure 5.5a). ΦED13 was capable of infecting both NCTC9343 and 638R (Table 5.1, Figure 5.3 ΦED13), therefore was incubated with NCTC9343 and 638R OMV, respectively. The average titre of ΦED13, pre-incubated with PBS and spotted onto NCTC9343 was $6x10^5$ PFU/ml. Following incubation with NCTC9343 OMV no plaques were visible at any dilution spotted onto NCTC9343 (Figure 5.5b). ΦED13 infection of 638R did not produce clear plaques of lysis, instead hazy zones were visible within the lawn of cells (Figure 5.3 ΦED13), possibly due to the phage undergoing a lysogenic lifecycle or host cell resistance. These zones were observed following pre-incubation of ΦED13 with PBS to a dilution of 10^{-4}, however no hazy zones were visible following pre-incubation of the phage with 638R OMV (Figure 5.5c). This indicated that the interaction of *B. fragilis* OMVs with bacteriophage reduced infectivity.

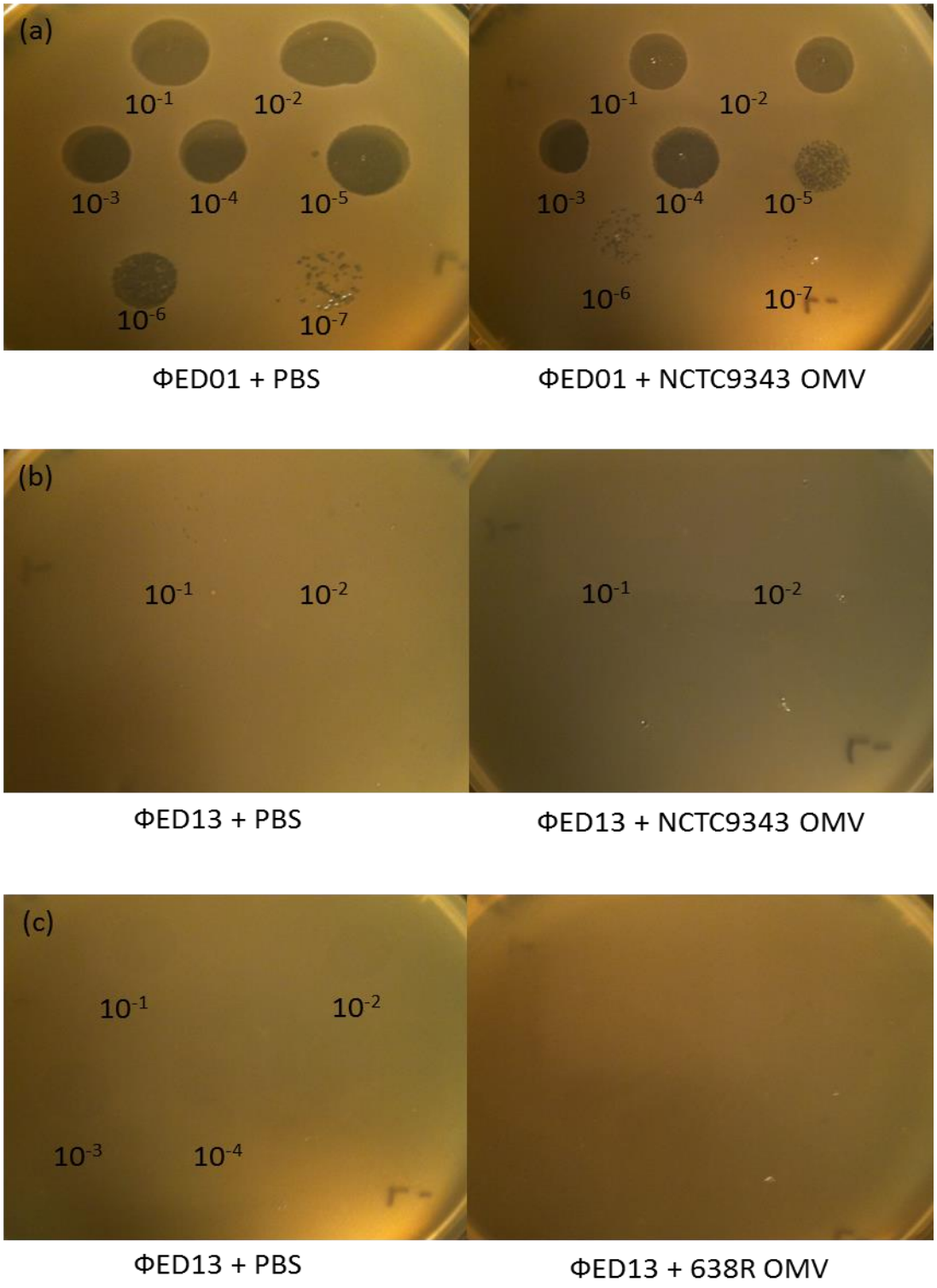


Figure 5.5: *B. fragilis* OMV interact with bacteriophage.

Ten-fold serial dilutions of bacteriophages pre-incubated with OMV spotted onto soft agar overlays containing NCTC9343 or 638R. (a) ΦED01 incubated with PBS and NCTC9343 OMV spotted onto NCTC9343, (b) ΦED13 incubated with PBS and NCTC9343 OMV spotted onto NCTC9343, (c) ΦED13 incubated with PBS and 638R OMV spotted onto 638R.

5.2.6. Development of a recombineering system in *B. fragilis*

Recombineering is a genetic engineering system that utilises bacteriophage encoded recombination proteins to generate recombinants. As it is based upon homologous recombination, it allows precise sequence alteration with only short segments of DNA homology (~ 50bp), eliminating the need for restriction enzymes and DNA ligase used in traditional genetic engineering (Sharan *et al.*, 2009). Usually, linear DNA with flanking regions encoding sequence homology to the target DNA is generated by PCR, and introduced to a strain containing the recombination system via electroporation (Yu *et al.*, 2000; Copeland *et al.*, 2001).

Recombineering in *E.coli* can be mediated either by the λ Red system, which involves three phage proteins, Exo, Beta and Gam, or the Rac-encoded RecET system, which contains two proteins, RecE and RecT (Court *et al.*, 2002). Exo and RecE are 5'→ 3'exonucleases which degrade linear double-stranded DNA (dsDNA) resulting in a 3' single-stranded DNA (ssDNA) overhang. The exposed 3' ends are bound by a ssDNA-annealing and strand invasion protein, Beta or RecT, which protect the target DNA and facilitate annealing to the homologous sequence (van Kessel *et al.*, 2008). The λ Red protein, Gam, enhances the efficiency of λ Red-mediated recombination by binding to the RecB subunit of the host RecBCD enzyme, therefore inhibiting the degradation of incoming linear dsDNA (Murphy, 1991; Copeland *et al.*, 2001). These systems have also been used for genetic manipulation of *Salmonella* and *Shigella* (Marinelli *et al.*, 2012), however they have been less successful in some other bacterial species (van Kessel & Hatfull, 2007; Swingle *et al.*, 2010).

Genes with homology to *bet* or *recT*, encoding putative DNA recombination proteins have been identified in phages and prophages of several Gram-positive and Gram-negative bacteria. A number of these genes are located next to a putative exonuclease gene, for example within *Bacillus subtillis*- and *Mycobacterium*-specific bacteriophages and in a prophage of *Pseudomonas syringae* (Datta *et al.*, 2008; van Kessel *et al.*, 2008; Swingle *et al.*, 2010). The use of these proteins in recombineering was demonstrated by van Kessel & Hartfull (2007), in which RecE and RecT homologues encoded by mycobacteriophage Che9c were utilised to create a recombination system, facilitating the genetic manipulation *of Mycobacterium* spp., including *M. smegmatis* and *M. tuberculosis.* Genes which encode proteins with homology to RecE and RecT have also been identified within a horizontally acquired prophage of *P. syringae* and are currently being used to develop a recombineering

system in this organism. It has been demonstrated that these proteins promote homologous recombination of linear dsDNA and ssDNA with the bacterial chromosome, following introduction to the cell via electroporation (Swingle *et al.*, 2010; Bao *et al.*, 2012).

Genetic manipulation of *B. fragilis* is problematic, mainly due to the extensive R-M systems which recognise and cleave foreign DNA. Therefore the development of alternative strategies, such as recombineering, to simplify manipulation of this organism would be advantageous. Analysis of the genome of *B. fragilis*-specific phage, ΦED01, revealed two genes which encoded proteins with similarity to RecE and RecT (Figure 5.6a) (G.W. Blakely, personal communication). The first of these genes, *phirecT* (CDS 42965-43609), encoded a 214aa protein, PhiRecT, with 95% amino acid sequence identity to a putative essential recombination protein of *Bacteroides* phage B40-8. PhiRecT belonged to the ssDNA annealing proteins ERF superfamily, of which RecT and λ Beta are also members. The second gene, *phirecE* (CDS 43610-44596), encoded a 328aa protein, PhiRecE, which contained a domain from the YqaJ family of phage-encoded exonucleases (Figure 5.6b). These genes were utilised with the aim of developing a recombineering system in *B. fragilis*.

5.6.1.1. Construction of a PhiRecET expression vector

To determine the ability of these proteins to mediate homologous recombination within *B. fragilis*, a plasmid containing *phirecE* and *phirecT* (pKJ03) was constructed. pKJ03 was created using the shuttle vector, pLYL01, which confers ampicillin resistance in *E.coli* and tetracycline resistance in *B. fragilis*. The bacteriophage genes were placed under the control of *B. fragilis* RpsF ribosomal protein promoter, P*rpsF.* Primers were designed to amplify PrpsF and *phirecE-phirecT.* The two amplicons were isolated, then fused and amplified via crossover PCR using the two outside primers. The PrpsF and phirecET outside primers contained KpnI and HindIII restriction sites, respectively. The crossover PCR product and pLYL01 were digested with HindIII and KpnI before the fragments were ligated using T4 DNA ligase. *E.coli* strain S-17 was transformed with the plasmid and ampicillin resistant colonies were selected. Plasmid DNA was extracted from four colonies and analysed by PCR using the two outside primers to amplify the bacteriophage genes and subsequent restriction endonuclease analysis was carried out with HindIII and KpnI (Figure 5.6c). Amplicons of ~1.8kb were detected following PCR with DNA extracted from clones 2-4 (Figure 5.6ci lanes 3-5). These fragments were similar in size to the digested insert DNA used in the ligation (1770 bp) (Figure 5.6ci lane 2). Following digestion of pKJ03 DNA,

extracted from clone 3, with KpnI and HindIII (Figure 5.5cii), two DNA fragments of approximately 8 kb and 1.8 kb were observed. This correlated with the size of pLYL01 (~8 kb) and the size of the DNA insert encoding PrpsF, phirecE and phirecT (1.77 kb). The construct was verified by Sanger sequencing. Transformation of *B. fragilis* NCTC9343 with pKJ03 was attempted via electroporation and conjugation (see section 2.6.5.) six times, however, no tetracycline resistant colonies were detected. *B. fragilis* NCTC9343 was transformed with pLYL01 as a control, on each occasion tetracycline resistant colonies were detected. As a result, the ability of PhiRecE and PhiRecT to mediate homologous recombination within *B. fragilis* could not be assessed.

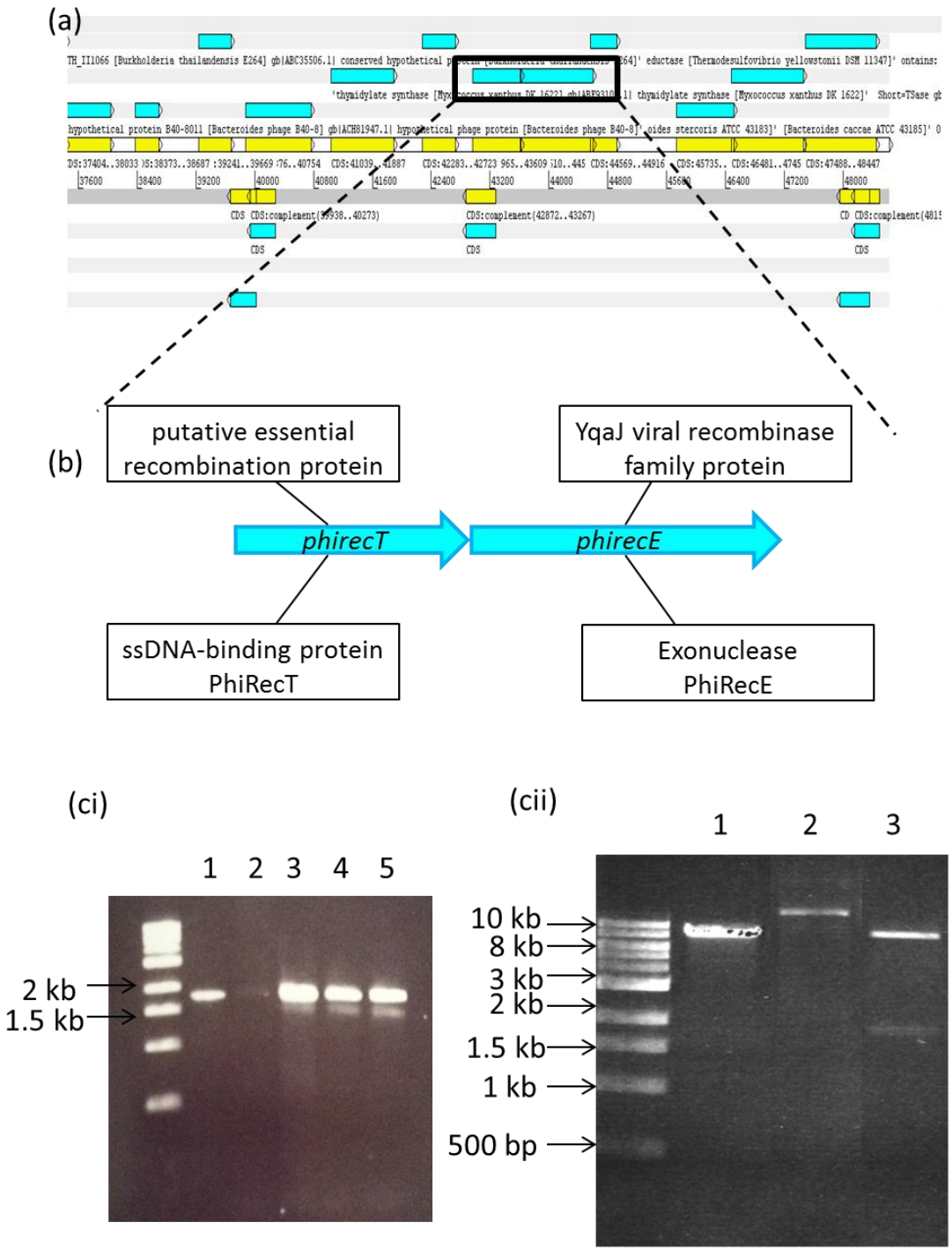


Figure 5.6: *phirecE* and *phirecT* genes of *B. fragilis*-specific bacteriophage ΦED01

(a) Artemis genome browser showing *phirecE* and *phirecT* encoded by ΦED01 (b) PhiRecE is a 328aa exonuclease, PhiRecT is a 214 aa, putative essential recombination protein with homology to RecT. (c) *E.coli* strain S-17 was transformed with pKJ03, plasmid DNA extracted from ampicillin resistant clones was analysed by (ci) PCR amplification (cii) restriction digest and agarose gel electrophoresis to confirm the presence of *phirecE-phirecT* insert. The first lane of each gel contains DNA molecular weight markers. (ci) Lanes: 1, *phirecE-phirecT*

insert DNA; 2, PCR product of clone 1; 3, PCR product of clone 2; 4, PCR product of clone 3; 5, PCR product of clone 4. (cii) restriction digest analysis Lanes: 1, uncut pLYL01 DNA; 2, uncut pKJ03 DNA; 3, pKJ03 DNA cut with HindIII and KpnI.

5.5. Conclusion

Advances in whole genome sequencing have revealed that bacteriophages play an important role in the evolution of bacteria. Lytic infection of a host by bacteriophages results in multiplication of the phage and lysis of the host cell, however, lysogenic infection results in the integration of the bacteriophage genome into the bacterial chromosome as a prophage (Canchaya *et al.*, 2003a). Many of the sequenced bacterial genomes contain multiple prophages, which can contribute up to 10% of the total DNA in some organisms (Brüssow & Hendrix, 2002; Canchaya *et al.*, 2003b). Bacteriophages can contribute to the pathogenicity of a bacterium since a number of bacterial virulence factors are phage encoded, for example the cholera toxin of *Vibrio cholera*e or the Shiga toxins of enterohemorrhagic *E. coli* (EHEC) (Wagner & Waldor, 2002). Bacteriophages are also important mediators of HGT between bacterial species but can also account for inter-strain genetic variability within the same species (Brüssow *et al.*, 2004). One extreme example of this was the comparison of the genome of foodborne pathogen *E.coli* O157 to the genome of the reference strain *E. coli* K12. The genomes of O157and K12 share approximately 4.1 Mb of highly conserved sequence, however O157 contains 1.3 Mb of DNA not found in the reference sequence. Bacteriophages have played a prominent role in diversity; the genome of O157 contains 18 prophages and prophage remnants which account for half of the DNA identified in this strain that is absent from K12 (Ohnishi *et al.*, 2001; Brüssow & Hendrix, 2002). Patrick *et a*l (2010) demonstrated that the genomes of *B. fragilis* NCTC9343, 638R and YCH46 contained a number of phage-related regions. This observation was also made for the genomes of the four newly sequenced strains described in Chapter 3. This demonstrates that bacteriophages have played a role in the within-strain variation of this organism.

Due to their ability to mediate the transfer of DNA between bacterial cells, bacteriophages have been utilised in the development of a number of molecular techniques. Transducing phages are a useful tool for genetic engineering and genomic analysis of their host bacteria and they have been identified for a number of bacterial species, including *E. coli* and *Salmonella*. Due to the difficulties of genetic manipulation of *B. fragilis*, the identification of a generalised transducing phage for this organism would be a useful tool.

Fifteen potentially new *B. fragilis*-specific bacteriophages were isolated from sewage from the Edinburgh area. Phage titres and restriction endonuclease cleavage profiles indicated

that seven of these bacteriophages may be similar to a previously identified phage, ΦED01. A further seven bacteriophages displayed unique restriction endonuclease profiles which differed to both ΦED01, suggesting they were potentially new bacteriophages. However, further analysis of this by genome sequencing would be required to confirm this. Unfortunately, none of the identified bacteriophage were capable of transduction, meaning to date, a generalised transducing phage has not been identified for *B. fragilis*.

In addition to transduction, bacteriophages have also been utilised to create other genetic tools, such as recombineering (see section 5.2.6). The genome of ΦED01 contained two genes which encoded homologues of the RecE and RecT proteins of the RecET system. An expression vector containing these genes was constructed to determine the potential to develop a recombineering system for *B. fragilis*. Attempts to transform *B. fragilis* NCTC9343 with this plasmid were unsuccessful, however, these proteins may have been toxic to host cells, especially since a potentially strong promoter was used for expression. The construct may have interfered with homologous recombination of replication forks within the cell. It is possible that the bacteriophages isolated in this study may also contain genes encoding RecET homologues, therefore the sequencing of these genomes would be a worthwhile endeavour.

6. General Discussion

The human gastrointestinal tract is colonised by a large population of diverse bacteria, which is estimated to be $>10^{14}$ organisms, representing >1000 different species (Liu *et al.*, 2008). Most of the interactions between these bacteria and the host are mutually beneficial rather than pathogenic. Commensal organisms inhabit a protected, nutrient-rich environment and in turn they play a role in the development of the host immune system and carry out essential functions within the host (Troy & Kasper, 2010). One function performed by the intestinal microbiota, including *B. fragilis,* is the degradation of complex plant polysaccharides, which are otherwise non-digestible by the host (Hooper & Macpherson, 2010). Degradation of these polysaccharides into short-chain fatty acids provides the host with ≥10% of their daily absorbed calories, and in return the bacteria are provided with an abundant carbon source (Hooper *et al.*, 2002). Commensal organisms, such as the *Bacteroides,* have evolved to harvest these nutrients as demonstrated by the presence of the large number of genes which encode carbohydrate-degrading enzymes (Xu & Gordon, 2003). For example, the genome of *B. thetaiotaomicron* encodes 64 enzymes known to degrade xylan, pectin or arabinose, which are common components of plant cell walls and are indigestible by the host (Sonnenburg *et al.*, 2005).

In addition to carbohydrate degradation, several members of the intestinal *Bacteroidales,* including *B. thetaiotaomicron, B. uniformis* and *B. fragilis,* encode a large number of polysaccharide biosynthesis loci, from which they can synthesise and express capsular polysaccharides (Coyne & Comstock, 2008). Polysaccharides produced by *B. fragilis* are important in the development and activation of the host immune system. PSA, a zwitterionic polysaccharide produced by some strains of *B. fragilis,* induces the splenic development of CD4+ T cells via IL-10 stimulation in mice (Mazmanian *et al.*, 2005; Mazmanian *et al.*, 2008; Round & Mazmanian, 2010). This induces an immunotolerance effect, which protects mice from experimental colitis and promotes colonisation by *B. fragilis* (Round *et al.*, 2011). PSA interacts with the host via OMV, which are internalised by dendritic cells. OMVs isolated from a PSA deletion strain do not stimulate anti-inflammatory cytokine production (Shen *et al.*, 2012). In addition to interactions with the host immune system, *B. fragilis* capsular polysaccharides are important in colonisation of the GI-tract. Coyne *et al* (2008) demonstrated that the presence of a single polysaccharide was sufficient for a mutant strain to effectively compete with wild-type *B. fragilis* when co-

inoculated into a gnotobiotic mouse. However, these findings were contradicted by evidence which suggested strains expressing a single polysaccharide were not competitive with wild-type *B. fragilis* for colonisation of the GI-tract (Liu *et al.*, 2008). These authors demonstrated that mutant strains expressing only PSA, PSB or PSC were quickly outcompeted by wild-type *B. fragilis*. Mutants expressing only one polysaccharide also displayed growth attenuation, suggesting multiple capsular polysaccharides are not only required for effective colonisation of the GI-tract by *B. fragilis* but also growth of the organism.

The genomes of three previously sequenced *B. fragilis* strains, NCTC9343, 638R and YCH46 each encoded ten polysaccharide biosynthesis loci, only two of which were conserved between two strains (Patrick *et al.*, 2010). The genomes of four further *B. fragilis* strains, RD48, BE1, GNAB92 and LS66 were sequenced and analysed during this study. BE1 and GNAB92 contained ten polysaccharide biosynthesis loci divergent from those previously identified in NCTC9343, 638R and YCH46, meaning, between five *B. fragilis* strains there is the potential to express 38 different polysaccharides. A further three possible divergent polysaccharide biosynthesis loci were identified within the genome of strain LS66. However, analysis of this strain was limited due to problems with genome assembly. The expression of such a large number of capsular polysaccharides has not been observed before. Coyne & Comstock (2008) demonstrated that the variation in polysaccharide expression is not conserved in members of the *Bacteroidales* which colonise the oral cavity, suggesting there is a selective pressure for polysaccharide diversity within the GI-tract.

Bacteriophages are estimated to outnumber their bacterial hosts ten-fold (Brüssow & Hendrix, 2002), therefore, to survive in the GI-tract, commensal organisms have evolved strategies to avoid bacteriophage infection. Due to the variation in bacteriophage genomes and their high rates of evolution, bacteria need diverse anti-phage mechanisms to prevent processes such as bacteriophage adsorption or entry and expression of foreign DNA into the bacterial cell. Finally, if bacteriophages circumvent these barriers then abortive infection systems trigger bacterial cell 'suicide', preventing bacteriophage multiplication (Bikard & Marraffini, 2012). The first step in bacteriophage infection is adsorption to the host cell, in which phages recognise a specific-host cell component (Weinbauer, 2004). Bacterial capsules can present an effective physical barrier to bacteriophage adsorption, for example, *E. coli* strains expressing the K1 polysaccharide capsule are resistant to infection

by bacteriophage T7. Removal of the capsule allows adsorption and replication (Scholl *et al.*, 2005). Some bacteriophages have evolved ways to penetrate the bacterial capsule, such as the expression of polysaccharide degrading enzymes (Sutherland *et al.*, 2004), however, it has been demonstrated that these enzymes are very specific and only act on a small number of closely related polysaccharide structures (Hughes *et al.*, 1998). Phase variation of capsular polysaccharides may be a mechanism to avoid the degradative enzymes encoded by bacteriophages. Variation in the surface components of the bacterial cell can also block host-specific bacteriophage receptors, therefore preventing adsorption. Different bacterial surface components have been identified as receptors for bacteriophages, including LPS and capsular polysaccharides (Xu *et al.*, 2013). LPS has been shown to act as receptor for phage which infect several Gram-negative bacteria, including *E. coli*, *Salmonella*, *V. cholerae* and *Campylobacter jejuni* (Lindberg & Hellerqvist, 1971; Zhang *et al.*, 2009; Sørensen *et al.*, 2011; Xu *et al.*, 2013). Lytic bacteriophage VP4 which infects *V. cholerae* serogroup O1 uses the O-antigen as its receptor. Xu *et al* (2013) demonstrated that mutations in the O-antigen biosynthesis genes resulted in changes to the O-antigen and prevented adsorption of the phage. The capsular polysaccharides of *C. jejuni* have been identified as the binding receptor of multiple *C. jejuni*-specific bacteriophages and mutations in several of the capsular polysaccharide genes resulted in phage resistance (Coward *et al.*, 2006). Sørensen *et al* (2012) demonstrated that the phase variable expression of a moiety present in the capsular polysaccharide of *C. jejuni* prevented the adsorption of bacteriophage F336. The variation in surface components of *B. fragilis* may similarly convey resistance to bacteriophage infection in the GI-tract by removing the host-specific bacteriophage receptors.

Following adsorption, bacteriophage DNA is injected into the bacterial cell. To prevent this, bacteria encode R-M systems which protect the cell from incoming foreign DNA. However, bacteriophages have evolved mechanisms to subvert these systems. Host cell DNA is methylated, therefore not recognised and cleaved by R-M systems, unlike unmethylated bacteriophage DNA. However, when phage DNA is methylated it is no longer recognised by the bacterial R-M systems, the new virions produced will be resistant to restriction and can easily infect cells containing the same R-M system. The bacteriophage DNA will remain methylated until the phage encounters a cell encoding a different R-M system, at which point the DNA once again becomes susceptible to restriction (Bikard & Marraffini, 2012). Some bacteriophages have acquired methylase encoding genes or can stimulate host

encoded methylases to confer protection to the bacteriophage DNA and so overcome host R-M systems (Kruger & Bickle, 1983; McGrath *et al.*, 1999). Diversity in R-M systems may be a mechanism by which bacteria can counteract this phage resistance mechanism. Diverse R-M systems increase the likelihood that methylated bacteriophage DNA will encounter a different R-M system and once again become susceptible to cleavage. Bacteriophages can also avoid restriction systems by changes to the DNA sequence which subsequently remove endonuclease restriction sites, usually via the accumulation of point mutations (Labrie *et al.*, 2010). A reduction in the number of sites can enable phage to avoid cleavage, for example, the enzyme EcoRII can only cleave when it binds two target sequences, however, the large distance between EcoRII sites in the genomes of phage T3 and phage T7 results in resistance to cleavage by the endonuclease (Tock & Dryden, 2005). Restriction endonuclease sites can also be completely absent from bacteriophage genomes, for example, *Staphylococcus* phage K has no Sau3A sites within its genome (Labrie *et al.*, 2010). Another bacteriophage anti-restriction system is the presence of unusual bases within the genome. Bacteriophage T4 is nearly devoid of cytosine residues, which are replaced by the unusual base hydroxymethylcytosine, T4 is therefore resistant to R-M systems that recognise sequences containing a cytosine (Stern & Sorek, 2011). Some bacteriophages encode proteins which target and inhibit host restriction endonuclease genes. Ocr (overcomes classical restriction activity) is one of the first proteins expressed by bacteriophage T7 following injection of its genome into the host cell. Ocr binds to the active site of type I R-M enzymes and inhibits their activity, therefore preventing restriction of the phage genome (Atanasiu *et al.*, 2002). A similar protein encoded by the bacteriophage T3 inhibits the activity of type I R-M systems and type III R-M systems (Labrie *et al.*, 2010).

B. fragilis strains NCTC9343, 638R and YCH46 encode a number of R-M systems which act as a barrier to bacteriophage infection. NCTC9343 encodes a type I R-M shufflon which can recognise eight different DNA specificities (Cerdeño-Tárraga *et al.*, 2005). The ability to quickly switch between DNA specificities could provide an advantage to the organism during phage infection. Type I R-M systems are composed of three subunits: a restriction enzyme (HsdR), a methylase (HsdM) and a specificity determinant (HsdS), which determines the DNA sequence to bind (Murray, 2000). The *B. fragilis* NCTC9343 shufflon in the *hsdS* gene has eight recombination sites which could potentially lead to the formation of eight different *hsdS* genes, therefore eight different DNA specificities. Another type I R-M shufflon was identified in *B. fragilis* 638R (Patrick *et al.*, 2010) and a similar shufflon was

identified in the genome of the newly sequenced strain GNAB92. In addition to the type I shufflons, the genomes of *B. fragilis* NCTC9343, 638R, YCH46 and strains sequenced in this study, GNAB92, BE1, RD48 and LS66 contain a number of other type I, type II and type III R-M systems. The variation in restriction systems present in *B. fragilis* may enhance survival of the organism by preventing bacteriophage infection. Another mechanism by which bacteria resist bacteriophage infection is through abortive infection systems, which are toxic and lead to the death of the bacterial cell thus preventing bacteriophage proliferation and protecting the population from infection (Labrie *et al.*, 2010). The Rex abortive infection system is found in *E. coli* strains which are susceptible to bacteriophage λ. Infection results in the activation of RexB, which produces an ion channel that reduces membrane potential. This leads to a reduction in cellular ATP and cell division is halted. In turn phage infection, which requires ATP, will also abort (Parma *et al.*, 1992).

Bacteria-encoded phage defence mechanisms display a high level of genetic variability, driven by the rapid evolution of bacteriophage genomes, which places a selective pressure upon bacteria to evolve strategies to evade infection and killing by bacteriophages (Stern & Sorek, 2011). The variation in polysaccharide biosynthesis loci and R-M systems observed in the three previously sequenced genomes of *B. fragilis,* and the four genomes sequenced and analysed in this study, may convey resistance to bacteriophage infection and therefore promote survival of the organism in the GI-tract.

In addition to being a member of the gut microbiota*, B. fragilis* is an opportunistic pathogen and the most commonly isolated anaerobic organism from intra-abdominal abscesses (Wexler, 2007). The capsular polysaccharides of *B. fragilis* are a major virulence factor and at least one of these polysaccharides had been linked to abscess formation. Coyne *et al* (2000) demonstrated that PSB, purified from *B. fragilis* NCTC9343, alone induced abscess formation. The MC is also involved in complement resistance and the large capsule has been associated with resistance to phagocytic uptake and killing (Patrick *et al.*, 2010). Variation in the cell surface components of *B. fragilis* may also enhance the survival of the organism when it escapes the GI-tract and causes infection, however, in comparison to other *Bacteroidetes* which also encode phase-variable polysaccharides*, B. fragilis* is more commonly isolated from opportunistic infections. This may be due in part to the presence of other horizontally acquired virulence factors.

Genome sequencing has revealed that a number of bacterial virulence factors are encoded by horizontally acquired mobile genetic elements such as prophages, plasmids and genomic islands (Pallen & Wren, 2007). For example, the anthrax toxin of *Bacillus anthracis* and the Type III secretion system of *Shigella* spp. are plasmid encoded (Hueck, 1998; Okinaka *et al.*, 1999), while the Type III secretion systems of enteropathogenic and enterohaemorrhagic *E. coli* are encoded by a pathogenicity island (Elliott *et al.*, 1998). A number of virulence factors are encoded by bacteriophage genomes which have integrated into the host chromosome such as diphtheria toxin of *Corynebacterium diphtheriae* and cholera toxin of *V. cholerae* (Freeman, 1951; Wagner & Waldor, 2002). In addition to capsular polysaccharides, *B. fragilis* has a number of other virulence factors which may play a role in its success as an opportunistic pathogen, including the release of degradative enzymes such a fibrinogenolysin and enterotoxin production (Patrick *et al.*, 2010). *B. fragilis* induces abscess formation, which is an important host mechanism to contain bacteria and limit the spread of infection. The main component involved in abscess formation is fibrinogen. *B. fragilis* strains NCTC9343, 638R and YCH46 which have been shown to bind human fibrinogen, encode a putative fibrinogen-binding protein and have fibrinogenolytic activity (Houston *et al.*, 2010). The genomes of the four strains sequenced in this study also encode putative fibrinogen-binding proteins. Houston *et al* (2010) demonstrated that supernatant isolated from LS66 degraded human fibrinogen, which demonstrates a potential mechanism by which *B. fragilis* evades host abscess formation so leading to dissemination of infection. A subset of *B. fragilis* strains, termed enterotoxigenic *B. fragilis,* encode an exotoxin (BFT) which has been associated with an acute form of disease, characterised by diarrhoea and loss of appetite in children, adults and livestock (Franco *et al.*, 1999). The enterotoxin gene has been identified in strains isolated in a number of different countries, including USA, Japan and Italy. *bft* was also detected in 18% of clinical isolates tested from the UK, Poland, France and Holland (Luczak *et al.*, 2001). The *bft* gene is found in a pathogenicity locus which is contained within a conjugative transposon (Franco, 2004). The genome of NCTC9343 encodes a similar conjugative transposon but does not contain the pathogenicity locus. Of the four newly sequenced genomes, only one, GNAB92 encoded the *bft* pathogenicity locus, while RD48 contained the same conjugative transposon identified in NCTC9343. The location of *bft* within a conjugative element implies this virulence factor was horizontally acquired and may contribute to *B. fragilis* pathogenicity. However, the

identification of the gene in only a small proportion of clinical isolates suggests there may not be a significant selective advantage to ensure dissemination of the gene.

B. fragilis encodes and expresses a unique homologue of the eukaryotic protein-modifier, ubiquitin. The *ubb* gene is located within an 11 kb, low GC content region in comparison to the backbone sequence and the closest DNA sequence homology is to a Grasshopper entomopox virus, suggesting acquisition of this gene was via inter-kingdom HGT. The *ubb* gene is predicted to encode a 76 amino acid protein, termed BfUbb, with 63% identity to human ubiquitin. This is the first example of ubiquitin being produced by a prokaryote. Ubiquitin plays a key role in a number of eukaryotic cellular processes, including protein degradation, endocytosis, DNA repair and regulation of the immune system. The role of BfUbb has yet to be determined, but *ubb* has evolved additional features which differentiate the encoded protein from human ubiquitin. BfUbb contains a signal sequence which results in the transport of the protein across the inner membrane to the periplasm of the cell; this in addition to the detection of BfUbb in concentrated supernatants suggests BfUbb may be packaged into OMV. Recently it has been shown that in addition to colonisation of the lumen of the GI-tract, *B. fragilis* also colonises the colonic crypts of mono-colonised mice (Round *et al.*, 2011), and *B. fragilis* OMV are phagocytosed by dendritic cells (Shen *et al.*, 2012). *B. fragilis* ubiquitin may be delivered to host cells by OMV and may interact with the host ubiquitylation pathway to aid the survival of the organism as a member of the resident microbiota or as an opportunistic pathogen.

Commensal bacteria have been shown to interfere with the host pro-inflammatory response, of which ubiquitin in an integral component, to maintain homeostasis in the GI-tract. Preserving the balance between regulation of the host immune response and inflammation is an essential function of the microbiota as overstimulation of anti-inflammatory pathways can lead to increased colonisation by pathogens and chronic inflammation can cause autoimmune diseases such as IBD (Kubinak & Round, 2012). Commensals can have an anti-inflammatory effect on the host immune system through the inhibition of NF-κB activation. For example, *B. thetaiotaomicron* and *Faecalibacterium prausnitzii* both act upon the NF-κB signalling pathway to reduce NF-κB activation, therefore dampening the host inflammatory response (Kelly *et al.*, 2004; Sokol *et al.*, 2008). BfUbb may have an anti-inflammatory effect of the host immune system, possibly through the inhibition of NF-κB activation, to aid in maintenance of gastrointestinal homeostasis. In

contrast, BfUbb may have a pro-inflammatory effect on the host immune system by interfering with the ubiquitylation pathway to stimulate NF-κB activation and therefore aid *B. fragilis* pathogenesis. In addition, the causes of IBD are not yet fully understood, although, a number of cellular signalling pathways, such as NOD2 have been implicated, in which ubiquitin plays an important role (Abbott *et al.*, 2004). Could BfUbb be a driver of autoimmune diseases by interfering with ubiquitin or ubiquitin-bound proteins within the host inflammatory pathways, therefore generating an autoimmune response? Furthermore, the mis-regulation of ubiquitin has been associated with a number of diseases such as cancer and neurodegenerative disorders. Could BfUbb be involved in the development of these diseases?

The organisms of the human GI-tract establish long-term, mutualistic relationships with the host. To survive within this specific niche, bacteria must successfully compete with other members of the microbiota for nutrients and space and withstand attacks from bacteriophages. *B. fragilis* has evolved a number of features which have aided the organism in establishing itself as a predominant member of the gastrointestinal microbiota. These features include the synthesis of phase-variable polysaccharides which play a role in colonisation and prevention of bacteriophage adsorption and diverse and variable R-M systems which may convey resistance to bacteriophage infection. *B. fragilis* has also evolved to utilise ingested plant-derived polysaccharides, allowing the organism access to nutrients not available to some other members of the microbiota. The presence of a number of mobile genetic elements such as transposons, insertion sequences and phage-related regions, within the genome of *B. fragilis* suggests this organism has undergone extensive horizontal gene transfer.

Bibliography

Abbott, D. W., A. Wilkins, J. M. Asara and L. C. Cantley (2004) The Crohn's Disease Protein, NOD2, Requires RIP2 in Order to Induce Ubiquitinylation of a Novel Site on NEMO. *Current Biology* **14**(24): 2217-2227.

Alaniz, R. C., B. L. Deatherage, J. C. Lara and B. T. Cookson (2007) Membrane vesicles are immunogenic facsimiles of *Salmonella* typhimurium that potently activate dendritic cells, prime B and T cell responses, and stimulate protective immunity in vivo. *J Immunol* **179**(11): 7692-7701.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389-3402.

Anderson, D. M. and D. W. Frank (2012) Five Mechanisms of Manipulation by Bacterial Effectors: A Ubiquitous Theme. *PLoS Pathog* **8**(8): e1002823.

Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Angot, A., A. Vergunst, S. Genin and N. Peeters (2007) Exploitation of Eukaryotic Ubiquitin Signaling Pathways by Effectors Translocated by Bacterial Type III and Type IV Secretion Systems. *PLoS Pathog* **3**(1): e3.

Atanasiu, C., T. J. Su, S. S. Sturrock and D. T. Dryden (2002) Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res* **30**(18): 3936-3944.

Austin, C. D., X. Wen, L. Gazzard, C. Nelson, R. H. Scheller and S. J. Scales (2005) Oxidizing potential of endosomes and lysosomes limits intracellular cleavage of disulfide-based antibody–drug conjugates. *Proceedings of the National Academy of Sciences of the United States of America* **102**(50): 17987-17992.

Aziz, R., D. Bartels, A. Best, M. DeJongh, T. Disz, R. Edwards, K. Formsma, S. Gerdes, E. Glass, M. Kubal, F. Meyer, G. Olsen, R. Olson, A. Osterman, R. Overbeek, L. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke and O. Zagnitko (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**(1): 75.

Backhed, F., R. E. Ley, J. L. Sonnenburg, D. A. Peterson and J. I. Gordon (2005) Host-bacterial mutualism in the human intestine. *Science* **307**(5717): 1915-1920.

Bao, Z., S. Cartinhour and B. Swingle (2012) Substrate and Target Sequence Length Influence RecTEpsy Recombineering Efficiency in *Pseudomonas syringae*. *PLoS One* **7**(11): e50617.

Beaber, J. W., B. Hochhut and M. K. Waldor (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J Bacteriol* **184**(15): 4259-4269.

Bentley, S. D., D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabbinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail, G. Samuel, I. C. Skovsted, M. S. Kaltoft, B. Barrell, P. R. Reeves, J. Parkhill and B. G. Spratt (2006) Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes. *PLoS Genet* **2**(3): e31.

Berg, J. O., L. Lindqvist, G. Andersson and C. E. Nord (1983) Neuraminidase in *Bacteroides fragilis*. *Appl Environ Microbiol* **46**(1): 75-80.

Betley, M. J. and J. J. Mekalanos (1985) Staphylococcal enterotoxin A is encoded by phage. *Science* **229**(4709): 185-187.

Bikard, D. and L. A. Marraffini (2012) Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. *Current Opinion in Immunology* **24**(1): 15-20.
Bomberger, J. M., D. P. MacEachran, B. A. Coutermarsh, S. Ye, G. A. O'Toole and B. A. Stanton (2009) Long-Distance Delivery of Bacterial Virulence Factors by *Pseudomonas aeruginosa* Outer Membrane Vesicles. *PLoS Pathog* **5**(4): e1000382.
Bomberger, J. M., S. Ye, D. P. MacEachran, K. Koeppen, R. L. Barnaby, G. A. O'Toole and B. A. Stanton (2011a) A <italic>Pseudomonas aeruginosa</italic> Toxin that Hijacks the Host Ubiquitin Proteolytic System. *PLoS Pathog* **7**(3): e1001325.
Bomberger, J. M., S. Ye, D. P. MacEachran, K. Koeppen, R. L. Barnaby, G. A. O'Toole and B. A. Stanton (2011b) A *Pseudomonas aeruginosa* Toxin that Hijacks the Host Ubiquitin Proteolytic System. *PLoS Pathog* **7**(3): e1001325.
Boname, J. M., M. Thomas, H. R. Stagg, P. Xu, J. Peng and P. J. Lehner (2010) Efficient Internalization of MHC I Requires Lysine-11 and Lysine-63 Mixed Linkage Polyubiquitin Chains. *Traffic* **11**(2): 210-220.
Boyd, E. F. and H. Brüssow (2002) Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends in Microbiology* **10**(11): 521-529.
Brandl, K., G. Plitas, B. Schnabl, R. P. DeMatteo and E. G. Pamer (2007) MyD88-mediated signals induce the bactericidal lectin RegIII gamma and protect mice against intestinal *Listeria monocytogenes* infection. *J Exp Med* **204**(8): 1891-1900.
Bremm, A. and D. Komander (2011) Emerging roles for Lys11-linked polyubiquitin in cellular regulation. *Trends in biochemical sciences* **36**(7): 355-363.
Brochet, M., E. Couvé, P. Glaser, G. Guédon and S. Payot (2008) Integrative Conjugative Elements and Related Elements Are Major Contributors to the Genome Diversity of *Streptococcus agalactiae*. *Journal of Bacteriology* **190**(20): 6913-6917.
Brüssow, H., C. Canchaya and W.-D. Hardt (2004) Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews* **68**(3): 560-602.
Brüssow, H. and R. W. Hendrix (2002) Phage Genomics: Small Is Beautiful. *Cell* **108**(1): 13-16.
Burrus, V. and M. K. Waldor (2004) Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology* **155**(5): 376-386.
Canchaya, C., G. Fournous, S. Chibani-Chennoufi, M.-L. Dillmann and H. Brüssow (2003a) Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* **6**(4): 417-424.
Canchaya, C., C. Proux, G. Fournous, A. Bruttin and H. Brüssow (2003b) Prophage Genomics. *Microbiology and Molecular Biology Reviews* **67**(2): 238-276.
Carver, T., N. Thomson, A. Bleasby, M. Berriman and J. Parkhill (2009) DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics (Oxford, England)* **25**(1): 119-120.
Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell and J. Parkhill (2005) ACT: the Artemis Comparison Tool. *Bioinformatics (Oxford, England)* **21**(16): 3422-3423.
Cash, H. L., C. V. Whitham, C. L. Behrendt and L. V. Hooper (2006) Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**(5790): 1126-1130.
Casjens, S. (2003) Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology* **49**(2): 277-300.
Cerdeno-Tarraga, A. M., S. Patrick, L. C. Crossman, G. Blakely, V. Abratt, N. Lennard, I. Poxton, B. Duerden, B. Harris, M. A. Quail, A. Barron, L. Clark, C. Corton, J. Doggett, M. T. Holden, N. Larke, A. Line, A. Lord, H. Norbertczak, D. Ormond, C. Price, E. Rabbinowitsch,

J. Woodward, B. Barrell and J. Parkhill (2005) Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**(5714): 1463-1465.
Cerdeño-Tárraga, A. M., S. Patrick, L. C. Crossman, G. Blakely, V. Abratt, N. Lennard, I. Poxton, B. Duerden, B. Harris, M. A. Quail, A. Barron, L. Clark, C. Corton, J. Doggett, M. T. G. Holden, N. Larke, A. Line, A. Lord, H. Norbertczak, D. Ormond, C. Price, E. Rabbinowitsch, J. Woodward, B. Barrell and J. Parkhill (2005) Extensive DNA Inversions in the *B. fragilis* Genome Control Variable Gene Expression. *Science* **307**(5714): 1463-1465.
Chalfie, M., Y. Tu, G. Euskirchen, W. Ward and D. Prasher (1994) Green fluorescent protein as a marker for gene expression. *Science* **263**(5148): 802-805.
Chatzidaki-Livanis, M., M. J. Coyne and L. E. Comstock (2009) A Family of Transcriptional Antitermination Factors Necessary for Synthesis of the Capsular Polysaccharides of *Bacteroides fragilis*. *Journal of Bacteriology* **191**(23): 7288-7295.
Chaudhuri, R. R., N. J. Loman, L. A. Snyder, C. M. Bailey, D. J. Stekel and M. J. Pallen (2008) xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Res* **36**(Database issue): D543-546.
Chen, I. and D. Dubnau (2004) DNA uptake during bacterial transformation. *Nat Rev Micro* **2**(3): 241-249.
Chen, J. and Z. J. Chen (2013) Regulation of NF-κB by ubiquitination. *Current Opinion in Immunology* **25**(1): 4-12.
Cheng, C. W., H. S. Lin, J. J. Ye, C. C. Yang, P. C. Chiang, T. S. Wu and M. H. Lee (2009) Clinical significance of and outcomes for *Bacteroides fragilis* bacteremia. *J Microbiol Immunol Infect* **42**(3): 243-250.
Comstock, L. E. (2009) Importance of glycans to the host-*Bacteroides* mutualism in the mammalian intestine. *Cell Host Microbe* **5**(6): 522-526.
Copeland, N. G., N. A. Jenkins and D. L. Court (2001) Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**(10): 769-779.
Court, D. L., J. A. Sawitzke and L. C. Thomason (2002) Genetic engineering using homologous recombination. *Annual Review Of Genetics* **36**: 361-388.
Coward, C., A. J. Grant, C. Swift, J. Philp, R. Towler, M. Heydarian, J. A. Frost and D. J. Maskell (2006) Phase-Variable Surface Structures Are Required for Infection of *Campylobacter jejuni* by Bacteriophages. *Applied and Environmental Microbiology* **72**(7): 4638-4647.
Coyne, M. J. and L. E. Comstock (2008) Niche-Specific Features of the Intestinal *Bacteroidales*. *Journal of Bacteriology* **190**(2): 736-742.
Cubitt, A. B., R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross and R. Y. Tsien (1995) Understanding, improving and using green fluorescent proteins. *Trends in biochemical sciences* **20**(11): 448-455.
Daniels, C., C. Vindurampulle and R. Morona (1998) Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy). *Molecular Microbiology* **28**(6): 1211-1222.
Darling, A. C. E., B. Mau, F. R. Blattner and N. T. Perna (2004) Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research* **14**(7): 1394-1403.
Darling, A. E., B. Mau and N. T. Perna (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS One* **5**(6): e11147.
Datta, S., N. Costantino, X. Zhou and D. L. Court (2008) Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proceedings of the National Academy of Sciences* **105**(5): 1626-1631.
Davies, J. and D. Davies (2010) Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews* **74**(3): 417-433.

Delcher, A. L., K. A. Bratke, E. C. Powers and S. L. Salzberg (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**(6): 673-679.
Dikic, I., S. Wakatsuki and K. J. Walters (2009) Ubiquitin-binding domains - from structures to functions. *Nat Rev Mol Cell Biol* **10**(10): 659-671.
Donath, M. Y. and S. E. Shoelson (2011) Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol* **11**(2): 98-107.
Dorward, D. W., C. F. Garon and R. C. Judd (1989) Export and intercellular transfer of DNA via membrane blebs of *Neisseria gonorrhoeae*. *Journal of Bacteriology* **171**(5): 2499-2505.
Duncan, L., M. Yoshioka, F. Chandad and D. Grenier (2004) Loss of lipopolysaccharide receptor CD14 from the surface of human macrophage-like cells mediated by *Porphyromonas gingivalis* outer membrane vesicles. *Microbial Pathogenesis* **36**(6): 319-325.
Dutta, S., K. Iida, A. Takade, Y. Meno, G. B. Nair and S. Yoshida (2004) Release of Shiga toxin by membrane vesicles in *Shigella dysenteriae* serotype 1 strains and in vitro effects of antimicrobials on toxin production and release. *Microbiol Immunol* **48**(12): 965-969.
Dynek, J. N., T. Goncharov, E. C. Dueber, A. V. Fedorova, A. Izrael-Tomasevic, L. Phu, E. Helgason, W. J. Fairbrother, K. Deshayes, D. S. Kirkpatrick and D. Vucic (2010) c-IAP1 and UbcH5 promote K11-linked polyubiquitination of RIP1 in TNF signalling. *EMBO J* **29**(24): 4198-4209.
Eklund, M. W., F. T. Poysky and S. M. Reed (1972) Bacteriophage and the toxigenicity of *Clostridium botulinum* type D. *Nat New Biol* **235**(53): 16-17.
Elliott, S. J., L. A. Wainwright, T. K. McDaniel, K. G. Jarvis, Y. K. Deng, L. C. Lai, B. P. McNamara, M. S. Donnenberg and J. B. Kaper (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* **28**(1): 1-4.
Ellis, T. N. and M. J. Kuehn (2010) Virulence and Immunomodulatory Roles of Bacterial Outer Membrane Vesicles. *Microbiology and Molecular Biology Reviews* **74**(1): 81-94.
Fabich, A. J., S. A. Jones, F. Z. Chowdhury, A. Cernosek, A. Anderson, D. Smalley, J. W. McHargue, G. A. Hightower, J. T. Smith, S. M. Autieri, M. P. Leatham, J. J. Lins, R. L. Allen, D. C. Laux, P. S. Cohen and T. Conway (2008) Comparison of Carbon Nutrition for Pathogenic and Commensal *Escherichia coli* Strains in the Mouse Intestine. *Infection and Immunity* **76**(3): 1143-1152.
Flint, H. J., E. A. Bayer, M. T. Rincon, R. Lamed and B. A. White (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Micro* **6**(2): 121-131.
Franco, A. A. (2004) The *Bacteroides fragilis* Pathogenicity Island Is Contained in a Putative Novel Conjugative Transposon. *Journal of Bacteriology* **186**(18): 6077-6092.
Franco, A. A., R. K. Cheng, G.-T. Chung, S. Wu, H.-B. Oh and C. L. Sears (1999) Molecular Evolution of the Pathogenicity Island of Enterotoxigenic *Bacteroides fragilis* Strains. *Journal of Bacteriology* **181**(21): 6623-6633.
Franco, A. A., L. M. Mundy, M. Trucksis, S. Wu, J. B. Kaper and C. L. Sears (1997) Cloning and characterization of the *Bacteroides fragilis* metalloprotease toxin gene. *Infection and Immunity* **65**(3): 1007-1013.
Freeman, V. J. (1951) Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *J Bacteriol* **61**(6): 675-688.
Frost, L. S., R. Leplae, A. O. Summers and A. Toussaint (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**(9): 722-732.
Galka, F., S. N. Wai, H. Kusch, S. Engelmann, M. Hecker, B. Schmeck, S. Hippenstiel, B. E. Uhlin and M. Steinert (2008) Proteomic characterization of the whole secretome of *Legionella pneumophila* and functional analysis of outer membrane vesicles. *Infect Immun* **76**(5): 1825-1836.

Gilbert, H. F. (1995) Thiol/disulfide exchange equilibria and disulfide bond stability. *Methods Enzymol* **251**: 8-28.
Goto, Y. and I. I. Ivanov (2013) Intestinal epithelial cells as mediators of the commensal-host immune crosstalk. *Immunol Cell Biol* **91**(3): 204-214.
Haglund, K., P. P. Di Fiore and I. Dikic (2003) Distinct monoubiquitin signals in receptor endocytosis. *Trends Biochem Sci* **28**(11): 598-603.
Hamoen, L. W., G. Venema and O. P. Kuipers (2003) Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* **149**(Pt 1): 9-17.
Hicke, L. and R. Dunn (2003) REGULATION OF MEMBRANE PROTEIN TRANSPORT BY UBIQUITIN AND UBIQUITIN-BINDING PROTEINS. *Annual Review of Cell and Developmental Biology* **19**(1): 141-172.
Hoeller, D., C.-M. Hecker and I. Dikic (2006) Ubiquitin and ubiquitin-like proteins in cancer pathogenesis. *Nat Rev Cancer* **6**(10): 776-788.
Hooper, L. V., D. R. Littman and A. J. Macpherson (2012) Interactions Between the Microbiota and the Immune System. *Science* **336**(6086): 1268-1273.
Hooper, L. V. and A. J. Macpherson (2010) Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat Rev Immunol* **10**(3): 159-169.
Hooper, L. V., T. Midtvedt and J. I. Gordon (2002) How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr* **22**: 283-307.
Hooper, L. V., T. S. Stappenbeck, C. V. Hong and J. I. Gordon (2003) Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol* **4**(3): 269-273.
Horstman, A. L. and M. J. Kuehn (2000) Enterotoxigenic *Escherichia coli* Secretes Active Heat-labile Enterotoxin via Outer Membrane Vesicles. *Journal of Biological Chemistry* **275**(17): 12489-12496.
Houston, S., G. W. Blakely, A. McDowell, L. Martin and S. Patrick (2010) Binding and degradation of fibrinogen by *Bacteroides fragilis* and characterization of a 54 kDa fibrinogen-binding protein. *Microbiology* **156**(Pt 8): 2516-2526.
Hu, D., B. Liu, L. Dijkshoorn, L. Wang and P. R. Reeves (2013) Diversity in the Major Polysaccharide Antigen of *Acinetobacter baumannii* Assessed by DNA Sequencing, and Development of a Molecular Serotyping Scheme. *PLoS One* **8**(7): e70329.
Huang, A., J. Friesen and J. L. Brunton (1987) Characterization of a bacteriophage that carries the genes for production of Shiga-like toxin 1 in *Escherichia coli*. *J Bacteriol* **169**(9): 4308-4312.
Hueck, C. J. (1998) Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. *Microbiology and Molecular Biology Reviews* **62**(2): 379-433.
Hughes, K. A., I. W. Sutherland and M. V. Jones (1998) Biofilm susceptibility to bacteriophage attack: the role of phage-borne polysaccharide depolymerase. *Microbiology* **144**(11): 3039-3047.
Hurley, J. H., S. Lee and G. Prag (2006) Ubiquitin-binding domains. *Biochem J* **399**(3): 361-372.
Inagaki, H., T. Suzuki, K. Nomoto and Y. Yoshikai (1996) Increased susceptibility to primary infection with *Listeria monocytogenes* in germfree mice may be due to lack of accumulation of L-selectin+ CD44+ T cells in sites of inflammation. *Infection and Immunity* **64**(8): 3280-3287.
Islam, S. T. and J. S. Lam (2013) Wzx flippase-mediated membrane translocation of sugar polymer precursors in bacteria. *Environmental Microbiology* **15**(4): 1001-1015.
Jackson, R. W., B. Vinatzer, D. L. Arnold, S. Dorus and J. Murillo (2011) The influence of the accessory genome on bacterial pathogen evolution. *Mob Genet Elements* **1**(1): 55-65.
Jin, L., A. Williamson, S. Banerjee, I. Philipp and M. Rape (2008) Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell* **133**(4): 653-665.

Johnsborg, O., V. Eldholm and L. S. Håvarstein (2007) Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology* **158**(10): 767-778.
Kadurugamuwa, J. L. and T. J. Beveridge (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**(14): 3998-4008.
Kahn, M. E., G. Maul and S. H. Goodgal (1982) Possible mechanism for donor DNA binding and transport in *Haemophilus*. *Proceedings of the National Academy of Sciences* **79**(20): 6370-6374.
Kamada, N., Y.-G. Kim, H. P. Sham, B. A. Vallance, J. L. Puente, E. C. Martens and G. Núñez (2012) Regulated Virulence Controls the Ability of a Pathogen to Compete with the Gut Microbiota. *Science* **336**(6086): 1325-1329.
Kamada, N., S. U. Seo, G. Y. Chen and G. Nunez (2013) Role of the gut microbiota in immunity and inflammatory disease. *Nat Rev Immunol* **13**(5): 321-335.
Karaolis, D. K., S. Somara, D. R. Maneval, Jr., J. A. Johnson and J. B. Kaper (1999) A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**(6734): 375-379.
Kelley, L. A. and M. J. Sternberg (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**(3): 363-371.
Kelly, D., J. I. Campbell, T. P. King, G. Grant, E. A. Jansson, A. G. P. Coutts, S. Pettersson and S. Conway (2004) Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-[gamma] and RelA. *Nat Immunol* **5**(1): 104-112.
Kesty, N. C., K. M. Mason, M. Reedy, S. E. Miller and M. J. Kuehn (2004) Enterotoxigenic *Escherichia coli* vesicles target toxin delivery into mammalian cells. *EMBO J* **23**(23): 4538-4549.
Komander, D. and M. Rape (2012) The ubiquitin code. *Annu Rev Biochem* **81**: 203-229.
Krogh, A., B. Larsson, G. von Heijne and E. L. Sonnhammer (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**(3): 567-580.
Kruger, D. H. and T. A. Bickle (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev* **47**(3): 345-360.
Kubinak, J. L. and J. L. Round (2012) Toll-Like Receptors Promote Mutually Beneficial Commensal-Host Interactions. *PLoS Pathogens* **8**(7): e1002785.
Kuehn, M. J. and N. C. Kesty (2005) Bacterial outer membrane vesicles and the host–pathogen interaction. *Genes & Development* **19**(22): 2645-2655.
Kulathu, Y. and D. Komander (2012) Atypical ubiquitylation — the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat Rev Mol Cell Biol* **13**(8): 508-523.
Kulp, A. and M. J. Kuehn (2010) Biological Functions and Biogenesis of Secreted Bacterial Outer Membrane Vesicles. *Annual Review of Microbiology* **64**(1): 163-184.
Kung, V. L., E. A. Ozer and A. R. Hauser (2010) The accessory genome of *Pseudomonas aeruginosa*. *Microbiol Mol Biol Rev* **74**(4): 621-641.
Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu and S. L. Salzberg (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5**(2): R12.
Labrie, S. J., J. E. Samson and S. Moineau (2010) Bacteriophage resistance mechanisms. *Nat Rev Micro* **8**(5): 317-327.

Lagesen, K., P. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes and D. W. Ussery (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**(9): 3100-3108.
Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2948.
Lee, S. M., G. P. Donaldson, Z. Mikulski, S. Boyajian, K. Ley and S. K. Mazmanian (2013) Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501**(7467): 426-429.
Lee, Y. K. and S. K. Mazmanian (2010) Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* **330**(6012): 1768-1773.
Lindberg, A. A. and C. G. Hellerqvist (1971) Bacteriophage Attachment Sites, Serological Specificity, and Chemical Composition of the Lipopolysaccharides of Semirough and Rough Mutants of *Salmonella* typhimurium. *Journal of Bacteriology* **105**(1): 57-64.
Liu, C. H., S. M. Lee, J. M. VanLare, D. L. Kasper and S. K. Mazmanian (2008) Regulation of surface architecture by symbiotic bacteria mediates host colonization. *Proceedings of the National Academy of Sciences* **105**(10): 3951-3956.
Locker, J. K. and G. Griffiths (1999) An unconventional role for cytoplasmic disulfide bonds in vaccinia virus proteins. *J Cell Biol* **144**(2): 267-279.
Lowe, T. M. and S. R. Eddy (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**(5): 955-964.
Luczak, M., P. Obuch-Woszczatynski, H. Pituch, P. Leszczynski, G. Martirosian, S. Patrick, I. Poxton, R. G. Wintermans, L. Dubreuil and F. Meisel-Mikolajczyk (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**(2): 222-225.
Lukjancenko, O., T. M. Wassenaar and D. W. Ussery (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* **60**(4): 708-720.
Lutton, D. A., S. Patrick, A. D. Crockard, L. D. Stewart, M. J. Larkin, E. Dermott and T. A. McNeill (1991) Flow cytometric analysis of within-strain variation in polysaccharide expression by *Bacteroides fragilis* by use of murine monoclonal antibodies. *J Med Microbiol* **35**(4): 229-237.
MacEachran, D. P., S. Ye, J. M. Bomberger, D. A. Hogan, A. Swiatecka-Urban, B. A. Stanton and G. A. O'Toole (2007) The *Pseudomonas aeruginosa* Secreted Protein PA2934 Decreases Apical Membrane Expression of the Cystic Fibrosis Transmembrane Conductance Regulator. *Infection and Immunity* **75**(8): 3902-3912.
Marinelli, L. J., G. F. Hatfull and M. Piuri (2012) Recombineering: A powerful tool for modification of bacteriophage genomes. *Bacteriophage* **2**(1): 5-14.
Marolda, C. L., J. Vicarioli and M. A. Valvano (2004) Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit. *Microbiology* **150**(12): 4095-4105.
Mazmanian, S. K., C. H. Liu, A. O. Tzianabos and D. L. Kasper (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**(1): 107-118.
Mazmanian, S. K., J. L. Round and D. L. Kasper (2008) A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**(7195): 620-625.
McBroom, A. J. and M. J. Kuehn (2007) Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response. *Molecular Microbiology* **63**(2): 545-558.

McGrath, S., J. F. M. L. Seegers, G. F. Fitzgerald and D. van Sinderen (1999) Molecular Characterization of a Phage-Encoded Resistance System in *Lactococcus lactis*. *Applied and Environmental Microbiology* **65**(5): 1891-1899.
Molineux, I. J. (2001) No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Molecular Microbiology* **40**(1): 1-8.
Moncrief, J. S., A. J. Duncan, R. L. Wright, L. A. Barroso and T. D. Wilkins (1998) Molecular Characterization of the Fragilysin Pathogenicity Islet of Enterotoxigenic *Bacteroides fragilis*. *Infection and Immunity* **66**(4): 1735-1739.
Murphy, K. C. (1991) Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of *Escherichia coli* RecBCD enzyme. *Journal of Bacteriology* **173**(18): 5808-5821.
Murray, N. E. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **64**(2): 412-434.
Norman, A., L. H. Hansen and S. J. Sorensen (2009) Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc Lond B Biol Sci* **364**(1527): 2275-2289.
Oeckinghaus, A., M. S. Hayden and S. Ghosh (2011) Crosstalk in NF-kappaB signaling pathways. *Nat Immunol* **12**(8): 695-708.
Ohnishi, M., K. Kurokawa and T. Hayashi (2001) Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends in Microbiology* **9**(10): 481-485.
Okinaka, R. T., K. Cloud, O. Hampton, A. R. Hoffmaster, K. K. Hill, P. Keim, T. M. Koehler, G. Lamke, S. Kumano, J. Mahillon, D. Manter, Y. Martinez, D. Ricke, R. Svensson and P. J. Jackson (1999) Sequence and Organization of pXO1, the Large *Bacillus anthracis* Plasmid Harboring the Anthrax Toxin Genes. *Journal of Bacteriology* **181**(20): 6509-6515.
Osborn, A. M. and D. Boltner (2002) When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* **48**(3): 202-212.
Pallen, M. J. and B. W. Wren (2007) Bacterial pathogenomics. *Nature* **449**(7164): 835-842.
Parker, H., K. Chitcholtan, M. B. Hampton and J. I. Keenan (2010) Uptake of *Helicobacter pylori* Outer Membrane Vesicles by Gastric Epithelial Cells. *Infection and Immunity* **78**(12): 5054-5061.
Parma, D. H., M. Snyder, S. Sobolevski, M. Nawroz, E. Brody and L. Gold (1992) The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes & Development* **6**(3): 497-510.
Patel, J. C., K. Hueffer, T. T. Lam and J. E. Galán (2009) Diversification of a *Salmonella* Virulence Protein Function by Ubiquitin-Dependent Differential Localization. *Cell* **137**(2): 283-294.
Patrick, S. (2002). *Bacteroides*. Molecular Medical Microbiology. M. Sussman. London, Academic Press**:** pp 1921-1948.
Patrick, S., G. W. Blakely, S. Houston, J. Moore, V. R. Abratt, M. Bertalan, A. M. Cerdeno-Tarraga, M. A. Quail, N. Corton, C. Corton, A. Bignell, A. Barron, L. Clark, S. D. Bentley and J. Parkhill (2010) Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiology* **156**(Pt 11): 3255-3269.
Patrick, S. and B. I. Duerden (2006). Gram-negative non-spore forming obligate anaerobes. Principles and Practice of Clinical Bacteriology**:** 541–556.
Patrick, S., S. Houston, Z. Thacker and G. W. Blakely (2009) Mutational analysis of genes implicated in LPS and capsular polysaccharide biosynthesis in the opportunistic pathogen *Bacteroides fragilis*. *Microbiology* **155**(Pt 4): 1039-1049.
Patrick, S., K. L. Jobling, D. O'Connor, Z. Thacker, D. T. Dryden and G. W. Blakely (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**(Pt 11): 3071-3078.
Patrick, S., J. P. McKenna, S. O»Hagan and E. Dermott (1996) A comparison of the haemagglutinating and enzymic activities of *Bacteroides fragilis* whole cells and outer membrane vesicles. *Microbial Pathogenesis* **20**(4): 191-202.
Patrick, S., J. Parkhill, L. J. McCoy, N. Lennard, M. J. Larkin, M. Collins, M. Sczaniecka and G. Blakely (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**(4): 915-924.
Patrick, S., J. H. Reid and A. Coffey (1986) Capsulation of in vitro and in vivo Grown *Bacteroides* Species. *Journal of General Microbiology* **132**(4): 1099-1109.
Peng, J., D. Schwartz, J. E. Elias, C. C. Thoreen, D. Cheng, G. Marsischky, J. Roelofs, D. Finley and S. P. Gygi (2003) A proteomics approach to understanding protein ubiquitination. *Nature Biotechnology* **21**(8): 921-926.
Petersson, J., O. Schreiber, G. C. Hansson, S. J. Gendler, A. Velcich, J. O. Lundberg, S. Roos, L. Holm and M. Phillipson (2011) Importance and regulation of the colonic mucus barrier in a mouse model of colitis. *American Journal of Physiology - Gastrointestinal and Liver Physiology* **300**(2): G327-G333.
Prindiville, T. P., R. A. Sheikh, S. H. Cohen, Y. J. Tang, M. C. Cantrell and J. Silva, Jr. (2000) *Bacteroides fragilis* enterotoxin gene sequences in patients with inflammatory bowel disease. *Emerg Infect Dis* **6**(2): 171-174.
Pumbwe, L., D. W. Wareham, J. Aduse-Opoku, J. S. Brazier and H. M. Wexler (2007) Genetic analysis of mechanisms of multidrug resistance in a clinical isolate of *Bacteroides fragilis*. *Clinical Microbiology and Infection* **13**(2): 183-189.
Qian, S. B., D. E. Ott, U. Schubert, J. R. Bennink and J. W. Yewdell (2002) Fusion proteins with COOH-terminal ubiquitin are stable and maintain dual functionality in vivo. *J Biol Chem* **277**(41): 38818-38826.
Quan, S., A. Hiniker, J.-F. Collet and J. A. Bardwell (2013). Isolation of Bacteria Envelope Proteins. Bacterial Cell Surfaces. A. H. Delcour, Humana Press. **966:** 359-366.
Rankin, D. J., E. P. C. Rocha and S. P. Brown (2011) What traits are carried on mobile genetic elements, and why[quest]. *Heredity* **106**(1): 1-10.
Reading, N. C. and D. L. Kasper (2011) The starting lineup: key microbial players in intestinal immunity and homeostasis. *Front Microbiol* **2**: 148.
Reid, J. H. and S. Patrick (1984) Phagocytic and serum killing of capsulate and non-capsulate *Bacteroides fragilis*. *Journal of Medical Microbiology* **17**(3): 247-258.
Renelli, M., V. Matias, R. Y. Lo and T. J. Beveridge (2004) DNA-containing membrane vesicles of *Pseudomonas aeruginosa* PAO1 and their genetic transformation potential. *Microbiology* **150**(Pt 7): 2161-2169.
Robertson, K. P., C. J. Smith, A. M. Gough and E. R. Rocha (2006) Characterization of *Bacteroides fragilis* Hemolysins and Regulation and Synergistic Interactions of HlyA and HlyB. *Infection and Immunity* **74**(4): 2304-2316.
Round, J. L., S. M. Lee, J. Li, G. Tran, B. Jabri, T. A. Chatila and S. K. Mazmanian (2011) The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332**(6032): 974-977.
Round, J. L. and S. K. Mazmanian (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**(5): 313-323.
Round, J. L. and S. K. Mazmanian (2010) Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci U S A* **107**(27): 12204-12209.

Rumbo, C., E. Fernandez-Moreira, M. Merino, M. Poza, J. A. Mendez, N. C. Soares, A. Mosquera, F. Chaves and G. Bou (2011) Horizontal transfer of the OXA-24 carbapenemase gene via outer membrane vesicles: a new mechanism of dissemination of carbapenem resistance genes in *Acinetobacter baumannii*. *Antimicrob Agents Chemother* **55**(7): 3084-3090.
Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream and B. Barrell (2000) Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England)* **16**(10): 944-945.
Rytkönen, A. and D. W. Holden (2007) Bacterial Interference of Ubiquitination and Deubiquitination. *Cell Host & Microbe* **1**(1): 13-22.
Sadowski, M., R. Suryadinata, A. R. Tan, S. N. Roesley and B. Sarcevic (2012) Protein monoubiquitination and polyubiquitination generate structural diversity to control distinct biological processes. *IUBMB Life* **64**(2): 136-142.
Samuel, G. and P. Reeves (2003) Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate Research* **338**(23): 2503-2519.
Sanada, T., M. Kim, H. Mimuro, M. Suzuki, M. Ogawa, A. Oyama, H. Ashida, T. Kobayashi, T. Koyama, S. Nagai, Y. Shibata, J. Gohda, J.-i. Inoue, T. Mizushima and C. Sasakawa (2012) The *Shigella flexneri* effector OspI deamidates UBC13 to dampen the inflammatory response. *Nature* **483**(7391): 623-626.
Schaar, V., S. P. W. de Vries, M. L. A. Perez Vidakovics, H. J. Bootsma, L. Larsson, P. W. M. Hermans, A. Bjartell, M. Mörgelin and K. Riesbeck (2011) Multicomponent *Moraxella catarrhalis* outer membrane vesicles induce an inflammatory response and are internalized by human epithelial cells. *Cellular Microbiology* **13**(3): 432-449.
Scheppach, W. (1994) Effects of short chain fatty acids on gut morphology and function. *Gut* **35**(1 Suppl): S35-38.
Schmukle, A. C. and H. Walczak (2012) No one can whistle a symphony alone – how different ubiquitin linkages cooperate to orchestrate NF-κB activity. *Journal of Cell Science* **125**(3): 549-559.
Scholl, D., S. Adhya and C. Merril (2005) *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Appl Environ Microbiol* **71**(8): 4872-4874.
Schwartz, A. L. and A. Ciechanover (2009) Targeting Proteins for Destruction by the Ubiquitin System: Implications for Human Pathobiology. *Annual Review of Pharmacology and Toxicology* **49**(1): 73-96.
Scott, K. P., D. K. Mercer, L. A. Glover and H. J. Flint (1998) The green fluorescent protein as a visible marker for lactic acid bacteria in complex ecosystems. *FEMS Microbiology Ecology* **26**(3): 219-230.
Sekirov, I., S. L. Russell, L. C. M. Antunes and B. B. Finlay (2010) Gut Microbiota in Health and Disease. *Physiological Reviews* **90**(3): 859-904.
Shao, Y. and I. N. Wang (2008) Bacteriophage adsorption rate and optimal lysis time. *Genetics* **180**(1): 471-482.
Sharan, S. K., L. C. Thomason, S. G. Kuznetsov and D. L. Court (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protocols* **4**(2): 206-223.
Shen, Y., Maria Letizia G. Torchia, Gregory W. Lawson, Christopher L. Karp, Jonathan D. Ashwell and Sarkis K. Mazmanian (2012) Outer Membrane Vesicles of a Human Commensal Mediate Immune Regulation and Disease Protection. *Cell Host & Microbe* **12**(4): 509-520.
Sherwood, J. E., S. Fraser, D. M. Citron, H. Wexler, G. Blakely, K. Jobling and S. Patrick (2011) Multi-drug resistant *Bacteroides fragilis* recovered from blood and severe leg

wounds caused by an improvised explosive device (IED) in Afghanistan. *Anaerobe* **17**(4): 152-155.
Shouldice, S. R., S.-H. Cho, D. Boyd, B. Heras, M. Eser, J. Beckwith, P. Riggs, J. L. Martin and M. Berkmen (2010) In vivo oxidative protein folding can be facilitated by oxidation–reduction cycling. *Molecular Microbiology* **75**(1): 13-28.
Sloper-Mould, K. E., J. C. Jemc, C. M. Pickart and L. Hicke (2001) Distinct functional surface regions on ubiquitin. *J Biol Chem* **276**(32): 30483-30489.
Sóki, J., E. Fodor, D. W. Hecht, R. Edwards, V. O. Rotimi, I. Kerekes, E. Urbán and E. Nagy (2004) Molecular characterization of imipenem-resistant, cfiA-positive *Bacteroides fragilis* isolates from the USA, Hungary and Kuwait. *Journal of Medical Microbiology* **53**(5): 413-419.
Sóki, J., D. W. Wareham, C. Rátkai, J. Aduse-Opoku, E. Urbán and E. Nagy (2010) Prevalence, nucleotide sequence and expression studies of two proteins of a Class III, *Bacteroides* plasmid frequently found in clinical isolates from European countries. *Plasmid* **63**(2): 86-97.
Sokol, H., B. Pigneur, L. Watterlot, O. Lakhdari, L. G. Bermúdez-Humarán, J.-J. Gratadoux, S. Blugeon, C. Bridonneau, J.-P. Furet, G. Corthier, C. Grangette, N. Vasquez, P. Pochart, G. Trugnan, G. Thomas, H. M. Blottière, J. Doré, P. Marteau, P. Seksik and P. Langella (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences* **105**(43): 16731-16736.
Solomon, J. M. and A. D. Grossman (1996) Who's competent and when: regulation of natural genetic competence in bacteria. *Trends Genet* **12**(4): 150-155.
Sommer, F. and F. Backhed (2013) The gut microbiota--masters of host development and physiology. *Nat Rev Microbiol* **11**(4): 227-238.
Sonnenburg, J. L., J. Xu, D. D. Leip, C.-H. Chen, B. P. Westover, J. Weatherford, J. D. Buhler and J. I. Gordon (2005) Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont. *Science* **307**(5717): 1955-1959.
Sonnhammer, E. L., G. von Heijne and A. Krogh (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175-182.
Sørensen, M. C. H., L. B. van Alphen, A. Harboe, J. Li, B. B. Christensen, C. M. Szymanski and L. Brøndsted (2011) Bacteriophage F336 Recognizes the Capsular Phosphoramidate Modification of *Campylobacter jejuni* NCTC11168. *Journal of Bacteriology* **193**(23): 6742-6749.
Sorensen, S. J., M. Bailey, L. H. Hansen, N. Kroer and S. Wuertz (2005) Studying plasmid horizontal transfer in situ: a critical review. *Nat Rev Micro* **3**(9): 700-710.
Sprinz, H., D. W. Kundel, G. J. Dammin, R. E. Horowitz, H. Schneider and S. B. Formal (1961) The response of the germfree guinea pig to oral bacterial challenge with *Escherichia coli* and *Shigella flexneri*. *Am J Pathol* **39**: 681-695.
Stern, A. and R. Sorek (2011) The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33**(1): 43-51.
Sutherland, I. W., K. A. Hughes, L. C. Skillman and K. Tait (2004) The interaction of phage and biofilms. *FEMS Microbiology Letters* **232**(1): 1-6.
Swiatecka-Urban, A., S. Moreau-Marquis, D. P. MacEachran, J. P. Connolly, C. R. Stanton, J. R. Su, R. Barnaby, G. A. O'Toole and B. A. Stanton (2006) *Pseudomonas aeruginosa* inhibits endocytic recycling of CFTR in polarized human airway epithelial cells. *American Journal of Physiology - Cell Physiology* **290**(3): C862-C872.

Swidsinski, A., V. Loening-Baucke, H. Lochs and L. P. Hale (2005) Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice. *World J Gastroenterol* **11**(8): 1131-1140.
Swingle, B., Z. Bao, E. Markel, A. Chambers and S. Cartinhour (2010) Recombineering Using RecTE from *Pseudomonas syringae*. *Applied and Environmental Microbiology* **76**(15): 4960-4968.
Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. DeBoy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. B. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli and C. M. Fraser (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America* **102**(39): 13950-13955.
Tettelin, H., D. Riley, C. Cattuto and D. Medini (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology* **11**(5): 472-477.
Thomas, C. M. and K. M. Nielsen (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**(9): 711-721.
Tigno-Aranjuez, J. T. and D. W. Abbott (2012) Ubiquitination and phosphorylation in the regulation of NOD2 signaling and NOD2-mediated disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**(11): 2022-2028.
Tock, M. R. and D. T. F. Dryden (2005) The biology of restriction and anti-restriction. *Current Opinion in Microbiology* **8**(4): 466-472.
Troy, E. B. and D. L. Kasper (2010) Beneficial effects of *Bacteroides fragilis* polysaccharides on the immune system. *Front Biosci (Landmark Ed)* **15**: 25-34.
Tzianabos, A., A. Onderdonk, B. Rosner, R. Cisneros and D. Kasper (1993) Structural features of polysaccharides that induce intra-abdominal abscesses. *Science* **262**(5132): 416-419.
van Kessel, J. C. and G. F. Hatfull (2007) Recombineering in *Mycobacterium tuberculosis*. *Nat Meth* **4**(2): 147-152.
van Kessel, J. C., L. J. Marinelli and G. F. Hatfull (2008) Recombineering mycobacteria and their phages. *Nat Rev Micro* **6**(11): 851-857.
Van Tassell, R. L. and T. D. Wilkins (1978) Isolation of auxotrophs of *Bacteroides fragilis*. *Can J Microbiol* **24**(12): 1619-1621.
Veiga, E. and P. Cossart (2005) *Listeria* hijacks the clathrin-dependent endocytic machinery to invade mammalian cells. *Nat Cell Biol* **7**(9): 894-900.
Vincze, T., J. Posfai and R. J. Roberts (2003) NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research* **31**(13): 3688-3691.
Wagner, P. L., J. Livny, M. N. Neely, D. W. Acheson, D. I. Friedman and M. K. Waldor (2002) Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli*. *Mol Microbiol* **44**(4): 957-970.
Wagner, P. L. and M. K. Waldor (2002) Bacteriophage Control of Bacterial Virulence. *Infection and Immunity* **70**(8): 3985-3993.
Wang, J. and M. A. Maldonado (2006) The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases. *Cell Mol Immunol* **3**(4): 255-261.
Wang, X., R. A. Herr, W.-J. Chua, L. Lybarger, E. J. H. J. Wiertz and T. H. Hansen (2007) Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-I by viral E3 ligase mK3. *The Journal of Cell Biology* **177**(4): 613-624.

Wareham, D. W., M. Wilks, D. Ahmed, J. S. Brazier and M. Millar (2005) Anaerobic sepsis due to multidrug-resistant *Bacteroides fragilis*: microbiological cure and clinical response with linezolid therapy. *Clin Infect Dis* **40**(7): e67-68.
Weeks, C. R. and J. J. Ferretti (1984) The gene for type A streptococcal exotoxin (erythrogenic toxin) is located in bacteriophage T12. *Infect Immun* **46**(2): 531-536.
Weinbauer, M. G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* **28**(2): 127-181.
Wexler, H. M. (2007) *Bacteroides*: the Good, the Bad, and the Nitty-Gritty. *Clinical Microbiology Reviews* **20**(4): 593-621.
Whitfield, C. (2006a) Biosynthesis and Assembly of Capsular Polysaccharides in *Escherichia coli*. *Annual Review of Biochemistry* **75**(1): 39-68.
Whitfield, C. (2006b) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* **75**: 39-68.
Whitfield, C. and A. Paiment (2003) Biosynthesis and assembly of Group 1 capsular polysaccharides in *Escherichia coli* and related extracellular polysaccharides in other bacteria. *Carbohydrate Research* **338**(23): 2491-2502.
Wostmann, B. S., C. Larkin, A. Moriarty and E. Bruckner-Kardoss (1983) Dietary intake, energy metabolism, and excretory losses of adult male germfree Wistar rats. *Lab Anim Sci* **33**(1): 46-50.
Wozniak, R. A. F. and M. K. Waldor (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Micro* **8**(8): 552-563.
Wu, S., K.-C. Lim, J. Huang, R. F. Saidi and C. L. Sears (1998) *Bacteroides fragilis* enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proceedings of the National Academy of Sciences* **95**(25): 14979-14984.
Xavier, R. J. and D. K. Podolsky (2007) Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**(7152): 427-434.
Xu, J. and J. I. Gordon (2003) Honor thy symbionts. *Proceedings of the National Academy of Sciences* **100**(18): 10452-10459.
Xu, J., J. Zhang, X. Lu, W. Liang, L. Zhang and B. Kan (2013) O Antigen Is the Receptor of *Vibrio cholerae* Serogroup O1 El Tor Typing Phage VP4. *Journal of Bacteriology* **195**(4): 798-806.
Yang, J., H. Chen, I. R. Vlahov, J.-X. Cheng and P. S. Low (2006) Evaluation of disulfide reduction during receptor-mediated endocytosis by using FRET imaging. *Proceedings of the National Academy of Sciences* **103**(37): 13872-13877.
Yaron, S., G. L. Kolling, L. Simon and K. R. Matthews (2000) Vesicle-Mediated Transfer of Virulence Genes from *Escherichia coli* O157:H7 to Other Enteric Bacteria. *Applied and Environmental Microbiology* **66**(10): 4414-4420.
Yu, D., H. M. Ellis, E.-C. Lee, N. A. Jenkins, N. G. Copeland and D. L. Court (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **97**(11): 5978-5983.
Zachar, Z. and D. C. Savage (1979) Microbial interference and colonization of the murine gastrointestinal tract by *Listeria monocytogenes*. *Infection and Immunity* **23**(1): 168-174.
Zhang, C., X.-H. Xing and K. Lou (2005) Rapid detection of a gfp-marked *Enterobacter aerogenes* under anaerobic conditions by aerobic fluorescence recovery. *FEMS Microbiology Letters* **249**(2): 211-218.
Zhang, J., W. Li, Q. Zhang, H. Wang, X. Xu, B. Diao, L. Zhang and B. Kan (2009) The Core Oligosaccharide and Thioredoxin of *Vibrio cholerae* Are Necessary for Binding and Propagation of Its Typing Phage VP3. *Journal of Bacteriology* **191**(8): 2622-2629.

Anaerobe 17 (2011) 152–155

Contents lists available at ScienceDirect

Anaerobe

journal homepage: www.elsevier.com/locate/anaerobe

Clinical Microbiology

Multi-drug resistant *Bacteroides fragilis* recovered from blood and severe leg wounds caused by an improvised explosive device (IED) in Afghanistan

Jeffrey E. Sherwood [a,*,1], Susan Fraser [a], Diane M. Citron [b], Hana Wexler [c], Garry Blakely [d], Kelly Jobling [d], Sheila Patrick [e]

[a] Walter Reed Army Medical Center, Department of Infectious Disease, Ward 63, 6900 Georgia Ave, NW, Washington, DC 20307-5001, USA
[b] R.M Alden Research Laboratory, 6133 Bristol Parkway #175 Culver City, CA 90230-6671, USA
[c] West Los Angeles VA Medical Center, Bldg, 304 Room 200, Los Angeles CA 90076, USA
[d] University of Edinburgh, Institute of Cell Biology, Darwin Building, Kings Buildings, Edinburgh EH9 3JR, UK
[e] Queen's University Belfast, Centre for Infection and Immunity, School of Medicine Dentistry and Biomedical Sciences, 97 Lisburn Road, Belfast BT9 7BL, UK

ARTICLE INFO

Article history:
Received 1 November 2010
Received in revised form
30 December 2010
Accepted 2 February 2011
Available online 3 March 2011

Keywords:
Bacteroides fragilis
Multi-drug resistance
nimE
Plasmid
Linezolid

ABSTRACT

This report summarizes the case of a 23 year-old otherwise healthy male that was injured in an improvised explosive device (IED) blast in support of Operation Enduring Freedom (OEF). He sustained bilateral open tibia and fibula fractures in the setting of being exposed to water contaminated with raw sewage. Despite long-term carbapenem therapy, the patient's wounds were repeatedly noted to have purulent drainage during surgical debridement and cultures from these wounds were persistently positive for *Bacteroides fragilis*. Apparent clinical failure persisted despite the addition of metronidazole to his regimen and an eventual trial of tigecycline.

Susceptibility testing of the *B. fragilis* isolate was performed and resistance to penicillin, clindamycin,metronidazole, cefoxitin, meropenem, imipenem, piperacillin/tazobactam, and tigecycline was confirmed. The presence of a *nimE* gene on a potentially transferrable plasmid was also confirmed by plasmid sequencing. The only antibiotics that displayed *in vitro* susceptibility were moxifloxacin and linezolid. These antibiotics were initiated in combination with aggressive irrigation and serial surgical debridement. Conversion to left-sided internal fixation became feasible and his left lower extremity was salvaged without residual evidence of infection. The patient completed an eight week course of combination moxifloxacin and linezolid therapy without adverse event.

This *B. fragilis* isolate displayed simultaneous high-level resistance to multiple antibiotics routinely utilized in anaerobic infections. This was evidenced by clinical failure, *in vitro* susceptibility testing, and demonstration of genes associated with resistance mechanisms. This case warrants review not only due to the rarity of this event but also the potential implications regarding anaerobic infections in traumatic wounds and the success of a novel treatment regimen utilizing combination therapy with moxifloxacin and linezolid.

Published by Elsevier Ltd.

1. Introduction

Bacteroides species are normally commensal in human colonic microbiota; however, infections with these bacteria can be associated with significant morbidity and mortality. The most commonly isolated Gram-negative anaerobic pathogen, *Bacteroides fragilis*, has potent virulence factors and has been associated with a variety of clinical syndromes including intra-abdominal sepsis and necrotizing skin and soft tissue infections. *B. fragilis* also has an underappreciated ability to develop several different mechanisms of resistance to universally utilized antimicrobials due to mobile genetic elements [1]. Susceptibility profiles can vary widely between different geographical locations and medical institutions. In the United States, multi-drug resistance is rare and anaerobic susceptibility testing is usually only pursued for specific surveillance purposes or due to extraordinary clinical circumstances.

In 2004, a national survey of the susceptibility patterns of *B. fragilis* was conducted in the United States. 5225 clinical isolates from 10 different medical centers were reviewed. Resistance to carbapenems and beta-lactam plus beta-lactamase inhibitor

* Corresponding author. Tel.: +1 808 258 1592; fax: +1 202 782 3765.
E-mail addresses: jeffrey.e.sherwood@us.army.mil (J.E. Sherwood), susan.fraser@us.army.mil (S. Fraser), d.m.citron@att.net (D.M. Citron), s.patrick@qub.ac.uk (S. Patrick).
[1] Present/Permanent Address: 3637 Childress Terrace, Burtonsville, MD 20866, USA.

1075-9964/$ – see front matter Published by Elsevier Ltd.
doi:10.1016/j.anaerobe.2011.02.007

combinations was found in less than 1% of the total isolates while only one isolate in this report was resistant to metronidazole. However, rates of resistance to carbapenems and beta-lactam plus beta-lactamase inhibitor combinations have been increasing in more recent reviews [2]. In 2005, Wareham et al. [3] reported the isolation of a *B. fragilis* strain (WI1) that was simultaneously resistant to carbapenems, beta-lactam plus beta-lactamase inhibitor combinations, and metronidazole in a patient who ultimately died from complications related to anaerobic sepsis. Subsequent genetic analysis of this isolate confirmed the presence of a carbapenemase gene (*cfiA*), macrolide-lincosamide-streptograminin resistance associated gene *ermF*, a plasmid-located tetracycline resistance *tetQ*, and increased efflux pump gene expression among several mechanisms of resistance [4]. 12 (6.9%) of 175 clinical *B. fragilis* isolates in another review from the United Kingdom were positive for carbapenemase gene (*cfiA*) indicating that these isolates could potentially convert to high-level resistance when placed under selective pressure [5]. In blood stream *B. fragilis* isolates from National Taiwan University Hospital rates of resistance to ampicillin plus sulbactam, clindamycin, and carbapenems have also been found to be increasing over time [6]. Given the clinical importance of appropriate empiric antimicrobial selection for the treatment of anaerobic infections, these reports of resistance in *B. fragilis* isolates are particularly worrisome.

2. Case report

The patient in this review was a 23 year-old previously healthy United States Army soldier who sustained multiple injuries in an improvised explosive device (IED) blast in Afghanistan during the summer of 2009. He was riding in an armored patrol when the explosion occurred. The vehicle he was in rolled over into a local stream believed to be contaminated by raw sewage. There were known open sewer systems in the area and the patient himself reported visible pollution. He was submerged in water up to his neck for several minutes while awaiting initial medical care. His injuries included a C5/C6 cervical fracture, a right femoral fracture, and bilateral open tibia and fibula fractures. By the time he arrived at Landstuhl Regional Medical Center (LRMC) in Germany two days after this initial injury his lower extremity wounds were noted to have foul odor. He underwent multiple surgeries while at LRMC to include right femoral intramedullary nail placement, external fixation of his bilateral tibia and fibula fractures, a left lower extremity fasciotomy, and several wound irrigation and debridement procedures. Meropenem 1 g intravenously every 8 h was initiated empirically shortly after his admission to LRMC and antibiotic beads containing vancomycin and tobramycin were placed intra-operatively. Wound cultures grew both an extended spectrum beta-lactamase (ESBL) *Escherichia coli* and *B. fragilis*.

After his arrival at Walter Reed Army Medical Center (WRAMC) in the United States approximately one week after his initial injury, the patient's bilateral lower extremity wounds continued to produce purulent drainage despite serial irrigation and debridement and continuation of broad-spectrum antibiotic therapy with meropenem. *B. fragilis* bacteremia was transiently detected on admission to WRAMC although *B. fragilis* was not found on several subsequent blood cultures. Metronidazole 500 mg intravenously every 6 h was added to his antibiotic regimen, but despite this intervention the patient's right lower extremity could not be salvaged and he required amputation below the right knee three weeks into his WRAMC hospital course. In order to salvage his left leg, internal fixation of his left tibia fracture was completed and continuous wound irrigation with Dakin's solution was initiated. The predominant organism from intra-operative wound cultures remained *B. fragilis*. Due to refractory anaerobic wound infection despite nearly four weeks of meropenem therapy, the patient was transitioned to empiric tigecycline 1 gm intravenously every 12 h and anaerobic susceptibility testing was pursued.

2.1. Microbiology

The patient's initial wound cultures from his left lower extremity were positive for an ESBL *E. coli* and *B. fragilis*. He had an isolated blood culture one week later that was also positive for *B. fragilis*, although several follow up blood cultures were negative indicating that this bacteremia was transient. In addition, antibiotic susceptibility testing of his blood stream isolate revealed a phenotypically different strain that was susceptible to carbapenems. However, a total of eight wound cultures from serial irrigation and debridement procedures from bilateral lower extremities were positive for *B. fragilis*. This occurred while the patient was on meropenem and metronidazole therapy.

The susceptibility testing performed by Quest Diagnostics by "E" test and at the R.M. Alden Research Laboratory by the agar dilution method [7] on one of the patient's *B. fragilis* wound isolates from WRAMC is summarized in Table 1.

A retrospective analysis of his very first wound *B. fragilis* isolate by the R.M Alden Research Laboratory revealed identical high-level multi-drug resistance (MDR). The isolate was also analyzed for plasmid content and found to contain two plasmids: a 5.5 kb plasmid identical to pHAG isolated from a multi-drug resistant *B. fragilis* in London UK [4] and the Class III plasmid pBFB35 [8] and also an 8.3 kb plasmid (designated pWAL610) with confirmed similarity to the sequenced fragment of pBF388c, an 8.3 kb plasmid present in a metronidazole resistant *B. fragilis* isolated in Kuwait [9]. The 8.3 kb pWAL610 was fully sequenced and annotated (Fig. 1). The nitroimidazole resistance gene *nimE* is 100% identical to the *nimE* gene in pBF388c. A transposase, 99% identical to tranposase Bf6 of pBF388c, was identified and also located within an IS4 family insertion sequence.

2.2. Clinical course

Based on ongoing clinical evidence of infection and available susceptibility data tigecycline was discontinued after only a few days of therapy in favor of combination therapy with both moxifloxacin and linezolid. Moxifloxacin 400 mg intravenously every day and linezolid 600 mg intravenously twice daily were initiated and transitioned to an oral route of administration when his scheduled surgeries became less frequent. Both agents were utilized simultaneously due to concerns regarding demonstration

Table 1
Minimum inhibitory concentrations (MIC) in μg/ml of 12 antibiotic/antibiotic combinations for a Bacteroides fragilis wound isolate.

Antibiotic	MIC	Interpretation
Penicillin	> 256	Resistant
Ampicillin and sulbactam	> 128/64	Resistant
Piperacillin and tazobactam	> 64/4	Resistant
Clindamycin	> 128	Resistant
Metronidazole	> 64	Resistant
Cefoxitin	> 64	Resistant
Imipenem	> 32	Resistant
Meropenem	> 32	Resistant
Tigecycline	> 16	Resistant
Chloramphenicol	> 32	Resistant
Moxifloxacin	< 0.5	Susceptible
Linezolid	< 2	Susceptible

Susceptibility testing was performed by E-test (which antibiotics) and by agar dilution (which antibiotics).

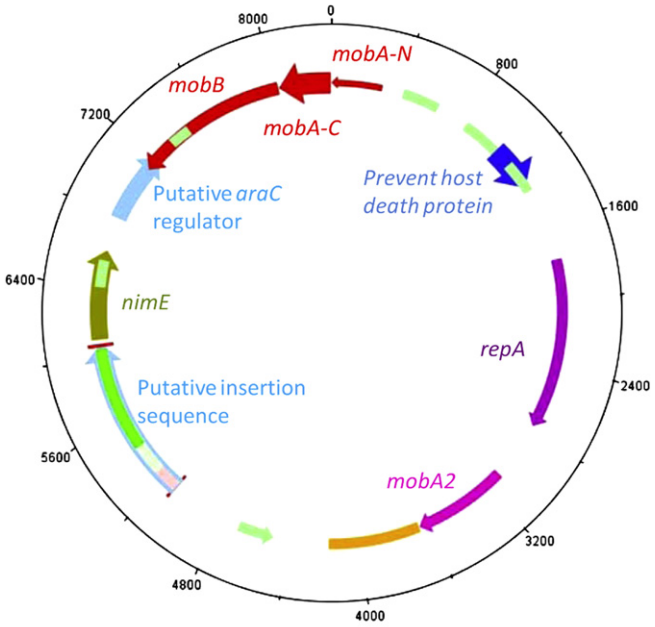


Fig. 1. Artemis DNA plotter image of the 8.3 kb pWAL610 isolated and sequenced from the patient's *B. fragilis* isolate (WAL610). Circular tracks from outside to inside: 1, DNA co-ordinates; 2, forward strand putative coding sequences; 3, reverse strand putative coding sequences; 4, insertion sequence IS*Bf6*; 5, inverted repeats.

of high-level resistance as well as the simple lack of additional therapeutic options. Due to the depth of invasion and the presence of hardware, this antibiotic regimen was continued for an 8 week total duration without adverse effect and with complete resolution of signs and symptoms of infection. The patient's left lower extremity was saved and upon discharge from the hospital he was ambulating with the use of right lower extremity prosthesis.

3. Discussion

B. fragilis infections are generally endogenous, arising from the patient's own normal microbiota [10]. In this case, however, it seems likely that the infection was environmentally acquired which is highly unusual. Although *B. fragilis* is an obligatory anaerobic bacterium that cannot be cultivated in the presence of oxygen, it is aero-tolerant and able to survive for a several hours under aerobic conditions [11]. This case demonstrates how underlying problems regarding the lack of infrastructure in Afghanistan have impacted the medical support mission in Operation Enduring Freedom (OEF). In the urban areas of Afghanistan specifically it is estimated that only 12–23% of the population has access to a safe source of drinking water. With the notable exception of one wastewater treatment facility near the city of Kabul, wastewater collection and treatment are virtually non-existent throughout the country. As a result, open sewer systems are commonplace [12]. The confirmation of identical high-level resistance in the patient's first clinical isolate of *B. fragilis* suggests an environmental source of the MDR isolate as opposed to treatment-induced resistance. It is highly likely that the MDR *B. fragilis* arose from within the local human population and that the water the patient was exposed to was contaminated through this population's utilization of an open sewer system. The presence of a 5.5 plasmid identical in sequence to a plasmid that is widespread across Europe and partial identity to an 8.3 plasmid identified in isolates from Europe and Kuwait is also intriguing.

B. fragilis and taxonomically related *Bacteroides* species are usually resistant to benzylpenicillin, other penicillins and, with the exception of cephamycins such as cefoxitin, many cephalosporins. Classically, resistance to penicillin in *B. fragilis* is associated with the production of active site serine beta-lactamases distantly related to Ambler's Class A. These beta-lactamases are susceptible to the inhibitory effects of compounds such as clavulanic acid, sulbactam and tazobactam. Resistance genes, for example *cepA*, may be plasmid or chromosomally located, and on mobile or mobilisable genetic elements. Similarly, the *cfxA* gene which encodes a cephalosporinase that degrades cephalosporins, penicillins and cefoxitin is located on a mobilisable transposon, *Tn4555*. Some *B. fragilis* strains produce a zinc requiring metallo-beta-lactamase of Ambler's Class B, associated with resistance to cephamycins and carbapenems. These enzymes are resistant to the beta-lactam inhibitors and are associated with the chromosomally encoded gene *cfiA* (*ccrA*). Furthermore, macrolide (e.g. erythromycin), lincosamide (e.g. clindamycin) and streptogramin (e.g. pristinamycin, virginiamycin) resistance is transferrable within and between *Bacteroides* spp. via conjugative plasmid transfer and also chromosomally located self-transmissible conjugative elements [13].A wide range of metronidazole resistance mechanisms have been described (e.g. Patel et al., 2009; Steffens et al., 2010) in addition to the *nim* (A-F) encoding genes [14–16]. Unlike the MDR *B. fragilis* WI1 isolated in London in 2004, the isolate presented in this review possesses a *nim* gene (*nimE)* located on a potentially mobile plasmid. The WI1 isolate demonstrated resistance to metronidazole without expression of a *nim* gene, an ability attributed to this isolate's enhanced efflux pump mechanisms. Similarly, efflux mechanisms and modifications of topoisomerase genes can contribute to fluoroquinolone resistance [4]. Of course, the WAL610 isolate did not demonstrate resistance to fluoroquinolones *in vitro* and a combination of both moxifloxacin and linezolid was the final successful treatment regimen.

Linezolid is traditionally thought of as a therapeutic option for resistant Gram-positive organisms such as methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus*. However, it may also have efficacy in treatment of resistant anaerobic infections. In one review evaluating the minimal inhibition concentrations (MIC's) for linezolid among 265 different anaerobes, the MIC range for the *B. fragilis* group was found to consistently range between 2 and 4 mg/l. The authors discovered no linezolid-resistant strains, suggesting that linezolid may be a potential candidate for the treatment of anaerobic infections [17]. Although most Gram-negative bacteria are resistant to linezolid as a result of endogenous efflux activity, another review examining MIC's for linezolid among *Bacteroides* species reported MIC's between 4 and 8 mg/L [18]. Of note, the WI1 case patient was also treated with linezolid therapy with apparent microbiological cure although the patient did not survive [3].

The reports over the past several years regarding drug-resistant *B. fragilis* isolates in clinical settings is concerning, especially when an isolate such as WAL610 reveals the ability to simultaneously express multiple different mechanisms of resistance. WAL610 closely resembles the MDR WI1 in this regard, although the mechanism of metronidazole resistance appears to be different. Further genetic analysis of the MDR *B. fragilis* WAL610 is underway to determine the mechanisms that generate resistance to the other antibiotics. While MDR *B. fragilis* have not previously been reported in United States, the very existence of isolates such as WAL610 and WI1 raises serious concerns regarding the eventual spread of resistance mechanisms and selection of empiric antibiotic therapy in anaerobic infections. The Clinical Laboratory Standards Institute (CLSI) does not recommend routine susceptibility testing for clinically encountered anaerobes although testing is recommended to

"assist in management of infection in individual patients with serious or life-threatening infections" and "persistence of infection despite adequate treatment with an appropriate therapeutic regimen" [7]. The diagnosis of a multi-drug resistant isolate in this case was delayed in part due to the time it took to arrange adequate anaerobic susceptibility testing through appropriate reference laboratories.

4. Conclusion

MDR *B. fragilis* is a truly rare clinical pathogen in the United States, although *Bacteroides* species certainly exhibit chromosomally mediated and plasmid-encoded resistance mechanisms and readily acquire resistance genes via mobile and mobilisable genetic elements. There are reports of increasing rates of resistance to beta-lactam and beta-lactamase inhibitor combinations and carbapenems throughout the world. While there has been much attention placed on MDR pathogens associated with traumatic war wounds from Iraq and Afghanistan such as *Acinetobacter baumannii* and the ESBL producing enterobactereciae, MDR *B. fragilis* should also be considered in patients with consistent exposure history especially when empiric antibiotics are failing. Furthermore, the oxazolidinone antibiotic linezolid represents a novel and potentially useful therapeutic option for resistant anaerobic infections.

Disclosures

There are no conflicts of interest. The views expressed in this case report are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the U.S. Government.

Acknowledgements

The authors would like to thank Drs. Itzhak Brook of Georgetown University, Ellie J.C. Goldstein of the R.M. Alden Research Laboratory, and Gerald Van Horn of Walter Reed Army Medical Center for their contributions, support, and clinical input regarding this case.

References

[1] Wexler H. *Bacteroides fragilis*: the good, the bad, and the nitty-gritty. Clin Microbiol Rev Oct 2007;20(4):593–621.
[2] Snydman DR, Jacobus NV, McDermott LA, Golan Y, Hecht DW, Goldstein EJC, et al. Lessons learned from the anaerobe survey: historical perspective and review of the most recent data (2005-2007). Clin Infect Dis 2010;50(Suppl. 1):S26–33.
[3] Wareham DW, et al. Anaerobic sepsis due to multidrug-resistant *Bacteroides fragilis*: microbiological cure and clinical response with linezolid therapy. Clin Infect Dis 2005;40:e67–8.
[4] Pumbwe L, Wareham DW, Aduse-Opoku J, et al. Genetic analysis of multidrug resistance in a clinical isolate of *Bacteroides fragilis*. Clin Microbiol Infect 2007;13:183–9.
[5] Edwards R, et al. Prevalence and degree of expression of the carbapenemase gene (cfiA) among clinical isolates of *Bacteroides fragilis* in Nottingham, UK. J Antimicrob Chemother 1999;43:273–6.
[6] Liu C-Y, et al. Increasing Trends in antimicrobial resistance among clinically Important anaerobes and *Bacteroides fragilis* isolates Causing Nosocomial infections: Emerging resistance to carbapenems. Antimicrob Agents Chemother Sept 2008;52(9):3161–8.
[7] Clinical Laboratory Standards Institute. Methods for antimicrobial susceptibility testing of anaerobic bacteria; approved standards. M11-A7. 7th ed. Wayne, PA: Clinical Laboratory Standards Institute; 2007.
[8] Soki J, Wareham DW, Ratkai C, et al. Prevalence, nucleotide sequence and expression studies of two proteins of a 5.6kb, class III, *Bacteroides* plasmid frequently found in clinical isolates from European countries. Plasmid Mar 2010;63(2):86–97. Epub Dec 21 2009.
[9] Soki J, Fodor E, Hecht DW, et al. Molecular characterization of imipenem-resistant, *cfia*- positive Bacteroides fragilis isolates from the USA, Hungary, and Kuwait. J Med Microbiol 2004;53:413–9.
[10] Patrick S, Duerden BL. Gram-negative non-spore forming obligate anaerobes. In: Gillespie SH, Hawkey P, editors. Principles and Practice of clinical Bacteriology. 2nd ed. London: Wiley; 2006. p. 541–56.
[11] Sund CJ, Rocha ER, Tzianabos AO, Wells WG, et al. The *Bacteroides fragilis* transcriptome response to oxygen and H2O2: the role of OxyR and its effect on survival and virulence. Mol Microbiol Jan 2008;67(1):129–42.
[12] United Nations Evironmental Programme. Afghanistan: Post-Conflict environmental Assessment. In: Davis T, editor; 2003. p. 1–180 [Switzerland].
[13] Patrick S. Bacteroides. pp. 1921–1948. In: Sussman M, editor. Molecular medical Microbiology. London: Academic Press; 2002.
[14] Patel EH, Paul LV, Casanueva AI, Patrick S, Abratt VR. Overexpression of the rhamnose catabolism regulatory protein, RhaR: a novel mechanism for metronidazole resistance in Bacteroides thetaiotaomicron. J Antimicrob Chemother 2009;64:267–73.
[15] Steffens LS, Nicholson S, Paul LV, Nord CE, Patrick S, Abratt VR. *Bacteroides fragilis* RecA protein overexpression causes resistance to metronidazole. Res Microbiol 2010;161:346–54.
[16] Lofmark S, Fang H, Hedberg M, Edlund C. Inducible metronidazole resistance and nim genes in clinical *Bacteroides fragilis* group isolates. Antimicrob Agents Chemother 2005;49:1253–6.
[17] Behra-Miellet J, Calvet L, Dubreuil L. Activity of linezolid against anaerobic bacteria. Int J Antimicrob Agents Jul 2003;22(1):28–34.
[18] Livermore DM. Linezolid *in vitro*: mechanism and antimicrobial spectrum. J Antimicrob Chemother 2003;51(Suppl. S2):ii9–16.

Microbiology (2011), **157**, 3071–3078 DOI 10.1099/mic.0.049940-0

A unique homologue of the eukaryotic protein-modifier ubiquitin present in the bacterium *Bacteroides fragilis*, a predominant resident of the human gastrointestinal tract

Sheila Patrick,[1] Kelly L. Jobling,[2] Danny O'Connor,[1] Zubin Thacker,[2] David T. F. Dryden[3] and Garry W. Blakely[2]

Correspondence
Garry W. Blakely
Garry.Blakely@ed.ac.uk

[1]Centre for Infection and Immunity, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Medical Biology Centre, Belfast BT9 7BL, UK

[2]Institute of Cell Biology, University of Edinburgh, Darwin Building, The Kings Buildings, Edinburgh EH9 3JR, UK

[3]EaStChem School of Chemistry, University of Edinburgh, The Kings Buildings, Edinburgh EH9 3JJ, UK

In the complete genome sequences of *Bacteroides fragilis* NCTC9343 and 638R, we have discovered a gene, *ubb,* the product of which has 63 % identity to human ubiquitin and cross-reacts with antibodies raised against bovine ubiquitin. The sequence of *ubb* is closest in identity (76 %) to the ubiquitin gene from a migratory grasshopper entomopoxvirus, suggesting acquisition by inter-kingdom horizontal gene transfer. We have screened clinical isolates of *B. fragilis* from diverse geographical regions and found that *ubb* is present in some, but not all, strains. The gene is transcribed and the mRNA is translated in *B. fragilis*, but deletion of *ubb* did not have a detrimental effect on growth. BfUbb has a predicted signal sequence; both full-length and processed forms were detected in whole-cell extracts, while the processed form was found in concentrated culture supernatants. Purified recombinant BfUbb inhibited *in vitro* ubiquitination and was able to covalently bind the human E1 activating enzyme, suggesting it could act as a suicide substrate *in vivo*. *B. fragilis* is one of the predominant members of the normal human gastrointestinal microbiota with estimates of up to $>10^{11}$ cells per g faeces by culture. These data indicate that the gastro-intestinal tract of some individuals could contain a significant amount of aberrant ubiquitin with the potential to inappropriately activate the host immune system and/or interfere with eukaryotic ubiquitin activity. This discovery could have profound implications in relation to our understanding of human diseases such as inflammatory bowel and autoimmune diseases.

Received 23 March 2011
Revised 16 August 2011
Accepted 22 August 2011

INTRODUCTION

The highly conserved protein-modifier ubiquitin has, to date, only been found in eukaryotes, where tagging of proteins with ubiquitin (ubiquitination) is intrinsic to control of diverse processes central to cell function. The first role identified for ubiquitin was the regulation of intracellular proteolysis via the 26S proteosome, in which ubiquitin becomes covalently bound to a substrate that is subsequently targeted for degradation (Hochstrasser, 2009). The first step of ubiquitination in eukaryotes is the covalent attachment of ubiquitin to the cognate activating enzyme E1, initially by adenylation of the C terminus of ubiquitin followed by formation of a thioester bond with the active site cysteine of E1. The covalently attached ubiquitin is subsequently transferred to the active site cysteine of the ubiquitin-conjugating enzyme E2, followed by formation of an isopeptide bond with the final substrate mediated by the ubiquitin ligase enzyme E3 (Komander, 2009).

It is now recognized that post-translational regulation by ubiquitination also plays an important role in modification of protein function, including cell cycle progression, membrane protein endocytosis, intracellular trafficking, ribosome biogenesis, signal transduction, DNA repair, stress responses, chromatin-mediated regulation of transcription and antigen presentation.

Ubiquitin is also a major factor involved in development and function of both the innate and adaptive immune

Abbreviations: GI, gastrointestinal; OMV, outer membrane vesicle.

systems. Mis-regulation of the ubiquitin pathway is therefore implicated in a wide range of diseases, including cancer, cardiac disease, neurodegenerative disorders and type 2 diabetes (Schwartz & Ciechanover, 2009; Rodríguez *et al.*, 2009; Komatsu & Ichimura, 2010; Vereecke *et al.*, 2009). Immune surveillance for invading pathogens is also controlled by ubiquitin. Extracellular pathogens are recognized by Toll-like receptors that activate a cascade leading to phosphorylation of IκBα, the inhibitor of NF-κB. Ubiquitination and subsequent proteolytic degradation of IκBα releases NF-κB for entry into the nucleus where it activates transcription of pro-inflammatory associated genes (Karin & Ben-Neriah, 2000). Similarly, regulatory events triggered by nucleotide oligomerization domain-like receptor recognition of intracellular pathogens involve ubiquitination (Shaw *et al.*, 2008). In addition, *Salmonella enterica* internalized by macrophages is recognized by a cytosolic mechanism that stimulates direct polyubiquitination of bacterial surface proteins with subsequent recruitment of the proteosome, followed by destruction of the invading prokaryote and presentation of derived peptides to class I major histocompatibility complex molecules (Perrin *et al.*, 2004).

While the evolutionary origin of eukaryotic ubiquitin is hypothesized to be rooted within a common prokaryotic ancestor, structural homologues of ubiquitin identified in bacteria do not have sequence similarity, apart from two terminal glycine residues; examples include the MoaD and ThiS proteins in *Escherichia coli* involved in thiamine and molybdopterin biosynthesis, respectively. These proteins are not involved in proteolysis but do utilize chemistry similar to ubiquitin conjugation, i.e. formation of a thioester bond between the C-terminal glycine and a cysteine catalytic residue in a complementary activating enzyme (Hochstrasser, 2009).

The primary mechanisms for regulation of protein half-life and translation quality control in bacteria do not involve homologues of eukaryotic ubiquitin. In *E. coli* the N-terminal rule pathway dictates the stability of a polypeptide via an adaptor protein, ClpS, which targets degradation by the ClpAP protease (Erbse *et al.*, 2006). Prematurely terminated translation products are marked at the C terminus with an 11 amino acid SsrA tag that directs the polypeptide for degradation by either ClpXP or ClpAP (Gottesman *et al.*, 1998). The analogous N-terminal rule pathway in eukaryotes uses the specificity of a class of ubiquitin E3 ligases, the N-recognins, to direct proteolysis by the proteosome (Tasaki *et al.*, 2005). A protein-tagging system that functions in a similar manner to the ubquitin–proteosome pathway has been identified in *Mycobacterium tuberculosis*; however, the Pup protein that becomes covalently attached to proteins destined for digestion does not share sequence homology with eukaryotic ubiquitin (Pearce *et al.*, 2008).

We now report evidence of horizontal gene transfer from a eukaryotic source into a bacterium of the normal human resident microbiota which results in expression of a secreted protein with 63 % identity to human ubiquitin. The Bacteroidetes are predominant members of the gastrointestinal (GI) tract microbiota. By culture, *B. fragilis* represents ~10–15 % of the members of the *Bacteroides* present in faeces, with estimates between 10^{11} and 10^{12} cells g^{-1} (Patrick, 2002). *B. fragilis* is also the most frequently isolated obligately anaerobic Gram-negative bacterium from life-threatening human infections, such as intra-abdominal, vaginal and brain abscesses (Patrick, 2002; Patrick & Duerden, 2006), and is a major cause of anaerobic bacteraemia, with a potential mortality rate of up to ~30 % (Cheng *et al.*, 2009). Such infections generally arise from faecal contamination of normally uncolonized body sites, for example, peritonitis following perforation of the bowel (Patrick & Duerden, 2006). The discovery that ubiquitin-positive *B. fragilis* is present in the GI tract, however, could have profound implications for our understanding of several human diseases in which ubiquitin malfunction is implicated, and indeed also in relation to the development of autoimmune disease.

METHODS

Bacterial strains and growth conditions. The isolates used in this study were *B. fragilis* NCTC9343, NCTC9344 and NCTC10584; *Bacteroides ovatus* Queen's University Belfast culture collection; and clinical isolates obtained from Craigavon Area Hospital Northern Ireland (designated 'LS'), the Royal Victoria Hospital Northern Ireland (designated 'JC'), the Free University of Amsterdam kindly supplied by J. van Doorn (designated 'BE'), the University of Edinburgh Scotland kindly supplied by I. Poxton (designated 'GNAB'); and a rifampicin-resistant mutant of an isolate from Chigaco, USA, kindly supplied by C. J. Smith (638R). Identification was confirmed with the API 20A system or PCR and sequencing. All *Bacteroides* strains were grown in an anaerobic atmosphere containing 10 % CO_2, 10 % H_2 and 80 % N_2 in either supplemented brain heart infusion broth or defined medium (DM) (Van Tassell & Wilkins, 1978).

Molecular and immunological techniques. *ubb* from genomic DNA was amplified by PCR from *Bacteroides* strains using *Pfu* polymerase (NEB) using *ubb*-specific oligonucleotide primers designed using the NCTC9343 DNA sequence. RNA was isolated from cultures grown in DM (samples taken at OD_{600} 0.4, 0.8 and 1.6) using the Qiagen RNeasy kit and quantified using a Hitachi U-2000 spectrophotometer. RT-PCR analysis of RNA samples used the Qiagen Omniscript RT kit with 20 ng RNA template. The absence of DNA contamination in RNA samples was confirmed by omitting the reverse transcription step and performing the PCR amplification. DNA was analysed by electrophoresis through 1 % agarose gels followed by staining with ethidium bromide and visualization by UV light. The *ubb* deletion mutant was generated in NCTC9343 by replacing *ubb* with an erythromycin resistance cassette using our previously described method (Patrick *et al.*, 2009).

The *ubb* gene was cloned as a 6 × His-tagged N-terminal fusion in pTRC99a. Recombinant (r)BfUbb was purified to homogeneity by overexpression in *E. coli* DH5α, followed by affinity purification using Qiagen Ni-NTA agarose resin and gel filtration through a Superose 12 10/300 GL column (GE Healthcare). Polyclonal antiserum was produced by inoculation of a New Zealand white rabbit with rBfUbb conjugated to keyhole limpet haemocyanin (KLH) carrier protein using an Imject Immunogen EDC kit with mcKLH (Thermo

Scientific). The conjugated antigen was suspended in PBS containing QuilA (0.125 mg ml^{-1}) adjuvant (Brenntag). rBfUbb (0.2 mg) was inoculated at approximately 4 week intervals over a 6 month period under UK Government Home Office Personal and Project Licences and with local ethical approval. Rabbit anti-bovine ubiquitin was obtained from Sigma-Aldrich. Whole cells for immunoblots were electrophoresed through 12 % polyacrylamide gels, transferred to a PVDF membrane and incubated with the appropriate primary and secondary antibodies before detection using the ECL+ chemiluminescent system (GE Healthcare). Supernatants from cultures grown in DM were prepared by centrifugation, filtration through a 0.45 μm membrane and concentrated by centrifugation through a membrane with a 100 kDa cut-off (Amicon Ultra; Millipore). Concentrated supernatant samples for immunoblotting were separated by electrophoresis through 16 % SDS Tris-Tricine polyacrylamide gels. Immunoblots were incubated with the appropriate primary antibody followed by anti-rabbit alkaline phosphate antibody conjugate and reacted with Sigma *Fast* BCIP/NBT(Sigma). Bacterial genome sequences were viewed with Artemis and the Artemis Comparison Tool (Carver *et al.*, 2008).

***In vitro* ubiquitination.** *In vitro* ubiquitination reactions were carried out in 25 mM HEPES and 0.5 mM DTT at 37 °C for 180 min and used the fraction II HeLa conjugation kit (Boston Biochemicals) with a biotinylated lysozyme substrate (Boston Biochemicals). Reactions contained 4 mg Fraction II HeLa ml^{-1}, 2.5 mg ubiquitin ml^{-1}, 4 μM ubiquitin aldehyde, 5 μM MG-132 and an ATP regeneration system. Reactions containing 280 μg BfUbb ml^{-1} were pre-incubated at 37 °C for 15 min prior to the addition of ubiquitin. Protein samples were separated by electrophoresis through 12 % polyacrylamide gels followed by transfer to a PVDF membrane and detection of biotinylated lysozyme using an avidin–HRP conjugate (Boston Biochemicals). Binding of BfUbb to human E1 was carried out in reactions containing 390 μg BfUbb ml^{-1} and either 55 μg or 27.5 μg E1 ml^{-1} (Boston Biochemicals) in 50 mM HEPES pH 8.0 at 37 °C for 60 min. Samples were analysed on a non-reducing 12 % polyacrylamide gel followed by transfer to a PVDF membrane and detection with polyclonal anti-eukaryotic ubiquitin antibodies (Chemicon) and the ECL+ chemiluminescent system (GE Healthcare).

RESULTS

A horizontally acquired gene encoding ubiquitin in *B. fragilis*

The annotation of the *B. fragilis* NCTC9343 genome indicates that it encodes a number of polypeptides with similarities to human proteins (Cerdeño-Tárraga *et al.*, 2005). The most striking homology, however, is found with a putative protein encoded by BF3883 that has 63 % identity (48 of 76 amino acids) to human ubiquitin, such as Uba52 or UbcEP2 (Fig. 1). While many of the amino acids important for interactions with the E1 activating enzyme are conserved in *B. fragilis* ubiquitin (BfUbb), it has at least two unusual features. Firstly, it contains a signal sequence and secondly, it does not possess the two terminal glycine residues required for covalent interactions with the E1 and E2 enzymes of the ubiquitin pathway (Fig. 1).

Despite the implication of its name, ubiquitin has to date only been found in eukaryotes. BF3883 is contained within an 11 kb sequence exhibiting a low GC content, with the closest DNA homology to the gene being 76 % identity (151/199 bp; $E=5\times10^{-29}$) to a ubiquitin-encoding sequence, MSV144, from a migratory grasshopper entomopoxvirus (GenBank accession no. AF063866.1). This suggests that BF3883 was acquired by inter-kingdom horizontal gene transfer. Hereafter, we refer to BF3883 as *ubb*.

The genome sequences of two other *B. fragilis* strains, 638R (isolated in the USA) and YCH46 (isolated in Japan), have also been determined (Patrick *et al.*, 2010; Kuwahara *et al.*, 2004). Strain 638R contains a similar low GC region that includes a putative ubiquitin-encoding gene (BF638R3923) which has 100 % sequence identity to *ubb*. In contrast, the YCH46 genome only has three genes from the equivalent low GC region and does not contain *ubb*, suggesting that the ubiquitin gene has been deleted. To determine the prevalence of *ubb* in other *B. fragilis* strains, we screened a number of isolates derived from different source materials and different geographical locations. Analysis by PCR demonstrated that *ubb* was present in clinical isolates from patients in Edinburgh, Belfast and Amsterdam (Fig. 2). Sequence comparison of the *ubb* amplicons with the NCTC9343 sequence showed 100 % conservation within the coding region (data not shown). While there was no obvious correlation with the presence or absence of the ubiquitin gene and the source of the sample, this demonstrates that the *ubb* gene is present in *B. fragilis* strains resident in the GI tract of individuals from diverse populations. Homologues of *ubb* were not found in other *Bacteroides* species, by either PCR or genome sequence analysis.

```
BfUbb MRFIKQVLLTITLCNIMLFALPSTVNAMQVFIKNRYGWTITLEVSPTDTVENVKQKIQDK
      :                            :.:.:.  : :::::: :.::.::: :::::
Ubc52 M--------------------------QIFVKTLTGKTITLEVEPSDTIENVKAKIQDK
                                          10        20        30
               *     *     *        *        *
BfUbb EGFPPDKIRLIYGGKQMEDGRTLADYNVQKDSTILICIRDVDC
      ::.:::. ::::.:::.::::::.:::.::.::. . .:
Ubc52 EGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
           40        50        60        70
```

Fig. 1. A homologue of ubiquitin present in the bacterium *B. fragilis*. Alignment of the predicted polypeptide encoded by the *ubb* gene in *B. fragilis* with human ubiquitin Ubc52. The first 28 residues of BfUbb represent the putative signal sequence. Identical and similar amino acids are indicated, with residues important for interactions with E1 activating enzyme shown in bold and with asterisks. The two C-terminal glycine residues of Ubc52 involved in covalent attachment to the E1-activating enzyme are underlined.

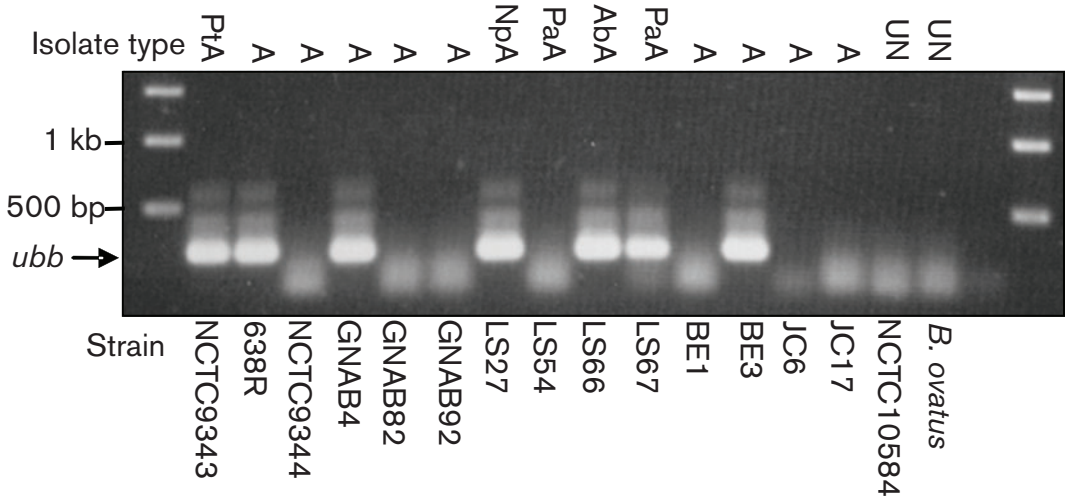


Fig. 2. Agarose gel showing detection of the *ubb* gene by PCR in isolates of *B. fragilis* from different geographical locations. Origins of *B. fragilis* samples are: PtA, peritoneal abscess; A, abscess; NpA, neoplasm abscess; PaA, perianal abscess; AbA, abdominal abscess; UN, unknown origin. The position of the *ubb* amplicon is shown adjacent to the image, with DNA molecular size markers in the first and last lanes. Note the absence of *ubb* in *B. ovatus* and some *B. fragilis* isolates.

To determine whether BfUbb has a central role in the metabolism of *B. fragilis* NCTC9343, as YCH46 which lacks *ubb* has a slower growth rate and attains lower cell density in stationary phase (Patrick *et al.*, 2010), we generated a strain in which *ubb* was deleted. Using our previously described method, *ubb* was replaced with an erythromycin resistance cassette (Patrick *et al.*, 2009). The resulting Δ*ubb*::*ermF* strain did not show any cell morphology defects and had the same growth rates compared with the parental strain in BHI-S medium and glucose defined medium (DM). This indicates that BfUbb is not essential for *B. fragilis* growth in laboratory medium, although it may be advantageous for survival within the host.

Expression of BfUbb in *B. fragilis*

Since it is possible that *ubb* had been acquired from a eukaryotic source, the gene may be an evolutionary relic and may or may not be integrated adjacent to a promoter. We confirmed active transcription of *ubb* in *B. fragilis* by detection of mRNA using RT-PCR (Fig. 3a). Total RNA was isolated from cultures grown in DM during early exponential growth and upon entry into stationary phase. RT-PCR products were detected from RNA templates derived from both growth phases, suggesting that *ubb* is transcribed by the housekeeping RNA polymerase of *B. fragilis*.

We cloned and expressed the predicted processed form of BfUbb as a 6×His-tagged fusion in *E. coli* and used this protein to generate rabbit polyclonal antiserum. Immunoblotting of cell extracts from cultures grown in glucose DM revealed BfUbb as a major protein band of ~9 kDa, plus a second band of ~12 kDa (Fig. 3b). This pattern is consistent with cleavage of the signal peptide having generated the abundant form of BfUbb and the ~12 kDa band, representing the holo-protein prior to export and processing. In addition, some lower-molecular-mass bands, interpreted as degradation products were observed. Interestingly, anti-bovine ubiquitin antibodies cross-reacted with purified recombinant 6×His-BfUbb, but did not detect BfUbb in whole-cell extracts of *B. fragilis*. The presence of a signal sequence and its apparent cleavage is consistent with BfUbb being transported to the periplasm of *B. fragilis*. To interact with the eukaryotic host, however, BfUbb would need to be secreted by the bacterium. *B. fragilis* is known to produce outer membrane vesicles (OMVs) that bleb from the cell surface (Patrick *et al.*, 1996). To determine whether BfUbb was potentially present in OMVs, we concentrated supernatants by filtration through membranes with a 100 kDa cut-off. Immunoblotting of these concentrated supernatants showed the presence of a smeared band of BfUbb that migrated close to the position of the purified 6×His-BfUbb (Fig. 3c; note these samples were separated on a higher percentage gel than that used in Fig. 3b). This band was not detected in concentrated supernatants from the Δ*ubb* strain (data not shown). Smeared bands are often characteristic of glycosylated proteins. Two triplet amino acid sequences (DTV and DST; Fig. 1) corresponding to the consensus recognition site for *B. fragilis O*-glycosylation (Fletcher *et al.*, 2011) are present in BfUbb. The potential glycosylation of BfUbb is being investigated.

In vitro activity of BfUbb

Since *ubb* is not present in all strains of *B. fragilis*, and deletion of the gene in NCTC9343 did not have an

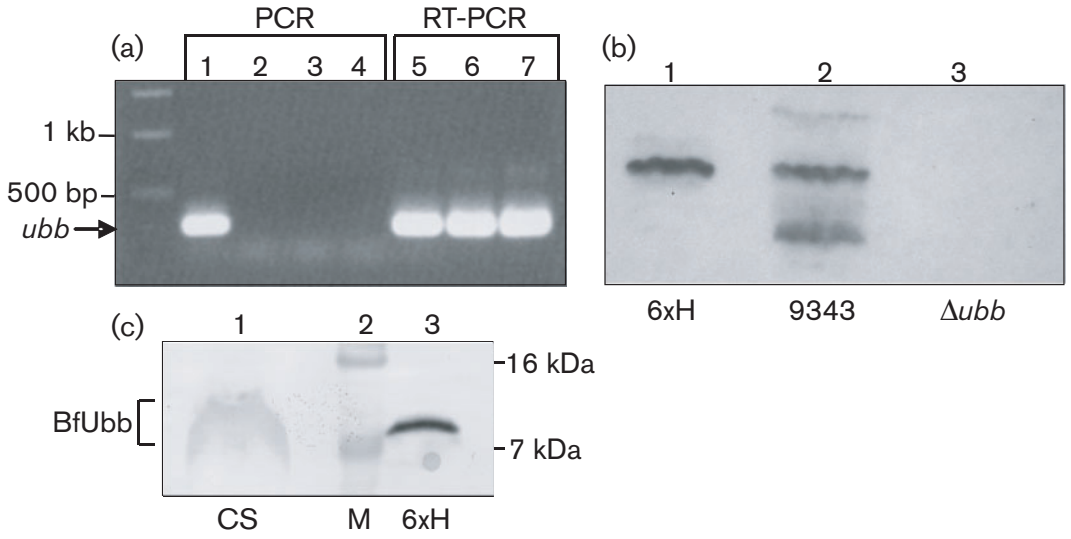


Fig. 3. Detecting expression of BfUbb. (a) Agarose gel showing transcription of *ubb* detected by RT-PCR. Lanes: 1, PCR using genomic DNA; 2–4, PCR using RNA from cultures at OD_{600} 0.4, 0.8 and 1.6 (indicates absence of DNA contamination); 5–7, RT-PCR using RNA from cultures at OD_{600} 0.4, 0.8 and 1.6. The position of *ubb* is indicated with an arrow. (b) Immunoblot using rabbit anti-BfUbb polyclonal antiserum. Lanes: 1, purified 6×His-BfUbb; 2, whole-cell extract of *B. fragilis* NCTC9343; 3, whole-cell extract of *B. fragilis* Δ*ubb*. The lowest band in lane 2 represents a degradation product of BfUbb. (c) Immunoblot of concentrated supernatant from a culture of NCTC9343 grown in DM (lane 1), compared with purified 6×His-BfUbb (lane 3). Molecular mass markers are shown in lane 2.

apparent effect on growth of the bacterium, we hypothesized that BfUbb might be a candidate protein for interaction with the host. To examine the effect of BfUbb on the ubiquitination pathway, we reconstituted the reaction using HeLa cell extracts (Fraction II) and a biotinylated lysozyme substrate. Pre-incubation of purified 6 × His-BfUbb with the HeLa extract, prior to addition of a 10-fold excess of eukaryotic ubiquitin, inhibited subsequent covalent attachment of ubiquitin to the substrate (Fig. 4a). This suggested that 6 × His-BfUbb was binding to the activating and conjugating enzymes of the ubiquitin pathway and inhibiting their catalytic functions. The first step of ubiquitination is the covalent attachment of ubiquitin to the E1 activating enzyme by formation of a thioester bond with the active site cysteine. The ubiquitin homologues encoded by all *B. fragilis* strains tested do not contain a C-terminal glycine but do have a C-terminal cysteine (Fig. 1). We hypothesized that BfUbb might form a disulphide bridge with human E1 that would lead to its inactivation. Under non-reducing conditions a covalent complex could be detected between 6 × His-BfUbb and human E1 by immunoblotting with anti-bovine ubiquitin antibodies (Fig. 4b). This complex was not detected under reducing conditions (data not shown). These data indicated that the two proteins are capable of forming a covalent intermediate in the absence of both ATP and a terminal glycine residue, and suggest that BfUbb could act as a suicide substrate leading to inactivation of E1 and blocking of the ubiquitination cascade.

DISCUSSION

The presence of *ubb* only in selected strains of *B. fragilis*, the similarity of BfUbb to human ubiquitin and its antigenic cross-reactivity, as well as the potential for BfUbb to interfere with the eukaryotic ubiquitination pathway could have profound implications for a range of human diseases. By culturing faeces, *Bacteroides* species numbers range from 10^9 to 10^{14} cells g^{-1}, of which *B. fragilis* estimates range from 4 to 14 % (reviewed by Patrick, 2002). This provides a potentially large reservoir of an aberrant form of ubiquitin in some humans. The resident GI tract microbiota has a well-proven role in the normal development of the immune system. In addition, inflammatory bowel disease is recognized as being caused by a deficient or abnormal mucosal immune response in individuals who are genetically susceptible, clearly related to the presence of intestinal microbes (Marks *et al.*, 2010; Xavier & Podolsky, 2007). In microscopic studies of the spatial organization of the adherent mucosal microbiota, *B. fragilis* was consistently associated with the mucosal surface of patients with inflammatory bowel disease. It was the predominant member of the adherent mucosal biofilm, accounting for >60 % of the biofilm mass (Swidsinski *et al.*, 2005). It will be intriguing to determine whether, given the intimate association of *B. fragilis* with the human GI tract, BfUbb plays a role in the development of these diseases. Furthermore, we postulate that BfUbb could be a driver of other autoimmune diseases, as a result of generating an

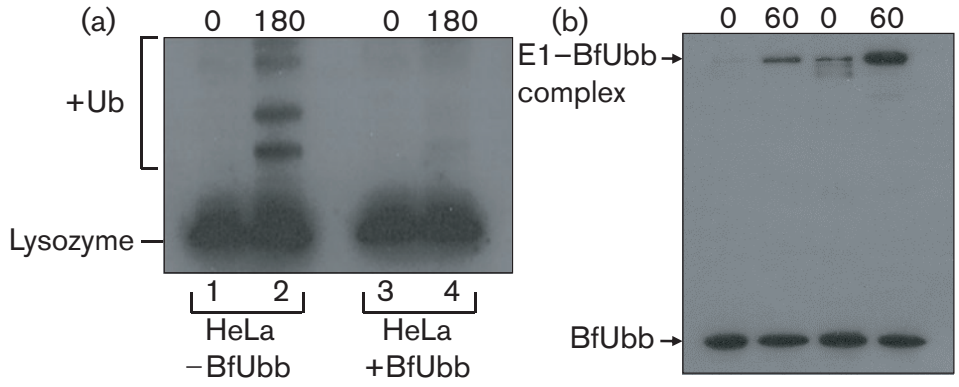


Fig. 4. *In vitro* activity of BfUbb. (a) Immunoblot of *in vitro* ubiquitination of lysozyme using HeLa cell extract in the absence (lanes 1 and 2) or presence (lanes 3 and 4) of BfUbb. Samples were taken at the start of the reaction (0) and at 180 min (180). Lane 2 shows the increase in molecular mass of the lysozyme substrate following covalent attachment of ubiquitin (+Ub). This covalent modification is inhibited by the addition of BfUbb (lane 4). (b) Covalent complexes between human E1 and BfUbb, under non-reducing conditions, were detected by immunoblotting using anti-bovine ubiquitin polyclonal serum. Two reactions using different concentrations of E1 are shown, with time points at 0 and 60 min. Note there was some antibody cross-reactivity with E1 at the higher concentration.

autoimmune reaction to either or both ubiquitin or proteins to which ubiqutin is bound.

The genome of *B. fragilis* does not encode the molecular components required to make a proteosome equivalent to the eukaryotic 26S proteosome; nor does it appear to contain the activating and conjugating enzymes of the ubiquitination pathway. The *ubb* gene has evolved from the eukaryotic equivalent to encode a protein containing a bacterial signal sequence predicted to specify export to the periplasm; also, the C-terminal glycine residues that are required for successful ubiquitination of target proteins in eukaryotes are absent. This suggests that BfUbb has been subverted by the bacterium to fulfil a unique role. So what is the function of BfUbb? The high degree of homology between BfUbb and human ubiquitin, with conservation of residues important for interaction with the E1-activating and E2-conjugating enzymes, implies that these interactions are important. Although we have shown that BfUbb can inhibit ubiquitination *in vitro*, it is possible that the C-terminal cysteine residue could form a thiol-ester bond with other amino acids on selected targets that lead to modified protein function. The isopeptide bond formed between eukaryotic ubiquitin and lysine residues is not the only way that conjugates can be generated; for example, the viral E3 ligase MIR1 can catalyse the formation of thiol-ester bonds between ubiquitin and cysteine residues on major histocompatibility complex I class proteins found on the surface of cytotoxic T lymphocytes (Cadwell & Coscoy, 2005).

The periplasmic localization of BfUbb would provide a means for mobilizing the protein into the environment. *B. fragilis* produces a large number of OMVs (Patrick *et al.*, 1996) that are free to diffuse throughout the gut lumen. The detection of BfUbb in supernatants that had been concentrated through a membrane with a 100 kDa cut-off suggests that the protein is associated with extra-cellular macromolecular complexes. OMVs are 'secretory vehicles' that can deliver bacterial components to host cells and tissues as well as other bacteria (Kuehn & Kesty, 2005; Bomberger *et al.*, 2009). The OMVs of *B. fragilis* agglutinate erythrocytes, via sodium periodate sensitive agglutinins, and are also known to contain degradative enzymes (Patrick *et al.*, 1996). The presence of unusual sphingolipids in the *B. fragilis* outer membrane may allow fusion between OMVs and eukaryotic cytoplasmic membranes that contain sphingolipids. Alternatively, the vesicles may enter epithelial cells by endocytosis. Ubiquitin plays a central role in the regulation of the pro-inflammatory response. Following exposure to a bacterial pathogen, a Toll-like receptor-initiated signalling cascade ultimately results in the phosphorylation of IκB that is subsequently ubiquitinated and degraded by the proteosome. Destruction of IκB releases NF-κB to enable its nuclear localization and initiation of transcription (Bhoj & Chen, 2009). If BfUbb were to enter the cytoplasm of an epithelial cell and act as a suicide substrate for ubiquitination, it could potentially inhibit activation of NF-κB, thus downregulating the inflammatory response.

BfUbb could also aid the success of *B. fragilis* as an opportunistic pathogen. Many pathogenic bacteria have evolved the ability to avoid or suppress the host immune response by interfering with the ubiquitin-proteosome pathway, although none of these mechanisms involve ubiquitin homologues (Boyer & Lemichez, 2004). For example, *Yersinia pestis* uses YopJ to prevent activation of the innate immune response, *Salmonella enterica* serovar Typhimurium produces a ubiquitin ligase and a deubiquitinating protease that are secreted into infected macrophages,

while intracellular *Chlamydia trachomatis* secretes two proteases that have deubiquitinating activities (Zhang *et al.*, 2006; Rytkönen *et al.*, 2007; Le Negrate *et al.*, 2008).

This discovery of a novel ubiquitin homologue in a bacterium, which is a key member of the normal human resident microbiota, could fundamentally alter our understanding of diseases that currently have an unknown aetiology. While we postulate a possible role in inflammatory bowel and autoimmune diseases, mis-regulation of ubiquitin is implicated in a wide range of diseases including cancer, neurodegenerative disorders and type 2 diabetes. The extent to which BfUbb may or may not be involved in any of these diseases remains to be determined, and will require careful consideration of BfUbb-positive *B. fragilis* colonization, the control of BfUbb expression and underlying human genetic predispositions.

ACKNOWLEDGEMENTS

We are grateful to John White and Laurie Cooper for assistance with protein purification. K. L. J. was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) studentship. D. O'C. was in receipt of a Department of Employment and Learning Northern Ireland Studentship. Z. T. was supported by a BBSRC grant (BB-C505875-1) awarded to G. W. B. This work was partly supported by a Wellcome Trust grant (WT090288MA) to D. T. F. D. and G. W. B.

REFERENCES

Bhoj, V. G. & Chen, Z. J. (2009). Ubiquitylation in innate and adaptive immunity. *Nature* **458**, 430–437.

Bomberger, J. M., Maceachran, D. P., Coutermarsh, B. A., Ye, S., O'Toole, G. A. & Stanton, B. A. (2009). Long-distance delivery of bacterial virulence factors by *Pseudomonas aeruginosa* outer membrane vesicles. *PLoS Pathog* **5**, e1000382.

Boyer, L. & Lemichez, E. (2004). Targeting of host-cell ubiquitin and ubiquitin-like pathways by bacterial factors. *Nat Rev Microbiol* **2**, 779–788.

Cadwell, K. & Coscoy, L. (2005). Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. *Science* **309**, 127–130.

Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B. G., Parkhill, J. & 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. & other authors (2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**, 1463–1465.

Cheng, C. W., Lin, H. S., Ye, J. J., Yang, C. C., Chiang, P. C., Wu, T. S. & Lee, M. H. (2009). Clinical significance of and outcomes for *Bacteroides fragilis* bacteremia. *J Microbiol Immunol Infect* **42**, 243–250.

Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D. A. & Bukau, B. (2006). ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. *Nature* **439**, 753–756.

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

Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**, 1338–1347.

Hochstrasser, M. (2009). Origin and function of ubiquitin-like proteins. *Nature* **458**, 422–429.

Karin, M. & Ben-Neriah, Y. (2000). Phosphorylation meets ubiquitination: the control of NF-κB activity. *Annu Rev Immunol* **18**, 621–663.

Komander, D. (2009). The emerging complexity of protein ubiquitination. *Biochem Soc Trans* **37**, 937–953.

Komatsu, M. & Ichimura, Y. (2010). Selective autophagy regulates various cellular functions. *Genes Cells* **15**, 923–933.

Kuehn, M. J. & Kesty, N. C. (2005). Bacterial outer membrane vesicles and the host–pathogen interaction. *Genes Dev* **19**, 2645–2655.

Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., Kuhara, S., Hattori, M., Hayashi, T. & 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.

Le Negrate, G., Krieg, A., Faustin, B., Loeffler, M., Godzik, A., Krajewski, S. & Reed, J. C. (2008). ChlaDub1 of *Chlamydia trachomatis* suppresses NF-κB activation and inhibits IκBα ubiquitination and degradation. *Cell Microbiol* **10**, 1879–1892.

Marks, D. J., Rahman, F. Z., Sewell, G. W. & Segal, A. W. (2010). Crohn's disease: an immune deficiency state. *Clin Rev Allergy Immunol* **38**, 20–31.

Patrick, S. (2002). Bacteroides. In *Molecular Medical Microbiology*, pp. 1921–1948. Edited by M. Sussman. London: Academic Press.

Patrick, S. & Duerden, B. I. (2006). Gram-negative non-spore forming obligate anaerobes. In *Principles and Practice of Clinical Bacteriology*, 2nd edn, pp. 541–556. Edited by S. H. Gillespie & P. Hawkey. London: Wiley.

Patrick, S., McKenna, J. P., O'Hagan, S. & 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., Houston, S., Thacker, Z. & 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. & other authors (2010). Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiology* **156**, 3255–3269.

Pearce, M. J., Mintseris, J., Ferreyra, J., Gygi, S. P. & Darwin, K. H. (2008). Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science* **322**, 1104–1107.

Perrin, A. J., Jiang, X., Birmingham, C. L., So, N. S. & Brumell, J. H. (2004). Recognition of bacteria in the cytosol of Mammalian cells by the ubiquitin system. *Curr Biol* **14**, 806–811.

Rodríguez, J. E., Schisler, J. C., Patterson, C. & Willis, M. S. (2009). Seek and destroy: the ubiquitin–proteasome system in cardiac disease. *Curr Hypertens Rep* **11**, 396–405.

Rytkönen, A., Poh, J., Garmendia, J., Boyle, C., Thompson, A., Liu, M., Freemont, P., Hinton, J. C. & Holden, D. W. (2007). SseL, a *Salmonella* deubiquitinase required for macrophage killing and virulence. *Proc Natl Acad Sci U S A* **104**, 3502–3507.

Schwartz, A. L. & Ciechanover, A. (2009). Targeting proteins for destruction by the ubiquitin system: implications for human pathobiology. *Annu Rev Pharmacol Toxicol* **49**, 73–96.

Shaw, M. H., Reimer, T., Kim, Y.-G. & Nuñez, G. (2008). NOD-like receptors (NLRs): bona fide intracellular microbial sensors. *Curr Opin Immunol* **20**, 377–382.

Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L. P. & Lochs, H. (2005). Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *J Clin Microbiol* **43**, 3380–3389.

Tasaki, T., Mulder, L. C., Iwamatsu, A., Lee, M. J., Davydov, I. V., Varshavsky, A., Muesing, M. & Kwon, Y. T. (2005). A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol* **25**, 7120–7136.

Van Tassell, R. L. & Wilkins, T. D. (1978). Isolation of auxotrophs of *Bacteroides fragilis*. *Can J Microbiol* **24**, 1619–1621.

Vereecke, L., Beyaert, R. & van Loo, G. (2009). The ubiquitin-editing enzyme A20 (TNFAIP3) is a central regulator of immunopathology. *Trends Immunol* **30**, 383–391.

Xavier, R. J. & Podolsky, D. K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**, 427–434.

Zhang, Y., Higashide, W. M., McCormick, B. A., Chen, J. & Zhou, D. (2006). The inflammation-associated *Salmonella* SopA is a HECT-like E3 ubiquitin ligase. *Mol Microbiol* **62**, 786–793.

Edited by: H. Flint