SURFACE LAYER PROTEINS AS VIRULENCE FACTORS IN *CLOSTRIDIUM DIFFICILE* INFECTION

A thesis submitted for the degree of Ph.D.

By

Mark Lynch, B.Sc.

JULY 2014

Based on research carried out at:
School of Biotechnology,
Dublin City University,
Dublin 9,
Ireland.

Under the supervision of

Dr. Christine Loscher and Dr. Mary O'Connell

DECLARATION

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ (Candidate) ID No.: _____ Date: _____

ACKNOWLEDGEMENTS

This PhD has been the most challenging and rewarding experience of my life, and completing it unscathed is not something I could have done without the amazing people around me. There are many people to thank for helping me along the epic journey. My incredible supervisor Christine has provided endless support and advice throughout my project, and has completely reaffirmed my love of all things science. Her ability to communicate complex ideas and develop hypotheses has been truly inspirational. I thank my second incredible supervisor, Mary, for introducing me to the amazing-yet-terrifying world of bioinformatics. Her unwavering enthusiasm and optimism has done so much to push me forward into unknown territories. Together they have set a gold standard for what it is to be a scientist, something I will strive to reach in my own career.

All the incredibly talented people in my lab have been instrumental in helping me hone my skills, and providing much needed fun times during the stressful weeks and months. I could not have asked to work with a better group, whose positive atmosphere and camaraderie made it a pleasure to go to work every morning. I've made lifelong friends, and hopefully future collaborators.

My awesome housemates for helping me forget the stresses of long days in the lab. I've never enjoyed having my room pranked as much as I have here. My friends at home, and abroad, for putting up with my long periods of absence, and for sticking with me through their constant reminders that I'm still a student. The weekends of playing music, writing and recording, even writing comics, made sure I didn't forget the other aspects of my life, long may they continue.

My grandmother, for all her encouragement and optimism. She has been an inspiration my entire life, and one of the hardest working people I know. And my parents, who have been instrumental in shaping the person I am today. They have always supported me in everything I've done. Their confidence in me, and their patience, has been boundless. I thank them for always believing in me, I couldn't have done it without them.

Table of Contents

ABSTRACT

Surface Layer Proteins as Virulence Factors in *Clostridium difficile* infection – Mark Lynch B.Sc.

Clostridium difficile is a nosocomial pathogen possessing Surface Layer Proteins (SLPs), which exhibit adherent and immunostimulatory properties. Using bioinformatics tools, evidence of positive selection (PS) was detected in the SLPs of hypervirulent ribotype (RT) 027, indicating a potential correlation between PS and immunostimulatory properties. The SLPs were purified, and examined for their effect on immune cells. Up-regulation of pro-inflammatory cytokines was observed, however the strength of the response differed between strains. RT 027 induced a strong inflammatory response relative to the RT 001, with which it exhibits sequence similarity. An animal model of *C. difficile* infection compared severity of RT 001 and RT 027. RT 001 induced mild disease with full recovery by day 7, while RT 027 exhibited more severe disease with less clearance. Tissue damage was visualised by histology staining. Further staining showed macrophage and neutrophil infiltration of the tissue. qPCR analysis showed induction of Th17-linked pro-inflammatory cytokines at day 3 in response to RT 001, with levels lowering by day 7. RT 027 infected animals showed later production of IL-10. This may contribute to suppressing clearance. A dendritic cell T cell co-culture was also carried out to fully determine the effect of SLPs on the adaptive immune response. These results strongly suggest the importance of SLPs in disease, and that positive selection may be playing a role in driving the emergence of hypervirulent strains.

ABBREVIATIONS

APC	Antigen Presenting Cell
BSA	bovine serum albumin
BMDC	Bone Marrow-derived Dendritic Cell
CD	Cluster of Differentiation
DC	Dendritic Cell
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
dNTP	Seoxyribonucleotide Triphosphates
DSS	Dextran Sulphate Sodium
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme-linked Immunosorbent Assay
FACS	Fluorescence Activated Cell Sorting
FBS	Foetal Bovine Serum
FCS	Foetal Calf Serum
FITC	Fluorescein Isothiocyanate
FSC	Forward Scatter
IBD	Inflammatory Bowel Disease
IFN	Interferon
IL	Interleukin
IRF3	Interferon Regulatory Factor 3
LPS	Lipopolysaccharide
MCP	Monocyte Chemoattractant Protein
MHC	Major Histocompatibility Complex
MIP	Macrophage Inflammatory Protein
mRNA	Messenger Ribonucleic Acid
MyD88	Myeloid Differentiation primary response gene 88
NF-κB	Nuclear Kactor Kappa-light-chain-enhancer of activated B cells
PAMP	Pathogen-associated Molecular Pattern
PBS	Phosphate Buffered Saline
PRR	Pattern Recognition Receptor
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid

RT	Ribotype
RPMI	Roswell Park Memorial Institute Medium
RT	Room Temperature
SLP	Surface Layer Protein
SSC	Side Scatter
TGF	Transforming Growth Factor
TLR	Toll-like Receptor
TNF	Tumour Necrosis Factor
UC	Ulcerative Colitis

PUBLICATIONS

A role for TLR4 in *Clostridium difficile* infection and the recognition of surface layer proteins (2011)

*Ryan A, **Lynch M**, Smith SM, Amu S, Nel HJ, McCoy CE, Dowling JK, Draper E, O'Reilly V, McCarthy C, O'Brien J, Ní Eidhin D, O'Connell MJ, Keogh B, Morton CO, Rogers TR, Fallon PG, O'Neill LA, Kelleher D, Loscher CE.*

PLoS Pathogens. Jun;7(6):e1002076

Colon cancer associated genes exhibit signatures of positive selection at functionally significant positions (2012)

*Morgan CC, Shakya K, Webb A, Walsh TA, **Lynch M**, Loscher CE, Ruskin HJ, O'Connell MJ.*

BMC Evolutionary Biology. Jul 12;12:114

Surface Layer Proteins Isolated from *Clostridium difficile* Induce Clearance Responses in Macrophages (2014)

*Collins LE, **Lynch M**, Marszalowska I, Kristek M, Rochfort K, O'Connell M, Windle H, Kelleher D, Loscher CE*

Microbes and Infection May;16(5):391-400

Conjugated linoleic acid suppresses dendritic cell activation and subsequent Th17 responses (2014)

*Draper E, DeCourcey J, Higgins SC, Canavan M, McEvoy F, **Lynch M**, Keogh B, Reynolds C, Roche HM, Mills KH, Loscher CE.*

Surface Layer Proteins from virulent *Clostridium difficile* ribotypes exhibit signatures of positive selection with consequences for innate immune response.

***Lynch M**, Walsh TA, Marszalowska I, Webb AE, MacAogain M, Rogers TR, Windle H, Kelleher D, O'Connell MJ and Loscher CE (submitted awaiting review)*

The role of macrophage TNFR2 in the resolution of intestinal inflammation

*Kristek M, DeCourcey J, **Lynch M**, Darby T, Quinlan A, Murphy CT, Casey P, Hill C, Melgar S, Loscher CE (manuscript in preparation)*

Syntaxin 11 negatively regulates IFN-γ secretion in early Th17 differentiation

*DeCourcey J, **Lynch M**, Kristek M, Collins LE, Darby T, Murphy CT, Quinlan A, Melgar S, Sur-Stadt U, Bulfone-Paus S and Loscher CE (manuscript in preparation)*

***Clostridium difficile* ribotype 027 induces more severe infection in vivo compared to ribotype 001**

***Lynch M**, Kristek M, DeCourcey J, Kennedy K, Marszalowska I, Casey P, MacAogain M, Rodgers T, Loscher CE (manuscript in preparation)*

Chapter 1

General Introduction

1.1 *Clostridium difficile*

Clostridium difficile is a spore-forming, anaerobic gram-positive bacteria and the leading cause of antibiotic-associated diarrhoea worldwide (Dawson et al. 2009). Infection usually occurs in hospitalised patients undergoing broad-spectrum antibiotic treatment, where treatment leads to a disruption of the protective microbiota of the gut (Rupnik et al. 2009). Therefore, this makes the intestine more susceptible to colonisation by pathogens (Sekirov et al. 2008; Buffie et al. 2012). In the absence of commensal bacteria, ingested *C. difficile* spores can germinate, colonise the intestinal walls, and begin producing toxins. With the aid of flagella, proteases and other surface proteins, they can penetrate deep into the mucus layer of the intestine (Denève et al. 2009). Depending on the strain causing the infection, symptoms can range from mild diarrhoea to life-threatening pseudomembranous colitis, sepsis and death (Rupnik et al. 2009).

1.1.1 *C. difficile* Toxin Production

C. difficile produces two main toxins; TcdA and TcdB. These cytotoxic proteins are secreted from the bacteria and endocytosed by host cells (Genth et al. 2008). They act by glycosylating small Rho GTPases in the intestinal cells, resulting in disorganisation of the actin cytoskeleton, apoptosis, and destruction of the intestinal epithelium (Kreitman & Akashi 1995). This damage can then allow the bacteria to access the underlying connective tissue, to which it can bind and invade deeper. A negative regulator, TcdC, is encoded in the same region as the two toxins. In total TcdA, TcdB and TcdC form the pathogenicity locus or PaLoc (Spigaglia & Mastrantonio 2002). It has been suggested that inactivation in the TcdC gene results in increased toxin production, giving rise to hypervirulent

strains such as ribotype (RT) 027 (Spigaglia & Mastrantonio 2002). Indeed, an 18bp deletion has been observed in the TcdC gene of 027 strains, and it has been proposed that unregulated toxin production was the cause of its hypervirulence (Spigaglia & Mastrantonio 2002). However, there are also various deletions observed in non-virulent strains (which may not result in loss of function), overall suggesting that increased virulence is not solely due to toxin production (Drudy, Fanning, et al. 2007).

Another hypervirulent strain, ribotype 017, was shown to produce a truncated, non-functional TcdA. Other $TcdA^{-}TcdB^{+}$ strains have been observed, and this toxin variance shows TcdB as the most important toxin in *C. diff* colonisation (Karjalainen et al. 2001). As diagnosis often relied on Toxin A assays to identify *C. diff* in stool samples, this strain may have long gone unnoticed. These toxin variant strains are of particular importance in Ireland. A study conducted in St. James Hospital determined that 81 out of 85 isolates were ribotype 017 (Karjalainen et al. 2001). These strains also show resistance to a wide variety of antibiotics, including fluoroquinolones (Karjalainen et al. 2001). While the toxins of *C. difficile* are obviously of great importance in the study of its pathogenesis, other virulence factors cannot be ignored.

1.1.2 *Clostridium difficile* Spores

Survival of *C. difficile* is dependent on the production of spores, and it is spore germination that is believed to be the first step in establishing CDI (Cohen et al. 2010). These spores are incredibly resilient and can survive many standard hospital cleaning agents, persisting on surfaces for extended periods of time

(Riggs et al. 2007; Setlow 2007). Once these spores are ingested and reach the intestinal tract, they can strongly adhere to intestinal epithelial cells (Paredes-Sabja & Sarker 2012). These spores also appear resistant to most antibiotics (Sarker & Paredes-Sabja 2012; Baines et al. 2009). Therefore, antibiotic treatment, which disrupts the protective microbiota, does nothing to target the spores, whilst eradicating their nutrient competitors. Specific primary bile salts, in the gut, such as taurocholate, promote germination of *C. difficile* spores, while chenodeoxycholate, a secondary bile salt, can inhibit spore germination as well as vegetative cell growth (Giel et al. 2010). The gut microbiota metabolise these primary bile salts into inhibitory secondary salts, providing a further protective function. However, upon antibiotic treatment, the disruption of the microbiota leads to an accumulation of primary bile salts (Hashimoto et al. 1996), promoting spore germination.

Vegetative cells induce a host immune response, with immune cells being recruited to the site of infection. Macrophages have been shown to phagocytose *C. difficile* spores *in vitro*, yet these phagocytosed spores remain dormant, and are not destroyed by the macrophage (Paredes-Sabja et al. 2012). In addition, proteins on the exterior of the spore directly interact with the membrane of the phagosome, and have the ability to induce cell death in macrophages. While vegetative cells are successfully inactivated, the survival and persistence of spores may be a contributing factor in recurrent CDI.

1.1.3 Surface layer protein SlpA

The main focus of study in *C. difficile* pathogenesis has been toxin production for some time, but there is increased interest in cell surface proteins as important virulence factors. Like many bacteria, *C. difficile* possess a surface layer, or S-Layer. S-Layers have many proposed functions, such as adherence, molecular sieves and evasion of the immune system. The S-Layer of *C. difficile* is composed of two main surface layer proteins (SLPs) termed the high molecular weight (HMW) SLP and the low molecular weight (LMW) SLP. These SLPs are the predominant surface antigen in *C. difficile*.

While many bacteria possess surface layers with multiple proteins, *C. difficile* is unique in that a single gene encodes for both major SLPs. The gene is termed *slpA* and codes for a protein precursor, consisting of three main regions, a short signal peptide of 26 amino acids in length, the LMW-coding region and the HMW-coding region. The signal peptide is thought to be responsible for transporting the immature SlpA protein to the cell surface, where it is then cleaved into the mature HMW and LMW SLPs. The structure of *slpA* can be seen in Figure 1.1.

Figure 1.1

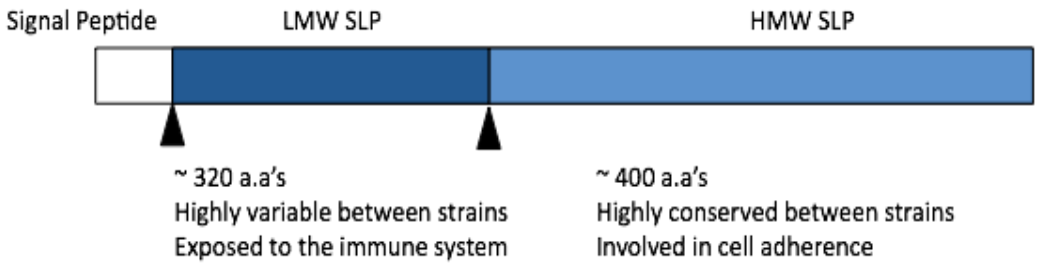


Figure 1.1 Structure of the *slpA* gene.

The signal peptide, which directs the protein to the exterior of the bacteria before being cleaved, is seen in white. The coding regions for the LMW and HMW proteins can be seen in dark and light blue respectively. Cleavage sites are represented by a black triangle.

The HMW SLP is highly conserved across strains, while the LMW SLP shows little sequence conservation (Calabi et al. 2002). The HMW SLP exhibits strong and specific binding to gastrointestinal tissues and human epithelial cells (Calabi et al. 2002). More recently it has been shown to be the major contributor to host cell adherence *in vitro* (Merrigan et al. 2013). Pre-coating host cells with SLP preparations successfully inhibited binding of *C. difficile* to the cells. In addition, adherence was also inhibited when host cells were incubated with anti-SLP serum prior to exposure with *C. difficile* (Merrigan et al. 2013). This study also suggested a prominent role for the LMW subunit in adherence. This is contrary to previously published studies (Calabi et al. 2002) which described the HMW protein as subunit involved in adherence. It seems likely that both SLP subunits play a role in adherence to the gastrointestinal tract.

In contrast, the LMW protein does not bind to purified extracellular matrix (ECM) components or mucosal connective tissues, but can weakly bind to rare gastrointestinal cells (Calabi et al. 2001). Therefore, the LMW protein may have a role other than adherence (Calabi et al. 2001). The binding of the HMW SLP to host epithelial cells may allow for the targeted delivery of the cytotoxins to the epithelial walls. This strongly suggests that when the epithelial layer has been damaged, the SLPs can allow the pathogen to invade deeper by binding to ECM components.

The LMW SLP is the immunodominant antigen. Because of high sequence variability between strains, and the fact that different strains induce different immune responses, it is plausible that this protein may be involved in masking the bacteria from the immune system. A change in structure or specific motifs in the LMW protein may result in host immune cells no longer being able to recognise *C. difficile* as a pathogen. Indeed SLPs have been shown to activate macrophages and dendritic cells *in vitro* (Ausiello et al. 2006).

The structure of these proteins and how they interact with each other has only recently been demonstrated (Fagan et al. 2009). A crystal structure was obtained only for a truncated form of the LMW SLP as C-terminal residues degraded at room temperature. The crystal structure of this LMW SLP revealed a two-domain configuration as seen in Figure 1.2. The larger domain, Domain 2, was composed roughly of 10% α-helices, 30% β-sheets and 60% loops. These loops in Domain 2 contain many hydrogen-bonding interactions. Domain 1 contains both the C- and N-terminal regions of the proteins and consists of 37% α-helices,

36% β-strands and only 27% loops. Domain 1 is also likely responsible for binding to the HMW protein. Deletion of 61 residues in Domain 1 (residues 260-321 of strain 630) resulted in a loss of binding. Despite the high variability in the LMW protein across strains, there are some strikingly conserved regions. For example, the area described above is well conserved, implying these residues are essential for binding to the HMW protein, and varying the sequence may result in a loss of binding. A common mechanism therefore must exist for the assembly of the mature SLP complex. The 24-peptide long signal sequence is also incredibly well conserved (Fagan et al. 2009).

The HMW protein is well conserved, and a loss of the first 40 residues in the HMW sequence resulted in loss of binding the LMW protein (Fagan et al. 2009). The mature SLP complex was found to have an "end-to-end" orientation with a specific binding domain between the HMW and LMW SLPs (Fagan et al. 2009). The orientation of the proteins can be inferred, as the HMW SLP possesses two PF04122 Pfam cell wall binding motifs, which are involved with binding to components of the bacterial cell wall (Fagan et al. 2009). It has been shown that the HMW protein is most likely anchored to the cell wall, and is "displaying" the LMW protein as the external part of the S-Layer. As the LWM protein is the outermost part of the bacteria, it will be exposed to the host immune system and therefore its high sequence variability may be a result of evolutionary selective pressures exerted by the host immune response acting on the bacteria to avoid a host response (Fagan et al. 2009).

Figure 1.2

Figure 1.2 Proposed Structure of the LMW/LMW SLP Complex

3D model of both HMW and LMW SLPs orientated exterior to the cell wall. Taken from Fagan et al. 2009.

The genes encoding *slpA* and all of the *cwp* genes are located on a 36.6-kb *cwp* gene cluster (Sebaihia et al. 2006). Despite *slpA* sequences exhibiting little sequence variability between serogroups (Karjalainen et al. 2002), a recent study has shown a potential role for recombination in *slpA* between different strains (Dingle et al. 2013). Five main clades of *C. difficile*, along with twelve stable variants of a 10-kb S-layer cassette, containing *slpA*, were identified. The twelve S-layer cassette variants unexpectedly behaved as independent components of the genome, associating randomly with the five clades. Dingle *et al* observed the S-layer cassette from hypervirulent RT 027 to be closely related to that of RT 001, and concluded that, because of this, hypervirulence may not be solely

influenced by the S-layer. They also show that isolates sharing the same PCR-ribotype can carry multiple distinct S-layer cassettes (Dingle et al. 2013).

1.1.4 Cell Wall Proteins

There is a family of genes (paralogs) (consisting of 29 genes in total) that code for cell surface proteins. All of these paralogs possess at least two copies of the Pfam 04122 motif, also known as "putative cell wall binding repeat 2". Fagan *et al* have recently suggested that these be designated cell wall protein, or *cwp*, genes (Fagan et al. 2011). The majority of these genes are not well characterised, and the functions of many of them remain somewhat elusive. Three of the somewhat characterised proteins are Cwp84, Cwp66 and CwpV.

Of particular interest is Cwp84, a cysteine protease whose function was initially hypothesised to be the degradation of the host ECM (Janoir et al. 2007). As the protein resides on the surface of the bacteria, it is in an optimal position to react with components of the ECM. It has been shown to cleave fibronectin, laminin and vitronectin, all of which are important components of the intestinal epithelium (Janoir et al. 2007). Unlike the highly variable SlpA, Cwp84 appears greatly conserved across strains, particularly sites in the protein active region. In addition, transcription of the *cwp84* gene is detected only in the early exponential growth phase of the bacteria (Savariau-Lacomme et al. 2003). Another function for Cwp84 has been suggested. Dang *et al* have shown that Cwp84 is the protease that cleaves the immature slpA into its HMW and LMW components (Dang et al. 2010). Inhibition of this protease results in accumulation of immature slpA on the cell wall, which leads to a decrease in viability of the

bacterium. In addition to this, upregulation of *slpA* is seen to occur under inhibition of *cwp84*, implying a greater need for SLPs when the protease is not performing its function (Dang et al. 2010). It is likely that both ECM degradation and SLP maturation are key functions of Cwp84. The high conservation of the SLP cleavage site between strains and low variability of Cwp84 itself imply a conserved process of SLP maturation across all strains of *C. difficile*. It is therefore an extremely important component of the S-layer.

Mutation of the *cwp84* gene was found to result in defective localisation of other Cwp proteins. As Cwp84 is necessary for the maturation of slpA, it is likely that this delocalisation is due to a disruption of the mature S-layer complex (de la Riva et al. 2011). In addition to the role of Cwp84 in slpA maturation, the protease Cwp13 also appears to play a role in the maturation of the S-layer. This protein can cleave and aid the maturation of Cwp84, therefore allowing the formation of the mature S-layer complex. In the absence of Cwp13, both immature and mature forms for Cwp84 were detected, indicating Cwp13 role in S-layer formation. Cwp84 was still present, albeit in lower numbers. Cwp13 also cleaves slpA at a site distinct from Cwp84. This suggests a further role in the degradation of mis-folded proteins (de la Riva et al. 2011). A diagrammatic representation of the maturation process of SlpA can be seen in Figure 1.3.

Figure1.3

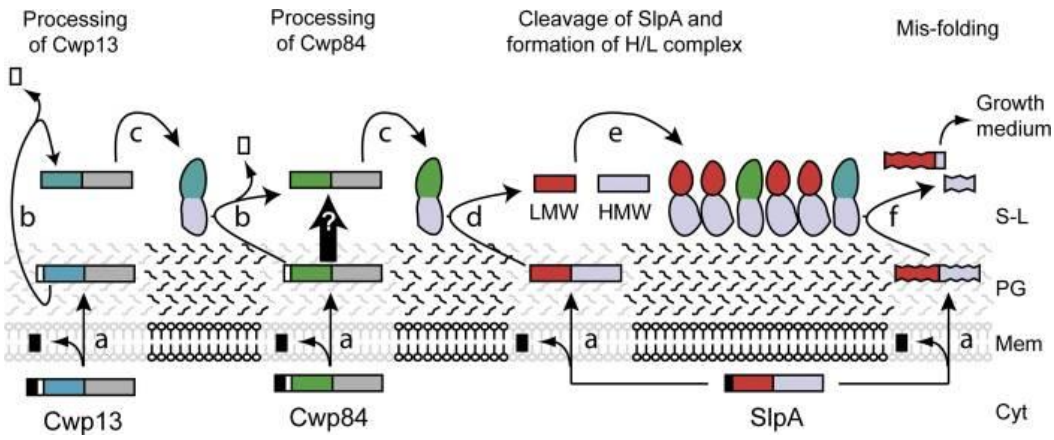


Figure 1.3 Maturation process of SlpA.

Cwp13 translocates to the cell membrane and processes it mature form in an autocatalytic fashion. The mature Cwp13 then cleaves Cwp84, which in turn cleaves SlpA into the mature HMW and LMW subunits. Taken from de la Riva et al. 2011.

Another characterised paralog of slpA is Cwp66. Unlike Cwp84, but similar to the HMW SLP, it seems to function mainly in adhesion. It is transcribed during the exponential phase of growth, and present in much lower levels than Cwp84 and slpA. Therefore this protein may not have the same level of importance as more major S-layer components. Waligora *et al* further demonstrated the role of Cwp66 in adhesion (Waligora et al. 2001). Similar to slpA, it possesses a 27 amino acid signal sequence, most likely responsible for transporting the protein to the cell surface. They have also shown that the protein is evenly distributed across the cell wall and that the C-terminal is surface exposed. Cwp66 shows more variability than Cwp84, exhibiting a highly conserved N-terminal, with a more variable C-terminal (~58% identity). With the N-terminal region of the protein most likely attached to the cell wall, there will be greater pressures on the

C-terminal region as it may be subject to immune selection, similar to the LMW SLP. Antibodies raised against the N-terminal domain of this protein resulted in partial loss of adherence, indicating Cwp66's role but also showing other adhesions being involved (Waligora et al. 2001).

1.2 Immunity in the Gastrointestinal Tract

The small intestine, cecum, large intestine, and rectum, together constitute the gastrointestinal tract. Shortly after birth, the intestine becomes colonised by vast amounts of non-pathogenic, commensal bacteria, with number estimating between 10^{13} – 10^{14} bacteria per gram of luminal content (Ley et al. 2006; Gill et al. 2006). In addition to these commensal bacteria, the presence of food antigens puts pressure on the intestinal immune system. It is of crucial importance that pathogens be recognised and an appropriate response mounted, while at the same time preventing persistent inflammation due to activation of immune cells by non-pathogenic antigens. Immune cells in the intestine are associated with the gut-associated lymphoid tissue (GALT), which comprises of Peyer's patches in the small intestine, and isolated lymphoid follicles in the large intestine, or colon (Jung et al. 2010). These areas of tissue contain high densities of immune cells, which survey antigens in the gut. The underlying tissue is termed the lamina propria, and contains a vast array of different immune cells, such as macrophages, dendritic cells, mast cells and lymphocytes (Doe 1989).

The first barrier pathogens face in the intestine is the intestinal epithelial cells. These cells form a monolayer between the lumen of the intestine and sub-epithelial tissue, location of the immune cells. These cells provide a necessary

physical barrier, but also can interact with both the bacteria present in the lumen of the gut, and the underlying host immune cells. Through direct recognition of bacteria by pattern recognition receptors, epithelial cells and interact with and influence the immune response.

1.2.1 Pattern Recognition Receptors

Pattern recognition receptors (PRRs) on the surface of epithelial and immune cells are essential in detecting common microbial ligands. Many intra- and extracellular PRRs exist, such as Nod-like receptors and Rig-I like receptors, capable of recognising a plethora of common antigens, termed pattern associated molecular patterns (PAMPs) (Goto & Ivanov 2013). Of particular interest are the Toll-like receptors (TLRs), which play an important role in bacterial recognition. There are ten functional TLRs in humans, and thirteen in mouse, each capable of recognising distinct PAMPs (Akira et al. 2001). TLR1, TLR2, TLR4, TLR5 and TLR6 are present on the cell surface. TLR2 can form heterodimers with both TLR1 and TLR6 to recognise bacterial lipopeptides and lipoteichoic acid from gram positive cell walls (Farhat et al. 2008). TLR4 recognises lipopolysaccharide (LPS) on the outer layer of gram negative bacteria (Takeuchi et al. 1999). TLR5 recognises flagellated bacteria (Hayashi et al. 2001). TLR3 and TLR9 recognise PAMPs from intracellular pathogens. TLR3 recognises both single and double stranded viral RNA (Alexopoulou et al. 2001), while TLR9 recognises unmethylated CpG from DNA viruses (Hemmi et al. 2000). Expression of TLR1, TLR2, TLR3, TLR4, TLR5 and TLR9 has been confirmed in epithelial cells of the small intestine (Otte et al. 2004), and have been shown to be involved in

maintaining homeostasis, with TLR4 and TLR2 deficient mice exhibiting increased susceptibility to colitis (Rakoff-Nahoum et al. 2004).

The main TLR of interest in this thesis is TLR4. Despite recognising LPS of gram negative bacteria, TLR4 is the receptor responsible for recognising SLPs and mounting a response to *C. difficile* (Ryan et al. 2011). There are two possible pathways for activation downstream of TLR4, the MyD88-dependent and MyD88-independent pathways. The MyD88-dependent pathway results in downstream activation of NF-κB, proinflammatory cytokine production and activation of T helper cells (Akira & Takeda 2004). Initiation of the MyD88-independent pathway, and subsequent IRF3 activation (Brikos & O'Neill 2008), requires the initial endocytosis of TLR4 in the presence of CD14 (Jiang et al. 2005). This pathway leads to induction of type one interferons (Kagan et al. 2008). In addition to this, activation of both these pathways is required for optimal IL-10 production in immune cells (Boonstra et al. 2006; Chang et al. 2007). While LPS induces both pathways downstream of TLR4, we have previously shown that SLPs from RT 001 only activate the MyD88-dependant pathway (Ryan et al. 2011). The main signalling molecules involved in TLR4 signalling can be seen in Figure 1.4.

Figure 1.4

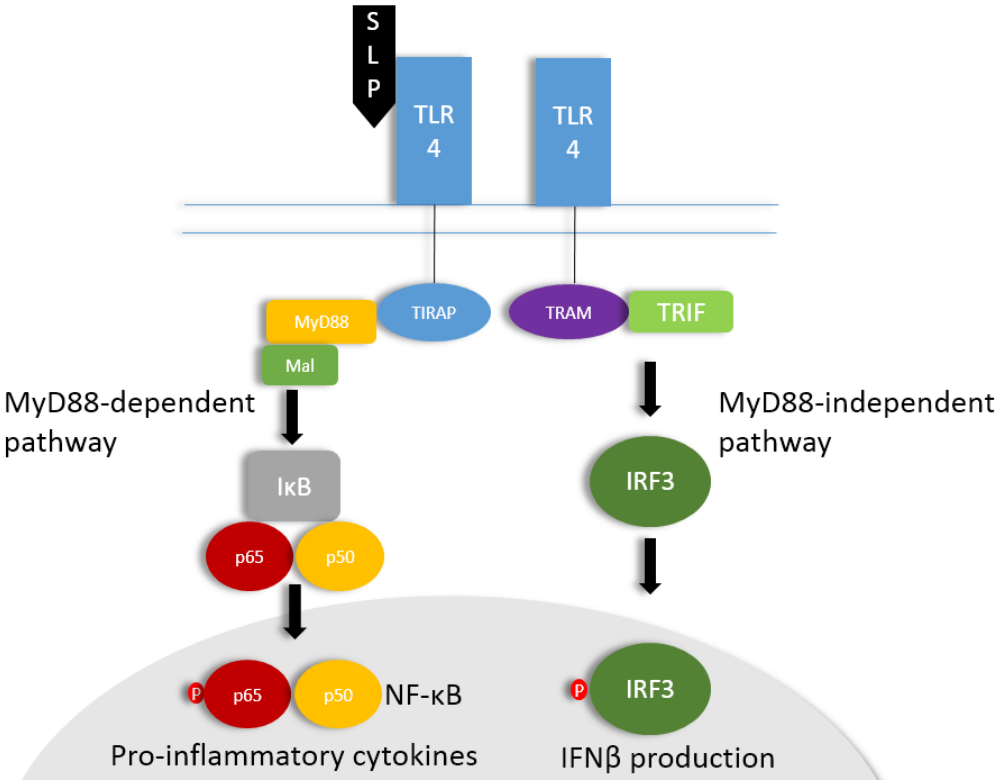


Figure 1.4 Toll-like Receptor 4 Signalling

TLR4 signalling in response to LPS. The MyD88-dependent pathway leads to downstream activation of NF-κB and pro-inflammatory cytokine production, while the MyD88-independent pathway results in IRF3 activation and type 1 interferon production.

1.2.2 Dendritic Cells

Dendritic cells (DCs) are the sentinels of the immune system and play an essential role in intestinal immune regulation. As the intestine possesses many resident inflammatory cells, an essential feature of immune cells in the gastrointestinal tract is to maintain homeostasis in the presence of commensal bacteria, while mounting a protective response against invading pathogens. Dendritic cells and macrophages both play important roles in immunity of the gastrointestinal tract, with populations of both cells residing in the lamina propria behind a wall of protective epithelial cells (Smith et al. 2005; Iwasaki 2007). Resident mucosal DCs sample the environment of the intestinal lumen in different ways. Specialised microfold cells, or M cells, present in the epithelium of Peyer's patches can uptake antigen from the lumen, and transcytose to underlying DCs (Coombes & Powrie 2008). DCs can also sample antigen directly by forming tight junction structures with intestinal epithelial cells and extending dendrites into the lumen (Rescigno et al. 2001). While this process occurs during normal conditions, sampling of resident commensal bacteria, the number of these trans-epithelial dendrites is increased during infection (Niess et al. 2005; Chieppa et al. 2006). This process requires MyD88-dependent signalling through TLRs to efficiently function (Niess et al. 2005).

Upon antigen uptake, intestinal DCs migrate to the mesenteric lymph nodes, where they can interact with naïve T and B cells. Here they can display antigens from commensal bacteria and apoptotic cells, helping drive tolerogenic responses to maintain a non-inflammatory state (Macpherson & Uhr 2004; Huang et al. 2000). In particular, DCs isolated from the mesenteric lymph nodes were found

to be capable of producing high levels of TGF-β, driving naïve T cells to an immunosuppressive Treg phenotype, in the absence of exogenous factors (Coombes et al. 2007).

As intestinal DCs recognise commensal bacteria without inducing an inflammatory response, the mechanism by which they can mount an effective response against pathogens is not yet fully known (Coombes & Powrie 2008). A distinguishing factor may be the ability of pathogens to cross the epithelial barrier and invade deeper into the gut. The presence of pathogenic bacteria in the gut may result in the recruitment of inflammatory cells by chemokine production by epithelial cells, where migrating DC-precursors in the blood may be recruited. These cells will not have the same conditioning as the resident intestinal DCs and therefore will mount an inflammatory response. This has been seen in response to pathogenic *Salmonella* spp., where the bacteria induced production of IL-8 from epithelial cells, attracting neutrophils and increasing the inflammatory state in the gut (Rimoldi et al. 2005). This increased inflammatory state will help recruit non-conditioned DCs to induce a protective response. In *S. typhimurium* infection, CCR6-mediated DC recruitment to the epithelium was found to be required to induce a protective T cell response (Salazar-Gonzalez et al. 2006). Therefore, circulating non-conditioned DCs play an essential role in inflammatory responses against pathogens, while resident intestinal DCs ensure homeostasis in the gut.

1.2.3 Macrophages

We have also previously shown that SLPs from *C. difficile* RT 001 activate macrophages and induce clearance responses (Collins et al. 2014). Macrophages, along with intestinal epithelial cells DCs, compose the first line of defence against invading pathogens, and their importance during bacterial infection has been previously reported (Murray & Wynn 2011; Mosser & Edwards 2008). Intestinal macrophages constitute the largest reservoir of macrophages in the body, with the highest numbers in the colon (Lee et al. 1985). Unlike inflammatory macrophages, they express lower levels of co-stimulatory molecules CD80, CD86 and CD40 (Rogler et al. 1998). While expressing TLRs at the mRNA and protein level, they appear unresponsive to many TLR ligands (Smith et al. 2005; Smith et al. 2011). Despite a general lack of cytokine secretion, they were found to produce high levels of TNFα, even in a steady state, without inducing inflammation (Bain et al. 2013). Intestinal macrophages are also capable of phagocytosing pathogens and exhibit bactericidal abilities (Smythies et al. 2005). They are also likely important in maintaining homeostasis in the gut, as intestinal macrophage depletion leads to inflammation (Qualls et al. 2006). The ability of SLPs to induce cell surface molecule and cytokine production in inflammatory macrophages likely comes into play when non-intestinal macrophages are recruited to the site of infection.

Despite the anergic nature of intestinal macrophages, non-intestinal macrophages recruited to the site of infection play an essential role in bacterial clearance. Macrophages can reduce pathogen numbers by phagocytosing bacteria (Aderem & Underhill 1999) and defective phagocytosis has been shown to have a

detrimental role in disease (Taylor et al. 2010). Once activated, the production of pro-inflammatory cytokines and chemokines will recruit more inflammatory cells to the infected area. Macrophages are known to produce high levels of proinflammatory TNFα, a cytokine previously associated with Crohn's Disease (Plevy et al. 1997). However, TNFα secretion from macrophages can also aid bacterial clearance, with TNFα knockout mice infected with mycobacteria recovering upon administration of recombinant TNFα (Bekker et al. 2000). Phagocytosis is likely the key mechanism in bacterial clearance by macrophages, as it physically lowers bacterial cell count by direct ingestion. As neutrophils have been shown to not affect bacterial cell numbers in the gut, phagocytosis by macrophages is likely determinate in disease resolution (Jarchum et al. 2012; Collins et al. 2014).

1.2.4 Neutrophils

Neutrophils are a key component in the innate immune response, and play a critical role in *C. difficile*-induced colitis (Jarchum et al. 2012). These innate cells have the ability to phagocytose pathogens, as well as releasing antimicrobial molecules such as lytic enzymes and reactive oxygen intermediates (ROIs) (Borregaard 2010; Nathan 2006). Neutrophils also express PRRs, including TLR4, this indicates their potential to directly recognise the SLPs on *C. difficile* (Hayashi et al. 2003). These cells are relatively short lived, and do not proliferate upon activation (Colotta et al. 1992). This short lifespan suggests that neutrophils only play a passive role in resolution of infection, with fewer cells being recruited as infection resolves, and cells dying off. They have also been shown to play a more direct role, with apoptotic neutrophils increasing

expression of CCR5, which act as decoys for the chemokines CCL3 and CCL5, limiting chemotaxis as the neutrophils die (Ariel et al. 2006).

Cytokine production by other cell types such as macrophages can stimulate neutrophils to prolong their lifespan (Soehnlein & Lindbom 2010). This inhibition of programmed cells death is likely an important factor in clearing pathogens from the site of infection. Other studies have shown crosstalk between neutrophils and dendritic cells, with supernatant from activated neutrophils driving the maturation of bone marrow-derived dendritic cells *in vitro* (Bennouna & Denkers 2005). Neutrophils can also directly recruit and modulate the adaptive immune response, releasing chemokines specific for Th17 responses (Pelletier et al. 2010). These recruited Th17 cells, along with regulatory T cells, can produce high levels of IL-8, which can further recruit neutrophils in a positive feedback loop (Himmel et al. 2011).

1.2.5 $CD4^+$ T cells

$CD4^+$ T cells are another major population of cells involved in both homeostasis and protective immunity of the gut. $CD4^+$ T cells are distributed throughout and around the gastrointestinal tract, and are found in the mesenteric lymph nodes (MLNs), gut-associated lymphoid tissue such as Peyer's patches, as well as in the lamina propria and intraepithelial compartments of the colon (Mowat 2003; Cheroutre et al. 2011; Beagley et al. 1995). As previously stated, intestinal DCs can migrate to the MLNs upon antigen uptake to activate $CD4^+$ T cells present at the MLN (Coombes & Powrie 2008). This occurs even in a steady state,

indicating the role of DCs in driving regulatory phenotypes in $CD4^+$ T cells in the absence of infection (Izcue et al. 2009).

Th1 cells are required for clearance of intracellular pathogens in the gut. This phenotype is driven by the presence of IFNγ and IL-12, and characterised by the expression of the transcription factor T-bet. Expression of T-bet then leads to a positive feedback loop, producing IL-12 and IFNγ (Zhu et al. 2012). This production of IFNγ was thought to drive colitis, with an IFNγ neutralising antibody treating disease in murine models of inflammatory bowel disease (Powrie et al. 1994). However, late administration of IFNγ failed to provide protection, and in addition, an anti-IFNγ was shown to be ineffective in treating Crohn's Disease (Reinisch et al. 2010). Treatment with an anti-IL-12p40 antibody, however was found to effectively treat established disease (Neurath et al. 1995). This suggests a pathogenic role for this phenotype in IBD.

The protective role of the Th17 subset of $CD4^+$ T cells in the gut has recently been described (Blaschitz & Raffatellu 2010). IL-23, IL-6 and TGFβ have all been shown to drive a Th17 phenotype (Blaschitz & Raffatellu 2010; Zúñiga et al. 2013). These cells are characterised by expression of the transcription factor retinoic acid-related orphan receptor γt (RORγt) and secretion of IL-17A, IL-17F, IL-21, IL-22 and GM-CSF (Korn et al. 2009). These cells represent a large population in both infection and non-infection of the intestine (Weaver et al. 2013). Studies have shown a link between a Th17 phenotype and increased disease (Feng et al. 2011), yet their persistence in the gut during a steady state suggests a more complex role in intestinal immunity. IL-17 promotes tight

junction formation, an important protective feature in preventing systemic inflammation (Kinugasa et al. 2000). Th17 cells are important in the clearance of extracellular bacteria, and have been shown to play protective roles against many pathogens (Khader et al. 2009; Zhang et al. 2009a; Sellge et al. 2010; Ishigame et al. 2009).

Another important subset of T cells in both gut homeostasis and disease is the $Foxp3^+$ Treg cells, and are characterised by high levels of IL-10 secretion (Fujio et al. 2010). These cells work in an immunosuppressive fashion, and can be protective during colitis, releasing IL-10 which dampens down the immune response (Ranatunga et al. 2012). Deficiency in IL-10 has also been shown to exacerbate symptoms of colitis (Kullberg et al. 1998). Despite this protective function of IL-10, many pathogens have the ability manipulate its production to impede clearance (Redford et al. 2011; O'Leary et al. 2011; Wilson et al. 2011; Xavier et al. 2013). This modulation of the host immune system by pathogens is likely an important component of disease severity.

1.2.6 Immune Responses to *C. difficile*

As previously mentioned, toxin production by *C. difficile* results in disruption of tight junctions and damage to the intestinal epithelium. This induces pro-inflammatory cytokine production and immune cell migration to the site of infection (Dawson et al. 2009). As SLPs are the predominant surface components of the bacteria, they are exposed to the host immune system. Previous studies have shown them to induce activation of dendritic cells (DCs), with upregulation of many important cell surface markers such as CD80, CD86 and MHC class II

(Ausiello et al. 2006). They have also been shown to upregulate pro-inflammatory cytokine production, with elevated levels of IL-1β and IL-6 detected (Ausiello et al. 2006). This inflammatory response to SLPs is likely to play an important role in CDI.

Host antibody responses against SLP also appear to play a role in *C. difficile* infection (Drudy et al. 2004). Despite the fact that no difference in antibody levels between patients, asymptomatic carriers and controls was observed, IgM seems to be an important determinate for recurring infection. Patients suffering from recurrent cases were seem to have lower levels of IgM, relative to patients who only experienced an isolated case of CDI. Passive immunisation of hamsters with anti-SLP antibody was seen to modulate disease and delay progress of infection *in vivo* (O'Brien et al. 2005). This study detected an increase in phagocytosis of *C. difficile* in response to anti-SLP antibodies, suggesting a mechanism by which immunisation occurs.

Our group has also shown the ability of SLPs from *C. difficile* RT 001 to induce clearance mechanisms in macrophages (Collins et al. 2014). SLPs from RT 001 upregulated expression of proinflammatory IL-6, IL-12p40 and TNFα, in addition to chemokines MIP-2 and MCP. Co-stimulatory surface molecule expression was also increased. The rate of phagocytosis in the macrophages was upregulated by SLPs. This is indicative of increased clearance of the pathogens, as phagocytosis will decrease bacterial load. While spores can survive within phagosomes, the macrophages can prevent them from germinating, by isolating them from co-germinants (Paredes-Sabja et al. 2012). Inhibition of p38, a

downstream component of TLR4 signalling, resulted in the inhibition of proinflammatory cytokine production, and direct phosphorylation of p38 was observed in response to SLP. This again shows the signalling of immune cells through TLR4 in response to SLPs from *C. difficile* (Collins et al. 2014).

A role for neutrophils in *C. difficile* infection has recently been shown, specifically involving MyD88-mediated neutrophil recruitment to the gut (Jarchum et al. 2012). As SLPs signal through TLR4 to activate MyD88, the proteins may have direct involvement in recruiting these cells. This study found that recruited neutrophils can form a protective barrier in the wake of epithelial cell damage during colitis, preventing dissemination of intestinal bacteria. When the receptor necessary for neutrophil recruitment (CCR2) was knocked out in mice, a 90% reduction in monocyte recruitment was detected in the gut. This drop did not induce increased mortality however (Jarchum et al. 2012). Neutrophils also did not decrease toxin concentration in the gut, or decrease bacterial density. Their main role in CDI is therefore to form a protective barrier as the epithelial layer is damaged, preventing dissemination of bacteria to the underlying tissues, as other cell types, such as macrophages and T cells, resolve infection (Jarchum et al. 2012).

1.3 Inflammatory Bowel Disease

Inflammatory bowel disease (IBD) is the general term for inflammatory disorders of the gastrointestinal tract, and many infections, such as CDI can mimic the associated symptoms. It is characterized by an abnormal inflammatory response to antigens present in the gut, inducing a persistent inflammatory state (Scaldaferri & Fiocchi 2007). The two major forms of IBD are Crohn's Disease and ulcerative colitis; Crohn's Disease affects the entire gastrointestinal tract, causing widespread inflammation, while ulcerative colitis is normally localised to the colon (Podolsky 2002). In both cases, inflammation is associated with a breakdown of the protective epithelial barrier, resulting in activation of the underlying immune cells, and exacerbated inflammation.

Patients with IBD exhibit increased numbers of inflammatory cells, particularly macrophages, in the mucosa, with the cells exhibiting increased co-stimulatory surface molecule and proinflammatory cytokine expression (Xavier & Podolsky 2007; Rugtveit et al. 1997; Kamada et al. 2005). Macrophages in IBD also increase expression of PRRs such as TLR4, in addition to tissue-degrading cathepsins, all of which exacerbate inflammation in the gut (Smith et al. 2005). Production of TNFα by macrophages is crucial in the pathogenesis of IDB, as it is involved in initiation of a proinflammatory cytokine cascade, which maintains the damaging inflamed state of the gut (Parameswaran & Patial 2010). Dysregulation of DCs can also increase the inflammatory state in the gut in IBD. In a mouse model, depletion of DCs was seen to ameliorate dextran-sulphate sodium salt (DSS)-induced colitis (Berndt et al. 2007). Depletion of macrophages and DCs in IL-10 deficient mice was also seen to ameliorate the

symptoms of colitis (Watanabe et al. 2003). While the causes of IBD are not entirely clear, several intestinal pathogens, including *C. difficile*, can induce a colitis-like state in the gut, causing symptoms similar to that of ulcerative colitis or Crohn's disease.

1.4 Aims and Objectives

C. difficile has the ability to induce severe inflammation in the gut, resulting in a colitis-like state. The severity of disease can also vary depending on the strain causing infection. The Surface Layer Proteins coating the outer layer of *C. difficile* also exhibit variability between strains, and we have previously shown that SLPs from RT 001 induce proinflammatory cytokine production *in vitro*, indicating their ability to induce an immune response during infection. As certain strains are described as hypervirulent, we wanted to determine the evolutionary history of *slpA*, the gene encoding the SLPs. Using the output of this phylogenetic analysis, we wanted to examine if positive selection (a form of evolution indicative of functional shift) is present in the sequences. We also hypothesised that any positive selection present may affect recognition by the host immune response, which in turn may affect disease severity. Our hypothesis suggests that the amino acid sequences of the SLPs are a crucial determinate in susceptibility and severity of infection, and that the SLPs of different strains have the ability to modulate the immune response for the benefit of the pathogen.

To test this hypothesis the aims of this thesis were:

- To determine the evolutionary history of *slpA* and examine for signs of positive selection
- To purify SLPs from different ribotypes of *C. difficile*, and compare their effects on immune cells *in vitro*.
- To investigate differences in disease severity between strains of *C. difficile* with and without positive selection in an animal model, and identify a potential link between disease and evolution of SLPs.

C. difficile is now one of the most common nosocomial pathogens in the world, and the emergence of new strains resulting in more severe disease highlights the need for a better understanding of the pathogenesis of CDI. Unravelling the role that SLPs play in the inflammatory response to *C. difficile* may potentially reveal therapeutic targets to help ensure the host immune system mounts an appropriate and effective response to clear the pathogen from the gut and resolve infection.

Chapter 2

Materials and Methods

2.1 MATERIALS

2.1.1 *C. difficile* Culture Reagents

Material	Source
Blood Agar Plates	BD
Inoculation Loops	Cruinn
Cryovials	Lennox
Glycerol	Sigma-Aldrich
DPBS	Biosciences
Precept Tablets	VWR
DMEM	Biosciences
Anaerobic Indicator	Sigma-Aldrich
Anaerobic Gas Packs	Fisher
Sodium Thioglycolate	Sigma-Aldrich
Vitamin K	Sigma-Aldrich
Hemin	Sigma-Aldrich

Table 2.1.1 All reagents used for *C. difficile* culture and corresponding sources.

2.1.2 S-Layer Extraction Reagents

Material	Source
50 mL unskirted centrifuge tubes	Sarstedt
BHI Broth	Cruinn
NaCl	Sigma-Aldrich
Tris Base	Sigma-Aldrich
Urea	Sigma-Aldrich
Complete Protease Inhibitor Cocktail	Roche
Slide-A-Lyzer Dialysis Cassettes	Pierce
2mL Screw Cap tubes	Sarstedt
Bis-acrylamide	Sigma-Aldrich
SDS	Sigma-Aldrich
APS	Sigma-Aldrich
TEMED	Sigma-Aldrich
Methanol	Lennox
Acetic Acid	Sigma-Aldrich
Glycine	Sigma-Aldrich
Brilliant Blue	Sigma-Aldrich

Table 2.1.2 All reagents used for Surface Layer extraction and SLP purification, and corresponding sources.

2.1.3 Cell Culture Reagents

Material	Source
Tissue culture flasks T-75 cm^2	Nunc
Tissue culture flask T-25 cm^2	Nunc
Sterile Petri dishes	Nunc
6-, 24-, 96-well tissue culture plates	Nunc
96-well round bottom plate	Starsedt
Trypan blue (0.4% w/v)	Sigma-Aldrich
Foetal Calf Serum	Sigma-Aldrich
RPMI-1640	Gibco
DMEM	Sigma-Aldrich
Penicillin Streptomycin/Glutamine	Life Sciences
LPS (from *E. coli*)	Enzo Lifesciences
DPBS	Biosciences

Table 2.1.3 All reagents used for immune cell culture and corresponding sources.

2.1.4 ELISA Reagents

Material	Source
96-well microtiter plate	Nunc
TMB	BD
Tween 20	Sigma-Aldrich
Bovine serum albumin	Sigma-Aldrich
ELISA Duoset kits	R&D Systems
DPBS	Biosciences

Table 2.1.4 All reagents used for ELISA and corresponding sources.

2.1.5 Flow Cytometry Reagents

Material	Source
FACS Clean	BD
FACS Flow	BD
FACS Rinse	BD
Sodium Azide	Sigma-Aldrich
EDTA	Sigma-Aldrich
Propidium Iodide Solution	Sigma-Aldrich
Anti-CD80 antibody	BD
Anti-CD86 antibody	BD
Anti-TLR4 antibody	BD
Anti-MHC II antibody	BD
Anti-CD40 antibody	BD
Anti-DAPI antibody	BD

Table 2.1.5 All reagents used for Flow Cytometry and corresponding sources.

2.1.6 RNA Isolation and cDNA Synthesis Reagents

Material	Source
Nucleospin RNA II Columns	Macherey-Nagel
DEPC-treated dH_2O	Invitrogen
cDNA Reverse Transcription kit	Applied Biosystems
Molecular grade ethanol	Sigma-Aldrich

Table 2.1.6 All reagents used for RNA isolation and cDNA synthesis, and corresponding sources.

2.1.7 qPCR Reagents

Material	Source
SYBR® Green Mastermix	Roche
Taqman® Gene Expression Mastermix	Applied Biosystems
MicroAmp® Optical 96-well plate	Applied Biosystems
MicroAmp® Optical Adhesive Film	Applied Biosystems
IL-6 PrimeTime qPCR Primers	IDT
IL-10 PrimeTime qPCR Primers	IDT
IL-12p40 PrimeTime qPCR Primers	IDT
IL-23 PrimeTime qPCR Primers	IDT
TNFα PrimeTime qPCR Primers	IDT
IFNɣ PrimeTime qPCR Primers	IDT
TGFβ PrimeTime qPCR Primers	IDT
RORɣ PrimeTime qPCR Primers	IDT
TLR4 PrimeTime qPCR Primers	IDT
CCL2 PrimeTime qRT-PCR Assay	IDT

CCL3 PrimeTime qRT-PCR Assay	IDT
CCL4 PrimeTime qRT-PCR Assay	IDT
S18 PrimeTime qRT-PCR Assay	IDT
S18 PrimeTime qPCR Primers	IDT
B2M PrimeTime qRT-PCR Assay	IDT
B2M PrimeTime qPCR Primers	IDT

Table 2.1.7 All reagents used for quantitative PCR and corresponding sources.

2.1.8 Immunohistochemistry Reagents

Material	Source
Acetone	Sigma-Aldrich
HCl	Sigma-Aldrich
Sodium Bicarbonate	Sigma-Aldrich
Ethanol	Lennox
Harris Haematoxylin	Sigma-Aldrich
Eosin	Sigma-Aldrich
Histo-clear	Fisher
OCT Solution	TissueTek
Superfrost Slides	Fisher
DPX Mountant	Sigma-Aldrich
Anti-NIMP Antibody	Abcam
Anti-CD4 Antibody	eBiosciences
Anti-F4/80 Antibody	eBiosciences
DAB Tissue Staining Kit	R&D Systems

Table 2.1.8 All reagents used for immunohistochemical staining and corresponding sources.

2.2 BIOINFORMATICS TOOLS

Using specific software and bioinformatics tools, it is possible to detect signatures of positive selection in the *slpA* sequences of multiple ribotypes of *C. difficile*. Sequences obtained from online databases can be used to construct a multiple sequence alignment (MSA) based on sequence similarity. Statistical tests are then performed to ensure sufficient phylogenetic signal is present before the phylogeny of the sequences can be resolved. Using the phylogenetic tree and MSA, sites of positive selection can be estimated by analysing the rate of synonymous mutations per synonymous site to non-synonymous mutations per non-synonymous site. The general workflow for the bioinformatics analysis is given in Figure 2.2.1.

Figure 2.2.1

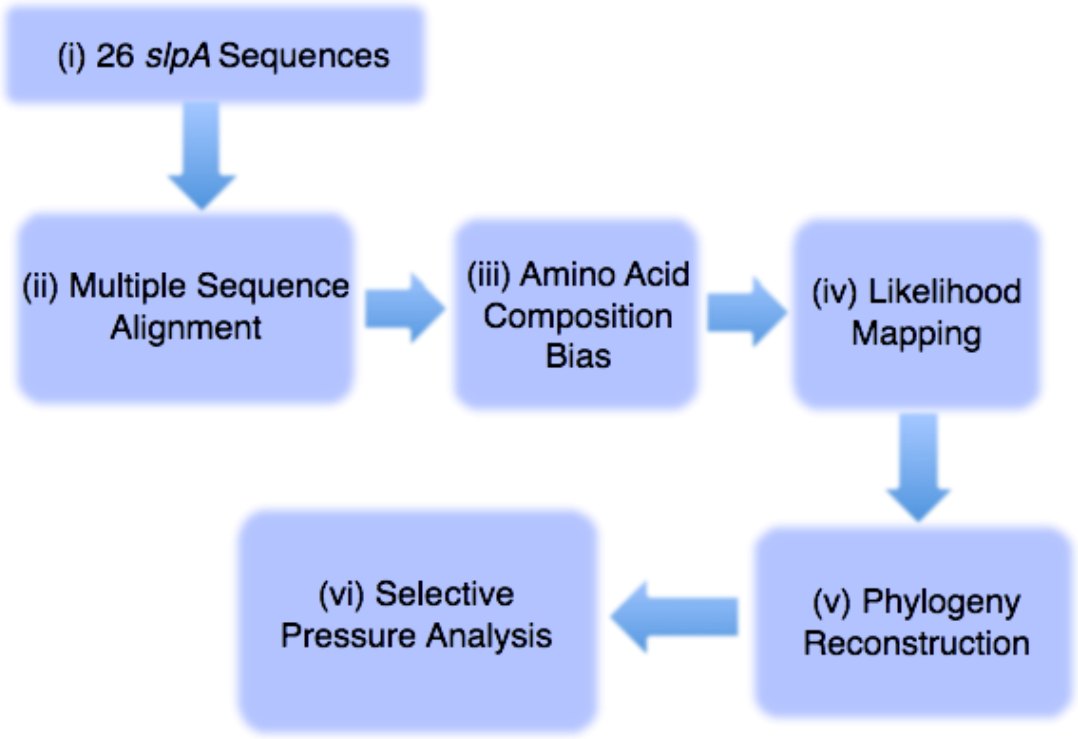


Figure 2.2.1 Workflow describing the process of biomolecular evolution analysis from sequence acquisition to selective pressure analysis.

2.2.1 Sequence Data Assembly

slpA gene sequences were obtained from 26 different strains of *C. difficile*, representing the majority of clinically important ribotypes and serogroups. A literature search was first conducted for complete DNA sequences of the *slpA* gene (Eidhin et al. 2006; Karjalainen et al. 2002). GenBank and EMBL accession numbers were obtained and used to retrieve the sequences (http://www.ncbi.nlm.nih.gov/genbank/). BLAST searches were performed to ensure all the coding DNA sequences (CDS) for *slpA* had been obtained. The sequences varied from 1833 to 2304 base pairs in length. The sequence data are given in Table 2.2.1.

2.2.2 Multiple Sequence Alignment

The nucleotide sequences were translated to amino acids using in house software "Translfas.c". Multiple sequence alignments (MSAs) for the amino acid sequences were then obtained using two software packages, ClustalX 1.83.1 (Jeanmougin et al. 1998) and MUSCLE 3.6 (Edgar 2004) with both alignments yielding the same result, the alignment produced from MUSCLE was chosen for all subsequent analyses. The alignment was edited by manual inspection, and ambiguous regions were removed, i.e. small identical regions on multiple strains that have been falsely misaligned, but are easily aligned by eye. The final alignment was then used to insert gaps into the nucleotide sequences in the positions corresponding from the protein alignment using in house software ("Putgaps.c"). The overall alignment length was 2,652 nucleotides, or 884 amino acids.

Table 2.2.1 – All *slpA* sequences used in bioinformatics analysis

Ribotype	Accession No.	TcdA/TcdB	Length of sequence	Serogroup
001	DQ060625 DQ060626 DQ060627 AJ300676	+/+	A.A: 756 Nuc: 2,271	G
002	DQ060628 DQ060629	+/+	A.A: 726 Nuc: 2,181	A2
005	DQ060630 DQ060631	+/+	A.A: 610 Nuc: 1,833	A1
010	DQ060633 AF478571	-/-	A.A: 767 Nuc: 2,304	D
012	DQ060634 DQ060635 AJ291709	+/+	A.A: 719 Nuc: 2,160	C
014	DQ060638	+/+	A.A: 732 Nuc: 2,199	H
016	AF478570	n/a	A.A: 610 Nuc: 1,833	A1
017	DQ060640 AJ300677	-/+	A.A: 714 Nuc: 2,145	F
027	R20291 CD196	+/+	A.A: 758 Nuc: 2,277	Unknown
031	DQ060641	-/-	A.A: 739 Nuc: 2,220	K

046	DQ060636	+/+	A.A: 719 Nuc: 2,160	C
054	DQ060632	+/+	A.A: 610 Nuc: 1,833	A1
066	DQ060639	-/-	A.A: 732 Nuc: 2,199	A9
078	DQ060643	+/+	A.A: 726 Nuc: 2,181	Unknown
092	DQ060637	+/+	A.A: 719 Nuc: 2,160	Unknown
094	DQ060642	+/+	A.A: 739 Nuc: 2,220	Unknown

Table 2.2.1 The table shown lists all the ribotypes used in the dataset. GenBank accession numbers for the *slpA* gene of each ribotype are given. In certain cases, *slpA* sequences from different strains within the same ribotype were included to allow as robust as possible analyses. Also included is information on sequence length, serogroup, and toxin production.

2.2.3 Amino Acid Composition Bias

Composition bias may result in distantly related sequences being falsely grouped together due to similar amino acid compositions. The TreePuzzle 5.2 software (Schmidt et al. 2002) performs a chi-squared test that compares the amino acid composition of each sequence in the dataset to the frequency distribution assumed in the maximum likelihood model. This distribution assumes homogeneity of composition. Every sequence should ideally pass this test and sequences that fail should be excluded from the dataset.

2.2.4 Likelihood Mapping Tests

Likelihood mapping involves dissembling the phylogenetic tree into all possible quartets and assessing the support for each possible quartet (Strimmer & von Haeseler 1997). If the data contains no phylogenetic signal, then the likelihood of all three possible relationships for that quartet will be equally likely, represented by three tips of a triangle. If sufficient phylogenetic signal is present, the majority of the signal will appear in these tip regions. If little or no phylogenetic signal is present, the majority of the signal will be in the vertices and central region of the triangle, representing a network or star phylogeny respectively. An example of good and bad signal can be seen in Figure 2.2.2 below.

Figure 2.2.2

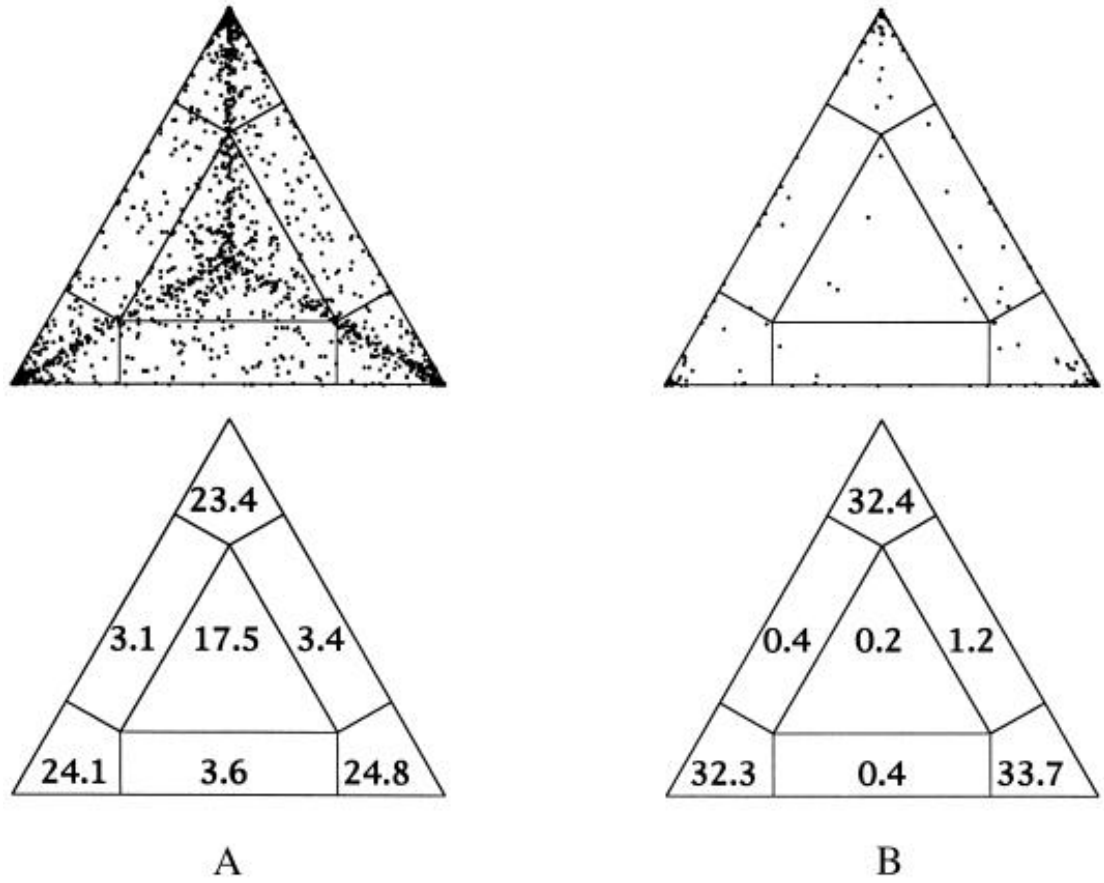


Figure 2.2.2 Examples of likelihood mapping where little signal is present (A) and high signal is present (B). In the top row the two triangles show two different patterns of signal. Each dot in these triangles represents a different quartet. The numbers in the bottom triangle represent the percentages of quartets in each section. The fewer dots in the centre of the triangle, and the more at the edges, the greater the amount of phylogenetic signal.

2.2.5 Phylogeny Reconstruction

To determine the most suitable model of evolution for the data, we employed Modelgenerator version 0.85 (Keane et al. 2006). In total, 88 models were tested with the data and compared with each other to see which was a better fit to the data. The protein model of substitution found to be the best fit for the data was the WAG substitution matrix (Whelan & Goldman 2001) with a discrete gamma model of rate heterogeneity (G), and an amino acid frequency estimated from the data (F). The phylogenetic tree for the *slpA* sequences was estimated under Bayesian inference using MrBayes v3.2.1 and the amino acid alignment (Huelsenbeck & Ronquist 2001). The tree was generated using the Nearest Neighbour Interchange (NNI) tree search algorithm and 100 bootstrap replicates implemented under the Akaike Information Criterion (AIC) statistic. The substitution model was WAG+G+F with discrete gamma model of rate heterogeneity. Clade supports were given as posterior probabilities. The settings used in MrBayes v3.2.1 are given in Table 2.2.2 below. A Shimodaira-Hasegawa, or SH test was performed on all trees generated to test for any statistically significant differences between the trees.

Table 2.2.2 MrBayes Settings

Parameter	Setting
Autoclose	Yes
lset applyto	All
nst	6
rates	gamma
prset aamodelpr	fixed (wag)
mcmcp ngen	500,000
printfreq	5,000
samplefreq	500
nchains	64
sacebrlens	yes
burnin	500

Table 2.2.2 Settings used for phylogenetic tree construction by MrBayes v3.2.1.

2.2.6 Selective Pressure Analysis

Codon models of evolution implemented in PAML 4.3 (Yang 2007) were used in the analysis of the SLPs. Site-specific and lineage-specific models were applied to the data, allowing for ω values across sites and along different branches. The models differed from each other in their complexity. Three models were applied for the branch-specific analysis and seven were used for the site-specific analysis, these represent all currently available codon models.

The first model M0 (also known as the Goldman and Yang model) (Goldman & Yang 1994) assumes that the rate of evolution is constant across all sites and lineages, and calculates a single value for ω across the entire alignment. The next model is known as M1 and allows for two classes of sites with $\omega_0 = 0$ and $\omega_1 = 1$; under this model purifying selection or neutral evolution are allowed, but positive selection is not permitted. Model M2, "the selection model", adds more parameters to M1 and allows for three classes of sites. The classes of sites can have ω = 1, ω = 0 and ω estimated entirely from the data (and free to be >1), and all associated proportions of sites fitting into each of these categories is estimated from the data (p0 etc). M1 and M2 can be compared with one another, as M2 is an extension of M1 with the addition of an extra parameter. An LRT can then be used to examine whether there is a statistically significant difference between the two models. The next model, M3, an extension of M0, contains two further parameters and allows ω values to be estimated entirely from the data. This model can allow two classes of sites to vary (k = 2) or three classes of sites to vary (k = 3). An LRT between M3 (k=2) and M0 can be used to determine if the

M3 (k=2) model better fits the data than M0. M3 (k=3) can be compared with M3 (k=2) (Yang & Swanson 2002).

The remaining models are different from those mentioned previously as they use discrete approximations to continuous distributions in order to model variability in ω at different sites along the sequence. M7 gives variation in ω a beta distribution. Under this model, ten classes of sites are assumed to exist with ω values constrained between zero and one. M8 is a similar model to M7, but it allows for an additional class of site with its ω value estimated entirely from the data and free to be greater than 1. M8 can be compared with M7 in an LRT, a model that does not allow for positive selection (Yang & Nielsen 2002). A final model, M8a, is the null model of M8 (Yang & Nielsen 2002). It restricts the additional site category that is estimated from the data to be ω = 1, and therefore does not allow for positive selection. If M8 then gets a statistically significantly better likelihood score than M8a, then any positive selection observed under M8 is more likely to be accurate (Yang & Nielsen 2002).

The two lineage-specific models currently being used are Model A and Model A null. Model A allows ω to vary across different lineages as well as across sites. Model A is a lineage-specific extension of M1. Model A null does not allow for positive selection, it can be compared with Model A to test whether or not Model A is a good fit for the data. The posterior probability of any given site in the alignment being under positive selection can be estimated using either Naive Empirical Bayes (NEB) or Bayes Empirical Bayes (BEB) (Yang & Nielsen 2002). NEB is less robust than BEB and is prone to error. In particular this

occurs in small datasets where ML estimates may have large sampling error and can result in false positives for sites under positive selection (Anisimova et al. 2002). BEB is more robust and reduces the rate of false positives for small sample sets (Yang et al. 2005).

A series of likelihood ratio tests (LRTs) were then carried out between models that are nested to determine the statistically significant model for the data for each branch in the tree and for the multiple sequence alignment. See Table 2.2.3 full list of all LRTs carried out. For each model, the log likelihood (lnL) values were recorded, with the lnL values closest to zero representing a closer fit to the data. χ^2 tests were then used to determine the significance of these models.

Table 2.2.3 Likelihood Ratio Tests

Comparison	df	Δl	Critical χ2 values
M0 v M3k2	2	X2	≥ 5.99
M3k2 v M3k3	-	X1	≥ 1.00
M1 v M2	2	X2	≥ 5.99
M7 v M8	2	X2	≥ 5.99
M8 v M8a*	1	X2	≥ 2.71 (@5%)
			≥ 5.41 (@1%)
M1 v model A	2	X2	≥ 5.99
Model A v Model A null	1	X2	≥ 3.84 (@5%)

Table 2.2.3 Table showing how the different models of evolution are compared to one another, along with the critical chi-square value that must be obtained for the particular model to pass the likelihood ratio test (LRT). Taken with permission from Morgan et al. 2010.

The difference between the lowest lnL values of models and lnL values of more complex models was calculated. This value, which is the difference in likelihood, was multiplied by the degrees of freedom given in the table below. If the resulting value exceeds the critical χ^2 value, then the result is significant. If this applies for complex models, then that model of evolution is a better fit and the associated parameter estimates for omega and proportion of sites are taken from that model.

Sites under positive selection were estimated using the empirical Bayes methods for both site and lineage-specific analyses. The methods used were naïve empirical Bayes (NEB) and Bayes empirical Bayes (BEB). Previously published knowledge on functionally important sites in the two proteins, HMW and LMW, were then combined with the positive selection results to determine if there is a correlation between positive selection and specific functions of the protein. A 3D structure of the LMW SLP was obtained from EMBL-EBI (http://www.ebi.ac.uk/). The PDB code for this structure is 3cvz. It was used to visualise positively selected residues on the protein.

2.3 *C. DIFFICILE* CULTURE AND SLP PURIFICATION

2.3.1 Culture on Solid Agar

Frozen *C. difficile* spore were acquired from the Central Pathology Laboratory at St. James's Hospital, Dublin. Spores were streaked on to Brucella Agar plates with 10% Horse Blood (BD) using sterile inoculation loops, and allowed to grow at 37°C for 48 hours. Plates were incubated anaerobically using gas-generating kits in a sealed anaerobic jar. Anaerobic indicator strips, which turn pink in the presence of oxygen, were used to confirm anaerobic conditions. After 48 hours, grey colonies, which were opaque, raised, and irregular in shape, were observed on the plate. A single colony was then used to streak a fresh agar plate using a sterile loop, which was then incubated in the same fashion as described previously. This was repeated for as long as colonies were required. All bacterial work was carried out in a class II safety cabinet.

2.3.2 Stock Preparation

To prepare fresh stocks for each new isolate, a single colony was streaked on to a fresh agar plate and incubated for 48 hours as described previously. One plate per frozen stock required was made. Two stock solutions were tested, a PBS/glycerol whole cell stock solution and a DMEM/glycerol spore stock solution. A solution of PBS with 15% glycerol was used for the whole cell stock solution. Two mL aliquots were added to sterile cryovials. After 48 hours, the plates were removed from the incubator. All the colonies from a single plate were collected using a sterile swab and added to the stock solution. This process was repeated as required. For the spore stock preparation, confluent cultures of each strain were grown anaerobically at 37°C for 5–7 days to generate spore

formation. The cultures were harvested with disposable loops into 1 mL of PBS in screw cap tubes. Tubes were then spun down at 8,000 rpm and the supernatant was removed. The cell pellet was washed in 1 mL PBS, and heat-shocked at 56°C for 10 min to kill surviving vegetative cells. The spores were centrifuged and re-suspended in 1 ml DMEM with 15% glycerol. All stocks were stored at -80°C.

2.3.3 Liquid Culture

C. difficile was also cultured in liquid Fastidious Anaerobe Broth (FAB) and Brain Heart Infusion (BHI) broth. BHI broth was autoclaved before use and supplemented with Sodium Thioglycolate, Hemin and Vitamin K. Sodium Thioglycolate (0.05%) was added to remove oxygen from the broth. Hemin (0.5%) and Vitamin K (0.1%) were added as additional nutrients for the bacteria.

2.3.4 Surface Layer Extraction

A blood agar plate was streaked as described above and *C. difficile* colonies were grown for 48 hours. BHI broth was added to 12 non-skirted 50 mL tubes and placed under anaerobic conditions for at least an hour to remove oxygen. One colony per 50 mL tube was used to inoculate the BHI broth, which was then incubated anaerobically for ~16 hours. As the growth characteristics for *C. difficile* varied between strains, this time needed to be adjusted and optimised for each strain used. When the broth appeared cloudy (with an OD of 0.8-1.0), the tubes were centrifuged at 3,000 rpm for 20 min to collect the cells. Supernatants were discarded and cell pellets were then resuspended in ice-cold 50 mM Tris:HCl pH 7.4. Ten mL was used per tube. The resuspended bacteria was

combined into 4 tubes containing 30 mL each. These tubes were again centrifuged at 3,000 rpm for 20 min and pellets were resuspended in ice-cold 50 mM Tris:HCl pH 7.4. This was then combined into 2 tubes of 20 mL each and centrifuged at 3,000 rpm for 20 min. Cell pellets were then resuspended in 20 mL room temperature 8 M urea/50 mM Tris:HCl pH 8.3, supplemented with complete protease inhibitor cocktail and incubated for 90 min at 37°C. The mixture was centrifuged at 12,000 rpm for 30 min and the supernatant, which contained the crude S-Layer, was collected. The extract was stored at -20°C until ready for dialysis.

2.3.5 Dialysis of crude S-layer proteins

The crude extract was dialysed against 20 mM Tris:HCl pH 8.5 to remove urea from the mixture. Slide-A-Lyzer® G2 Dialysis Cassettes with a Membrane Molecular-Weight Cut-off (MWCO) of 10 K were used, being submerged in 5 L of buffer. Buffer was changed every 2 hours, with a total of 4 buffer changes. The dialysed sample was then filtered using a Polyethersulfone (PES) Membrane filter with a pore size of 0.2 µm and aliquoted into 10 mL tubes. The sample was stored at -20°C.

2.3.6 Fast Protein Liquid Chromatography (FPLC)

Fast Protein Liquid Chromatography (FPLC) is a form of liquid chromatography used to separate specific proteins from a mixture by the concept of anion exchange. The crude protein mixture is injected into an anion exchange column, where the proteins bind tightly. Buffer is then run through the column with an increasing salt gradient, and over time the proteins are displaced from the

column, at a rate dependent on their affinity to the charges on the column. Therefore different proteins are eluted at different time points.

Surface Layer Proteins (SLPs) were purified by FPLC using the MonoQ HR10/10 anion exchange column attached to the AKTA FPLC system. SLPs were eluted with a linear gradient of 0-0.3 M NaCl in 20 mM Tris HCl pH 8.5 at a flow rate of 4 ml/min. Fractions of 2 mL were collected. Large peaks in UV readings on the column related to collected fractions that contained a significant amount of protein. As the SLPs of each strain used varied in size, adjustment of the parameters in the UNICORN software were required for each strain to ensure optimal elution of the SLPs. The parameters are listed in Table 2.3.1 below.

Table 2.3.1

Parameter	Setting
Pressure Limit	4.00 MPa
UV Averaging Time	5.10
Flow Rate	4.00 mL/min
Starting Conc. Buffer B	0.00% B
Equilibrate with	1 CV (Column Volume)
Flowthrough Fraction Size	0.00 mL
Start Flowthrough at	NextTube
Empty loop with	10.0 mL
Wash Column with	5 CV
Start Frac at	60% B
Eluate Fraction Size	2.0 mL
Starte Eluate Frac at	FirstTube
Peak Frac Size	0.00 mL
Start Peak Frac At	NextTube
Peak Start Slope	100.00 mAU/min
Peak End Slope	75.00 mAU/min
Minimun Peak Width	0.31 min
End Frac At	100% B
Target ConcB 1	60% B
Length of gradient 1	20 CV
Target Conc B 2	100% B
Length of Gradient 2	4 CV
Target Conc B 3	100% B
Length of Gradient 3	2 CV
Conc of Eluent B	100% B
Clean with	5 CV
Reequilibrate with	2 CV

Table 2.3.1 Table of all variable paramaters in FPLC analysis

2.3.7 SLP Purification – SDS-PAGE

Polyacrylamide Gel Electrophoresis was used to visualise the presence of proteins in the collected fractions and determine those in which purified SLP was present. Protein samples were diluted in 5x Loading Buffer supplemented with Dithiothreitol (DTT). Samples were heated to 96°C for 5 min to denature any protein structures and allowed to cool to room temperature. Stacking and Separating Gels were made. Gels were 10% acrylamide. APS (10%) and TEMED were added last with gentle swirling to induce setting. The PageRuler™ Plus Prestained Protein Ladder (Fermentas) was run alongside all protein samples as a size indicator. Two bands appearing at around 32 kDa and 44 kDa in the same lane indicate the presence of SLPs. Gels were submerged in 1x Running buffer and electrophoresis was carried out at 30 mA per gel for approximately one hour, until the dye front reached the base of the gel. Recipes for all solutions are given in the Appendix.

2.2.8 Coomassie Staining

Once electrophoresis was complete, gels were washed with dH_2O and submerged in 10 mL of Coomassie Blue stain. The gels were then left on a rocker for 1 hour to stain any protein bands present, then the coomassie stain was removed and the gel was washed briefly with dH_2O. 10 mL of destain solution was added to the gels. They were placed on the rocker for 10 min before replacing the buffer with fresh destain. This was repeated four times. The gels were then left rocking in destain overnight at 4°C. Gels were then examined for the presence of SLPs.

2.2.9 Concentration of purified SLPs

Relevant fractions containing pure SLPs were collected and pooled together. They were then concentrated down to a volume of ~200 μL using Amicon Ultra-4 filter unit centrifuge tubes. Samples were spun at 4000 rpm for 30 min to achieve this reduction in volume. Samples were then collected and exposed to a UV light for 15 min to ensure sterilisation. SLPs were then aliquoted and stored at -20°C until ready for use.

2.2.10 Confirmation of *slpA* Sequences

The strains used in this study included R13537 (RT 001) and R12885 (RT 014). In these strains the sequence of the *slpA* gene has been previously determined (accession numbers DQ060626 and DQ060638 respectively). To determine the *slpA* gene sequences of our clinical strains belonging to ribotypes 027 and 078, whole-genome sequencing was performed at Trinity College Dublin.

2.3.11 LAL Endotoxin Testing of SLPs

SLP samples were tested for the presence of endotoxin, to ensure any observed immune response was not due to contaminants. The ToxinSensor™ Chromogenic LAL Endotoxin Assay Kit from GenScript was used, as Limulus amebocyte lysate (LAL) reacts with bacterial endotoxin and lipopolysaccharide. A series of standards were made from 1 EU/mL (endotoxin unit/mL) endotoxin stock solution, yielding standards of 0.1, 0.05, 0.025 and 0.01 EU/mL. Standards and samples (100 μL) were placed in specific endotoxin-free vials in duplicate. A blank of LAL reagent water was also prepared. LAL (100 μL) was added to each vial and the samples were incubated at 37°C for 10 min. Chromogenic substrate

solution (500 μL) was then added, the samples were gently mixed and incubated for 6 min. Stop solution (500 μL) and colour stabiliser were added and samples were gently swirled to avoid generation of bubbles. Absorbance was read at 545 nm and the amount of endotoxin present in the samples determined from the standard curve. Assays were run in duplicate and included a reagent diluent only well. Endotoxin testing was carried out on two representative preparations of SLPs and subsequent SLP preparations were carried out under the same conditions.

2.3.12 SLP Quantification

The concentration of SLP samples was determined using the Pierce® BCA (bicinchoninic acid) Protein Assay Kit. The assay uses the well-documented reduction of Cu^{+2} to Cu^{+1} by protein in a base, along with the colourimetric detection of Cu^{+1} using bicinchoninic acid (BCA). A purple colour is observed in the presence of protein, with darker colour signifing a higher concentration. A BSA (Bovine Serum Albumin) standard curve was prepared in 20 mM Tris:HCl pH 8.5 from a 2,000μg/mL BSA stock solution using serial dilutions to yield standards at 2,000, 1,500, 1,000, 750, 500, 250, 125, 25 and 0 μg/mL BSA. Each standard and sample (25 μL) were added to a 96 well plate in triplicate. BCA Working Reagent (200 μL) was added to each of the wells and the plate was placed on a shaker for 30 seconds. The plate was covered and incubated at 37°C for 30 min. Samples were allowed to cool to room temperature and absorbance was read at 562 nm on a plate reader. A standard curve was constructed using the blank-corrected absorbance values for each standard vs. its concentration in μg/mL. Unknown samples were calculated from the standard curve.

2.4 CELL CULTURE AND STIMULATION

2.4.1 J774 Macrophage culture

The J774 macrophage murine cell line were grown in T75 cell culture flasks using DMEM (Dulbecco's Modified Eagle Medium), supplemented with 10% FCS (Fetal Calf Serum) and 2% Pen/Strep (a mixture of penicillin and streptomycin). FCS was heat-inactivated to inactivate complement. Cells were grown at 37°C in 5% CO_2 in a humidified incubator. When cells reached ~90% confluence, adherent cells were removed gently scraping the flask with a sterile cell scraper. Cell suspension was collect in a 50 mL centrifuge tube and centrifuged at 1,200 rpm for 5 min. The resulting cell pellet was resuspended in 5 mL of fully supplemented RPMI and split into new culture flask with fresh media.

2.4.2 Bone Marrow-derived Dendritic Cell Culture

BALB/c mice (10-14 weeks of age) were sacrificed and hind legs were removed. Flesh and muscle were carefully removed from the leg bones and bones were stored in RPMI (Roswell Park Memorial Institute) media on ice. All flesh was removed from the bones, the femur was cut above the knee and below the hip to expose bone marrow. A syringe with a 27.5 G needle was filled with RPMI and used to flush the bone marrow out of the femur into a sterile petri dish. The same process was repeated for the tibia. All bone marrow was collected in a single petri dish. A fresh sterile 10 mL syringe with a 19 G needle was used to collect the bone marrow, which was placed in a 50 mL centrifuge tube. This was centrifuged at 1,200 rpm for 5 min. Cells were resuspended in 5 mL of media and a cell count was performed. Enough media was added to the cells to load 1

mL of cells per plate. RPMI supplemented with GM-CSF (Granulocyte macrophage colony-stimulating factor) was prepared, with 9 mL required for every plate used. Cells were plated and incubated at 37°C in 5% CO_2 in a humidified incubator for 3 days. Media (7 mL) was removed from each plate and 10mL of fresh RPMI/GM-CSF was added. Cells were incubated for 4 days. Cells were then removed from dishes with a sterile cell scraper, collected in a single 50 mL centrifuge tube and centrifuged at 1,200 rpm for five min. A cell count was performed and cells were plated at $0.5x10^6$ cells/mL.

2.4.3 $CD4^+$ T cell Culture

BALB/c mice (10-14 weeks of age) were sacrificed and the spleen was removed. All fat was removed from the spleen, which was then placed on a 40 µm nylon cell strainer (BD) above a 50 mL tube. The strainer was moistened with RPMI media. A sterile Pasteur pipette was filled with media, and was used to apply pressure to the spleen, breaking apart the tissue. Media (20-30 mL) was used to homogenise the tissue through the cell strainer. Homogenised splenocytes were then decanted through a fresh 40 µm strainer to remove any excess fat or debris. Splenocytes were centrifuged at 1,200 rpm for 5 min at 4˚C and the supernatant was aspirated. Cells were resuspended in 10 mL media and a cell count was performed as previously described. Cells were adjusted to a concentration of $1x10^8$ cells/mL using recommended media (PBS/2%FBS/1 mM EDTA).

The EasySep™ Mouse $CD4^+$ isolation kit was then used to purify CD4+ cells. Two mL of cells were placed in a 5 mL polystyrene tube and 100 µL of Normal Rat Serum was added (50 µL/mL of cells). 100 µL of EasySep™ Mouse $CD4^+$ T

Cell Isolation Cocktail was added and mixed well. Cells were incubated at room temperature for 10 min. After vigorous vortexing, EasySep™ Streptavidin RapidSpheres™ were added at a concentration of 75 µL/mL of cells. The suspension was mixed well and incubated at room temperature for 2.5 min. The final volume of the cell suspension was brought to 2.5 mL. The polystyrene tube was placed in the EasySep™ Magnet and incubated for a further 2.5 min. The magnet was then inverted in one continuous motion, and the cells were collected in a new sterile polystyrene tube. A cell count was then preformed to determine concentration.

2.4.4 $CD4^+$ T cell/Dendritic Cell co-culture

Dendritic cells were grown as previously described. At Day 7 of the culture, Cells were plated at $1x10^6$ cells/mL and stimulated with SLPs (20 µg/mL) and LPS (100 ng/mL). Cells were then activated with OVA peptide (5 µg/mL). Cells were then incubated for a further 24 hours at 36˚C in 5% CO_2. After 24 hours, cells were washed in PBS/2%FBS, and irradiated (at Trinity College Dublin, under supervision) with 40 Gy (4000 rads) using a gamma irradiator with a Cesium-137 source. Cells were then resuspended at a concentration of $2x10^5$ cells/mL. OVA-specific $CD4^+$ T cells were isolated from OVA transgenic D011.10 mice as previously described. These T cells are specific to OVA peptide and can be activated by dendritic cells displaying OVA. OVA-specific T cells were adjusted to a final concentration of $2x10^6$ cells/mL. Dendritic cells (100 µL) and 100 µL of T cells were added to a 96 well plate to give a final volume of 200 µL/well. Cells were cultured for 5 days at 37˚C in 5% CO_2. At Day 5, cells were centrifuged and 200 µL of media was carefully removed. Cell pellets were

resuspended in 200 µL of fresh media, and incubated further until day seven. At day six, fresh dendritic cells were stimulated with LPS, SLP and OVA in the same manner as before, and incubated for 24 hours. At day seven, cells were again centrifuged and 100 µL of media was removed. The dendritic cells were again irradiated and 100 µL (at $2x10^5$ cells/mL) was added to the wells for the second round of T cell stimulation. Cells were incubated until day 10, when media was removed for cytokine analysis.

2.4.5 Cell Enumeration

Cell counts and viability tests were performed using a haemocytometer and trypan blue dye. When a cell population is mixed with the dye, live cells with intact membranes can exclude the dye from the cytoplasm, while dying cells with permeable membranes allow the dye to enter the cell. Live cells will appear with a bright halo around them, while dead/dying cells appear dark blue. The number of viable cells, as well as cell viability can then be calculated. A moistened coverslip is placed over the haemocytometer. Cell suspension (100 µL) was added to 250 µL Trypan blue dye and 150µL PBS and incubated at room temperature for 2 min. The cell suspension was then applied to the Brightline™ Neubauer haemocytometer (Sigma) and the cells were drawn into the chamber by capillary action. Cells inside the 1 mm^2 grid were counted and averaged under 10x magnification using a phase contrast microscope (Figure 2.4.1). Cell number was then counted using the formula: Cells/mL = N x D ($x10^4$), where N = average cell number and D = dilution factor.

Figure 2.4.1

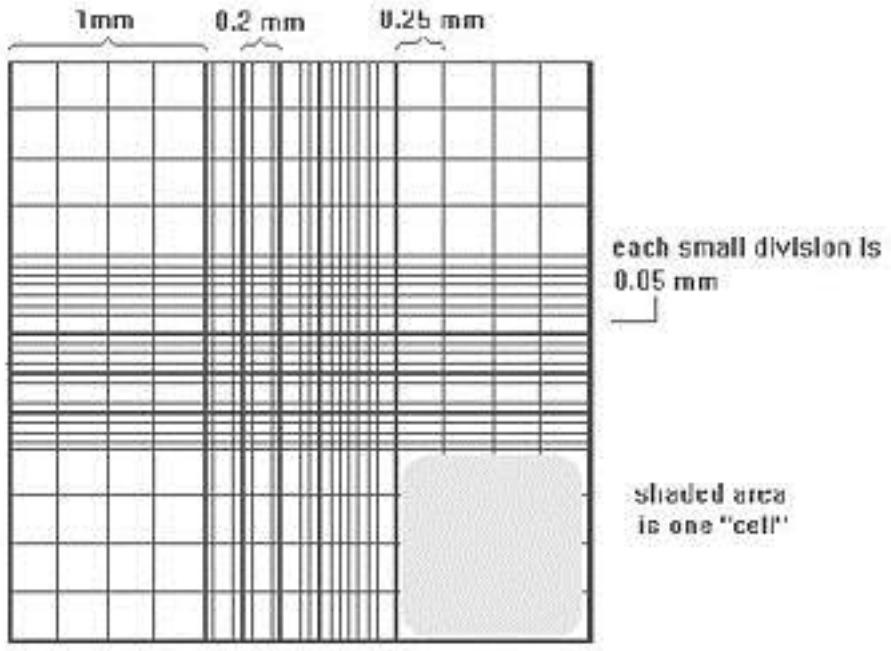


Figure 2.4.1 Haemocytometer used for cell enumeration.

2.4.6 Cell Stimulation

As previously described, cells were plated at a density of $0.5x10^6$ cells/mL, and incubated until adhered to the base of the well. Cells were stimulated with SLPs from all available ribotypes at a concentration of 20 μg/mL for 24 hours. Cells were also stimulated with LPS at a concentration of 100 ng/mL as a positive control. Control cells were incubated in media alone. After 24 hours, supernatant and cells were collected. Supernatant was stored at -20°C until required.

2.4.7 Enzyme Linked Immunosorbant Assay (ELISA)

Enzyme Linked Immunosorbant Assay (ELISA) is a technique for used for the detection and quantification of proteins in cell supernatant. Sandwich ELISAs were used for all studies. Immunoassays for the detection of IL-1β, IL-2, IL-4,

IL-10, IL-12p40, IL-12p70, IL-23 and TNFα were carried out. R&D Systems DuoSet ELISA kits were used in each case, with the designated protocol being optimised where necessary. Capture antibody, specific for the target molecule, was diluted in PBS to a concentration of 2 μg/mL, and coated on a 96-well plate. Plates were covered and incubated overnight at room temperature. Plates were then washed three times with wash buffer (PBS/0.05% Tween 20) and blocked for two hours with 300 μL of blocking buffer per well. This ensured that any areas of non-specific binding on the antibody were effectively blocked. A series of standards for each cytokine being assayed was prepared from a stock solution, diluted in reagent diluent (1% BSA in PBS). Plates were washed 3 times and samples and standards were added to the plate in duplicate. ELISA plates were then incubated overnight at 4°C. Biotinylated detection antibody was diluted in reagent diluent to a working concentration of 50 ng/mL. The plates were washed three times and 50 μL of Detection Antibody was added per well. Plates were incubated for two hours at room temperature to ensure maximum binding of the detection antibody to the target molecule. Plates were again washed and a 1/200 dilution of streptavidin-HRP was prepared. Streptavidin-HRP (50 μL) was added per well and plates were covered and incubated for 20 min at room temperature and then washed. Streptavidin has a high affinity for biotin, therefore the amount of HRP present after washing is proportional to the amount of target protein HRP-substrate (50 μL), Tetramethylbenzidine (TMB), was added per well. Plates were incubated in the dark at room temperature for 20 min. HRP catalyses the oxidation of the TMB substrate, and a blue compound is formed. The intensity of the colour is proportional to the concentration of the target protein. The reaction was stopped by addition 25 μL of 2 N H_2SO_4 per well to prevent further

oxidation. Absorbance was read at 450 nm. A standard curve was constructed using absorbance values for each standard vs. its concentration in μg/mL. Unknown samples were calculated from the standard curve. In the case of IL-1β, TBS instead of PBS was used.

2.4.8 Flow Cytometry

Flow cytometry is a technique that allows cells in solution to be individually analysed by focusing a stream of cells through a laser, one cell at a time. Cell size and density can be determined by the nature of the scattered light. Specific cell surface markers can be targeted and analysed using monoclonal antibodies conjugated to flurochromes, which absorb and emit light of specific wavelengths. Excitation of the flurochromes is caused by absorption of laser light and the resultant scattered light is collected in photomultiplier tubes (PMT). These PMTs measure fluorescence intensity and side scatter, a measure of cell granularity and density. Cell size is detected at a photodiode in the direct path of the laser in the form of forward scatter, where it detects light intensity from the laser directly. This process allows for the analysis of individual cells within a mixture of multiple populations.

Cells were plated in 6-well plates at a concentration of $0.5x10^6$ cells/mL to a total volume of 2 mL per well. Cells were stimulated with SLPs as previously described with the addition of LPS as a positive control and media alone as a negative control. A cell scraper was used to gently detach cells from the well, and cells and media were collected in a 15 mL centrifuge tube. An equal amount of FCS was added, and cells were incubated at room temperature for 15 min.

Cells were then centrifuged at 1,200 rpm for 5 min and resuspended in FACS buffer (PBS/2%FCS/0.05%NaN_3). One mL of buffer was used per 2 million cells. 200 µL of each sample were transferred to a round bottom 96-well plate. Plates were centrifuged at 2,000 rpm for 5 min. Relevant antibody was added to each well at a concentration of 0.5 µg/1x10^6 cells. Antibody stock was diluted in FACS buffer as required. Treated cells were incubated at 4˚C for 30 min. Cells were then washed with 200 µL FACS buffer, and centrifuged at 2,000 rpm for 5 min. This was repeated three times. Cells were resuspended in FACS buffer and transferred to labelled FACS tubes for analysis.

2.5 *IN VIVO* TECHNIQUES

2.5.1 *C. difficile* infection of mice

C57BL/6J mice were infected with *C. difficile* using an antibiotic-induced model of mouse infection (Chen et al. 2008). Mice were treated for three days with an antibiotic mixture of kanamycin (400 μg/ml), gentamicin (35 μg/ml), colistin (850 U/ml), metronidazole (215 μg/ml) and vancomycin (45 μg/ml) in the drinking water. Mice were subsequently given autoclaved water. On day 5, mice were injected intraperitoneally with clindamycin (10 mg/kg). Mice were infected with 10^3 *C. difficile* spores on day six by oral gavage. Initial studies determined infection with 10^3 spores of *C. difficile* R13537 caused mild transient weight loss and diarrhoea in wild-type C57BL/6J strain mice. Animals were weighed daily and monitored for overt disease, including diarrhoea. Moribund animals with >15% loss in body weight were humanely killed. The cecum was harvested from uninfected and infected mice at days three and seven and the contents were removed for CFU counts.

2.5.2 CFU counts

The contents of cecum were recovered from infected and uninfected mice, weighed and stored frozen. Each sample of cecum material was thawed and homogenised in 1 mL PBS (pH 7.4) by vortex mixing in a 1.5 mL microcentrifuge tube. The suspension was serially diluted (10^{-1} to 10^{-4}) and 50 μL of each dilution was spread in duplicate onto quadrants of Brazier's CCEY plates (Lab M). Plates were incubated under anaerobic conditions at 37ºC for 30 hours. Colonies were counted and CFU/g determined for each sample.

2.5.3 Tissue preparation

Tissue samples were collected for histological staining, as well as RNA isolation. Squares of tissue from the cecum and the distal colon roughly 5 mm^3 were cut for the preparation of RNA. This tissue was stored in RA1 buffer until required. The remaining colon was then cut in half, and the distal colon tissue prepared for tissue staining. The colon was cut open along its length and cleaned in PBS. Using two pairs of forceps, the tissue was rolled over itself, creating a "swiss roll" (Moolenbeek & Ruitenberg 1981), for easily obtaining a cross-section of the whole length of tissue. The rolled tissue was embedded in OCT (optimal cooling temperature) compound and flash frozen in liquid nitrogen. The tissue was stored at -80°C until required for use.

2.5.4 Frozen tissue section preparation

A drop of OCT was placed on the metal mount. Frozen distal colon sample in OCT was carefully placed on the mount and more OCT was added around it to adhere it to the mount and left to dry. The mount was placed on the holder and set to 10µm thickness. Sample was positioned correctly with arrow buttons. The wheel was spun to begin cutting tissue. Once tissue was exposed, the thickness was set to 6 µm and samples were taken.

2.5.5 RNA isolation and quantification

All nucleic acid work was carried out in an RNase-free environment, to ensure the quality of RNA starting product. RNase degrades RNA; therefore its removal from the environment is essential. A dedicated RNA area was established, with specific gloves and only pipettes with filter tips being used. PCR grade water was used at all times to ensure quality of nucleic acids.

Frozen tissue samples were homogenised using a rotor-stator homogeniser in RA1 buffer. RNA was then isolated using the NucleoSpin® RNA II Total RNA Isolation Kit (Macherey-Nagel). Homogenised tissue in RA1 buffer (350 µL) was placed in a fresh tube and 3.5 µL of ß-mercaptoethanol (ß-ME) was added to the sample. The samples were vortexed vigorously. In some cases, the tissue remains viscous, and the sample was passed through a 20 G needle to reduce viscosity. The samples were placed in Nucleospin® Filters (violet ring) and centrifuged for 1 min at 11,000 x g. The filter was then discarded and 350 µL of 70% ethanol was added and mixed with the samples by pipetting up and down. The lysate was then loaded onto a NucleoSpin® RNA II Column (light blue ring) and centrifuged for 30 seconds at 11,000 x g. The membrane on the filter was then desalted by adding 350 µL Membrane Desalting Buffer (MDB) and centrifuged for 11,000 x g for 1 min. Next, the DNA was digested with DNase reaction mixture. Ninety-five µL was added to each sample and incubated at room temperature for 15 min. RA2 buffer (200 µL) was added to wash the silica membrane, and centrifuged for 30 seconds at 11,000 x g. RA3 buffer (600 µL) was added to the column and was centrifuged for 30 seconds at 11,000 x g. A further 250 µL of RA3 buffer was again added to the column and centrifuged for

2 min at 11,000 x g. Finally, RNase-free H_2O was pre-heated to 65ºC, and 60 µL was added to each sample to elute the RNA. The samples were centrifuged at 11,000 x g for one minute and the RNA was collected in sterile 1.5 mL RNase-free centrifuge tubes.

The RNA was then quantified using a NanoDrop Spectrophotometer. The absorbance from 230 nm to 600 nm was read for all RNA samples. Sample (1 µL) was loaded to give quantitative assessment and purity information. An A260/A280 ratio close to two is required for optimum RNA purity (due to the absorbance of uracil to thymidine).

Examining other absorbance spectra ratios, for example the A260/A230 ratio, can give an indication of the presence of contaminants in the sample. For example, phenol absorbance will lower the samples A260/A230 ratio. An A260/A230 ratio of 1.8-2.2 indicates non-contaminated nucleic acid.

2.5.6 Reverse Transcription-cDNA synthesis

Synthesis of complementary DNA (cDNA) from mRNA gives a DNA template for qPCR. Once quantified, the RNA was then reverse transcribed into cDNA for use in qPCR. 1 µg of RNA was used to synthesise cDNA. The components of the cDNA master mix are given in the table below. A final volume of 20 µL is required for each reaction. A total of 5.8 µL of master mix is required, therefore if 1 µg of RNA equates to less than 14.2 µL, the remaining volume is made up with RNase-free H_2O. Reactions were made up in RNase-free 0.2 mL tubes and

placed in the thermal cycler. The reaction is described in Table 2.5.1. cDNA was stored at -20˚C until required.

Table 2.5.1 Parameters for cDNA synthesis from RNA.

	Step 1	**Step 2**	**Step 3**	**Step 4**
Temp (˚C)	25	37	85	4
Time	10 min	120 min	5 min	∞

2.5.7 Quantitative Polymerase Chain Reaction

RNA was isolated from colonic tissue to examine gene expression by real-time quantitative Polymerase Chain Reaction (qPCR). This process uses primers to bind to a specific region of cDNA and amplify it exponentially using a heat-stable DNA polymerase. The sequences of all primers used are given in Appendix D. The amount of cDNA amplified is quantified in real time and allows quantification of the target gene. Quantification can be absolute (total copies) or relative. Relative quantification was primarily used in this study, where expression is normalised to a control gene. SYBR green intercalating dye non-specifically binds to double-stranded DNA, and when it forms a complex with qPCR-amplified DNA, it is excited by light of wavelength 492 nm and emits light of 520 nm. This allows for the detection of PCR product. The non-specific nature of SYBR green dye can result in the fluorescence of double-stranded contaminants, such as genomic DNA, and result in false positives or high background. The use of gene-specific probes can prevent this from happening. These probes are labelled with a reporter dye and quencher, and

exploit the 5' – 3' nuclease activity of DNA polymerase to cleave the probe and release the reporter dye from the quenching protein, resulting in signal. The cost of running a patent protected technique lessens its usability when screening large amounts of target genes.

The synthesised cDNA was used for all proceeding PCR reactions using specific primers for specific genes. Before gene analysis of cDNA was carried out, standard curves were obtained for each set of primers to ensure their quality. A 1:10 serial dilution of cDNA was made from neat to 10^{-4}. Using the primer of interest, a reaction was carried out on the serial dilution of cDNA. The log of the dilution factor was plotted against the C_q value to obtain a standard curve. The efficiency of the PCR reaction was determined with the slope of the standard curve and the equation below.

$$\textbf{Efficiency} = 10^{\wedge(-1/\text{Slope})} - 1 \times 100$$

An efficiency of 100% means the PCR product is doubling every cycle during the logarithmic phase of the reaction. Slopes ranging from -3.60 to -3.20 equate to efficiencies of 90-110%, along with an R^2 value of >0.95, were considered acceptable.

SYBR-Green based detection was used for the majority of qPCR reactions, as an initial screening of colonic tissue was required. Initially, the cDNA is heated to 95˚C to denature the DNA double helix into two separate strands. The mixture is then cooled to 55˚C to allow primers to anneal to the target sequences. The

temperature is raised to 72˚C, the optimal working temperature of Taq polymerase, which then copies the template sequence, forming a double stranded DNA sequence. This process is repeated for 40 cycles. The SYBR dye then binds to this newly formed double stranded product, and fluoresces at intensity directly proportional to the amount of new DNA generated. All samples were normalised to a normalisation gene, which is an endogenous control that maintains expression levels despite treatment. Two suitable normalisation genes were selected to ensure maximum accuracy of the qPCR assays. Ribosomal protein s18 and GusB were the normalisation genes selected. All samples were normalised to the geometric mean of the normalisation genes C_q values. Samples were compared using the comparative C_q method:

$$\mathbf{ddC_q = dC_q\ sample - dC_q\ reference}$$

dC_q sample is the normalised C_q value of the treated sample and dC_q reference is the normalised C_q value of the control sample. Primers were added at a final concentration of 500 mM, and ROX reporter dye was used an internal reference signal to account for non-PCR fluorescence occurring in individual wells. All samples were run on the ABI Prism 7500 under the following conditions:

Table 2.5.2 Parameters for qPCR reaction.

	Step 1	**Step 2**	**Step 3(a)**	**Step 3(b)**
Temp (˚C)	50	95	95	60
Time	2 min	10 min	15 sec*	1 min*
			* repeat for 40 cycles	

2.5.8 Haematoxylin & Eosin Staining

Tissue slides were mounted on the tissue holder and fixed in 75% acetone/alcohol for 5 min at room temperature. Slide were then washed in PBS for 5 min by dipping up and down, and placed in haematoxylin for 10 min to stain the tissue. Excess stain was then removed by placing the tissue under running tap water for 5 min, before dipping the slides in 1% acid alcohol for 30 seconds. The slides were again washed under tap water for 1 min and placed in 0.1% Sodium bicarbonate for 1 min. Slides were washed under tap water for 5 min and then rinsed in 95% ethanol for 10 dips. The tissue was then counterstained with eosin for 1 min. The sections were then dehydrated by dipping in 75% ethanol for 3 min, 95% ethanol for 3 min (x2), 100% ethanol for 3 min, and then finally in Histoclear for 3 min (x2). Slides were left to dry. 2 drops of DPX mountant (Sigma-Aldrich) were applied to coverslips and the slides were firmly pressed down on them. The slides were then stored at room temperature until ready for use.

2.5.9 Neutrophil/Macrophage/CD4+ T cell tissue staining

An HRP-DAB staining kit (R&D) was used to detect the presence of neutrophils in mouse colonic tissue. Tissue sections secured on slides were placed in a humidity chamber to prevent drying. The sample was covered in 1-3 drops of Peroxidase Blocking reagent for 5 min and then washed with PBS for 5 min. Serum Blocking Reagent was added to the tissue for 15 min. Slides were drained, but not rinsed, and incubated for 15 min with Avidin Blocking Reagent. Sections were then rinsed again with PBS and drained. Biotin blocking reagent was added and incubated for a further 15 min, before rinsing again in PBS. A 1/100 dilution

of primary antibody (Abcam, anti-NIMP-R14, 15µg/mL) was made and 100 µL was added to each tissue section. Samples were incubated for one hour at room temperature. The tissue was rinsed with PBS and washed 3 times (15 min/wash). Slides were then carefully drained to remove excess buffer. Biotinylated Secondary antibody was added to the samples and incubated for 45 min, and the tissue was rinsed with PBS and washed 3 times (15 min/wash) with the excess buffer then being removed. Samples were incubated for 30 min with Hanks Salt Solution with HRP (HSS-HRP), the tissue was rinsed with PBS and washed 3 times (2 min/wash) with the excess buffer then being removed. 150 µL of DAB chromogen was added to each tissue section and incubated for 15 min. Samples were rinsed with distilled water, and washed with fresh distilled water for 5 min. Excess water was removed and sections were counterstained with haematoxylin for 30 seconds, before being rinsed and washed with water. Slides were placed vertically on filter paper and mounted with coverslips using DPX mountant. Slides were then stored at room temperature until required.

Chapter 3

Analysis of sequence patterns of Surface Layer Proteins from different strains of *Clostridium difficile*

3.1 INTRODUCTION

3.1.1 Modern Synthesis of Evolution – Molecular Evolution:

Evolution is the change in inherited characteristics of living systems over time and through successive generations. The first to propose it were Darwin and Wallace, publishing a joint letter in 1858, detailing observed variation within species and a natural means of selection (Darwin & Wallace 1858). It was proposed that variation within a population would lead to individuals possessing traits better suited to the environment than others. These individuals with increased fitness would then be selected and not only would these individuals survive but they would contribute these advantageous traits to the next generation at a greater rate than those less fit. This idea was based on observations of morphological characteristics between individuals of a species. For example, different sub-species of finches living on the Galapagos displayed variation in beak morphology. In each case, the subspecies of finch possessed a beak that was well suited for accessing the main food source of that island. Darwin postulated that the finches of the Galapagos descended from a common ancestral species native to mainland Equador, and that the standing variation in the founder population provided the substrate for selection to act on individuals in the population of each island (Darwin 1859). This diversification into various ecological niches is known as an adaptive radiation event. The discovery of DNA as the material of inheritance in the early 20th century and the seminal work by Zuckerkandl and Pauling in the 1960's (Zuckerkandl & Pauling 1965a; Zuckerkandl & Pauling 1965b) marked the emergence of a new field of evolutionary study that incorporated the use of molecular data as well as phenotypic/morphological characters.

3.1.2 Mutation Rate, Genetic Drift and Effective Population Size:

The basis of molecular evolution lies in genetic mutations, where incorrect nucleotides added during DNA replication give rise to amino acid substitutions which can alter the 3D structure and therefore the function of the protein. Zuckerkandl and Pauling proposed the "Molecular Clock" to describe the steady accumulation of amino acid substitutions over time (as judged by the fossil record) (Zuckerkandl & Pauling 1965b). For the globin gene across a group of vertebrate species they recorded the number of amino acid substitutions and found a direct correlation between the number of substitutions and the amount of time that had elapsed since they last shared a common ancestor (Zuckerkandl & Pauling 1965b). The presence of a global molecular clock would have had distinct advantages for the modelling of protein change over time, for the extrapolation of divergence times in the absence of fossil data, and for the prediction of the trajectory of sequence change into the future (e.g. epidemiological implications). However, this hypothesis has since been disproven and there is no global molecular clock. Many recent studies show mutations occurring at variable rates across different species, different proteins, different regions of the same protein, and indeed different genomes in the same cell (Shuying Li, Dennis K. Pearl 2000; Gu & Li 1992; Yang & Nielsen 1998).

The accumulation of mutations, in the absence of selection, is known as random genetic drift. Under the Neutral Theory of evolution, genetic drift is the causative agent of evolutionary change (Kimura 1968). The rate at which these neutral mutations become fixed in a population is dependent on the effective population

size, or N_e (Wright 1938). The probability of a mutation becoming fixed (P_x) is given by **Equation 1**:

Equation 1: $P_x = ½(N_e)$

If N_e is large, then the rate of fixation of random mutations in the population will be small. A small N_e will increase the rate at which random mutations become fixed (Kimura 1968). In certain cases where population size is severely reduced, such as in the event of a population bottleneck, genetic variability within that population is reduced (Lynch & Hill 1986). Therefore any mutations that arise are more likely to rapidly become fixed in the population (See Figure 3.1). To accommodate the accumulation and acceptance of slightly deleterious mutations in populations, the Neutral theory was expanded to the "Nearly Neutral Theory of Molecular Evolution" (Ohta 1973; Ohta & Gillespie 1996), which is central to evolutionary theory today.

Figure 3.1

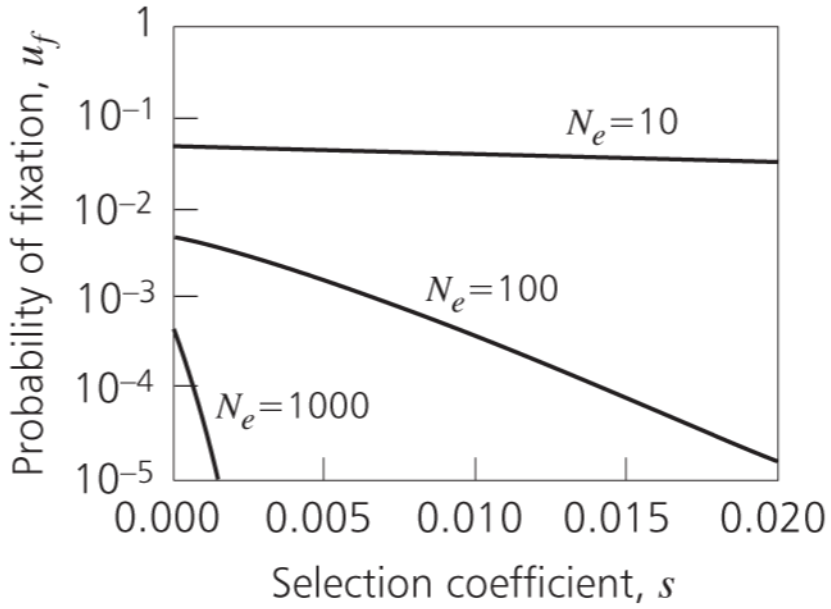


Figure 3.1 Fixation of deleterious mutations based on population size

The probability of fixation of deleterious alleles (*Uf*) is affected by the strength of selection against *Uf* (s) and the effective population size (*Ne*). The allele is initially present at frequency 1/(2*Ne*). Deleterious mutations will fix more readily in smaller populations (*Ne* = 10) compared to larger populations (*Ne* = 1000). Taken from (Whitlock & Bürger 2004).

3.1.3 Natural Selection:

Natural selection acts on individuals of a population. Any substitution that increases the fitness and reproductive success of an individual in a given environment or set of conditions - will result in increased propagation of those traits. Over time, and in the absence of a change in selective pressure such as environmental shift, the genetic mutation that gave rise to the trait will spread to fixation in the population.

Selective pressures can be classified as positive (advantageous to the host) or purifying (disadvantageous). Purifying selection works to remove deleterious mutations from a population. It is generally held that the majority of coding positions are evolving under purifying selection (Hughes 1999). The importance of amino acid sequence on protein structure, and therefore function, places an evolutionary restraint on the sequences, limiting the number of viable amino acid substitutions (Peterson et al. 2009). However, when N_e is low, and genetic variation is reduced, slightly deleterious mutations can become fixed due to genetic drift (Figure 3.1) (Whitlock and Bürger 2004). Positive selection occurs in cases where an amino acid substitution increases the fitness of an organism, by improving the function of the gene on which the mutation occurred. These mutations can only occur at specific points on a gene (Nielsen et al. 2005) This increased fitness will allow the individual to propagate its genes, and the population retains these advantageous mutations in a given set of environmental conditions. This has been observed in many mammalian genes to date (Kosiol et al. 2008).

There has long been debate as to the role of natural selection as the main driving force in molecular evolution (Nei 2005). Much of the literature describes the majority of amino acid substitutions as neutral or nearly neutral, and "positively selected" sites are, in fact, the result of the random fixation of amino acid altering substitutions due to population size and genetic drift (Ohta & Gillespie 1996; Ohta 1973; Kimura 1968). King & Jukes (1969) proposed that the main function of natural selection was to eliminate deleterious mutations from a population, while advantageous mutations accumulated by the random fixation of neutral or nearly neutral mutations. Evidence for positive selection does exist however. Many examples are seen in immune-specific genes (Tanaka & Nei 1989; Su & Nei 2001). For example, positive selection was detected in the peptide binding sites of the MHC Class I molecule, with purifying selection detected in the non-peptide binding sites (Hughes & Nei 1989). As this molecule is essential is distinguishing self-peptides from non-self peptides, greater variation in the peptide binding sites will infer an advantage by allowing immune cells to display a larger range of pathogenic peptides.

3.1.4 Relationship between Sequence, Structure and Function of a Protein:

Degeneracy exists in the genetic code, with multiple codons coding for each amino acid. Many mutations at the DNA level therefore will not result in an amino acid substitution, and these are known as silent mutations. When silent mutations occur in the protein coding sequence of a gene they are referred to as synonymous substitutions. Mutations that do give rise to an amino acid substitution are known as non-synonymous substitutions.

The specific 3D conformation of a protein determined by its sequence is of extreme importance to its function. It has been shown that certain non-synonymous substitutions have a measurable effect on the overall structure and function of the protein (Peterson et al. 2009; Loughran et al. 2012). As proteins tend to have very specific functions, the majority of substitutions will result in a partial or complete loss of function (MacArthur et al. 2012). This structure-function relationship gives rise to an evolutionary constraint where much of the protein sequence cannot change is the protein is to remain functional (Peterson et al. 2009). Therefore many non-synonymous substitutions will be under purifying selection and will be removed from the population. For example a non-synonymous point mutation in the haemoglobin beta gene will result in sickle cell anaemia (Wishner et al. 1975).

The level of positive selection in any given genome is debated and from comparative genomic studies in a wide range of species it has been estimated that between 20% and 40% of protein coding genes show signatures of positive selection (Smith & Eyre-Walker 2002; Kosiol et al. 2008). Recent publications have shown a definitive link between positive selection as predicted from protein sequences and protein functional shift (Moury & Simon 2011; Loughran et al. 2012; Levasseur et al. 2006). Loughran *et al* clearly show positive selection resulting in protein functional shift through site-directed mutagenesis on positively selected sites in human myeloperoxidase (MPO). They show that the novel chlorination activity of MPO was obliterated on mutating the positively selected residues (Loughran et al. 2012). This study is particularly salient to this chapter as the models employed by Loughran *et al* (Loughran et al. 2012) to

detect selective pressure heterogeneity are identical to those used in this chapter. Other studies have also shown this clear link between molecular predictions of positive selection and protein functional shift, including the study of the fungal lipase/feruloyl esterase A family involving site-directed mutagenesis on sites under positive selection showing a loss of function phenotype (Levasseur et al. 2006). (Moury & Simon 2011) also found a clear link between positive selection and functional shift in Potato Virus Y (PVY), where mutation of a specific codon resulted in increased virus accumulation in the host. Therefore, protein functional shift is synonymous with the signature for positive selection in coding DNA sequences (Hughes 1999; Loughran et al. 2012; Levasseur et al. 2006).

3.1.5 Measuring Selective Pressure Heterogeneity:

Selective pressure heterogeneity is estimated using the ratio of the rate of non-synonymous substitutions per non-synonymous site (Dn) to synonymous substitutions per synonymous site (Ds), i.e., Dn/Ds or ω as it will be referred to throughout this chapter. An $\omega = 1$ indicates neutral evolution, $\omega < 1$ indicates purifying selective pressure, and, $\omega > 1$ indicates positive selection. This method of estimation of selective pressure is based on the assumption that the rate of synonymous substitution acts as a proxy for the rate of neutral evolution or genetic drift as these sites are undetectable by natural selection. The assumption that Ds is a proxy for genetic drift has been questioned (Du et al. 2014; Dimitrieva & Anisimova 2014) and evidence to the contrary has been proposed from a comparative genomic analysis of the *BRCA1* gene (Hurst & Pál 2001) . In this example an elevated ω value was shown to result from purifying selection on silent sites (Hurst & Pál 2001). Using sliding window analysis on mouse-rat and

human-dog comparisons of the *BRCA1* gene, a particularly low Ds was observed to coincide with sites where $\omega > 1$. Therefore it is possible that an estimate of $\omega > 1$ may not be indicative of true positive selection, but may result from a decrease in Ds as a result of purifying selection on silent sites (Hughes 2007). Different sequence alignment methods can give different outputs, impacting any reconstructed phylogeny and subsequent positive selection analysis. When testing for positive selection, alignment methods with a higher tendency to place non-homologous together were found to be error prone, resulting in false positives (Blackburne & Whelan 2013). When using Codeml (Yang 2007; Yang 1998), a small dataset, high sequence similarity, low sequence length can all result in false positives (Anisimova et al. 2001; Anisimova et al. 2002). Methods of phylogeny construction, such as neighbouring-joining (Saitou & Nei 1987), are also much less accurate than Bayesian inference, and can therefore impact on positive selection results (Kuhner & Felsenstein 1994). Avoidance of such methods will minimise the risk of obtaining false positives.

3.1.6 Methods for Assessing Selective Pressure Heterogeneity:

Initial methods to detect positive selection included averaging methods for estimating Dn/Ds for an entire alignment (Nei & Gojobori 1986; Li et al. 1985) and sliding window approaches analysing overlapping regions of the protein thereby accounting for variation in mutational rates across regions of the protein. Both of these approaches have subsequently proven to be misleading in their estimates of the selective pressure at work (Schmid & Yang 2008). For example, the sliding window approach involves calculating Dn/Ds ratios across a small section, or window, of the alignment (Schmid & Yang 2008). This window then

advances by a predefined window shift size, and the Dn and Ds values are recorded in each window across the entire protein (Fares et al. 2002). While simple in terms of underlying concept and computational requirements, many analyses using these approaches have proven over simplistic averaging over all sites or all lineages (Endo et al. 1996; Fares et al. 2002). Indeed the issue with the sliding window approach is that is it inconsistent, and has been shown to generate artifactual trends of synonymous and non-synonymous rate variation (Schmid & Yang 2008).

It is clear that modelling a simple Dn/Ds ratio over the entire protein, or over all lineages in the alignment, is not biologically realistic and can lead to underestimation from over simplification of the process of molecular evolution. Given the heterogeneity in mutational rates and selective pressures across proteins and lineages and given the lack of precision, accuracy and concern over statistical robustness provided by averaging methods and sliding window approaches, we have employed a suite of codon models of evolution modelled in a maximum likelihood framework (Yang 2007). These models allow for variation in selective pressure across sites and across lineages (Goldman & Yang 1994; Yang & Nielsen 1998; Yang & Nielsen 2002). The various models employed are summarized below and are detailed in the materials and methods chapter in more detail (Chapter 2). The models are nested and as such they can be compared to one another (under strict guidelines) using LRTs (Anisimova et al. 2001; Yang 1998). These LRTs compare nested models of evolution and allow for variable Dn/Ds ratios among sites in a sequence ("site-specific models"), along branches in a phylogenetic tree ("branch models"), or a

combination of both ("site-branch models"). Each of the models differs in complexity by the addition of extra parameters. Each model allows for a number of classes of sites in the sequences, for example, the nearly neutral model M1a allows two site classes where $\omega = 0$ and $\omega = 1$. The model that best fits the data (i.e. best describes the observed pattern of substitution in the sequences) is identified by the LRT. In general, one model is constrained so that $\omega \leq 1$. Another model will allow at least one class of sites where ω is estimated from the data. Under these circumstances, if $\omega > 1$, and following the appropriate LRT there is a statistically significant increase in likelihood value relative to the simpler model, then the simpler model is rejected and positive selection is invoked (statistical significance is estimated using the chi-squared distribution). The significance of the difference in likelihood score between the models (Δln) can be tested by comparing 2 Δln with χ^2_{υ}, where υ is equal to the degrees of freedom (df). The df corresponds to the number of free parameters between the models being compared.

3.1.7 Positive selective pressure as a predictor of functional change between *C. difficile* subtypes.

In the case of *C. difficile* infection (CDI), the severity of disease, and observed symptoms, can vary greatly, depending on the strain causing the infection (Dawson et al. 2009). Particular strains, such as ribotypes 027 and 078, have been associated with more severe disease, and as a result have been dubbed 'hypervirulent' (Goorhuis et al. 2008; Marsh et al. 2012; Clements et al. 2010). These strains have also been associated with higher rates of recurrent infection (Marsh et al. 2012). The SLPs, encoded by the *slpA* gene, exhibit variability

between strains (McCoubrey & Poxton 2001) We have previously shown the ability of SLPs from RT 001 to induce an inflammatory response *in vitro* (Ryan et al. 2011; Collins et al. 2014). We set out to elucidate if patterns of substitution within the SLPs contributed to the fitness of the pathogen as measured by disease/infection profile of the strain.

In patients infected with *C. difficile*, recovery usually occurs within seven days, with the host immune response successfully clearing the pathogen (Kachrimanidou & Malisiovas 2011). An efficient host response relies on recognition of the pathogen, and this may be achieved through specific pathogen recognition signatures on the SLP sequence. We know from a previous study of *C. difficile* RT 001 that SLPs activate the immune system through TLR4 (Ryan et al. 2011). Given the variability observed in the SLP coding sequence between ribotypes it is possible that the TLR4 affinity for SLP may also vary across ribotypes. A similar relationship between SLP sequence and immune recognition has been observed by Thompson (2002). If the host response is initiated through recognition of a specific protein signature on the bacterial SLP, then altering the sequence of this protein may impact upon recognition of the pathogen. By determining if the genes coding for SLPs from different strains are undergoing positive selection, we may gain further insight into the pathogenic role of SLPs in infection**.**

By identifying specific proteins, or regions within proteins, under positive selection in *C. difficile* we are pinpointing the specific amino acid adaptations that enable the organism to better survive its environment. Our hypothesis is that

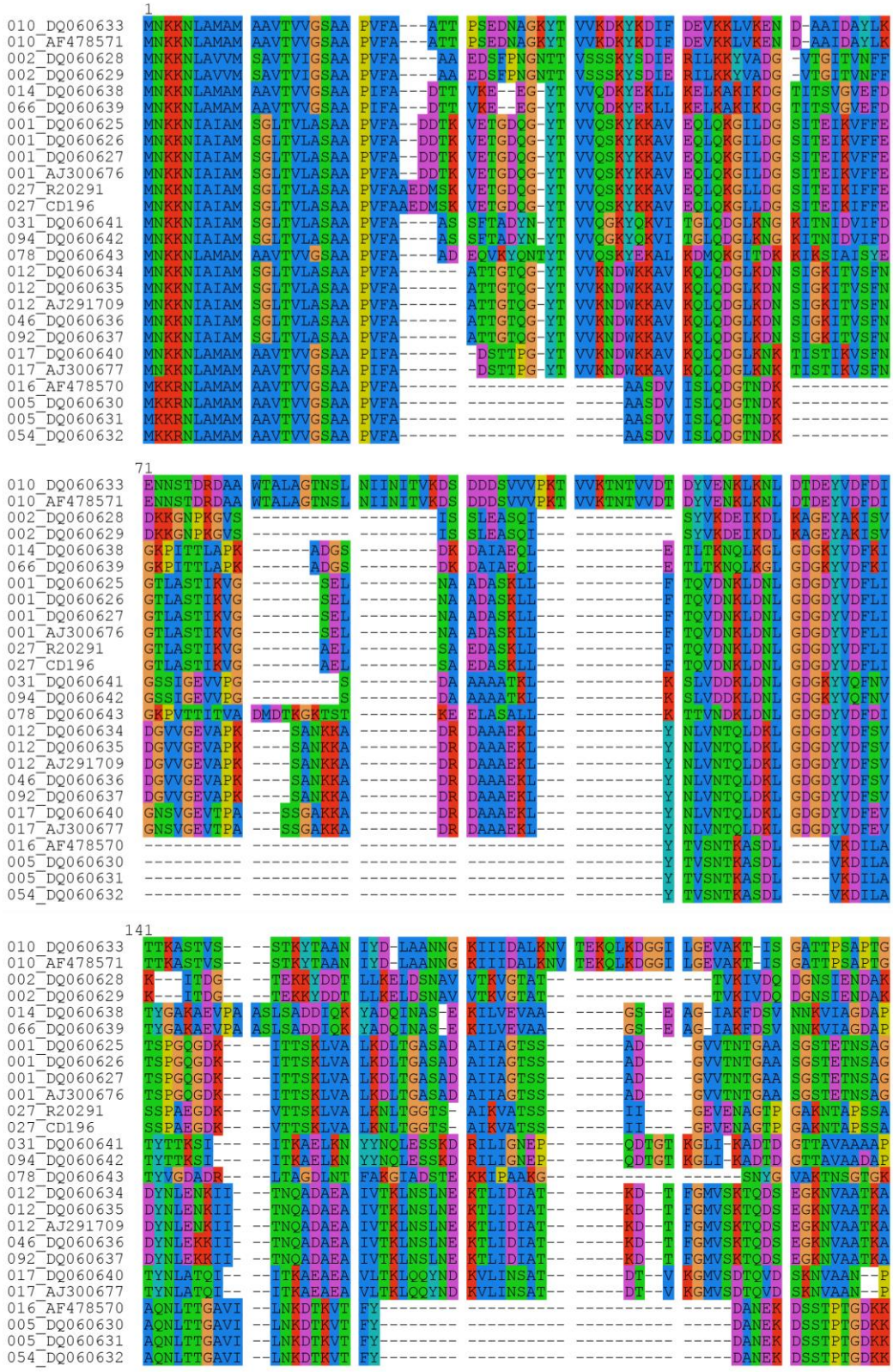
the variation in the SLPs from *C. difficile* is not solely due to genetic drift, and that this variation is increasing the fitness of the organism. As we know SLPs activate the immune system to induce an inflammatory response, we propose that positive selection of SLPs from specific strains will modulate the recognition and response to the pathogen. Therefore SLP sequence will play a role in disease severity. To test this, we first compile a dataset of *slpA* sequences from different strains of *C. difficile*, align these sequences and construct a phylogenetic gene tree, after which we test from signatures of positive selection in the *slpA* genes. If our hypothesis is true, we will expect to detect positive selection in the more virulent strains of *C. difficile*, suggesting a role for slpA sequence in disease severity. If it is false, we would not expect to detect positive selection in these strains, with sequence variation being a product of random genetic drift.

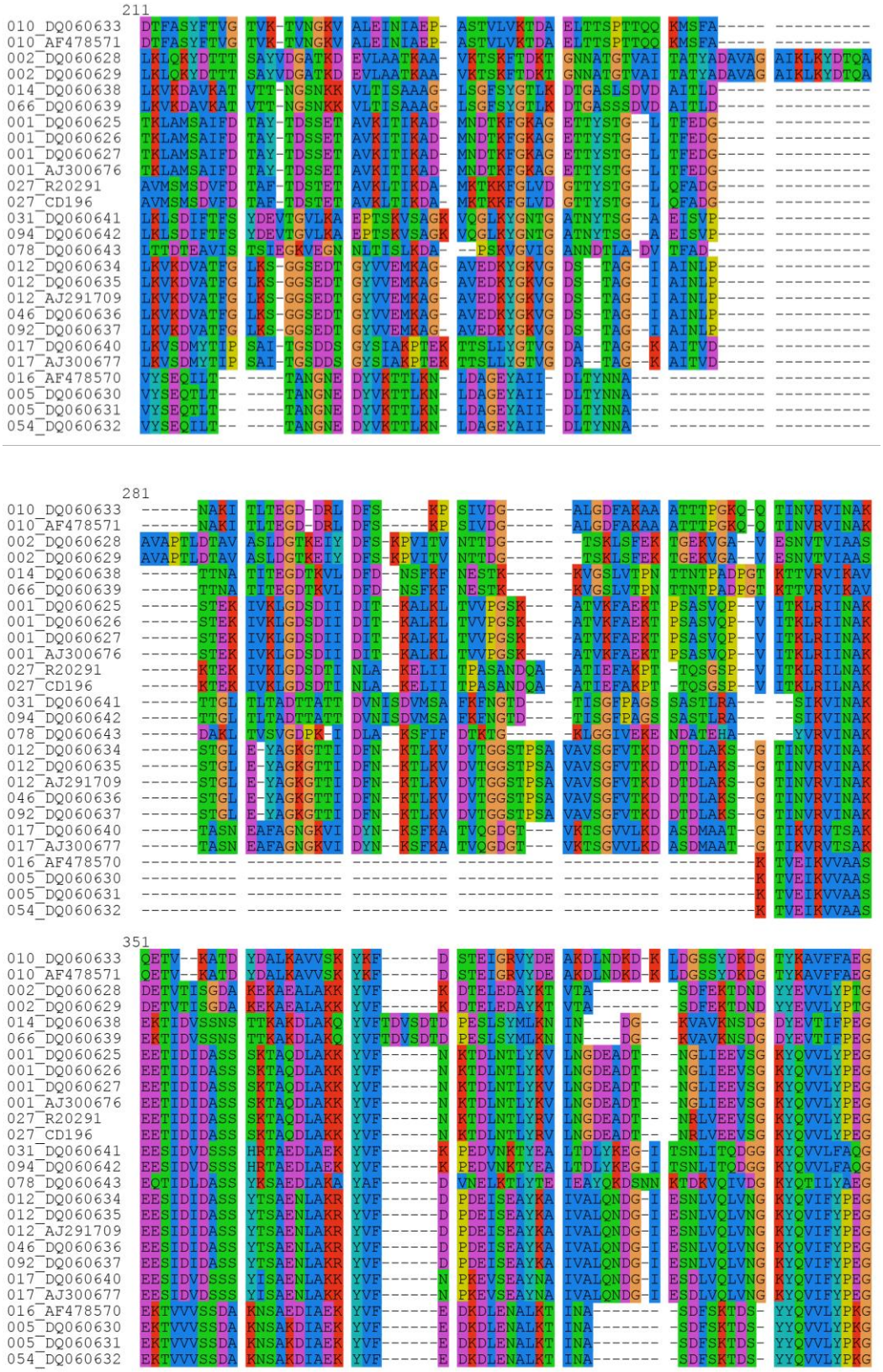
3.2 RESULTS

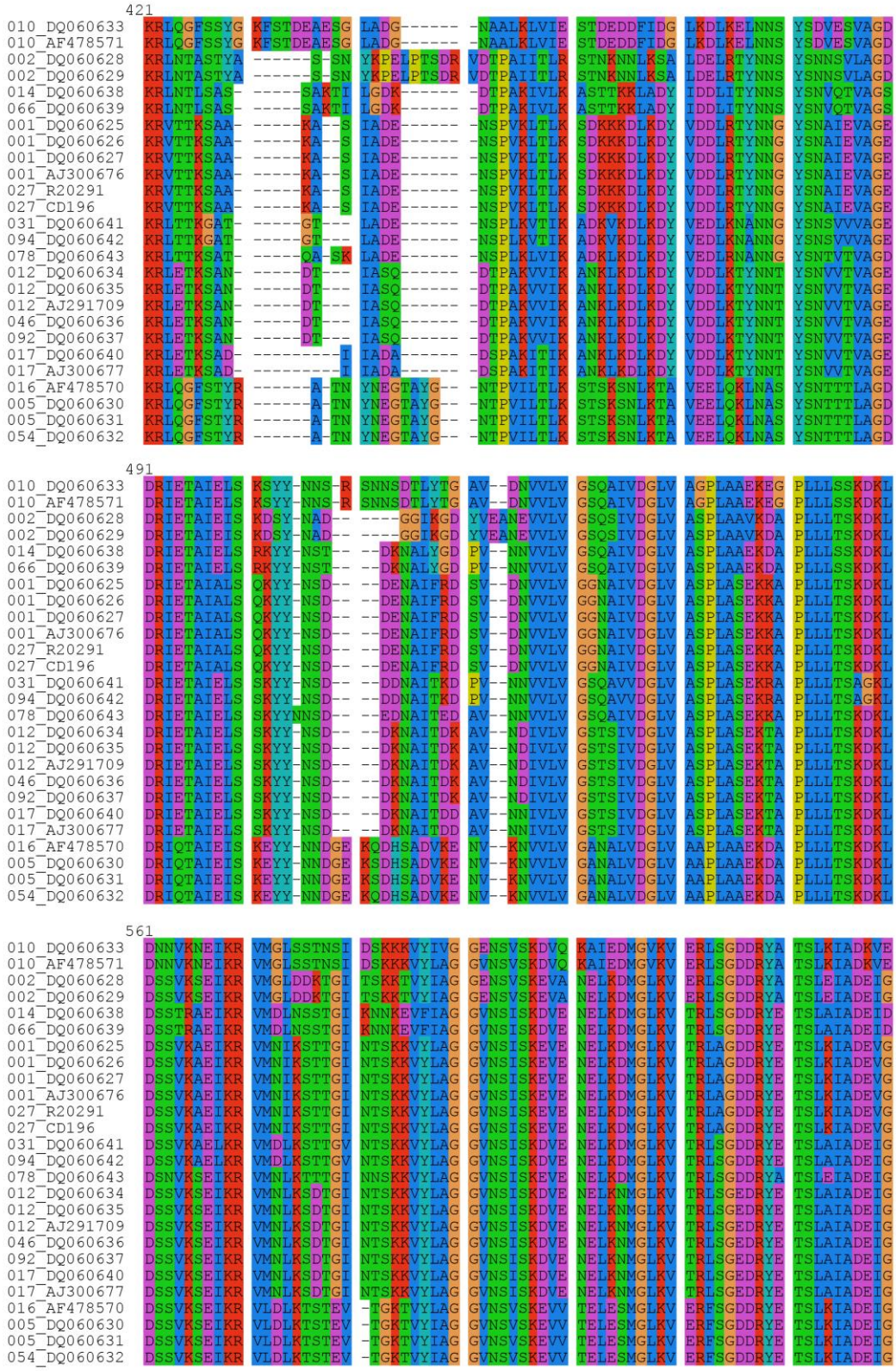
3.2.1 SlpA Multiple Sequence Alignment

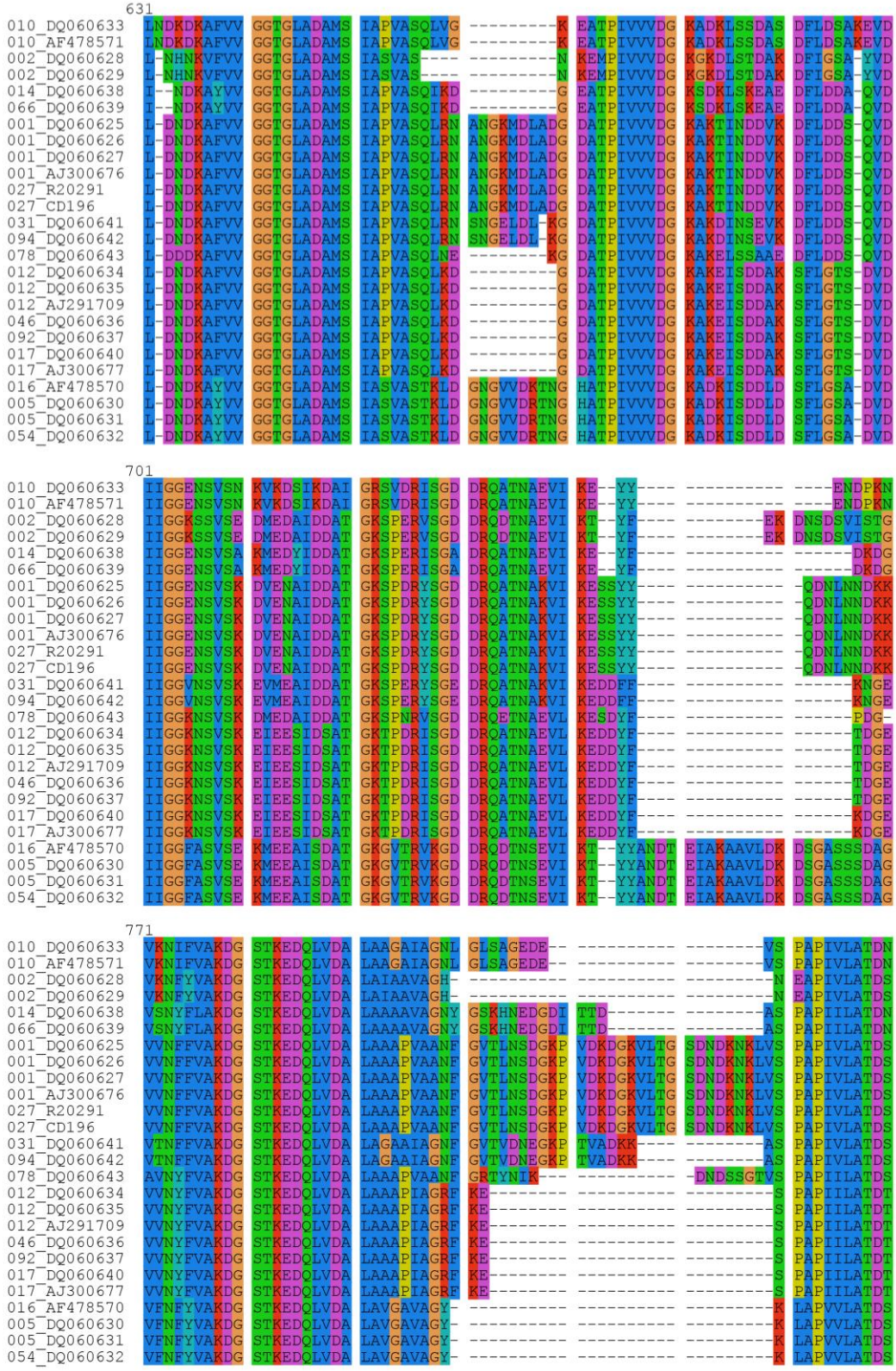
Sequences used in the analysis are detailed previously in Table 2.2.1. In some cases, there are multiple identical representatives for the same ribotype. The multiple sequence alignment of the *slpA* gene resulted in 2652 aligned nucleotide positions (884 codon positions). The first 24 residues in the alignment appear highly conserved, representing the signal peptide involved in transport of the SLPs to the surface of the cell, as seen in Figure 3.1. The next 300 residues represent the LMW protein, and a high level of variability is seen, with many indel events. In comparison, positions 350 to 430 in the alignment are more highly conserved and represent the HMW SLP-binding domain. Region 450 – 650 represents a highly conserved region of the HMW subunit, while region 750 – 800 displays a more variable region of the HMW SLP. Position 27 to 430 represents the LMW portion of slpA and 431 to 883 represent the HMW SLP, Figure 3.1 displays the full alignment.

Figure 3.1




211
010_DQ060633
010_AF478571
002_DQ060628
002_DQ060629
014_DQ060638
066_DQ060639
001_DQ060625
001_DQ060626
001_DQ060627
001_AJ300676
027_R20291
027_CD196
031_DQ060641
094_DQ060642
078_DQ060643
012_DQ060634
012_DQ060635
012_AJ291709
046_DQ060636
092_DQ060637
017_DQ060640
017_AJ300677
016_AF478570
005_DQ060630
005_DQ060631
054_DQ060632
281
351


421
010_DQ060633 KRLQGFSSYG KFSTDEAESG LADG------ -NAALKLVIE STDEDDFIDG LKDLKELNNS YSDVESVAGD
010_AF478571 KRLQGFSSYG KFSTDEAESG LADG------ -NAALKLVIE STDEDDFIDG LKDLKELNNS YSDVESVAGD
002_DQ060628 KRLNTASTYA ------S-SN YKPELPTSDR VDTPAIITLR STNKNNLKSA LDELRTYNNS YSNNSVLAGD
002_DQ060629 KRLNTASTYA ------S-SN YKPELPTSDR VDTPAIITLR STNKNNLKSA LDELRTYNNS YSNNSVLAGD
014_DQ060638 KRLNTLSAS- -----SAKTI LGDK------ -DTPAKIVLK ASTTKKLADY IDDLITYNNS YSNVQTVAGS
066_DQ060639 KRLNTLSAS- -----SAKTI LGDK------ -DTPAKIVLK ASTTKKLADY IDDLITYNNS YSNVQTVAGS
001_DQ060625 KRVTTKSAA- -----KA--S IADE------ -NSPVKLTLK SDKKKDLKDY VDDLRTYNNG YSNAIEVAGE
001_DQ060626 KRVTTKSAA- -----KA--S IADE------ -NSPVKLTLK SDKKKDLKDY VDDLRTYNNG YSNAIEVAGE
001_DQ060627 KRVTTKSAA- -----KA--S IADE------ -NSPVKLTLK SDKKKDLKDY VDDLRTYNNG YSNAIEVAGE
001_AJ300676 KRVTTKSAA- -----KA--S IADE------ -NSPVKLTLK SDKKKDLKDY VDDLRTYNNG YSNAIEVAGE
027_R20291 KRVTTKSAA- -----KA--S IADE------ -NSPVKLTLK SDKKKDLKDY VDDLRTYNNG YSNAIEVAGE
027_CD196 KRVTTKSAA- -----KA--S IADE------ -NSPVKLTLK SDKKKDLKDY VDDLRTYNNG YSNAIEVAGE
031_DQ060641 KRLTTKGAT- -----GT--- LADE------ -NSPLKVTIK ADKVKDLKDY VEDLKNANNG YSNSVVVAGE
094_DQ060642 KRLTTKGAT- -----GT--- LADE------ -NSPLKVTIK ADKVKDLKDY VEDLKNANNG YSNSVVVAGE
078_DQ060643 KRLTTKSAT- -----QA-SK LADE------ -NSPLKLVIK ADKLKDLKDY VEDLRNANNG YSNTVTVAGD
012_DQ060634 KRLETKSAN- -----DT--- IASQ------ -DTPAKVVIK ANKLKDLKDY VDDLKTYNNT YSNVVTVAGE
012_DQ060635 KRLETKSAN- -----DT--- IASQ------ -DTPAKVVIK ANKLKDLKDY VDDLKTYNNT YSNVVTVAGE
012_AJ291709 KRLETKSAN- -----DT--- IASQ------ -DTPAKVVIK ANKLKDLKDY VDDLKTYNNT YSNVVTVAGE
046_DQ060636 KRLETKSAN- -----DT--- IASQ------ -DTPAKVVIK ANKLKDLKDY VDDLKTYNNT YSNVVTVAGE
092_DQ060637 KRLETKSAN- -----DT--- IASQ------ -DTPAKVVIK ANKLKDLKDY VDDLKTYNNT YSNVVTVAGE
017_DQ060640 KRLETKSAD- ---------I IADA------ -DSPAKITIK ANKLKDLKDY VDDLKTYNNT YSNVVTVAGE
017_AJ300677 KRLETKSAD- ---------I IADA------ -DSPAKITIK ANKLKDLKDY VDDLKTYNNT YSNVVTVAGE
016_AF478570 KRLQGFSTYR ------A-TN YNEGTAYG-- -NTPVILTLK STSKSNLKTA VEELQKLNAS YSNTTTLAGD
005_DQ060630 KRLQGFSTYR ------A-TN YNEGTAYG-- -NTPVILTLK STSKSNLKTA VEELQKLNAS YSNTTTLAGD
005_DQ060631 KRLQGFSTYR ------A-TN YNEGTAYG-- -NTPVILTLK STSKSNLKTA VEELQKLNAS YSNTTTLAGD
054_DQ060632 KRLQGFSTYR ------A-TN YNEGTAYG-- -NTPVILTLK STSKSNLKTA VEELQKLNAS YSNTTTLAGD
491
010_DQ060633 DRIETAIELS KSYY-NNS-R SNNSDTLYTG AV--DNVVLV GSQAIVDGLV AGPLAAEKEG PLLLSSKDKL
010_AF478571 DRIETAIELS KSYY-NNS-R SNNSDTLYTG AV--DNVVLV GSQAIVDGLV AGPLAAEKEG PLLLSSKDKL
002_DQ060628 DRIETAIEIS KDSY-NAD-- ----GGIKGD YVEANEVVLV GSQSIVDGLV ASPLAAVKDA PLLLTSKDKL
002_DQ060629 DRIETAIEIS KDSY-NAD-- ----GGIKGD YVEANEVVLV GSQSIVDGLV ASPLAAVKDA PLLLTSKDKL
014_DQ060638 DRIETAIELS RKYY-NST-- --DKNALYGD PV--NNVVLV GSQAIVDGLV ASPLAAEKDA PLLLSSKDKL
066_DQ060639 DRIETAIELS RKYY-NST-- --DKNALYGD PV--NNVVLV GSQAIVDGLV ASPLAAEKDA PLLLSSKDKL
001_DQ060625 DRIETAIALS QKYY-NSD-- --DENAIFRD SV--DNVVLV GGNAIVDGLV ASPLASEKKA PLLLTSKDKL
001_DQ060626 DRIETAIALS QKYY-NSD-- --DENAIFRD SV--DNVVLV GGNAIVDGLV ASPLASEKKA PLLLTSKDKL
001_DQ060627 DRIETAIALS QKYY-NSD-- --DENAIFRD SV--DNVVLV GGNAIVDGLV ASPLASEKKA PLLLTSKDKL
001_AJ300676 DRIETAIALS QKYY-NSD-- --DENAIFRD SV--DNVVLV GGNAIVDGLV ASPLASEKKA PLLLTSKDKL
027_R20291 DRIETAIALS QKYY-NSD-- --DENAIFRD SV--DNVVLV GGNAIVDGLV ASPLASEKKA PLLLTSKDKL
027_CD196 DRIETAIALS QKYY-NSD-- --DENAIFRD SV--DNVVLV GGNAIVDGLV ASPLASEKKA PLLLTSKDKL
031_DQ060641 DRIETAIELS SKYY-NSD-- --DDNAITKD PV--NNVVLV GSQAVVDGLV ASPLASEKRA PLLLTSAGKL
094_DQ060642 DRIETAIELS SKYY-NSD-- --DDNAITKD PV--NNVVLV GSQAVVDGLV ASPLASEKRA PLLLTSAGKL
078_DQ060643 DRIETAIELS SKYYNNSD-- ---EDNAITED AV--NNVVLV GSQAIVDGLV ASPLASEKKA PLLLTSKDKL
012_DQ060634 DRIETAIELS SKYY-NSD-- --DKNAITDK AV--NDIVLV GSTSIVDGLV ASPLASEKTA PLLLTSKDKL
012_DQ060635 DRIETAIELS SKYY-NSD-- --DKNAITDK AV--NDIVLV GSTSIVDGLV ASPLASEKTA PLLLTSKDKL
012_AJ291709 DRIETAIELS SKYY-NSD-- --DKNAITDK AV--NDIVLV GSTSIVDGLV ASPLASEKTA PLLLTSKDKL
046_DQ060636 DRIETAIELS SKYY-NSD-- --DKNAITDK AV--NDIVLV GSTSIVDGLV ASPLASEKTA PLLLTSKDKL
092_DQ060637 DRIETAIELS SKYY-NSD-- --DKNAITDK AV--NDIVLV GSTSIVDGLV ASPLASEKTA PLLLTSKDKL
017_DQ060640 DRIETAIELS SKYY-NSD-- --DKNAITDD AV--NNIVLV GSTSIVDGLV ASPLASEKTA PLLLTSKDKL
017_AJ300677 DRIETAIELS SKYY-NSD-- --DKNAITDD AV--NNIVLV GSTSIVDGLV ASPLASEKTA PLLLTSKDKL
016_AF478570 DRIQTAIEIS KEYY-NNDGE KQDHSADVKE NV--KNVVLV GANALVDGLV AAPLAAEKDA PLLLTSKDKL
005_DQ060630 DRIQTAIEIS KEYY-NNDGE KSDHSADVKE NV--KNVVLV GANALVDGLV AAPLAAEKDA PLLLTSKDKL
005_DQ060631 DRIQTAIEIS KEYY-NNDGE KSDHSADVKE NV--KNVVLV GANALVDGLV AAPLAAEKDA PLLLTSKDKL
054_DQ060632 DRIQTAIEIS KEYY-NNDGE KQDHSADVKE NV--KNVVLV GANALVDGLV AAPLAAEKDA PLLLTSKDKL
561
010_DQ060633 DNNVKNEIKR VMGLSSTNSI DSKKKVYIVG GENSVSKDVQ KAIEDMGVKV ERLSGDDRYA TSLKIADKVE
010_AF478571 DNNVKNEIKR VMGLSSTNSI DSKKKVYLAG GVNSVSKDVQ KAIEDMGVKV ERLSGDDRYA TSLKIADKVE
002_DQ060628 DSSVKSEIKR VMGLDDKTGI TSKKTVYIAG GENSVSKEVA NELKDMGLKV ERLSGDDRYA TSLEIADEIG
002_DQ060629 DSSVKSEIKR VMGLDDKTGI TSKKTVYIAG GENSVSKEVA NELKDMGLKV ERLSGDDRYA TSLEIADEIG
014_DQ060638 DSSTRAEIKR VMDLNSSTGI KNNKEVFIAG GVNSISKDVE NELKDMGLKV TRLSGDDRYE TSLAIADEID
066_DQ060639 DSSTRAEIKR VMDLNSSTGI KNNKEVFIAG GVNSISKDVE NELKDMGLKV TRLSGDDRYE TSLAIADEID
001_DQ060625 DSSVKAEIKR VMNIKSTTGI NTSKKVYLAG GVNSISKEVE NELKDMGLKV TRLAGDDRYE TSLKIADEVG
001_DQ060626 DSSVKAEIKR VMNIKSTTGI NTSKKVYLAG GVNSISKEVE NELKDMGLKV TRLAGDDRYE TSLKIADEVG
001_DQ060627 DSSVKAEIKR VMNIKSTTGI NTSKKVYLAG GVNSISKEVE NELKDMGLKV TRLAGDDRYE TSLKIADEVG
001_AJ300676 DSSVKAEIKR VMNIKSTTGI NTSKKVYLAG GVNSISKEVE NELKDMGLKV TRLAGDDRYE TSLKIADEVG
027_R20291 DSSVKAEIKR VMNIKSTTGI NTSKKVYLAG GVNSISKEVE NELKDMGLKV TRLAGDDRYE TSLKIADEVG
027_CD196 DSSVKAEIKR VMNIKSTTGI NTSKKVYLAG GVNSISKEVE NELKDMGLKV TRLAGDDRYE TSLKIADEVG
031_DQ060641 DSSVKAELKR VMDLKSTTGV NTSKKVYLAG GVNSISKDVE NELKDMGLKV TRLSGDDRYE TSLAIADEIG
094_DQ060642 DSSVKAELKR VMDLKSTTGV NTSKKVYLAG GVNSISKDVE NELKDMGLKV TRLSGDDRYE TSLAIADEIG
078_DQ060643 DSNVKSEIKR VMNLKTTTGI NNSKKVYLAG GVNSISKEVE NELKDMGLKV TRLSGDDRYA TSLEIADEIG
012_DQ060634 DSSVKSEIKR VMNLKSDTGI NTSKKVYLAG GVNSISKDVE NELKNMGLKV TRLSGEDRYE TSLAIADEIG
012_DQ060635 DSSVKSEIKR VMNLKSDTGI NTSKKVYLAG GVNSISKDVE NELKNMGLKV TRLSGEDRYE TSLAIADEIG
012_AJ291709 DSSVKSEIKR VMNLKSDTGI NTSKKVYLAG GVNSISKDVE NELKNMGLKV TRLSGEDRYE TSLAIADEIG
046_DQ060636 DSSVKSEIKR VMNLKSDTGI NTSKKVYLAG GVNSISKDVE NELKNMGLKV TRLSGEDRYE TSLAIADEIG
092_DQ060637 DSSVKSEIKR VMNLKSDTGI NTSKKVYLAG GVNSISKDVE NELKNMGLKV TRLSGEDRYE TSLAIADEIG
017_DQ060640 DSSVKSEIKR VMNLKSDTGI NTSKKVYLAG GVNSISKDVE NELKNMGLKV TRLSGEDRYE TSLAIADEIG
017_AJ300677 DSSVKSEIKR VMNLKSDTGI NTSKKVYLAG GVNSISKDVE NELKNMGLKV TRLSGEDRYE TSLAIADEIG
016_AF478570 DSSVKSEIKR VLDLKTSTEV -TGKTVYLAG GVNSVSKEVV TELESMGLKV ERFSGDDRYE TSLKIADEIG
005_DQ060630 DSSVKSEIKR VLDLKTSTEV -TGKTVYIAG GVNSVSKEVV TELESMGLKV ERFSGDDRYE TSLKIADEIG
005_DQ060631 DSSVKSEIKR VLDLKTSTEV -TGKTVYIAG GVNSVSKEVV TELESMGLKV ERFSGDDRYE TSLKIADEIG
054_DQ060632 DSSVKSEIKR VLDLKTSTEV -TGKTVYIAG GVNSVSKEVV TELESMGLKV ERFSGDDRYE TSLKIADEIG


631
010_DQ060633
010_AF478571
002_DQ060628
002_DQ060629
014_DQ060638
066_DQ060639
001_DQ060625
001_DQ060626
001_DQ060627
001_AJ300676
027_R20291
027_CD196
031_DQ060641
094_DQ060642
078_DQ060643
012_DQ060634
012_DQ060635
012_AJ291709
046_DQ060636
092_DQ060637
017_DQ060640
017_AJ300677
016_AF478570
005_DQ060630
005_DQ060631
054_DQ060632
701
771

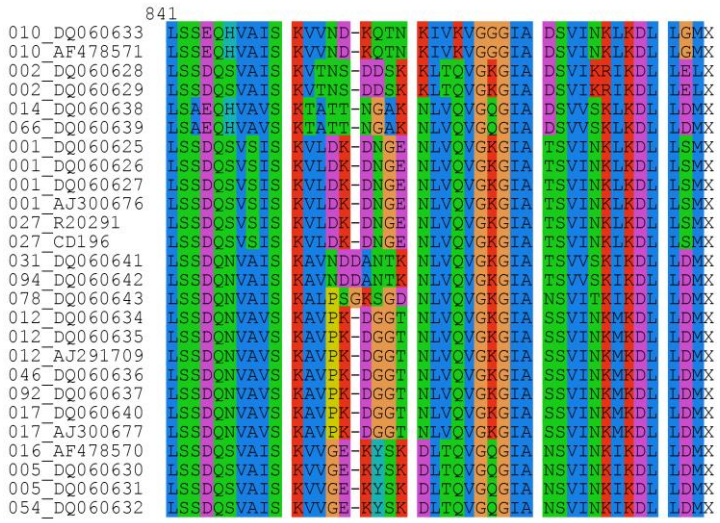


Figure 3.1 Multiple Sequence Alignment of SlpA The complete amino acid sequences for the slpA protein from 26 different strains of *C. difficile* aligned using MUSCLE 3.6 (Edgar 2004). Sequences from ribotypes 001, 002, 005, 010, 012, 014, 016, 017, 027, 031, 046, 054, 066, 092 and 094 are present. The first 24 peptides represent the conserved signal peptide. Residues 25 – 430 represent the variable LMW SLP, while positions 430 – 883 represent the HMW SLP.

3.2.2 *slpA* Likelihood Mapping

Likelihood mapping tests were performed using TreePuzzle v5.2 (Schmidt et al. 2002) to determine if there was sufficient phylogenetic signal present. This software breaks the phylogenetic tree into all possible quartets and assesses the support for each quartet (Strimmer & von Haeseler 1997). If the data contains no phylogenetic signal, then the likelihood of all three possible relationships for that quartet will be equally likely, represented by populations of the three tips of a triangle, seen in Figure 3.2. If sufficient phylogenetic signal is present, the majority of the signal will appear in these tip regions. If little or no phylogenetic signal is present, the majority of the signal will be in the vertices and central region of the triangle, representing a network or star phylogeny respectively. 96.8% of signal appeared in the corners of the triangle, indicating sufficient signal for phylogenetic analysis to be carried out on the sequences. The results can be seen in Figure 3.2.

Analysis of amino acid composition bias was also carried out on the sequences. This was to determine if sequences were being falsely grouped together on the phylogeny due to similar amino acid frequencies rather than recent shared history. There is much evidence for this problem in the literature (Loomis & Smith 1990; Penny et al. 1990; Hasegawa & Hashimoto 1993). The composition of each sequence was compared to the frequency distribution assumed in the model of evolution (Schmidt et al. 2002). All sequences passed this test, indicating that there was no composition bias to negatively impact the phylogeny. The results of the amino acid composition bias tests can be seen in Table 3.1.

Figure 3.2

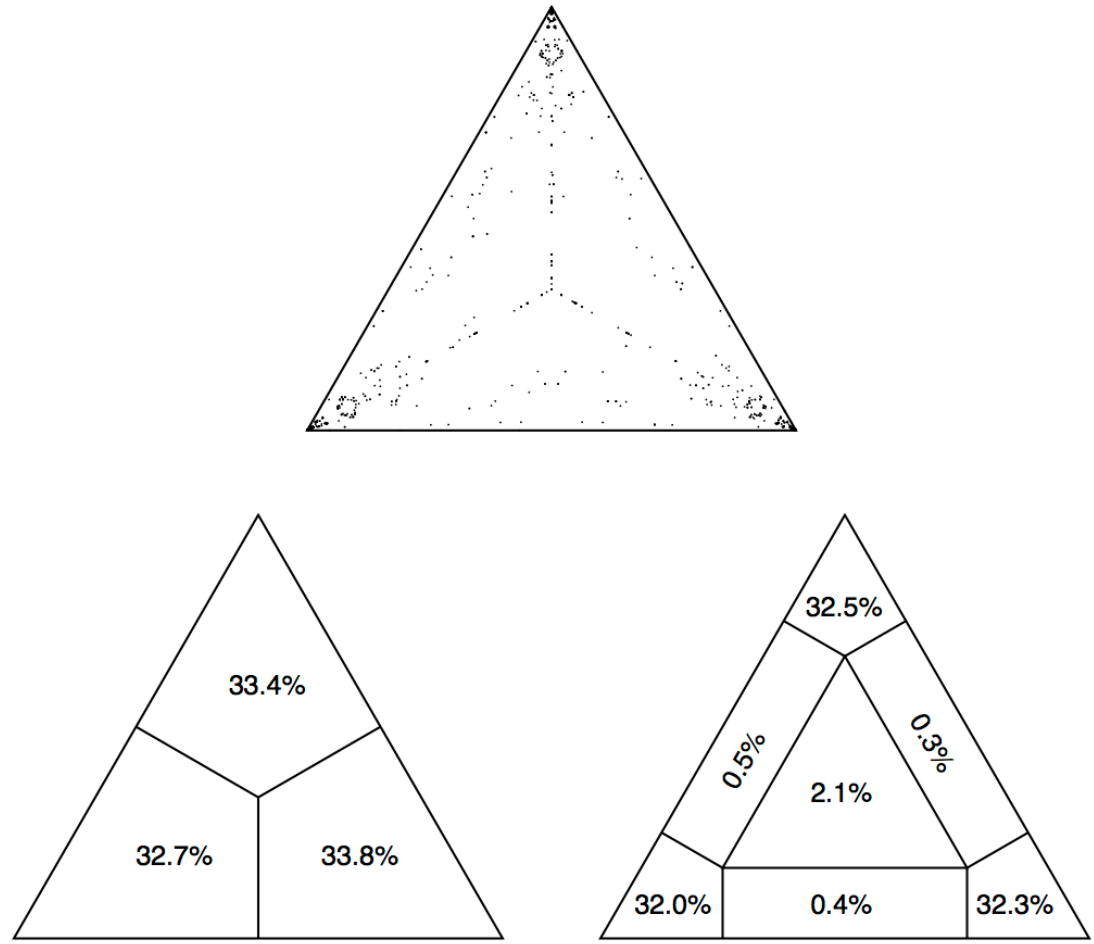


Figure 3.2 Likelihood Mapping Results for SlpA

Results of the likelihood mapping tests for the level of phylogenetic signal in the dataset. The vast majority of the signal (> 96%) appears in the corners of the triangle, and are evenly distributed across all three corners, indicating there is sufficient phylogenetic signal within the dataset for the analysis to be carried out.

Table 3.1

Sequence	5% Chi-Squared Test	p-value
DQ060633	Passed	99.99%
AF478571	Passed	100.00%
DQ060628	Passed	92.86%
DQ060629	Passed	92.86%
AF478570	Passed	95.22%
DQ060630	Passed	95.22%
DQ060631	Passed	95.22%
DQ060632	Passed	96.48%
DQ060638	Passed	99.95%
DQ060639	Passed	99.95%
DQ060625	Passed	99.99%
DQ060626	Passed	99.99%
DQ060627	Passed	99.99%
AJ300676	Passed	99.99%
R20291	Passed	99.92%
CD196	Passed	99.92%
DQ060641	Passed	99.99%
DQ060642	Passed	99.99%
DQ060643	Passed	100.00%
DQ060634	Passed	100.00%
DQ060635	Passed	100.00%
AJ291709	Passed	100.00%
DQ060636	Passed	100.00%
DQ060637	Passed	100.00%
DQ060640	Passed	99.95%
AJ300677	Passed	99.95%

Table 3.1 Results of test for compositional heterogeneity. The amino acid composition of all sequences was calculated and compared with the frequency distribution assumed in the maximum likelihood model.

3.2.3 Phylogenetic Trees

The dataset consisted of *slpA* genes from 29 isolates of *C. difficile* representing 16 ribotypes and 886 aligned amino acid positions. The Bayesian phylogenetic trees of the gene were constructed using both MrBayes v3.1.2 (Huelsenbeck & Ronquist 2001) and PhyloBayes v3 (Lartillot et al. 2009). Both programs produced statistically identical trees (See Appendix for statistical analysis). A radial phylogram was produced using the Dendroscope software (Huson et al. 2007) and is shown in Figure 3.3 below. Ribotypes with very short branch lengths since divergence from their common ancestor are given as clusters at the tips of branches. Branch lengths are proportional to amount of evolutionary change. Support values for each branch are given as posterior probabilities (PP), and were found to be the maximum possible value (i.e. 1.00), except where otherwise specified on the tree. The phylogeny shows that the hypervirulent RT 027 is closely related to RT 001, with RT 001 being a particularly common strain in Ireland.

Figure 3.3

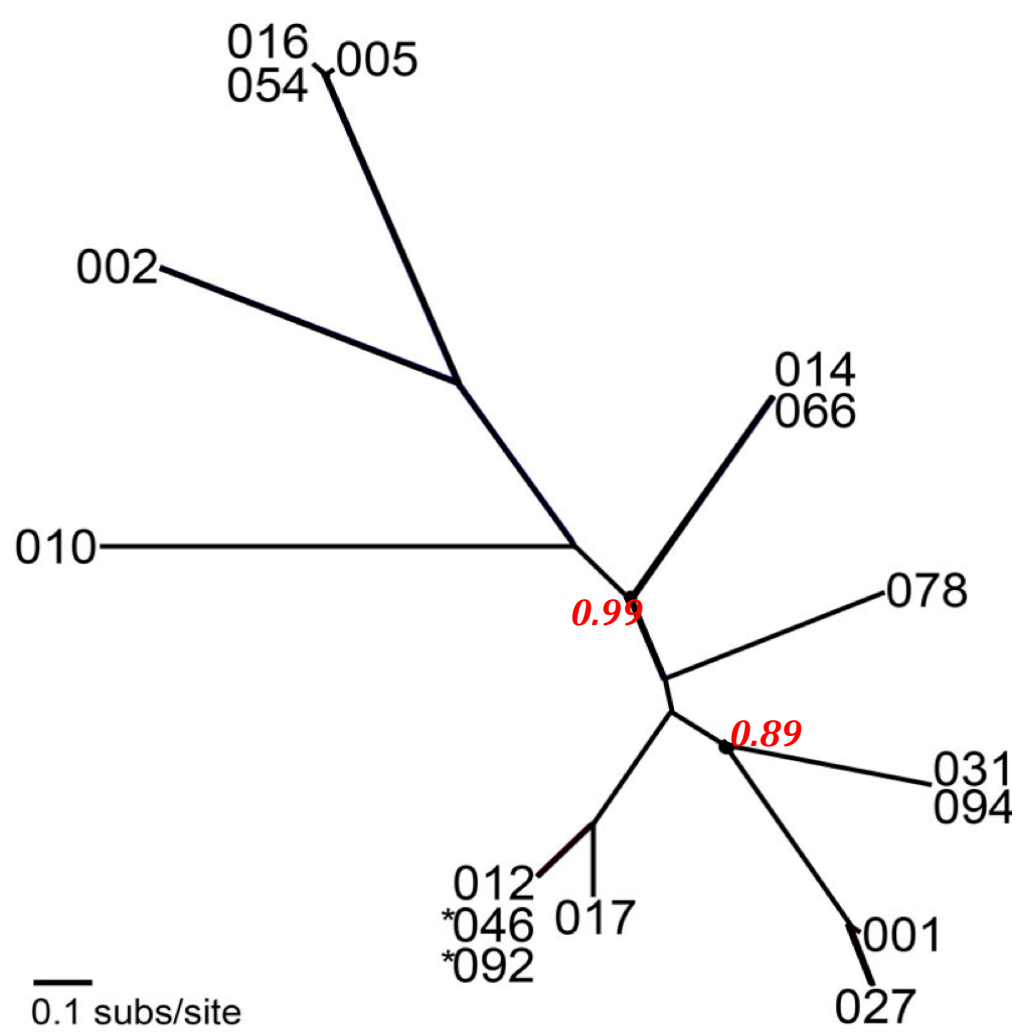


Figure 3.3 Phylogenetic tree of SlpA

Phylogenetic tree of the SlpA protein from 16 major ribotypes of *C. difficile*, reconstructed using MrBayes (Huelsenbeck & Ronquist 2001). The ribotypes are labelled using conventional numbering system. Ribotypes with very short branch lengths since divergence from their common ancestor are given as clusters at the tips of branches. Branch lengths are proportional to amount of evolutionary change. All but three nodes had the maximum posterior probability value (PP) = 1.00: two of these values are shown in red italics on the tree - 0.99 and 0.89. The final exception was the node joining 046 and 092, where * denotes PP of 0.82.

3.2.4 Selective Pressure Analysis

Both the alignment and phylogenetic tree were used to determine selective pressure variation along the alignment. D_n/D_s ratios were estimated using maximum likelihood at both site-specific and lineage-specific levels. The site-specific analysis used 8 models. M1 gave a InL value of -16,292.610. Model M2, had an InL = -16,242.808. When an LRT was carried out between M1 and M2, M2 was found to be a significantly better fit for that data then M1 ($2\Delta l$ = 99.604, P<0.05). As M2 predicted sites of positive selection on *slpA*, this shows that a model allowing for, and predicting, positive selection (M2) fit the data better than neutral models (M1). M0 had an InL value = -16,877.369 and an ω = 0.1440, meaning that under this model purifying selection is invoked. Both M3 (*k=2*) and M3 (*k=3*) gave better likelihood scores than the previous models, InL = -16,210.101 for M3 (*k=2*) and InL = -16,098.846 for M3 (*k=3*). The difference in likelihood between M3 (*k=2*) and M0 was statistically significant ($2\ \Delta l$ = 1334.535, P<0.05), as was the difference in likelihood values between M3 (*k=3*) and M3 (*k=2*) (Δl = 111.254). Model M8 gave the lowest InL value of all the models at InL = -16,052.155. M8 was compared with both M7 and the null model M8a with two independent LRTs, and proved statistically more significant than either of the rested models. A $2\Delta l$ = 106.64 for M7 v M8, and a $2\Delta l$ = 87.17 for M8a v M8 was calculated. Therefore, M8 is the site-specific model that best fits the data and gives the highest likelihood. Under this model, 44 amino acid sites in total were estimated to be undergoing positive selection (ω = 0.5). A summary of all LRTs is given in Table 3.2. A summary of the models used along with estimated sites under positive selection are given in Table 3.3 for site-specific models and in Table 3.5 for lineage-specific. A 3-D model of the LMW

SLP was obtained from PDB (accession number: 3cvz) to visualise the spatial positioning of positively selected amino acid residues (Figure 3.4) (Fagan et al. 2009).

The portion of the alignment representing the LMW protein-coding region was highly variable between strains. Site-specific analysis showed that residues estimated to be positively selected were largely located within a loop-rich region in domain 2 of the LMW portion of the gene.

The lineage-specific analyses yielded a more complex story. All lineages under selection can be seen in Figure 3.5. On a number of branches, positive selection was seen to be occurring in the HMW protein-coding region, an area of the gene that is well conserved across the strains in the dataset, see alignment (Figure 3.1). The vast majority of branches with evidence of positive selection have their positively sites restricted to the HMW protein, however 4 branches (branch numbers 7, 9, 11 and 12) exhibited positive selection in the LMW protein. Branches 9, 11 and 12 all estimate a small number of sites in the LMW SLP to be under selection, with branches 11 and 12 leading to hypervirulent RT 017 and 078 (Table 3.4 and Figure 3.5). Branch 3 (RT 002) estimated 11 sites under positive selection in the LMW and 41 sites in the HMW of the SLP (*slp*A).

Of particular importance here are the results for branch 7, leading to RT 027 (Table 3.4). RT027 is of clinical importance due to the fact that it is hypervirulent (Marsh et al. 2012), and positive selective pressure (and

subsequent protein functional shift) may be a contributing factor to its pathogenesis.

In total, eight branches show positive selection occurring in the HMW protein, including branches leading directly to RTs 010, 002, 005, 031 and 094. This indicates that these RTs have been under selective pressure to adapt, and given the function of the HMW protein the pressure may have been for improved adhesion to the host epithelium (Calabi et al. 2002; Merrigan et al. 2013). There are relatively few sites in the LMW region under positive selection for these branches. There was no evidence of positive selection in either LMW or HMW regions in the common RT 001.

Table 3.2

Branch Numbered as per Fig.3.5	LRT	Null Model lnL	Alt Model lnL	ΔInL	Critical Value	Null Rejected ?
-	m0 vs m3Discrtk2	-16877.36	-16210.10	1334.5	5.99	Yes
-	m1Neutral vs m2Selection	-16292.61	-16242.80	99.60	5.99	Yes
-	m3Discrtk2 vs m3Discrtk3	-16210.10	-16098.84	111.25	1	Yes
-	m7 vs m8	-16105.47	-16052.15	106.64	5.99	Yes
	m8a vs m8	-16095.74	-16052.15	87.17	2.71	Yes
1	m1Neutral vs modelA	-16292.61	-16283.29	18.64	5.99	Yes
1	modelAnull vs modelA	-16287.47	-16283.29	8.36	3.84	Yes
2	m1Neutral vs modelA	-16292.61	-16287.40	10.40	5.99	Yes
2	modelAnull vs modelA	-16291.58	-16287.40	8.35	3.84	Yes
3	m1Neutral vs modelA	-16292.61	-16273.15	38.90	5.99	Yes
3	modelAnull vs modelA	-16280.23	-16273.15	14.14	3.84	Yes
4	m1Neutral vs modelA	-16292.61	-16254.53	76.14	5.99	Yes
4	modelAnull vs modelA	-16271.07	-16254.53	33.07	3.84	Yes
5	m1Neutral vs modelA	-16292.61	-16279.22	26.76	5.99	Yes
5	modelAnull vs modelA	-16279.95	-16279.22	1.45	3.84	No
6	m1Neutral vs modelA	-16292.61	-16292.61	0	5.99	No
6	modelAnull vs modelA	-16292.61	-16292.61	$2x10^{-6}$	3.84	No
7	m1Neutral vs modelA	-16292.61	-16267.49	50.235	5.99	Yes
7	modelAnull vs modelA	-16292.69	-16267.49	50.40	3.84	Yes
8	m1Neutral vs modelA	-16292.61	-16286.51	12.18	5.99	Yes
8	modelAnull vs modelA	-16289.93	-16286.51	6.84	3.84	Yes
9	m1Neutral vs modelA	-16292.61	-16280.18	24.84	5.99	Yes

Branch	LRT	Null Model lnL	Alt Model lnL	ΔInL	Critical Value	Null Rejected ?
9	modelAnull vs modelA	-16290.96	-16280.18	21.55	3.84	Yes
10	m1Neutral vs modelA	-16292.61	-16292.61	0	5.99	No
10	modelAnull vs modelA	-16292.61	-16292.61	$2x10^{-6}$	3.84	No
11	m1Neutral vs modelA	-16292.61	-16289.17	6.88	5.99	Yes
11	modelAnull vs modelA	-16292.61	-16289.17	6.88	3.84	Yes
12	m1Neutral vs modelA	-16292.61	-16288.18	8.84	5.99	Yes
12	modelAnull vs modelA	-16291.48	-16288.18	6.59	3.84	Yes
13	m1Neutral vs modelA	-16292.61	-16280.74	23.73	5.99	Yes
13	modelAnull vs modelA	-16289.44	-16280.74	17.39	3.84	Yes
14	m1Neutral vs modelA	-16292.61	-16287.18	10.84	5.99	Yes
14	modelAnull vs modelA	-16289.57	-16287.18	4.78	3.84	Yes
15	m1Neutral vs modelA	-16292.61	-16278.29	28.62	5.99	Yes
15	modelAnull vs modelA	-16283.81	-16278.29	11.02	3.84	Yes
16	m1Neutral vs modelA	-16292.61	-16246.06	93.09	5.99	Yes
16	modelAnull vs modelA	-16269.49	-16246.06	46.86	3.84	Yes
17	m1Neutral vs modelA	-16292.61	-16291.19	2.83	5.99	No
17	modelAnull vs modelA	-16291.19	-16291.19	0	3.84	No
18	m1Neutral vs modelA	-16292.61	-16277.20	30.81	5.99	Yes
18	modelAnull vs modelA	-16282.83	-16277.20	11.26	3.84	Yes

Table 3.2 Table of Likelihood Ratio test (LRT) results for site-specific and lineage-specific models for each branch on the tree. Likelihood values are given for each null and alternative model.

Table 3.3

Model	P	L	Estimates of parameters	PS sites
M0 : one ratio	1	-16877.369	$\omega = 0.144$	None
M1:Neutral	1	-16292.610	$p_0 = 0.617$ $p_1 = 0.382$ $\omega_0 = 0.082$, $\omega_1 = 1.000$	Not allowed
M2:Selection	4	-16242.808	$p_0 = 0.593$ $p_1 = 0.327$ ($p_2 = 0.078$) $\omega_0 = 0.086$ $\omega_1 = 1.00$ $\omega_2 = 68.753$	BEB 46 > 0.50 12 > 0.95 4 > 0.99
M3:Discrete(K = 2)	3	-16210.101	$p_0 = 0.460$ $p_1 = 0.539$ $\omega_0 = 0.031$ $\omega_1 = 0.355$	None
M3:Discrete(K = 3)	5	-16098.846	$p_0 = 0.382$ $p_1 = 0.515$ $p_2 = 0.101$ $\omega_0 = 0.018$ $\omega_1 = 0.245$ $\omega_2 = 20.897$	None
M7: Beta	2	-16105.477	$p = 0.418$ $q = 1.298$	Not allowed
M8: Beta&Omega > 1	4	-16052.155	$p_0 = 0.920$ $p = 0.476$ $q = 1.960$ $p_1 = 0.079$ $\omega = 29.209$	BEB 44 > 0.50 11 > 0.95 4 > 0.99
M8a: Beta&Omega = 1	3	-16095.794	$p_0 = 0.871$ $p = 0.523$ $q = 2.949$ $p_1 = 0.128$, $\omega = 1.000$	Not allowed

Table 3.3 Results of Site-specific analysis. All models tested are displayed, along with number of parameters, log likelihood scores, estimates of parameters, and number of positively selected sites.

Table 3.4

Branch	Model	P	lnL	Estimation of parameters	PS Sites
1	A	3	-16283.290	$p_0 = 0.612$ $p_1 = 0.379$ $p_2 = 0.004$ $p_3 = 0.002$ **Background:** $\omega_0 = 0.081$ $\omega_1 = 1.000$ $\omega_2 = 1.000$ **Foreground:** $\omega_0 = 0.081$ $\omega_1 = 1.000$ $\omega_2 = 999.0$	**BEB** 3>0.99
2	A	3	-16287.406	$p_0 = 0.613$ $p_1 = 0.379$ $p_2 = 0.004$ $p_3 = 0.002$ **Background**: $\omega_0 = 0.081$ $\omega_1 = 1.000$ $\omega_2 = 1.000$ **Foreground**: $\omega_0 = 0.081$ $\omega_1 = 1.000$ $\omega_2 = 999.0$	**BEB** 3>0.99
3	A	3	-16273.156	$p_0 = 0.546$ $p_1 = 0.325$ $p_2 = 0.080$ $p_3 = 0.047$ **Background**: $\omega_0 = 0.077$ $\omega_1 = 1.000$ $\omega_2 = 1.000$ **Foreground**: $\omega_0 = 0.078$ $\omega_1 = 1.000$ $\omega_2 = 305.086$	**BEB** 52>0.50 11>0.95 1>0.99
4	A	3	-16254.536	$p_0 = 0.534$ $p_1 = 0.326$ $p_2 = 0.086$ $p_3 = 0.052$ **Background**: $\omega_0 = 0.076$ $\omega_1 = 1.000$ $\omega_2 = 1.000$ **Foreground**: $\omega_0 = 0.078$ $\omega_1 = 1.000$ $\omega_2 = 49.513$	**BEB** 51>0.50 21>0.95 8>0.99
7	A	3	-16267.493	$p_0 = 0.577$ $p_1 = 0.364$ $p_2 = 0.035$ $p_3 = 0.022$ **Background**: $\omega_0 = 0.082$ $\omega_1 = 1.000$ $\omega_2 = 1.000$ **Foreground**: $\omega_0 = 0.082$ $\omega_1 = 1.000$ $\omega_2 = 999.000$	**BEB** 15>0.50
8	A	3	-16286.517	$p_0 = 0.599$ $p_1 = 0.369$ $p_2 = 0.019$ $p_3 = 0.012$ **Background**: $\omega_0 = 0.080$ $\omega_1 = 1.000$ $\omega_2 = 1.000$ **Foreground**: $\omega_0 = 0.080$ $\omega_1 = 1.000$ $\omega_2 = 65.618$	**BEB** 9>0.50 1>0.95
9	A	3	-16280.189	$p_0 = 0.585$ $p_1 = 0.366$ $p_2 = 0.029$ $p_3 = 0.018$ **Background**: $\omega_0 = 0.080$ $\omega_1 = 1.000$ $\omega_2 = 1.000$ **Foreground**: $\omega_0 = 0.081$ $\omega_1 = 1.000$ $\omega_2 = 998.998$	**BEB** 11>0.50 1>0.95
11	A	3	-16289.170	$p_0 = 0.593$ $p_1 = 0.372$ $p_2 = 0.020$ $p_3 = 0.013$ **Background**: $\omega_0 = 0.082$ $\omega_1 = 1.000$ $\omega_2 = 1.000$ **Foreground**: $\omega_0 = 0.081$ $\omega_1 = 1.000$ $\omega_2 = 998.975$	**BEB** 3>0.50

Branch	Model	P	lnL	Estimation of parameters	PS Sites
12	A	3	-16288.187	p_0=0.604 p_1=0.367 p_2=0.017 p_3=0.010 **Background**: ω_0=0.080 ω_1=1.000 ω_2=1.000 **Foreground**: ω_0=0.080 ω_1=1.000 ω_2=998.952	**BEB** 6>0.50 1>0.95
13	A	3	-16280.744	p_0=0.601 p_1=0.341 p_2=0.036 p_3=0.020 **Background**: ω_0=0.079 ω_1=1.000 ω_2=1.000 **Foreground**: ω_0=0.079 ω_1=1.000 ω_2=32.021	**BEB** 18>0.50 1>0.99
14	A	3	-16287.185	p_0=0.606 p_1=0.376 p_2=0.010 p_3=0.006 **Background**: ω_0=0.079 ω_1=1.000 ω_2=1.000 **Foreground**: ω_0=0.081 ω_1=1.000 ω_2=67.858	**BEB** 3>0.50 2>0.95
15	A	3	-16278.296	p_0=0.598 p_1=0.354 p_2=0.029 p_3=0.017 **Background**: ω_0=0.077 ω_1=1.000 ω_2=1.000 **Foreground**: ω_0=0.078 ω_1=1.000 ω_2=8.993	**BEB** 12>0.50 4>0.95 1>0.99
18	A	3	-16277.205	p_0=0.560 p_1=0.337 p_2=0.063 p_3=0.038 **Background**: ω_0=0.078 ω_1=1.000 ω_2=1.000 **Foreground**: ω_0=0.077 ω_1=1.000 ω_2=94.127	**BEB** 45>0.50 7>0.95 2>0.99

Table 3.4 Results of the lineage-specific analysis. All models tested, log likelihood, estimates of parameters and sites under positive selection are given for each branch.

Figure 3.4

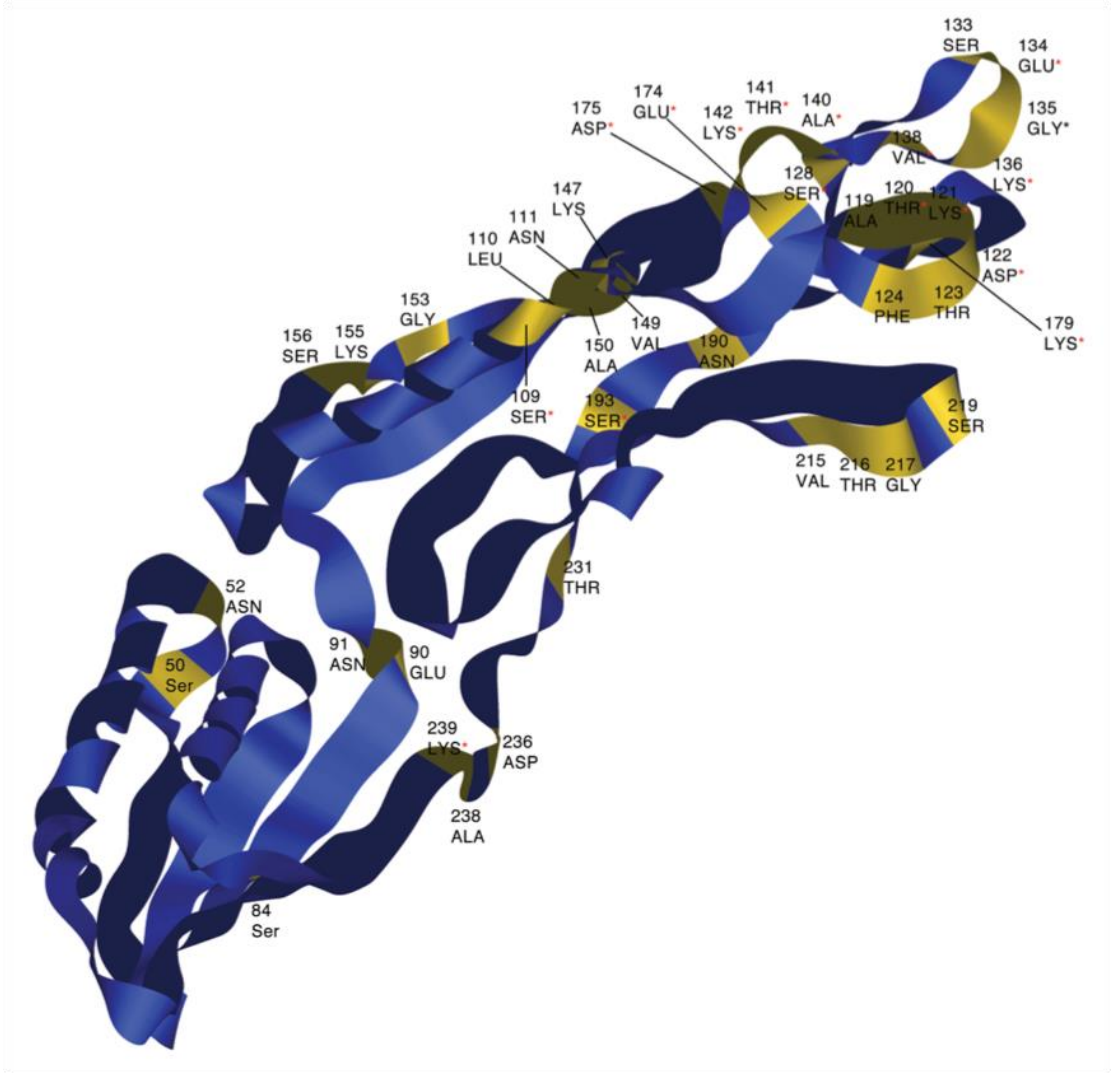


Figure 3.4. Positively selected residues highlighted on the 3D structure of the LMW SLP.

The LMW SLP crystal structure was obtained from the EMBL database and positively selected sites highlighted. α-helices and β-sheets can clearly be seen along with loop regions. Specific amino acids under positive selection are labelled in gold. The majority of sites under positive selection can be seen in Domain 2, which is rich in loops. Red asterisks indicate residues with probability of greater than 0.9 being under positive selection.

Figure 3.5

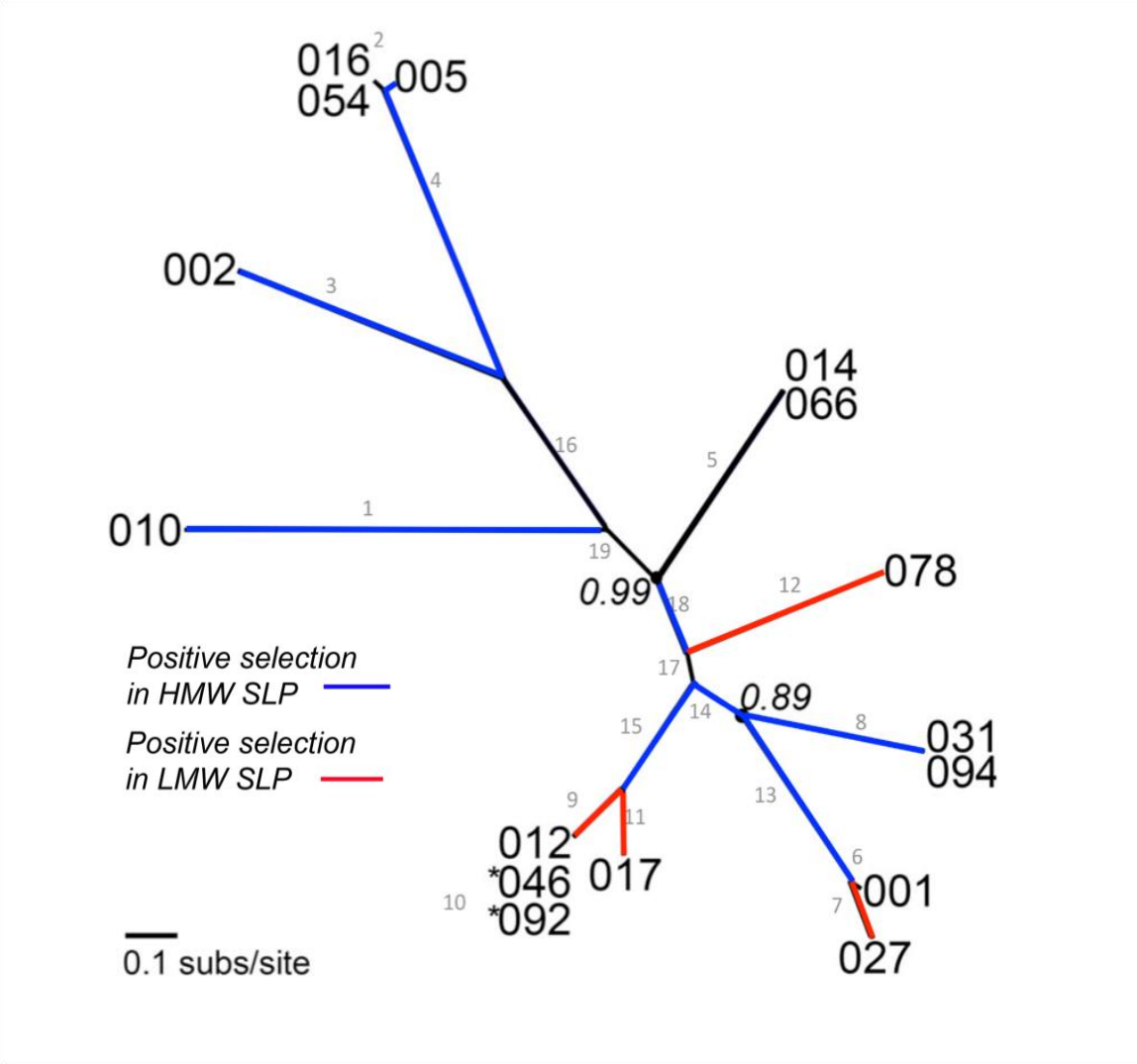


Figure 3.5. The phylogenetic tree of the SLPs analysed and summary of lineages under positive selection. Branches in blue correspond to positively selected sites detected in the High Molecular Weight (HMW) protein, those highlighted in red correspond to positively selected sites in the Low Molecular Weight (LMW) protein. Positive selection was estimated on branches leading to ribotypes 012, 017, 027 and 078 for the LMW SLP, while selection in the HMW SLP was detected in branches leading to 002, 005, 010, 031 and 094. In all cases the parameters for selective pressure analysis were chosen from the codon substitution model that fitted the data best following likelihood ratio tests.

3.3 DISCUSSION

C. difficile is an increasingly problematic nosocomial pathogen and the main focus of research into its pathogenesis has been cytotoxic protein production. Deletions in the toxin regulator TcdC were thought to account for the increased disease severity in patients infected with RT 027 (Carter et al. 2011), yet this same deletion has been observed to result in a functional TcdC (Dupuy et al. 2008). Therefore toxin production alone cannot account for such a wide range of symptoms observed in different hospitals. We have previously shown that SLPs from RT 001 can induce an inflammatory response, specifically through TLR4 (Ryan et al. 2011). While we know that these proteins do indeed have the capacity to induce inflammation, to date no comparative studies have been performed on the SLPs of different strains. Little is known of how this sequence variability affects the ability of the immune system to recognise and respond to *C. difficile*. The results of this study indicate that the observed sequence variation in the SLPs of *C. difficile* is not entirely random, and many of the observed variable sites along the *slpA* gene are under positive selection.

The 26 isolates in our dataset exhibit vastly different sequences in the *slpA* gene, particularly in the area encoding the LMW protein. The HMW protein was much more conserved (Karjalainen et al. 2001) suggesting the importance of structure in its role in adherence (Calabi et al. 2002; Merrigan et al. 2013). As the LMW subunit is the outermost component of the SLP complex (Fagan et al. 2009), it is likely the first part of the bacteria with which immune cells will come into contact. Sequence variation between strains suggests that the immune system may not be able to recognise SLPs from different strains to the same extent, and

this sequence variation may play a role in the vast spectrum of symptoms observed in *C. difficile* Infection. Here we propose that SLP sequence is an essential determinant of susceptibility and severity of infection.

Signatures of positive selection were indeed detected in the SLP sequences in our dataset, strongly suggesting this variation was not random, and inferred a survival benefit to the pathogen. The site-specific analysis showed positive selection occurring in the LMW-coding region of the gene. The majority of sites estimated with the highest probability (PP>0.95) occur on the outermost region of the protein, Domain 2. While the exact residues that interact with TLR4 are not known, the position of these positively selected sites make them a potential candidate. This region of the protein is likely the most accessible to immune cells, as the S-Layer lattice may make other regions inaccessible for recognition (Fagan et al. 2009). Amino acid substitutions, particularly at important regions involved in recognition, may affect the ability of the host to detect the bacteria through SLP-TLR4 interaction. Variation in sequence will likely perturb recognition motifs, thereby affecting the recognition and interaction of immune cells. The SLPs can therefore be thought of as a molecular "cloak" to mask the bacteria from the pattern recognition receptors (PRRs) of the host. S-Layers of bacteria have previously been reported to play a role in immune evasion (Thompson 2002; Kern & Schneewind 2010; Bahl et al. 1997; Sara et al. 2000; Sleytr 1997), and the SLPs of *C. difficile* may be acting in a similar fashion. Conversely, mutations at these sites may result in stronger activation of an inflammatory response. In the case of a gut pathogen like *C. difficile*, this may be beneficial. Increased inflammation will result in a breakdown of the epithelial

lining, and as SLPs bind to extracellular matrix components (Calabi et al. 2002), may allow the pathogen to gain a stronger foothold.

As we were interested in comparing the SLPs from different strains of *C. difficile*, we performed lineage specific analysis for selective pressure heterogeneity. The results indicate that many branches along the tree exhibit evidence of positive selection. Unexpectedly, positive selection was detected in the relatively conserved HMW subunit of many ribotypes on the tree. The HMW SLP is mainly involved in adherence to gastrointestinal tissues and ECM components (Calabi et al. 2002); the results suggest that SLPs from ribotypes 010, 002, 005, 031 and 094 may be evolving to alter adherence properties of the protein. Improved adherence will allow for deeper penetration into the tissues and may contribute to persistent disease. Further experimentation would be required to test this hypothesis however. Positive selection is also detected in the LMW region of the gene for RT 002, 031 and 094, mainly in the area necessary for HMW-binding. This may suggest a co-evolution of the two proteins. A possible scenario is that as the HMW protein changes its sequence, its confirmation may change. As a result, the area of the LMW protein involved in binding to the HMW protein may have to modify its sequence to account for the conformational change in the HMW SLP.

Only four branches of the tree estimated positive selection in the LMW subunit. Two of these branches led to the "hypervirulent" 027 and 078 ribotypes. Of particular interest is the close relationship between RT 001 and 027. RT 001 is not associated with severe disease, and we have previously shown SLPs from

this ribotype induce clearance responses in macrophage to help moderate and clear infection (Collins et al. 2014). The SLP sequence from RT 001 is closely related to that of RT 027, yet the sequence differences between the two ribotypes are under positive selection. In total, 15 amino acid resides were estimated to be under positive selective pressure in RT 027 SLPs. Many studies describe 027 as being associated with more severe and persistent disease, and toxin production alone cannot account for this (Carter et al. 2011; Clements et al. 2010; Valiente et al. 2014; Marsh et al. 2012). As the SLPs in RT 027 have undergone a predicted functional shift, they may now be modulating the immune response to ensure the pathogen is not cleared, while also increasing inflammation to cause tissue damage and allow the bacteria to strongly adhere and give rise to persistent and recurrent infection.

When the pattern of positive selection was mapped on to the phylogenetic tree of the 26 *slpA* sequences, two distinct clades can be seen. Hypervirulent strains 027 and 078 appeared on one side of the tree, with the epidemic RT 027 most closely related to RT 001 (Figure 3.5). If the SLPs in this group are evolving in a unique way to avoid recognition by the host immune system, then this would prevent, or at least slow down, clearance of this strain in the host. The bacteria can remain producing toxins undisturbed and so lead to increased pathogenesis. While it cannot be denied that the toxins play a major role in virulence, the SLPs are likely making a significant contribution to pathogenesis.

The existence of horizontal gene transfer (HGT) in prokaryotic cells has recently questioned the validity of phylogenetic trees when examining the evolutionary

history of bacteria (Doolittle 2000). By exchanging genes between species and strains, bacterial genomes can become a mosaic over time, strongly suggesting that linear evolution of individual species does not occur. Indeed, a recent study has shown that recombinational switching can occur in the S-Layer cassette of *C. difficile* (Dingle et al. 2013), and that a single PCR ribotype can carry multiple S-Layer cassettes. For our analysis however, the *slpA* sequences for each isolate in a specific ribotype were identical, and all ribotypes displayed unique sequences. Because of this, we make an initial assumption that the SLPs of *C. difficile* are not undergoing horizontal gene transfer, and therefore are evolving in a linear manner. In addition, while HGT may still be occurring in the genome of *C. difficile*, specifically in the region coding for SLPs, HGT has no bearing on this analysis. As this study is examining only a single gene from *C. difficile*, we are resolving the phylogenetic relationships between *slpA* sequences alone, regardless of any HGT that is occurring. As a result however, we cannot assume that the *slpA* sequences in this dataset are globally universal for each given ribotype. For the proceeding chapters, *slpA* sequences were confirmed by whole genome sequencing where required, and were found to be identical to those in this analysis.

In conclusion, we have resolved the phylogeny of the *slpA* gene from 26 strains of *C. difficile*, and detected signatures of positive selection in multiple sequences, most notably in 'hypervirulent' strains 027 and 078. Variations in these sequences that are undergoing positive selection are inferring a survival benefit on the pathogen, and may play a role in the increased disease severity associated with RT 027 and 078. While the molecular evolutionary analysis of these

sequences gives some intriguing insights into their role in severity of infection and the inflammatory response, further *in vitro* and *in vivo* analyses will be needed to directly confirm the role these proteins have in modulating the immune system. By isolating SLPs from different strains of *C. difficile*, and examining the response of immune cells to the proteins, it will be possible to determine how this positive selection and sequence variation plays a role in the response to the pathogen and clearance of disease.

Chapter 4

Purification and Immune Characterisation of SLPs from different ribotypes of *Clostridium difficile*

4.1 INTRODUCTION

In recent years, the emergence of hypervirulent *C. difficile* strains has elicited a renewed interest in this pathogen. Many previous studies have suggested a role for the SLPs as important virulence factors in infection (Ryan et al. 2011; Vohra & Poxton 2012; Ausiello et al. 2006; Collins et al. 2014; Madan & Jr 2012). For example, human monocytes and human monocyte-derived dendritic cells (MDDCs) have been shown to become activated in the presence of SLPs, resulting in the upregulation of pro-inflammatory cytokines IL-1β and IL-6, as well as cell surface molecules CD80, CD83 and MHC class II (Ausiello et al. 2006).

We have previously shown that SLP activates immune cells through TLR4 (Ryan et al. 2011). Bone marrow-derived dendritic cells (BMDCs), from C3H/HeN mice, stimulated with SLP from RT 001 showed upregulation of IL-12p70, IL-23, TNFα and IL-10, and increased expression of CD40, MHCII, CD80 and CD86. We showed that BMDCs isolated from C3H/HeJ mice, a strain naturally defective in TLR4 signalling, did not become activated by SLP and had no induced cytokine secretion or cell surface marker expression (Ryan et al. 2011; Eidhin et al. 2006). *In vivo* studies further corroborated these findings, and demonstrated that $TLR4^{-/-}$ and $MyD88^{-/-}$ mice were more susceptible to infection, and displayed greater weight loss and greater CFU counts than wild type, $TLR2^{-/-}$ and $TRIF^{-/-}$ mice. $TLR4^{-/-}$ mice were also unable to mount an antibody response to SLP four days post-infection, with low levels of IgG antibody present in the circulation relative to wild type.

Another indicator of the importance of SLPs is demonstrated by studies showing that immunisation of hamsters with anti-SLP antibody significantly decreases the rate of infection, increasing rates of phagocytosis and lowering bacterial load in the gut. As SLPs are the main surface antigen of *C. difficile*, antibody binding may also affect the bacteria's ability to bind to the gastrointestinal tract (O'Brien et al. 2005).

In the previous chapter, we show that evidence of positive selection was detected in the SLPs of certain strains. Sequence differences between SLPs of other *C. difficile* strains were also observed, with some of these sequence differences being positively selected for. To further examine this, and to determine if these observed sequence differences affect the immune response, we sought to purify SLPs from a variety of *C. difficile* ribotypes, and examine the ability of these SLPs to elicit an immune response.

Given that we wanted to isolate, purify and quantify SLPs from a range of different strains, we need to consider that the varying size of SLPs (Eidhin et al. 2006) requires unique parameters for SLP purification by FPLC. Therefore, this study first aimed to fully optimise *C. difficile* growth, and SLP purification protocols for a variety of *C. difficile* ribotypes. This would then allow us to then compare SLPs between ribotypes, for their effect on the immune response. Furthermore, it will enable us to compare ribotypes undergoing positive selection with all other ribotypes, thus helping to determine if positive selection plays a role in modulating the immune response. To this end, we choose to purify ribotypes 027 and 078, which displayed evidence of positive selection, and

ribotypes 001 and 014, which did not, for comparison. This will greatly increase our understanding of how *C. difficile* modulates the immune response, specifically through its outer layer of SLPs.

4.2 RESULTS

4.2.1 *C. difficile* growth times on blood agar and liquid broth

For each ribotype stock used, plates were examined after 24 hours incubation for the presence of *C. difficile* colonies. *C. difficile* appeared as circular cream/grey colonies with irregular or jagged edges on the blood agar. In each case, sparse colonies were visible after 24 hours. After 48 hours colonies appeared larger and more confluent.

The growth of *C. difficile* in liquid broth required optimisation. Colonies were grown and incubated in BHI (Brain Heart Infusion) broth until an OD value of 0.8~1.0 was achieved. An incubation time of 16 hours yielded satisfactory results for RT 001, with broth appearing cloudy. A 16-hour incubation time for RT 078 resulted in the formation of pellets in the bottom of the tubes, which were easily re-suspended upon agitation. This build-up of cellular material indicated a shortened incubation time was required. Incubation for 8 hours yielded the necessary OD value of one for RT 078. RT 027 behaved in a similar manner to RT 001, with the bacteria growing in BHI after 16 hours of incubation with no problems. The incubation time for each SLP ribotype is given in Table 4.1.

Table 4.1

***C. difficile* Ribotype**	**Incubation Time**
001	16 Hours
027	16 Hours
078	8 Hours
014	16 Hours

Table 4.1 Required growth times for C. difficile in BHI broth. RT 078 grew at a much more rapid rate than RT 001, 014 or 027.

4.2.2 Purification Optimisation of SLPs

Crude S-Layer extract was collected from each ribotype and dialysed before being run through the FPLC machine. For ribotype 001 a gradient of 0-0.3M NaCl in 20mM Tris-HCl pH 8.5 at a flow rate of 4 ml/min was used. The increase in NaCl concentration occurred over 20 column volumes (CV). The gradient was visualised on a graph alongside UV peaks, which represented proteins detaching from the column and eluting into 2mL fractions. Fractions collected on or near high UV peaks on the graph were chosen to run on gels as they likely contained protein (Figures 4.1 and 4.8).

The initial parameters resulted in purified SLPs collected across many fractions, however the initial 30 fractions contained no protein. This was due to heavy binding to the column, and an insufficient concentration of salt to remove the bound proteins. When the length of the gradient was lessened to 10 CV (Figures 4.2 and 4.8), pure SLPs were produced, but with fewer empty fractions. Further shortening of the gradient to 8 CV eluted fractions from A1-B10 and gave rise to extra bands visualised with the SLPs (Figures 4.3 and 4.9). This indicated the slope of the gradient was too high, and many proteins were eluting from the column at once. The dark SLP bands were visible from A5 – A11 with other protein bands. A gradient of 10 CV was used as standard for all further RT 001 purifications. The SLPs were eluted in the later fractions B14 – C11. Various other bands were detected in later fractions, representing lesser proteins of the surface layer.

SLPs from RT 014 were also eluted at a gradient of 10CV (Figure 4.4 and Figure 4.10). Fractions were collected later in the run for this ribotype, therefore no other proteins were eluted during the purification, and only a single UV peak was observed on the graph. When crude S-Layer extract from RT 078 was run with a 10 CV gradient of 0-0.3M NaCl, many fractions contained no protein. For this ribotype, as well as 027, shortening the length to 8 CV improved the run time (Figures 4.5 and 4.12 for 027, 4.6 and 4.13 for 078). FPLC graphs for the four different ribotypes are shown in Figures 4.1-4.6.

Table 4.2

Ribotype	Gradient used
001	10 CV
014	10 CV
027	8 CV
078	8 CV

Table 4.2 Salt gradients (in Column Volumes) required to elute purified SLPs.

Figure 4.1 – Ribotype 001 Eluted at 20CV

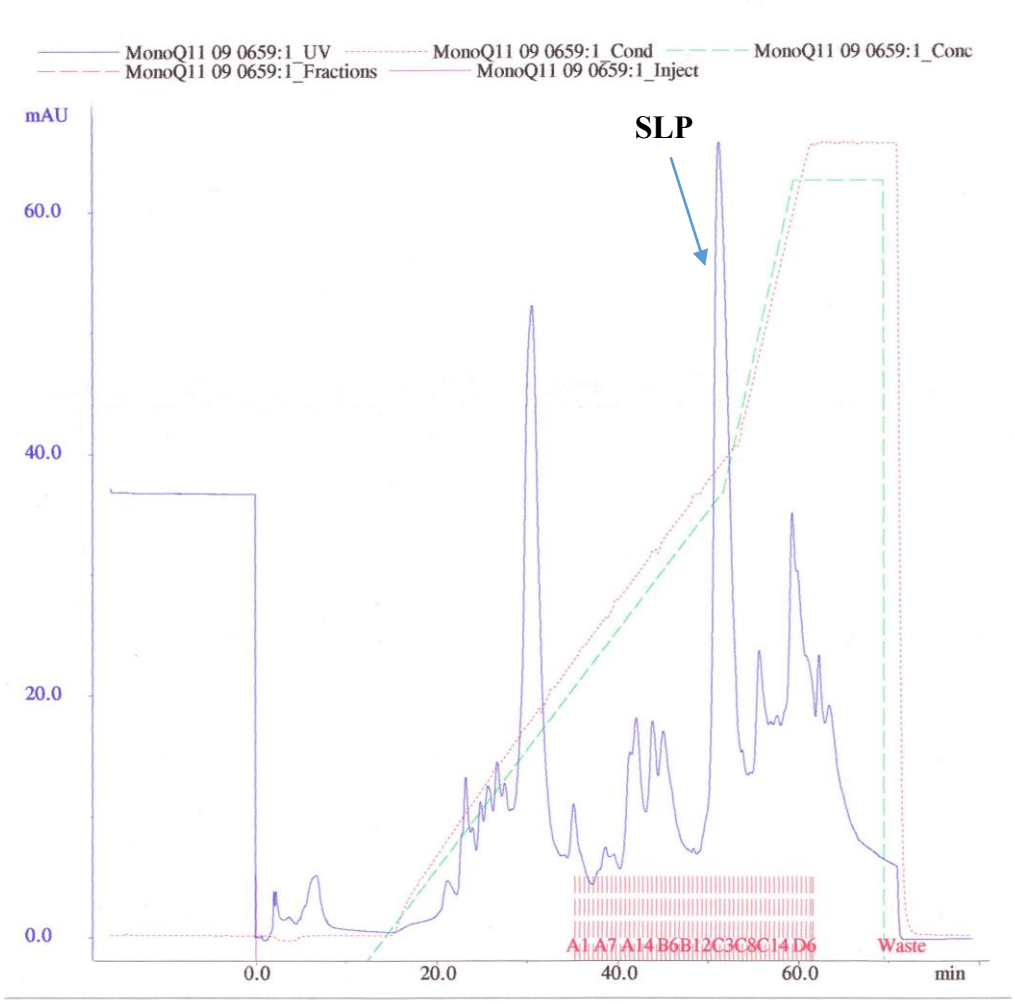


Figure 4.1 FPLC graph from crude S-Layer of ribotype 001 eluted at 20CV.

Peaks represent light absorption by displaced proteins in mili-absorption units (mAU), quantified on the y-axis. Time in minutes is given on the x-axis. The largest peak on the graph at 50 mins represents the purified SLPs, while other peaks represent various surface proteins. The green slope represents increasing concentration of salt, while the red vertical lines represent individual fractions collected.

Figure 4.2 – Ribotype 001 Eluted at 10CV

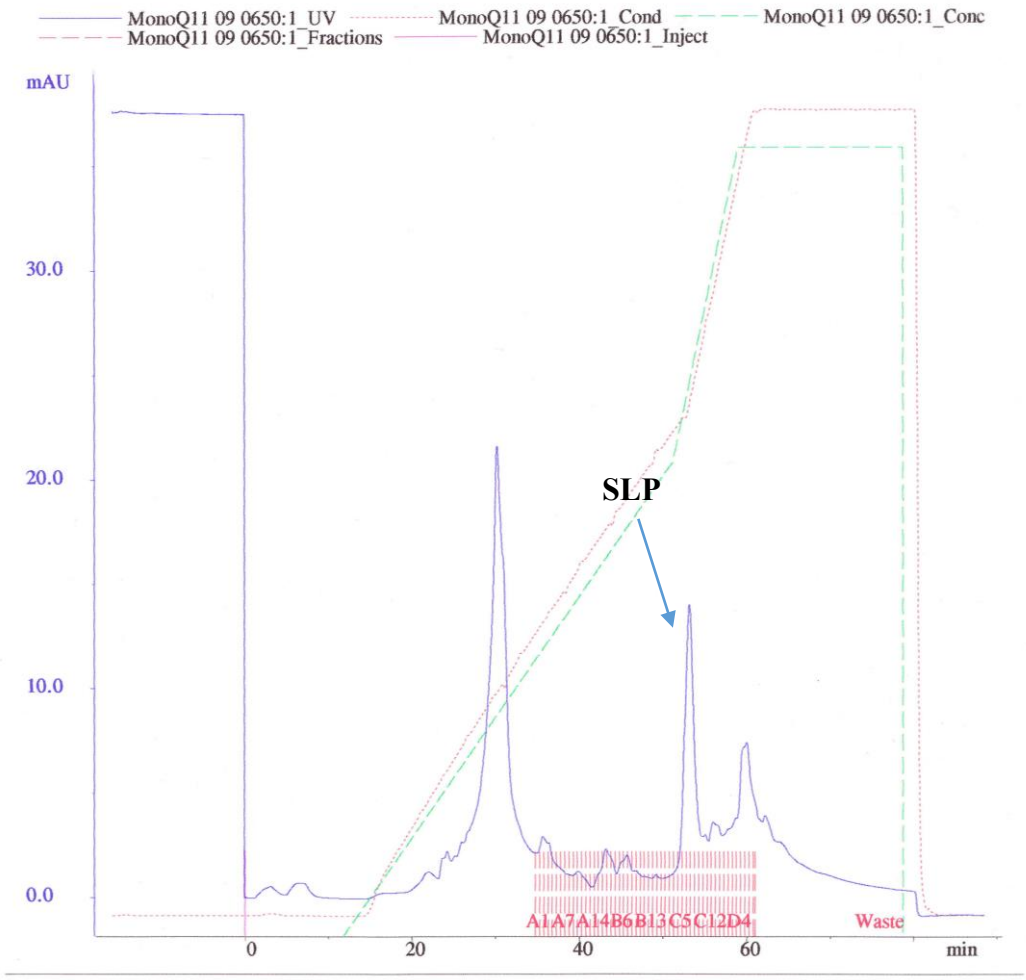


Figure 4.2 FPLC graph from crude S-Layer of ribotype 001 eluted at 10CV.

Peaks represent light absorption by displaced proteins in mili absorption units (mAU), quantified on the y-axis. Time in minutes is given on the x-axis. The largest peak on the graph at 50 mins represents the purified SLPs, while other peaks represent various surface proteins. Fewer peaks are observed under these conditions, as the shorter gradient elutes proteins at a faster rate. The green slope represents increasing concentration of salt, while the red vertical lines represent individual fractions collected.

Figure 4.3 – Ribotype 001 Eluted at 8CV

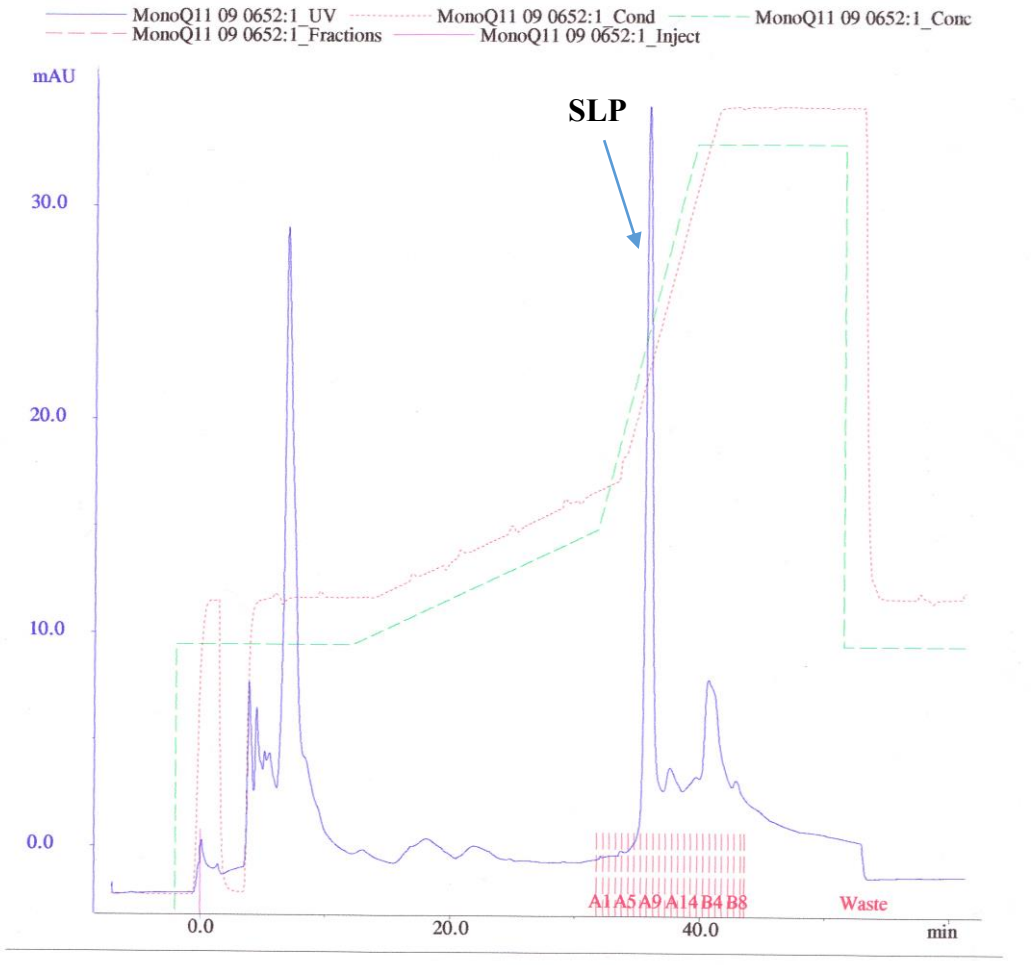


Figure 4.3 FPLC graph from crude S-Layer of ribotype 001 eluted at 8CV.

Peaks represent light absorption by displaced proteins in mili absorption units (mAU), quantified on the y-axis. Time in minutes is given on the x-axis. The largest peak on the graph at 35 mins represents the SLPs. Due to the increased salt gradient; several proteins are eluted at this point. Other peaks represent various surface proteins. Fewer peaks are observed under these conditions, as the shorter gradient elutes proteins at a faster rate. The green slope represents increasing concentration of salt, while the red vertical lines represent individual fractions collected.

Figure 4.4 – Ribotype 014 Eluted at 10CV

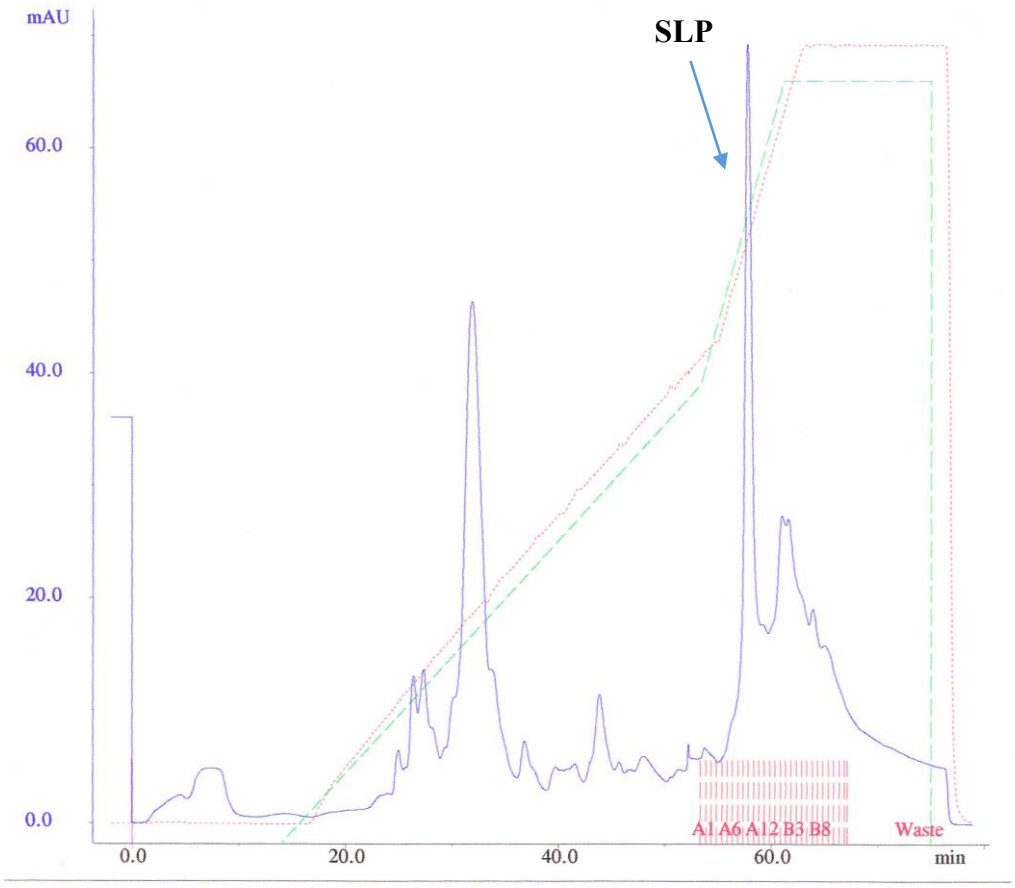


Figure 4.4 FPLC graph from crude S-Layer of ribotype 014 eluted at 10CV.

Peaks represent light absorption by displaced proteins in mili absorption units (mAU), quantified on the y-axis. Time in minutes is given on the x-axis. The largest peak on the graph at 55 mins represents the purified SLPs, while other peaks represent various surface proteins. The green slope represents increasing concentration of salt, while the red vertical lines represent individual fractions collected.

Figure 4.5 – Ribotype 027 Eluted at 8CV

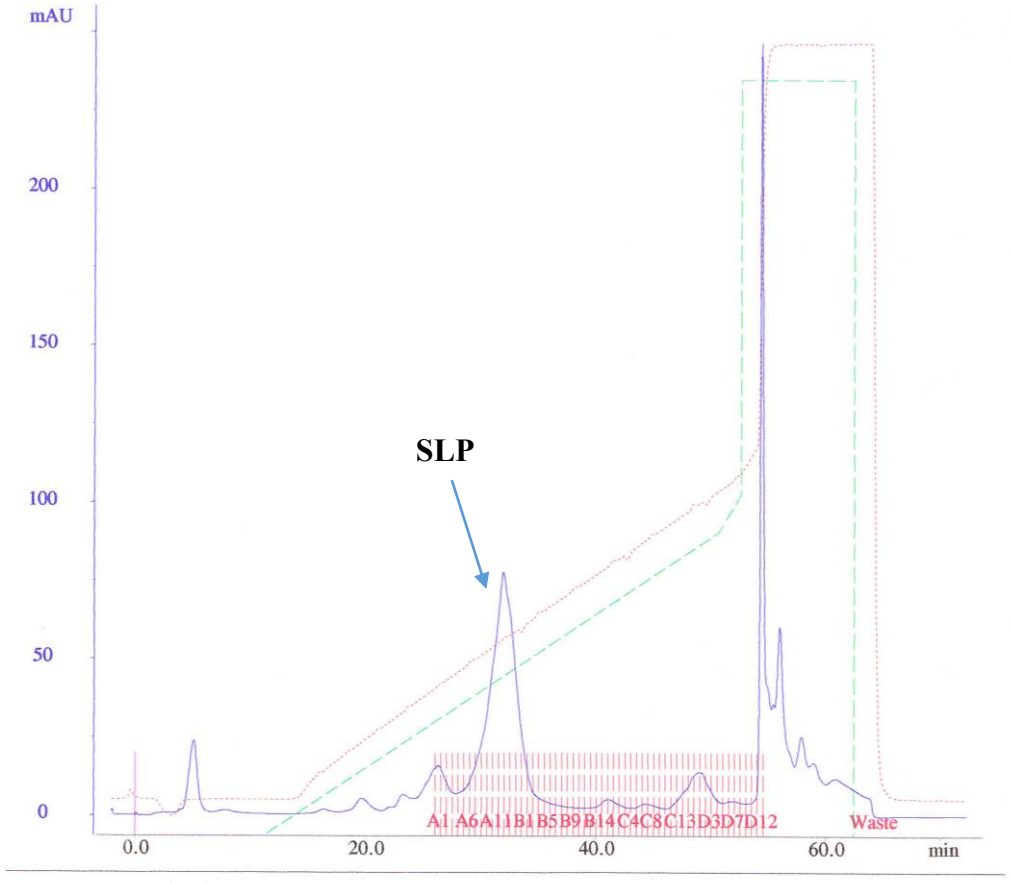


Figure 4.5 FPLC graph from crude S-Layer of ribotype 027 eluted at 8CV.

Peaks represent light absorption by displaced proteins in mili absorption units (mAU), quantified on the y-axis. Time in minutes is given on the x-axis. The largest peak on the graph at 30 mins represents the SLPs. Other peaks represent various surface proteins. The green slope represents increasing concentration of salt, while the red vertical lines represent individual fractions collected.

Figure 4.6 – Ribotype 078 Eluted at 8CV

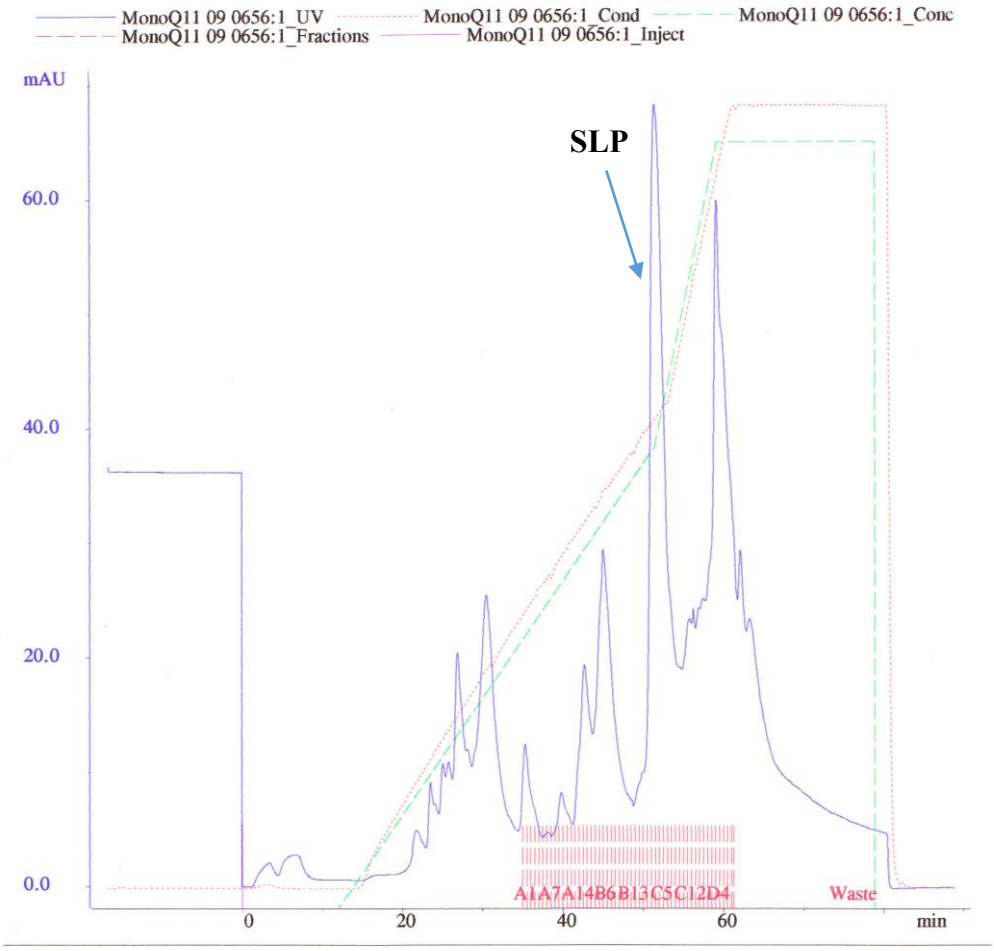


Figure 4.6 FPLC graph from crude S-Layer of ribotype 078 eluted at 8CV.

Peaks represent light absorption by displaced proteins in mili-absorption units (mAU), quantified on the y-axis. Time in minutes is given on the x-axis. The largest peak on the graph at 55 mins represents the purified SLPs, while other peaks represent various surface proteins. The green slope represents increasing concentration of salt, while the red vertical lines represent individual fractions collected.

4.2.3 SDS-PAGE

Fractions (2 mL) were collected during FPLC. The fractions were labelled A1-A15, B1-B15, and C1-C15 etc. Of these, 100µL aliquots of every second sample were taken to run on polyacrylamide gels. Gels were stained as described in Materials and Methods. Any SLPs present were visualised as two bands appearing at roughly 32kDa and 44kDa. Figure 4.7 shows bands present in an undialysed crude S-Layer extract, a dialysed extract and purified SLPs respectively. The figures below show purified SLPs for RT 001 (Figures 4.8 and 4.9), 014 (Figure 4.10), 027 (Figure 4.11) and 078 (Figure 4.12).

Figure 4.7

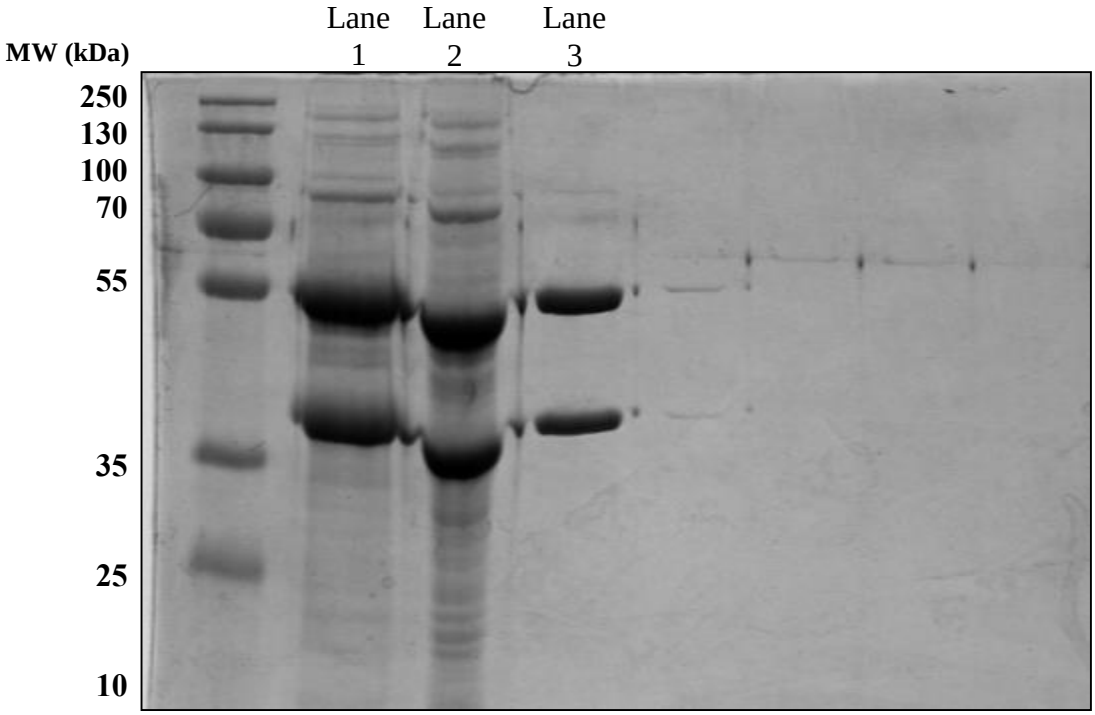


Figure 4.7 SDS-PAGE gels of eluted crude undialysed, dialysed and purified SLP.

Many bands are present in the undialysed crude extract (Lane 1), representing other components of the S-Layer that exist in much lower concentrations. Once dialysed, the same proteins are present, possibly due to small spore sizes in the dialysis cassettes, whose main purpose is to remove urea from the sample (Lane 2). Purified samples appear in Lane 3 as two bands representing the HMW and LMW SLPs, appearing at 32kDa and 44kDa. The molecular weight ladder is given on the left in kDa.

Figure 4.8

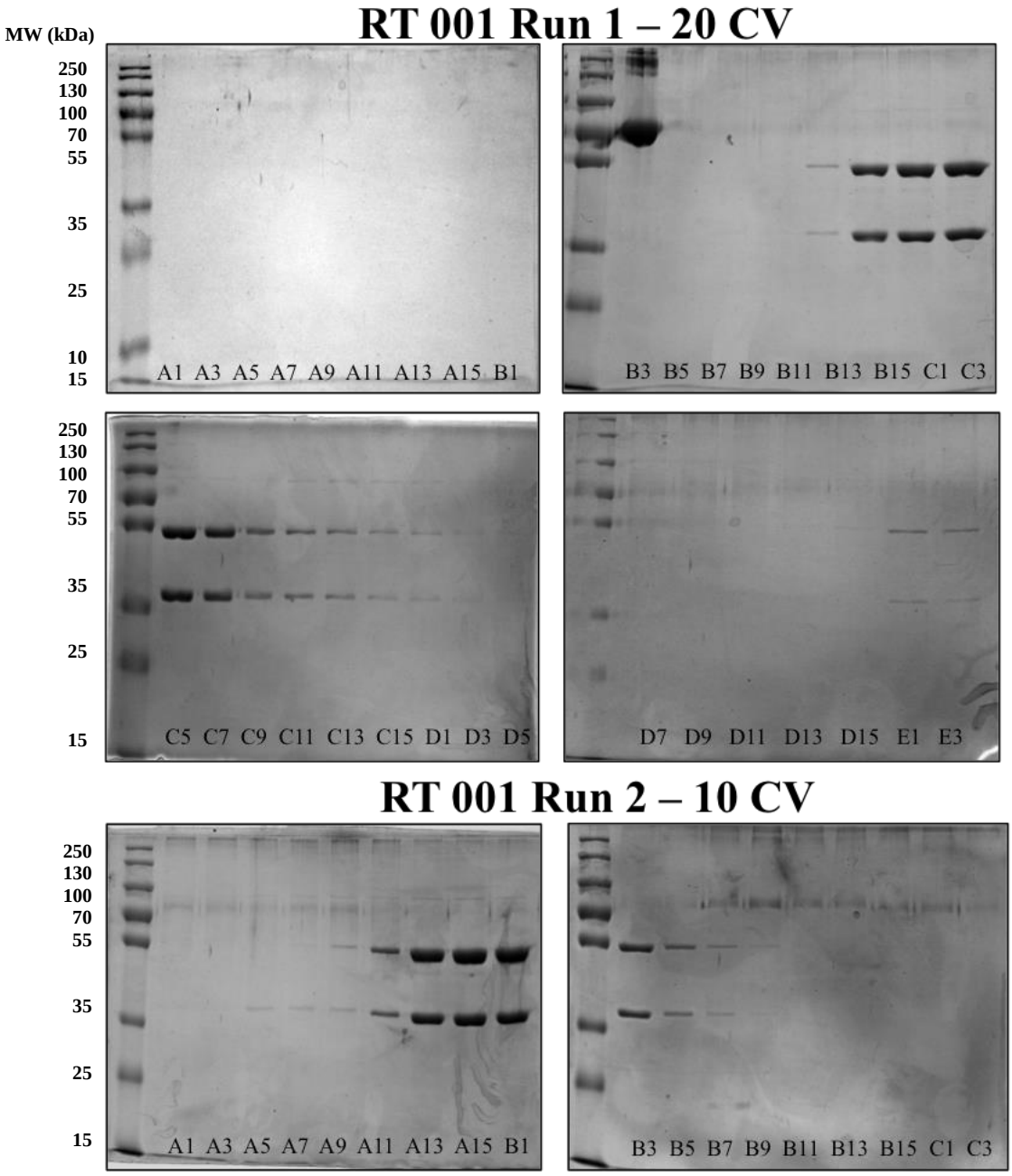


Figure 4.8 SDS-PAGE gels of eluted S-Layer from RT 001 at gradients of 20 and 10 CV.

The SLPs were eluted into fewer fractions at 10CV, but at higher concentrations. The time frame of the FPLC run was also reduced. The HMW and LMW SLPs are identified by two dark bands appearing around 32kDa and 44kDa respectively. The molecular weight ladder is given on the left in kDa.

Figure 4.9

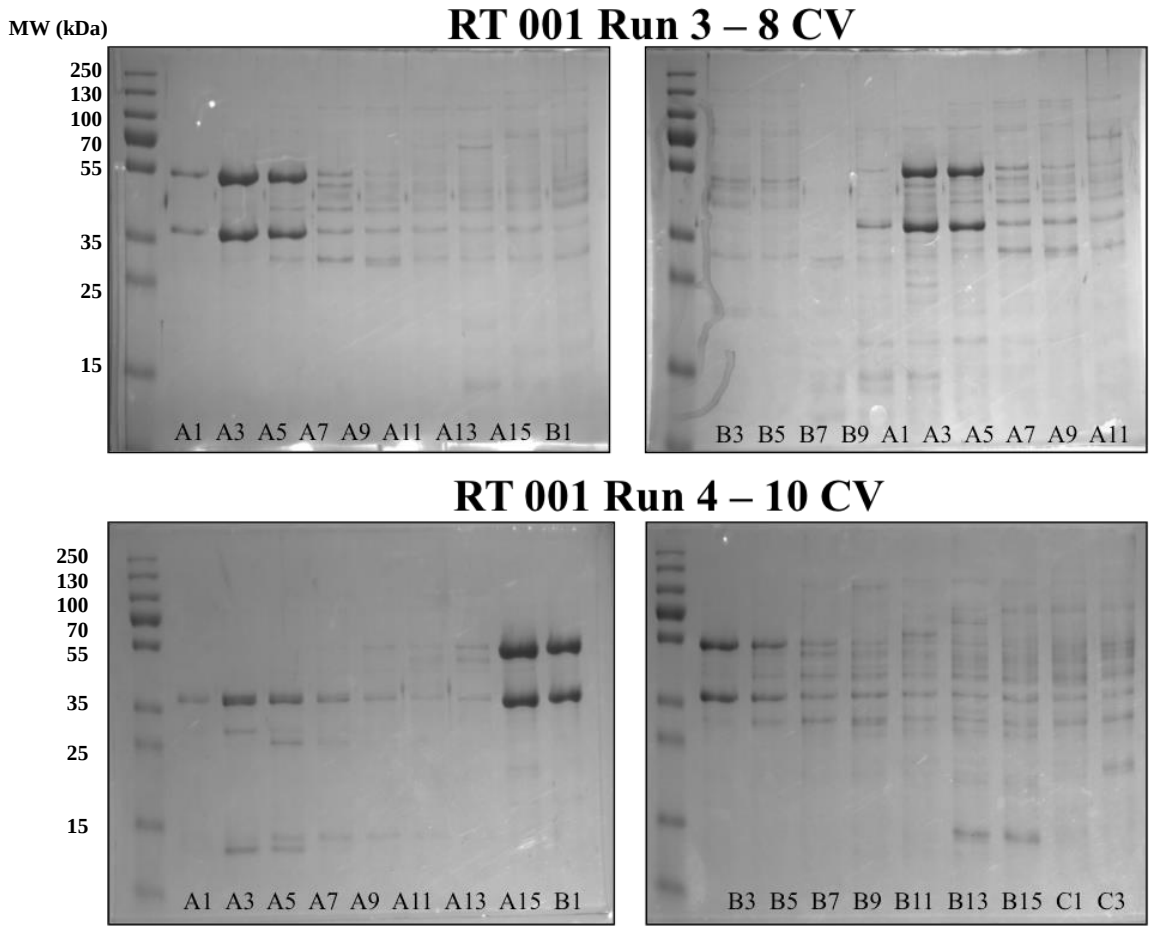


Figure 4.9 SDS-PAGE gels of eluted S-Layer from RT001 eluted at 8 and 10 CV.

Fractions are labelled A1-E3. Run 3, with a salt gradient of 8CV was found to be ineffective, with multiple bands appearing in the same fraction as the SLPs. Fractions A3 and A5 clearly show multiple bands with the darker SLPs. A gradient of 10CV was therefore used for all further RT 001 extractions. SLPs were then observed in fractions A15 to B3.The HMW and LMW SLPs are identified by two dark bands appearing around 32kDa and 44kDa respectively. The molecular weight ladder is given on the left in kDa.

Figure 4.10

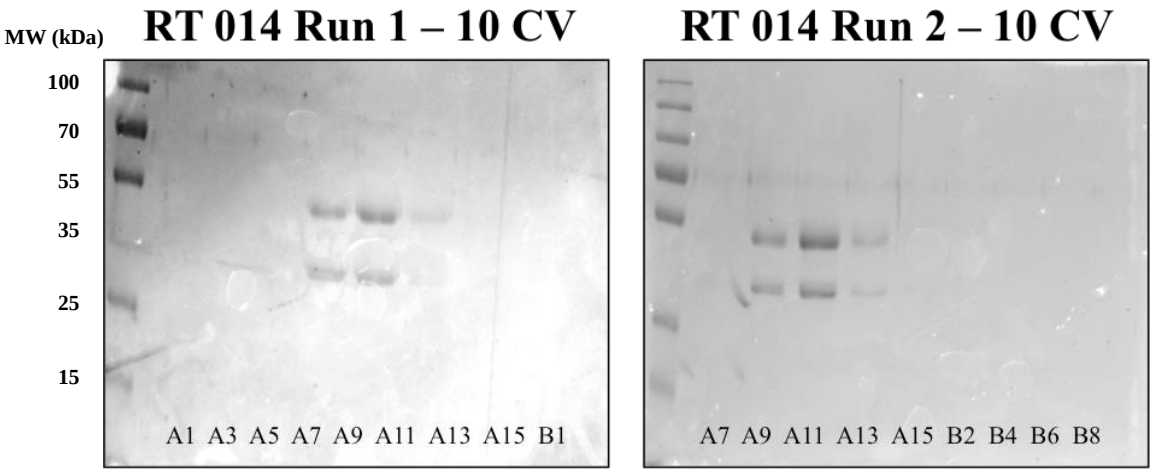


Figure 4.10 SDS-PAGE gels of eluted S-Layer from RT 014 at 10CV.

Fractions are labelled A1 – B8. Elution of RT 014 was carried out with an 8 CV gradient, both 0-0.3M NaCl in 20 mM Tris HCl pH 8.5. Two dark bands appearing around 32kDa and 44kDa respectively identify the HMW and LMW SLPs. SLP bands can be clearly seen at A7 – A11. Similar to RT 002 and 010, no other surface proteins were observed on these gels. The molecular weight ladder is given on the left in kDa.

Figure 4.11

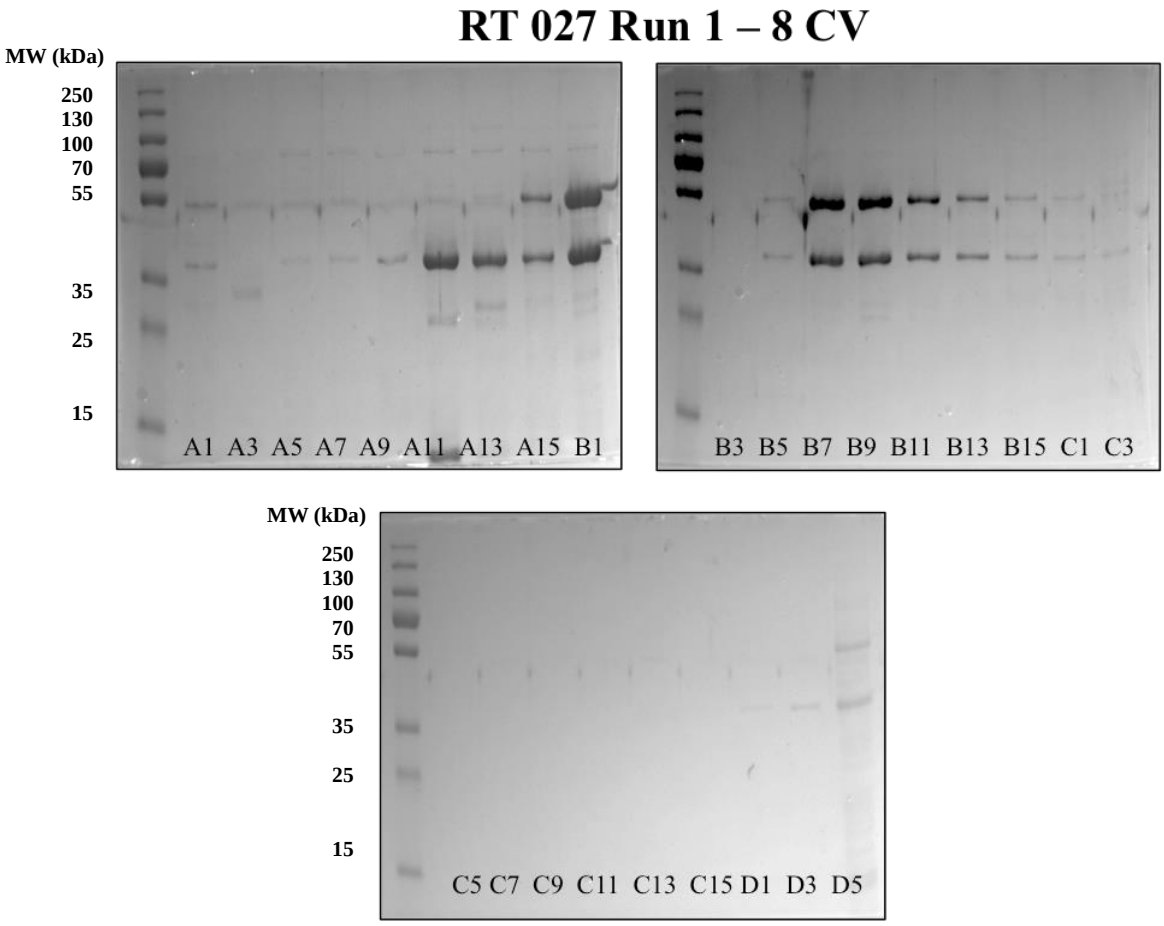


Figure 4.11 SDS-PAGE gels of eluted S-Layer from RT 027 at 8CV

SDS-PAGE gel images showing eluted fractions of crude S-Layer from RT 027. Fractions are labelled A1 – D5. Two dark bands appearing around 32kDa and 44kDa respectively identify the HMW and LMW SLPs. Similar to 002, 010 and 014 no other surface proteins were observed on these gels. Purified SLPs were observed in fractions B5 – C1. RT 027 SLPs were eluted in high amounts as observed by the many fractions displaying the double bands. The molecular weight ladder is given on the left in kDa.

Figure 4.12

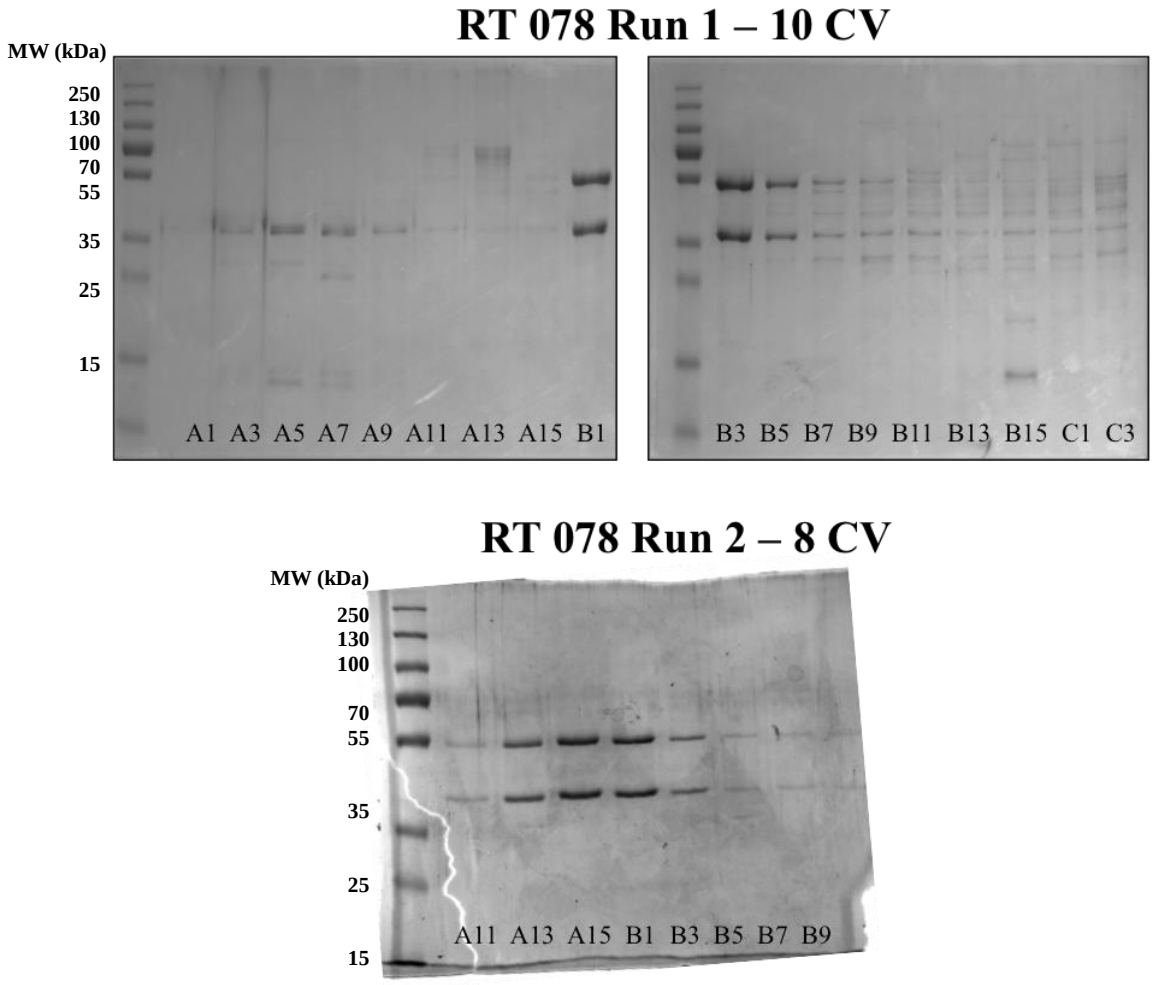


Figure 4.12 SDS-PAGE gels of eluted S-Layer from RT 078 at 10CV and 8CV.

SDS-PAGE gel images showing eluted fractions of crude S-Layer from RT 078. Fractions are labelled A1 – c3. Two dark bands appearing around 32kDa and 44kDa respectively identify the HMW and LMW SLPs. Adjusting the salt gradient to 8 CV resulted in purified SLPs being eluted. SLP bands were observed in fractions B1 – B5 at a gradient of 10 CV, and A11 – B5 at a gradient of 8 CV. The molecular weight ladder is given on the left in kDa.

4.2.4 SLPs from Ribotype 001 induce cell surface marker upregulation.

To confirm the immunostimulatory activity of the proteins, SLPs from RT 001 were used to stimulate BMDCs, with LPS acting as a positive control. Cells were collected and analysed for expression of important cell surface markers. Levels of CD80, CD86 and CD40 were all highly upregulated by SLPs from RT 001 (Figure 4.13). This confirmed what we had previously published for RT 001 (Ryan et al. 2011). LPS, a positive control, also highly upregulated these surface markers. TLR4 was more highly upregulated by LPS, with only a small increase observed for SLP stimulation. The results from Flow Cytometry analysis can be seen in Figure 4.13.

Figure 4.13

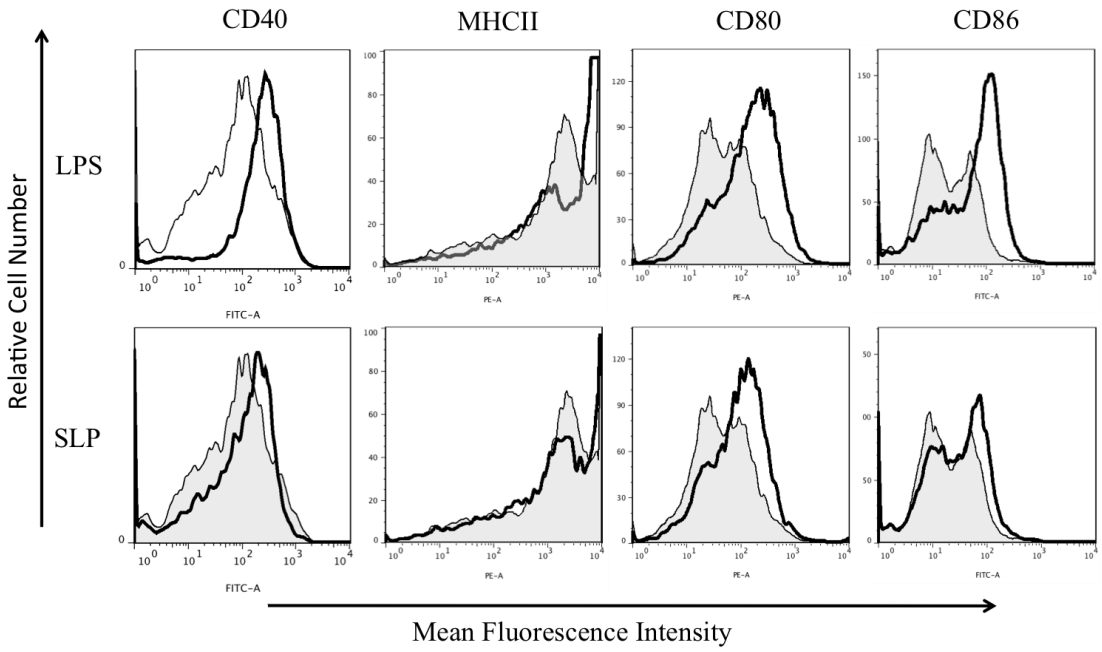


Figure 4.13 Cell surface marker expression by BMDC.

Cells were incubated with either LPS or SLP for 24 hours and were then washed and stained with antibodies specific for cell surface markers shown above. Histograms indicate fluorescence intensity, with a shift to the right indicative of upregulation. Shaded histogram represents control cells, dark line histograms represents cells stimulated by LPS/SLP. Results are indicative of three independent experiments.

4.2.5 SLPs from Ribotype 001 induce proinflammatory cytokine production.

BMDCs were stimulated with either LPS or SLP in order to detect cytokines induced by ELISA. In this experiment, irrelevant Surface proteins, eluted during FPLC, were also added to the cells, as a negative control. LPS and SLP were also added together in the same well, to examine whether the presence of both antigens compounded activation. Figure 4.14 shows that, as expected, IL-12p40, IL-6, IL-1β and anti-inflammatory IL-10 were all found to be upregulated in dendritic cells stimulated with SLP. Proteins eluted during FPLC, which were not SLPs (UN), did not induce any cytokine response from DCs, suggesting the SLPs are the immunodominant surface antigen. In addition, there were no synergistic or additive effects observed when SLP and LPS were added together. The results can be seen in Figure 4.14.

Figure 4.14

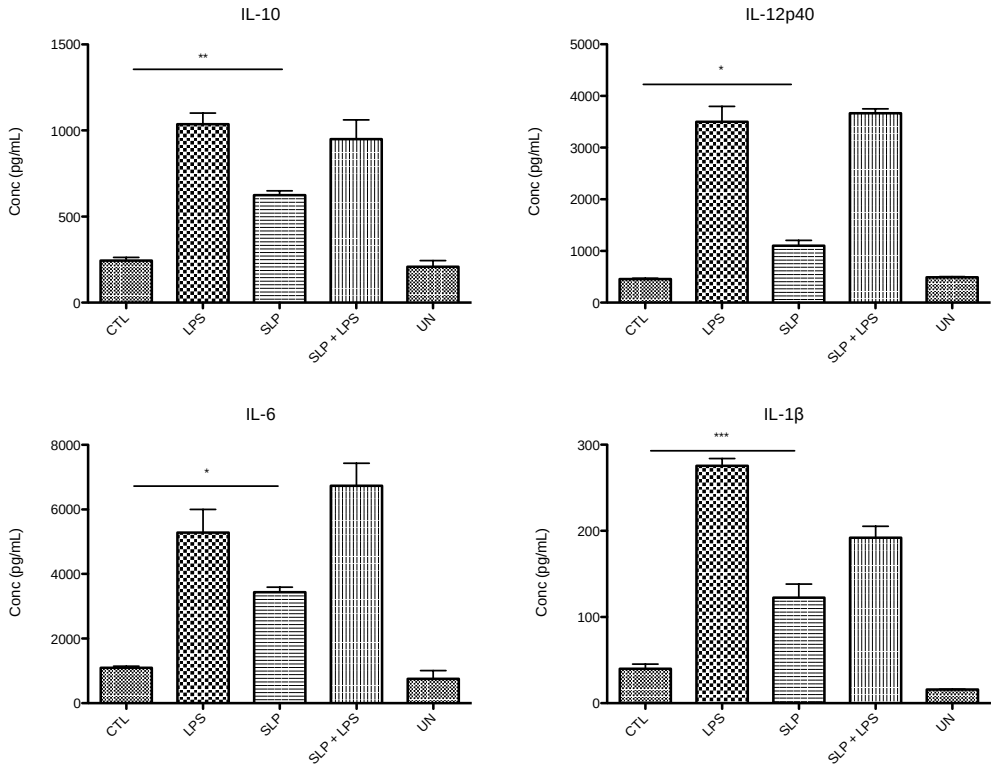


Figure 4.14 Cytokine production by BMDCs.

BMDC were stimulated with LPS (100ng/mL), SLP (20µg/mL) or irrelevant unknown proteins (UN) (20µg/mL). Supernatants were collected and cytokine levels were analysed by ELISA. SLP from RT 001 stimulated cytokine production, albeit to a lesser extent to LPS. Alternate cell wall proteins eluted during FPLC did not stimulate immune cells, nor did LPS and SLP together have a greater effect. The results show the mean (±SEM) for n = 3. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, determined by one-way ANOVA, followed by Newman-Keuls analysis comparing all groups. Results are indicative of three independent experiments.

4.2.6 SLPs from different ribotypes of *C. difficile* have differential effects on expression of cell surface markers on macrophages.

With the immunostimulatory properties of SLP confirmed by activation of dendritic cells *in vitro,* a comparative analysis of all available SLPs was preformed to determine if sequence variation resulted in different immune responses. Of all ribotypes in our dataset we chose four to represent each area of the phylogenetic tree. RT 001 was previously well characterised by our group, and others. From Chapter 3, we detected evidence of positive selection on the SLPs of RT 027, as well as RT 078, both being linked with hypervirulence. We also choose SLPs from RT 014, a strain not associated with hypervirulence, and on which we detected no evidence of positive selection. SLPs from these ribotypes were used for all further analysis. J774 murine macrophages were used for this initial screening, due to their low cost, ease of use and lack of need for animals. Furthermore, we have already detected the effect of RT 001 on macrophages (Collins et al. 2014). The macrophages were stimulated with SLP of various ribotypes, with LPS being used as a positive control. For each ribotype, SLPs were added at a concentration of 20µg/mL, and LPS at a concentration of 100ng/mL.

We examined the effects of the SLPs on the expression of cell surface markers important in antigen presentation and interaction with other immune cells. There was a strong upregulation of CD40, CD80 and MHC II expression on macrophages in response to LPS. SLPs from RT 027 and 078 all strongly upregulated these surface markers at levels comparable to LPS (Figures 4.15 – 4.18). RT 001 and 014 SLPs also increased expression of CD40, CD80 and

MHC II on macrophages, but to a lesser extent than LPS. Cells stimulated with SLPs from RT 027 and 078 induced a higher expression of CD40, CD80 and MHC II than either RT 001 or 014.

Figure 4.15

CD40

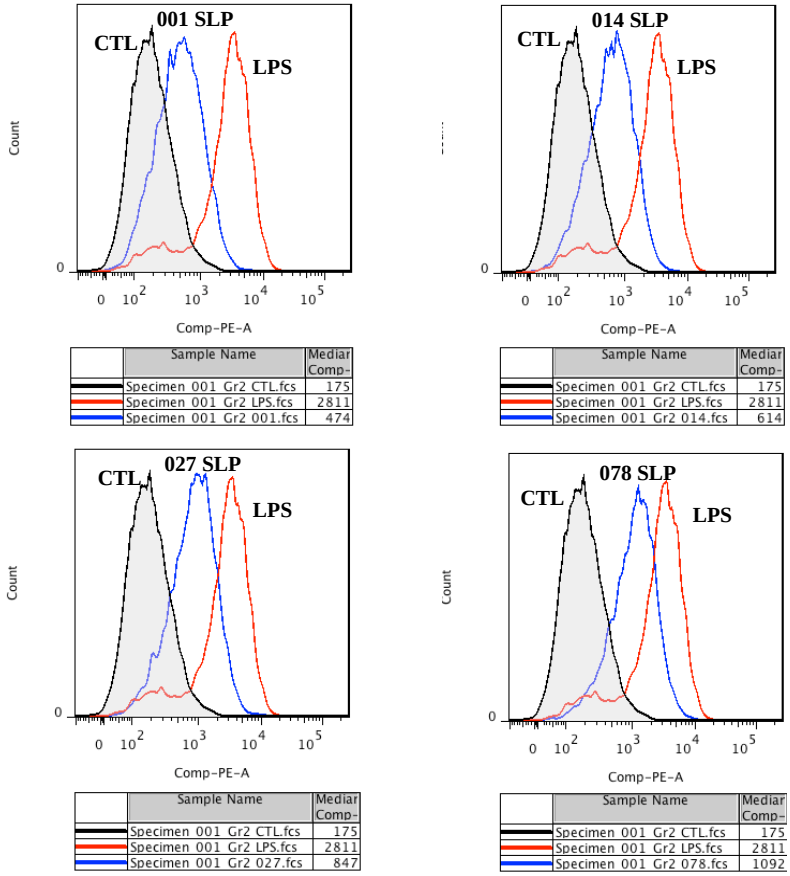


Figure 4.15 Surface marker CD40 expression on macrophages in response to SLPs and LPS

J774 macrophages were stimulated with SLP from ribotypes 001, 014, 027 and 078 (20µg/mL) for 24 hours. LPS (100ng/mL) was used as a positive control. Results show expression of the cell surface marker CD40. Histograms indicate fluorescence intensity, with a shift to the right indicative of upregulation. Control cells are shaded in grey, SLP-stimulated cells are labelled in blue, and LPS-stimulated cells are labelled in red. All SLPs induced induction of CD40, with SLPs from RT 027 and 078 being most potent. Results are indicative of three independent experiments.

Figure 4.16

CD80

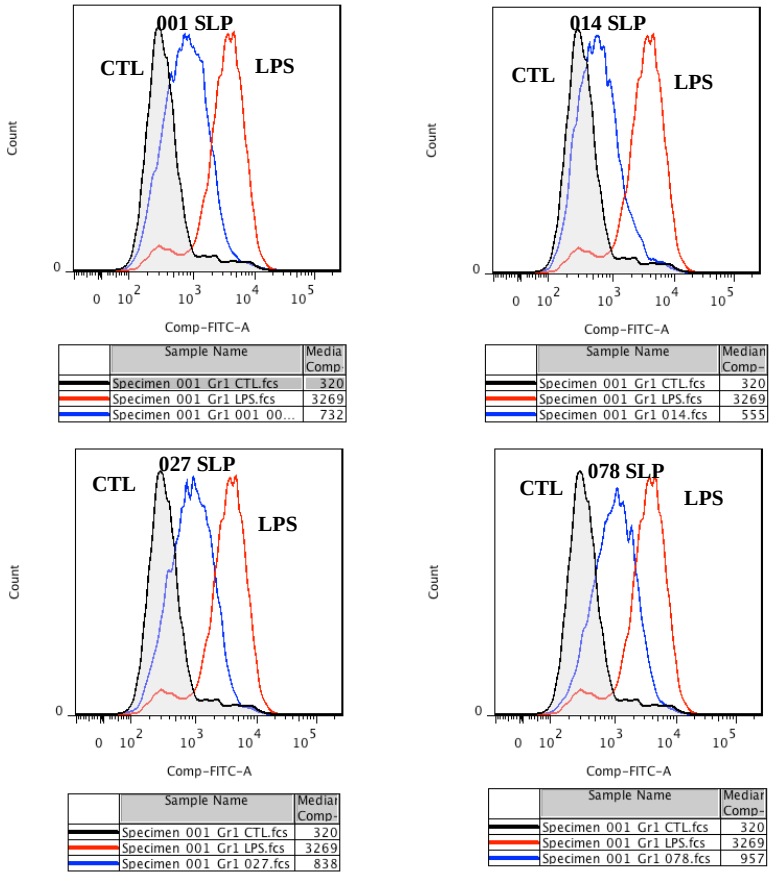


Figure 4.16 Surface marker CD80 expression on macrophages in response to SLPs and LPS

J774 macrophages were stimulated with SLP from ribotypes 001, 014, 027 and 078 (20µg/mL) for 24 hours. LPS (100ng/mL) was used as a positive control. Results show expression of the cell surface marker CD80. Histograms indicate fluorescence intensity, with a shift to the right indicative of upregulation. Control cells are shaded in grey, SLP-stimulated cells are labelled in blue, and LPS-stimulated cells are labelled in red. All SLPs induced induction of CD80, with SLPs from RT 027 and 078 being most potent. Results are indicative of three independent experiments.

Figure 4.17

CD86

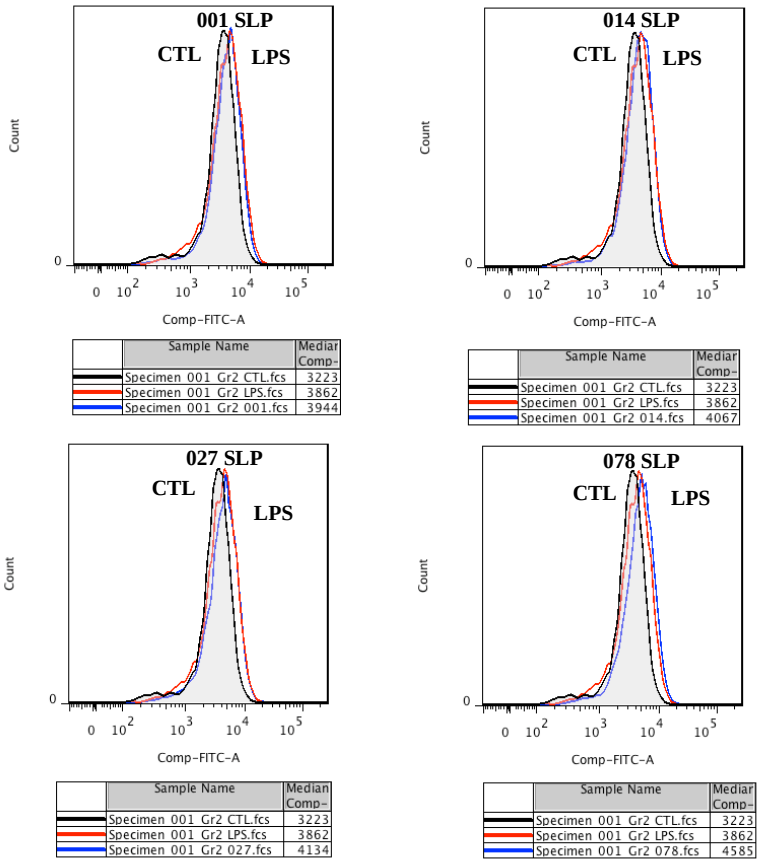


Figure 4.17 Surface marker CD86 expression on macrophages in response to SLPs and LPS

J774 macrophages were stimulated with SLP from ribotypes 001, 014, 027 and 078 (20µg/mL) for 24 hours. LPS (100ng/mL) was used as a positive control. Results show expression of the cell surface marker CD86. Histograms indicate fluorescence intensity, with a shift to the right indicative of upregulation. Control cells are shaded in grey, SLP-stimulated cells are labelled in blue, and LPS-stimulated cells are labelled in red. All SLPs induced marginal induction of CD80, with SLPs from RT 027 and 078 being most potent. Results are indicative of three independent experiments.

Figure 4.18

MHC Class II

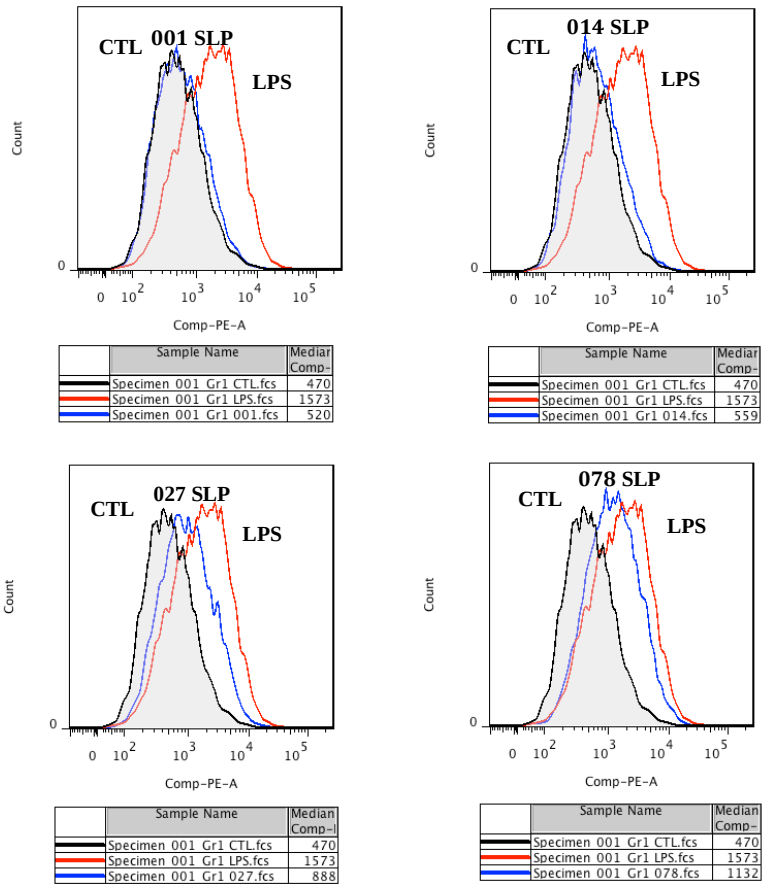


Figure 4.18 Surface marker MHC Class II expression on macrophages in response to SLPs and LPS

J774 macrophages were stimulated with SLP from ribotypes 001, 014, 027 and 078 (20µg/mL) for 24 hours. LPS (100ng/mL) was used as a positive control. Results show expression of the cell surface marker MHC Class II. Histograms indicate fluorescence intensity, with a shift to the right indicative of upregulation. Control cells are shaded in grey, SLP-stimulated cells are labelled in blue, and LPS-stimulated cells are labelled in red. All SLPs induced induction of MHC Class II, with SLPs from RT 027 and 078 being most potent. Results are indicative of three independent experiments.

4.2.7 SLPs from different ribotypes of *C. difficile* have differential effects on the production of cytokines by macrophages.

Our four ribotypes, 001, 014, 027 and 078, were used to stimulate J774 macrophages and the cytokine profile of these cells was analysed. J774 macrophages were stimulated with SLP and LPS (positive control) for 24hrs, and ELISA was used to measure the levels of IL-6, IL-10, IL-12p40, IL-23, IL-27 and TNFα. SLPs from each ribotype induced cytokine production at higher levels than control cells. SLPs from RT 001 and 014 consistently induced the lowest levels of cytokines relative to other ribotypes, with no significant difference being observed between the cytokine levels in response to either ribotype (Figure 4.19). SLPs from both ribotypes produced higher levels of TNFα ($p<0.001$), IL-12p40 ($p<0.05$), IL-6 ($p<0.001$), and IL-27 ($p<0.01$) relative to control cells.

Activation of macrophages with RT 027 SLPs consistently induced higher cytokine levels, with two-fold increases in of IL-12p40, TNFα, IL-6 IL-23, IL-27 and IL-10 observed relative to RT 001 (Figure 4.19; $p<0.001$). As previously stated, RT 027 has been dubbed a "hypervirulent" strain, and is associated with more severe disease progression and recurrence. SLPs from RT 078 also induced a strong inflammatory response, in the case of TNFα, producing higher levels than SLPs from 027 ($p<0.05$). Relative to RT 027, SLPs from RT 078 induced lower levels of IL-12p40 ($p<0.01$), IL-6 ($p<0.001$) and IL-10 ($p<0.001$) (Figure 4.19). These results show SLPs from RT 027 to be extremely potent activators of the inflammatory response. This increased potency may play a role in the observed severity of disease caused by RT 027.

Figure 4.19

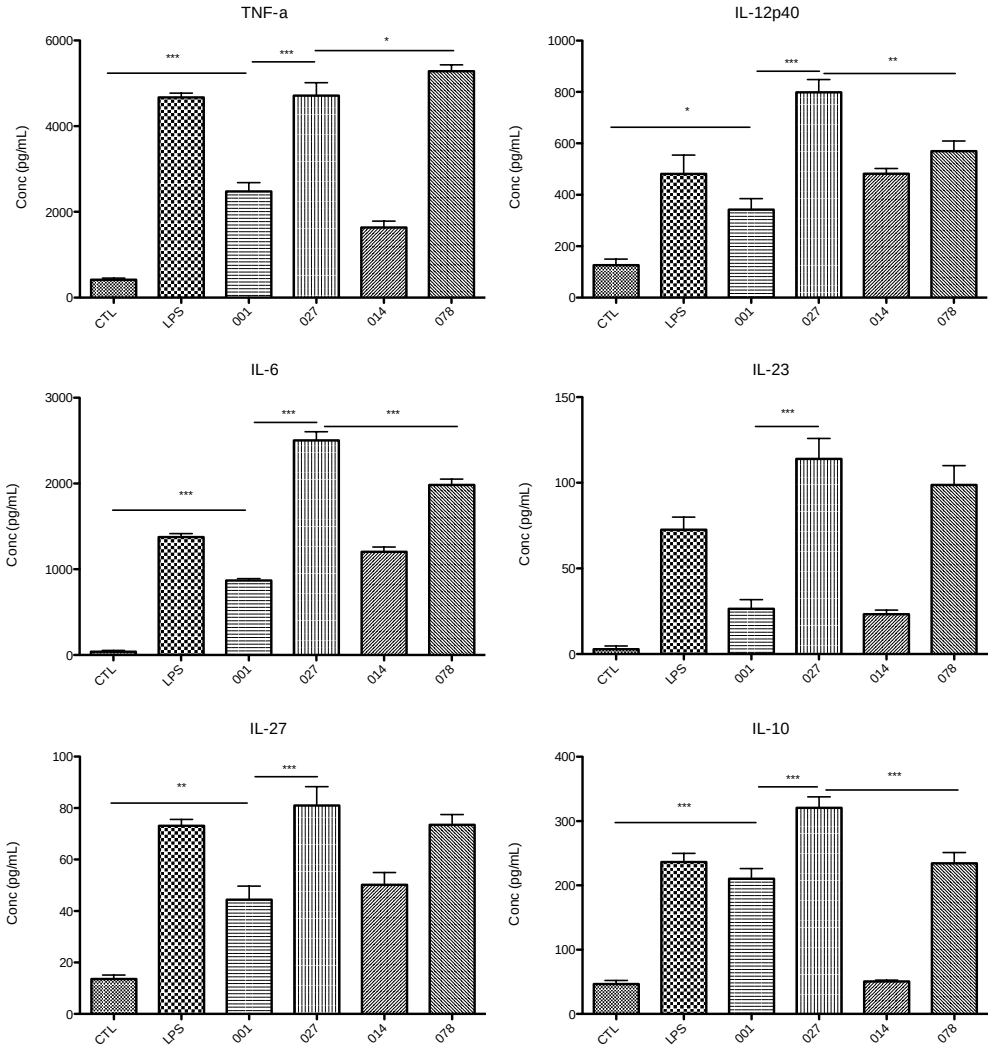


Figure 4.19 Cytokine production by macrophages in response to SLP and LPS.

J774 macrophages were stimulated with LPS (100ng/mL) and SLP (20µg/mL) from ribotypes 001, 027, 014 and 078 for a period of 24 hours. Supernatants were collected and cytokine levels were analysed by ELISA. SLPs from RT 027 and 078 consistently induce higher levels of cytokine production relative to RT 001 and 014. The results show the mean (±SEM) for n = 3. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, determined by one-way ANOVA, followed by Newman-Keuls analysis comparing all groups. Results are indicative of three independent experiments.

4.2.8 SLPs from different ribotypes of *C. difficile* have differential effects on the production of chemokines by macrophages.

The production of chemokines in response to SLPs from different ribotypes was also analysed. Levels of MIP-1α, MIP-2, MCP and RANTES were analysed in response to RT 001, 014, 027 and 078 SLPs (Figure 4.20). SLPs from RT 001 and RT 014 induced the lower chemokine levels, yet MIP-1α ($p<0.01$), MIP-2 ($p<0.001$), MCP ($p<0.01$) levels were significantly higher than control cells. RANTES levels were not significantly upregulated in response to SLPs from these ribotypes. SLPs from 027 and 078 consistently produced the highest levels of chemokines, comparable to LPS-stimulation. RT 027 induced significantly higher levels of MIP-1α ($p<0.001$), MIP-2 ($p<0.001$), MCP ($p<0.05$) and RANTES ($p<0.001$) compared to RT 001-stimulation. In the case of MIP-1α, MIP-2 and MCP, SLPs from RT 078 induced higher levels than 027 ($p<0.05$, $p<0.001$ and $p<0.001$ respectively). Conversely, SLPs from RT 014 were weak chemokine stimulators, with the exception of MIP-1α, where levels were similar to that of LPS. These results once again show the strong potency of SLPs from RT 027 and 078 to induce high levels of inflammatory molecules.

Figure 4.20

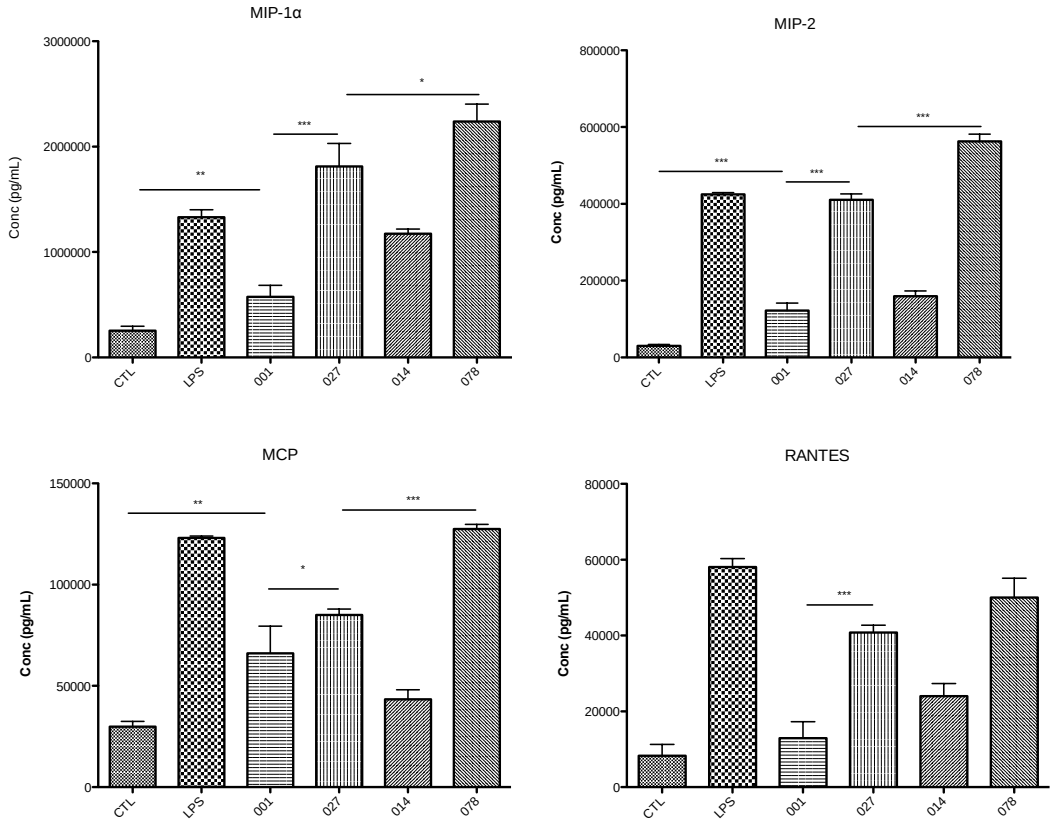


Figure 4.20 Chemokine production by macrophages in response to SLP and LPS.

J774 macrophages were stimulated with LPS (100ng/mL) and SLP (20µg/mL) from ribotypes 001, 027, 010, 014 and 078 for a period of 24 hours. Supernatants were collected and chemokine levels were analysed by ELISA. SLPs from RT 027 and 078 consistently induce higher levels of chemokine production relative to RT 001 and 014. The results show the mean (±SEM) for n = 3. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, determined by one-way ANOVA, followed by Newman-Keuls analysis comparing all groups. Results are indicative of three independent experiments.

4.2.9 SLPs from different ribotypes of *C. difficile* have differential effects on the production of cytokines by dendritic cells.

Given the importance of dendritic cells in the immune response, we next examined whether the differences we observed between ribotypes in macrophages also we evident in dendritic cells. Similar to the trend of the previous experiments, SLPs from RT 027 and 078 generally induced higher levels of proinflammatory cytokines than RT 001 and 014 (Figure 4.21). Two exceptions to this were IL-6 and IL-23. No significant difference was observed in IL-6 levels between RT 001 and 027, yet both were significantly higher than control ($p<0.001$). Levels of IL-23 were higher in SLP RT 001-stimulated cells compared to 027- and 078 SLP stimulation ($p<0.001$). This is the only instance of RT 001 inducing the highest response relative to SLPs from the other ribotypes. 027 SLPs induced high levels of the immunosuppressive IL-10 in addition to proinflammatory cytokines ($p<0.001$). Chemokine levels also followed a similar trend. There is an almost two-fold increase in levels of MIP-1α, MIP-2 and MCP in 027-stimulated cells relative to RT 001 ($p<0.05$ for MIP-1α, MIP-2, $p<0.01$ for MCP). Both RT 027 and 078 induced the highest levels of chemokine production, at levels higher or equal to LPS stimulation. RT 001 and 014 induced levels higher than control cells, but lower than LPS (MIP-1α $p<0.01$, MIP-2 $p<0.05$ and MCP $p<0.001$). The results can be seen in Figure 4.21.

Figure 4.21

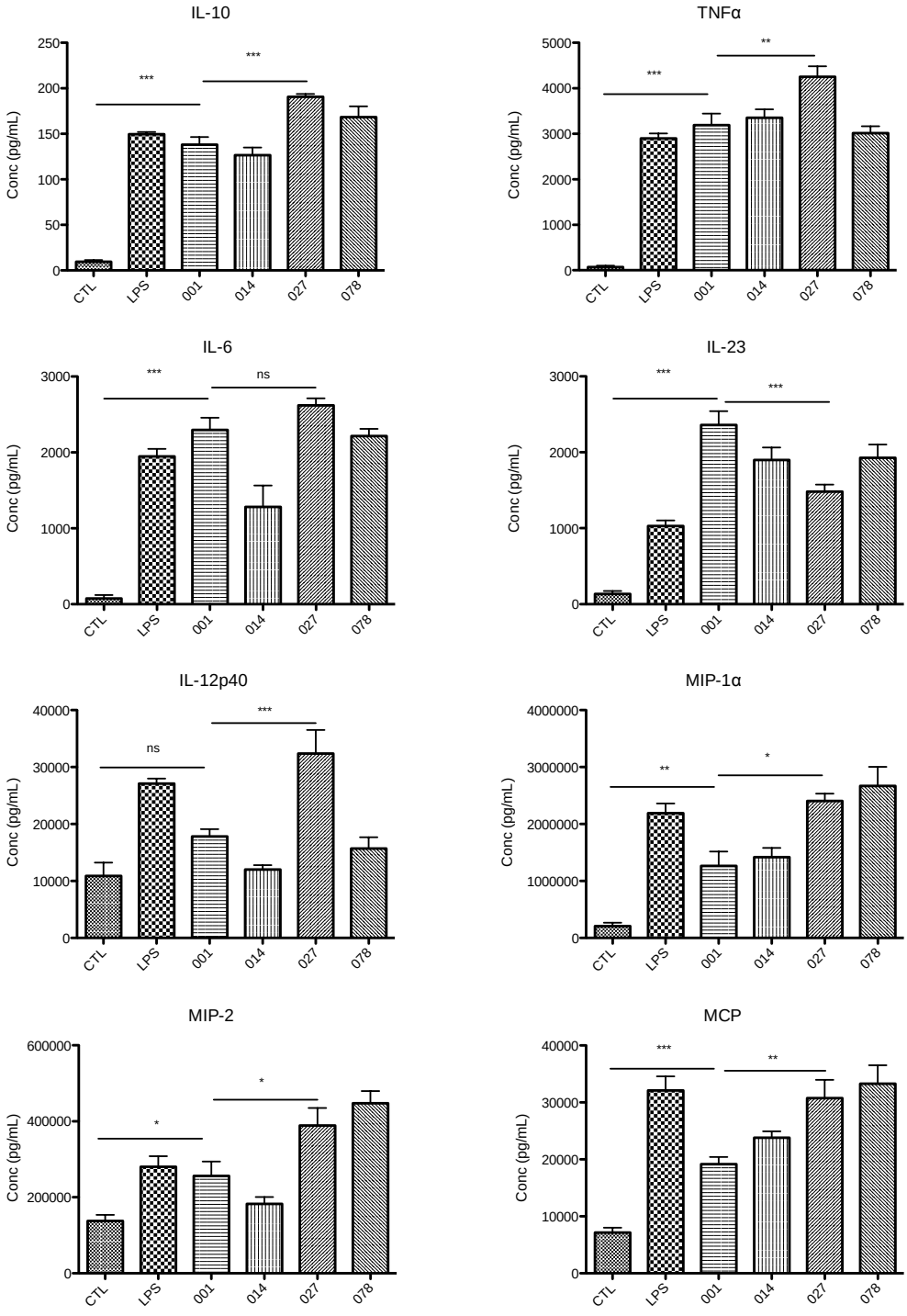


Figure 4.21 Differential effects of SLPs on BMDCs

BMDCs taken from BALB/c mice were stimulated with LPS (100ng/mL) and SLP (20µg/mL) from ribotypes 001, 027, 014 and 078 for a period of 24 hours. Supernatants were collected and cytokine levels were analysed by ELISA. The results show the mean (±SEM) for n = 3. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, determined by one-way ANOVA, followed by Newman-Keuls analysis comparing all groups. Results are indicative of three independent experiments.

4.2.10 SLPs from different ribotypes induce variable levels of phagocytosis in macrophages.

A key difference in the infection caused by different ribotypes of *C. difficile* is the effective clearance of the infection. We therefore next examined the ability of the SLPs from RT 001, 014, 027 and 078 to induce phagocytosis in macrophages. FITC-labelled polystyrene beads were added to cells that were stimulated with SLPs for 24hrs. Cells stimulated with LPS were used as a positive control. Cells were analysed at 30 minutes, 1 hour and 2 hours by flow cytometry to determine the rate of phagocytosis of the beads in the presence of SLPs. Cells positive for FITC indicated phagocytosing cells. Cells not stimulated with SLPs but exposed to the polystyrene beads were used as a negative control.

Control cells had a low level of phagocytosis at 30 minutes; less than 5% of the population contained beads (Figure 4.22 and 4.23). After 1 hour, this had increased marginally to 7% and by 2 hours, 25% of the population of cells had phagocytosed the beads. Phagocytosis was significantly increased for LPS-stimulated cells, with 17%, 25% and 53% of the population phagocytosing beads at the 30 minute, 1 hour and 2 hour time points respectively. SLPs from RT 001 induced phagocytosis, but at a decreased level when compared to LPS-stimulated cells at each time point, and RT 014 SLPs resulted in an even poorer response with only 22% of cells phagocytosing at 2 hours. RT 027 and 078 SLP-treated cells displayed a similar level of phagocytosis to the LPS controls at 30mins but phagocytosis was slightly lower in these groups when compared to LPS at 1 hour and 2 hours. RT 078 induces higher levels of phagocytosis than RT 027 at one hour, but this is reversed at two hours, with the phagocytosing population at

almost 39% for RT 027 SLPs and 34% for RT 078. At one hour it has a higher level than RT 027 SLPs, but again at two hours, phagocytosis levels are lower than that on RT 001 and 027. These results did not follow the previous trend of a RT 001/014 versus RT 027/078 axis. RT 001 and 027 were the strongest initiators of phagocytosis, with RT 078 less potent and RT 014 the weakest. The results of the phagocytosis can be seen in Figures 4.22 and 4.23.

Figure 4.22

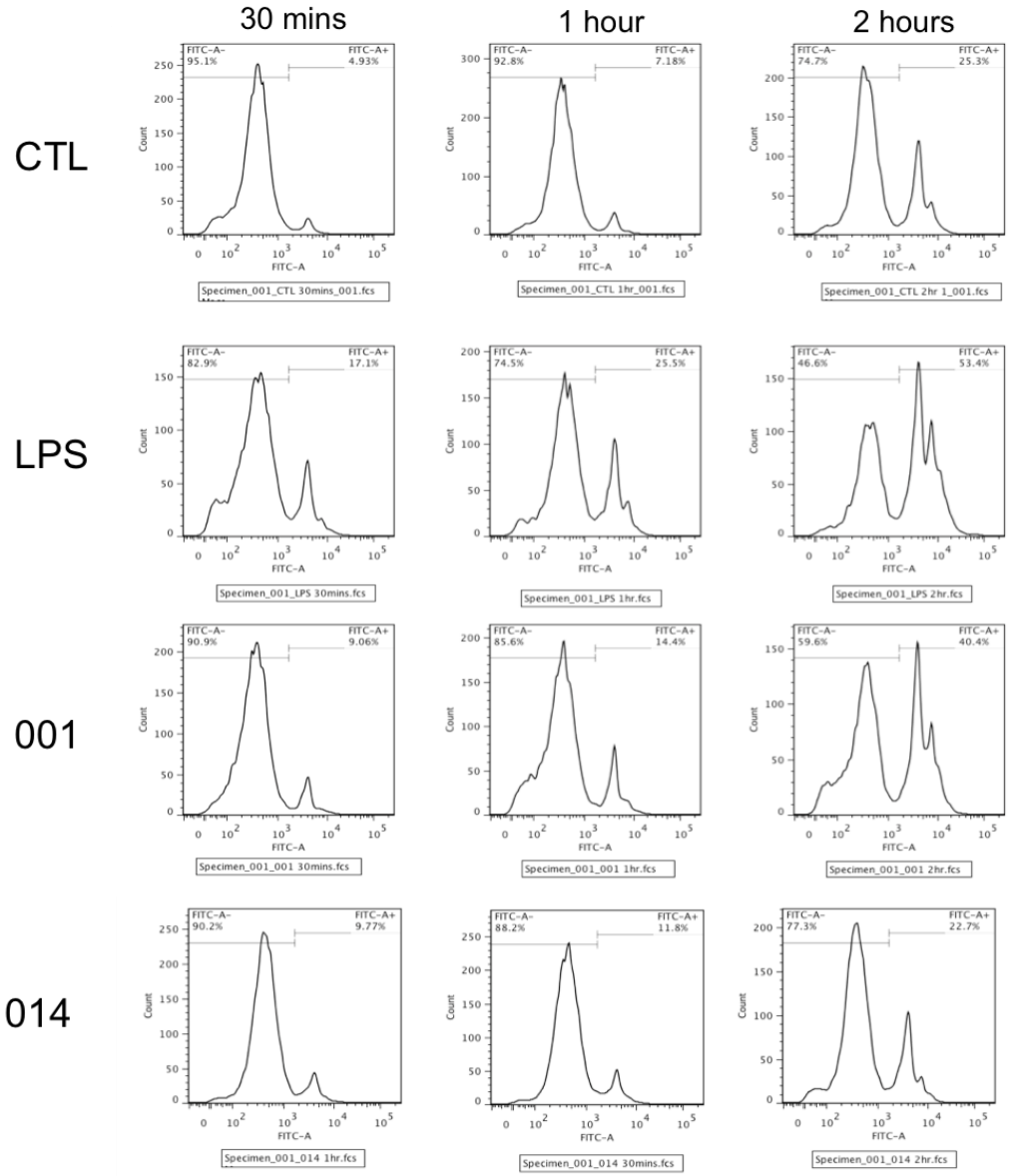


Figure 4.22 SLPs from ribotypes 001 and 014 induce phagocytosis in macrophages.

Phagocytosis of FITC-labelled fluorescent beads by J774 macrophages in the presence of LPS and SLP ribotypes 001, and 014. Cells were stimulated for 24hrs with SLP (20µg/mL) or LPS (100ng/mL), and beads were then added (10 beads/cell) for 30 minutes, 1 hour, and 2 hours. Percentage of the cell population was measured by the quantity of FITC signal from the cells using flow cytometry. SLPs from RT 001 induced higher levels of phagocytosis that those of RT 014. Results are indicative of three independent experiments.

Figure 4.23

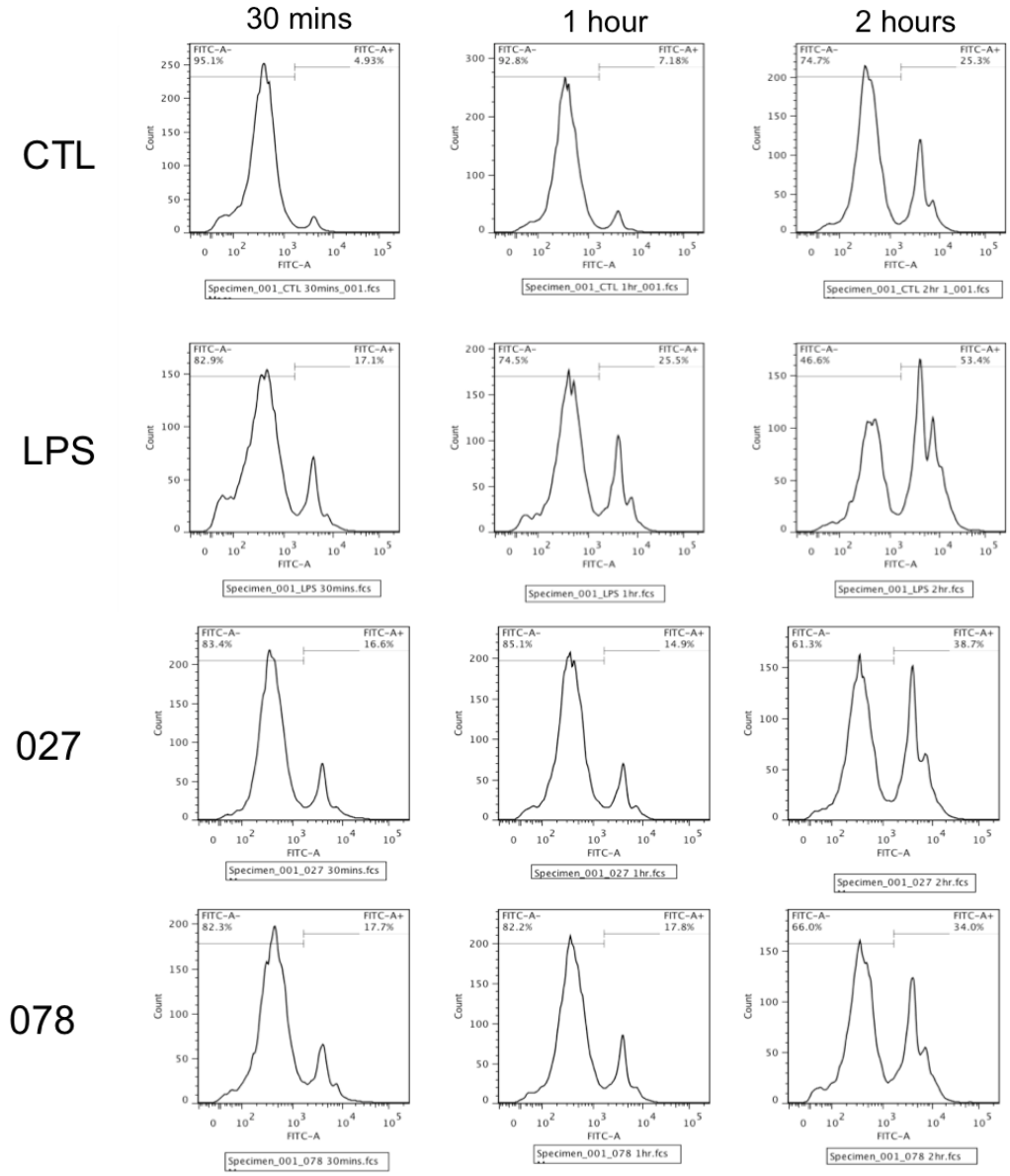


Figure 4.23 SLPs from ribotypes 027 and 078 induce phagocytosis in macrophages.

Phagocytosis of FITC-labelled fluorescent beads by J774 macrophages in the presence of LPS and SLP ribotypes 027, and 078. Cells were stimulated for 24hrs with SLP (20µg/mL) or LPS (100ng/mL), and beads were then added (10 beads/cell) for 30 minutes, 1 hour, and 2 hours. Percentage of the cell population was measured by the quantity of FITC signal from the cells using flow cytometry. SLPs from RT 027 induced similar levels of phagocytosis that those of RT 001. SLPs from RT 078 induced lower levels. Results are indicative of three independent experiments.

4.2.11 SLPs from ribotype 027 induce a more potent inflammatory response relative to RT 001.

We next wanted to more closely compare the SLPs of RT 001 and 027. They share phylogenetically similar sequences, yet positive selection was predicted on the SLPs of RT 027 and not RT 001. BMDCs were isolated from the bone marrow of BALB/C mice and culured for seven days in the presence of GMCSF. Cells were stimulated with SLPs from RT 001 and 027, with LPS being used as a positive control. Cytokine levels in response to the SLPs or LPS were analysed by ELISA. SLPs from 001 induced higher levels of IL-23 compared to RT 027 SLPs (Figure 4.24, $p<0.01$). This cytokine, along with IL-6, is known to be important for Th17 responses, involved in clearing extracellular pathogens. Proinflammatory cytokines IL-6 ($P<0.05$), IL-12p40 ($p<0.001$), TNFα ($p<0.001$) and IL-1β ($p<0.05$) were all secreted at higher levels in response to RT 027 SLPs relative to RT 001, indicating a more severe inflammatory response to this ribotype. High levels of IL-10 were again detected in response to SLPs from RT 027 ($p<0.001$). IL-27 is known to augment IL-10 production in T cells, and its increased levels in response to RT 027 SLPs ($p<0.001$) suggests an important role for IL-10 in infection with this ribotype.

Figure 4.24

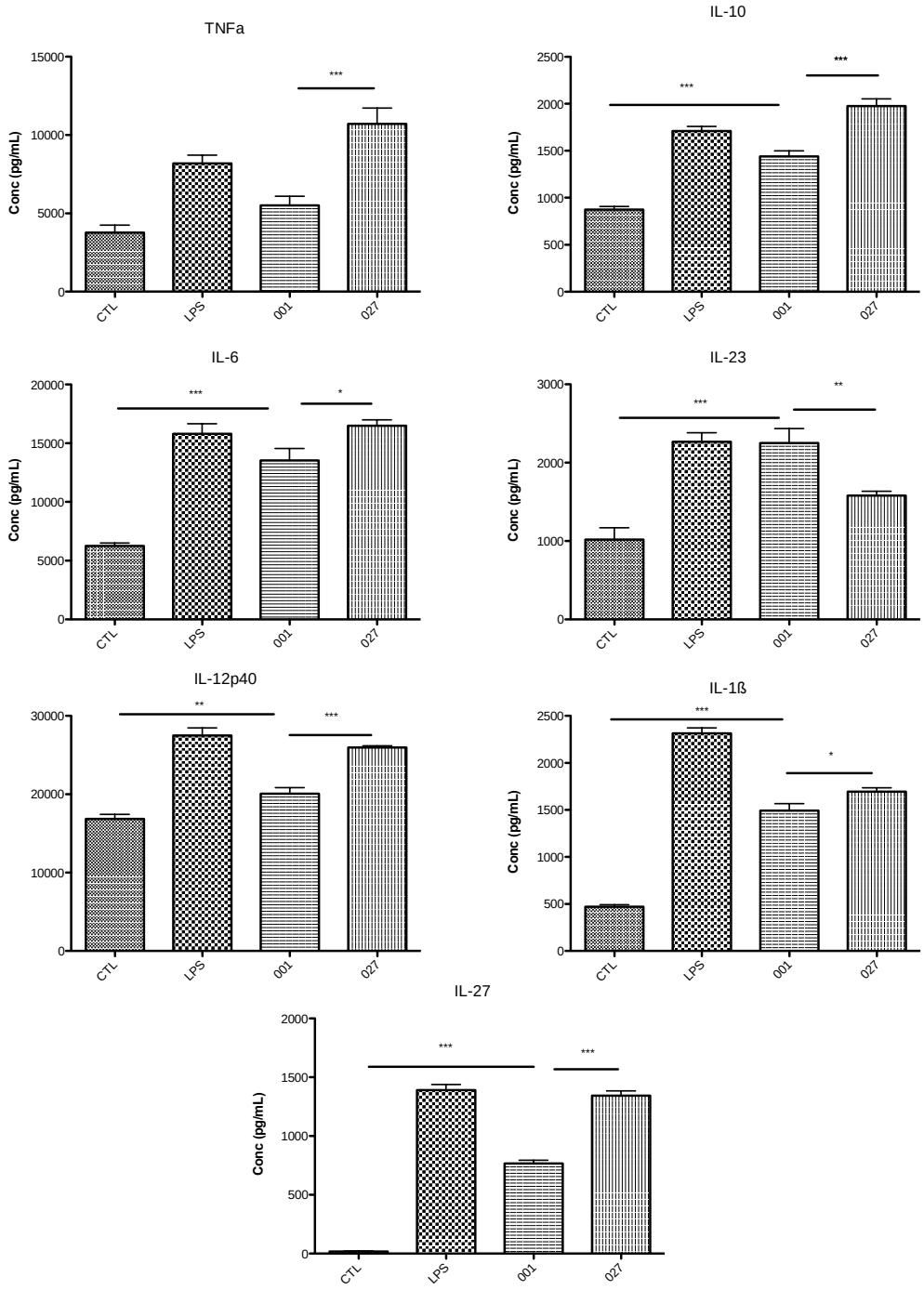


Figure 4.24 Comparisons of SLP responses from RT 001 and 027 on BMDCs.

BMDCs taken from BALB/c mice were stimulated with LPS (100ng/mL) and SLP (20µg/mL) from ribotypes 001, 027, 014 and 078 for a period of 24 hours. Supernatants were collected and cytokine levels were analysed by ELISA. With the exception of IL-23, SLPs from RT 027 induced higher cytokine levels than RT 001. The results show the mean (±SEM) for n = 3. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, determined by one-way ANOVA, followed by Newman-Kuels analysis comparing all groups. Results are indicative of three independent experiments.

4.3 DISCUSSION

The focus of this chapter was to determine if positive selection on the SLPs of *C. difficile* had an impact on their ability to activate an immune response. Evidence of positive selection was detected in several strains, and of particular interest were RT 078 and 027. These ribotypes have previously been described as hypervirulent (Marsh et al. 2012; Goorhuis et al. 2008), and the presence of positive selection on the SLPs of these strains indicates a potential role for these proteins in hypervirulence and severe disease. For comparison, we also choose two strains that did not show evidence of positive selection. RT 001 has been previously characterised by our group and others (Ryan et al. 2011; Collins et al. 2014; Eidhin et al. 2006), and was known to induce an inflammatory response. We also choose RT 014 as there was also no evidence for positive selection on this strain, but has previously been reported to be common in humans (Janezic et al. 2012). A European study also determined ribotypes 001, 014, 027 and 078 as the most prevalent in a hospital-based survey (Bauer et al. 2011).

The ability of SLPs to induce macrophages to produce cytokines and chemokines is an important indicator of how potently they activate the immune system. Our group has previously shown that SLPs from RT 001 activate the immune response, specifically through TLR4, and induce clearance responses in macrophages (Ryan et al. 2011; Collins et al. 2014). A comparative study of SLPs based of positive selection has not previously been performed. We found that SLPs from four ribotypes in our dataset activated macrophages and DCs to produce proinflammatory cytokines. The profile of cytokine production was

variable across ribotypes, with SLPs from some strains proving more potent than LPS, a positive control for TLR4 activation.

We observed up-regulation of cell surface molecules such as CD80, CD86, MHC II, and CD40 in macrophages in response to SLP from RT 001, 014, 027 and 078. Cells stimulated with SLPs from RT 027 and 078 expressed higher levels of CD80, CD40 and MHC II than those induced by RT 001 and 014. These molecules are critical in recruiting and interacting with cells of the adaptive immune system. CD80/86 acting as co-stimulatory molecules for T cell receptor (TCR) activation, inducing stimulating proliferation and cytokine production (Lanier et al. 1995). CD40 interacts with CD40L on T helper cells to further activate APCs and increase co-stimulatory signals (Grewal & Flavell 1996), and MHC Class II is responsible for displaying antigen to cells of the adaptive immune system (König et al. 1992). Higher levels of these molecules in response to RT 027 and 078 suggest that these strains induce a more potent inflammatory response relative to SLPs from RT 001 and 014. These results suggest that SLPs undergoing positive selection induce higher levels of surface molecule expression for the proliferation and activation of T cells. This was not expected, as our hypothesis for positive selection on the SLPs of *C. difficile* was to act as a molecular cloak, shielding itself from an inflammatory response and causing more persistent and severe disease. Therefore we expected a weaker response, however, we see a weaker response to SLPs on which no positive selection was detected. Perhaps this weaker response is sufficient for clearance without exacerbating inflammation. In the case of the positively selected SLPs, it is possible that increased inflammatory response benefits the bacteria, inducing

more tissue damage to allow greater binding to the gut, if the pathogen can avoid clearance by the host immune system.

While upregulation of cell surface molecules are important in inducing an adaptive immune response, the local cytokine environment can influence the nature of that adaptive response. SLPs from RT 001, 014, 027 and 078 all activate immune cells to produce proinflammatory cytokines. Variation in the response of macrophages and DCs to SLP was observed depending on the ribotype from which the SLP were purified. RT 001 has been frequently used in the past to show immune reactivity, and this was observed in this study. However it proved to induce one of the weakest responses relative to other ribotypes. SLPs from RT 027 consistently induced high levels of proinflammatory cytokine secretion in macrophages, exhibiting up to twofold increases in cytokine production in response to proteins from this ribotype.

High levels of proinflammatory cytokines TNFα, IL-12p40, IL-6 and IL-23 were all observed in response to SLPs from RT 027 and 078. TNFα is a particularly important cytokine in inflammatory bowel disease, with increased levels having previously been associated with ulcerative colitis (UC), and TNFα exhibiting a pathogenic effect (Neurath et al. 1997; Braegger et al. 1992). In addition many anti-TNF therapies are being explored for the treatment of UC and Crohn's disease (Lv et al. 2014). IL-12p40 is also associated with diseases such as UC (Kullberg et al. 1998). While IL-6 and IL-23 are associated with induction of a protective Th17 response (discussed below), the presence of high levels of IL-10 may negate the protective effects.

SLPs from RT 001 and 014 induced lower levels TNFα, IL-12p40, IL-6 and IL-23 relative to SLPs from RT 027 and 078. Once again, we see SLPs under positive selection inducing high levels of inflammation. This inflammatory state must be beneficial for the pathogen, as increased tissue damage may allow the bacteria to invade deeper into the gut, resulting in persistent disease. The positively selected SLPs of hypervirulent strains also induce high levels of chemokines, necessary for recruiting neutrophils and cells of the adaptive immune system. RT 027 and 078 SLPs induced up to two-fold higher levels of MIP-1, MIP-2, MCP and RANTES in macrophages. These chemokines are essential in recruiting inflammatory cells to the site of infection, and can directly affect the nature of the immune response. These SLPs may recruit many inflammatory cells to the site of infection by stimulating these chemokines, exacerbating inflammation. Decreased chemokine levels have previously been linked to aggravated disease (Kurihara et al. 1997), yet high levels are detected in response to the SLPs under positive selection. It is possible that these ribotypes are recruiting inflammatory cells to the site of infection, but the cytokine environment blocks an effective response, resulting in more severe disease symptoms.

Dendritic cells play an essential role in recognising pathogens and stimulating an adaptive immune response, and when stimulated with SLPs, showed a similar response to that of macrophages. Interestingly levels of IL-23 were higher in RT 001 SLP-stimulated DCs compared to 027 SLP-stimulated cells. Levels of IL-6 were also similar to that of 027 SLP-stimulation. Both IL-6 and IL-23 are required to drive a Th17 response. This T cell phenotype plays an important role

in the clearance of extracellular pathogens in the gut (Khader et al. 2009; Zhang et al. 2009b; Blaschitz & Raffatellu 2010) and their high levels is response to RT 001 SLPs may hint at a role for Th17 cells in *C. difficile* clearance. Th17 cells have not previously been associated with *C. difficile* clearance, and suggests a novel mechanism by which the immune system can clear the pathogen. If the positive selection in the SLPs of 027 and 078 results in decreased Th17 differentiation, these strains may be able to avoid clearance while inducing higher levels of inflammation. SLPs from RT 027 and 078 induced high levels of TNFα and IL-10. Again we see strong induction of damage-inducing TNFα and immune-suppressive IL-10 by SLPs undergoing positive selection. This suggests evolution of the SLPs to induce an inflammatory state, while suppressing the immune systems attempt to clear the bacteria with IL-10.

High levels of IL-10 were also observed in macrophages in response to SLPs from 027 and 078. This cytokine exhibits immunosuppressive effects, so high levels in response to strains inducing severe inflammation was not expected. As high levels of IL-10 were produced by SLPs under positive selection, it is possible that *C. difficile* is modulating the immune response to dampen any clearance mechanisms and evade destruction. We have previously shown RT 001 SLPs to effectively induce clearance responses to resolve infection, and these SLPs produce low levels of IL-10 relative to the other ribotypes. IL-10 has been shown to block Th17, Th1 and Th2 development (O'Garra & Vieira 2007; Maynard & Weaver 2008), and other organisms have been shown utilise IL-10 to avoid clearance (Xavier et al. 2013; Redford et al. 2011), and it is possible that positive selection on the SLPs modulates immune cells to obtain a similar result.

RT 027 SLPs also induce high levels of IL-27. Together with IL-6, this cytokine has been shown to promote IL-10 production in T helper cells (Stumhofer et al. 2007). These cytokines, along with high levels of T cell-activating surface molecules, may drive an adaptive response producing large amounts of IL-10, increasing inflammation and tissue damage, while impairing clearance.

The ability of macrophages to phagocytose pathogens is an important determinant in clearance of disease (Taylor et al. 2010), and SLPs from our four ribotypes induced phagocytosis in macrophages. Once again, the rate at which cells phagocytosed varied between ribotypes. SLPs from RT 001, 027 and 078 induced high levels of phagocytosis, with 014 being less potent. SLPs from RT 001 and 027 induced similar levels, with RT 078 being less potent. This suggests an impaired ability of RT 014 to activate the immune response through its SLPs. RT 078 and 027 have been well documented as hypervirulent strains (Stabler et al. 2009; Goorhuis et al. 2008), and we have detected positive selection on the SLPs of both 027 and 078. As selection infers greater survival benefit on the organism, it might be expected that SLPs under positive selection induce lower levels of phagocytosis. The observed increase was therefore not expected. It has previously been shown that phagocytosed *C. difficile* spores can readily survive inside the phagosomes of macrophages (Paredes-Sabja et al. 2012). This increased phagocytosis may be due to an increased inflammatory state, but as SLPs from RT 001 also strongly induce phagocytosis, this does not explain the lack of clearance in response to RT 027 and 078. It is likely that phagocytosis is not as important a determinate in disease severity as other factors, such as cytokine secretion.

The SLPs of *C. difficile* are only one component in the mechanisms inducing disease in the gut, and care must be taken in interpreting the response of isolated cell populations *in vitro* to purified SLPs. To fully understand the significance of the differences observed in immune response between *C. difficile* of different ribotypes, infection models using the whole bacteria are required. Animal models will help determine if the results in this chapter translate to *C. difficile* infection. This study clearly highlights the SLPs ability to induce an immune response, and that variations in the immune response exist between strains. In particular, there appears to be a role for positive selection of the SLPs in modulating the host immune system. It seems the ability of *C. difficile* to induce a Th17 promoting response through the SLPs, aids clearance, while positive selection seems to result in SLPs producing higher amounts of IL-10, thus suppressing the immune system and inhibiting an effective T cell response. Due to the observed strong inflammatory response to SLPs from RT 027, the presence of positively selected sites on its *slpA* gene, and its close phylogenetic relationship with the SLPs from RT 001, further exploration of the comparison between these two ribotypes may give further information on the nature of this hypervirulence, and its relation to positive selection.

Chapter 5

A comparison of *Clostridium difficile* ribotypes 001 and 027 *in vivo*.

5.1 INTRODUCTION

Recent years have seen an increase in the incidence of *C. difficile* infection (CDI), as well as in its severity and recurrence (Rupnik et al. 2009). It has been previously reported that the strain of *C. difficile* causing the infection can be an important determinate in predicting severity (Goorhuis et al. 2008; Ghose 2013; Marsh et al. 2012). Infection with certain strains may result in asymptomatic conditions, while others cause more severe conditions, such as colitis, sepsis, and even death. The recent emergence of 'hypervirulent' strains, such as ribotypes 027 and 078 (Goorhuis et al. 2008; Marsh et al. 2012), has renewed interest in CDI, with a number of these strains exhibiting increased toxin production (Warny et al. 2005). RT 027 in particular has been associated with more severe and recurrent disease. There has been debate as to whether ribotypes can be indicative of disease severity. Recently a study (Walk et al. 2012) has argued that ribotype does not predict severity of infection when results were adjusted for covariates. This idea has been challenged, and resulted in much discussion in the literature (A S Walker, Eyre, Crook, et al. 2013a; A Sarah Walker, Eyre, Crook, et al. 2013; Walk et al. 2013a). Walkers group have conducted the largest population study of genotype and CDI, who show that RT 027 is consistently associated with more severe disease (A Sarah Walker, Eyre, Wyllie, et al. 2013). The exact molecular mechanisms of this increased virulence are not fully known.

Many studies have examined the importance of SLPs as virulence factors (Ausiello et al. 2006; Cerquetti et al. 2000; Drudy et al. 2004; Merrigan et al. 2013). SLPs can vary greatly in sequence between strains, and this sequence variation may play a role in disease severity (McCoubrey & Poxton 2001; Eidhin

et al. 2006). We have previously shown that SLPs from RT 001 activate an immune response, specifically through TLR4, inducing proinflammatory cytokine production (Ryan et al. 2011). We have also shown the ability of SLPs from RT 001 to induce clearance responses in macrophages *in vitro*, essential for resolution of infection (Collins et al. 2014). As SLP sequences vary between strains, this sequence variability may be playing a role in disease severity. We have previously shown the SLPs from *C. difficile* RT 001 to signal through the MyD88-depenent pathway alone, activating NF-κB (Ryan et al. 2011), but not IRF3. It is not yet known if the sequence differences that exist between the SLPs of *C. difficile* have the potential to induce variable signalling pathways downstream of TLR4.

The type of immune response that is mounted in the gut against invading pathogens can greatly influence disease outcome. For example, Th17 cells have been shown to be essential in the clearance of extracellular pathogens in the gut (Blaschitz & Raffatellu 2010). This T cell subset also plays a role in neutrophil recruitment, as Th17-deficiency results in impaired recruitment of the cells to the mucosa in *S. typhimurium* infection (Raffatellu et al. 2008). A neutrophil response is also important in *C. difficile* infection (Jarchum et al. 2012), further implying a role for Th17 cells in CDI. Cytokines IL-6 and IL-23 are important in driving Th17 adaptive response, and these cytokines have previously been observed to be upregulated in response to SLPs from RT 001 (Ryan et al. 2011; Ausiello et al. 2006). Inhibition of such a response may result in impaired clearance and exacerbated disease.

A variety of animal models exist for the study of CDI, utilising animals ranging from rodents such as mice and guinea pigs, to larger animals such as Rhesus monkeys and foals (Lawley & Young 2013; Best et al. 2011). Each model has unique advantages, and all have been used to greatly improve our understanding of the pathogen. As the host responses to *C. difficile* has been reported to vary between animals (Keel & Songer 2006), the choice of animal model is critical, especially in respect to comparing observations to human disease. One of the first animal models utilised hamsters, administering the cecal contents from antibiotic-treated, diseased animals to the cecum of healthy animals (Bartlett et al. 1977). This model proved very successful, linking the role of toxins to CDI (Bartlett et al. 1978; Libby et al. 1982), and demonstrating disease transmissibility (Bartlett et al. 1977). There are drawbacks for this model however, infection occurs in the cecum of hamsters, but in the colon of humans (Price et al. 1979). In addition, relative to mice, there is a lack of immunological reagents available to study the host response to infection.

Mouse models for CDI have been continuously refined. The location of infection in the colon, similar to humans, and the wide availability of mouse-specific antibodies, make it a very suitable animal in which to study CDI. Like humans, healthy mice are resistant to *C. difficile* due to the protective nature of the microbiota (Lawley et al. 2009), therefore susceptibility needs to be induced by the administration of antibiotics. The antibiotic cocktail mouse model of CDI (Chen et al. 2008) mirrors the majority of disease factors observed in humans by disrupting the commensal bacteria of the gut, and for this reason we chose this model for the study.

The aim of this study was to confirm variable immune responses to different *C. difficile* ribotypes *in vivo* and to examine whether these responses induced different disease states. By using the mouse antibiotic model of *C. difficile* infection and examining progress of infection, along with cytokine analysis of RNA from the colon, we hope to elucidate the mechanisms of disease severity in CDI.

5.2 Results

5.2.1 Mice infected with *C. difficile* exhibit weight loss not observed in control animals

The experimental groups for the *C. difficile* animal model are given in Table 5.1. All observed weight changes were plotted as percentage weight change. No weight loss was observed in control mice administered with antibiotics alone. Animals infected with ribotypes 001 and 027 exhibit similar weights as control animals one day after infection. At Day 2, a decrease in weight is observed in both groups indicating progression of infection ($P<0.001$). This can be seen in Figure 5.1. After Day 2, both groups begin to recover weight. RT 001-infected animals seem to recover at a more rapid rate, and by Day 6, their weight is indistinguishable from the Control group. RT 027-infected animals, however, appear to recover more slowly ($p<0.05$ at Day 5), implying a more persistent infection, not being cleared by the immune system.

Table 5.1

Group	Day 3	Day 7	
Control	4	4	
001	6	6	
027	6	6	
Total	**16**	**16**	**32**

Table 5.1. The table above indicates the animal numbers per group in the experiment. Four animals were present in each control group, and 6 for each infection group. A total of 32 mice were used in the study.

5.2.2 Ribotype 027-infected mice possess a higher bacterial load than RT 001-infected mice.

The contents of the cecum were collected from mice in each group at Days 3 and 7 and used to determine the number of colony forming units (CFU), a measure of CFUs indicative of bacterial load in the gut. At Day 3, RT 001-infected animals had an average of $1x10^6$ CFU/g, while 027-infected animals had 5 times this amount, averaging at $5x10^6$ CFU/g ($p<0.001$). As both infected groups were initially treated with 10^3 *C. difficile* spores, this suggests increased virulence in RT 027, with a higher bacterial load three days after infection. At Day 7, numbers were lowered for both groups, but 027-infected animals still contained a large amount of bacteria in the gut (over $3x10^6$ CFU/g, $p<0.01$). This is a further indication of RT 027s increased virulence, and suggests an inability in the host to effectively clear the pathogen. The CFU counts can be seen in Figure 5.1.

Figure 5.1

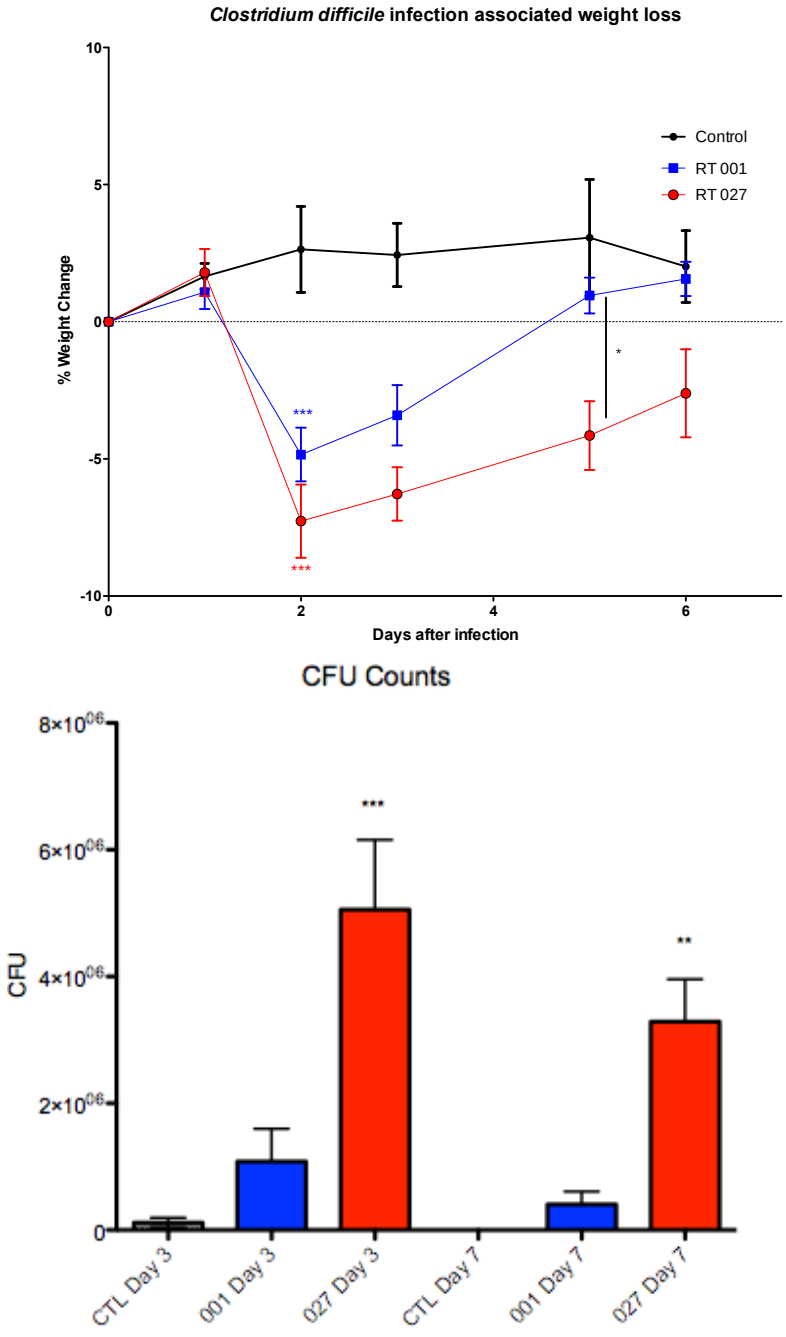


Figure 5.1 *Clostridium difficile* infection appears more severe in 027-infected mice relative to RT 001 infection. Weight change and CFU counts of mice infected with 10^3 *C. difficile* spores, showing the progression of disease. The control group is labelled black, RT 001-infected group blue and RT 027-infected red. Weight-loss is observed in both 001- and 027-infected animals, with greater recovery in RT 001. RT 027-infected animals also exhibit higher bacterial load at both time points. Statistical significance between the groups was determined by one-way ANOVA * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

5.2.3 *Clostridium difficile* ribotype 027 causes more severe tissue damage than 001, with less recovery by Day 7.

H&E staining of colonic tissue from RT 001- and 027-infected animals revealed the extent of tissue damage caused by *C. difficile* infection. Tissue from control mice clearly showed the structure of the intestinal wall. Increased oedema and disruption of villi structure is seen in mice infected with RT 001 (Figure 5.2). This damage is largely repaired at Day 7. These data suggest that the animal can effectively mount a response against *C. difficile* RT 001, and reverse tissue damage within seven days. Tissue damage was also observed in RT 027-infected animals. The extent of this damage appeared similar to that of RT 001 at day 3, but clearance had not occurred by day 7. More severe damage to the mucosa was observed at this time point, indicating an inability of the animals to effectively clear the pathogen, indicating that *C. difficile* RT 027 causes more severe disease than RT 001. Histology images can be seen in Figure 5.2.

Figure 5.2

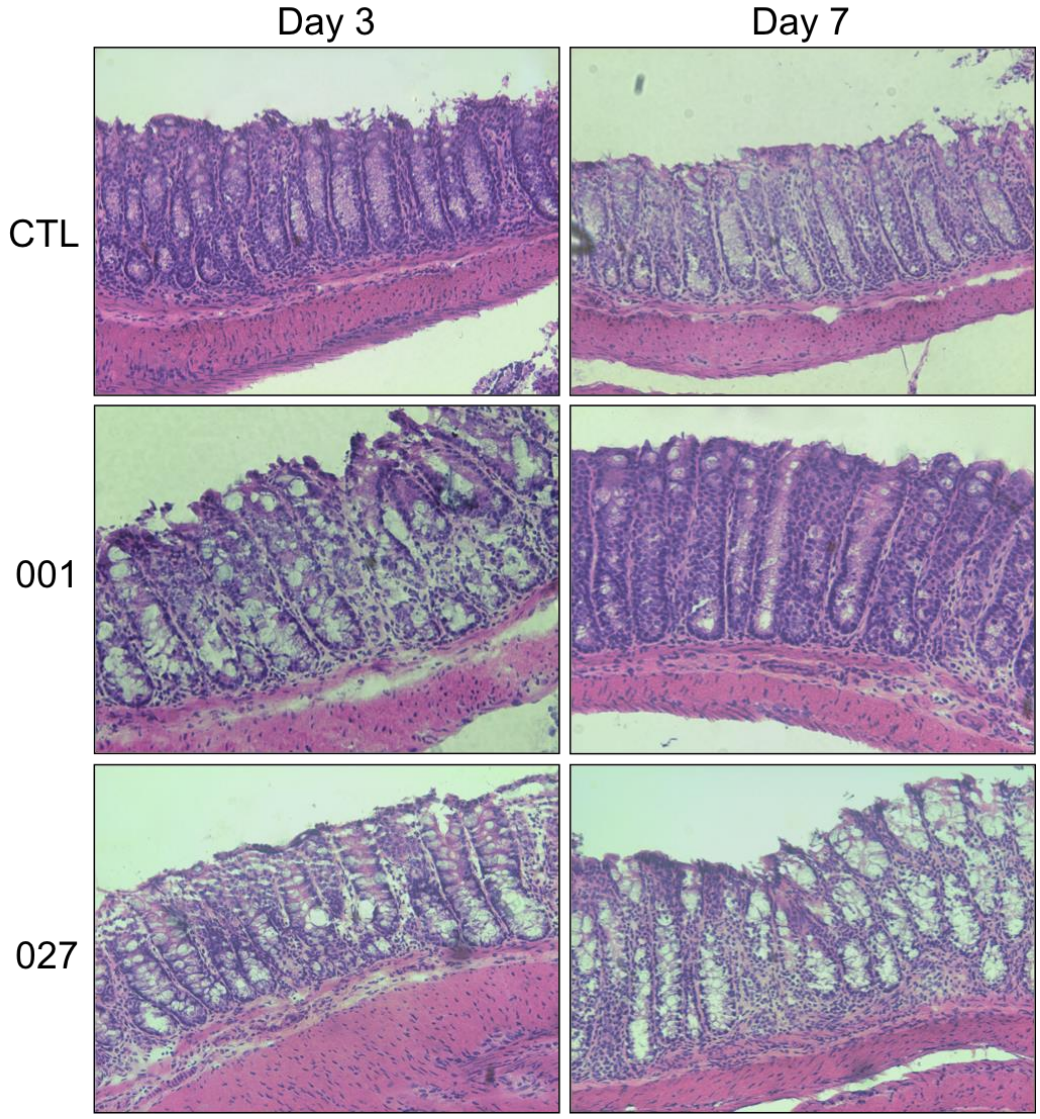


Figure 5.2. Histology reveals increased tissue damage in mice infected with ribotype 027.

Cross-sections of distal colon are shown for antibiotic-treated control mice, as well as RT 001- and 027-infected mice at day three and day seven post-infection. Nuclei are stained dark purple, while cytoplasm and muscle fibre are stained pink. Damage to intestinal structure can clearly be seen for both infected groups, with 001-infected animals recovering by day seven. As disease progresses in 027-infected animals, damage appears more severe, indicating a lack of clearance. Shown are representative images from four sampled mice per group.

5.2.4 *Clostridium difficile* ribotypes 001 and 027 induces recruitment of neutrophils *in vivo*, with neutrophils persisting in ribotype 027-induced infection.

Tissue sections of the distal colon were stained with a neutrophil-specific antibody and DAB staining kit, with the neutrophils being visible as brown spots. The brown colour of the DAB staining was converted to red using ImageJ software, for ease of visualisation. Surrounding tissue was counterstained with haematoxylin to also allow for visualisation of surrounding gut tissue. No positive staining was observed in the tissue from control animals, indicating no neutrophils present in a non-infected state (Figure 5.3). A high infiltration of neutrophils was observed in RT 001-infected animals at Day 3, with brown spots being observed in the lamina propria and near the lumen. Neutrophil infiltration has been previously reported in CDI, so this was not surprising. At Day 7, no recruited neutrophils were detected in the colon. For 027-infected animals, neutrophils were observed at both days 3 and 7 post-infection. The presence of neutrophils at the later time-point suggests persistent inflammation, which is in line with the observed tissue damage. Representative images of neutrophil infiltration can be seen in Figure 5.3.

Figure 5.3

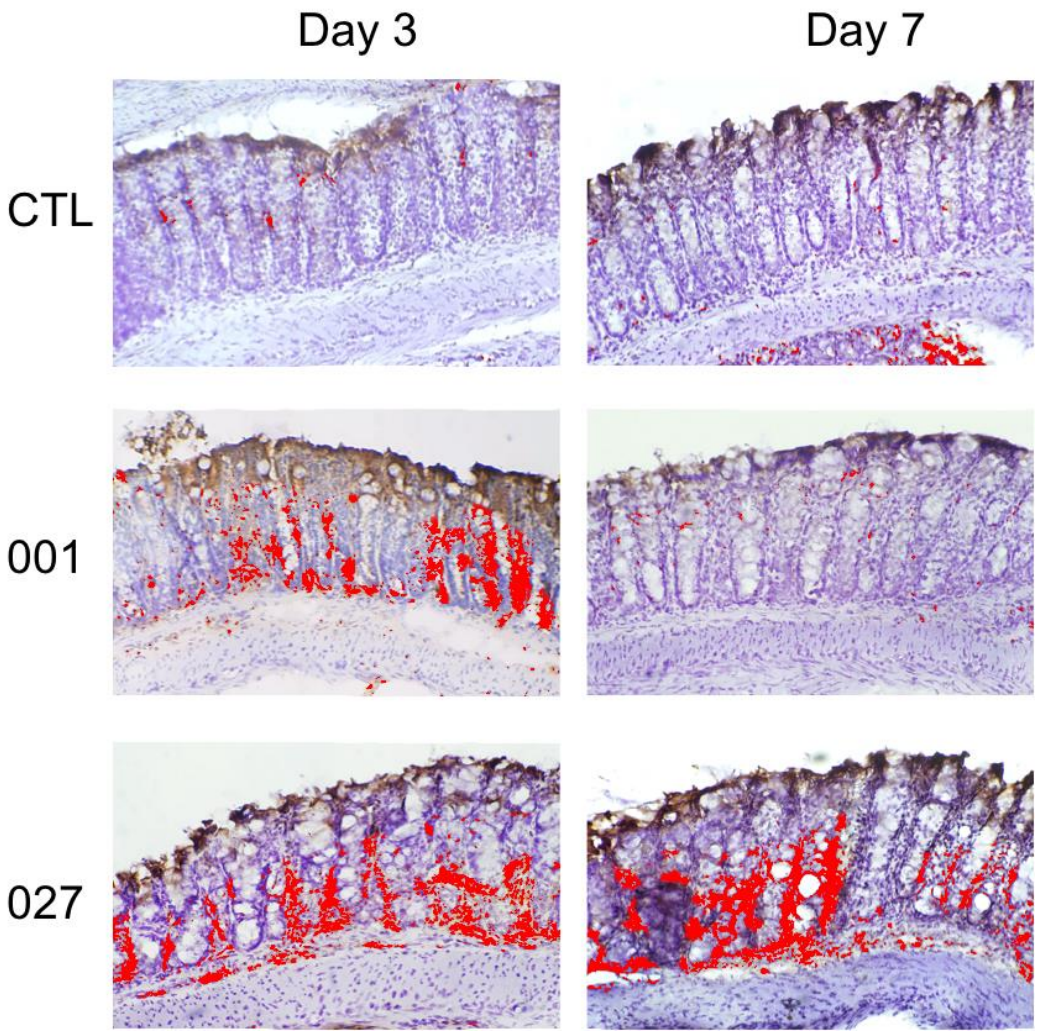


Figure 5.3 *C. difficile* infection induces neutrophil recruitment to the gut.

Neutrophils targeted with anti-NIMP-R14 antibody and visualised with DAB chromogen substrate were detected in the colon of mice infected with both ribotypes 001 and 027. For 001 infection, neutrophils are recruited and present at day three, but had cleared by day seven. This clearance was not observed in 027-infection, with neutrophils persisting at day seven, possibility contributing to continued inflammation. Shown are representative images from four sampled mice per group.

5.2.5 Macrophages were detected in colonic tissue of both 001- and 027-infected animals.

Anti-F4/80 antibody was used to stain for the presence of macrophages in the colonic tissue of infected mice. The brown colour of the DAB staining was converted to red using ImageJ software, for ease of visualisation. Small numbers of macrophages were observed in the lamina propria of non-infected control tissue (Figure 5.4). This was expected as intestinal macrophages are resident in the gut (Doe 1989). A much higher population of macrophages were observed in 001-infected mice at day 3, suggesting increased recruitment due to the inflammatory environment of the infected gut. Fewer numbers were present at Day 7, again suggesting that the host immune response is successfully controlling and resolving infection at this time point. Tissue damage was evident at both days 3 and 7 in 027-infected animals. Macrophage numbers appeared in lower numbers than 001-infected tissue, suggesting impaired clearance in response to this strain. At Day 7, macrophage numbers remained high, with no recovery in tissue structure. This is consistent with the observed late inflammation and more severe disease observed in RT 027-infection. Representative images of macrophage staining can be seen in Figure 5.4.

Figure 5.4

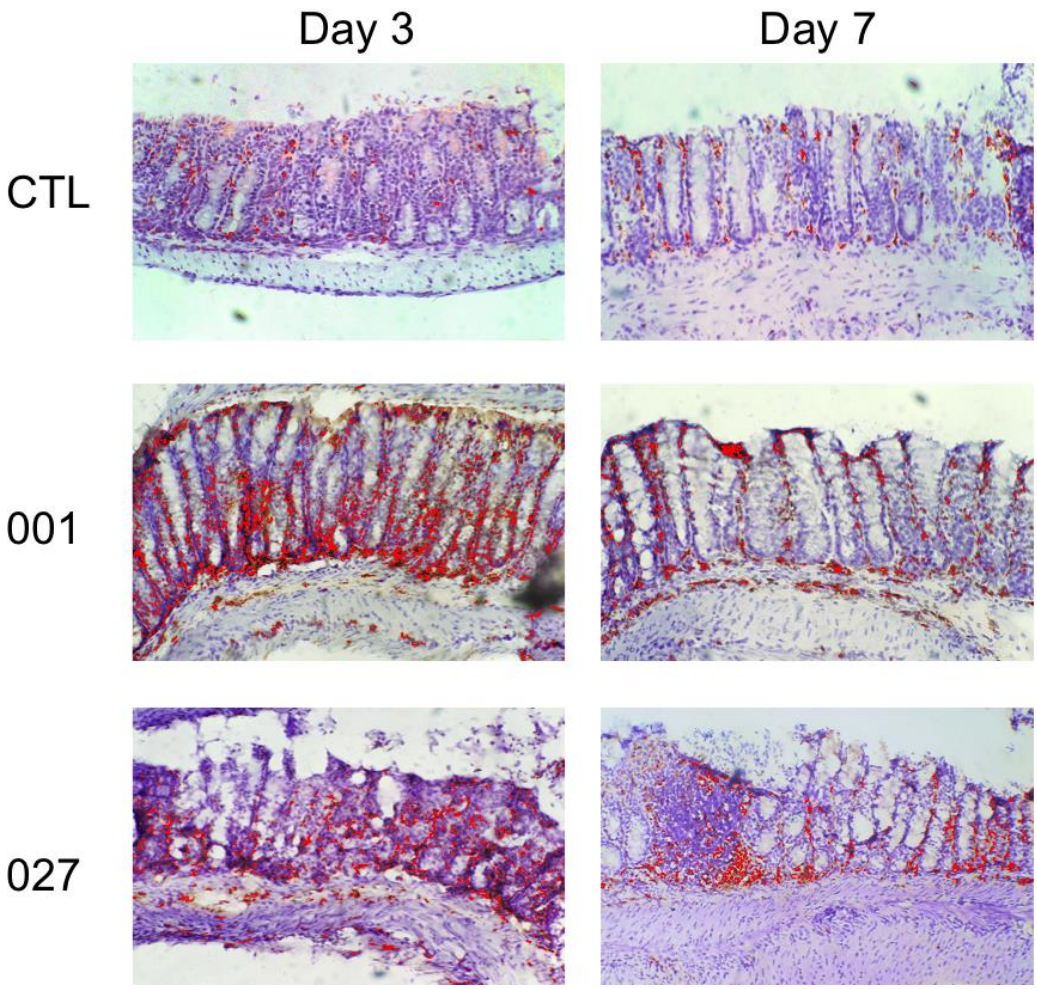


Figure 5.4 *C. difficile* infection induces macrophage recruitment to the gut.

Macrophages stained with anti-F4/80 antibody and visualised with DAB chromogen substrate were detected in the colon of control mice, as well as those infected with both ribotypes 001 and 027. Macrophages were detected in control tissue, due to the presence of resident intestinal macrophages. *C. difficile* RT 001 induced recruitment of a large amount of macrophages to the gut, can be seen by their higher numbers at day three. By day seven, this number had decreased. Macrophage numbers were also increased in RT 027-infection, with higher numbers at day seven relative to day three. Shown are representative images from four sampled mice per group.

5.2.6 *Clostridium difficile* ribotypes 001 and 027 induces recruitment of CD4+ T cells *in vivo.*

Anti-CD4 antibody was used to stain for the presence of CD4+ T cells in the colonic tissue of infected mice. The brown colour of the DAB staining was converted to red using ImageJ software, for ease of visualisation. No T cells are detected in control tissue at either Day 3 or Day 7 after infection, suggesting a lack of resident T helper cells in the gut (Figure 5.5). CD4+ cells were detected in the colonic tissue from 001-infected mice at Day 3, albeit in lower numbers than either neutrophils or macrophages. At Day 7, similar numbers of CD4+ cells were observed, as would be expected for a successful adaptive immune response. Fewer CD4+ cells were detected at Day 3 of 027-infected mice. As adaptive responses can take longer to be induced, this was not unexpected. By Day 7, numbers of CD4+ cells had dramatically increased, corresponding to an increased inflammatory state. Representative images of macrophage staining can be seen in Figure 5.5.

Figure 5.5

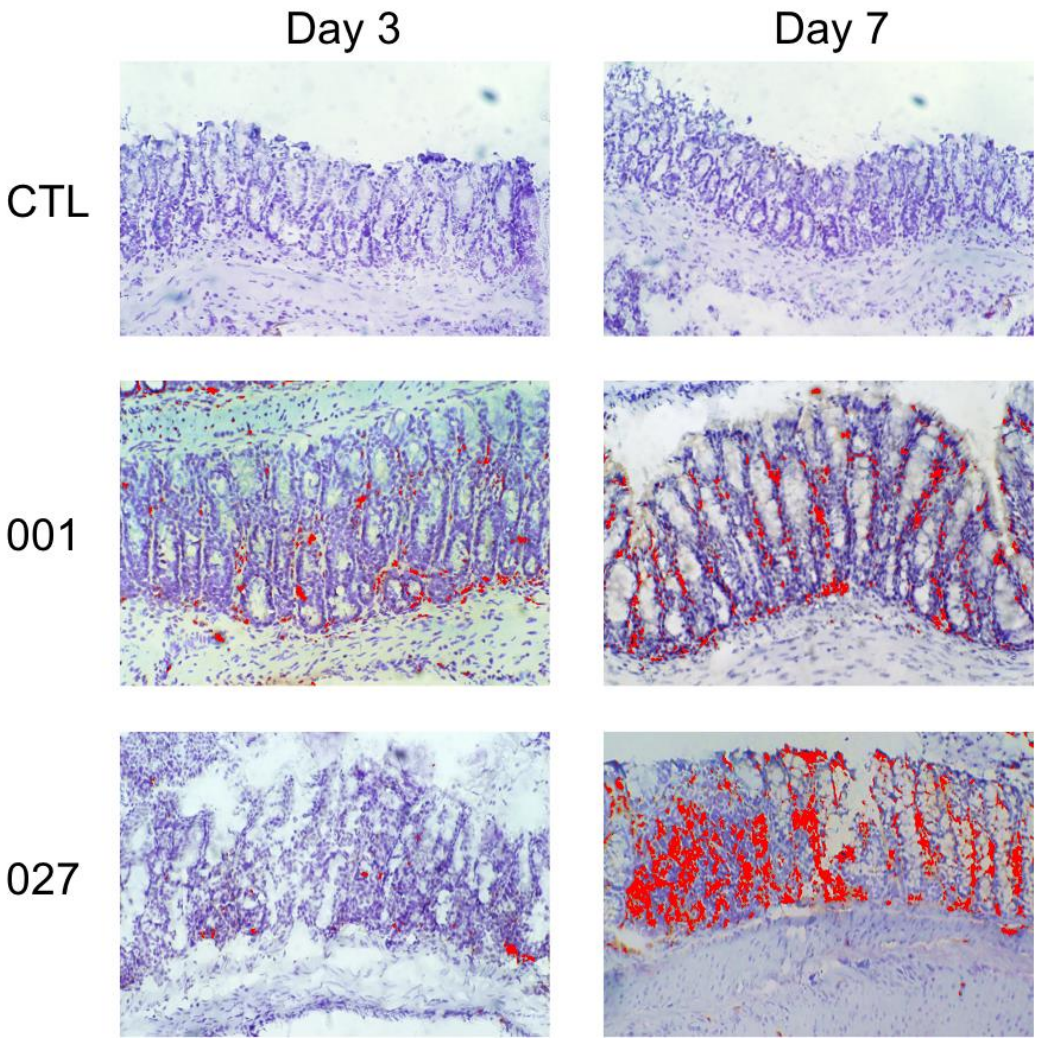


Figure 5.5 *C. difficile* infection induces CD4+ T helper cells recruitment to the gut.

T helper cells stained with anti-CD4 antibody and visualised with DAB chromogen substrate were detected in the colon of mice infected with both ribotypes 001 and 027. CD4+ cells were not detected in control tissue. *C. difficile* RT 001 induced recruitment of CD4+ cells to the gut, with positive staining observed at both days three and seven post infection. CD4+ cell levels were lower in 027-infected tissue at day three, relative to RT 001-infection. This number had increased by day seven. Shown are representative images from four sampled mice per group.

5.2.7 *Clostridium difficile* ribotype 001 induces early production of proinflammatory cytokines, including those necessary for a Th17 response in the colon.

qPCR analysis of homogenised colonic tissue revealed the cytokine environment of the gut during infection. At Day 3, IL-23, TGF-β and IL-6 were all strongly induced by RT 001, with a 10-fold increase in both IL-23 ($p<0.01$) and TGF-β ($p<0.001$), and an 8-fold increase in IL-6 ($p<0.05$). This can be seen in Figure 5.6. There was no statistically significant increase of IL-23 or TGF-β in response to RT 027. At day 7 the expression of these cytokines had dropped to control levels in RT 001-infected animals. An exception to this is was an increase in RORγ expression in RT 001-infected animals at Day 7 ($p<0.05$). IL-23, TGF-β and IL-6 are essential in the development of an adaptive Th17 response, which has been previously been shown to play an important role in the clearance of extracellular bacterial infections. RORγ is the master regulator of Th17 response, and its presence in RT 001-infected animals suggests an important role for Th17 responses in *C. difficile* infection. Additionally, a ~13-fold increase in IL-17 expression was detected at Day 3 for 001-infection ($p<0.001$), again indicative of a Th17 response. At Day 7, IL-17 expression remained high in 001-infected animals, producing 5 times more than control mice ($p<0.05$). At this time point, 027-infected animals produced higher levels of IL-17 (8-fold higher relative to control, $p<0.01$), but no statistically significant difference was found between RT 001 and 027. RT 027 did not induce increased expression of IL-23 or TGF-β, yet a 4-fold increase in IL-6 was observed in response to this ribotype at Day 3 ($p<0.01$). The results can be seen in Figure 5.6.

Figure 5.6

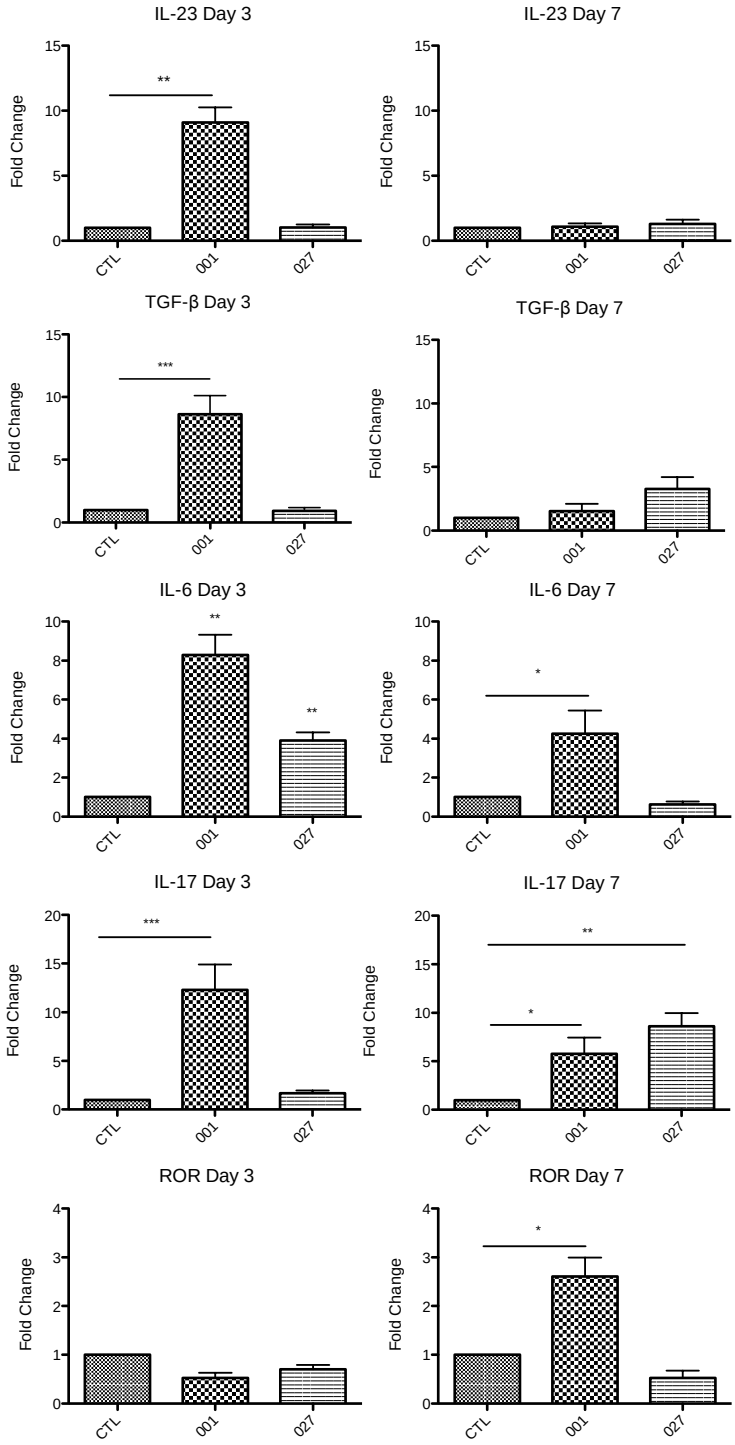


Figure 5.6. *Clostridium difficile* ribotype 001 induces early production of proinflammatory cytokines, including those necessary for a Th17 response. qPCR analysis of colonic tissue isolated from all animals in control, RT 001 and RT 027 groups at Day 3 and Day 7 post-infection. Expression of IL-6, IL-23, TGFβ, IL-17 and RORγ is upregulated by RT 001 at day three. This is not seen for RT 027. Cytokine production given in fold change relative to control. Student t tests were carried out on ΔCq values, * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

5.2.8 *Clostridium difficile* ribotype 027 exhibits late induction of proinflammatory cytokines, with high levels of IL-10 in the colon.

Three days after infection, we observed no significant increase in expression of pro-inflammatory cytokines IL-12p40, TNFα or IFNγ in either 001- or 027 infected mice relative to control animals. Observed tissue damage is likely due to production of toxins that act on the cytoskeleton of host cells. Anti-inflammatory IL-10 expression was significantly upregulated by RT 027 at both Day 3 and Day 7 post-infection; a ~2.5-fold increase in IL-10 levels was seen in 027-infected animals at Day 3 ($p<0.01$), and a ~1.5-fold increase was seen at Day 7 ($p<0.05$). IL-12p40 was upregulated by both RT 001 and 027 at Day 7, with a 2-fold increase in response to RT 001 ($p<0.05$) and 3-fold increase in response to 027 ($p<0.001$). No significant up regulation of TNFα was detected in response to RT 001, but RT 027 induced ~4-fold increase in the levels of this cytokine at Day 7 ($p<0.05$). IFNγ remained consistent between the three experimental groups at Day 3, with a non-significant decrease caused by RT 027. At Day 7, levels of IFNγ were comparable between control and 001-infected animals, but a decrease was seen in response to RT 027 ($p<0.05$). The results can be seen in Figure 5.7

Figure 5.7

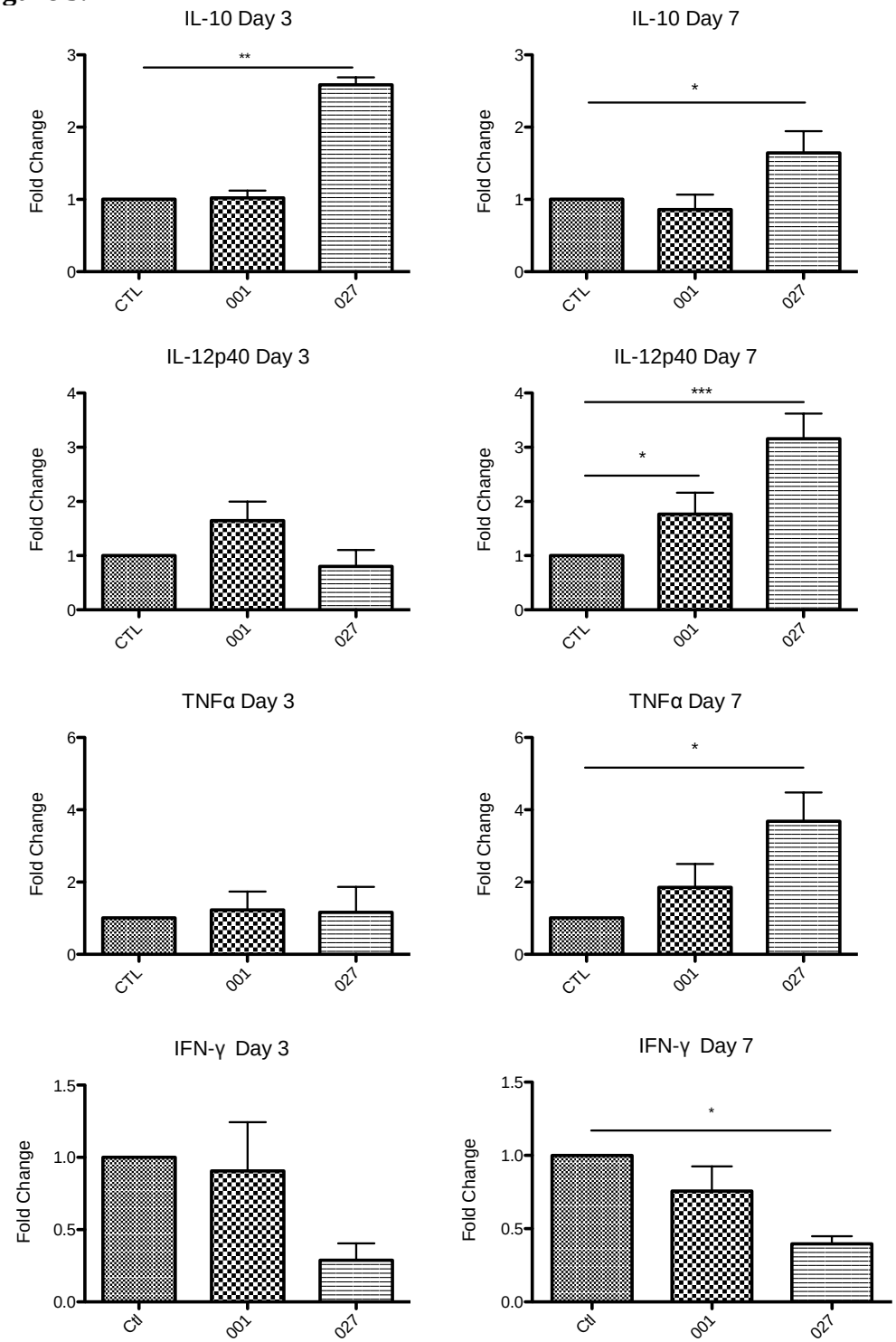


Figure 5.7. *Clostridium difficile* ribotype 027 exhibits late induction of proinflammatory cytokines, with high levels of IL-10. qPCR analysis of colonic tissue isolated from all animals in control, RT 001 and RT 027 groups at Day 3 and Day 7 post-infection. Expression of IL-10, IL-12p40, TNFα, and IFNγ is modulated by RT 027. Cytokine production is given in fold change relative to control animals. Student t tests were carried out on ΔCq values, * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

5.2.9 Chemokine levels are similarly up regulated in both 001- and 027-infected mice, with TLR4 being downregulated by 027.

TLR4 expression is responsible for recognising the SLPs of *C. difficile* (Ryan et al. 2011), so increased expression of this receptor may playing a role in initiating an immune response and clearing the pathogen. High levels of TLR4 (2-fold increase compared to control, $p<0.01$) were detected in 001-infected mice at Day 3. This increase was not observed in 027-infected mice; there was a small, but statistically significant, decrease relative to RT 001 ($p<0.05$), indicating there may be an altered response to this ribotype. Expression of chemokines MIP-1α and MIP-1β and MCP were all upregulated in response to *C. difficile*. There was an almost 3-fold increase in MCP-1 expression in response to RT 001 at Day 3 relative to control mice ($p<0.01$); no statistically significant increase was seen in 027-infected animals at either time point. MIP-1α expression was not upregulated by either RT 001 or 027 at Day 3. Increased levels 1α were seen, however, at Day 7 in response to both RT 001 and 027, with 2.5-fold ($p<0.001$) and 3-fold increases ($p<0.001$) being observed respectively. Both RT 001 and 027 induced a ~3-fold upregulation of MIP-1β at Day 3 relative to control animals ($p<0.01$), and at Day 7 these levels were statistically similar to control levels. The results can be seen in Figure 5.8.

Figure 5.8

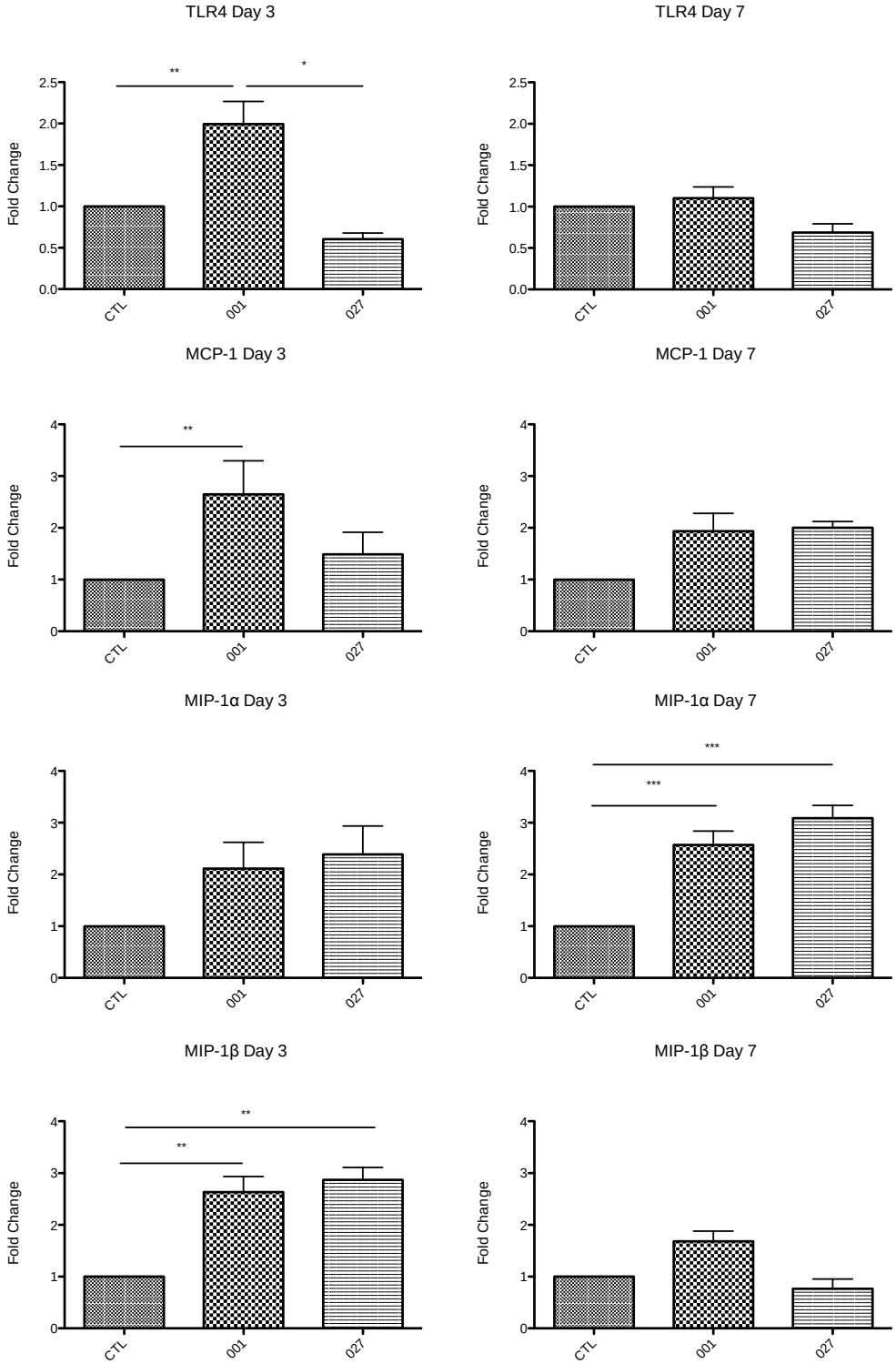


Figure 5.8. *Clostridium difficile* induces chemokine production and modulates TLR4 expression. qPCR analysis of colonic tissue isolated from all animals in control, RT 001 and RT 027 groups at Day 3 and Day 7 post-infection. Expression of TLR4, MCP-1, MIP-1α and MIP-1β is shown. Cytokine production is given in fold change relative to control animals. Student t tests were carried out on ΔCq values, * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

5.2.10 *Clostridium difficile* ribotype 001 induces early production of proinflammatory cytokines, including those necessary for a Th17 response in the cecum.

cDNA was also synthesised from RNA isolated from the cecum of control and infected animals. A similar trend of cytokine expression was observed in the cecum relative to the colon. Expression of IL-6, IL-23 and TGF-β was all increased in response to RT 001, but not RT 027, at Day 3. A two-fold increase in both IL-23 ($p<0.05$) and TGF-β ($p<0.01$) expression was detected in 001-infected animals relative to control. There was also a three-fold increase in IL-6 ($p<0.05$). No significant increase for these cytokines was detected in response to RT 027 at either time point. In addition, IL-17 expression was not increased by either ribotype at Day 3 in the cecum. An increase in IL-17 was seen at Day 7 in 001-infected mice, but this increase was found to be statistically insignificant. No RORγ was detected in the cecum. For each of these cytokines, expression was comparable to control animals at Day 7. Cytokine expression in the cecum can be seen in Figure 5.9.

Figure 5.9

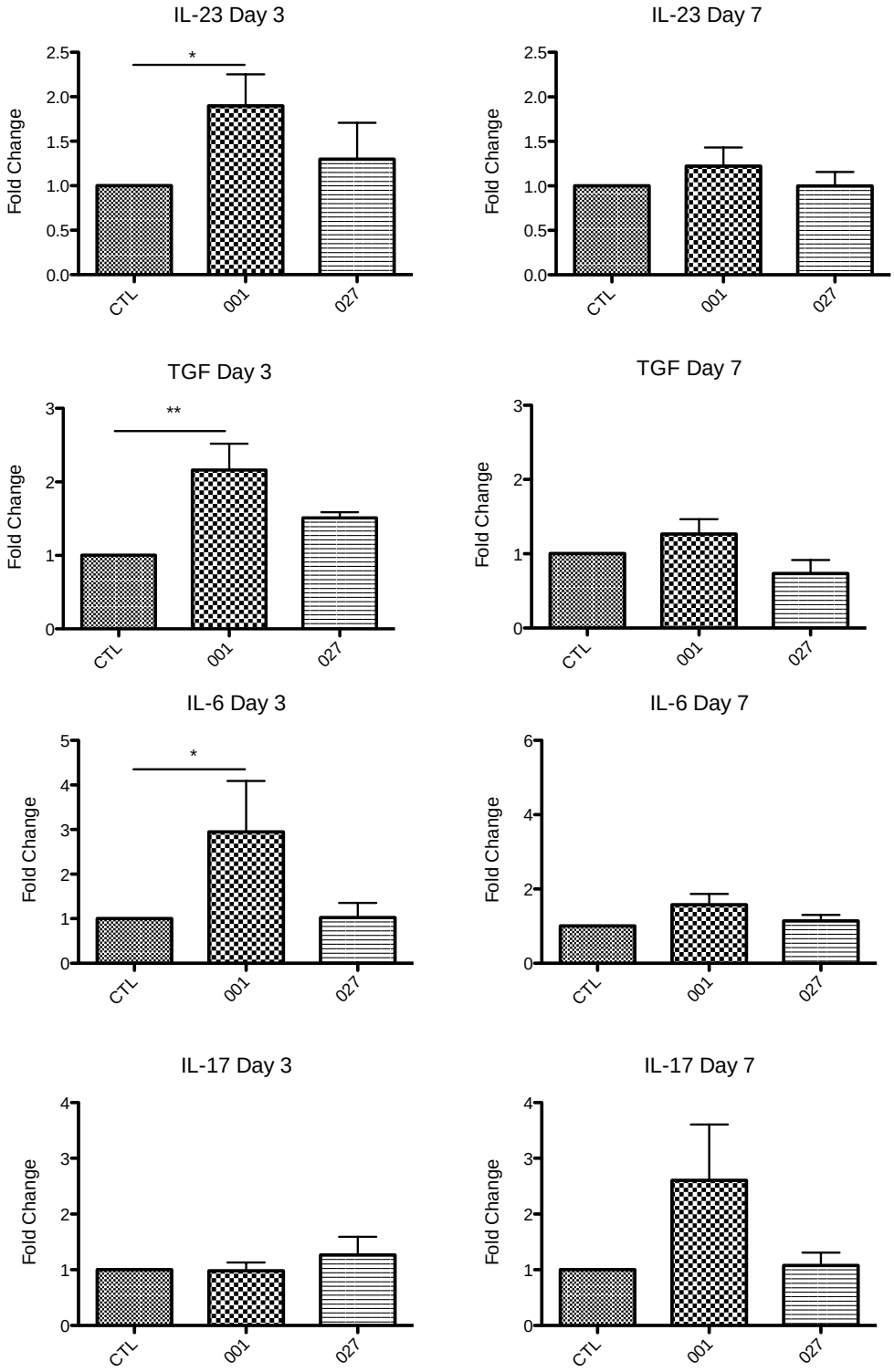


Figure 5.9. *Clostridium difficile* ribotype 001 modulates Th17-specific cytokines in the cecum. qPCR analysis of cecum tissue isolated from all animals in control, RT 001 and RT 027 groups at Day 3 and Day 7 post-infection.. Expression of IL-23, TGFβ, IL-6 and IL-17 is upregulated by RT 001. Cytokine production is given in fold change relative to control animals. Student t tests were carried out on ΔCq values, * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

5.2.11 *Clostridium difficile* ribotype 027 exhibits late induction of proinflammatory cytokines, with increased levels of IL-10 in the cecum.

cDNA was synthesised from RNA isolated from the cecum of control and infected animals. At Day 3, no significant increase in IL-10 was detected in response to either RT 001 or 027. RT 027 induced a 1.5-fold increase in IL-10 production at Day 7 ($p<0.001$). There was also no statistically significant increase in levels of IL-12p40 for both infection groups at either time points. TNFα was not upregulated by either ribotype in the cecum at Day 3. At Day 7, however, RT 027 induced a ~4-fold increase in TNFα levels relative to control ($p<0.01$). The results can be seen in Figure 5.10.

Figure 5.10

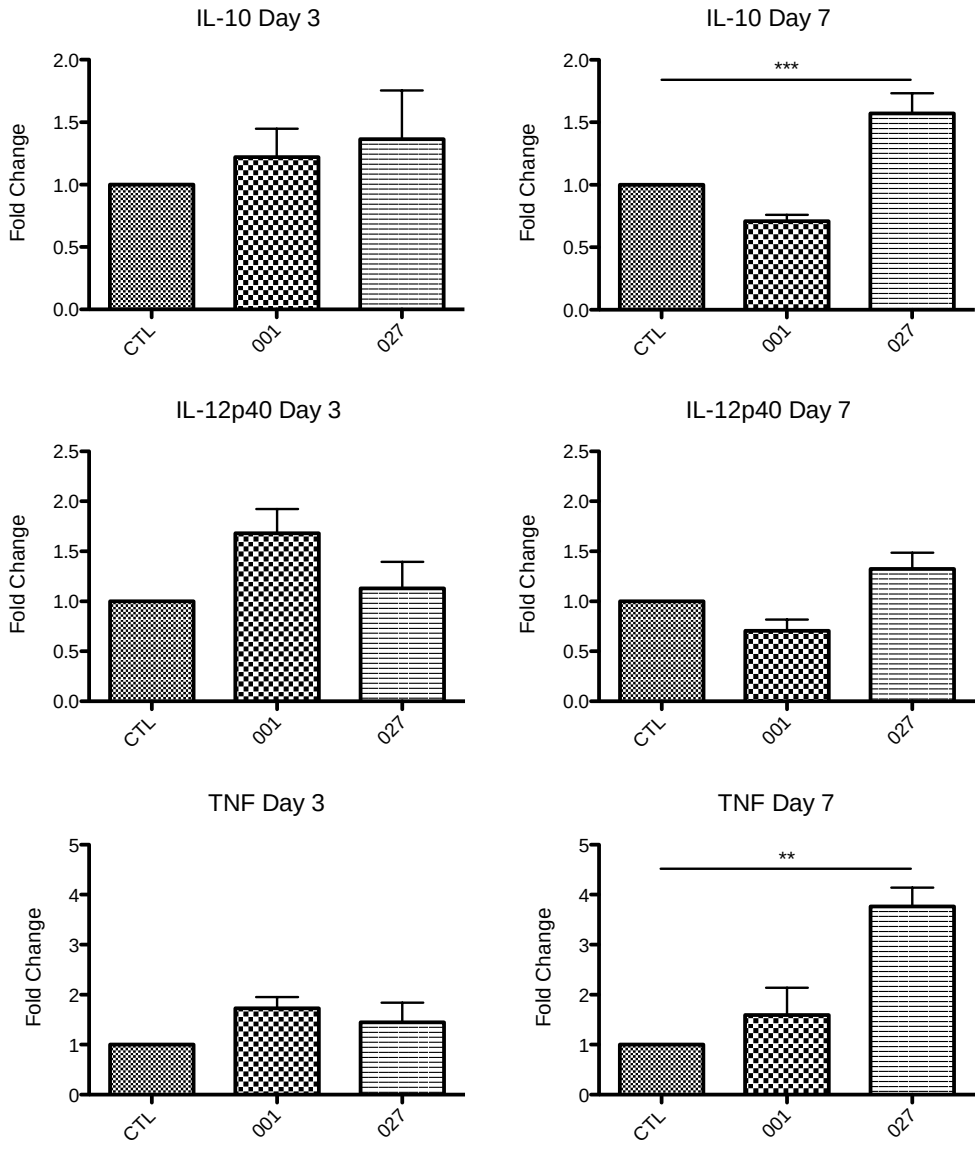


Figure 5.10. *Clostridium difficile* ribotype 027 exhibits late induction of proinflammatory cytokine TNFα, with increased levels of IL-10. qPCR analysis of cecum tissue isolated from all animals in control, RT 001 and RT 027 groups at Day 3 and Day 7 post-infection.. Expression of IL-10, IL-12p40 and TNFα is shown. Cytokine production is given in fold change relative to control animals. Student t tests were carried out on ΔCq values, * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

5.2.12 SLPs isolated from *Clostridium difficile* ribotypes 001 and 027 differentially drive high levels of IL-17 and IL-10, respectively, *in vitro*.

IL-10 production is upregulated by SLPs from RT 027, while IL-17 is upregulated by RT 001 SLPs, *in vitro*. Our group has recently shown the importance in surface layer proteins (SLPs) in the immune response to *C. difficile*, activating immune cells through TLR4. As we have detected high levels of IL-10 in 027-infected mice, correlating to impaired bacterial clearance, determining if SLPs play a role in this response is crucial. In order to do this, we employed a DC/T cell co-culture model. BMDCs activated with SLPs of RT 001 and 027, in the presence of OVA peptide, were grown in culture with CD4+ T cells isolated from OVA-transgenic mice. Dendritic cells were irradiated before co-culture, ensuring all cytokine production originated from the BMDC-activated T cells. Cytokine analysis revealed high levels of IL-17 in response to BMDCs stimulated with SLPs from RT 001 ($p<0.001$). This correlates with observations from the animal model showing Th17-linked cytokine production in response to this ribotype. Also correlating with the infection study was high levels of IL-10 from T cells in response to 027 SLP-stimulated BMDCs. $CD4^+$ T cells showed a two-fold increase in IL-10 production in response to 027 SLP-activated DCs relative to RT 001 ($p<0.001$). This high IL-10 production in response to 027 SLPs is also observed in dendritic cells isolated from BALB/c mice, as well as in the J774 macrophage cell line.

Figure 5.10

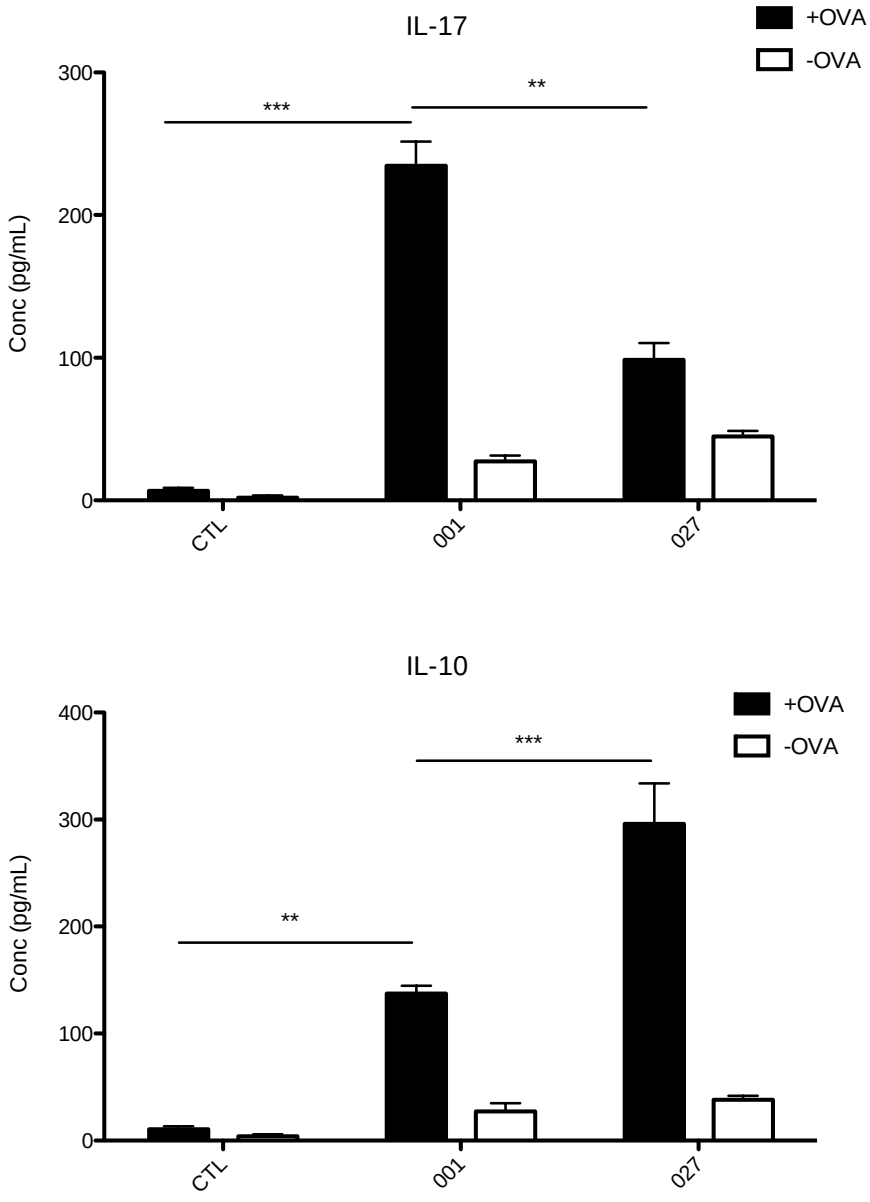


Figure 5.10 IL-10 production is upregulated by SLPs from ribotype 027, while IL-17 is upregulated by RT 001 SLPs *in vitro*.

SLPs isolated from *C. difficile* can specifically induce cytokine profiles similar to those observed *in vivo*. Specifically, SLPs from RT 027 induced high levels of IL-10, while SLPs from 001 induced high levels of IL-17. Statistical significance between the groups was determined by one-way ANOVA * $p<0.05$; ** $p<0.01$, *** $p<0.001$. Results are representative of three independent experiments.

5.2.13 SLPs from ribotype 027, but not 001, initiate both MyD88-dependent and -independent pathways downstream of TLR4.

In order to address how the different immune responses to RT 001 and 027 SLPs were being mediated, we examined whether they both activated similar signalling pathways downstream of TLR4. SLPs from RT 001 and 027 both activated NF-κB downstream of TLR4 (Figure 5.11), and subsequently, NF-κB-dependent IL-8 production. IRF3 is a transcription factor associated with the MyD88-independent pathway, and simultaneous activation of both these pathways has been associated with high IL-10 production. As previously observed, we did not see IRF3 activation with SLPs from RT 001, but RT 027 SLPs induce high levels of IRF3, and subsequently RANTES. A five-fold increase in IRF3 expression was detected relative to control cells ($p<0.001$), with RANTES also significantly increased by SLPs from 027 ($p<0.001$). The activation of both NF-κB and IRF3 by SLPs from RT 027 may explain the higher levels of IL-10 induced by RT 027.

Figure 5.11

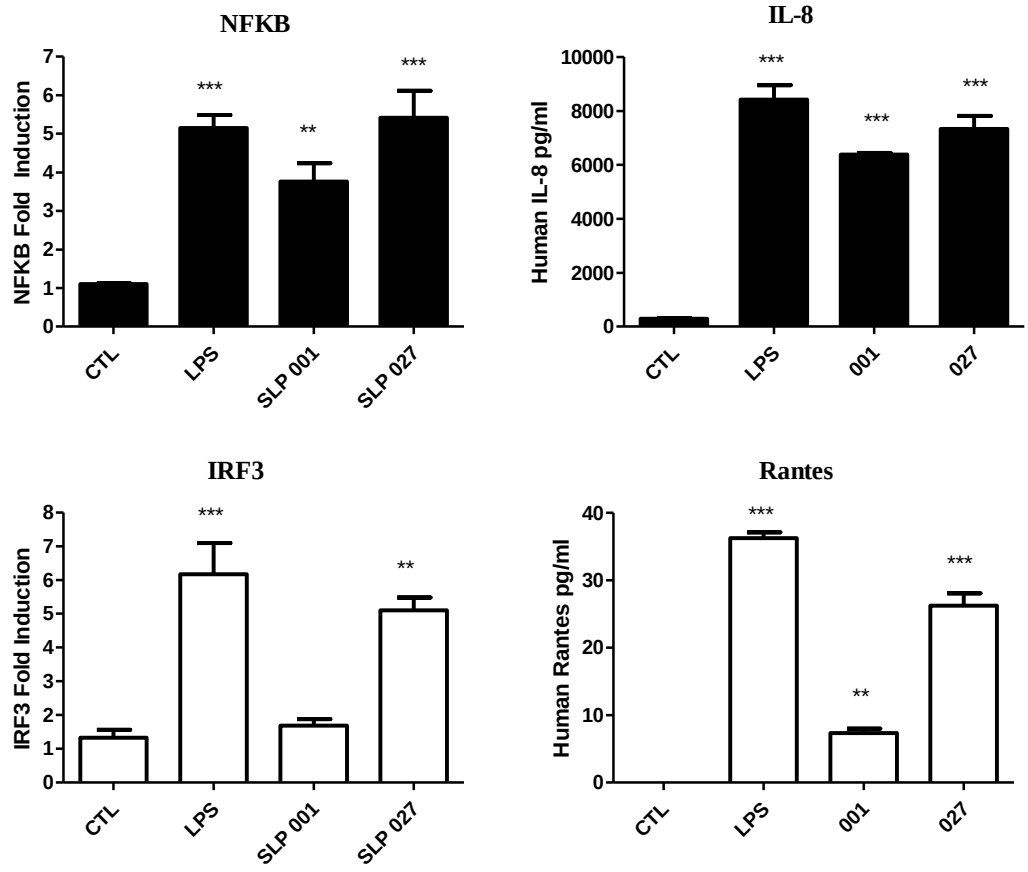


Figure 5.11 SLPs from ribotype 027 induce high levels of IRF3, indicating activation of the MyD88-independent pathway.

HEK-293 cells transfected with TLR4 were stimulated with SLPs from ribotypes 001 and 027. Levels of NF-kB and IRF3 were detected by luciferase assay, while IL-8 and RANTES levels were detected by ELISA. SLPs from both ribotypes stimulated production of NF-kB and IL-8. High levels of IRF3 and RANTES were observed in response to SLPs from 027, not 001. Statistical significance between the groups was determined by one-way ANOVA * $p<0.05$; ** $p<0.01$, *** $p<0.001$. Results are representative of three independent experiments.

5.3 DISCUSSION

In this chapter we set out to confirm our hypothesis that *C. difficile* RT 027 causes more severe disease than RT 001, by inducing higher levels of inflammation, and blocking the necessary clearance mechanisms required for resolving infection. To do this we used the antibiotic cocktail mouse model of CDI, which successfully mimics many of the symptoms associated with the disease (Chen et al. 2008).

Despite some controversy in the literature regarding ribotype and prediction of disease severity (Walk et al. 2012; A Sarah Walker, Eyre, Wyllie, et al. 2013; A S Walker, Eyre, Crook, et al. 2013b; A Sarah Walker, Eyre, Crook, et al. 2013), this study shows a strong link between *C. difficile* RT 027 and severe disease. Animals infected with this ribotype showed greater weight loss, more tissue damage, and higher bacterial load relative to control and 001-infected mice. These observations suggest that *C. difficile* RT 027 cannot be effectively cleared by the host immune system, and as a result, causes persistent disease. There is evidence in the literature that RT 001 results in less severe infection than 027 (Saxton et al. 2009), and mice infected with this ribotype were able to effectively recover after seven days, with lowered CFU counts and weights comparable to control mice. Histology of colonic tissue supported this analysis, revealing 001-infected animals with healthy gut morphology seven days after infection, while RT 027 induced more severe, persistent damage.

Induction of proinflammatory cytokines is likely a critical determinate in effectively clearing *C. difficile*. Upregulation of cytokines necessary for a Th17

response, specifically IL-6, IL-23 and TGF-β, was observed in RT 001-infected animals, but not in 027-infection. Increased levels of IL-17 were also detected at both Days 3 and 7 in 001-infected animals. Later induction of RORγ in this group is indicative of a Th17-driven adaptive immune response necessary for clearance. It has previously been established that Th17 responses play a crucial role in the clearance of extracellular pathogens (Khader et al. 2009; Sellge et al. 2010; Zhang et al. 2009a). Th17 cells have been shown to play a role in tight junction formation, so their accumulation in the gut may aid in reversing tissue damage (Kinugasa et al. 2000). Reversal of tissue damage, lower CFU counts, and lower proinflammatory cytokine levels at Day 7 all strongly suggest that RT 001 induces a Th17 response, which efficiently clears the pathogen from the gut. The lack of Th17-specific cytokines in 027-infected mice may contribute to persistent infection and lack of clearance. Indeed, impaired Th17 responses have been associated with disease in both mice and humans. IL-17A is associated with increased neutrophil recruitment, which is necessary for sufficient clearance of *C. difficile* (Roussel et al. 2010). Previous studies in a *Cytrobacter rodentium* model of intestinal infection showed IL-17 exhibit protective effects (Ishigame et al. 2009). However, Th17 cells have also been associated with pathology of inflammatory bowel disease, particularly seen to have detrimental effects for DSS colitis models (Feng et al. 2011). From our observations, it seems likely that the lack of a Th17 response in RT 027 infection is a contributing factor in its severe infection.

RT 027 induces few proinflammatory cytokines three days after infection. At this time point, bacterial load (CFU/g) in the gut is at its highest and histological

analysis reveals tissue damage induced by the infection. This observed tissue damage is likely due to toxin production by the bacteria, acting on the actin cytoskeleton of intestinal cells and destroying their structure. High levels of TNFα levels are associated with increased colitis in UC patients (Murch et al. 1991), and blocking of TNFα has been shown to alleviate symptoms of UC (Lv et al. 2014). IL-12p40 is also associated with more severe UC (Kullberg et al. 1998). High levels of proinflammatory TNFα, IL-12p40, and IL-6, are all observed in 027-infected animals at day 7. Interestingly, high levels of anti-inflammatory IL-10 are also observed. These cytokines will induce more damage may play a role in persistent infection by increasing local inflammation, while IL-10 may supress surrounding immune cell activity, impeding clearance.

IL-10 is an anti-inflammatory cytokine produced by many cells of the immune system. It is known to induce many anti-inflammatory actions, such as blocking phagosome maturation and proinflammatory cytokine production, and blocking antigen presentation in innate cells. It has been associated with inhibition of Th1, Th2, and Th17 differentiation, as well as protection against UC (O'Garra & Vieira 2007; Maynard & Weaver 2008; Ranatunga et al. 2012). In contrast, many recent infection studies have shown that an absence of IL-10 can lead to increased bacterial clearance (Redford et al. 2011). This suggests that IL-10 production results in different outcomes from different infections. IL-10 is found to play a deleterious role in *M. tuberculosis* infection, specifically by blocking IFNγ-mediated phagosome maturation, allowing phagocytosed bacteria to survive (O'Leary et al. 2011). The intracellular bacteria *Brucella abortus* has also been shown to utilise IL-10 to evade an immune response. It was found that

B. abortus induced early IL-10 production by $CD25^{+}CD4^{+}$ T cells to inhibit macrophage function, causing a balance in pro-inflammatory and anti-inflammatory cytokines which benefited the bacteria, and caused persistent infection (Xavier et al. 2013). Parasites have also been seen to recruit IL-10, with *Leishmania major* also inducing early IL-10 production by T cells, making BALB/c mice more susceptible to infection (Schwarz et al. 2013). Blocking of IL-10 was also found to increase immunity to *Schistosoma mansoni*, again indicating this pathogen was using the anti-inflammatory effects of IL-10 to its advantage (Wilson et al. 2011). Taken together, there is a strong possibility that *C. difficile* RT 027 is modulate cells T cells to produce high levels of IL-10, suppressing clearance of the pathogen.

IFNγ production was downregulated by RT 027. High levels of IFNγ are associated with UC, with anti-IFNγ antibodies alleviating disease in scid mice (Powrie et al. 1994). The down-regulation of IFNγ by RT 027 was therefore not expected. However, as anti-IFNγ antibodies have been shown to be ineffective in conditions such as Crohn's disease, it is likely this cytokine in not required for colitis to occur. IFNγ levels have previously been associated with chemokine production and neutrophil recruitment (McLoughlin et al. 2008), and low levels which may contribute to decreased cell migration and lack of clearance. IFNγ can also inhibit IL-10 production (Hu et al. 2006), so low levels in RT 027-infection correlate with increased IL-10.

PCRs carried out on RNA isolated from the cecum showed a similar trend in cytokine production in response to the two ribotypes. The fold changes of each

of these cytokines were much lower than those observed in the colon. This may be due to the main infection occurring in the distal colon of the animals, while the cecum likely contains many ungerminated spores. There is however, statistically significant increases in IL-23, TGF-β and IL-6 in response to RT 001, along with increased IL-10 in response to 027, suggesting that any vegetative cells present in the cecum are inducing a similar immune response to that of the colon.

Cellular infiltration into the gut is indicative of an inflammatory, colitis-like state. We observed infiltration of neutrophils, macrophages and CD4+ T helper cells into the gut at different time points for both RT 001 and 027 infections. Neutrophil recruitment has previously been described as a necessary step in the clearance of *C. difficile*, specifically being mediated through MyD88 (Jarchum et al. 2012). No neutrophils were detected in the colon of control mice that were administered antibiotics. We detected neutrophil infiltration at Day 3 in both 001- and 027-infected animals, showing an immune response to the pathogen. By day 7, neutrophil numbers had lowered in RT 001-infected animals but not in 027-infection. It is possible that the persistent levels of neutrophils in the gut in response to RT 027 are contributing to an increased inflammatory state, and therefore playing a role disease severity. Levels of MCP were higher in 001-infected animals relative to control at Day 3. This cytokine is responsible for recruiting neutrophils, and its expression explains the observed recruitment. This upregulation was not observed in response to RT 027 however. IFNγ has previously been shown to be involved in neutrophil recruitment, and is downregulated in response to RT 027. These observations suggest that RT 027

should decrease neutrophil recruitment to the gut by decreasing IFNγ expression and not inducing MCP production. This was not observed however, indicating a more complex mechanism for neutrophil recruitment.

We have previously shown a role for macrophages *in vitro*, specifically responding to the SLPs of *C. difficile* RT 001 and inducing clearance mechanisms (Collins et al. 2014). In this study we detected macrophages in the colon of each of the animal groups. Their detection in control groups is likely due to the presence of resident intestinal macrophages. An increase in macrophage numbers in both infection groups is expected and likely contributing to the inflammatory environment. At Day 3, larger numbers are seen in mice infected with RT 001. This correlates with increased expression of chemokines MCP and MIP-1β in the colon at Day 3. Combined with the observed cytokine profile in response to RT 001, it is likely these macrophages are inducing an environment favouring a Th17 adaptive response, and as the SLPs activate clearance responses in macrophages, their early recruitment may be an explanation for the less severe disease associated with this ribotype. RT 027 induces later recruitment, with macrophage numbers much higher at Day 7. Levels of MIP-1α are increased at both time points, but there is no increased expression of MCP or MIP-1β, which may explain impaired/delayed recruitment. The lack of proinflammatory cytokine production at Day 3 may result in a beneficial environment for the pathogen, causing the cells to produce IL-10 and therefore suppressing any clearance responses, resulting in more severe inflammation and disease.

CD4+ T cells were also recruited to the site of infection. No positive staining was observed in the colons of control animals at either time point, indicating a lack of CD4+ cells in the non-infected/susceptible state of the gut. Infiltration of these cells was observed in response to RT 001 at Day 3 and Day 7. Given the cytokine environment of the colon at Day 3, as determined by PCR analysis, it is likely that these CD4+ T helper cells are being driven to a Th17 phenotype, a phenotype previously shown to be necessary in bacterial clearance. The presence of increased levels of RORγ, the master regulator of Th17 cells, further corroborates this. Lower numbers of these cells at Day 7 also indicate the infection is being resolved. CD4+ cell infiltration is also observed in animals infected with RT 027, but with very low numbers at Day 3. This had increased by Day 7. The cytokines present in the gut during 027-infection do not promote a Th17 response, and this may be an indication as to why RT 027 induces more severe disease. Further to this, high levels of IL-10 in the colon of 027-infected mice suggest T cells may be acquiring a Treg phenotype. While these cells can play important roles in dampening down an immune response during clearance, they may also be recruited by *C. difficile* to inadvertently supress the immune response to the pathogen and inhibit clearance. As clearance is inhibited, the high levels of TNFα and IL-12p40 may be contributing to the inflammation induced by toxin production. This will further break down the epithelial lining of the intestine, resulting in the bacteria gaining a greater foothold, invading deeper into the gut.

While the responses to *C. difficile in vivo* are intriguing, and give insight into the mechanisms by which RT 027 induces more severe and persistent disease, we

must be careful in correlating this data with human disease. The common ancestor of humans and mice existed between 65 and 75 million years ago, which means there is ~150 million years of independent evolution between the two organisms (Waterston et al. 2002; Madsen et al. 2001). Due to this, it is not surprising that many differences exist between the immune systems of mouse and human (Mestas & Hughes 2004). These differences may mean that the responses observed in mice to not hold true for human disease. However, as a model for CDI, all the observations in this chapter correlate to expected findings for CDI, and consistently show RT 027 inducing more severe disease. More severe disease is consistently observed for *C. difficile* RT 027, while RT 001 has been associated with more mild disease. The recent debates in the literature (A Sarah Walker, Eyre, Wyllie, et al. 2013; Walk et al. 2012; A S Walker, Eyre, Crook, et al. 2013a; Walk et al. 2013b) suggest that ribotype alone cannot predict disease severity when other factors are taken into account. In this study, the animal groups were treated with the same antibiotics before being administered 10^3 *C. difficile* spores. They were kept in the same conditions for seven days. We observed a clear difference in severity between mice infected with RT 001 and 027.

Experiments conducted on purified SLPs *in vitro* also corroborated with the results from the animal study. Specifically, dendritic cells displaying OVA peptide, and activated with SLPs from RT 001 and 027 were grown in co-culture with OVA-specific T cells. As the dendritic cells were irradiated before the addition of T cells, all cytokines produced were from the T cells alone. BMDCs exposed to SLPs from RT 001 induced production of IL-17 in the T cells. This is

in line with what we observed in the animal model with high levels of IL-17 in the colon of 001-infected mice. This suggests that the SLPs may be playing a direct role in inducing this response *in vivo*. This IL-17 production was not observed in response to SLPs from 027, with high levels of IL-10 being observed. Again, this correlates to the findings in the animal model study. It appears that dendritic cells activated with SLPs from RT 001 drive T cells to produce IL-17, while 027 SLP-activated dendritic cells induce production of IL-10 in T cells. These SLPs may therefore be playing an important role in disease severity.

The mechanism for the increased levels of IL-10 we have associated with RT 027 may lie in the downstream signalling of TLR4. As a common pattern recognition receptor, TLR4 signalling is well understood and results in downstream activation of the transcription factor NF-κB. We have previously shown the SLPs from RT 001 to activate the MyD88-dependant pathway downstream of TLR4, inducing NF-κB transcription. It has previously been shown that optimal IL-10 production is induced when both MyD88-dependent and –independent pathways are activated, and expression of IRF3 would be indicative of the MyD88-independent pathway being active (Boonstra et al. 2006; Chang et al. 2007). We observed high expression of IRF3 in response to SLPs from 027, but not RT 001. SLPs from this ribotype also induced high levels of RANTES, a chemokine downstream of this arm of the pathway. Therefore, it appears that the SLPs from 027 are directly inducing immune cells to produce higher levels of IL-10 through TLR4 activation of the MyD88-independent pathway, which is not seen with SLPs from RT 001.

In conclusion, we have seen that *C. difficile* RT 027 induces more severe infection than RT 001. This ribotype caused more severe tissue damage, increased cellular infiltration, late induction of proinflammatory cytokines, and high levels of immune-suppressing IL-10. These factors act together to inhibit an appropriate immune response and clearance of the pathogen, while causing severe inflammation and tissue damage. RT 001 also induced tissue damage during the early stages of infection, yet generated a cytokine environment in the colon beneficial for the derivation of Th17 cells, an important T cell subset for the clearing of intestinal pathogens.

We have also seen that the SLPs from *C. difficile* have the potential to directly induce these environments. Purified SLPs from RT 001 induce high levels of IL-23, IL-6 and IL-17 *in vitro* (see previous chapter), while SLPs from 027 induce high levels of IL-10, in addition to many proinflammatory cytokines. Dendritic cells activated with SLPs also induce T cells to produce IL-17 in response to RT 001, and IL-10 in response to 027. It also appears that the SLPs from 027 drive the MyD88-independent pathway downstream of TLR4, which is associated with high IL-10 production. Therefore, the SLPs from *C. difficile* may be directly modulating the immune response, in the case of RT 027, by inducing IL-10 production and suppressing clearance mechanisms, while toxin production causes further tissue damage and disease.

Chapter 6

General Discussion

6.1 GENERAL DISCUSSION

Many studies have reported that the symptoms and severity of CDI can vary greatly depending on the strain causing infection (Kato et al. 2009; Drudy, Harnedy, et al. 2007; Drudy, Fanning, et al. 2007; Arvand et al. 2009; Bauer et al. 2011; Dawson et al. 2009; Warny et al. 2005). Some strains may induce mild disease, others cases may be asymptomatic, while certain strains can induce severe colitis, leading to sepsis and even death. Various virulence factors have been studied over the years, with a particular interest in toxin production. Toxins produced by *C. difficile* interact and damage the actin cytoskeleton of the epithelial cells of the intestine, causing infiltration of inflammatory cells into the gut (Libby et al. 1982). Therefore the ability of a particular strain to produce higher levels of toxins may explain the vast spectrum of infections observed. While the role of toxins in CDI has been heavily studied (Young & Hanna 2014; Carter et al. 2011; Steele et al. 2012; Warny et al. 2005), less is known about the role of the SLPs coating the outer layer of the pathogen. SLPs, consisting of low molecular weight (LMW) and high molecular weight (HMW) subunits cover the vast majority of the outer S-Layer of *C. difficile*. Our hypothesis states that different ribotypes of *C. difficile* cause variable disease, and that the sequences of the SLPs are a crucial determinate in severity of infection. The sequence of these proteins will determine how well the host immune system recognises the pathogen, and the type of immune response initiated.

SLPs exhibit vast sequence variability between strains (Calabi et al. 2001; Eidhin et al. 2006), and have been shown to induce an inflammatory response (Ausiello et al. 2006). We have reported that the proteins signal specifically through TLR4

to induce inflammation (Ryan et al. 2011). This receptor is normally associated with recognising LPS on the surface of gram-negative bacteria, so the ability of TLR4 to also recognise proteins from a gram-positive pathogen shows the robustness of the innate immune response. Our group have shown that the SLPs from ribotype 001 activate immune cells *in vitro*, and induce clearance responses in macrophages (Collins et al. 2014). The sequence variability of the SLPs between strains, along with the importance of sequence in recognition by TLR4, suggests that proteins from different ribotypes may not activate immune cells in the same way, and large sequence differences may even result in a loss of recognition of the SLPs by the immune system.

To determine if a link exists between sequence variation of SLPs and disease, we performed bioinformatics analysis of the gene that codes for SLPs, *slpA*. S-Layers are present on the surface of many bacterial species, and sequence variation of surface proteins has been reported to result in immune evasion for many pathogens (Higgins & Carrington 2014; van der Woude & Bäumler 2004). The low sequence similarity between the SLPs of different ribotypes suggests *C. difficile* may have evolved a similar method for survival. Analysis of the *slpA* gene from the 16 distinct ribotypes revealed many amino acid residues undergoing positive selection, giving a fitness advantage to that particular ribotype.

We detected evidence of positive selection in both the LMW and HMW SLP subunits. The LMW SLP subunit has previously been described as the immunodominant antigen of the S-Layer (Fagan et al. 2009). While it was not

surprising to find evidence of positive selection here, due to high sequence variability, the majority of ribotypes exhibiting positive selection in this subunit have been previously reported as "hypervirulent", for example ribotypes 017, 027 and 078 (Arvand et al. 2009; Goorhuis et al. 2008; Marsh et al. 2012). Our *in vitro* studies determined that the SLPs from RT 027 and 078 (where selection on the LMW SLP was detected) induced high levels of proinflammatory cytokine production, while SLPs from RT 001 and 014 (where no positive selection was detected) produced much lower levels. It is therefore possible that amino acid altering mutations in the SLPs from these ribotypes are contributing to hypervirulence, specifically by interacting with TLR4 to modulate the host immune response to the advantage of the pathogen.

Of particular interest is the relationship between the SLPs of ribotypes 001 and 027. The *slpA* sequences from these ribotypes exhibit high sequence similarity, and are sister taxa on our phylogenetic tree. Therefore, a similar immune response to both SLPs might be expected. Positive selection on the branch leading to 027 suggests that the small sequence variations that do exist are important. RT 027 is described as hypervirulent in many studies, being associated with more severe colitis and recurrent disease (Marsh et al. 2012; Valiente et al. 2014; Dawson et al. 2011). It was initially hypothesised that the variations observed in SLP of RT 027 may shield the bacteria from the immune system. Sequence variation may result in a loss of recognition by the immune system and the SLPs will therefore act as a molecular "cloak". The observed increased inflammation may contribute to disease severity in addition to the increased toxin production preciously described in RT 027 (Carter et al. 2011).

Positive selection has previously been associated with functional shift, where positively selected mutations alter the structure and function of a protein. Loughran *et al* have shown protein functional shift by positive selection in human myeloperoxidase (MPO) (Loughran et al. 2012). Using site directed mutagenesis on the amino acid residues under positive selection, they were able to show a complete loss in the novel chlorination function of MPO. The methods used for the detection of positive selection in MPO are identical to those used in this thesis (Loughran et al. 2012). This link between positive selection and protein functional shift has also been shown in other studies. Mutation of positively selected sites in the lipase/feruloyl esterase A family of proteins resulted in loss of function in fungi (Levasseur et al. 2006), and similar results were found for Potato Virus Y (Moury & Simon 2011). These functional studies of positively selected residues corroborate our findings of positive selection on the SLPs of *C. difficile*, and strongly suggest a role for protein functional shift in severe CDI.

In addition to showing the link between positive selection and protein functional shift, phylogenetic analysis have been used to directly determine the role of positive selection in other infections. Positive selection has previously been detected on the pilus adhesion, FimH, of uropathogenic *Escherichia coli* (UPEC) (Chen et al. 2009). The receptor binding domain of FimH contains a mannose-specific binding pocket, yet positive selection was detected in residues away from this area. Despite maintaining binding to mannosylated receptors on the uroepithelium, mutation of two residues under positive selection resulted in a

10,000-fold reduction in colonisation of UPEC. This mutation also resulted in UPEC being unable to form intracellular bacterial communities (Chen et al. 2009). The PorB surface protein of *Neisseria meningitidis* has also been shown to exhibit signatures of positive selection, the residues under positive selection are predicted to be exposed to the host immune response with implications for pathogenesis (Urwin et al. 2002). Urwin *et al.* postulated that these positively selected sites were under strong immune selection (Urwin et al. 2002). These studies show a direct link between positive selection and pathogenicity, and highlight the importance of understanding these links in infections like CDI.

SLPs from ribotypes 027 and 078 consistently induced higher levels of cytokine production in macrophages and dendritic cells than SLPs of RT 001 and 014. This strongly suggests that the positively selected amino acid substitutions in *slpA* facilitate the production a more severe inflammatory response. The SLPs may therefore be contributing to the inflammatory state in the gut through interaction with the host immune system. As the SLPs are only a single component of the bacteria however, we used an animal model to see if these results directly translate to infection and disease.

The antibiotic cocktail mouse model for CDI (Chen et al. 2008) was used to determine if this trend of increased inflammation by positive selection on SLPs was evident in infection with the whole bacteria. Again, we observed high levels of pro-inflammatory cytokines in response to RT 027, with increased tissue damage and bacterial load in the gut relative to RT 001. Animals cleared infection of RT 001 after seven days, but it appears the immune response could

not mount an appropriate response against 027. This more severe infection in response to RT 027 agrees with our *in vitro* findings. The cytokine environment of the gut is a crucial determinate in how the adaptive immune system responds to pathogens, and may be the key to how *C. difficile* modulates the host to evade clearance while causing severe inflammation. RT 001 induces IL-6, IL-23 and TGF-β, all indicative of a Th17 response, and many studies have shown the protective role Th17 cells play in clearing extracellular pathogens, particularly in the gut (Khader et al. 2009; Zhang et al. 2009b; Blaschitz & Raffatellu 2010; Ishigame et al. 2009).

Positive selection acting on the SLPs of RT 027 may be altering the sequence motif responsible for activating TLR4 and mounting an immune response. High levels of TNFα in particular may be driving disease, having been associated with severe colitis in many studies (Plevy et al. 1997; Braegger et al. 1992), and anti-TNFα therapies being utilised as a viable treatment for inflammatory bowel disease (Lv et al. 2014; Su et al. 2002). High levels of IL-10 may be working to suppress components of the immune system necessary for clearance. IL-10 has been shown to inhibit Th17 development (Maynard & Weaver 2008), and the lack of Th17 specific cytokines, along with RORγ in the gut of 027-infected animals suggests that this is the case. Other pathogens are known to manipulate the immune system with IL-10 to avoid clearance (O'Leary et al. 2011), and *C. difficile* may use a similar mechanism. IFNγ has been shown to play a role in phagosome maturation (O'Leary et al. 2011; Hostetter et al. 2002), so its downregulation by 027 may prevent macrophages from destroying phagocytosed bacteria. Low IFNγ levels, in addition to high levels of IL-10 and TNFα all may

contribute to more severe and persistent disease. Activation of the MyD88-independent pathway leads to downstream activation of IRF3, and activation of both MyD88-dependent and -independent pathways results in higher levels of IL-10 (Boonstra et al. 2006; Chang et al. 2007). We found that SLPs from RT 027 induced expression of IRF3. The motif present in the SLPs of 027 likely modulates TLR4 signalling, increasing IL-10 production.

While the specific sequence motifs that TLR4 binds to on the SLPs are unknown, likely candidates are the areas under positive selection. As the SLPs from ribotypes 001 and 027 are the most closely related sequences in our dataset, only 027 SLPs can activate the MyD88-independent pathway, it appears that the small sequence differences under positive selection are playing a significant role in how the host mounts an immune response. It is also important to note that the responses observed *in vivo* cannot be attributed to SLPs alone. The production of toxins, along with the role of other adhesions plays a role in the pathology of *C. difficile*. However, our results *in vitro* showing high IL-17 induced by RT 001 SLPs and high IL-10 by 027 SLPs are consistent with the infection studies *in vivo*, and it is likely that SLP sequence is a crucial determinate in driving adaptive T cell responses to clear infection. Positive selection on SLPs from 027 appears to play a role in inhibiting the protective Th17 response observed in 001-infection, specifically by activating the MyD88-independent pathway downstream of TLR4, and increasing IL-10 expression. This shows us a link between positive selection of the SLPs and severity of infection.

Future studies on site-directed mutagenesis of the positively selected regions, and

the use of recombinant SLPs may help to further confirm the role of positively selected residues on the *slpA* gene in modulating the immune response. Such studies could also contribute to determining the amino acid motif by which TLR4 recognises *C. difficile* and mounts a response, thereby unmasking a potential therapeutic target.

The continued prevalence of antibiotic-resistant strains, along with the emergence of new hypervirulent strains, reveals the need for new forms of treatment against CDI. If a sequence motif with signatures of positive selection is responsible for modulating TLR4 to increase IL-10 production, then blocking this interaction may be a key therapeutic strategy. However, it is important to consider that a complete loss of recognition by TLR4 could have detrimental effects on infection. TLR5 also plays a role in mounting a response to *C. difficile* infection, specifically by recognising bacterial flagella (Yoshino et al. 2013). TLR5 stimulation has also been shown to have protective effects against *C. difficile* –induced colitis in mice (Jarchum et al. 2011), therefore in the case of strains undergoing positive selection on *slpA*, blocking TLR4-SLP interaction may help ameliorate severe symptoms of disease, while enhanced TLR5 activation could provide improved protective effects.

Antibody therapies against SLPs could potentially be a viable treatment for CDI, indeed anti-SLP antibody has previously been detected in human sera (Drudy et al. 2004). However, due to the high sequence variation between strains, a universal treatment targeting all variations of SLPs may prove difficult. Inhibiting the activity of the protease Cwp84 could potentially reduce severity of

CDI. As this cell wall protein is responsible for cleaving SlpA in the mature LMW and HMW subunits of the S-Layer (de la Riva et al. 2011), blocking the protease activity would result in the accumulation of immature SLP precursors at the cell wall. This would alter the conformation of the S-Layer, potentially inhibiting adherence of the gastrointestinal tract (Merrigan et al. 2013), and blocking SLP-TLR4 interaction.

Another potential therapy is the blocking of IL-10 to ensure an appropriate immune response is initiated. As this cytokine suppresses adaptive immune responses, anti-IL-10 antibody could reverse this immune suppression and help clear disease. Neutralisation of IL-10 has been shown to be beneficial in both bacterial and viral infections. Blocking of IL-10 has been shown to increase survival in mice infected with *Klebsiella pneumonia* (Greenberger et al. 1995), and administration of anti-IL-10 antibody effectively enhances resistance to *Mycobacterium avium* in mice (Bermudez & Champsi 1993). Similar results are seen in viral infections, with IL-$10^{-/-}$ mice exhibiting less severe disease in response to West Nile Virus, along with anti-IL-10 antibody providing protective effects (Bai et al. 2009). Blocking of IL-10 has even been shown to be effective in the treatment of parasitic infections, specifically in human visceral leishmaniasis (Gautam et al. 2011). These studies together highlight the potential importance of anti-IL-10 therapy in the treatment of severe CDI.

The idea that ribotype can be a reliable indicator of disease severity has been controversial, and conflicting positions are debated heavily in the literature (Walk et al. 2012; A Sarah Walker, Eyre, Crook, et al. 2013; A Sarah Walker,

Eyre, Wyllie, et al. 2013). However, we have shown a direct link between RT 027 and increased inflammation leading to severe disease. Specifically, we suggest that positive selection on the SLPs of RT 027 contributes significantly to increased severity of disease. We must exercise caution in comparing these results to global infections as recent studies have shown potential for recombination and horizontal gene transfer between the SLPs of different ribotypes (Dingle et al. 2013). This means that individual isolates of a particular ribotype may possess unique SLP sequences. We confirmed by sequencing that all SLPs analysed in this thesis correspond directly to the sequences from our bioinformatics analysis. We can therefore directly correlate positive selection analysis with our *in vitro* and *in vivo* findings. While these finding may not be true for all strains designated RT 027, they provide an important insight into the role of SLPs in infection, along with the mechanisms by which they modulate the immune response.

In conclusion, positive selection on the *slpA* sequences of *C. difficile* appears to induce a protein functional shift, and modulate TLR4 signalling. In the case of hypervirulent RT 027, this manipulation activates IRF3, increasing IL-10 production in the gut, and suppressing adaptive responses such as a protective Th17 response. In the absence of this functional shift, infection appears less severe, with successful clearance by the host immune system. Identification of these positively selected regions of *slpA* gives us greater insight into the molecular mechanisms of CDI, and reveals novel potential therapeutic targets for future treatment of the disease. A diagram summarising this process can be seen in Figure 6.1.

Figure 6.1

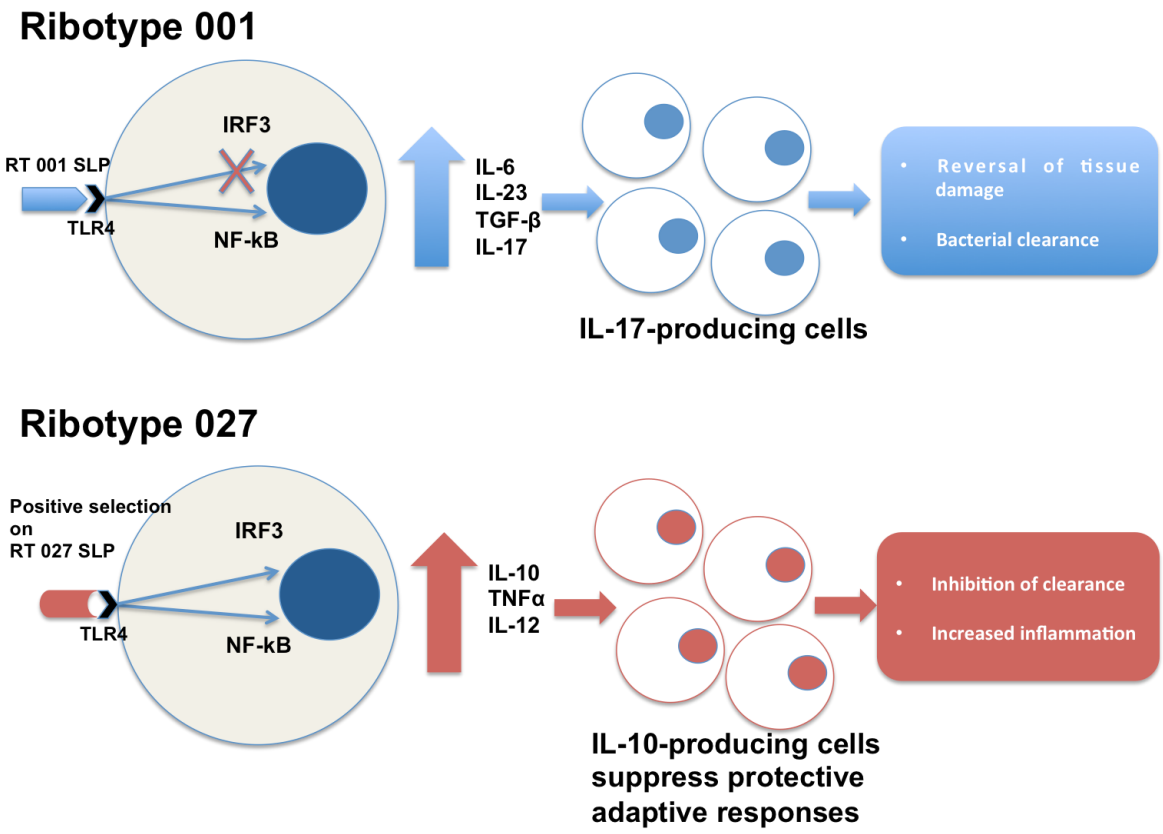


Figure 6.1 Summary of immune modulation by SLPs

SLPs from RT 001 activate NF-kB via TLR4 activation to stimulate proinflammatory cytokine production. High levels of IL-6, IL-23 and TGF-β may help drive an IL-17-heavy response, which, along with phagocytosis, aids bacterial clearance and tissue repair. Positive selection on the SLPs of RT 027 modulates TLR4 signalling to induce IRF3 activation, inducing high levels of IL-10, which can block adaptive immune responses. Later induction of proinflammatory cytokines exacerbates tissue damage, while IL-10 suppresses clearance, leading to persistent disease.

Chapter 7

Bibliography

Aderem, A. & Underhill, D.M., 1999. Mechanisms of phagocytosis in macrophages. *Annual review of immunology*, 17, pp.593–623. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10358769 [Accessed February 18, 2014].

Akira, S. & Takeda, K., 2004. Toll-like receptor signalling. *Nature reviews. Immunology*, 4(7), pp.499–511. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15229469 [Accessed May 23, 2014].

Akira, S., Takeda, K. & Kaisho, T., 2001. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nature immunology*, 2(8), pp.675–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11477402 [Accessed June 16, 2014].

Alexopoulou, L. et al., 2001. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature*, 413(6857), pp.732–8. Available at: http://dx.doi.org/10.1038/35099560 [Accessed June 3, 2014].

Anisimova, M., Bielawski, J.P. & Yang, Z., 2002. Accuracy and power of bayes prediction of amino acid sites under positive selection. *Molecular biology and evolution*, 19(6), pp.950–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12032251 [Accessed February 24, 2014].

Anisimova, M., Bielawski, J.P. & Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular biology and evolution*, 18(8), pp.1585–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11470850.

Ariel, A. et al., 2006. Apoptotic neutrophils and T cells sequester chemokines during immune response resolution through modulation of CCR5 expression. *Nature immunology*, 7(11), pp.1209–16. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1797066&tool=pmcentrez&rendertype=abstract [Accessed June 17, 2014].

Arvand, M. et al., 2009. Clostridium difficile ribotypes 001, 017, and 027 are associated with lethal C. difficile infection in Hesse, Germany. *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, 14(45). Available at: http://www.ncbi.nlm.nih.gov/pubmed/19941785 [Accessed July 2, 2014].

Ausiello, C.M. et al., 2006. Surface layer proteins from Clostridium difficile induce inflammatory and regulatory cytokines in human monocytes and dendritic cells. *Microbes and infection / Institut Pasteur*, 8(11), pp.2640–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16935543 [Accessed October 17, 2012].

Bahl, H. et al., 1997. Molecular biology of S-layers. *FEMS microbiology reviews*, 20(1-2), pp.47–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9276928 [Accessed June 18, 2014].

Bai, F. et al., 2009. IL-10 signaling blockade controls murine West Nile virus infection. M. Gale, ed. *PLoS pathogens*, 5(10), p.e1000610. Available at: http://dx.plos.org/10.1371/journal.ppat.1000610 [Accessed July 15, 2014].

Bain, C.C. et al., 2013. Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6Chi monocyte precursors. *Mucosal immunology*, 6(3), pp.498–510. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3629381&tool=pmcentrez&rendertype=abstract [Accessed June 3, 2014].

Baines, S.D. et al., 2009. Activity of vancomycin against epidemic Clostridium difficile strains in a human gut model. *The Journal of antimicrobial chemotherapy*, 63(3), pp.520–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19112083 [Accessed June 20, 2014].

Bartlett, J.G. et al., 1978. Antibiotic-induced lethal enterocolitis in hamsters: studies with eleven agents and evidence to support the pathogenic role of toxin-producing Clostridia. *American journal of veterinary research*, 39(9), pp.1525–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/697162 [Accessed June 23, 2014].

Bartlett, J.G. et al., 1977. Clindamycin-associated colitis due to a toxin-producing species of Clostridium in hamsters. *The Journal of infectious diseases*, 136(5), pp.701–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/915343 [Accessed June 23, 2014].

Bauer, M.P. et al., 2011. Clostridium difficile infection in Europe: a hospital-based survey. *Lancet*, 377(9759), pp.63–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21084111 [Accessed June 13, 2014].

Beagley, K.W. et al., 1995. Differences in intraepithelial lymphocyte T cell subsets isolated from murine small versus large intestine. *Journal of immunology (Baltimore, Md. : 1950)*, 154(11), pp.5611–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7751614 [Accessed July 8, 2014].

Bekker, L.G. et al., 2000. Immunopathologic effects of tumor necrosis factor alpha in murine mycobacterial infection are dose dependent. *Infection and immunity*, 68(12), pp.6954–61. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=97804&tool=pmcentrez&rendertype=abstract.

Bennouna, S. & Denkers, E.Y., 2005. Microbial antigen triggers rapid mobilization of TNF-alpha to the surface of mouse neutrophils transforming them into inducers of high-level dendritic cell TNF-alpha production. *Journal of immunology (Baltimore, Md. : 1950)*, 174(8), pp.4845–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15814711 [Accessed July 7, 2014].

Bermudez, L.E. & Champsi, J., 1993. Infection with Mycobacterium avium induces production of interleukin-10 (IL-10), and administration of anti-IL-10 antibody is associated with enhanced resistance to infection in mice. *Infect. Immun.*, 61(7), pp.3093–3097. Available at: http://iai.asm.org/content/61/7/3093.short [Accessed July 15, 2014].

Berndt, B.E. et al., 2007. The role of dendritic cells in the development of acute dextran sulfate sodium colitis. *Journal of immunology (Baltimore, Md. : 1950)*, 179(9), pp.6255–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17947701 [Accessed July 8, 2014].

Best, E.L., Freeman, J. & Wilcox, M.H., 2011. Models for the study of Clostridium difficile infection. *Gut microbes*, 3(2), pp.145–67. Available at: /pmcc/articles/PMC3370947/?report=abstract [Accessed May 14, 2014].

Blackburne, B.P. & Whelan, S., 2013. Class of multiple sequence alignment algorithm affects genomic analysis. *Molecular biology and evolution*, 30(3), pp.642–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23144040 [Accessed May 26, 2014].

Blaschitz, C. & Raffatellu, M., 2010. Th17 cytokines and the gut mucosal barrier. *Journal of clinical immunology*, 30(2), pp.196–203. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2842875&tool=pmcentrez&rendertype=abstract [Accessed June 6, 2014].

Boonstra, A. et al., 2006. Macrophages and myeloid dendritic cells, but not plasmacytoid dendritic cells, produce IL-10 in response to MyD88- and TRIF-dependent TLR signals, and TLR-independent signals. *Journal of immunology (Baltimore, Md. : 1950)*, 177(11), pp.7551–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17114424 [Accessed July 3, 2014].

Borregaard, N., 2010. Neutrophils, from marrow to microbes. *Immunity*, 33(5), pp.657–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21094463 [Accessed May 27, 2014].

Braegger, C.P. et al., 1992. Tumour necrosis factor alpha in stool as a marker of intestinal inflammation. *Lancet*, 339(8785), pp.89–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1345871 [Accessed July 2, 2014].

Brikos, C. & O'Neill, L.A.J., 2008. Signalling of toll-like receptors. *Handbook of experimental pharmacology*, (183), pp.21–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18071653 [Accessed June 30, 2014].

Buffie, C.G. et al., 2012. Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. *Infection and immunity*, 80(1), pp.62–73. Available at: http://iai.asm.org/content/80/1/62.full [Accessed June 18, 2014].

Calabi, E. et al., 2002. Binding of Clostridium difficile Surface Layer Proteins to Gastrointestinal Tissues Binding of Clostridium difficile Surface Layer Proteins to Gastrointestinal Tissues. , 70(10), pp.5770–5778.

Calabi, E. et al., 2001. Molecular characterization of the surface layer proteins from Clostridium difficile. *Molecular microbiology*, 40(5), pp.1187–99. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11401722.

Carter, G.P. et al., 2011. The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of Clostridium difficile. T. M. Koehler, ed. *PLoS pathogens*, 7(10), p.e1002317. Available at: http://dx.plos.org/10.1371/journal.ppat.1002317 [Accessed June 4, 2014].

Cerquetti, M. et al., 2000. Characterization of surface layer proteins from different Clostridium difficile clinical isolates. *Microbial pathogenesis*, 28(6), pp.363–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10839973 [Accessed December 7, 2012].

Chang, E.Y. et al., 2007. Cutting edge: involvement of the type I IFN production and signaling pathway in lipopolysaccharide-induced IL-10 production. *Journal of immunology (Baltimore, Md. : 1950)*, 178(11), pp.6705–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17513714 [Accessed July 3, 2014].

Chen, S.L. et al., 2009. Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52), pp.22439–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2794649&tool=pmcentrez&rendertype=abstract [Accessed July 16, 2014].

Chen, X. et al., 2008. A mouse model of Clostridium difficile-associated disease. *Gastroenterology*, 135(6), pp.1984–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18848941 [Accessed April 30, 2014].

Cheroutre, H., Lambolez, F. & Mucida, D., 2011. The light and dark sides of intestinal intraepithelial lymphocytes. *Nature reviews. Immunology*, 11(7), pp.445–56. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3140792&tool=pmcentrez&rendertype=abstract [Accessed May 26, 2014].

Chieppa, M. et al., 2006. Dynamic imaging of dendritic cell extension into the small bowel lumen in response to epithelial cell TLR engagement. *The Journal of experimental medicine*, 203(13), pp.2841–52. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2118178&tool=pmcentrez&rendertype=abstract [Accessed May 23, 2014].

Clements, A.C.A. et al., 2010. Clostridium difficile PCR ribotype 027: assessing the risks of further worldwide spread. *The Lancet infectious …*, 10(C), pp.395–404. Available at: http://www.thelancet.com/journals/a/article/PIIS1473-3099(10)70080-3/fulltext [Accessed January 30, 2014].

Cohen, S.H. et al., 2010. Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America*, 31(5), pp.431–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20307191 [Accessed May 25, 2014].

Collins, L.E. et al., 2014. Surface layer proteins isolated from Clostridium difficile induce clearance responses in macrophages. *Microbes and Infection*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24560642 [Accessed February 24, 2014].

Colotta, F. et al., 1992. Modulation of granulocyte survival and programmed cell death by cytokines and bacterial products. *Blood*, 80(8), pp.2012–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1382715 [Accessed June 8, 2014].

Coombes, J.L. et al., 2007. A functionally specialized population of mucosal CD103+ DCs induces Foxp3+ regulatory T cells via a TGF-beta and retinoic acid-dependent mechanism. *The Journal of experimental medicine*, 204(8), pp.1757–64. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2118683&tool=pmcentrez&rendertype=abstract [Accessed May 23, 2014].

Coombes, J.L. & Powrie, F., 2008. Dendritic cells in intestinal immune regulation. *Nature reviews. Immunology*, 8(6), pp.435–46. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2674208&tool=pmcentrez&rendertype=abstract [Accessed May 23, 2014].

Dang, T.H.T. et al., 2010. Chemical probes of surface layer biogenesis in Clostridium difficile. *ACS chemical biology*, 5(3), pp.279–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20067320.

Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection*,

Darwin, C. & Wallace, A., 1858. On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *Journal of the Proceedings of the Linnean Society of London. Zoology*, 3(9), pp.45–62. Available at: http://doi.wiley.com/10.1111/j.1096-3642.1858.tb02500.x [Accessed February 25, 2014].

Dawson, L.F. et al., 2011. Hypervirulent Clostridium difficile PCR-ribotypes exhibit resistance to widely used disinfectants. *PloS one*, 6(10), p.e25754. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3201945&tool=pmcentrez&rendertype=abstract [Accessed October 18, 2012].

Dawson, L.F., Valiente, E. & Wren, B.W., 2009. Clostridium difficile--a continually evolving and problematic pathogen. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 9(6), pp.1410–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19539054 [Accessed October 17, 2012].

Denève, C. et al., 2009. New trends in Clostridium difficile virulence and pathogenesis. *International journal of antimicrobial agents*, 33 Suppl 1, pp.S24–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19303565.

Dimitrieva, S. & Anisimova, M., 2014. Unraveling Patterns of Site-to-Site Synonymous Rates Variation and Associated Gene Properties of Protein Domains and Families. *PloS one*, 9(6), p.e95034. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4045579&tool=pmcentrez&rendertype=abstract [Accessed June 5, 2014].

Dingle, K.E. et al., 2013. Recombinational switching of the Clostridium difficile S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing. *The Journal of infectious diseases*, 207(4), pp.675–86. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3549603&tool=pmcentrez&rendertype=abstract [Accessed May 27, 2014].

Doe, W.F., 1989. The intestinal immune system. *Gut*, 30(12), pp.1679–85. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1434429&tool=pmcentrez&rendertype=abstract [Accessed June 26, 2014].

Doolittle, W.F., 2000. Uprooting the tree of life. *Scientific American*, 282(2), pp.90–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10710791 [Accessed May 26, 2014].

Drudy, D., Harnedy, N., et al., 2007. Emergence and control of fluoroquinolone-resistant, toxin A-negative, toxin B-positive Clostridium difficile. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America*, 28(8), pp.932–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17620240 [Accessed October 18, 2012].

Drudy, D. et al., 2004. Human antibody response to surface layer proteins in Clostridium difficile infection. *FEMS immunology and medical microbiology*, 41(3), pp.237–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15196573 [Accessed October 17, 2012].

Drudy, D., Fanning, S. & Kyne, L., 2007. Toxin A-negative, toxin B-positive Clostridium difficile. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 11(1), pp.5–10. Available at: http://www.sciencedirect.com/science/article/pii/S1201971206000919 [Accessed February 15, 2014].

Du, J. et al., 2014. Selection on synonymous codons in mammalian rhodopsins: a possible role in optimizing translational processes. *BMC evolutionary biology*, 14, p.96. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4021273&tool=pmcentrez&rendertype=abstract [Accessed June 5, 2014].

Dupuy, B. et al., 2008. Clostridium difficile toxin synthesis is negatively regulated by TcdC. *Journal of medical microbiology*, 57(Pt 6), pp.685–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18480323 [Accessed October 5, 2012].

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, 5(1), p.113. Available at: http://www.biomedcentral.com/1471-2105/5/113 [Accessed February 20, 2014].

Eidhin, D.N. et al., 2006. Sequence and phylogenetic analysis of the gene for surface layer protein, slpA, from 14 PCR ribotypes of Clostridium difficile. *Journal of medical microbiology*, 55(Pt 1), pp.69–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16388033 [Accessed December 22, 2012].

Endo, T., Ikeo, K. & Gojobori, T., 1996. Large-scale search for genes on which positive selection may operate. *Molecular biology and evolution*, 13(5), pp.685–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8676743 [Accessed February 25, 2014].

Fagan, R.P. et al., 2011. A proposed nomenclature for cell wall proteins of Clostridium difficile. *Journal of medical microbiology*, 60(Pt 8), pp.1225–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21252271 [Accessed October 18, 2012].

Fagan, R.P. et al., 2009. Structural insights into the molecular organization of the S-layer from Clostridium difficile. *Molecular microbiology*, 71(5), pp.1308–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19183279 [Accessed February 11, 2014].

Fares, M.A. et al., 2002. A sliding window-based method to detect selective constraints in protein-coding genes and its application to RNA viruses. *Journal of molecular evolution*, 55(5), pp.509–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12399925 [Accessed February 25, 2014].

Farhat, K. et al., 2008. Heterodimerization of TLR2 with TLR1 or TLR6 expands the ligand spectrum but does not lead to differential signaling. *Journal of leukocyte biology*, 83(3), pp.692–701. Available at: http://www.jleukbio.org/content/83/3/692.full [Accessed May 23, 2014].

Feng, T. et al., 2011. Th17 cells induce colitis and promote Th1 cell responses through IL-17 induction of innate IL-12 and IL-23 production. *Journal of immunology (Baltimore, Md. : 1950)*, 186(11), pp.6313–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3249225&tool=pmcentrez&rendertype=abstract [Accessed June 11, 2014].

Fujio, K., Okamura, T. & Yamamoto, K., 2010. The Family of IL-10-secreting CD4+ T cells. *Advances in immunology*, 105, pp.99–130. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20510731 [Accessed July 8, 2014].

Gautam, S. et al., 2011. IL-10 neutralization promotes parasite clearance in splenic aspirate cells from patients with visceral leishmaniasis. *The Journal of infectious diseases*, 204(7), pp.1134–7. Available at: http://jid.oxfordjournals.org/content/204/7/1134.short [Accessed July 16, 2014].

Genth, H. et al., 2008. Clostridium difficile toxins: more than mere inhibitors of Rho proteins. *The international journal of biochemistry & cell biology*, 40(4), pp.592–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18289919 [Accessed October 18, 2012].

Ghose, C., 2013. Clostridium difficile infection in the twenty-first century. *Emerging Microbes & Infections*, 2(9), p.e62. Available at: http://dx.doi.org/10.1038/emi.2013.62 [Accessed May 14, 2014].

Giel, J.L. et al., 2010. Metabolism of bile salts in mice influences spore germination in Clostridium difficile. A. J. Ratner, ed. *PloS one*, 5(1), p.e8740. Available at: http://dx.plos.org/10.1371/journal.pone.0008740 [Accessed June 20, 2014].

Gill, S.R. et al., 2006. Metagenomic analysis of the human distal gut microbiome. *Science (New York, N.Y.)*, 312(5778), pp.1355–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3027896&tool=pmcentrez&rendertype=abstract [Accessed May 26, 2014].

Goldman, N. & Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular biology and evolution*, 11(5), pp.725–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7968486 [Accessed February 16, 2014].

Goorhuis, A. et al., 2008. Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 47(9), pp.1162–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18808358 [Accessed February 27, 2014].

Goto, Y. & Ivanov, I.I., 2013. Intestinal epithelial cells as mediators of the commensal-host immune crosstalk. *Immunology and cell biology*, 91(3), pp.204–14. Available at: http://dx.doi.org/10.1038/icb.2012.80 [Accessed May 29, 2014].

Greenberger, M.J. et al., 1995. Neutralization of IL-10 increases survival in a murine model of Klebsiella pneumonia. *Journal of immunology (Baltimore, Md. : 1950)*, 155(2), pp.722–9. Available at: http://www.jimmunol.org/content/155/2/722.abstract [Accessed July 15, 2014].

Grewal, I.S. & Flavell, R.A., 1996. The role of CD40 ligand in costimulation and T-cell activation. *Immunological reviews*, 153, pp.85–106. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9010720 [Accessed July 2, 2014].

Gu, X. & Li, W.H., 1992. Higher rates of amino acid substitution in rodents than in humans. *Molecular phylogenetics and evolution*, 1(3), pp.211–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1342937 [Accessed February 25, 2014].

Hasegawa, M. & Hashimoto, T., 1993. Ribosomal RNA trees misleading? *Nature*, 361(6407), p.23. Available at: http://dx.doi.org/10.1038/361023b0 [Accessed June 17, 2014].

Hashimoto, S. et al., 1996. Effects of beta-lactam antibiotics on intestinal microflora and bile acid metabolism in rats. *Lipids*, 31(6), pp.601–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8784740 [Accessed June 20, 2014].

Hayashi, F. et al., 2001. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*, 410(6832), pp.1099–103. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11323673 [Accessed July 8, 2014].

Hayashi, F., Means, T.K. & Luster, A.D., 2003. Toll-like receptors stimulate human neutrophil function. *Blood*, 102(7), pp.2660–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12829592 [Accessed June 25, 2014].

Hemmi, H. et al., 2000. A Toll-like receptor recognizes bacterial DNA. *Nature*, 408(6813), pp.740–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11130078 [Accessed June 9, 2014].

Higgins, M.K. & Carrington, M., 2014. Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface protein families. *Protein science : a publication of the Protein Society*, 23(4), pp.354–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24442723 [Accessed May 30, 2014].

Himmel, M.E. et al., 2011. Human CD4+ FOXP3+ regulatory T cells produce CXCL8 and recruit neutrophils. *European journal of immunology*, 41(2), pp.306–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21268001 [Accessed July 7, 2014].

Hostetter, J.M. et al., 2002. Cytokine effects on maturation of the phagosomes containing Mycobacteria avium subspecies paratuberculosis in J774 cells. *FEMS immunology and medical microbiology*, 34(2), pp.127–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12381463 [Accessed June 29, 2014].

Hu, X. et al., 2006. IFN-gamma suppresses IL-10 production and synergizes with TLR2 by regulating GSK3 and CREB/AP-1 proteins. *Immunity*, 24(5), pp.563–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16713974 [Accessed June 16, 2014].

Huang, F.P. et al., 2000. A discrete subpopulation of dendritic cells transports apoptotic intestinal epithelial cells to T cell areas of mesenteric lymph nodes. *The Journal of experimental medicine*, 191(3), pp.435–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2195813&tool=pmcentrez&rendertype=abstract [Accessed July 8, 2014].

Huelsenbeck, J.P. & Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)*, 17(8), pp.754–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11524383 [Accessed February 21, 2014].

Hughes, A.L., 1999. *Adaptive Evolution of Genes and Genomes*, Oxford University Press.

Hughes, A.L., 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity*, 99(4), pp.364–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17622265 [Accessed June 9, 2014].

Hughes, A.L. & Nei, M., 1989. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Molecular biology and evolution*, 6(6), pp.559–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2484936 [Accessed June 18, 2014].

Hurst, L.D. & Pál, C., 2001. Evidence for purifying selection acting on silent sites in BRCA1. *Trends in genetics : TIG*, 17(2), pp.62–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11173101 [Accessed February 25, 2014].

Huson, D.H. et al., 2007. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC bioinformatics*, 8(1), p.460. Available at: http://www.biomedcentral.com/1471-2105/8/460 [Accessed February 20, 2014].

Ishigame, H. et al., 2009. Differential roles of interleukin-17A and -17F in host defense against mucoepithelial bacterial infection and allergic responses. *Immunity*, 30(1), pp.108–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19144317 [Accessed June 12, 2014].

Iwasaki, A., 2007. Mucosal dendritic cells. *Annual review of immunology*, 25, pp.381–418. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17378762 [Accessed May 26, 2014].

Izcue, A., Coombes, J.L. & Powrie, F., 2009. Regulatory lymphocytes and intestinal inflammation. *Annual review of immunology*, 27, pp.313–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19302043 [Accessed May 27, 2014].

Janezic, S. et al., 2012. Clostridium difficile genotypes other than ribotype 078 that are prevalent among human, animal and environmental isolates. *BMC microbiology*, 12(1), p.48. Available at: http://www.biomedcentral.com/1471-2180/12/48 [Accessed June 28, 2014].

Janoir, C. et al., 2007. Cwp84, a surface-associated protein of Clostridium difficile, is a cysteine protease with degrading activity on extracellular matrix proteins. *Journal of bacteriology*, 189(20), pp.7174–80. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2168428&tool=pmcentrez&rendertype=abstract [Accessed October 18, 2012].

Jarchum, I. et al., 2012. Critical role for MyD88-mediated neutrophil recruitment during Clostridium difficile colitis. *Infection and immunity*, 80(9), pp.2989–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22689818 [Accessed October 18, 2012].

Jarchum, I. et al., 2011. Toll-like receptor 5 stimulation protects mice from acute Clostridium difficile colitis. *Infection and immunity*, 79(4), pp.1498–503. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3067529&tool=pmcentrez&rendertype=abstract [Accessed July 11, 2014].

Jeanmougin, F. et al., 1998. Multiple sequence alignment with Clustal X. *Trends in biochemical sciences*, 23(10), pp.403–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9810230 [Accessed February 25, 2014].

Jiang, Z. et al., 2005. CD14 is required for MyD88-independent LPS signaling. *Nature immunology*, 6(6), pp.565–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15895089 [Accessed May 27, 2014].

Jung, C., Hugot, J.-P. & Barreau, F., 2010. Peyer's Patches: The Immune Sensors of the Intestine. *International journal of inflammation*, 2010, p.823710. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3004000&tool=pmcentrez&rendertype=abstract [Accessed June 6, 2014].

Kachrimanidou, M. & Malisiovas, N., 2011. Clostridium difficile infection: a comprehensive review. *Critical reviews in microbiology*, 37(3), pp.178–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21609252 [Accessed February 24, 2014].

Kagan, J.C. et al., 2008. TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. *Nature immunology*, 9(4), pp.361–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18297073 [Accessed May 23, 2014].

Kamada, N. et al., 2005. Abnormally differentiated subsets of intestinal macrophage play a key role in Th1-dominant chronic colitis through excess production of IL-12 and IL-23 in response to bacteria. *Journal of immunology (Baltimore, Md. : 1950)*, 175(10), pp.6900–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16272349 [Accessed June 16, 2014].

Karjalainen, T. et al., 2002. Clostridium difficile genotyping based on slpA variable region in S-layer gene sequence: an alternative to serotyping. *Journal of clinical …*, 40(7), pp.2452–2458. Available at: http://jcm.asm.org/content/40/7/2452.long [Accessed February 12, 2014].

Karjalainen, T. et al., 2001. Molecular and genomic analysis of genes encoding surface-anchored proteins from Clostridium difficile. *Infection and …*, 69(5), pp.3442–3446. Available at: http://iai.asm.org/content/69/5/3442.full [Accessed December 7, 2012].

Kato, H. et al., 2009. Rapid analysis of Clostridium difficile strains recovered from hospitalized patients by using the slpA sequence typing system. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*, 15(3), pp.199–202. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19554407 [Accessed October 18, 2012].

Keane, T.M. et al., 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology*, 6, p.29. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1435933&tool=pmcentrez&rendertype=abstract [Accessed January 21, 2014].

Keel, M.K. & Songer, J.G., 2006. The comparative pathology of Clostridium difficile-associated disease. *Veterinary pathology*, 43(3), pp.225–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16672570 [Accessed June 23, 2014].

Kern, J. & Schneewind, O., 2010. BslA, the S-layer adhesin of B. anthracis, is a virulence factor for anthrax pathogenesis. *Molecular microbiology*, 75(2), pp.324–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2828814&tool=pmcentrez&rendertype=abstract [Accessed June 18, 2014].

Khader, S.A., Gaffen, S.L. & Kolls, J.K., 2009. Th17 cells at the crossroads of innate and adaptive immunity against infectious diseases at the mucosa. *Mucosal immunology*, 2(5), pp.403–11. Available at: http://dx.doi.org/10.1038/mi.2009.100 [Accessed June 12, 2014].

Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature*, 217(5129), pp.624–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5637732 [Accessed February 25, 2014].

King, J.L. & Jukes, T.H., 1969. Non-Darwinian evolution. *Science (New York, N.Y.)*, 164(3881), pp.788–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5767777 [Accessed June 18, 2014].

Kinugasa, T. et al., 2000. Claudins regulate the intestinal barrier in response to immune mediators. *Gastroenterology*, 118(6), pp.1001–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10833473 [Accessed May 29, 2014].

König, R., Huang, L.Y. & Germain, R.N., 1992. MHC class II interaction with CD4 mediated by a region analogous to the MHC class I binding site for CD8. *Nature*, 356(6372), pp.796–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1574118 [Accessed July 2, 2014].

Korn, T. et al., 2009. IL-17 and Th17 Cells. *Annual review of immunology*, 27, pp.485–517. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.immunol.021908.132710 [Accessed May 26, 2014].

Kosiol, C. et al., 2008. Patterns of positive selection in six Mammalian genomes. *PLoS genetics*, 4(8), p.e1000144. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2483296&tool=pmcentrez&rendertype=abstract [Accessed January 21, 2014].

Kreitman, M. & Akashi, H., 1995. Molecular Evidence for Natural Selection. *Annual Review of Ecology and Systematics*, 26(1), pp.403–422. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.es.26.110195.002155 [Accessed February 25, 2014].

Kuhner, M.K. & Felsenstein, J., 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular biology and evolution*, 11(3), pp.459–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8015439 [Accessed June 18, 2014].

Kullberg, M.C. et al., 1998. Helicobacter hepaticus triggers colitis in specific-pathogen-free interleukin-10 (IL-10)-deficient mice through an IL-12- and gamma interferon-dependent mechanism. *Infection and immunity*, 66(11), pp.5157–66. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=108643&tool=pmcentrez&rendertype=abstract [Accessed June 16, 2014].

Kurihara, T. et al., 1997. Defects in Macrophage Recruitment and Host Defense in Mice Lacking the CCR2 Chemokine Receptor. *The Journal of experimental …*, 186(10). Available at: http://jem.rupress.org/content/186/10/1757.abstract [Accessed July 31, 2013].

De la Riva, L. et al., 2011. Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of Clostridium difficile. *Journal of bacteriology*, 193(13), pp.3276–85. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3133288&tool=pmcentrez&rendertype=abstract [Accessed October 18, 2012].

Lanier, L.L. et al., 1995. CD80 (B7) and CD86 (B70) provide similar costimulatory signals for T cell proliferation, cytokine production, and generation of CTL. *Journal of immunology (Baltimore, Md. : 1950)*, 154(1), pp.97–105. Available at: http://www.jimmunol.org/content/154/1/97.abstract [Accessed July 2, 2014].

Lartillot, N., Lepage, T. & Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics (Oxford, England)*, 25(17), pp.2286–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19535536 [Accessed February 20, 2014].

Lawley, T.D. et al., 2009. Antibiotic treatment of clostridium difficile carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infection and immunity*, 77(9), pp.3661–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2737984&tool=pmcentrez&rendertype=abstract [Accessed January 29, 2014].

Lawley, T.D. & Young, V.B., 2013. Murine models to study Clostridium difficile infection and transmission. *Anaerobe*, 24, pp.94–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24076318 [Accessed June 19, 2014].

Lee, S.H., Starkey, P.M. & Gordon, S., 1985. Quantitative analysis of total macrophage content in adult mouse tissues. Immunochemical studies with monoclonal antibody F4/80. *The Journal of experimental medicine*, 161(3), pp.475–89. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2187577&tool=pmcentrez&rendertype=abstract [Accessed July 7, 2014].

Levasseur, A. et al., 2006. Tracking the connection between evolutionary and functional shifts using the fungal lipase/feruloyl esterase A family. *BMC evolutionary biology*, 6, p.92. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1660568&tool=pmcentrez&rendertype=abstract [Accessed February 11, 2014].

Ley, R.E., Peterson, D.A. & Gordon, J.I., 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, 124(4), pp.837–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16497592 [Accessed May 23, 2014].

Li, W.H., Wu, C.I. & Luo, C.C., 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular biology and evolution*, 2(2), pp.150–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3916709 [Accessed February 24, 2014].

Libby, J.M., Jortner, B.S. & Wilkins, T.D., 1982. Effects of the two toxins of Clostridium difficile in antibiotic-associated cecitis in hamsters. *Infection and immunity*, 36(2), pp.822–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=351302&tool=pmcentrez&rendertype=abstract [Accessed June 23, 2014].

Loomis, W.F. & Smith, D.W., 1990. Molecular phylogeny of Dictyostelium discoideum by protein sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, 87(23), pp.9093–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=55110&tool=pmcentrez&rendertype=abstract [Accessed June 17, 2014].

Loughran, N.B. et al., 2012. Functional consequence of positive selection revealed through rational mutagenesis of human myeloperoxidase. *Molecular biology and evolution*, 29(8), pp.2039–46. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3408071&tool=pmcentrez&rendertype=abstract [Accessed February 19, 2014].

Lv, R. et al., 2014. Tumor necrosis factor alpha blocking agents as treatment for ulcerative colitis intolerant or refractory to conventional medical therapy: a meta-analysis. *PloS one*, 9(1), p.e86692. Available at: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0086692#pone.0086692-Braegger1 [Accessed June 16, 2014].

Lynch, M. & Hill, W.G., 1986. Phenotypic Evolution by Neutral Mutation. *Evolution*, 40(5), pp.915–935.

MacArthur, D.G. et al., 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, N.Y.)*, 335(6070), pp.823–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3299548&tool=pmcentrez&rendertype=abstract [Accessed June 2, 2014].

Macpherson, A.J. & Uhr, T., 2004. Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science (New York, N.Y.)*, 303(5664), pp.1662–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15016999 [Accessed May 28, 2014].

Madan, R. & Jr, W. a P., 2012. Immune responses to Clostridium difficile infection. *Trends in molecular medicine*, 18(11), pp.658–666. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23084763 [Accessed October 29, 2012].

Madsen, O. et al., 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature*, 409(6820), pp.610–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11214318 [Accessed June 26, 2014].

Marsh, J.W. et al., 2012. Association of relapse of Clostridium difficile disease with BI/NAP1/027. *Journal of clinical microbiology*, 50(12), pp.4078–82. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3502988&tool=pmcentrez&rendertype=abstract [Accessed May 18, 2014].

Maynard, C.L. & Weaver, C.T., 2008. Diversity in the contribution of interleukin-10 to T-cell-mediated immune regulation. *Immunological reviews*, 226, pp.219–33. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2630587&tool=pmcentrez&rendertype=abstract [Accessed June 16, 2014].

McCoubrey, J. & Poxton, I.R.I., 2001. Variation in the surface layer proteins of Clostridium difficile. *FEMS Immunology & Medical …*, 31(2), pp.131–135. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1574-695X.2001.tb00509.x/full [Accessed February 12, 2014].

McLoughlin, R.M. et al., 2008. IFN-gamma regulated chemokine production determines the outcome of Staphylococcus aureus infection. *Journal of immunology (Baltimore, Md. : 1950)*, 181(2), pp.1323–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18606687 [Accessed June 16, 2014].

Merrigan, M.M. et al., 2013. Surface-layer protein A (SlpA) is a major contributor to host-cell adherence of Clostridium difficile. M. R. Popoff, ed. *PloS one*, 8(11), p.e78404. Available at: http://dx.plos.org/10.1371/journal.pone.0078404 [Accessed March 21, 2014].

Mestas, J. & Hughes, C.C.W., 2004. Of Mice and Not Men: Differences between Mouse and Human Immunology. *The Journal of Immunology*, 172(5), pp.2731–2738. Available at: http://www.jimmunol.org/content/172/5/2731.full [Accessed June 4, 2014].

Moolenbeek, C. & Ruitenberg, E.J., 1981. The "Swiss roll": a simple technique for histological studies of the rodent intestine. *Laboratory animals*, 15(1), pp.57–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7022018 [Accessed May 27, 2014].

Morgan, C.C. et al., 2010. Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. *BMC evolutionary biology*, 10(1), p.39. Available at: http://www.biomedcentral.com/1471-2148/10/39 [Accessed February 24, 2014].

Mosser, D.M. & Edwards, J.P., 2008. Exploring the full spectrum of macrophage activation. *Nature reviews. Immunology*, 8(12), pp.958–69. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2724991&tool=pmcentrez&rendertype=abstract [Accessed October 5, 2012].

Moury, B. & Simon, V., 2011. dN/dS-based methods detect positive selection linked to trade-offs between different fitness traits in the coat protein of potato virus Y. *Molecular biology and evolution*, 28(9), pp.2707–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21498601 [Accessed June 3, 2014].

Mowat, A.M., 2003. Anatomical basis of tolerance and immunity to intestinal antigens. *Nature reviews. Immunology*, 3(4), pp.331–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12669023 [Accessed May 23, 2014].

Murch, S.H. et al., 1991. Serum concentrations of tumour necrosis factor alpha in childhood chronic inflammatory bowel disease. *Gut*, 32(8), pp.913–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1378961&tool=pmcentrez&rendertype=abstract [Accessed June 16, 2014].

Murray, P.J. & Wynn, T. a, 2011. Protective and pathogenic functions of macrophage subsets. *Nature reviews. Immunology*, 11(11), pp.723–37. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3422549&tool=pmcentrez&rendertype=abstract [Accessed July 30, 2013].

Nathan, C., 2006. Neutrophils and immunity: challenges and opportunities. *Nature reviews. Immunology*, 6(3), pp.173–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16498448 [Accessed May 27, 2014].

Nei, M., 2005. Selectionism and neutralism in molecular evolution. *Molecular biology and evolution*, 22(12), pp.2318–42. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1513187&tool=pmcentrez&rendertype=abstract [Accessed May 30, 2014].

Nei, M. & Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution*, 3(5), pp.418–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3444411 [Accessed January 30, 2014].

Neurath, M.F. et al., 1995. Antibodies to interleukin 12 abrogate established experimental colitis in mice. *The Journal of experimental medicine*, 182(5), pp.1281–90. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2192205&tool=pmcentrez&rendertype=abstract [Accessed July 8, 2014].

Neurath, M.F. et al., 1997. Predominant pathogenic role of tumor necrosis factor in experimental colitis in mice. *European journal of immunology*, 27(7), pp.1743–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9247586 [Accessed July 2, 2014].

Nielsen, R.G., Husby, S. & Kruse-Andersen, S., 2005. Premature closure of the upper esophageal sphincter as a cause of severe deglutition disorder in infancy. *Journal of pediatric surgery*, 40(4), pp.721–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15852289 [Accessed June 18, 2014].

Niess, J.H. et al., 2005. CX3CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. *Science (New York, N.Y.)*, 307(5707), pp.254–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15653504 [Accessed May 26, 2014].

O'Brien, J.B. et al., 2005. Passive immunisation of hamsters against Clostridium difficile infection using antibodies to surface layer proteins. *FEMS microbiology letters*, 246(2), pp.199–205. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15899406 [Accessed October 18, 2012].

O'Garra, A. & Vieira, P., 2007. T(H)1 cells control themselves by producing interleukin-10. *Nature reviews. Immunology*, 7(6), pp.425–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17525751 [Accessed June 12, 2014].

O'Leary, S., O'Sullivan, M.P. & Keane, J., 2011. IL-10 blocks phagosome maturation in mycobacterium tuberculosis-infected human macrophages. *American journal of respiratory cell and molecular biology*, 45(1), pp.172–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20889800 [Accessed June 30, 2014].

Ohta, T., 1973. Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428), pp.96–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4585855 [Accessed February 25, 2014].

Ohta, T. & Gillespie, J., 1996. Development of Neutral and Nearly Neutral Theories. *Theoretical population biology*, 49(2), pp.128–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8813019 [Accessed January 22, 2014].

Otte, J.-M., Cario, E. & Podolsky, D.K., 2004. Mechanisms of cross hyporesponsiveness to Toll-like receptor bacterial ligands in intestinal epithelial cells. *Gastroenterology*, 126(4), pp.1054–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15057745 [Accessed July 8, 2014].

Parameswaran, N. & Patial, S., 2010. Tumor necrosis factor-α signaling in macrophages. *Critical reviews in eukaryotic gene expression*, 20(2), pp.87–103. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3066460&tool=pmcentrez&rendertype=abstract [Accessed July 16, 2014].

Paredes-Sabja, D. et al., 2012. Clostridium difficile Spore-Macrophage Interactions: Spore Survival. D. Chakravortty, ed. *PloS one*, 7(8), p.e43635. Available at: http://dx.plos.org/10.1371/journal.pone.0043635 [Accessed October 18, 2012].

Paredes-Sabja, D. & Sarker, M.R., 2012. Adherence of Clostridium difficile spores to Caco-2 cells in culture. *Journal of medical microbiology*, 61(Pt 9), pp.1208–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22595914 [Accessed June 20, 2014].

Pelletier, M. et al., 2010. Evidence for a cross-talk between human neutrophils and Th17 cells. *Blood*, 115(2), pp.335–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19890092 [Accessed May 27, 2014].

Penny, D. et al., 1990. Trees from sequences: panacea or pandora's box? *Australian Systematic Botany*, 3(1), p.21. Available at: http://www.publish.csiro.au/view/journals/dsp_journal_fulltext.cfm?nid=150&f=SB9900021 [Accessed June 17, 2014].

Peterson, M.E. et al., 2009. Evolutionary constraints on structural similarity in orthologs and paralogs. *Protein science : a publication of the Protein Society*, 18(6), pp.1306–15. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2774440&tool=pmcentrez&rendertype=abstract [Accessed February 25, 2014].

Plevy, S.E. et al., 1997. A role for TNF-alpha and mucosal T helper-1 cytokines in the pathogenesis of Crohn's disease. *Journal of immunology (Baltimore, Md. : 1950)*, 159(12), pp.6276–82. Available at: http://www.jimmunol.org/content/159/12/6276.abstract [Accessed July 3, 2014].

Podolsky, D.K., 2002. Inflammatory bowel disease. *The New England journal of medicine*, 347(6), pp.417–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12167685 [Accessed June 18, 2014].

Powrie, F. et al., 1994. Inhibition of Th1 responses prevents inflammatory bowel disease in scid mice reconstituted with CD45RBhi CD4+ T cells. *Immunity*, 1(7), pp.553–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7600284 [Accessed June 16, 2014].

Price, A.B., Larson, H.E. & Crow, J., 1979. Morphology of experimental antibiotic-associated enterocolitis in the hamster: a model for human pseudomembranous colitis and antibiotic-associated diarrhoea. *Gut*, 20(6), pp.467–75. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1412461&tool=pmcentrez&rendertype=abstract [Accessed June 23, 2014].

Qualls, J.E. et al., 2006. Suppression of experimental colitis by intestinal mononuclear phagocytes. *Journal of leukocyte biology*, 80(4), pp.802–15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16888083 [Accessed July 7, 2014].

Raffatellu, M. et al., 2008. Simian immunodeficiency virus-induced mucosal interleukin-17 deficiency promotes Salmonella dissemination from the gut. *Nature medicine*, 14(4), pp.421–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2901863&tool=pmcentrez&rendertype=abstract [Accessed June 6, 2014].

Rakoff-Nahoum, S. et al., 2004. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell*, 118(2), pp.229–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15260992 [Accessed June 2, 2014].

Ranatunga, D.C. et al., 2012. A protective role for human IL-10-expressing CD4+ T cells in colitis. *Journal of immunology (Baltimore, Md. : 1950)*, 189(3), pp.1243–52. Available at: http://www.jimmunol.org/content/189/3/1243.full [Accessed June 9, 2014].

Redford, P.S., Murray, P.J. & O'Garra, A., 2011. The role of IL-10 in immune regulation during M. tuberculosis infection. *Mucosal immunology*, 4(3), pp.261–70. Available at: http://dx.doi.org/10.1038/mi.2011.7 [Accessed March 20, 2014].

Reinisch, W. et al., 2010. Fontolizumab in moderate to severe Crohn's disease: a phase 2, randomized, double-blind, placebo-controlled, multiple-dose study. *Inflammatory bowel diseases*, 16(2), pp.233–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19637334 [Accessed May 30, 2014].

Rescigno, M. et al., 2001. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nature immunology*, 2(4), pp.361–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11276208 [Accessed June 17, 2014].

Riggs, M.M. et al., 2007. Asymptomatic carriers are a potential source for transmission of epidemic and nonepidemic Clostridium difficile strains among long-term care facility residents. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 45(8), pp.992–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17879913 [Accessed May 27, 2014].

Rimoldi, M. et al., 2005. Monocyte-derived dendritic cells activated by bacteria or by bacteria-stimulated epithelial cells are functionally different. *Blood*, 106(8), pp.2818–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16030185 [Accessed June 25, 2014].

Rogler, G. et al., 1998. Isolation and phenotypic characterization of colonic macrophages. *Clinical and experimental immunology*, 112(2), pp.205–15. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1904962&tool=pmcentrez&rendertype=abstract [Accessed July 7, 2014].

Roussel, L. et al., 2010. IL-17 promotes p38 MAPK-dependent endothelial activation enhancing neutrophil recruitment to sites of inflammation. *Journal of immunology (Baltimore, Md. : 1950)*, 184(8), pp.4531–7. Available at: http://www.jimmunol.org/content/184/8/4531.full [Accessed May 29, 2014].

Rugtveit, J. et al., 1997. Cytokine profiles differ in newly recruited and resident subsets of mucosal macrophages from inflammatory bowel disease. *Gastroenterology*, 112(5), pp.1493–505. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9136827 [Accessed July 8, 2014].

Rupnik, M., Wilcox, M.H. & Gerding, D.N., 2009. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nature reviews. Microbiology*, 7(7), pp.526–36. Available at: http://www.nature.com/doifinder/10.1038/nrmicro2164 [Accessed October 4, 2012].

Ryan, A. et al., 2011. A role for TLR4 in Clostridium difficile infection and the recognition of surface layer proteins. *PLoS pathogens*, 7(6), p.e1002076. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3128122&tool=pmcentrez&rendertype=abstract [Accessed October 17, 2012].

Saitou, N. & Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4), pp.406–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3447015 [Accessed May 26, 2014].

Salazar-Gonzalez, R.M. et al., 2006. CCR6-mediated dendritic cell activation of pathogen-specific T cells in Peyer's patches. *Immunity*, 24(5), pp.623–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2855652&tool=pmcentrez&rendertype=abstract [Accessed July 8, 2014].

Sara, M., Sleytr, U.B.U. & Sára, M., 2000. S-Layer Proteins. *Journal of Bacteriology*, 182(4), pp.859–868. Available at: http://jb.asm.org/content/182/4/859.full [Accessed December 7, 2012].

Sarker, M.R. & Paredes-Sabja, D., 2012. Molecular basis of early stages of Clostridium difficile infection: germination and colonization. *Future microbiology*, 7(8), pp.933–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22913353 [Accessed June 20, 2014].

Savariau-Lacomme, M.M.-P. et al., 2003. Transcription and analysis of polymorphism in a cluster of genes encoding surface-associated proteins of Clostridium difficile. *Journal of …*, 185(15), pp.4461–4470. Available at: http://jb.asm.org/content/185/15/4461.long [Accessed February 12, 2014].

Saxton, K. et al., 2009. Effects of exposure of Clostridium difficile PCR ribotypes 027 and 001 to fluoroquinolones in a human gut model. *Antimicrobial agents and chemotherapy*, 53(2), pp.412–20. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2630646&tool=pmcentrez&rendertype=abstract [Accessed October 18, 2012].

Scaldaferri, F. & Fiocchi, C., 2007. Inflammatory bowel disease: progress and current concepts of etiopathogenesis. *Journal of digestive diseases*, 8(4), pp.171–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17970872 [Accessed May 27, 2014].

Schmid, K. & Yang, Z., 2008. The trouble with sliding windows and the selective pressure in BRCA1. *PloS one*, 3(11), p.e3746. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2581807&tool=pmcentrez&rendertype=abstract [Accessed February 25, 2014].

Schmidt, H.A. et al., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, 18(3), pp.502–504. Available at: http://bioinformatics.oxfordjournals.org/content/18/3/502.short [Accessed February 19, 2014].

Schwarz, T. et al., 2013. T cell-derived IL-10 determines leishmaniasis disease outcome and is suppressed by a dendritic cell based vaccine. I. Müller, ed. *PLoS pathogens*, 9(6), p.e1003476. Available at: http://dx.plos.org/10.1371/journal.ppat.1003476 [Accessed May 23, 2014].

Sebaihia, M. et al., 2006. The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature genetics*, 38(7), pp.779–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16804543 [Accessed October 4, 2012].

Sekirov, I. et al., 2008. Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infection and immunity*, 76(10), pp.4726–36. Available at: http://iai.asm.org/content/76/10/4726.full [Accessed May 28, 2014].

Sellge, G. et al., 2010. Th17 cells are the dominant T cell subtype primed by Shigella flexneri mediating protective immunity. *Journal of immunology (Baltimore, Md. : 1950)*, 184(4), pp.2076–85. Available at: http://www.jimmunol.org/content/184/4/2076.full [Accessed June 5, 2014].

Setlow, P., 2007. I will survive: DNA protection in bacterial spores. *Trends in microbiology*, 15(4), pp.172–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17336071 [Accessed June 20, 2014].

Shuying Li, Dennis K. Pearl, H.D., 2000. Phylogenetic Tree Construction Using Markov Chain Monte Carlo. *Journal of the American Statistical Association*, 95(450), pp.493–508. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.4461 [Accessed February 25, 2014].

Sleytr, U., 1997. I. Basic and applied S-layer research: an overview. *FEMS Microbiology Reviews*, 20(1-2), pp.5–12. Available at: http://www.sciencedirect.com/science/article/pii/S0168644597000399 [Accessed June 18, 2014].

Smith, N.G.C. & Eyre-Walker, A., 2002. Adaptive protein evolution in Drosophila. *Nature*, 415(6875), pp.1022–4. Available at: http://www.readcube.com/articles/10.1038/4151022a [Accessed February 25, 2014].

Smith, P.D. et al., 2011. Intestinal macrophages and response to microbial encroachment. *Mucosal immunology*, 4(1), pp.31–42. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3821935&tool=pmcentrez&rendertype=abstract [Accessed May 23, 2014].

Smith, P.D., Ochsenbauer-Jambor, C. & Smythies, L.E., 2005. Intestinal macrophages: unique effector cells of the innate immune system. *Immunological reviews*, 206, pp.149–59. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16048547 [Accessed July 7, 2014].

Smythies, L.E. et al., 2005. Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *The Journal of clinical investigation*, 115(1), pp.66–75. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=539188&tool=pmcentrez&rendertype=abstract [Accessed July 7, 2014].

Soehnlein, O. & Lindbom, L., 2010. Phagocyte partnership during the onset and resolution of inflammation. *Nature reviews. Immunology*, 10(6), pp.427–39. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20498669 [Accessed May 27, 2014].

Spigaglia, P. & Mastrantonio, P., 2002. Molecular analysis of the pathogenicity locus and polymorphism in the putative negative regulator of toxin production (TcdC) among Clostridium difficile clinical. *Journal of clinical microbiology*, 40(9), pp.3470–3475. Available at: http://jcm.asm.org/content/40/9/3470.short [Accessed February 13, 2014].

Stabler, R. a et al., 2009. Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome biology*, 10(9), p.R102. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2768977&tool=pmcentrez&rendertype=abstract [Accessed October 18, 2012].

Steele, J. et al., 2012. Antibody Against TcdB, but Not TcdA, Prevents Development of Gastrointestinal and Systemic Clostridium difficile Disease. *The Journal of infectious diseases*, pp.1–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23125448 [Accessed December 7, 2012].

Strimmer, K. & von Haeseler, A., 1997. Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences*, 94(13), pp.6815–6819. Available at: http://www.pnas.org/content/94/13/6815 [Accessed February 24, 2014].

Stumhofer, J.S. et al., 2007. Interleukins 27 and 6 induce STAT3-mediated T cell production of interleukin 10. *Nature immunology*, 8(12), pp.1363–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17994025 [Accessed June 2, 2014].

Su, C. et al., 2002. Efficacy of anti-tumor necrosis factor therapy in patients with ulcerative colitis. *The American journal of gastroenterology*, 97(10), pp.2577–84. Available at: http://dx.doi.org/10.1111/j.1572-0241.2002.06026.x [Accessed July 3, 2014].

Su, C. & Nei, M., 2001. Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences. *Molecular biology and evolution*, 18(4), pp.503–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11264401 [Accessed June 18, 2014].

Takeuchi, O. et al., 1999. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity*, 11(4), pp.443–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10549626 [Accessed June 7, 2014].

Tanaka, T. & Nei, M., 1989. Positive darwinian selection observed at the variable-region genes of immunoglobulins. *Molecular biology and evolution*, 6(5), pp.447–59. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2796726 [Accessed June 18, 2014].

Taylor, a E. et al., 2010. Defective macrophage phagocytosis of bacteria in COPD. *The European respiratory journal*, 35(5), pp.1039–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19897561 [Accessed July 31, 2013].

Thompson, S.A., 2002. Campylobacter surface-layers (S-layers) and immune evasion. *Annals of periodontology / the American Academy of Periodontology*, 7(1), pp.43–53. Available at: http://www.joponline.org/doi/abs/10.1902/annals.2002.7.1.43 [Accessed June 18, 2014].

Urwin, R. et al., 2002. Phylogenetic Evidence for Frequent Positive Selection and Recombination in the Meningococcal Surface Antigen PorB. *Molecular Biology and Evolution*, 19(10), pp.1686–1694. Available at: http://mbe.oxfordjournals.org/content/19/10/1686.short [Accessed July 16, 2014].

Valiente, E., Cairns, M.D. & Wren, B.W., 2014. The Clostridium difficile PCR ribotype 027 lineage: a pathogen on the move. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24621128 [Accessed March 18, 2014].

Vohra, P. & Poxton, I.R., 2012. Induction of cytokines in a macrophage cell line by proteins of Clostridium difficile. *FEMS immunology and medical microbiology*, 65(1), pp.96–104. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22409477 [Accessed December 7, 2012].

Waligora, A.J. et al., 2001. Characterization of a cell surface protein of Clostridium difficile with adhesive properties. *Infection and immunity*, 69(4), pp.2144–53. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=98141&tool=pmcentrez&rendertype=abstract [Accessed February 14, 2014].

Walk, S.T. et al., 2012. Clostridium difficile ribotype does not predict severe infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 55(12), pp.1661–8. Available at: http://cid.oxfordjournals.org/content/55/12/1661.long [Accessed February 24, 2014].

Walk, S.T. et al., 2013a. Understanding increased mortality in Clostridium difficile-infected older adults. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 57(4), pp.625–6. Available at: http://cid.oxfordjournals.org/content/57/4/625.full?ijkey=4398148d4e1a93f8b9a39e4d205f1403ec04377c&keytype2=tf_ipsecsha [Accessed May 14, 2014].

Walk, S.T. et al., 2013b. Understanding increased mortality in Clostridium difficile-infected older adults. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 57(4), pp.625–6. Available at: http://cid.oxfordjournals.org/content/57/4/625.extract [Accessed June 3, 2014].

Walker, A.S., Eyre, D.W., Crook, D.W., et al., 2013. Regarding "Clostridium difficile ribotype does not predict severe infection". *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 56(12), pp.1845–6. Available at: http://cid.oxfordjournals.org/content/56/12/1845.1.full [Accessed March 12, 2014].

Walker, A.S., Eyre, D.W., Wyllie, D.H., et al., 2013. Relationship between bacterial strain type, host biomarkers, and mortality in Clostridium difficile infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 56(11), pp.1589–600. Available at: http://cid.oxfordjournals.org/content/56/11/1589.full [Accessed March 12, 2014].

Walker, A.S., Eyre, D.W., Crook, D.W., et al., 2013a. Reply to Walk et al. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 57(4), pp.626–7. Available at: http://cid.oxfordjournals.org/content/57/4/626.full#ref-1 [Accessed May 14, 2014].

Walker, A.S., Eyre, D.W., Crook, D.W., et al., 2013b. Reply to Walk et al. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 57(4), pp.626–7. Available at: http://cid.oxfordjournals.org/content/57/4/626.full [Accessed March 12, 2014].

Warny, M. et al., 2005. Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe. *Lancet*, 366(9491), pp.1079–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16182895 [Accessed January 22, 2014].

Watanabe, N. et al., 2003. Elimination of local macrophages in intestine prevents chronic colitis in interleukin-10-deficient mice. *Digestive diseases and sciences*, 48(2), pp.408–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12643623 [Accessed July 16, 2014].

Waterston, R.H. et al., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), pp.520–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12466850 [Accessed May 27, 2014].

Weaver, C.T. et al., 2013. The Th17 pathway and inflammatory diseases of the intestines, lungs, and skin. *Annual review of pathology*, 8, pp.477–512. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3965671&tool=pmcentrez&rendertype=abstract [Accessed May 23, 2014].

Whelan, S. & Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution*, 18(5), pp.691–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11319253.

Whitlock, M.C. & Bürger, R., 2004. Fixation of New Mutations in Small Populations. *Evolutionary Conservation Biology*, pp.155–170.

Wilson, M.S. et al., 2011. IL-10 blocks the development of resistance to re-infection with Schistosoma mansoni. E. J. Pearce, ed. *PLoS pathogens*, 7(8), p.e1002171. Available at: http://dx.plos.org/10.1371/journal.ppat.1002171 [Accessed May 23, 2014].

Wishner, B.C. et al., 1975. Crystal structure of sickle-cell deoxyhemoglobin at 5 Å resolution. *Journal of Molecular Biology*, 98(1), pp.179–194. Available at: http://www.sciencedirect.com/science/article/pii/S0022283675801082 [Accessed June 18, 2014].

Van der Woude, M.W. & Bäumler, A.J., 2004. Phase and antigenic variation in bacteria. *Clinical microbiology reviews*, 17(3), pp.581–611, table of contents. Available at: http://cmr.asm.org/content/17/3/581.full [Accessed July 2, 2014].

Wright, S., 1938. Size of population and breeding structure in relation to evolution. *Science*, 87(2263), pp.430–431.

Xavier, M.N. et al., 2013. CD4+ T cell-derived IL-10 promotes Brucella abortus persistence via modulation of macrophage function. C. R. Roy, ed. *PLoS pathogens*, 9(6), p.e1003454. Available at: http://dx.plos.org/10.1371/journal.ppat.1003454 [Accessed May 23, 2014].

Xavier, R.J. & Podolsky, D.K., 2007. Unravelling the pathogenesis of inflammatory bowel disease. *Nature*, 448(7152), pp.427–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17653185 [Accessed October 30, 2012].

Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular biology and evolution*, 15(5), pp.568–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9580986.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24(8), pp.1586–91. Available at: http://mbe.oxfordjournals.org/content/24/8/1586.abstract [Accessed February 19, 2014].

Yang, Z. & Nielsen, R., 2002. Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Molecular Biology and Evolution*, 19(6), pp.908–917. Available at: http://mbe.oxfordjournals.org/content/19/6/908.long [Accessed February 20, 2014].

Yang, Z. & Nielsen, R., 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of molecular evolution*, 46(4), pp.409–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9541535 [Accessed February 25, 2014].

Yang, Z. & Swanson, W.J., 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular biology and evolution*, 19(1), pp.49–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11752189.

Yang, Z., Wong, W.S.W. & Nielsen, R., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular biology and evolution*, 22(4), pp.1107–18. Available at: http://mbe.oxfordjournals.org/content/22/4/1107 [Accessed February 19, 2014].

Yoshino, Y. et al., 2013. Clostridium difficile flagellin stimulates toll-like receptor 5, and toxin B promotes flagellin-induced chemokine production via TLR5. *Life sciences*, 92(3), pp.211–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23261530 [Accessed July 16, 2014].

Young, V.B. & Hanna, P.C., 2014. Overlapping roles for toxins in Clostridium difficile infection. *The Journal of infectious diseases*, 209(1), pp.9–11. Available at: http://jid.oxfordjournals.org/content/early/2013/09/26/infdis.jit461.full [Accessed May 14, 2014].

Zhang, Z., Clarke, T.B. & Weiser, J.N., 2009a. Cellular effectors mediating Th17-dependent clearance of pneumococcal colonization in mice. *The Journal of clinical investigation*, 119(7), pp.1899–909. Available at: http://www.jci.org/articles/view/36731/version/2 [Accessed June 11, 2014].

Zhang, Z., Clarke, T.B. & Weiser, J.N., 2009b. Cellular effectors mediating Th17-dependent clearance of pneumococcal colonization in mice. *The Journal of clinical investigation*, 119(7), pp.1899–909. Available at: http://www.jci.org/articles/view/36731 [Accessed June 11, 2014].

Zhu, J. et al., 2012. The transcription factor T-bet is induced by multiple pathways and prevents an endogenous Th2 cell program during Th1 cell responses. *Immunity*, 37(4), pp.660–73. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3717271&tool=pmcentrez&rendertype=abstract [Accessed May 26, 2014].

Zuckerkandl, E. & Pauling, L., 1965a. Evolutionary divergence and convergence in proteins. *Evolving Genes and Proteins*, pp.97 – 166.

Zuckerkandl, E. & Pauling, L., 1965b. Molecules as documents of evolutionary history. *Journal of theoretical biology*, 8(2), pp.357–66. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5876245 [Accessed February 24, 2014].

Zúñiga, L. a et al., 2013. Th17 cell development: from the cradle to the grave. *Immunological reviews*, 252(1), pp.78–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23405896.

Appendices

APPENDIX A – MEDIA AND BUFFERS

5x Loading Buffer

Glycerol	15mL
10% SDS	10mL
Tris pH6.8 (0.5M)	6.25mL
dH_2O	18.75mL
Bromoblue	tipful

250µL of DTT was added to 1mL of Loading Buffer to make final buffer. 3µL of this 5x loading buffer was added to 12µL of each sample.

10% Separating Gel

dH_2O	10mL
Bis-acrylamide	8.38mL
Tris pH 8.8 (1.5M)	6.25mL
10% SDS	250µL
10% APS	250µL
TEMED	30µL

Components required for four 10% separating gels.

10% Stacking Gel

dH_2O	3.5mL
Bis-acrylamide	1.062mL
Tris pH 6.8 (0.5M)	1.56mL
10% SDS	62.5µL
10% APS	62.5µL
TEMED	4µL

Components required for four 10% stacking gels.

10x Running Buffer

Tris base	15g
Glycine	72g
SDS	5g
dH_2O	500mL

Components required for 10x Running Buffer.

Coomassie Stain

0.2% Brilliant Blue	0.2g
45% Methanol	45mL
45% dH_2O	45mL
10% Acetic Acid	10mL

Components required for 100mL of Coomassie Blue Stain.

Destain Solution

25% Methanol	50mL
65% dH_2O	130mL
10% Acetic Acid	20mL

Components required for 200mL Destain solution.

10X PBS

NaCl	160g
Na_2HPO_4	23.2g
KH_2PO_4	4g
KCl	4g
dH_2O	2L

List of salts required for 2L of 10x PBS.

Wash Buffer

10X PBS	500mL
dH_2O	4,500mL
0.05% Tween	2.5mL

List of components required for 5L of 1x Wash Buffer.

TBS Reagent Diluent

Trizma	6.1g
NaCl	1.74g
BSA	0.2g
Tween	0.1mL
dH_2O	200mL

List of components required for 200mL of TBS Reagent Diluent.

cDNA Master Mix

Component	**Amount**
10X RT buffer	2µL
10X RT Random Primers	2µL
25X dNTP Mix (100µM)	0.8µL
MultiScribe™Reverse Transcriptase	1µL

APPENDIX B – SLP CHARACTERISATION

A Bicinchoninic acid (BCA) assay was performed on all samples to determine the protein concentration for each batch of SLPs. This was of great importance as a standard volume of each SLP was required for cell stimulation. The standard curve is given in Figure 7.1. Sample Concentrations calculated using equation of line are given in Table 4.2.3 below.

Figure 7.1

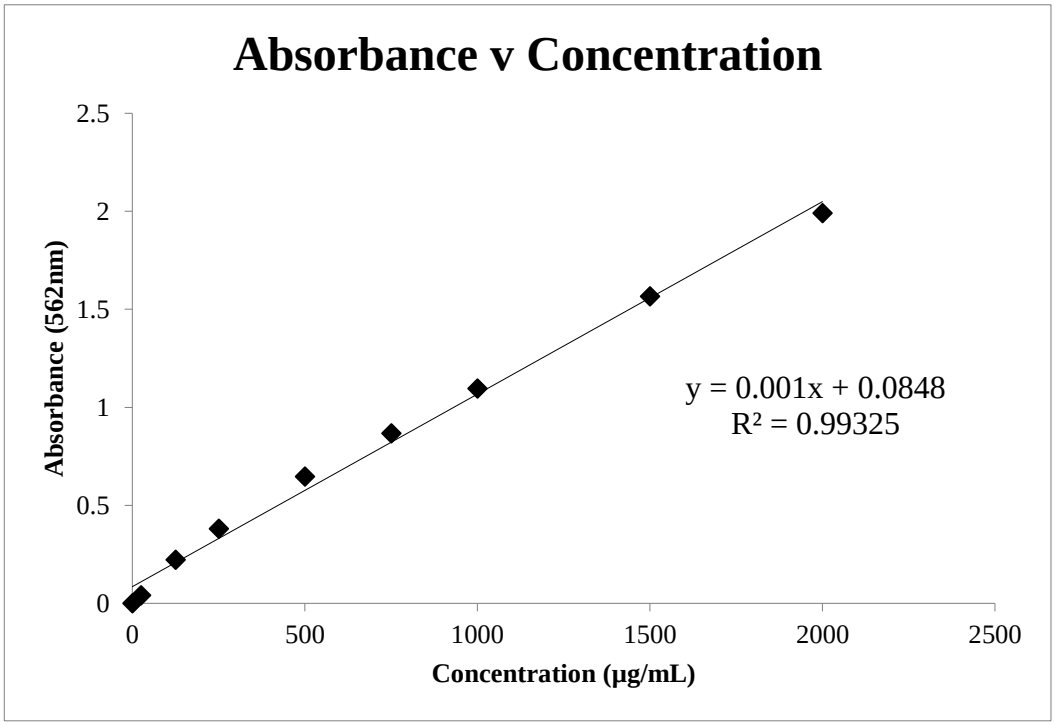


Figure 7.1

Plot of standard concentration vs. absorbance at 562nm. Equation of the line and R^2 value are also given.

Concentration of SLPs from different ribotypes of *C. difficile*.

Ribotype	Conc (mg/mL)
001	3.640
014	3.3315
078	7.243
027	8.365

LAL Endotoxin Assay

To ensure that any observed immune response was indeed due to the SLP and not due to potential contaminating endotoxins. The assay results are in Figure 7.2 given below. LPS is used as a positive control. All samples tested negative for endotoxin, showing no contamination.

Figure 7.2

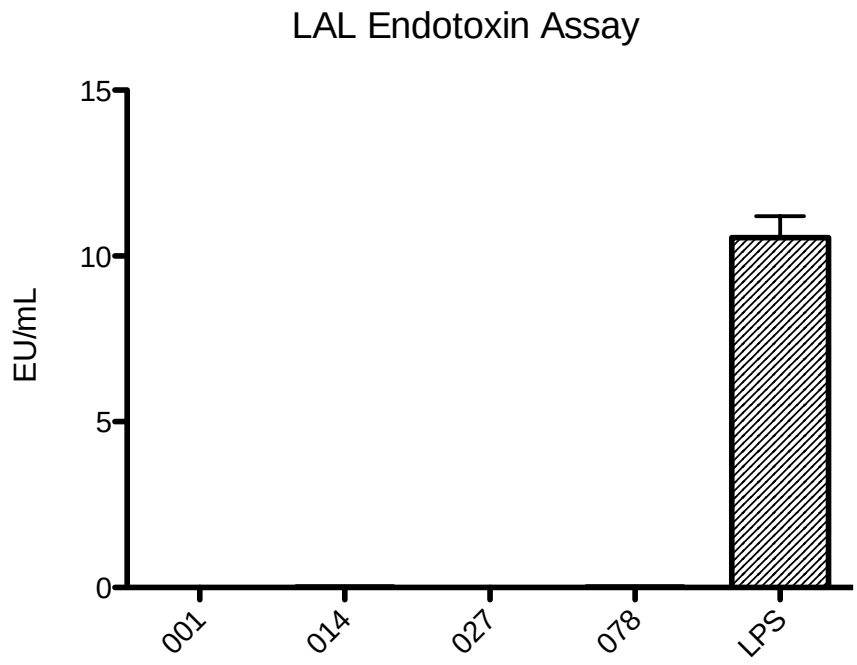
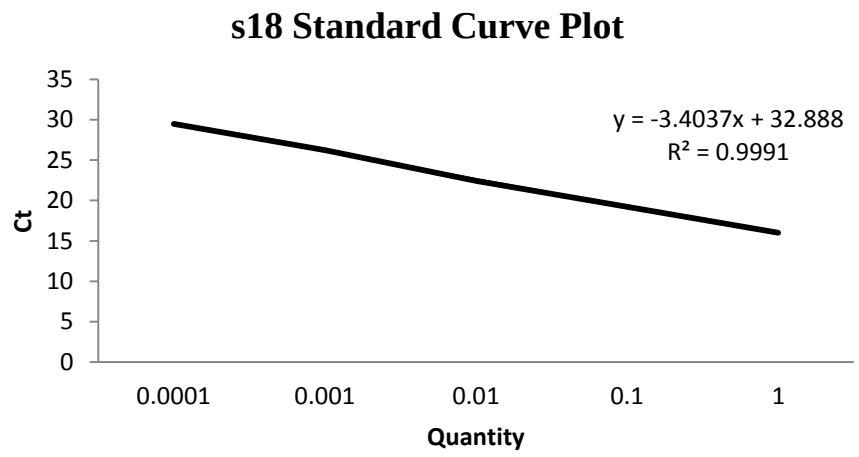


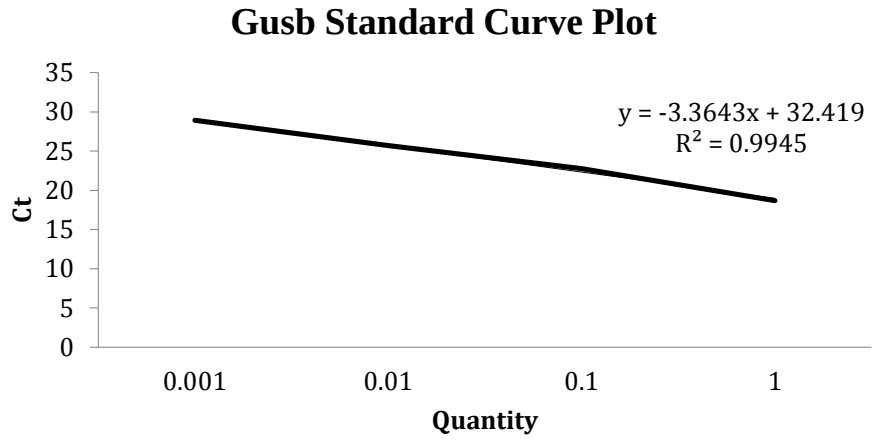
Figure 7.2

LAL assay results showing Endotoxin activity in samples. LPS (10ng/mL) was used as a positive control. SLPs were tested at a concentration of 20µg/mL. Results show that no Endotoxin is present in the SLP samples, and observed responses are due to activation with SLP.

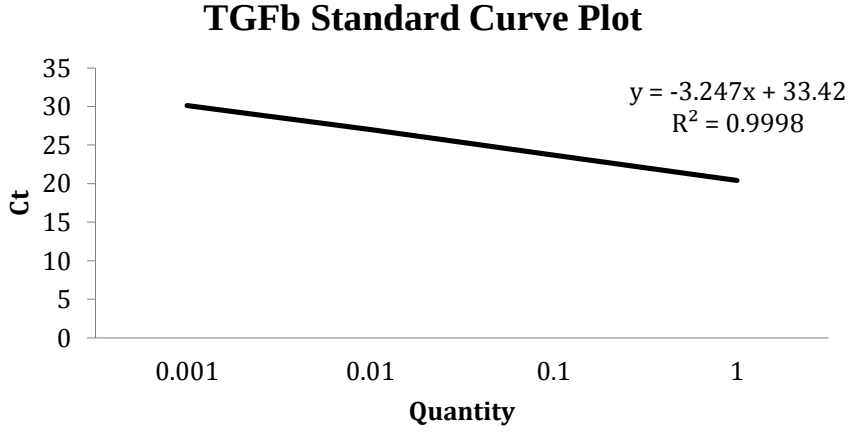
APPENDIX C – PRIMER EFFICIENCY CURVES



Efficiency = 97% Dynamic Range = 16-29 C_q

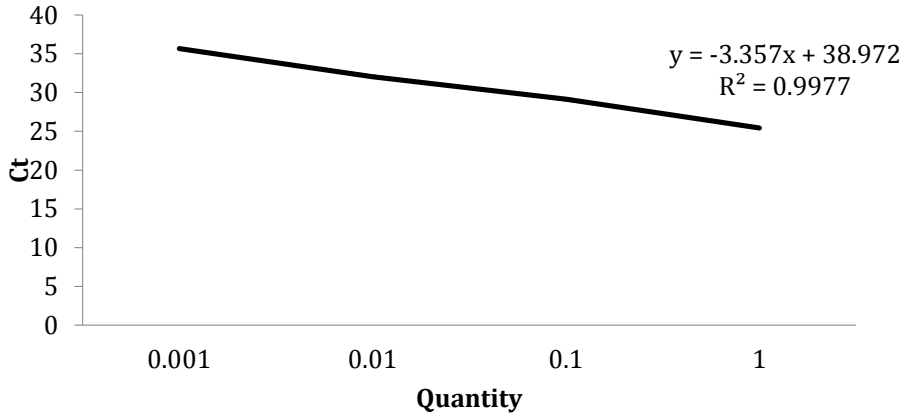


Efficiency = 98% Dynamic Range = 18-28 C_q



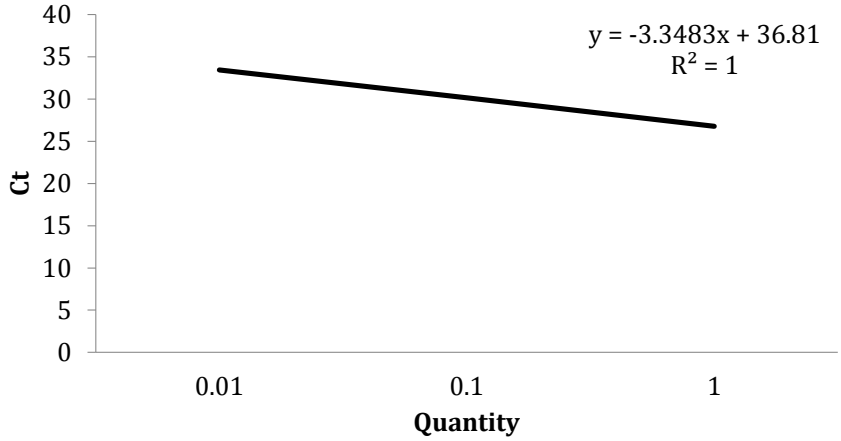
Efficiency = 100% Dynamic Range = 21-32 C_q

IFNg Standard Curve Plot



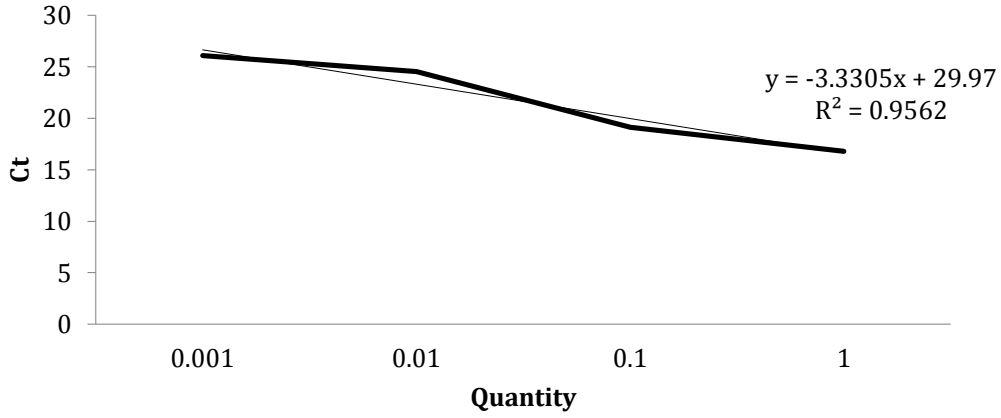
Efficiency = 99% **Dynamic Range = 24-34 C_q**

IL-23 Standard Curve Plot



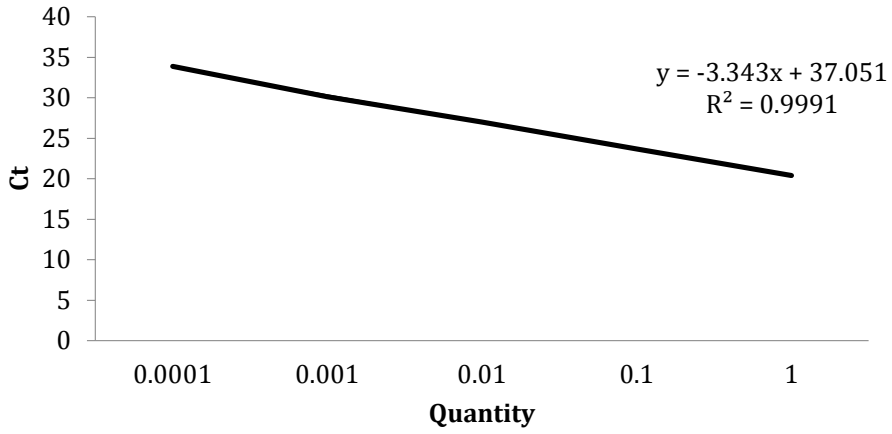
Efficiency = 99% **Dynamic Range = 26-35 C_q**

IL-6 Standard Curve Plot



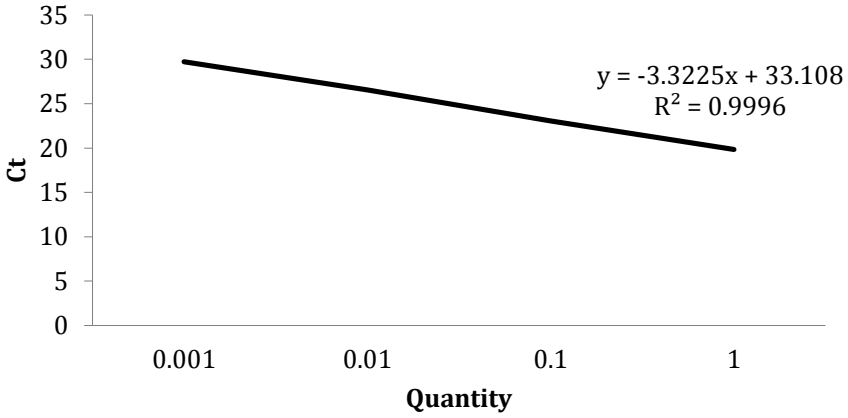
Efficiency = 100% **Dynamic Range = 17-26 C_q**

IL-10 Standard Curve Plot



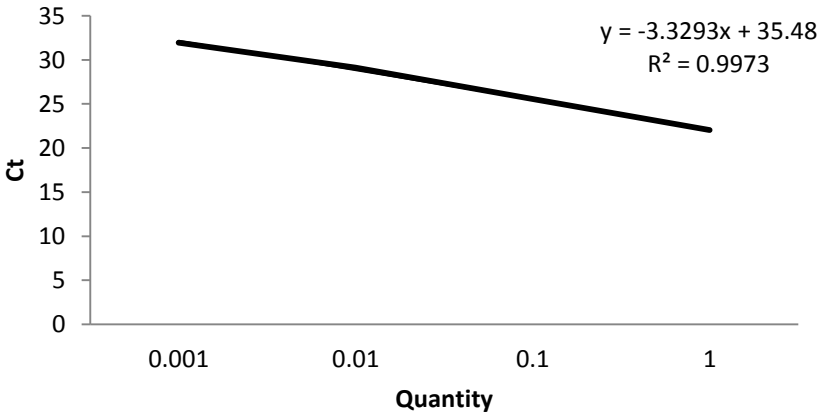
Efficiency = 99% **Dynamic Range = 20-33 C_q**

IL-12b Standard Curve Plot



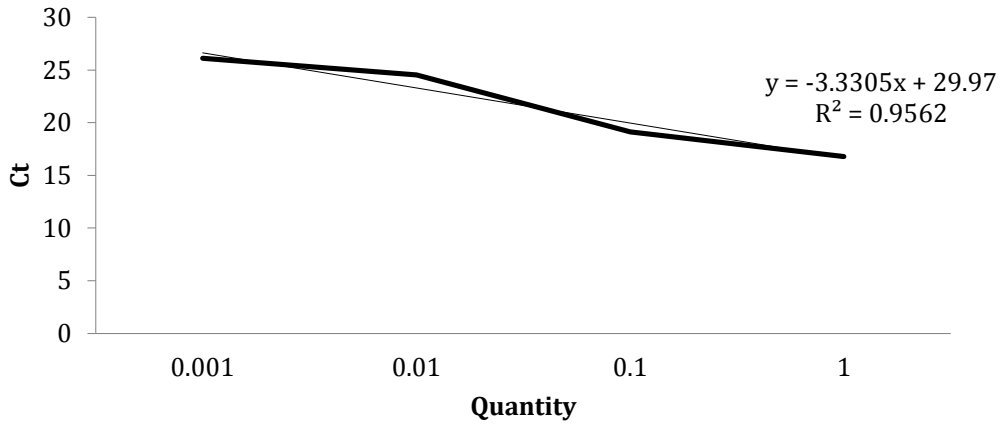
Efficiency = 100% **Dynamic Range = 19-30 C_q**

TNFa Standard Curve Plot



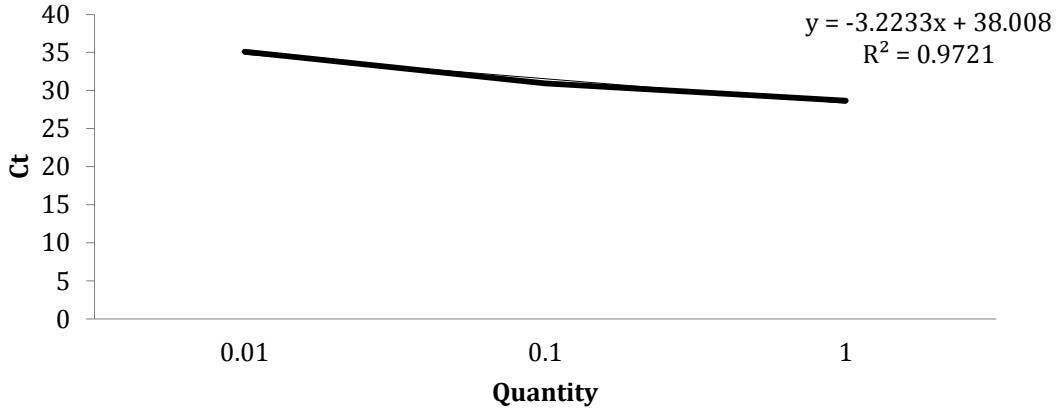
Efficiency = 100% **Dynamic Range = 20-34 C_q**

IL-17a Standard Curve Plot



Efficiency = 97% **Dynamic Range = 26-36 C_q**

ROR Standard Curve Plot



Efficiency = 104% **Dynamic Range = 28-35 C_q**

APPENDIX D – PRIMER SEQUENCES

Gusb
Primer 1 5’- CCA ATG AGC CTT CCT CTG C-3’
Primer 2 5’- ATC ATA TTT GGC GTT GCT CAC-3’

Rps18
Primer 1 5’- CCT GAG AAG TTC CAG CAC AT-3’
Primer 2 5’- ACA CCA CAT GAG CAT ATC TCC-3’

IL-6
Primer 1 5’- AGC CAG AGT CCT TCA GAG A-3’
Primer 2 5’- TCC TTA GCC ACT CCT TCT GT-3’

IL-10
Primer 1 5’- GGC ATC ACT TCT ACC AGG TAA-3’
Primer 2 5’- TCA GCC AGG TGA AGA CTT TC-3’

IL-12b (p40)
Primer 1 5’- TGT CCT CAG AAG CTA ACC ATC-3’
Primer 2 5’- TCC AGT CCA CCT CTA CAA CA-3’

IL-17a
Primer 1 5’- AGA CTA CCT CAA CCG TTC CA-3’
Primer 2 5’- GAG CTT CCC AGA TCA CAG AG-3’

IL-23
Primer 1 5’- ACC AGC GGG ACA TAT GAA TC-3’
Primer 2 5’- GAT CCT TTG CAA GCA GAA CTG-3’

IFNγ
Primer 1 5’- ATG AAC GCT ACA CAC TGC ATC-3’
Primer 2 5’- CCA TCC TTT TGC CAG TTC CTC -3’

TGFb1
Primer 1 5’- GCG GAC TAC TAT GCT AAA GAG G-3’
Primer 2 5’- CCG AAT GTC TGA CGT ATT GAA GA-3’

TNFa
Primer 1 5’- AGA CCC TCA CAC TCA GAT CA-3’
Primer 2 5’- TCT TTG AGA TCC ATG CCG TTG-3’

RORg
Primer 1 5’- TCT GCA AGA CTC ATC GAC AAG-3’
Primer 2 5’- GAG GTG CTG GAA GAT CTG C-3’