Lactobacillus fermentum 3872 genome sequencing and analysis

Burhan Lehri

The thesis is being submitted in partial fulfilment of the requirements of the University for the award of Doctor of Philosophy (PhD)

October 2017

Declaration

I declare that the work reported in this thesis is entirely my own and has been carried out at Kingston University, UK.

This thesis has not been submitted, in whole or in part, for any other degree at this or any other University.

Burhan Lehri

Abstract

In recent years, there has been a rise in antimicrobial-resistant bacteria caused by overdependence on, and misuse of, antibiotics. This has led to an increase in research for identifying alternatives to combat pathogens. One promising means of combating pathogenic bacteria, particularly for those residing in the gastrointestinal tract (GIT), is the use of probiotics. This thesis focuses on a potential probiotic strain *Lactobacillus fermentum* 3872, the genome sequence of which was circularised during the study, identifying genes that may contribute to probiotic activity. Several genes involved in GIT survival, such as acid symporters were discovered, along with genes that encode adhesion proteins such as those involved in mucus, fibronectin and collagen binding. The genes mentioned above may contribute to *L. fermentum* 3872 survivability within the GIT and have an antagonistic effect on enteric pathogens via competitive exclusion. Other interesting genes identified in *L. fermentum* 3872 were potentially involved in bacterial aggregation, exopolysaccharide and vitamin synthesis, along with four prophage encoding regions. Genes that encode a class III bacteriocin was also identified. An additional gene encoding a collagen binding protein (CBP) of a newly discovered plasmid pLF3872, was recognised. The chromosomal sequence also had a partial CBP encoding gene. pLF3872 has a toxin-antitoxin gene pair that ensures stable maintenance of the plasmid, along with conjugation-related genes. Functional analysis of the recombinant CBP via ELISA experiments found that the protein had the ability to bind to collagen I, a protein present on the epithelial lining of cells of the GIT. ELISA experiments also demonstrated that a common gastrointestinal pathogen, *Campylobacter jejuni,* can bind to collagen I in a concentration-dependent manner. In addition, mass spectrometry analysis identified that *C. jejuni* strains 11168H and 81-176 may utilise flagellar components (FlaA and FlaB) for adhesion. Furthermore, *C. jejuni* 11168H and 81-176 binding to collagen I was inhibited in the presence of either *L. fermentum* 3872 or CBP,

thus reducing *C. jejuni* adherence via competitive exclusion. Using an *in vitro* assay, it was also demonstrated that *L. fermentum* 3872 cell-free supernatant could inhibit the growth of *C. jejuni*, due to the acidic environment brought about by *L. fermentum* 3872.

During the completion of the genome sequence of *L. fermentum* 3872, comparison of various sequence assembly techniques which focused on the quality of the genome assembly was conducted. The results showed that further extension of the genome sequence during sequence assembly may lead to assembly errors when over-relying on a commonly-used sequence quality indicator, referred to as read mapping. It is suggested that care must also be taken when using long read technology to complete the genome sequence of a bacteria, as this may result in nucleotide sequence redundancies.

Acknowledgements

I would predominantly like to thank my director of studies Professor Andrey Karlyshev, for providing the guidance and help required throughout the course, and the acceptance of my studentship. I would also like to thank my second supervisor Dr Alan Seddon for providing additional assistance and help when required. Moreover, I would like to thank my parents for their support throughout the course of the PhD, along with my colleagues Ana Viera and Amritha Ramesh for their support and assistance, and Dr Ali Ryan for the use of laboratory equipment's present within his lab.

Contents

List of Figures

List of Tables

List of Abbreviations

ACT	Artemis Comparison Tool
BASys	Bacterial annotation system
BLASTN	Basic local alignment search tool nucleotide
bp	Base pair
BSA	bovine serum albumin
CAM	Chloramphenicol
CD	Crohn's Disease
CDD	Conserved domain database
CFU	Colony forming units
CISA	Contig integrator for sequence assembly
CLC GWB	CLC genomics workbench
CLRs	C-type lectin receptors
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
ELISA	Enzyme linked immune absorbance assay
ESPGHAN	European Society for Paediatric Gastroenterology Hepatology and Nutrition
FD	Functional Dyspepsia
GBS	Guillain-Barré syndrome
GIT	Gastrointestinal tract
HCl	Hydrochloric acid
IBD	Inflammatory Bowel Disease
IBS	Irritable Bowel Syndrome

ICORN	Iterative correction of reference nucleotides
IMAGE	Iterative mapping and assembly for gap elimination
ISP	Ion sphere particle
kb	Kilobase
kDa	Kilodaltons
LASTZ	Large scale genome alignment tool
LOS	Lipooligosaccharide
MALT lymphoma	Mucosal Lymphoid Tissue Lymphoma
Mb	Mega base
MIRA	Mimicking Intelligent Read Assembly
mRNA	Messenger ribonucleic acid
MSCRAMMS	Microbial Surface Adhesive Matrix Molecules
NCBI	National Center for Biotechnology Information
NLRs	Nucleotide-binding oligomerization domain-like receptors
OD	Optical density
PacBio	Pacific Biosciences
PAMPs	Pathogen-associated molecular patterns
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline with tween 20
PCR	Polymerase chain reaction
PROKKA	Rapid prokaryotic genome annotation
PRRs	Pattern recognition receptors
PVDF	Polyvinylidene difluoride
RAST	Rapid annotation server
REAPR	Recognition of errors in assemblies using paired end reads

RNA	Ribonucleic acid
S-D	Shine-Dalgarno
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	Standard error
SOC	Super optimal broth with catabolite repression
SPAdes	St. Petersburg genome assembler
TBE	Tris/Borate/EDTA
TBST	Tris-buffered saline Tween
TLRs	Toll-like receptors
TMB	3,3′,5,5′-Tetramethylbenzidine substrate
UC	Ulcerative Colitis
UV	Ultraviolet
WAO	World Allergy Organisation

Chapter 1: Introduction

1.1 The early concept and definition of health-promoting bacteria

Even before the discovery of microorganisms, fermented food products have been used to promote health in humans. For example, both the Greek physician Hippocrates and the Roman historian Plinius advocated fermented milk as a medicinal aid (Vijaya *et al.*, 2013; Ho and Prasad, 2013). The actual concept of consuming live microorganisms to promote health and longevity was proposed by immunologist Elie Metchnikoff, who deduced that not all bacteria are harmful (Butel, 2014). Metchnikoff suggested that the consumption of fermented food products with their active ingredient of live bacteria may help displace pathogenic gastrointestinal organisms (Butel, 2014; Sánchez *et al.*, 2017; Mackowiak, 2013). The term 'probiotic', meaning 'for life', was first used in 1965 by Lilly and Stillwell (Lilly and Stillwell, 1965). They defined the term as 'growth promoting factors produced by microorganisms'. This definition was adapted by the World Health Organisation (WHO) and the Food and Agriculture Organisation (FAO) in 2001 to incline more towards Metchnikoff's concept of probiotics to 'Live microorganisms which when administered in adequate amounts, confer a health benefit to the host' (World Health Organization and, Food and Agriculture Organization of the United Nations, 2001). The definition took into consideration that the benefits brought about by probiotics are not restricted to a specific delivery system, or limited to a specific organ/body part of the host (Sánchez *et al.*, 2017). Nonetheless, some studies have shown that bacteria do not necessarily have to be viable for them to provide a beneficial effect to its host (Sanders, 2008; Lahtinen, 2012).

1.2 Categorising a microbe as a probiotic

For a microbe to be classified as a probiotic, it should be taxonomically identified using molecular techniques, must be deposited in an internationally recognised culture collection bank, and must not be able to transmit certain antibiotic-resistance genes

(Moal and Servin, 2014). Furthermore, according to WHO and FAO, the microbe must not provide a harmful effect to the host when administered, and its beneficial effect on the host must be proven (Butel, 2014). The microbe should be able to survive and reach the site of its action when administered (i.e. ingested microbes must survive the harsh conditions of the GIT) (Butel, 2014). The microbe must also remain stable during the manufacturing and packaging process and maintain its probiotic characteristic (Butel, 2014).

1.3 Probiotics and their mechanisms of action

A probiotic can benefit the host either by directly interacting with an infectious agent and/or disease (Butel, 2014). A probiotic can also interact with the host to enhance its defence against infectious agents, improving the host's barrier function, and/or benefit a host by modulating the host's immune system (Butel, 2014). The molecular means by which a probiotic provides its beneficial effect(s), known as its mechanism(s) of action, is diverse and has not been determined completely (Bermudez-Brito *et al.*, 2012). Mechanisms of action of probiotics vary and are dependent on the genus, species, and strains of microbes (Islam, 2015). To use probiotics more effectively, it is important to know the mechanism of action, as in the past the European Food Safety Authority (EFSA) had refused the health claims of marketed probiotics mainly due to a lack of information about their mechanism of action (Butel, 2014). As previously noted, many studies focus on the effect a probiotic has in treating a condition rather than on how it provides its beneficial effect (Butel, 2014). Understanding a probiotics mechanism of action would enable more efficient utilisation of probiotics as tools for improving health.

1.3.1 Synthesis of antimicrobials and competitive exclusion

As discussed above, one of the ways probiotics benefit the host is by interacting with an infectious agent and/or competing with a pathogen for nutrients and space (Butel,

2014). A probiotic could release antimicrobials such as bacteriocins, metabolites or biosurfactants (Butel, 2014). A probiotic could also compete for, and inhibit, adhesion sites utilised by microbes for host colonisation (Butel, 2014). Probiotics may also neutralise microbial toxins and play a role in the detoxification of host food products, preventing damage to the host (Oelschlaeger, 2010).

The most common type of antimicrobial substances synthesised by probiotics is short chain fatty acids, such as lactic acid (Oelschlaeger, 2010). Fatty acids can lower the local pH environment of the host, harming pathogens that are not adapted to a low pH environment. Bacterial toxins can also be neutralised by short chain fatty acids, such as the inhibition of Shiga toxin produced by enterohaemorrhagic *Escherichia coli* (EHEC) O157: H7 via the production of butyric and lactic acid by probiotic *Clostridium butyricum* MIYAIRI (Takahashi *et al.*, 2004). Probiotics such as *Lactobacillus rhamnosus* GG can also bind to mycotoxin deoxynivalenol and reduce its effect on the host (Turner *et al.*, 2008). Another example of an antimicrobial substance produced by probiotics is hydrogen peroxide, which can cause oxidative stress on other microbes (Berstad *et al.*, 2016). Hydrogen peroxide is particularly effective against anaerobes, which are more sensitive to the effects of reactive oxygen species (Berstad *et al.*, 2016). The production of hydrogen peroxide by various *Lactobacillus* strains has been implicated in anti-*Candida* infection activity (Verdenelli *et al.*, 2014).

The production of bacteriocins, which are a diverse group of antimicrobial peptides, has also been of interest in combating pathogenic bacteria (Dobson *et al.*, 2012). A bacteriocin can act against closely related species of bacteria that synthesise it (narrow-spectrum) or can act against a wide range of bacteria (broad-spectrum) (Yang *et al.*, 2014). These peptides do not harm the host bacteria due to the presence of immunity proteins (Yang *et al.*, 2014). Bacteriocins can be low molecular weight (< 10 kDa) and

heat stable, which are grouped as Class I lantibiotic peptides that either disrupt the membrane structure or have an enzymatic action against microbes (Yang *et al.*, 2014; Sonomoto and Yokota, 2011). Bacteriocins can also be Class II non-lantibiotic peptides, which may cause microbial death via depolarization of the cell wall, due to their amphiphilic structure (Yang *et al.*, 2014; Cotter, Hill and Ross, 2005). Bacteriocins can be high molecular weight (> 30kDa) and heat labile Class III bacteriocins, which can break the peptide bonds of a peptidoglycan layer (Yang *et al.*, 2014; Sonomoto and Yokota, 2011). The latter classes of bacteriocins are further subdivided into various groups (Yang *et al.*, 2014). An example of the potential use of bacteriocin can be seen in a broad-spectrum class II bacteriocin Abp118 produced by the probiotic *Lactobacillus salivarius* UCC118 that protected mice from *Listeria monocytogenes* infection. However, the mutant *L. salivarius* UCC118 strain had no protective effect against *L. monocytogenes* (Corr *et al.*, 2007).

As mentioned previously, competitive exclusion of pathogenic bacteria may also contribute to probiotic function. Probiotics may compete for limited resources against pathogens to make it difficult for pathogens to colonise a host (Oelschlaeger, 2010). For example, *Lactobacilli* sp such as *L. acidophilus* bind to ferric hydroxide, thus preventing pathogens from obtaining iron (Oelschlaeger, 2010). Another important competition tool utilised by probiotic bacteria is competition for adhesion ligands. Probiotics can protect the host from infectious diseases by competing for the same adhesion ligand as that of a pathogen, thus reducing the pathogen's potential to colonise a host (Oelschlaeger, 2010). Bacterial surface proteins, such as mucin binding proteins, are the most commonly studied proteins for competition experiments for attachment to adhesion ligands (Oelschlaeger, 2010). Probiotics may also form biofilms protecting the host against pathogen colonisation (Oelschlaeger, 2010). Adhesion is also an important precursor of other beneficial effects on the host as it allows the probiotic bacteria to

attach and colonise a host. Probiotics may also produce biosurfactants that could act as antimicrobials or help in probiotic microbial adhesion and biofilm formation (Oelschlaeger, 2010; Singh and Cameotra, 2004; Sharma and Saharan, 2016).

1.3.2 Improvement of the host barrier function

Improvement of the hosts barrier function includes enhancement/stabilisation of the host epithelial cell tight junctions, such as those of the gut. The latter was observed in human cell lines (T_{84}) when the probiotic *E. coli* strain Nissle 1917 maintained the epithelial barrier and protected against the enteropathogenic *E. coli* strain E23486/69 disruption of the gut's barrier (Zyrek *et al.*, 2007). The study concluded that the probiotic *E. coli* strain Nissle 1917 promoted the production of ZO-2 protein, which helped stabilise and maintain the epithelial tight junction structure of the gut (Zyrek *et al.*, 2007). Improvement of the host barrier function can also involve promotion of the host mucus expression (Butel, 2014). For example, studies have shown that multiple *Lactobacillus* strains could induce MUC3 production after attachment to epithelial cells, reducing the potential of the pathogen, *E. coli* E2348/69, adhering to epithelial cells (Mack *et al.*, 2003). Probiotics may also deconjugate bile salts produced in the gut to make them more potent antimicrobials, thus protecting the host from infectious pathogens (Oelschlaeger, 2010).

1.3.3 Modulation of the host immune system

A probiotic may also modulate the host immune system. Immune modulation could be used for the prevention and treatment of infectious and inflammatory diseases, as well as for the removal of neoplastic cells (Oelschlaeger, 2010). Immune modulation can involve both the adaptive and innate immune response and could be brought about by the interaction of bacterial products with epithelial and dendritic cells, along with macrophages or lymphocytes (Bermudez-Brito *et al.*, 2012). The most common mammalian host cells that interact with probiotics to provide an immune response are

intestinal epithelial cells and, to a lesser extent, dendritic cells (Bermudez-Brito *et al.*, 2012). An immune response could be mediated by the detection of pathogen-associated molecular patterns (PAMPs), which are conserved products such as bacterial metabolites, cell wall components, and/or DNA that are recognised by pattern recognition receptors (PRRs) of the host cells (Bermudez-Brito *et al.*, 2012; Oelschlaeger, 2010). PRRs could be Toll-like receptors (TLRs), extracellular C-type lectin receptors (CLRs), and Nucleotide-binding oligomerization domain-like receptors or NOD-like receptors (NLRs) (Bermudez-Brito *et al.*, 2012). An example of utilising probiotics for the treatment of infectious disease is the administration of probiotic bacterium *Lactobacillus casei* CRL 431. This probiotic was found to induce TLR2, TLR4 and TLR9, regulating the production of cytokines TNFα, IFNγ and IL-10, and reducing the severity of *Salmonella enterica* serovar Typhimurium infection in murine models (Castillo, Perdigon and de Moreno, 2011). Studies have also shown that *Lactobacillus casei* Shirota enhances natural killer cell activity through the action of interleukin 1, which in turn can be used against neoplastic cells (Takeda *et al.*, 2006). In addition, probiotic *E. coli* Nissle 1917 and *Lactobacilli* strains can induce the production of antimicrobials such as human beta-defensin-2 in Caco-2 intestinal epithelial cells. Defensins tend to have a broad spectrum of antimicrobial activity and can be used against pathogenic bacteria and fungi (Wehkamp *et al.*, 2004).

1.4 Potential uses of probiotics

There has been increasing interest in probiotic research, both for clinical and industrial purposes (Evivie et al., 2017; Mackowiak, 2013). The rise in interest in probiotic research for clinical purposes can be attributed to the realisation of the importance of commensal organisms in human health, disease, and immunology (Sánchez *et al.*, 2017). This realisation was made more evident by the human microbiome (Hsiao and Fraser-Liggett, 2009), Metagenomics of the human intestinal tract (MetaHit)

(http://www.metahit.eu/index.php?id=410, 2017), and MyNewGut (http://www.mynewgut.eu/, 2017) projects, which helped characterise microorganisms associated with healthy and diseased humans (Sánchez *et al.*, 2017). In the past, interest in clinical probiotic research had decreased due to the advent of antibiotics (Butel, 2014), but with the current rise in antimicrobial-resistant bacteria, probiotics are being studied as potential pathogen antagonists.

1.4.1 Infectious diarrhoea

More than 20 microbes can cause infectious diarrhoea, the most common of which are rotavirus, *Salmonella*, *Campylobacter*, *Shigella* and enterotoxigenic *E. coli* (Sánchez *et al.*, 2017; Allen *et al.*, 2010). Many probiotics are ideal candidates for surviving the harsh environment of the GIT and thus are being considered for the treatment of infectious diarrhoea. As already noted, some of the ways in which probiotics can tackle pathogens of the gut are by competing for limited nutrients, and binding sites (Sánchez *et al.*, 2017). Probiotics can also produce antimicrobials such as lactic acid and enhance the antimicrobial agents of the GIT itself (i.e. by deconjugating bile salts) (Sánchez *et al.*, 2017). Probiotics can also generate a host immune response to tackle pathogens (Sánchez *et al.*, 2017). A meta-analysis of 63 studies, conducted by Allen *et al*. (2010), concluded that the use of probiotics for the treatment of acute infectious diarrhoea can help shorten the duration of diarrhoea and fever, particularly in children, although the effectiveness of probiotic treatment varied greatly between each study (Allen *et al.*, 2010). This conclusion was mirrored by Salari *et al*. (2012), who added that the efficacy of the use of probiotics in the treatment of infectious diarrhoea is still unknown (Salari, Nikfar and Abdollahi, 2012). The variability in treatment effectiveness may have been caused by the fact that both of the studies mentioned above did not consider the strain/species of the microbes being tested and treated against. The European Society for Paediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) also stated that

the use of probiotics *Lactobacillus rhamnosus* GG, *Saccharomyces boulardii*, *Lactobacillus reuteri* DSM 17938, and heat-killed *Lactobacillus acidophilus* LB can be used in children with acute gastroenteritis as adjuncts to rehydration therapy (Szajewska *et al.*, 2014).

A promising use of probiotics is in the treatment of antibiotic-associated diarrhoea, as antibiotic treatment tends to cause dysbiosis of the GIT microbiota. Dysbiosis can result in an increase of *Clostridium difficile* and thus *C. difficile* associated diarrhoea (Sánchez *et al.*, 2017). Studies have shown that probiotics such as *Lactobacillus rhamnosus* GG and *Saccharomyces boulardii* can protect against *C. difficile* related diarrhoea (Sánchez *et al.*, 2017). The use of faecal transplant, where a healthy individual's microbiome is transferred to a person being treated with antibiotics, is also an effective tool in the prevention of *C. difficle* related diarrhoea (Hevia *et al.*, 2015). Faecal transplantation may contain probiotic bacteria.

1.4.2 Inflammatory bowel disease

Crohn's disease (CD), pouchitis and ulcerative colitis (UC) are conditions grouped as inflammatory bowel disease (IBD), which is represented by chronic inflammation of the GIT (Mattos *et al.*, 2015). It is hypothesised that the conditions are caused by a deregulation of the immune system resulting in a host inflammatory response and that dysbiosis of the GIT microbiota plays a major role in this (Sánchez *et al.*, 2017). Currently, many of the treatments for IBD modulate the immune response and inflammatory cytokines (Mattos *et al.*, 2015). A probiotic’s ability to modulate the host immune system to reduce inflammation and normalise a GIT that has undergone dysbiosis are hopeful signs for the treatment of IBD (Sánchez *et al.*, 2017). The use of probiotics has lead to promising results in the treatment of UC, and more so for pouchitis (Sánchez *et al.*, 2017; Lichtenstein, Avni-Biron and Ben-Bassat, 2016). Probiotic #VSL3, which consists of a combination of probiotic bacteria, is known to be

particularly useful in the treatment of IBD (Penner and Fedorak, 2005). However, the effectiveness of probiotics in the treatment of CD has not been promising to date (Lichtenstein, Avni-Biron and Ben-Bassat, 2016).

1.4.3 Irritable bowel syndrome

Irritable bowel syndrome (IBS) is a chronic gastrointestinal disorder that causes abdominal discomfort along with bowel dysfunction, with current treatments having low efficacy (Hoveyda *et al.*, 2009). The aetiology of the condition is unknown, although it is thought to be multifactorial, with indications that dysbiosis of the GIT microbiota, inflammation, stress, infectious diseases, and/or abnormal gut movement may all play roles in its manifestation (Butel, 2014). Probiotics are thought to be a promising tool in the treatment of IBS, as they can strengthen the gut barrier, normalise the gut microbiota, and improve GIT dysmotility (Cozma-Petrut *et al.*, 2017). A meta-analysis of 14 randomised control trials found that probiotics may be useful in treating IBS (Hoveyda *et al.*, 2009). Studies conducted by Zhang *et al.* (2016) that looked at 21 randomised control trials, and Ford *et al.* (2014), which examined 43 randomised controls trials, also found probiotics to be effective in the treatment of IBS. However, it must be noted that it is hard to compare different randomised control trials as study parameters vary (for example, the duration of study/treatment, age, dosage differences) (Hoveyda *et al.*, 2009; Cozma-Petrut *et al.*, 2017). Comparison is further complicated by the fact that different strains, species, and genera of probiotics have different effects on IBS, as some probiotics may help against specific symptoms while others would have an effect on a range of symptoms (Hoveyda *et al.*, 2009; Cozma-Petrut *et al.*, 2017).

1.4.4 *Helicobacter pylori* infections

Probiotics are being investigated for the treatment of *Helicobacter pylori* infections, which can cause peptic ulcers, gastritis (inflammation of the stomach lining), stomach

cancer, and cancer of the mucosal lymphoid tissue (MALT lymphoma) (Homan and Orel, 2015). One of the ways that probiotics can be used against *H. pylori* infections is by reducing the inflammatory response induced by *H. pylori* via pro-inflammatory cytokines such as IL-8, IL6, IL-2, IL-1ß, and thus alleviate the symptoms of the diseases caused by *H. pylori* (Homan and Orel, 2015). For example, *in vitro* experiments have shown that probiotic *Lactobacillus bulgaricus* could reduce IL-8 production by inhibiting the TLR4 pathway utilised by *H. pylori* SS1-LPS to upregulate IL-8 in SGC-7901 cells (Zhou *et al.*, 2008). As already mentioned, probiotics can also produce antimicrobials, such as short-chain fatty acids, which can reduce the pH environment of the host and have a negative effect on *H. pylori* colonisation (Lesbros-Pantoflickova, Corthesy-Theulaz and Blum, 2007). *H. pylori* reduces host epithelial mucin production by suppressing the expression of MUC5AC and MUC1. Probiotics such as *Lactobacillus plantarum* strain 299v and *Lactobacillus rhamnosus* GG can stimulate mucin production by increasing expression of MUC2 and MUC3 genes, which may prevent *H. pylori* adhesion to gastric epithelial cells by improving the epithelial barrier (Homan and Orel, 2015). Probiotics, such as *Lactobacillus reuteri*, can also compete for adhesion sites, preventing *H. pylori* adhesion and thus colonisation (Homan and Orel, 2015). Meta-analysis studies on the effects of probiotics against *H. pylori* have concluded that probiotics are effective adjuncts to antibiotic therapy in the treatment of *H. pylori* infections, as the addition of probiotics can improve the efficacy of the treatment and reduce the side effects introduced when using antibiotics (Ruggiero, 2014).

1.4.5 Allergy

An allergic reaction is the result of a highly sensitised immune system that overreacts in the presence of a foreign antigen. According to the hygiene hypothesis, the lack of exposure to infectious agents and the absence of commensal microorganisms at an early

age, along with an overuse of antibiotics, prevent the immune system from developing normally (Waligora-Dupriet and Butel, 2012). This hypothesis also coincides with the fact that most allergic conditions are found in industrialised countries where hygiene and the use of antibiotics are prevalent (Waligora-Dupriet and Butel, 2012).

Some of the first microbial colonisers of infants (i.e. *Lactobacilli* spp) originate from the mother and reside in the gastrointestinal tract. These microbes may combat pathogens and modulate the immune system, thus aiding its natural development (Waligora-Dupriet and Butel, 2012). Many probiotics are also commensal microorganisms (Papadimitriou *et al.*, 2015), and it is proposed that the use of these bacteria may aid the natural development of the immune system and thus reduce the risk of allergies (Renz-Polster *et al.*, 2005). For example, probiotics may be used on children born by caesarean section as studies have shown that these infants do not receive microbes that would normally be obtained from their mother's vaginal tract and thus, they are highly likely to have allergic conditions (Renz-Polster *et al.*, 2005).

A study conducted by the World Allergy Organisation (WAO) stated that the evidence currently available on the supplementation of probiotics in infants does not indicate a decrease in risk of developing an allergy. Nonetheless, there are indications that the administration of probiotics is beneficial in preventing eczema (Fiocchi *et al.*, 2015). A double-blind study consisting of 62 pregnant women with a family history of atopic dermatitis found that the group that received probiotic *Lactobacillus rhamnosus* GG had offspring with lower risks of contracting atopic dermatitis as opposed to the placebo group (Rautava, Kalliomaki and Isolauri, 2002). Furthermore, a double-blind trial on 27 infants with eczema also found that in contrast to the control group, the condition of infants that had received *L. rhamnosus* GG and *Bifidobacterium lactis* improved (Islam, 2015). The WAO guideline recommends the use of probiotics on pregnant women,

women who are breastfeeding and infants with a high risk of developing allergic diseases (Fiocchi *et al.*, 2015).

1.5 An introduction to *Lactobacillus fermentum* 3872, a lactic acid bacterium

Bacteria are the predominant type of microbes that are classified as probiotics (Alok *et al.*, 2015). Among bacteria, the most common genera of probiotics are *Lactobacillus* and *Bifidobacterium* (Bermudez-Brito *et al.*, 2012; Islam, 2015). *Lactobacillus fermentum* belongs to the order Lactobacillales, which is the largest order in the phylum Firmicutes, and the class Bacilli (Claesson, Sinderen and O'Toole, 2007). *Lactobacillus* also belongs to the family Lactobacillacea (Claesson, Sinderen and O'Toole, 2007). Lactic acid bacteria (LAB), especially those that belong to *Lactobacillus* species, are often used in commercial food products such as in the meat, dairy and vegetable industries (Tajabadi *et al.*, 2013). *Lactobacilli* helps make up the normal microflora of many mammals including humans, contributing to the microbiota at several body sites, including the GIT (Bermudez-Brito *et al.*, 2012). Therefore, *Lactobacillus* have a generally regarded as safe (GRAS) status, which makes them good candidates for probiotic research (Anderson *et al.*, 2013). Within the GIT, *Lactobacilli* are commensal in the stomach (pH of 2), colon (pH 5–5.7) and the small intestine (pH 5–7), which also makes them useful in probiotic research relating to the GIT, as they can be ingested to reach their site of action (Jandhyala *et al.*, 2015).

As already noted, *Lactobacilli* are one of the earliest colonisers of the GIT and play a crucial role in reducing the risk of allergy in infants with a predisposition to heredity allergy, and thus may play a role in immune maturation (Johansson *et al.*, 2011). LABs also commonly produce antimicrobials such as bacteriocins and lactic acid which, as previously discussed, can be useful against infectious pathogens (Bermudez-Brito *et al.*, 2012). Nonetheless, *Lactobacillus fermentum* AGR1487 can lower the intestinal barrier

in vitro, and thus a safety profile of any potential probiotic must always be generated prior to its use (Anderson *et al.*, 2013).

The probiotic under investigation for this study is *Lactobacillus fermentum* 3872. The strain has been found in healthy human milk, in both mothers' and children's faecal matter, along with vaginal secretions. Thus, the strain is commensal to humans and can colonise various parts of the human body without adverse effects (Abramov *et al*., 2014; Lehri, Seddon and Karlyshev, 2017a). *In vitro* experiments have demonstrated that *L. fermentum* 3872 is resistant to gastric and intestinal stress and have strong adhesion to human HeLa and buccal cells. The bacteria also produce antimicrobials such as hydrogen peroxide and lactic acid (Abramov *et al*., 2014). The strain's commensal nature and its ability to produce antimicrobials, along with its adhesive traits makes *L. fermentum* 3872 a potential probiotic candidate. The strain, as a mixture of other probiotics, has been patented due to its ability to combat mastitis, which is an infection of the breast caused by *Staphylococcus* sp (Abramov *et al*., 2014).

1.6 OMIC technologies and their importance in probiotic research

OMIC technologies are an umbrella of tools, which include genomics, transcriptomics, proteomics, metabolomics, glycomics and lipomics. There has been an increase in interest in using OMIC approaches for probiotic research and to collect better evidence on how probiotics provide benefit to a host (Johnson and Klaenhammer, 2014). Genomics is the study of the genetic makeup of an organism (Horgan and Kenny, 2011). Transcriptomics measures the expression of RNA at a given time or under a certain environmental condition or infection (Horgan and Kenny, 2011). Proteomics is the study of proteins expressed by the genes of an organism including microbes (Horgan and Kenny, 2011). Glycomics and lipomics are involved in the study of cellular carbohydrates and lipids respectively. Metabolomics is the identification of the entire metabolite (end product) profile of an organism for a particular environmental

condition. An organism's metabolite profile can be useful in providing information about the interaction between the probiotic and host (Horgan and Kenny, 2011; Vernocchi, Chierico and Putignani, 2016).

The current study focuses on genomic analysis. Sequencing a probiotic's genome assists in characterising the microbe as a potential probiotic and helps to accurately categorise a probiotics phylogeny (Oelschlaeger, 2010). Identifying genomic relatedness via comparative genomics to another probiotic that has previously been studied may help direct the potential use of a probiotic and aid in determining the type of studies to be conducted. For example, comparative analysis was performed on the genomes of 34 strains of *Lactobacillus acidophilus*. The analysis found that there was very little genetic diversity between the strains and that they were almost identical at a nucleotide level, indicating that the strains, by and large, would provide similar health benefits (Bull *et al.*, 2014; Walsh *et al.*, 2016).

Genome sequencing may also aid in building a safety profile for a probiotic through the identification of genes involved in antibiotic resistance (Papadimitriou *et al.*, 2015). Genome sequencing may provide an indication of the likelihood of a probiotic's ability to transfer genetic material to another microbe by looking for genes involved in conjugation (Doron and Snydman, 2015). This would allow for precautions to be taken in preventing other microbes from acquiring virulence genes. Sequencing also permits the selection of candidate genes for functional analysis, such as the selection and expression of a candidate gene for *in vitro* experiments, or gene knockout experiments (Johnson and Klaenhammer, 2014). This approach may also help tailor the products of probiotics for pharmaceutical use (i.e. drug development) and aid in identifying a probiotics mechanism of action (Papadimitriou *et al.*, 2015; Johnson and Klaenhammer, 2014).

As previously mentioned, stability and viability of a probiotic is recommended by the WHO and FAO. Thus, sequencing, followed by re-sequencing of a microbe's genome, could be conducted after storage and/or administration to a host. This would help determine whether genomic changes such as gene loss or gain have been introduced during storage or after administration to a host, as genomic changes may result in a phenotypic effect (Walsh *et al.*, 2016). The ability of a probiotic to reach the site of action and survive long enough to be able to administer a beneficial effect is important. Genome sequencing can provide an idea of the probiotic's survivability for a specific host anatomical location by identifying genes related to bacterial survival (Walsh *et al.*, 2016), such as those involved in acid resistance systems (e.g. F_1F_0-ATPase), which is involved in the homeostasis of cellular pH (Papadimitriou *et al.*, 2015). Genes involved in adhesion provide information about survival and help determine whether a microbe can colonise the host at a particular anatomical location (Johnson and Klaenhammer, 2014). Identifying adhesion genes would also provide information on the ability of the probiotic to compete for host adhesion ligands against a pathogen, preventing it from colonising a host via competitive exclusion.

Comparative analysis of other probiotic genome sequences would allow further studies to be conducted to determine an ideal combination of probiotics that would synergistically provide a beneficial effect to a host (Johnson and Klaenhammer, 2014). Comparative genomics can also be used to identify why different strains of probiotics impose varying effects, such as a study that identified why some *Lactobacillus* spp are associated with host weight gain and others with weight loss (Drissi *et al.*, 2014). Genome sequencing, may be used for other OMIC technologies. For example, proteomics and metabolomic studies often require mass spectrometry data, which is dependent on a genome sequence and annotations of the sequence (Johnson and Klaenhammer, 2014), while transcriptomics needs an understanding of the genetic

makeup of an organism for measuring expression of transcribed genes (Johnson and Klaenhammer, 2014; Horgan and Kenny, 2011).

In summary, OMIC technologies can provide a holistic view of an organism and the effects it has on its host (Johnson and Klaenhammer, 2014). The data generated can be used to build a profile of a microbe by identifying both genotypic and phenotypic markers that comply with traits that are attributed to its probiotic properties (Johnson and Klaenhammer, 2014). Such information could be used to identify a probiotic's capacity to survive and reach its target region, its safety to the host, and its stability and possible mechanism of action, as recommended by the WHO and FAO (Johnson and Klaenhammer, 2014).

1.6.1 Genome sequencing and its complexities

There has been an exponential rise in the amount of genome sequencing data being generated and made publicly available via data repositories such as GenBank (Land *et al.*, 2014). This increase in sequencing data is mainly due to a reduction in the cost of sequencing technologies (Reuter, Spacek and Snyder, 2015). Nonetheless, despite the rapid accumulation of genomic data, most relate only to draft or incomplete genome sequences (Land *et al.*, 2014). This is seen by the fact that, since October 2013, more than eighty percent of the microbial sequencing data deposited in GenBank were of draft quality, with even relatively small genomes of around 1 MB left as incomplete sequences (Land *et al.*, 2014). Draft genome sequences are useful in obtaining minimal genetic information of a microbial genome, which can often miss out genes that may play a role in the phenotypic traits of microbes. It is essential to obtain a complete genome sequence of a microbe, particularly for probiotic research, in order to get a more complete picture of the probiotics genotype. This is because a probiotic's trait can

be strain-specific (Islam, 2015) and thus is reliant on genes that may not be discovered simply through a draft genome sequence.

Currently, even though there is a reduction in the cost of obtaining sequencing data, the cost and complexity involved in completing a genome sequence is a major reason for bacterial genome sequences being left in a draft state (Land *et al.*, 2014). To understand the complexities of a genome sequencing project, a brief overview and explanation of the sequencing assembly process is given. Initially, DNA from an organism is fragmented and the fragments are then sequenced. The sequence data generated is referred to as a 'read' (Ekblom and Wolf, 2014). The length of each read can vary from hundreds of bases to kilo-bases depending on the sequencing technology used (Ekblom and Wolf, 2014). Before submission to a data repository (e.g. Genbank), the read data generated needs to be assembled into longer genomic fragments, referred to as contigs. To assemble the reads, contiguation can be achieved using *de novo* assemblers, which utilise algorithms that find matches between reads and merge them to extend the reads (Ekblom and Wolf, 2014). This process does not require a reference genome.

An alternative assembly process is a reference-based assembly in which reads are mapped onto a closely related pre-assembled reference genome. Often the drawback of a reference-based assembly can be that novel genomic features of a sequenced genome are absent (Silva *et al.*, 2013). A reference-based assembly is also reliant on the quality of the reference genome assembly. Often assemblers do not contiguate beyond a repetitive region as it is difficult to determine the location of the repetitive regions on the genome (Treangen and Salzberg, 2011). Repetitive regions are commonly RNAs and/or transposons, where multiple copies of the same sequence are present in the genome. Many genome sequences deposited in data repositories use high throughput sequencers due to their affordability. High throughput sequencers such as the Ion Torrent Personal Genome Machine (PGM) generate high quality short sequencing reads

of about 400 bp. The sequencing machine also produces large amounts of the reads, resulting in a high coverage, which can help with nucleotide variant calling during genome assembly (Hunt *et al.*, 2013). The drawback of high throughput sequencing technology is that the *de novo* assembly of short reads often generate short contigs that fail to span across large repetitive regions of a genome, making genome sequence completion difficult (Hunt *et al.*, 2013). Certain manual bioinformatic techniques, such as BLASTN extension and read extension, can be used to extend contigs and merge contigs further, but this is often time-consuming.

Another genome completion method is to use sequencing machines, such as the PacBio RSII sequencer (Pacific Biosciences), which can generate reads of kilobases in length (i.e. >14 kb) (Reuter, Spacek and Snyder, 2015), although the technology is relatively new and expensive. Long reads also have a high sequencing error rate and generate little coverage, which can result in low-quality base calling and thus assembly (Reuter, Spacek and Snyder, 2015; Antipov *et al.*, 2016). Currently, a good approach is to use a hybrid of the two technologies, known as a hybrid sequence assembly, which uses contigs generated from long low-quality reads as a scaffolding tool and then maps short high-quality reads to identify nucleotide variations (Antipov *et al.*, 2016). Nonetheless, there could still be assembly errors introduced as described in this study, and thus a sequence-free tool that uses a restriction digest map of the genome, such as an optical map can help to validate the genome assembly (Mendelowitz and Pop, 2014).

1.6.2 Currently sequenced *L. fermentum* spp

At the start of this study, two complete and 16 draft genome sequences of *L. fermentum* strains were deposited in the Genbank database. The GenBank database had classified *L. fermentum* CECT 5716 as complete and circular, but this strain consisted of many nucleotide ambiguities and thus was classified as incomplete for this study (Lehri, Seddon and Karlyshev, 2017a). Currently, there are five complete genome sequences

with the inclusion of *L. fermentum* 3872, and 19 draft genome sequences of *L. fermentum* strains.

1.7 An introduction to *Campylobacter*

As previously mentioned, many *Lactobacilli* spp, being one of the earliest colonisers of the human GIT, are well-suited for GIT survival. For the current study, after sequencing the genome of *L. fermentum* 3872 and identifying a potential protein that could be utilised for competitive exclusion studies, functional analysis of the antagonistic effect of *L. fermentum* 3872 on the enteric pathogen *Campylobacter jejuni* was conducted.

Campylobacter is one of the leading causes of food-related illness worldwide (Kaakoush *et al.*, 2015; Johnson, Shank and Johnson, 2017) and is categorised by the WHO as a pathogen that urgently requires antimicrobial alternatives due to an emerging number of antibiotic-resistant species (World Health Organisation, 2017). Once consumed, *Campylobacter* spp binds to the mucus layer or the epithelial cell lining of the gastrointestinal tract. It has been theorised that after adhesion, clinical symptoms occur due to invasion of intestinal mucosa by *Campylobacter* spp (Janssen *et al*., 2008), leading to an inflammatory response and diarrhoea (Janssen *et al*., 2008), or by the prevention of intestinal fluid reabsorption (Janssen *et al*., 2008; Johnson, Shank and Johnson, 2017). *Campylobacter jejuni* causes Campylobacteriosis (an infection by *Campylobacter* spp), contributing to 90% of *Campylobacter* infections in the UK (Johnson, Shank and Johnson, 2017; Acheson and Allos, 2001). *C. jejuni* infections is one of the most frequent cause of acute gastroenteritis worldwide, with symptoms such as loose, watery or bloody diarrhoea, abdominal cramps and fever that can last up to a week (Kaakoush *et al.*, 2015). The onset of symptoms is dependent on the dose of *C. jejuni*, although it tends to be around 24–72 hours (Kaakoush *et al.*, 2015). Symptoms of C*ampylobacter* infections are also similar to other gastrointestinal infections such as

those caused by *Salmonella* and *Shigella*, making it difficult to distinguish between the various types of infections (Acheson and Allos, 2001).

C. jejuni infections have been implicated with IBD. As previously described, IBD is characterised by chronic inflammation of the GIT, including CD and UC (Kaakoush *et al.*, 2015). There is promising research being conducted on using probiotics to tackle IBD (Sánchez *et al.*, 2017). *Campylobacter* infections have also been associated with colorectal cancer and oesophageal disease, which can lead to oesophageal adenocarcinoma (Kaakoush *et al.*, 2015). Post *C. jejuni* infection can lead to IBS and functional dyspepsia (FD). IBS is characterised by recurrent discomfort or pain within the abdominal region for at least three months, for a minimum of three days each month (Kaakoush *et al.*, 2015). Studies have shown that between 3%–36% of IBS is related to enteric infection (Spiller and Garsed, 2009), with 9%-13.8% of *C. jejuni* infections leading to post-infectious IBS (Kaakoush *et al.*, 2015). Studies have also found that *C. jejuni* associated IBS can persist for up to 10 years (Schwille-Kiuntke *et al.*, 2011). FD is characterised by abdominal pain or discomfort in the upper abdomen for three months with the absence of observable structural abnormality of the upper abdomen during endoscopy examination (Kaakoush *et al.*, 2015). Studies have shown that between 9.6%-30.5% of individuals infected with enteric pathogens tend to get post-infectious FD (Kaakoush *et al.*, 2015).

Campylobacteriosis also increases the risk of celiac disease (Riddle *et al.*, 2013). Celiac disease is a disorder where an individual is unable to digest gluten, which causes damage to the lining of the small intestine. *C. jejuni* infections may also rarely cause extra-gastrointestinal disorders such as reactive arthritis, occurring in nine of every 1000 *Campylobacter* infections (Ajene, Walker and Black, 2013). *Campylobacter*-induced reactive arthritis can exhibit itself a month after infection, and can resolve within 1–5 years (Kaakoush *et al.*, 2015). Although extra-gastrointestinal infections are rare,

bacteraemia and septicaemia are common types of extra-gastrointestinal disorders caused by *Campylobacter* infections. A study by Nielsen *et al.* (2010) showed that bacteraemia and septicaemia could occur in the Danish population in 2.9 cases per 1,000,000 people every year. Extra-gastrointestinal conditions mainly occur in immunocompromised patients, or very young/old patients (Kaakoush *et al.*, 2015; Acheson and Allos, 2001).

C. jejuni can also be dangerous for pregnant females with infections linked to septic abortion and neonatal sepsis (Kaakoush *et al.*, 2015). They are also associated with autoimmune disorders such as Guillain-Barré syndrome (GBS) and Miller Fisher syndrome (Kaakoush *et al.*, 2015). GBS is a disorder characterised by a weakness or loss of sensation in the legs and arms caused by damage to the nervous system (Van Doorn, Ruts and Jacobs, 2008). The incidence of GBS is 1.2-2.3 cases per 100,000 people per year, with *C. jejuni* infection being a major trigger for the disease (Van Doorn, Ruts and Jacobs, 2008). The risk of GBS infection increases with age and is more common in men (Van Doorn, Ruts and Jacobs, 2008). GBS is thought to be caused by cross-reactivity of host antibodies to *C. jejuni* lipooligosaccharide and human gangliosides (Kaakoush *et al.*, 2015; Van Doorn, Ruts and Jacobs, 2008). Miller Fisher syndrome is a variant of GBS; and causes paralysis or defect of the eye muscle and is thought to be a result of *C. jejuni* molecular mimicry (Kaakoush *et al.*, 2015).

1.7.1 Epidemiology of *Campylobacter* infections

There is a global rise in *Campylobacter* infections (Kaakoush *et al.*, 2015). In industrialised countries, the annual incidence of *Campylobacter* infections ranges from 4.4-9.3 per 1000 people (World Health Organization, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, 2013). In many industrialised countries, *Campylobacter* infections often occur in individuals less than a year old and during young adulthood (between the ages of 15 and 44 years)

(Acheson and Allos, 2001). However, according to a recent public health report, the majority of UK *Campylobacter* infections in 2015 occurred in people aged between 50 and 59 years (Public Health England, 2017). *Campylobacter* infection rates show a seasonal increase during the spring/summer months in the UK (Public Health England, 2017; Nichols *et al.*, 2012). As already noted, in the UK, males are more likely to suffer from *Campylobacter* infections, with 14% more male cases of Campylobacteriosis than female cases reported (Nichols *et al.*, 2012).

An important risk factor for *Campylobacter* infection is international travel, with *C. jejuni* contributing significantly to travellers' diarrhoea (Kaakoush *et al.*, 2015). In Europe, the cases of travellers' diarrhoea related to *Campylobacter* infections has increased from 7% (2008) to 12% (2010) (Gautret *et al.*, 2012), and in England, 17% of *Campylobacter* infections are related to travel (Zenner and Gillespie, 2011). In the UK as a whole, one-fifth of reported *Campylobacter* spp cases are travel-related (Nichols *et al.*, 2012). In developing countries, *Campylobacter* infections are endemic and tend to occur in children less than two years old, although they are rare among adults (Acheson and Allos, 2001; Platts-Mills and Kosek, 2014). *Campylobacter* infections are also an important cause of childhood morbidity due to diarrheal symptoms in the developing world (Acheson and Allos, 2001). In the developing world, infants infected with *Campylobacter* spp also exhibit stunting (Amour *et al.*, 2016). Furthermore, in contrast to industrialised countries, *Campylobacter* infections do not show a seasonal increase in developing countries and can be asymptomatic (Nichols *et al.*, 2012).

1.7.2 Reservoir of *Campylobacter* infections

Campylobacter transmission to humans is commonly zoonotic, with 50%–80% of infections related to chickens (EFSA Panel on Biological Hazards, 2011). This is mainly caused by a high *Campylobacter* spp carriage rate in broiler chickens, resulting in a large number of poultry farms being contaminated with *Campylobacter* spp

(Kaakoush *et al.*, 2015). *C. jejuni* has also been known to survive in chicken faeces for about six days, which can lead to soil and water contamination, especially if manure from poultry is used as fertiliser (Kaakoush *et al.*, 2015). A survey conducted by the UK Food Standards Agency found that the prevalence of infection with *Campylobacter* spp was 73.3% in chickens sold in the UK from 2014 to 2015 (Jorgensen et al., 2017). *Campylobacter* spp abundance is more common in organic chickens than non-organic sources, with the risk of infection being 1.7 times higher from organic sources (Kaakoush *et al.*, 2015). Furthermore, frozen and fresh meats tend to be more frequently contaminated with *Campylobacter* spp than commercially cooked meat, with *C. jejuni* accounting for 69% of *Campylobacter* spp contamination (Moore *et al.*, 2002). Apart from chicken, other forms of domestic animals, such as cattle, pigs, sheep, goat, cats and dogs, contribute to *Campylobacter* infections (Kaakoush *et al.*, 2015). Another leading risk of *Campylobacter* infection is untreated water. Contaminated water is a more common vector of transmission than poultry in developing nations (Kaakoush *et al.*, 2015). *Campylobacter* infections can also result from drinking unpasteurized milk (Kaakoush *et al.*, 2015) and, according to a study conducted by the UK Health Protection Agency, 3% of infections are also a result of person-to-person contact caused by poor hygiene (Kaakoush *et al.*, 2015).

1.7.3 Current treatments

Campylobacter infections are commonly treated with hydration therapy to maintain normal body fluid and electrolyte balance (Kaakoush *et al.*, 2015; Acheson and Allos, 2001). Nonetheless, antibiotics are administered to individuals who are immunocompromised or who show severe/persistent or extra-gastrointestinal symptoms (Kaakoush *et al.*, 2015; Acheson and Allos, 2001), and individuals who are pregnant, have high fevers or have a prolonged illness (Kaakoush *et al.*, 2015; Acheson and Allos, 2001). When antibiotics are needed, fluoroquinolones such as ciprofloxacin are usually

given. These target topoisomerase IV and DNA gyrase. The latter enzymes are essential for DNA replication. Fluoroquinolones are favourable as they are effective against various pathogens that cause gastroenteritis and thus can be quickly prescribed, although the rate of ciprofloxacin antibiotic resistance is high and rising amongst *Campylobacter* species (Kaakoush *et al.*, 2015; Acheson and Allos, 2001). In the US, ciprofloxacin resistance rose from 11% in 1997 to 25% in 2011 (Hampton, 2013), while in South Korea and Thailand, ciprofloxacin resistance is 92% (Han *et al.*, 2007) and 100% (Serichantalergs *et al.*, 2007) respectively. In European states, resistance to fluoroquinolones is up to 91.5% (European Food Safety Authority and European Centre for Disease Prevention and Control, 2016). This increase in resistance is due to the misuse of ciprofloxacin in poultry farming, animal husbandry and, to a lesser extent, due to suboptimal use in the treatment of human disease (Kaakoush *et al.*, 2015). If Campylobacteriosis is diagnosed, then macrolides are preferred which block translocation required for protein synthesis (Kaakoush *et al.*, 2015). A study conducted by Tribble *et al.* (2007) showed that treatment of travellers' diarrhoea caused by *C. jejuni* and *C. coli* with levoflaxin (fluoroquinolone) had a 72% cure rate, which was lower than treatment with azithromycin (macrolide), which had a 98% cure rate. Erythromycin, another macrolide, is also an effective antibiotic against *Campylobacter*, as there is low antibiotic resistance, and the antibiotic can be given to children and pregnant women (Acheson and Allos, 2001). Nonetheless, macrolides are being used in the food production industry to promote the growth of animals, as well as for therapeutic reasons, leading to macrolide-resistant strains of *Campylobacter* (Kaakoush *et al.*, 2015). Ciprofloxacin has been recommended for macrolide-resistant *Campylobacter* (Kaakoush *et al.*, 2015). Various strains of *Campylobacter* resistant to tetracycline, amoxicillin, cephalosporins, ampicillin, metronidazole, along with trimethoprim, vancomycin and rifampin have been found (Acheson and Allos, 2001). In

rare cases, *Campylobacter* infections are treated with aminoglycosides such as gentamicin, which affect protein synthesis (Kaakoush *et al.*, 2015). Therefore, due to a rise in antimicrobial resistance, alternative means of combating *Campylobacter* infections are needed (Johnson, Shank and Johnson, 2017).

1.7.4 Possible methods to tackle *C. jejuni* infections

One of the possible ways to reduce *Campylobacter* infections is to introduce hygiene measures in farms, such as the introduction of boot dips, house-specific boots/overshoes and hand washing facilities (Kaakoush *et al.*, 2015). This would prevent the spread of *Campylobacter* between broiler houses, ultimately reducing the number of chickens infected with *Campylobacter* and thus prevent its spread to humans (Kaakoush *et al.*, 2015). These methods are effective in reducing *Campylobacter* in chicks, although the downside of the aforementioned strategies is the willingness of the workers to adhere to the hygiene measures (Kaakoush *et al.*, 2015).

Another approach to combating *Campylobacter* infections is the use of bacteriocins. Many anti-*Campylobacter* bacteriocins have been isolated from *Lactobacillus* bacteria. For example, bacteriocin L-1077 isolated from *Lactobacillus salivarius* 1077 was found to reduce *C. jejuni* infection in chickens (Svetoch *et al.*, 2011). Furthermore, a class II bacteriocin produced by *Paenibacillus polymyxa* NRRL B-3059 also reduced *C. jejuni* colonisation of chickens (Johnson, Shank and Johnson, 2017). Nonetheless, *C. jejuni* and *C. coli* have been known to develop resistance against bacteriocins, as was the case for bacteriocins OR-7 and E-760 (Johnson, Shank and Johnson, 2017). A study by Johnson, Shank and Johnson (2017) showed that the mechanism of resistance was due to the multidrug CmeABC efflux pump. In addition, studies have also shown phages, many of which have been isolated from sources such as sewage, poultry, livestock and manure, to be effective against *Campylobacter* (Johnson, Shank and Johnson, 2017). Bacteriophages added to a chicken's water source have been found to successfully

reduce the *Campylobacter* load (Kaakoush *et al.*, 2015), although prolonged studies show that *Campylobacter* levels return to their original level over time after constant use of a particular bacteriophage (Kaakoush *et al.*, 2015). This loss of effectiveness may be caused by *Campylobacter* bacteria having a high rate of genomic variation, which would make the bacteria resistant to a phage (Johnson, Shank and Johnson, 2017). Nonetheless, phages can evolve consistently to attack the host (Johnson, Shank and Johnson, 2017). So far, studies that utilise phage therapy to treat *Campylobacter* infections in humans have not been conducted. A cocktail of phages may be used to treat *Campylobacter* infections, as currently there are no individual phages known to be effective against all *Campylobacter* strains (Kaakoush *et al.*, 2015; Johnson, Shank and Johnson, 2017). There have also been strategies to vaccinate humans and chickens against *Campylobacter* infections, such as delivering flagellar components of *C. jejuni* and using *Campylobacter* inner membrane antigens, although none of the approaches have successfully prevented *Campylobacter* colonisation in chickens or protected humans against *Campylobacter* infections (Kaakoush *et al.*, 2015; Johnson, Shank and Johnson, 2017). Removal of *Campylobacter* during food processing, by using organic acids, sodium chlorite, or UV may successfully prevent human cases of campylobacteriosis. However, either the costs are high and/or there is a potential effect on the taste of the food product (Kaakoush *et al.*, 2015). Studies have also shown that probiotics can be an effective tool in reducing or preventing *Campylobacter* infections as mentioned in a review article by Johnson, Shank and Johnson, (2017). These studies have predominantly utilised competitive exclusion as a tool for reducing *Campylobacter* infections with promising results (Johnson, Shank and Johnson, 2017). Probiotics may be utilised as a prophylaxis against *Campylobacter* associated travellers' diarrhoea, and can be useful in countries where *Campylobacter* infections are endemic (Johnson, Shank and Johnson, 2017). The most common types of probiotics used for anti-*C. jejuni*

studies are *Lactobacillus* and *Enterococcus* due to their common presence within the human GIT, along with *Bifidobacterium* and *Saccharomyces cerevisiae* (Johnson, Shank and Johnson, 2017). A study conducted by Nishiyama *et al.* (2014) treated chicks with daily doses of *Lactobacillus gasseri* SBT2055 (LG2005) over a 14-day period. The results showed a significant decrease in *C. jejuni* 81-176 in the treated chicks compared to controls (Nishiyama *et al.*, 2014). This study concluded that the probiotic strain inhibited *C. jejuni* 81-176 adhesion to human intestinal epithelial cells and that there was co-aggregation between the probiotic and *C. jejuni* 81-176 (Nishiyama *et al.*, 2014). Another study, conducted by Wine *et al.* (2009) determined that probiotic *Lactobacillus helveticus* R0052 could reduce *C. jejuni* NCTC 11168 and 81-176 invasion to human colon T84 cells and *C. jejuni* NCTC 11168 invasion to intestine 407 cells. The authors deduced that competitive exclusion as a result of *L. helveticus* R0052 adhesion to the latter cells played a role in the reduction of invasion of the two *C. jejuni* strains (Wine *et al.*, 2009).

Another means of utilising probiotics to combat *Campylobacter* is through the release of organic acids commonly produced by *Lactobacillus* spp, as studies such as that conducted by Neal-McKinney *et al.* (2012) have shown that species of *Lactobacillus* inhibit *C. jejuni* growth *in vitro* by disrupting the membrane of *C. jejuni* by releasing lactic acid. The study demonstrated that low pH may play a role in inhibiting *C. jejuni* growth, as after neutralising the pH, *C. jejuni* growth was not affected. Interestingly, the study also demonstrated that reducing the pH environment with hydrochloric acid had a weaker inhibitory effect than when using lactic acid, indicating that the pH is not the only inhibitory factor (Neal-McKinney *et al.*, 2012). Nonetheless, the effect of lactic acid within a gut environment may be limited due to the buffering effect of bicarbonate released by the pancreas (Johnson, Shank and Johnson, 2017). Along with combating *Campylobacter*, probiotics may also have a beneficial effect on the growth of livestock.

Studies have shown that probiotics in chicken feed increased livestock's nutrient transport capacity, metabolic rate, and protein production (Johnson, Shank and Johnson, 2017). Probiotics also have an added benefit in that they are easy to administer as food, they are cheap, and they may stay within their host for a prolonged period (Neal-McKinney *et al.*, 2012).

1.7.5 *C. jejuni* adhesins

Adhesion is an important tool utilised by pathogenic bacteria to colonise and invade a host and thus for pathogenesis (Rubinchik, Seddon and Karlyshev, 2012). For this reason, a better understanding of the means by which *C. jejuni* adheres and thus colonises a host is important. Some examples of *C. jejuni* adhesins are provided in this section.

The most important adhesins of *Campylobacter* are the cell surface membrane proteins CadF, FlpA and JlpA. FlpA and CadF which attach to fibronectin, a component of the extracellular matrix proteins (ECM) (Rubinchik, Seddon and Karlyshev, 2012; Konkel, Larson and Flanagan, 2010). Proteins that attach to the ECM are referred to as microbial surface adhesive matrix molecules (MSCRAMMS), and are often utilised as primary tools for host pathogenesis and invasion; thus, they are good targets for competitive exclusion experiments (Kline *et al.*, 2009). A study by Kuusela *et al.* (1989) showed that *C. jejuni* binds to various other ECM proteins such as collagen I, III, IV and V. These proteins are found in the GIT although, to the authors' knowledge, the proteins involved in collagen adhesion have not been identified for *C. jejuni*. The JlpA protein of *C. jejuni* binds to Hep-2 epithelial cells (Rubinchik, Seddon and Karlyshev, 2012). JlpA also starts a host inflammatory response by initiating the NF- κB and p38 MAP kinase production (Kawai *et al.*, 2012). The gene *Cj1349*, contains fibronectin (Fn) type III domains, indicating a potential ability to adhere to fibronectin. This protein plays a role in adhesion to chicken epithelial cells (Flanagan *et al.*, 2009). The Campylobacter

protein A (CapA protein) of *C. jejuni* plays a role in adhesion and invasion, with *C. jejuni capA* mutant genes showing reduced adhesion and invasion potential to Caco-2 cells (Ó Cróinín and Backert, 2012). The Cj0091 protein of *C. jejuni* plays a role in adhesion to INT 407 cells, with mutants showing reduced adhesion ability (Oakland *et al.*, 2011). The major outer membrane protein (MOMP) and a flagella subunit protein (FlaA) of *C. jejuni* also play a role in adhesion to INT 407 cells (Rubinchik, Seddon and Karlyshev, 2012). MOMP, as well as FlaA, is associated with adhesion to human histo-blood group antigens (Mahdavi *et al.*, 2014). Mutation analysis of the *tylA* gene of *C. jejuni* 81-176 resulted in a reduction in adhesion to human Caco-2 cells, and thus, it is also a potential adhesin. The Peb1 protein of *C. jejuni* is found to play a role in adhesion, with *peb1* mutants showing a reduction in adherence to HeLa cells and a reduction in invasion to INT-607 cells (Ó Cróinín and Backert, 2012). The Peb4 protein is thought to be a molecular chaperone for adhesion proteins such as CadF (Ó Cróinín and Backert, 2012). Other potential adhesion proteins utilised by *C. jejuni* are CapB, which has amino acid sequence similarity to CapA, and P95. These have been associated with adhesion to Caco 2 cells in other pathogenic bacteria such as *Haemophilus* and *Bordetella* (Rubinchik, Seddon and Karlyshev, 2012).

1.7.6 *C. jejuni* sequencing information

To date, 33 species belonging to the genus *Campylobacter* have been recorded. *Campylobacter* belongs to the class Epsilonproteobacteria and the order Campylobacteria. The order includes the genera *Helicobacter* and *Wolinella*. The *Campylobacter* genera have genomes ranging from 1.6–2.0 Mb. The *C. jejuni* genome sequence tends to have high levels of genetic variations with hyper variable regions, many of which are involved in the biosynthesis of capsule, lipooligosaccharide (LOS) and flagella structures (Parkhill *et al.*, 2000; Young, Davis and DiRita, 2007). Genomic variations tend to occur due to phase variation or insertions, deletions, duplications,

point or frameshift mutations (Parkhill *et al*., 2000; Young, Davis and DiRita, 2007). *C. jejuni* can also acquire DNA via transformation, transduction and conjugation (Johnson, Shank and Johnson, 2017). *C. jejuni* prefers taking DNA from closely related strains/species, rather than other bacterial populations (Young, Davis and DiRita, 2007). This is more common under low carbon dioxide and high cell density levels (Johnson, Shank and Johnson, 2017; Young, Davis and DiRita, 2007), indicating that environmental factors control the increase in DNA uptake, although the exact reasoning for increased horizontal gene transfer is unknown (Johnson, Shank and Johnson, 2017). The high genomic variability potentially contributes to the bacterium's propensity to develop antibiotic resistance. (Young, Davis and DiRita, 2007).

1.8 Aims and objectives

The aim of this study was to complete the genome sequence of *L. fermentum* 3872 in order to identify potential genes attributed to probiotic properties of the bacterium. The study also aimed to select a candidate gene that could be used for competitive exclusion experiments against *C. jejuni* strains 11168H and 81-176.

Objectives of the study were to:

1. Circularise the genome sequence of *L. fermentum* 3872.
2. Check the quality of the *L. fermentum* 3872 genome assembly.
3. Identify genes of *L. fermentum* 3872 potentially involved in the probiotic activity of the bacterium.
4. Investigate whether putative CBP of *L. fermentum* 3872 binds to collagen I, a protein found on the epithelial lining of the GIT.
5. Investigate the binding of *C. jejuni* strains 11168H and 81-176 to collagen I and identify the protein(s) involved in collagen I adhesion.

6. Determine whether recombinant CBP protein and *L. fermentum* 3872 inhibit binding of *C. jejuni* strains 11168H and 81-176 to collagen I.
7. Identify whether *C. jejuni* strains 11168H and 81-176 growth is inhibited in the presence of *L. fermentum* 3872.

Chapter 2: Materials and Method

2.1 Laboratory consumables:

Laboratory consumables were purchased from Fisher Scientific and Sigma-Aldrich.

2.2 Water

Distilled water for microbiology work was obtained from the PureLab Option ELGA DV35 water purification device. For molecular biology work distilled water was obtained from Milli-Q Millipore Direct-Q UV.

2.3 Growth media and supplements

2.3.1 Growth media

All media were made sterile by autoclaving at 121°C for 15 minutes unless otherwise stated. Nutrient media (powdered) were all suspended in water (ELGA DV35) and made using manufacturers' guidelines. Additional supplements added during media preparation are indicated alongside the manufacturer of the media used below:

De man, Rogosa, Sharpe (MRS) agar (Oxoid) and De man, Rogosa, Sharpe (MRS) broth (Oxoid) were prepared following the manufacturer's instructions.

Luria-Bertani (LB) Lennox broth (Fisher) was developed in accordance with the manufacturer's guidelines. Luria-Bertani (LB) agar was prepared by mixing 10 g LB broth Lennox (Fisher) and 7.5 g agar (Fluka) in 500 ml water (ELGA DV35). For some experiments, chloramphenicol (CAM; Sigma) was added to make a final concentration of 25 µg/ml.

Columbia blood agar base (CBA; Oxoid) was prepared according to the manufacturer's instructions. After sterilisation, the agar was cooled to about 50°C and supplemented with 5% defibrinated horse blood (Oxoid) and 0.2% Skirrow supplement (Oxoid).

Mueller-Hinton (MH) broth (Fluka) and Muller-Hinton (MH) agar (Fluka) were prepared in accordance with the manufacturer's instructions in water (ELGA DV35). For agar well diffusion assay experiments, MH soft agar was prepared by mixing MH broth (Fluka) and 0.75% agar (Fluka). To store bacterial cultures at -80°C MH broth (Fluka) was mixed with 15% glycerol.

For NEB express *E. coli* transformation, Super optimal broth with catabolite repression (SOC) medium (NEB) was used following manufacturers guidelines. The SOC medium was pre-sterilised by the manufacturer.

2.3.2 Antibiotic supplements

Premade chloramphenicol (CAM) in a concentration of 34 mg/ml (purchased from Sigma) was stored at -20°C. A final concentration of 25 µg/ml was used for experiments. Skirrow supplement (Oxoid) was made in accordance with the manufacturer's instructions using Milli-Q water. Milli-Q water was used as it contained a 0.22 µm filter for sterilisation purposes. The Skirrow supplement was stored at -20°C.

2.4 Buffers and reagents

2.4.1 PBS buffer

Phosphate buffered saline (PBS) (Fisher) was made to 1x in water (ELGA DV35) and sterilised by autoclaving.

To store bacterial culture, glycerol PBS solution was made by mixing 1x PBS (Fisher) with a final concentration of 15% glycerol (Fisher).

2.4.2 Western blot buffers

Tris-buffered saline (TBS) 10x concentrate was made by adding 18.2 g of Tris(hydroxymethyl)amino-methane (Sigma) with 87 g NaCl (Sigma) in 900 ml in water (ELGA DV35). The buffer was then adjusted to a pH of 8 using 1M hydrochloric

acid (HCl; Sigma) and made to 1 L in water (ELGA DV35). A 1x dilution was used for all experiments.

Tris-buffered saline Tween (TBST) 1x was made by adding 0.1% Tween 20 (Sigma) to 1x TBS.

Transfer buffer was made by mixing 3 g Tris(hydroxymethyl)amino-methane (Sigma) to 15 g of Ultra-Pure Glycine (National Diagnostics) to which 800 ml of water (ELGA DV35) was added. Two hundred millilitres of methanol was also added to the buffer to make a final volume of 1 L. The buffer was stored at 4°C.

2.4.3 Enzyme linked immune absorbance assay (ELISA) buffers and consumables

ELISA coating buffer was prepared by mixing 0.19 g Sodium Carbonate (Fisher) and 0.37 g Sodium Bicarbonate (Sigma) in 125 ml Milli-Q water; pH 9.6. The buffer was filter sterilised using a 0.22 µm filter (Fisher).

Phosphate buffered saline with tween 20 (PBST), used as a wash buffer for ELISA experiments, was prepared by adding tween 20 (Sigma) to a final concentration of 0.1% in sterile 1x PBS (Fisher).

For ELISA experiments the plates used were 96-well flat bottom and transparent. The plates were either UV-treated Nunclon Delta (ThermoFisher Scientific), UV-treated High Bind Corning Costar and Untreated Med Bind Corning Costar plates.

2.4.4 Silver stain reagents

Gel fix reagent was made by mixing Milli-Q water with molecular biology grade ethanol (Fisher) and acetic acid (Fisher) at a ratio of 6:3:1 respectively.

Ethanol wash was made by adding molecular biology grade ethanol (Fisher) to Milli-Q water for a final ethanol concentration of 10%.

Reaction stop reagent was prepared by adding acetic acid (Fisher) to Milli-Q water to get a 5% acetic acid concentration.

2.4.5 Preparation of Collagen I

Collagen I Bovine (Sigma) was dissolved in 0.1 M acetic acid (Fisher) for 1-3 hours at room temperature to obtain a concentration of 1 mg/ml collagen I. The solution was stored at 4°C before use.

2.5 Bacterial growth and storage

2.5.1 Bacterial strains and plasmid used for experiments

L. fermentum 3872 was isolated from a healthy human female's milk. As recorded by Lehri, Seddon and Karlyshev (2017a), the bacterium was deposited in the All-Russian Collection of Microorganisms at the Moscow regions in Pushchino, with the accession number VKM B-2793D.

C. jejuni strains used for this study were 11168H and 81-176. *C. jejuni* 11168H is a hypermotile variant of *C. jejuni* NCTC 11168 and was isolated from human faeces (Karlyshev *et al.*, 2002). *C. jejuni* 81-176 was originally isolated from raw milk, and is a highly virulent strain (Hu and Kopecko, 1999).

NEB express *E. coli*, a derivative of *E. coli* BL21, was used for cloning experiments. NEB express *E. coli* is deficient in proteases Lon and OmpT and is not resistant to chloramphenicol (CAM) used for cloning experiments.

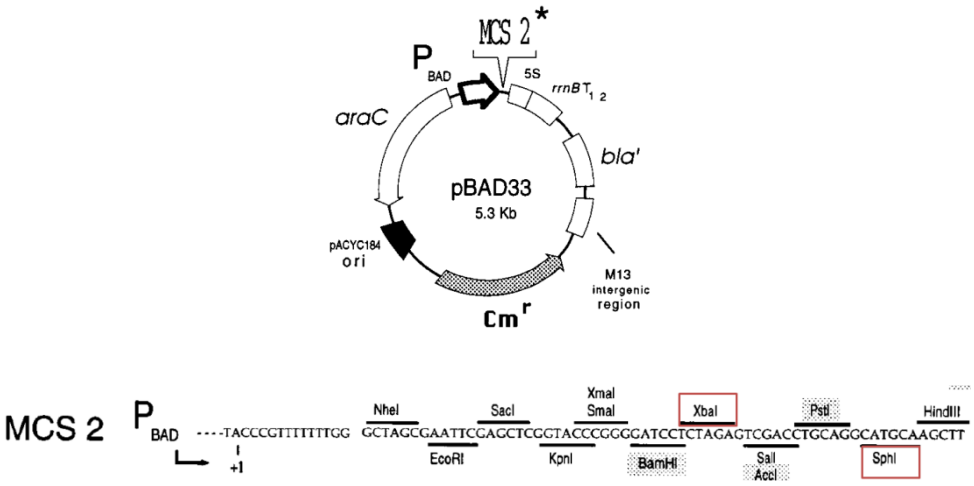


Figure 2.1. Diagram of the expression vector pBAD33. The recombinant gene is to be placed between the red boxed restriction enzymes *Xba*I and *Sph*I on the MCS2 region.

The expression vector pBAD33 was used for the protein expression experiment (Figure 2.1). The circular plasmid is 5.3 kb in size and contains a P_{BAD} promoter region that allows control of protein expression via regulation by arabinose. The plasmid also contains a chloramphenicol resistance gene for selection of transformed cells.

2.5.2 Bacterial growth conditions

Lyophilised *L. fermentum* 3872 was re-suspended in 1 ml PBS (Fisher) 1x, 50 µl of which was streaked onto MRS (Oxoid) agar to isolate single colonies. The bacteria were grown overnight at 37°C in an anaerobic container (BD GasPak), under anaerobic conditions. Three single colonies were re-streaked onto MRS agar (Oxoid) and grown overnight. Each of the three samples were stored in 15% glycerol-PBS (Fisher) at -80°C.

For experiments, ten microliters of *L. fermentum* 3872 culture stocks were plated onto MRS agar (Oxoid) plates and grown overnight at 37°C under anaerobic conditions. For agar well diffusion assay, *L. fermentum* 3872 was initially grown on MRS agar (Oxoid) under anaerobic conditions and then suspended in 10 ml MRS broth (Oxoid), the culture was then grown on a shaker at 120 rpm overnight.

C. jejuni strains 11168H and 81-176 culture stocks were stored in MH broth (Oxoid) with 15% glycerol (Fisher) at -80°C. Twenty microliters of the stock were aliquoted onto CBA agar (Oxoid) for growth. The strains were grown at 37°C in a microaerobic incubator (Don Whitley) overnight, under conditions 10 % CO_2, 5% O_2, and 85 % N_2**.** *C. jejuni* was again restreaked onto CBA agar (Oxoid) plates and grown for 24 hours before use.

NEB express *E. coli* was grown overnight on LB agar (Fisher) or in LB broth (Fisher) at 37°C. For some cultures, the LB broth/agar was supplemented with CAM (25 µg/ml). Transformed *E. coli* culture stocks were stored at -80°C in 15% glycerol-PBS.

All culture stocks were prepared from fresh one-day-old bacterial growth.

2.6 Agarose gel electrophoresis

Tris/Borate/EDTA (TBE) buffer 10x (BIO-RAD) was diluted to 1x in water (ELGA DV35), which was added to agarose (Fisher) to make a 1% agarose gel. The mixture was heated in a microwave and left to cool. Once cool, ethidium bromide 0.5 µg/ml (Fisher) was added to the solution, which was swirled to mix. The solution was poured into a gel tank (Fisher Scientific) with a 12 well comb and left to solidify. Electrophoresis was conducted in 1x TBE buffer (BIO-RAD) at 120 V for one hour using a VWR power unit. Blue gel loading dye (NEB) diluted to 1x was used to help load samples onto each well. For all agarose gel electrophoresis experiments, the DNA size molecular marker used was 2-log DNA Ladder (NEB). The gel was visualised using a G: BOX UV-transilluminator (Syngene) to excite the ethidium bromide. Image processing was conducted using GenSnap software (Syngene).

2.7 DNA extraction

Qiagen Gentra Puregene Yst/Bct kit was used to extract DNA from *L. fermentum* 3872, following the manufacturer's protocol. For Ion Torrent (Life Technologies) sequencing,

DNA was extracted from the confluent growth of two *L. fermentum* 3872 clonal isolates labelled clonal isolate 1 and 2.

For PacBio RSII (Pacific Biosciences) sequencing, DNA was extracted from clonal isolate 1. The Qiagen Gentra Puregene Yst/Bct kit protocol was also modified, with *L. fermentum* 3872 being incubated for 30 minutes at 37°C in a customised lysozyme buffer. The buffer consisted of 20 mg/ml egg white lysozyme (Sigma) with 1% Triton (Sigma) and 10% TE (Sigma) in Milli-Q water. After incubation, the sample was spun down using an Eppendorf mini spin plus micro-centrifuge at 14,104 x g for two minutes, and then the supernatant was discarded. Cell lysis solution provided by the Qiagen Yst/Bct kit was supplemented with 20 µl of proteinase K (Invitrogen), 300 µl of which was added to the sample and incubated at 80°C for five minutes as per standard protocol. The sample was incubated in protein precipitation solution provided by the Qiagen Gentra Puregene Yst/Bct kit for one hour on ice followed by one hour incubation with RNase (Qiagen). The latter was done to remove protein and RNA contaminants as very pure samples are required for PacBioRSII sequencing due to the absence of a DNA amplification step during template preparation.

E. coli plasmid extraction was conducted for cloning experiments. DNA was extracted using the QIAprep Spin Miniprep kit (Qiagen) following the manufacturer's protocol.

2.8 DNA quality assessment

DNA quality and concentration were determined by using a NanoVue spectrophotometer (GE Healthcare Life Sciences), gel electrophoresis, Qubit fluorometer (Invitrogen), DropSense (Trinean) and Bioanalyzer 2100 (Agilent Technologies).

2.9 Next-generation sequencing

2.9.1 Fragmentation and library preparation for Ion Torrent PGM sequencing

L. fermentum 3872 clonal isolate 1 and 2 were sequenced using the Ion Torrent Personal Genome Machine (PGM; Life Technologies). DNA fragmentation and library preparation were conducted using NEBNext fast DNA fragmentation and library preparation kit for Ion Torrent (NEB). During the fragmentation step, the samples were incubated in NEBNext fragmentation master mix at 25°C for 10 minutes, followed by 10 minutes heat inactivation of the fragmentation enzyme at 70°C. The rest of the protocol was followed in accordance with the kit's guidelines.

The protocol for 100 ng - 1 µg was followed for adapter ligation as the concentration of the DNA was above 100 ng. Standard protocol for clean-up of adapter ligated DNA was followed using AMPure XP Beads (Invitrogen).

A 2% E-Gel cassette (Invitrogen) and an E-gel imaging device (Invitrogen) were used to select fragmented DNA of around the 490 bp mark, as recommended by the Ion Torrent 400 bp sequencing kit guidelines. Twenty microliters of *L. fermentum* 3872 DNA from clonal isolate 1 and 2 were mixed with bromophenol blue loading dye (NEB) 1x. The samples were added to each E-Gel well. A 2-log DNA Ladder (NEB) was used as a molecular marker. The lower wells of the E-Gel cassette were filled with 20 µl of Milli-Q water. The E-Gel was run until the ladder reached the 490 bp mark at which time the lower wells were filled with 20 µl of Milli-Q water. The E-Gel was run for another two minutes to extract DNA fragments of around 490 bp in size.

The fragmented samples, with adapters attached, were amplified using polymerase chain reaction (PCR). The protocol for 100 ng - 1 µg was followed as recommended by the NEBNEXT fast DNA fragmentation and library preparation kit's instructions (NEB) and the reaction was made to 50 µl. The PCR cycling conditions are shown in Table 2.1.

Table 2.1. PCR cycling conditions used for adapter ligated DNA.

Initial Denaturation	98°C (30seconds)	1 cycle
Denaturation **Annealing** **Extension**	98°C (10seconds) 58°C (30seconds) 72°C (30seconds)	10 cycles
Final extension	72°C (5minutes)	1 cycle

The amplified library was cleaned using 50 µl of Ampure Xp beads (Invitrogen), following the NEBNEXT fast DNA fragmentation and library preparation kit protocol.

Bioanalyzer 2100 (Agilent Technologies) and the Agilent DNA 1000 kit (Agilent Technologies) were used to determine the size of the DNA fragment and the concentration. Two readings were taken for each amplified sample in order to take pipetting error into consideration. The samples were prepared in accordance with the manufacturer's protocol.

2.9.2 Template preparation and sequencing for Ion Torrent PGM

Three sequencing runs were conducted on the Ion Torrent PGM (Life Technologies) for *L. fermentum* 3872. Two sequencing runs were for clonal isolate 1 and one sequencing run was for clonal isolate 2.

Ion PGM template OT2 400 bp kit (Life Technologies) was used for template preparation. Clonal 1 sequencing run 1 concentration was initially made to 26 pM/L as recommended by the kit's protocol. This was reduced to 20 pM/L for sequencing run 2. The DNA for clonal isolate 2 was made to 26 pM/L. The Ion One Touch 2 (Life Technologies) was used for emulsion PCR and sample recovery. The Ion Torrent ES system (Life Technologies) was used for sample enrichment. For sequencing, the Ion

PGM 400 bp sequencing kit (Life Technologies) and the Ion 314v2 chip (Life Technologies) were used following the manufacturer's protocol.

2.9.3 PacBio RSII sequencing

PacBio RSII sequencer (Pacific Biosciences) was used to sequence *L. fermentum* 3872 clonal isolate 1. This generated long reads used for genome scaffolding. The P6-C4 sequencing kit (Pacific Biosciences) and a single SMRT-Cell chip (Pacific Biosciences) were used. Template preparation and sequencing was conducted by the TGAC genome analysis centre.

2.9.4 Optical mapping

A restriction digest map (optical map) of *L. fermentum* 3872 clonal isolate 1 was created by OpGen (http://www.opgen.com/sequencing/, 2017) using an Argus Optical mapping system with restriction enzyme *Spe*I. The data generated was used for genome assembly validation for *L. fermentum* 3872 and to aid in genome completion.

2.10 Bioinformatics

2.10.1 Genome sequence assemblers used

For *L. fermentum* 3872 genome assembly, the Ion Torrent generated reads were *de novo* assembled using SPAdes V3.5.0 (Bankevich *et al.*, 2012) and MIRA V3.4.2.0 (Chevreux *et al.*, 2004) on the Ion Torrent server. The reads were also *de novo* assembled on the CLC Genomics Workbench (CLC GWB) programme. Default CLC GWB assembler parameters (mismatch cost 2, length fraction 0.5, insertion cost 3, deletion cost 3, and similarity fraction 0.8) were used, the minimum contig size was set to 1 kb. To assess genome assembly quality, read mapping was conducted using CLC GWB with default, stringent and very stringent read mapping parameters. As mentioned by Lehri, Seddon and Karlyshev (2017b), stringent read mapping parameters had a length fraction value of 0.8 and a similarity fraction of 0.9. Very stringent read mapping

parameters used a length fraction of 0.9 and a similarity fraction of 1, with all other parameters kept at default. Mapsolver software was used to align sequence-generated contigs to the *Sph*I restriction digest map produced by OpGen. The software was run using default parameters.

2.10.2 Identifying and assembling the plasmid pLF3872

Contigs generated by MIRA V3.4.2.0 (Chevreux *et al.*, 2004) and CLC GWB assemblers were combined using CISA contig integrator (Lin and Liao, 2013). The quality of the assembly was determined by read mapping using CLC GWB with default parameters. Using the non-redundant NCBI database, the NCBI BLASTN similarity search tool identified that some of the contigs aligned to plasmid sequences plca36 (CP000935.1; *Lactobacillus casei str.* Zhang), pWCFS103 (CR377166.1; *Lactobacillus plantarum* WCFS1) and plasmid 1 (CP002392.1) from *Lactobacillus paracasei subsp*. Reads were mapped onto the closest plasmid sequence plca36 to generate consensus sequences. To circularise the contig, the consensus sequence was merged with the Ion Torrent assembled contig. To confirm that the plasmid sequence was circular, sequence homology between reads from one end of the sequence was found with the other end of the sequence.

2.10.3 Assembly of the chromosomal sequence (misassembled)

The Ion Torrent PGM (Life Technologies) generated sequence read data were *de novo* assembled using CLC GWB, SPAdes V3.5.0 (Bankevich *et al.*, 2012) and MIRA V3.4.2.0 (Chevreux *et al.*, 2004) assemblers. The contigs produced by the sequencing runs were combined using CISA contig integrator (Lin and Liao, 2013). Default read mapping parameters were used to determine sequence quality. Contigs were split if read mapping indicated regions where the nucleotide sequence consisted of unaligned read ends, gaps in read mapping, or low read coverage. The nucleotide sequence ends of each contig were then selected and run using BLASTN and a non-redundant database. If

a match was identified, the matching sequence with the highest hit was extended and added to its corresponding contig. After extension, some contigs could be joined by finding overlaps between the ends of contigs. Each extended/joined contig was verified using read mapping (default parameters). Further contiguation was also conducted by *de novo* assembling unmapped reads (collected after read mapping) and extending/joining contigs. The remaining contigs were verified using default, stringent and very stringent read mapping parameters using CLC GWB. After contiguation, the remaining contigs were aligned to the closest correctly assembled reference sequence *Lactobacillus fermentum* IFO 3956 using Contiguator 2.7.4 (Galardini *et al.*, 2011) to design primers for Sanger sequencing, in order to fill in the remaining gaps. Contiguator orientates and aligns contigs by assuming that even if there are strain-to-strain nucleotide variations, the overall genome sequence organisation is similar to a closely-related reference sequence. BLASTN with a cut-off value of 1e-20; and a .ptt file obtained from the NCBI GenBank website for *L. fermentum* IFO 3956 was used for alignment. The .ptt file improves sequence alignment accuracy by also looking for amino acid homology between sequences.

2.10.4 Correct chromosomal assembly

The genome sequence was completed using a hybrid sequencing approach. This combined long low coverage reads generated by the PacBio RSII (Pacific Biosciences) sequencing machine with short high coverage reads generated by the Ion Torrent PGM (Life Technologies). The size of the long reads, which are on average >14 kb in length, results in less sequencing gaps, although these reads are error-prone due to low read coverage and poor base calling capability (Reuter, Spacek and Snyder, 2015). The Ion Torrent PGM reads are less error-prone because of high read coverage but generate multiple contigs as the machine can only accurately sequence for a length of approximately 400 bp.

As described in our articles (Lehri, Seddon and Karlyshev, 2017a; Lehri, Seddon and Karlyshev, 2017b), sequencing reads generated by the PacBio RSII (Pacific Biosciences) sequencer were assembled using the HGAP3 (hierarchical genome assembly process) pipeline, which utilises the CELERA assembly programme (Myers *et al.*, 2000). Large redundant sequenced regions for the chromosomal sequence and plasmid were determined and removed using Artemis Comparison Tool (ACT) (Carver *et al.*, 2005). To confirm that the chromosomal sequence was circular, homology between reads from one end of the sequence to the other end of the sequence was found. Ion Torrent reads were mapped onto the assembled sequence using default read mapping parameters on CLC GWB to correct nucleotide errors. The assembly was also verified using stringent and very stringent read mapping parameters on the CLC GWB. Reads from clonal isolate 2 from the Ion Torrent sequencing run were mapped onto the complete genome sequence using CLC GWB with default read mapping parameters to determine if there were clonal variations.

2.10.5 Annotation of the genome

As described in Lehri, Seddon and Karlyshev (2017a), once the genome of *L. fermentum* 3872 was complete, it was annotated using RAST (Overbeek *et al.*, 2014), PROKKA (Seemann, 2014), BASys (Van Domselaar *et al.*, 2005), and the NCBI GenBank annotation pipeline (Angiuoli *et al.*, 2008). Annotation irregularities (e.g. truncated coding sequences, or missing gene annotations) between the tools mentioned above were detected and corrected using Geneious software (Kearse *et al.*, 2012).

2.10.6 Genome sequence analysis

The LASTZ programme (Harris RS, 2007), run on Geneious software (Kearse *et al.*, 2012) was used to compare the chromosomal sequence of *L. fermentum* 3872 to that of other circular *L. fermentum* spp genomes, namely *L. fermentum* strains F6, IFO 3956, and CECT 5716. This would allow identification of unique regions and nucleotide

variations between *L. fermentum* 3872 and *L. fermentum* spp used for comparison. Spine/AGEnt Pan-core analysis software with default parameters (Ozer, Allen and Hauser, 2014) was used to identify unique genes of *L. fermentum* 3872 along with genes that are common among *L. fermentum* strains F6, and IFO 3956. PHAST (Zhou *et al.*, 2011) software was used to identify any prophage-related regions of the *L. fermentum* 3872 genome, while BAGEL3 (Heel *et al.*, 2013) was used to identify bacteriocin encoding regions.

2.11 Cloning techniques used for expression of collagen binding protein

2.11.1 Primer design

Primer pairs were designed to PCR amplify the *cbp* gene of plasmid pLF3872 (Table 2.2). Primers were also designed to check for potential PCR errors (Table 2.3). Sigma Genosys synthesised all primers.

To generate sticky ends, *Xba*I and *Sph*I restriction enzyme sites and a Shine-Dalgarno (S-D) sequence were incorporated to the *cbp* sequence. Initially, the *cbp* gene had a C-terminal hexa histidine tag. This was then changed to an N-terminal hexa histidine tag and by the removal of a leader peptide encoding region. Primers were also made to amplify the adhesion domain of the CBP protein (partial CBP), with an N-terminal hexa histidine tag and no leader peptide. To identify and remove the leader peptide sequence, Softberry software ProtCompB (http://www.softberry.com/berry.phtml?topic=pcompb&group=programs&subgroup=proloc, 2017) was used. NCBI CDD was also used to determine the CBP 'A domain', for construction of the partial *cbp* gene sequence. The forward primer used to construct the partial *cbp* gene sequence was the same as that used for the construction of the full *cbp* gene sequence, labelled CBPFOR2 in Table 2.2.

Table 2.2. Primers used for cloning experiment.

Primer name	**Primer description**	**primer sequence (5'-3')**
CBPREV1	CBP reverse primer (C-terminal histidine tag)	ATGAGCATGCTCAATGGTGATGGTGATGGTGAATAGTAAATCTACTTATAACTACTAAACC
CBPFOR1	CBP forward primer (C-terminal histidine tag)	AACAATCTAGAAGAAGGAGGCAACAGTATGAGTAATAAATCTAGGCATCGTCGAC
	Primer Name	**primer sequence (5'-3')**
CBPREV2	CBP reverse primer (N-terminal histidine tag)	ATGAGCATGCTCAAATAGTAAATCTACTTATAACTACTAAACC
CBPFOR2	CBP forward primer (N-terminal histidine tag)	ATATGCTTCTAGAAGAAGGAGGCAACAGTATGCACCATCACCATCACCATGATAGCAAGACAAATATTACTCAGAACGGTACG
CBPREV3	Partial CBP reverse primer (N- terminal histidine tag)	CTTTGCATGCTCAATTTATAGTTAAACTATGATTATATCCATTAG

Table 2.3. Primers used for recombinant *cbp* gene sequence nucleotide validation.

Primer Name	**primer sequence (5'-3')**
pBADFor	GATTAGCGGATCCTACCTGACGC
pBADRev	GCTTCTGCGTTCTGATTTAATC
CBPseq1	CTCAGTTCCAGATTCAGCCA
CBPseq2	CTGAGGGTTATGCAATTGTCTAC
CBPseq3	GGCATCAGACAACTGGAAGTAC

2.11.2 Preparing primers for PCR

Lyophilised primers were spun down and then reconstituted in Milli-Q water to make 100 µM stock. Working primer solutions were diluted to 10 µM in Milli-Q water. All primers were stored at -20°C.

2.11.3 PCR amplification

For cloning experiments, PCR was conducted using Q5 High-Fidelity DNA polymerase (NEB). The reaction was prepared as described in Table 2.4. The dNTPs were purchased from Fisher. Thermocycler conditions are described in Table 2.5.

Table 2.4. PCR reaction mixture for *cbp* gene amplification

5x Q5 reaction buffer	10 µl
10 mM dNTPs	1 µl
10 µM forward primer	1 µl
10 µM reverse primer	1 µl
DNA	1 µl
Q5 High-Fidelity DNA polymerase	0.5 µl
Milli-Q water	35.5 µl
Total	50 µl

Table 2.5. PCR thermocycler conditions for *cbp* gene amplification.

	Temp	**Time**
Initial denaturation	98°C	30 seconds
20 cycles	98°C 55°C 72°C	10 seconds 30 seconds 4 minutes
Final extension	72°C	2 minutes

PCR products were analysed and sample quality was determined by gel electrophoresis. For cloning experiments, PCR products were purified using the QIAquick PCR purification kit (Qiagen) following standard manufacturer instructions. The sample was eluted using 30 µl of elution buffer.

2.11.4 Restriction analysis

Sticky ends were generated using restriction enzymes *Xba*I and *Sph*I on both the insert DNA and pBAD33 vector with NEB buffer 4 diluted to 1x. The enzymes were

incubated for one hour at 37°C, and then heat inactivated at 65°C for 20 minutes. The reaction mixture is described in Table 2.6.

Table 2.6. Reaction mixture for restriction digestion of insert and vector DNA used for cloning experiments.

	Insert DNA	**Vector DNA**
NEB buffer 4 10x	5 µl	2.5 µl
***Xba*I**	1 µl	1 µl
***Sph*I**	1 µl	1 µl
Sample DNA	20 µl	10 µl
dH_2O	23 µl	10.5 µl
Total	50 µl	25 µl

2.11.5 Gel extraction

Gel extraction was conducted using the Invitrogen Purelink Quick Gel extraction kit. Restriction digested samples were run on a 1% agarose gel to separate and remove unwanted oligonucleotide fragments. A scalpel was used to excise a band from the gel using long UV wavelength on G: BOX (Syngene). To determine the appropriate amount of gel solubilisation buffer, the cut gel sample was weighed in a Lo-Bind 1.5 ml Eppendorf. The rest of the protocol was followed in accordance with the manufacturer's guidelines. After gel extraction, sample concentration was determined using a NanoVue spectrophotometer (GE Healthcare Life Sciences).

2.11.6 Ligation and transformation

Ligation of vector and insert was conducted using the NEB quick ligation kit following the manufacturer's protocol.

NEB express *E. coli* was transformed with the ligation mixture. Standard manufacturer protocol for transformation was followed. As per the manufacturer’s protocol, after heat shock, SOC medium (NEB) was added to the mixture of vector and competent *E. coli*.

The mixture was then diluted 10x with SOC medium (NEB). Fifty microliters of both the diluted and undiluted mixture were grown overnight on separate 25 µg/ml CAM supplemented LB agar (Fisher) plates. After growth, single colonies were selected and grown to confluence for plasmid extraction and analysis.

2.11.7 Analysis of transformants

To confirm the presence of the *cbp* gene on vector pBAD33. The plasmid DNA was extracted using QIAprep spin miniprep kit (Qiagen) and a restriction digest was conducted by mixing 12 µl NEB buffer 4 diluted to 1x with 1 µl of *Xba*I and *Sph*I, along with 1 µl of plasmid DNA. The samples were incubated for 30 minutes at 37°C and then run on an 1% agarose gel.

2.11.8 Validation of the recombinant *cbp* gene

The recombinant *cbp* gene sequence was validated using Sanger sequencing performed by GENEWIZ. The primers used are shown in Table 2.3. Chromas software (Technelysium) was used to determine the quality of the Sanger sequencing data, after which the Sanger sequenced data was compared to that generated by high-throughput sequencing.

2.12 CBP protein expression

Transformed NEB express *E. coli* was grown overnight at 37°C on LB agar (Fisher), supplemented with 25 µg/ml CAM. The *E. coli* was suspended in 10 ml LB broth (Fisher) supplemented with 25 µg/ml CAM and grown overnight in a shaker at 120 rpm and 37°C. The 10 ml culture was transferred into 250 ml of LB broth (Fisher) and allowed to reach an OD_{600} of 0.6 at 37°C on a shaker at 120 rpm. The sample was induced with 0.1% L-arabinose (Acros Organics) for three hours. Before and after induction, OD_{600} measurements were taken along with one millilitre aliquots of *E. coli*. The aliquots were spun down using an Eppendorf mini spin plus micro-centrifuge at

14,104 x g for one minute and then the supernatant was removed. NuPAGE LDS buffer x1 (Invitrogen) was added to the sample for lysis and SDS-PAGE analysis. The stock samples were spun down using an Eppendorf centrifuge 5810R at 3220 x g for 10 minutes at 4°C to remove the LB broth. The spun down sample was either stored at -80°C or protein purification was conducted immediately. Initially, a small-scale protein purification using the MagneHis protein purification kit (Promega) was conducted following manufacturer's guidelines. This was preceded by large-scale protein purification. For large scale purification, NEB express *E. coli* was initially lysed in 5 ml of a mixture of 1 mg/ml egg white lysozyme (Sigma), protease inhibitor cocktail (Sigma) of which 1 ml was added for every 20 g of *E. coli*, and 3 µl DNAse I (Promega) in wash buffer (QIAGEN fast start kit) for 30 minutes at room temperature. Ultimately, for large scale protein purification, 5 ml of 10x Promega fastbreak cell lysis reagent supplemented with 3 µl DNAse I was added to the induced *E. coli* as the initial custom lysis buffer resulted in degraded protein. For every 10 ml of *E. coli* suspension, 1 ml of 10x Promega fastbreak cell lysis reagent was added, with 20 minutes' incubation at room temperature.

Protein purification was conducted using His60 Ni Superflow gravity column (Clonetech). The column was equilibrated at room temperature and then washed with 2.5 ml wash solution (Qiagen). Clarified lysate was added to the column and inverted three times. The lysate was incubated in the column for five minutes at room temperature. The column was then washed twice with 5 ml of wash solution (Qiagen). The protein was eluted using 3 ml of elution buffer (Qiagen), collecting in 1 ml fractions.

2.13 Protein detection

2.13.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

NuPAGE LDS 1x sample buffer (Invitrogen) was added to samples and heated for 10 minutes at 70°C. The samples were then loaded on to NuPAGE Novex 4%–12% Bis-tris gel (Invitrogen). The gel was run at 150 V for one hour in NuPAGE MOPS SDS running buffer 1x (Invitrogen). An X-cell sure-lock mini gel electrophoresis unit (Invitrogen) was used for running the gel. The molecular marker used was PageRuler Plus Prestained protein ladder (FisherScientific).

2.13.2 Coomassie staining

Simply-blue safe stain (Invitrogen) was used after running SDS-PAGE for Coomassie staining following standard manufacturer's guidelines. Gel images were taken using G: Box (Syngene) with the GenSnap (Syngene) software.

2.13.3 Western blot

After running an SDS-PAGE, the NuPAGE Novex 4%-12% Bis-tris gel (Invitrogen) was placed in transfer buffer (see 2.4.2 for recipe) for 15 minutes, the amount added was enough to cover the gel. An Immobilon –P 0.45 µM polyvinylidene difluoride membrane (PVDF; Millipore) used for blotting was activated in 100% methanol for 15 seconds and washed twice for two minutes in water (ELGA DV35). The membrane was placed in transfer buffer for five minutes. Six western blot filter papers (ThermoFisher Scientific) were incubated in transfer buffer for 15 minutes. A semi-dry transfer unit (Sigma) was used at 15 V for 90 minutes to blot the membrane with protein(s). Once the transfer was complete, the membrane was washed twice for 10 minutes in 15 ml TBS buffer diluted to 1x (see 2.4.2 for recipe). The membrane was then blocked using 3% BSA (Sigma) in 20 ml 1x TBS and incubated for one hour at room temperature. The membrane was washed twice in 15 ml TBST buffer (see section 2.4.2 for recipe), and then once in 15 ml TBS 1x for 10 minutes. Then Penta-His primary antibody (Qiagen)

(0.2mg/ml) was diluted to 1:1000 in TBS 1x and 3% BSA (Sigma). The membrane was incubated with 10 ml of the primary antibody solution for one hour at room temperature. After incubation, the membrane was washed twice in 15 ml TBST followed by a single 15 ml TBS 1x wash for 10 minutes. The membrane was incubated with Anti-mouse IgG, HRP-linked secondary antibody (cell signalling technologies) diluted to 1:1000 in 10 ml TBS 1x and 10% non-fat dried milk powder (Marvel original). The membrane was incubated with secondary antibody for one hour at room temperature. Once incubated, the membrane was washed four times with 15 ml TBST for 10 minutes each. SuperSignal West Pico Chemiluminescent substrate (ThermoFisher Scientific) was used for detection of protein following manufacturer's guidelines. Syngene G: Box was used for detection of protein(s) using GenSnap (Syngene) software.

2.13.4 Silver staining

Silver staining was used to visualise protein samples obtained from Co-Immunoprecipitation (Co-IP) experiments. Silver staining is a highly sensitive visualisation procedure that allows for the detection of proteins of less than 0.25 ng. The sample was initially prepared by adding 7.5 µl of sample to 2.5 µl of 4x NuPAGE LDS sample buffer (Invitrogen), and then heated for 10 minutes at 70°C. Ten microliters of the sample were loaded onto a NuPAGE Novex 4%–12% Bis-tris gel (Invitrogen). The gel was run following the standard SDS-PAGE protocol described above. Samples were stained using a Pierce silver stain kit for mass spectrometry (ThermoFisher Scientific) following the manufacturer's protocol. The gel was visualised using a Syngene G: Box (Syngene) using GenSnap (Syngene) software.

2.14 BCA assay

Protein concentration was determined using a Pierce BCA protein assay kit (ThermoFisher Scientific) following the manufacturer's guidelines. Bovine serum

albumin (BSA) provided by the kit was diluted to varying concentrations and used as a protein standard for determining concentration. A line of best fit of the absorbance generated by the BSA against the concentration of BSA used was made using a Microsoft Excel spreadsheet. The absorbance of the query protein was then used against the line of best fit to determine the concentration of the query protein. The absorbance was read at 562 nm using a Tecan Infinite M200 Pro microplate reader.

2.15 Enzyme-linked immuno absorbance assay experiments

For all ELISA experiments, 96-well plates were coated with 100 µl (0.36 µg/well) of collagen I (Sigma) at 4°C overnight. The wells were washed twice with 200 µl PBST (section 2.4.3 for recipe) and then blocked with 200 µl 2% BSA (Sigma) in PBS for one hour at room temperature. The wells were washed three times with 200 µl PBST, and then 100 µl of samples (CBP, bacteria or PBS) were incubated on the plates at 37°C for one hour in an incubator. If *C. jejuni* was used a microaerobic incubator (Don Whitley) was used. The wells were washed four times with 200 µl PBST and then incubated with 100 µl primary antibody diluted to 1:1000 in PBS with 0.05% Tween 20 (Sigma) and 1 mg/ml BSA (Sigma), see Table 2.7 for antibodies used. The plate was incubated at 37°C and washed four times with 200 µl PBST. This was followed by incubation with 100 µl secondary antibody diluted to a ratio of 1:1000 in PBS with 0.05% Tween 20 (Sigma) and 3% BSA (Sigma) for one hour at 37°C, see Table 2.7 for antibodies used. The wells were washed another four times using 200 µl PBST, followed by incubation with 100 µl of 3,3′,5,5′-Tetramethylbenzidine substrate (TMB) at room temperature for 15 minutes. The reaction was stopped with 1M H_2SO_4 and the optical density was read at 450 nm using a Tecan Infinite M200 Pro microplate reader. After each washing step, the wells were tapped gently three times on tissue paper.

Table 2.7. Antibodies used for each ELISA experiment.

ELISA experiments	**Primary antibody**	**Secondary antibody**
CBP detection	Thermo Scientific Pierce 6x-His Epitope Tag monoclonal primary antibody (MA1-21315)	Cell Signaling Technology anti-mouse IgG, HRP-linked secondary polyclonal antibody (#7075)
***C. jejuni* binding assay, and competition between *C. jejuni* and CBP in binding to collagen I**	Bio-Rad; *Campylobacter jejuni* monoclonal primary antibody (BGN/2E10)	Bio-Rad; Goat anti-mouse polyclonal secondary antibody (STAR87P)
Whole cell competition between *C. jejuni* and *L. fermentum* 3872 in binding to collagen I	Antibodies-online; anti-*Campylobacter jejuni* (PEB1) polyclonal primary antibody (ABIN488113)	SAB; Goat-anti-rabbit IgG polyclonal HRP conjugated secondary antibody (L3012)

2.15.1 Sample preparation for ELISA experiment

For ELISA experiments, bacteria were suspended in PBS. The amount of bacteria added to each well was determined by the optical density at 600 nm, measured using a cell density meter (WPA Biowave-CO800).

2.15.2 Choosing optimal ELISA plates

ELISA experiments were initially conducted to determine the ideal type of plates and BSA to use for future experiments. The plates tested were UV-treated Nunclon Delta (ThermoFisher Scientific), UV-treated High Bind Corning Costar and Untreated Med Bind Corning Costar plates. Standard ELISA protocol was followed, as described in section 2.15. CBP (0.36 µg/well) was prepared in PBS for the sample incubation step of the ELISA experiment. BSA from Sigma (99% purity) and Acros Organics (western blot grade) were coated on wells and used as negative controls. For all future ELISA experiments, untreated Med Bind Corning Costar plates and BSA (Sigma) were used.

2.15.3 Choosing buffers for coating

To determine the optimal type of buffer for coating Collagen I (Sigma) onto 96-well plates for ELISA experiments, coating was performed under acidic and basic conditions. To achieve a pH of 3.6, collagen I (Sigma) was made to 3.6 µg/ml in

distilled Milli-Q water. To achieve a pH of 9.6, collagen I (Sigma) was made to 3.6 µg/ml in ELISA coating buffer (see section 2.4.3 for recipe). One hundred microliters of the samples were added to each well. For the experiment, standard ELISA protocol was followed. For the incubation step, CBP 0.05, 0.2, and 0.36 µg/wells were prepared in PBS. BSA (Sigma) coated wells were used as negative controls. For all future ELISA experiments, the ELISA coating buffer (pH 9.6) was used.

2.15.4 CBP binding assay

ELISA experiments were conducted to determine whether CBP would bind to collagen I. Wells were coated with collagen I and BSA (negative control) using 0.36 µg/well prepared in PBS. ELISA protocol as described in section 2.15 was followed. CBP in PBS of 0.05, 0.2, and 0.36 µg/wells were used for the incubation step.

2.15.5 *C. jejuni* 11168H and 81-176 binding assay

To determine whether *C. jejuni* binds to collagen I, *C. jejuni* suspensions were made to an OD_{600} of 1, 0.1 and 0.01 in PBS, of which 100 µl were added to each well. The final amount of *C. jejuni* bacterial suspension added to each well was 2×10^8 CFU/well, 2×10^7 CFU/well and 2×10^6 CFU/well. Negative controls used BSA (Sigma) coated wells incubated with *C. jejuni* strains 11168H and 81-176.

2.15.6 Determining whether CBP and *C. jejuni* strains 11168H and 81-176 compete with each other in binding to collagen I

To discover whether CBP inhibited *C. jejuni* 11168H and 81-176 binding to collagen I, 0.36 or 2 µg/well of CBP was mixed with either *C. jejuni* 11168H or 81-176 strains at 2×10^7 CFU/well. Controls used for the experiment included collagen I or BSA (Sigma) coated wells with CBP, *C. jejuni* or PBS incubated to each well.

2.15.7 Whole cell competition between *L. fermentum* 3872 and *C. jejuni* strains 11168H and 81-176

To determine whether *L. fermentum* 3872 inhibits *C. jejuni* attachment to collagen I, *L. fermentum* 3872 suspension was made to an OD_{600} of 0.5 (1 x 10^8 CFU/ml), 1 (2 x 10^8 CFU/ml), 5 (1 x 10^9 CFU/ml), and 9 (2 x 10^9 CFU/ml) mixed with *C. jejuni* strains of OD_{600} 0.1 (2 x 10^8 CFU/ml) in PBS. One hundred microliters of the samples were added to wells with a ratio of *L. fermentum* 3872 to *C. jejuni* of 1:2, 1:1, 5:1 and 10:1. The controls used were *L. fermentum* 3872, *C. jejuni*, and PBS incubated in collagen I coated wells where appropriate.

2.16 Agar well diffusion assay

An agar well diffusion assay was conducted to find out if *L. fermentum* 3872 cell-free supernatant inhibited the growth of *C. jejuni* 11168H and 81-176 strains. Following the method described by Nishiyama *et al.* (2014). One-day *C. jejuni* cultures were made to an OD_{600} of 1 in PBS and 300 µl of the suspension was added on to 15 ml soft MH agar (0.75%) at 41°C. The agar was poured on top of 20 ml solidified MH agar. Overnight cultures of *L. fermentum* 3872 were grown in 10 ml MRS broth in a microaerobic incubator at 37 °C. The *L. fermentum* 3872 culture was then filter sterilised using a 0.22 µm filter (Fisher). Four 10 mm wells were made on the agar plate. The MH agar wells were filled with 300 µl of MRS broth (pH 6.3) used as a control, *L. fermentum* 3872 supernatant (pH 4.2), pH adjusted *L. fermentum* 3872 supernatant adjusted by 1M NaOH (pH 6.3), or boiled *L. fermentum* 3872 supernatant at 100°C for five minutes (pH 4.2).

2.17 Detection of *C. jejuni* proteins involved in collagen I adhesion

2.17.1 Co-immunoprecipitation assay

A Dynabeads Co-Immunoprecipitation Kit (Co-IP; Life Technologies) was used to investigate any collagen I interacting proteins expressed by *C. jejuni* strains 11168H and

81-176. Each mg of beads was coated with 15 µg of collagen I (Sigma). To couple collagen I to the beads, they were incubated for 24 hours on a rotator at 37°C. The rest of the protocol was followed as recommended by the manufacturer.

To prepare bacterial lysates for incubation with collagen I-coated beads, 20 ml suspensions of OD_{600} 1 in PBS was made for *C. jejuni* strains 11168H and 81-176. To remove the supernatant, the suspensions were spun down in an Eppendorf centrifuge 5810R at 3220 x g for 10 minutes at 4°C. IP buffer 4x supplied with the Dynabeads Co-IP Kit was diluted to 1x in Milli-Q water, to which 100 mM molecular biology grade NaCl (Sigma) and 5 µl of DNAse 1 (Promega) were added to make a lysis buffer. A ratio of cell weight to lysis buffer volume of 1:9 was mixed as recommended by the manufacturer and incubated on ice for 10 minutes and then sonicated (Soniprep 159) on ice with 10-second bursts of sonication and 30-second rest cycles for 10 cycles. The clarified lysate was spun down at 4°C at 3220 x g using an Eppendorf centrifuge 5810R for five minutes. For each experiment, the clarified lysate was mixed with 2 mg of collagen I-coupled beads. The clarified lysates of *C. jejuni* 11168H and 81-176 were incubated for one or three hours at room temperature at 30 rpm in a rotator (ThermoFisher Scientific). The rest of the protocol was completed as recommended by the manufacturer.

2.17.2 Mass spectrometry analysis

Mass spectrometry on Co-IP eluates for *C. jejuni* strains 11168H and 81-176 was performed by the Cambridge Centre for Proteomics Core Service.

SDS-PAGE was run using MOPS SDS running buffer 1x prepared with Milli-Q water. Initially, staining was conducted using Coomassie, although the approach was unsuccessful. Therefore, a silver stain was run using the Pierce Silver Stain kit for mass spectrometry (ThermoFisher Scientific) following the manufacturer's protocol. All

buffer exchanges and gel handling were conducted with fresh gloves, and under a laminar flow hood to avoid contamination. Bands observed for *C. jejuni* eluates were cut using sterile scalpels and placed into Lo-bind Eppendorf containing 50 µl of Milli-Q water and then sent for mass spectrometry analysis. The instrument used was the LC MS/MS ESI-ORBITRAP-HCD.

To recognise and assemble protein sequences, the data output was run through MASCOT software. The data was run against protein fasta files of *C. jejuni* NCTC 11168 and 81-176, retrieved from the NCBI GenBank database.

2.18 Statistical analysis

Standard deviation was calculated using one way analysis of variance (ANOVA). For ELISA experiments, the relevant negative control values were subtracted from the test values. Each of the error bars indicated on graphs for ELISA experiments represent standard error of the mean (SEM). The p values on the graphs for ELISA experiments are labelled as: * for $0.005<p\leq0.05$, ** for $0.001<p\leq0.005$ and *** for $p\leq0.001$. For agar well diffusion assay standard deviation (SD) is shown.

Chapter 3: Results

3.1 Sequencing the genome of *L. fermentum* 3872

3.1.1 Ion Torrent PGM sequencing

Initially, the genome of *L. fermentum* 3872 was sequenced using the Ion Torrent PGM sequencer. The first step of the sequencing process was to isolate DNA. Two clonal isolates of *L. fermentum* 3872 were grown to confluence and sequenced. One of these was used for genome assembly and the other was used to determine sequence variations between the clonal isolates of *L. fermentum* 3872.

3.1.2 DNA extraction for Ion Torrent PGM sequencing

Once DNA had been extracted from the *L. fermentum* 3872 clonal isolates 1 and 2, the samples were run on an agarose gel to obtain a primary yield estimate, as well as to see if there was any rRNA contamination. The above mentioned cases may affect the quality of base calling, which in turn would affect genome assembly. The concentration and purity of the DNA samples were further validated using a NanoVue spectrophotometer.

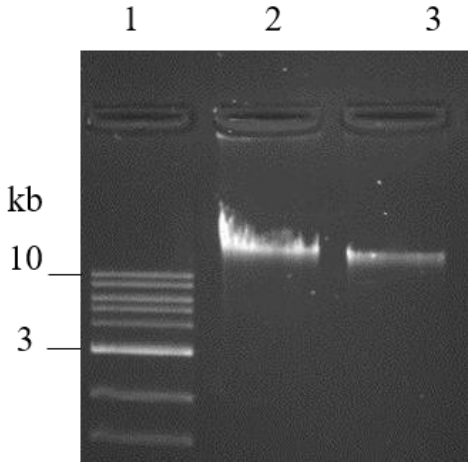


Figure 3.1. Agarose gel electrophoresis image of purified DNA from *L. fermentum* 3872 used for Ion Torrent PGM sequencing;1, 2-log DNA ladder (NEB); 2, DNA clonal isolate 2; 3, DNA clonal isolate 1.

Table 3.1. NanoVue spectrophotometer readings for DNA isolated from *L. fermentum* 3872 clonal isolate 1 and 2.

Clonal isolate 1		**Clonal isolate 2**	
concentration:	322.8 µg/ul	**concentration:**	347 µg/ul
A260/A280	1.838	**A260/A280**	1.922
A260/A230	1.528	**A260/A230**	1.881

DNA extracted from clonal isolates 1 and 2 were of high concentration and quality (Figure 3.1 and Table 3.1). Both samples had an A260/A280 reading of above 1.8 (Table 3.1). The A260/A230 value of clonal isolate 1 was 1.528, which is lower than the standard 1.8–2.0 (Table 3.1). The lower A260/A230 value was likely caused by reagents used during the DNA extraction procedure, such as ethanol, which would not influence DNA sequencing as the minute quantities of reagents would be removed during the various sample preparation steps.

3.1.3 DNA size selection for Ion Torrent library preparation

After determining the quality, concentration and purity of the DNA for *L. fermentum* 3872, the NEBNext fast DNA fragmentation and library preparation kit for Ion Torrent was used to prepare the DNA samples for sequencing (see section 2.9). The purified DNA samples of *L. fermentum* 3873 were first enzymatically fragmented, followed by sequence end repair and nick translation. The 5’ prime ends of the fragmented DNA were then phosphorylated for blunt end ligation of adapters. PCR of the adapter ligated DNA was conducted to amplify the DNA, which increases sample concentration and selects fragments with correctly attached adapters used for Ion Torrent sequencing. The samples were then size selected using a 2% E-Gel to select the manufacturer recommended DNA size for the 400 bp Ion Torrent sequencing kit. The recommended size is around 490 bp as both adapters attached to the ends of the fragmented DNA are 45 bp in length.

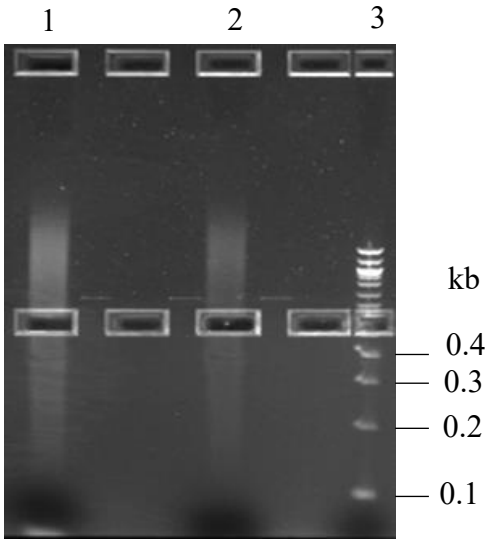


Figure 3.2. E-Gel cassette used for DNA size selection. The image was taken after extracting size selected DNA. The wells are: 1, fragmented DNA from clonal isolate 1; 2, fragmented DNA from clonal isolate 2; 3, 2-log DNA ladder (NEB).

The DNA samples have been sheared, as can be seen by the smearing on the E-Gel cassette image (Figure 3.2). The lower wells of the 2-log DNA marker are around 490 bp. Thus, fragmented DNA of clonal isolates 1 and 2 were extracted at approximately 490bp.

3.1.4 Bioanalyzer result

To accurately determine DNA fragment size and concentration, a Bioanalyzer 2100 was used after DNA size selection. Accurately determining the concentration of DNA is important in order to have the correct Ion Sphere particle (ISP) to DNA ratio. This would help clonality of DNA after emulsion PCR. The analysis was conducted in duplicate and the average concentration of the sample was used for template preparation, which would help account for pipetting errors.

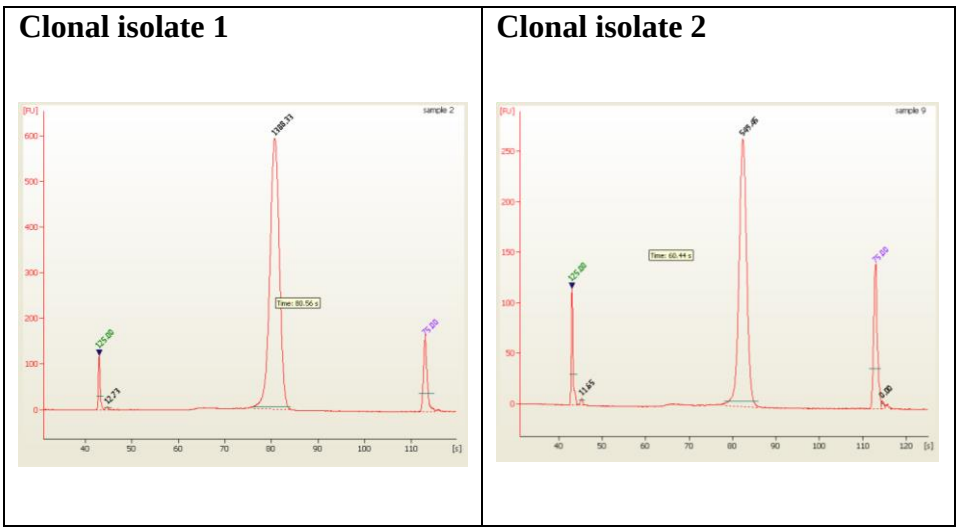


Figure 3.3. Graph showing Bioanalyzer 2100 result of size selected *L. fermentum* 3872 DNA of clonal isolates 1 and 2.

Table 3.2. Bioanalyzer 2100 result of size selected DNA of *L. fermentum* 3872 clonal isolates 1 and 2.

	Clonal isolate 1		**Clonal isolate 2**	
	Size [bp]	Molarity [pmol/l]	Size [bp]	Molarity [pmol/l]
Reading 1	458	4,592.30	488	1,705.80
Reading 2	459	4,130.30	495	1,416.30

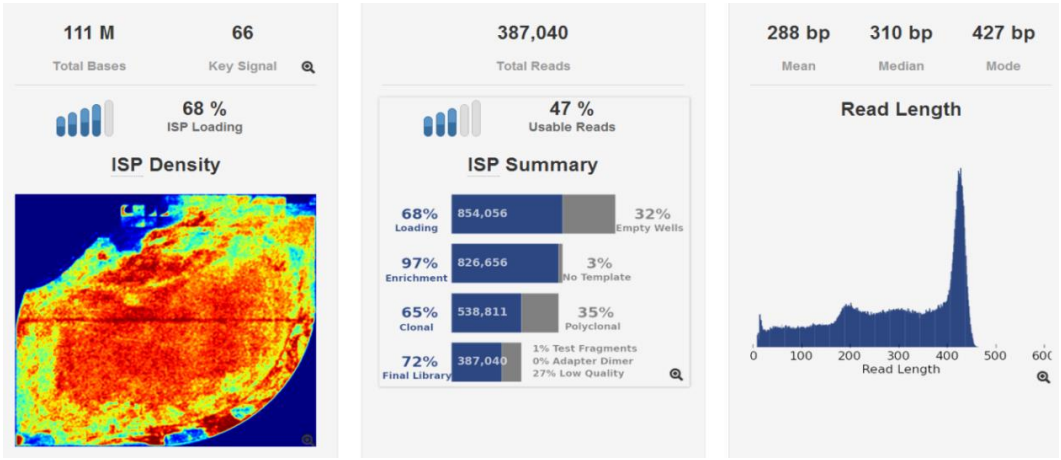
The molarity for both size selected DNA samples was high enough for template preparation (Table 3.2). The sizes of the fragmented DNA samples were smaller than 490 bp, but were acceptable for the Ion Torrent 400 bp sequencing kit (Table 3.2 and Figure 3.3).

3.1.5 Quality of Ion Torrent PGM sequencing

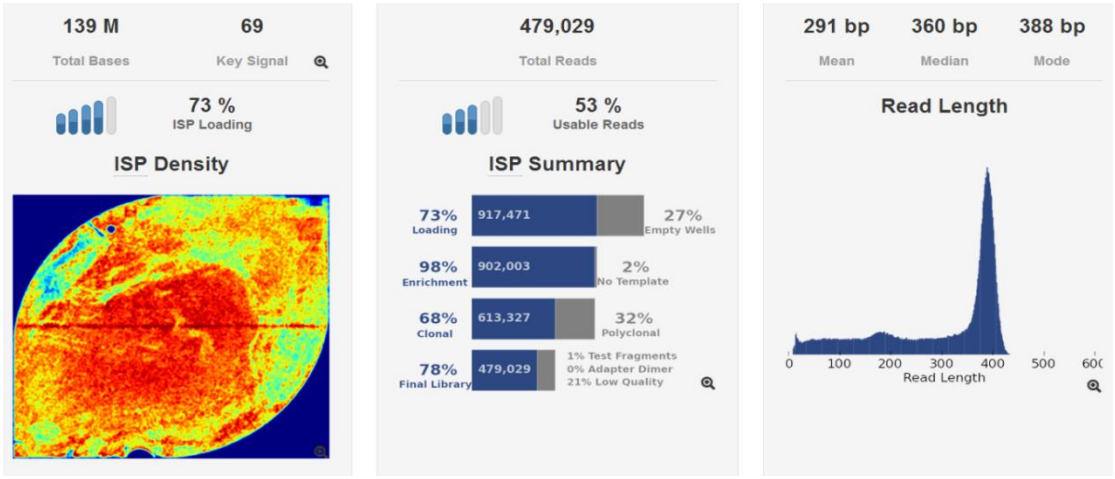
Once the fragment size and concentration were determined using the Bioanalyzer 2100 equipment, emulsion PCR was conducted. During emulsion PCR, adapter-ligated DNA were denatured and attached to complementary sequences located on ISPs. The DNA attached to each ISP was amplified. Once emulsion PCR was completed, enrichment of the samples was conducted to remove beads without amplified DNA, followed by

loading of the samples on to an Ion 314v2 chip for sequencing. Two sequencing runs were conducted on clonal isolate 1 to provide sufficient data necessary for genome assembly. One sequencing run was carried out for clonal isolate 2 for sequence validation.

(A) Sequencing run 1 (clonal isolate 1)



(B) Sequencing run 2 (clonal isolate 2)



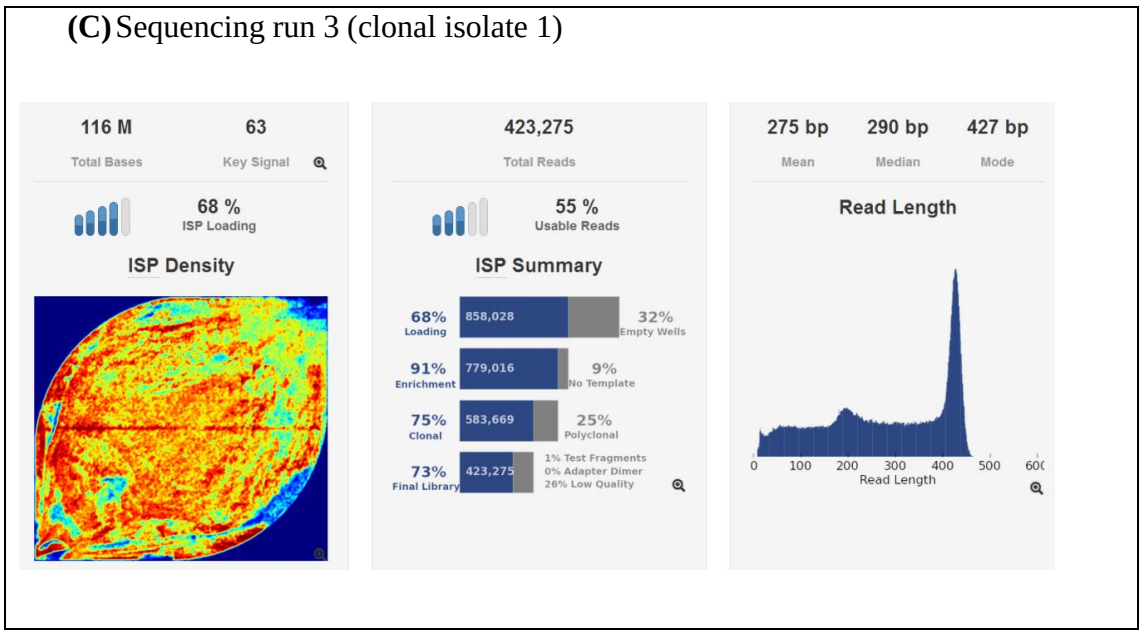


Figure 3.4. Summary of the Ion Torrent PGM sequencing output for *L. fermentum* 3872 sequencing runs 1 (A), 2 (B), and 3 (C).

The ISP loading indicates the number of wells on the Ion 314v2 chip loaded with samples. For sequencing run 1, the ISP loading was 68%. The low ISP loading would not affect the quality of sequenced reads, and as sequencing was performed on the samples multiple times, the final coverage would still be high. For sequencing run 1, the enrichment was good at 97%. Thus, many DNA fragments with correctly primed adapters were sequenced. The mean was 288 bp, less than the desired read length of 400 bp, with the most common read length being 427 bp. Sequencing run 2 had better ISP loading at 73%, with an enrichment of 98%. The mode read length was 388 bp with a mean of 291 bp, both being smaller than the preferred read length of 400 bp. For sequencing run 3, the ISP loading was 68%, and 91% of the reads were enriched. The number of polyclonals was reduced considerably to 25% when compared to sequencing runs 1 and 2 due to a reduction in the concentration of DNA used during the template preparation step. For sequencing run 3, the mean read length was 275 bp and the mode was 427 bp. Overall, sequence runs 1, 2 and 3 generated 47%, 53% and 55% useable data. Although the data generated may not have the desired read length it is still sufficient for genome assembly as the read lengths are longer than that generated by the

Ion Torrent 200 bp kit (Life Technologies) also used for genome assembly projects. Data from sequence run 2 with DNA from clonal isolate 1 were also sufficient for validating sequence variations between DNA from the two clonal isolates.

3.2 Assembling the genome of *L. fermentum* 3872

3.2.1 *L. fermentum* 3872 genome sequence assembly using Ion Torrent PGM generated reads

The plasmid pLF3872 sequence was circularised using the Ion Torrent PGM generated reads, which were *de novo* assembled using MIRA V3.4.2.0 (Chevreux *et al.*, 2004) and CLC GWB, as described in section 2.10.2.

Table 3.3. The data output of CLC, MIRA and SPAdes *de novo* assemblers for the three Ion Torrent PGM sequencing runs, along with CISA contig integrator and HGAP3 data output of PacBio RSII sequencer. The table was obtained from Lehri, Seddon and Karlyshev, (2017b).

	Ion Torrent										**PacBio RSII**
	Run 1			**Run 2**			**Run 3**				
Assembly tool	**CLC**	**MIRA**	**SPAdes**	**CLC**	**MIRA**	**SPAdes**	**CLC**	**MIRA**	**SPAdes**	**CISA**	**HGAP3**
Number of contigs	162	162	288	152	178	291	180	139	288	43	7
Total assembly (bp)	2103399	2321294	2185334	2106119	2313595	2184846	2079293	2328129	2184268	2497362	2430608
Largest contig (bp)	89244	153205	123470	89537	124270	123504	55652	126522	123449	152374	2302236

Prior to using a combination of PacBio RSII and Ion Torrent PGM generated reads, an initial attempt was made in only using Ion Torrent reads to complete the chromosomal sequence of *L. fermentum* 3872. As described in section 2.10.3 and article by Lehri, Seddon and Karlyshev (2017b), each of the Ion Torrent sequenced reads were *de novo* assembled using SPAdes V3.5.0 (Bankevich *et al.*, 2012), MIRA V3.4.2.0 (Chevreux *et al.*, 2004) and CLC GWB. The contigs generated were merged using CISA contig integrator (Lin and Liao, 2013). After the merger, the number of contigs indicated in

Table 3.3 was reduced to 43. Read mapping using default parameters showed that 88.16% of nucleotides mapped onto the sequence. The total number of contigs was increased to 48 by identifying and splitting contigs where read mapping did not confer with the contig assembly. The latter was done by looking for regions on a contig that had unaligned read sequencing ends, low read coverage, and regions where reads did not map onto a contig. After further contiguation using BLASTN followed by unmapped read extension, the total number of contigs was reduced to 18. The estimated genome size of *L. fermentum* 3872 had also increased to 2.5 Mb, which was larger than the genome size estimated by the *de novo* assemblers initially used for the Ion Torrent reads (Table 3.3, on row 'Total assembly (bp)'). Read mapping using default parameters showed an improvement in read map coverage with 99.65% of nucleotides mapped onto the assembly, as opposed to 88.16%, indicating a near complete genome.

The contigs were aligned to a closely related reference sequence *L. fermentum* IFO 3956 using Contiguator (Galardini *et al.*, 2011) to generate primers for Sanger sequencing.

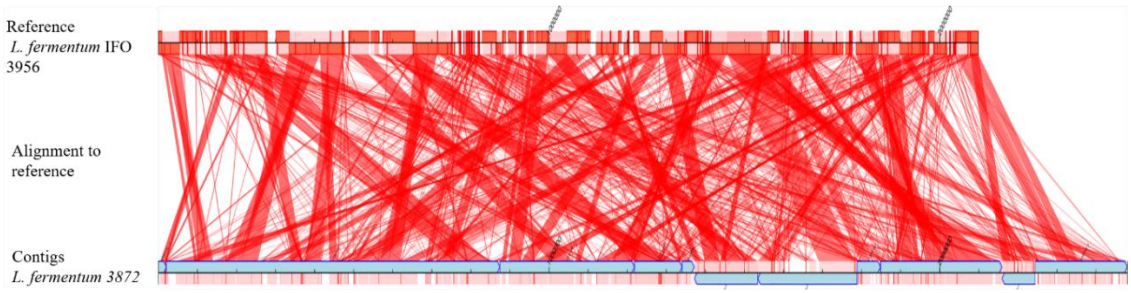


Figure 3.5. Contiguator alignment result for the 18 contigs of *L. fermentum* 3872 mapped to reference sequence *L. fermentum* IFO 3956. BLASTN was used with default parameters. Only contigs that aligned to the reference sequence in blue are shown. The red lines indicate matching alignments between the latter two strains. The diagram was obtained from Lehri, Seddon and Karlyshev (2017b).

Figure 3.5 shows that regions of each contig from *L. fermentum* 3872 align to multiple areas of the reference strain *L. fermentum* IFO 3956. Thus, there may either be strain to strain variation between *L. fermentum* 3872 and *L. fermentum* IFO 3956, or the *L. fermentum* 3872 contigs are misassembled. Misassembly was unlikely as the reads mapped well to the chromosomal sequence with no signs of unaligned read ends, gaps,

and low coverage when using default, stringent and very stringent read mapping parameters. Nonetheless, an optical map of the chromosomal sequence for *L. fermentum* 3872 was generated to confirm genome assembly quality.

3.2.2 Optical map data to determine initial assembly quality

An optical map uses a restriction enzyme to cleave DNA at specific sites. A spectrum of the cleavage sites is used to generate a genetic fingerprint of the genome. The genome sequence is then mapped onto the restriction enzyme map (optical map). The method is suitable for deducing assembly quality of the nucleotide sequence without being dependent on traditional quality indicators that require nucleotide sequence data such as read mapping.

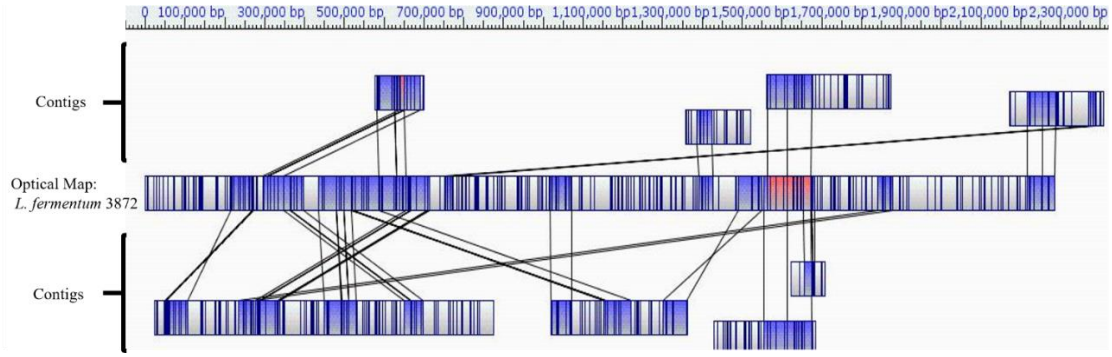


Figure 3.6. Alignment result of the 18 contigs of *L. fermentum* 3872 to the restriction map of *L. fermentum* 3872. The blue regions on the optical map indicate areas of each contig that are correctly assembled, while the white regions indicate areas that are misassembled. The red regions are areas that are duplicated. The diagram was obtained from Lehri, Seddon and Karlyshev (2017b).

The 18 assembled contigs of *L. fermentum* 3872 were aligned to the optical map generated by OpGen using Mapsolver software. Figure 3.6 shows that of the 18 contigs, one of which was the plasmid pLF3872, only eight align to the optical map and none were correctly assembled.

3.2.3 Determining the point of assembly error

To identify the point at which misassembly occurred, along with its cause, the CLC GWB, SPAdes V3.5.0 (Bankevich *et al.*, 2012), MIRA V3.4.2.0 (Chevreux *et al.*, 2004) and CISA-generated contigs used to produce the 18 contigs were aligned to the optical map using Mapsolver software. Read map analysis of the point at which the genome assembly differed from the optical map, along with its surrounding areas, were also examined to determine whether read mapping had identified misassembly of the *L. fermentum* 3872 genome sequence.

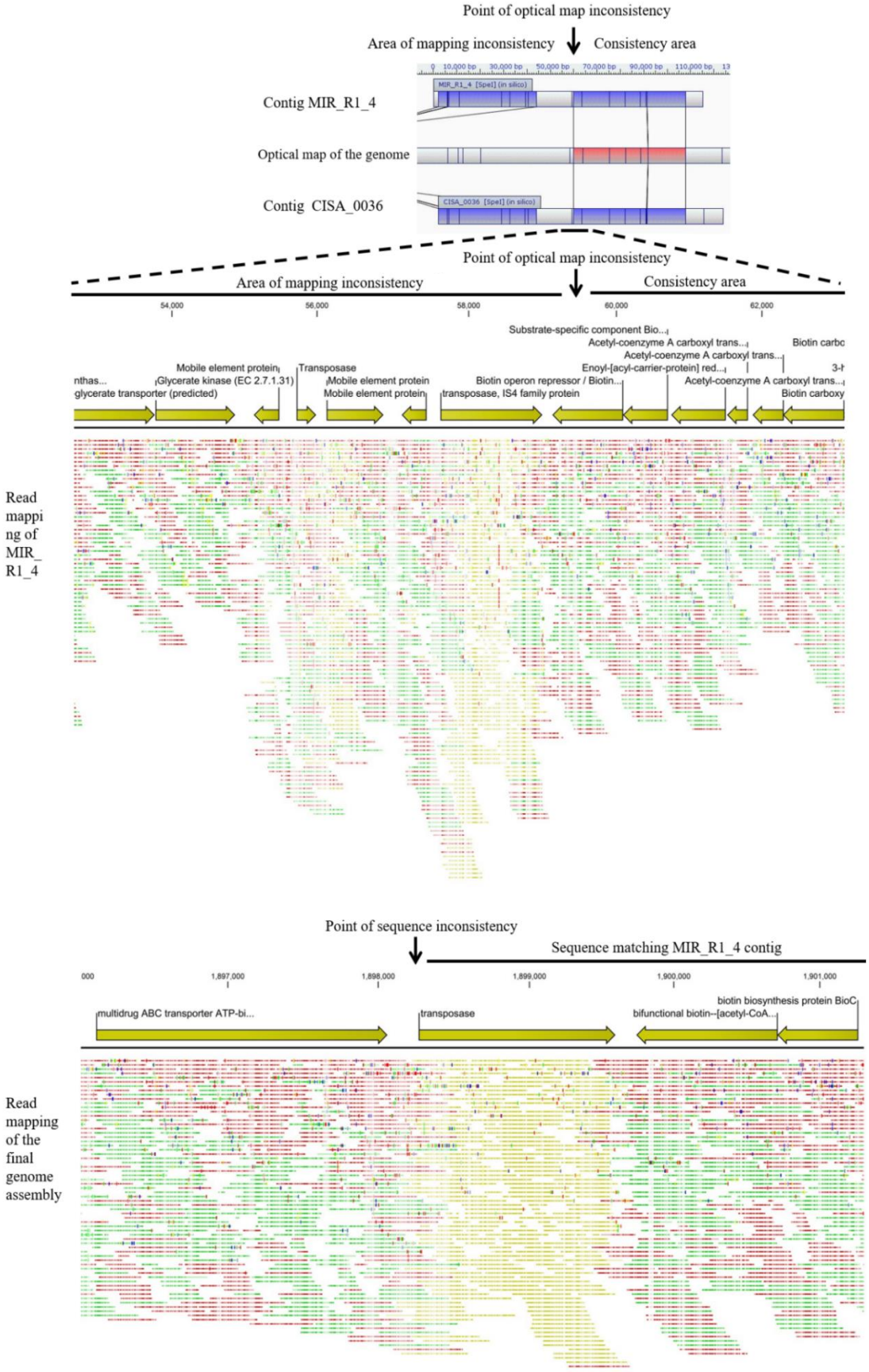


Figure 3.7. Optical map and read map comparison in determining genome assembly error. The read colours red, green and yellow represent forward, reverse and complementary reads respectively. The blue regions on the optical map indicate areas of

the sequenced genome that align to the optical map, while red regions indicate duplicate alignments and white regions indicate areas where there is no sequence alignment to the optical map. The diagram was obtained from Lehri, Seddon and Karlyshev (2017b).

Table 3.4. The number of contigs that aligned to the optical map for the various assemblers used, along with the number of contigs that were identified as misassembled. The table was obtained from Lehri, Seddon and Karlyshev (2017b).

	Ion Torrent											**PacBio RSII**
	Run 1			**Run 2**			**Run 3**					
	CLC	**MIRA**	**SPAdes**	**CLC**	**MIRA**	**SPAdes**	**CLC**	**MIRA**	**SPAdes**	**CISA**	**18 Contigs**	**Final assembly (HGAP3)**
Contigs aligned to the optical map	9	7	12	8	12	10	5	11	11	12	8	1
Misassembled contigs	1	2	2	1	1	1	1	4	1	4	8	0

Many of the assembly errors occurred near repetitive regions. An example is provided in Figure 3.7, where the point at which the optical map differs from the incorrect genome assembly is either near a transposon/mobile element or an rRNA region (Lehri, Seddon and Karlyshev, 2017b). Figure 3.7 also demonstrates that read mapping did not identify the misassembly, and thus can sometimes fail as a determinant for sequence assembly quality. Read mapping of the region where misassembly occurred (point of misassembly) shows no indication of misassembly, yet the optical map clearly shows a misassembled genome. Analysis of the *de novo* assembled contigs that aligned to the optical map also showed that some of the errors occurred during *de novo* assembly and were software specific while others were specific to a sequence region of the genome. For example, aligning the CISA-generated contigs to the optical map showed that assembly errors were transferred from *de novo* assembled contigs to CISA contigs. This can be observed in Figure 3.7 where the *de novo* assembled contig MIR_R1_4 transferred its misassembled region to CISA_0036. As previously mentioned, of the 18 contigs, none aligned correctly to the optical map (Figure 3.6, Table 3.4). Thus, additional contiguation steps on the CISA generated contigs resulted in a far worse

genome assembly. A summary of the misassembled contigs is provided in Table 3.4, which also shows that some *de novo* assemblers had more misassembled contigs than others.

3.2.4 PacBio RSII sequencing result

To complete the genome sequence of *L. fermentum* 3872, a hybrid sequencing approach was needed. This combined short reads generated by the Ion Torrent PGM and long reads generated by the PacBio RSII sequencer. For PacBio RSII sequencing, fresh DNA that had not been stored at -20°C was required to avoid shearing caused by cycles of freezing/thawing. Thus, DNA was isolated from *L. fermentum* 3872 clonal isolate 1 as described in section 2.7. DNA from clonal isolate 1 was used for PacBio RSII sequencing and for Ion Torrent PGM genome assembly.

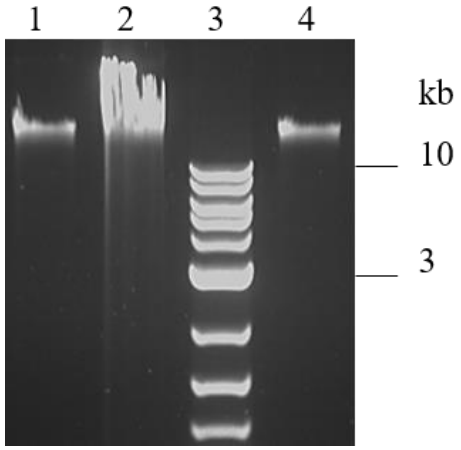


Figure 3.8. Agarose gel image of DNA isolated from *L. fermentum* 3872 used for PacBio RSII sequencing. The wells are: 1, Sample 1 (1:10 dilution); 2, Sample 1 (neat); 3, 2-log DNA marker (NEB); 4, Sample 2.

Both samples looked concentrated and of good quality. There is a light smear observed below the bright band for Sample 1 (Figure 3.8), this may be caused by the higher concentration of the DNA as the smearing is not evenly distributed. Both samples were sent to TGAC genomics for PacBio RSII sequencing.

3.2.5 Pacbio RSII sequencing sample quality

TGAC genomics conducted PacBio RSII sequencing and sample quality analysis. Unlike Ion Torrent PGM, the PacBio RSII sequencer does not amplify DNA and thus, must be of high quality and free from contaminants, such as RNA and proteins, before sequencing. Multiple DNA quality analysis tools such as a Qubit fluorometer, and DropSense spectrophotometer were used to accurately determine DNA purity. A Tapestation was used to determine DNA fragment size after fragmentation of DNA.

Table 3.5. Concentration, purity, and fragment size of the isolated DNA of *L. fermentum* 3872 for PacBio RSII sequencing.

Sample	**Qubit Values**		**DropSense Values**				**Tapestation Values**		**Total Amount (µg)**	**OVERALL P/F**
	ng/µl	**Pass/Fail**	**ng/µl**	**260/280**	**260/230**	**Overall P/F**	**Average insert size kb**	**Pass/Fail**		
1	318	Pass	38.95	1.78	1.43	Pass	>60	Pass	30.21	Pass
2	618	Pass	8.78	1.48	1.19	Fail	>60	Pass	358.44	Fail

For Sample 2, the A260/280 value was 1.48, which was lower than the 1.8 recommended by TGAC genomics (Table 3.5). The A260/230 value of 1.19 was also lower than the recommended 1.8-2.0. Sample 1 had an A260/280 value of 1.78, which was almost within the acceptable range of 1.8-2.0, although the A260/230 value was 1.43 (Table 3.5). As previously mentioned, small amounts of DNA extraction reagents used, commonly cause low A260/230 values, which would be removed during PacBio RSII library preparation as a clean-up is performed before sequencing, and thus is not much of a concern. The low A260/280 value is concerning as this indicates nucleic acid

purity. Contaminants such as proteins may cause a lower value, the presence of proteins would affect the average read length of the sequencing data. Thus, Sample 1 was chosen for PacBio RSII sequencing.

3.2.6 The final assembly of the *L. fermentum* 3872 chromosomal sequence

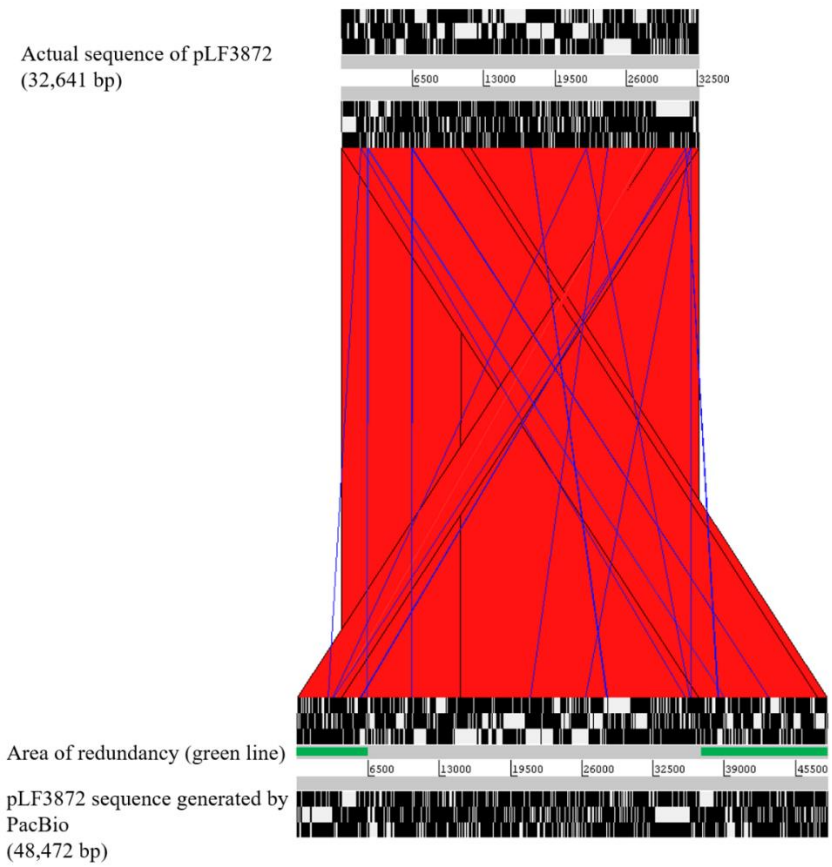


Figure 3.9. pLF3872 assembly comparison of the Ion Torrent PGM generated plasmid sequence (top) and the PacBio RSII generated plasmid (bottom). The red regions show alignment to similar areas between the two plasmid assemblies, with the redundant region of the pLF3872 PacBio RSII assembly underlined in green. The diagram was obtained from Lehri, Seddon and Karlyshev (2017b).

As the initial attempt in trying to complete the chromosomal sequence of *L. fermentum* 3872 was unsuccessful, PacBio RSII reads were *de novo* assembled using HGAP3 software. The 74,588 reads produced by the PacBio RSII sequencer generated seven contigs. The smallest contig was 14,320 bp in length, while the largest contig was 2,302,236 bp. The largest contig aligned to the optical map of the chromosomal sequence of *L. fermentum* 3872. The size of the chromosomal sequence generated by the HGAP3 assembly tool was larger than that of the optical map. The map and Artemis

Comparison Tool (ACT) (Carver *et al.*, 2005) indicated that the ends of the sequence had large redundancies, with both ends aligning to each other. Merging overlapping regions resulted in a circular chromosomal sequence of 2,297,851 bp. This was also observed for the plasmid pLF3872, as the PacBio RSII generated contig of 48,472 bp represented the already completed plasmid sequence pLF3872. The plasmid size was larger than the predetermined pLF3872 size of 32,641 bp. Comparison of the two sequences using the Artemis comparison tool (ACT) (Carver *et al.*, 2005) identified redundant overlapping regions for the plasmid sequence of 48,472 bp in length (Figure 3.9). Once the redundant sequence was merged together, the plasmid was circularised and matched with the pLF3872 plasmid generated using the Ion Torrent PGM. The remaining five contigs generated by the PacBio RSII sequencing data ranged from 14-18 kb in size. BLASTN comparison of the five contigs to the completed chromosomal sequence of *L. fermentum* 3872 identified that these sequences were fragments of the chromosome matching to repetitive sequences, namely transposons or integrase encoding genes. Mapping of PacBio RSII reads to the contigs resulted in only a single read aligning to the map. Thus, the five sequences were low quality reads that were inaccurately classified as contigs by the HGAP3 software.

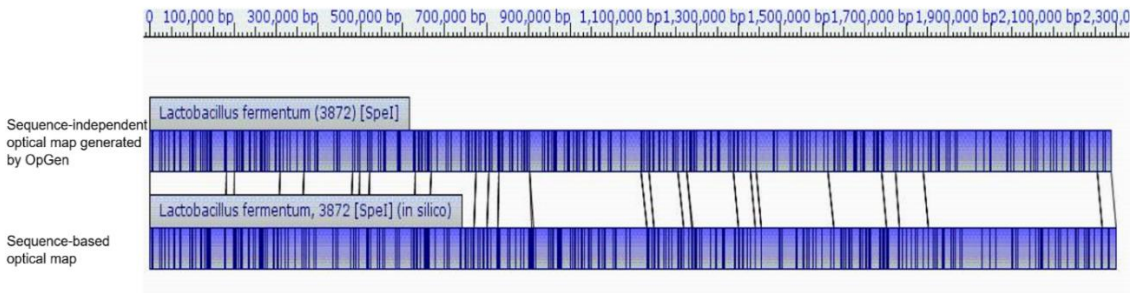


Figure 3.10. Final assembly of the *L. fermentum* 3872 genome sequence aligned to the optical map of *L. fermentum* 3872. Alignment was conducted using Mapsolver software with default parameters. The blue regions indicate regions of the chromosomal sequence that aligned to the optical map. The diagram was obtained from Lehri, Seddon and Karlyshev (2017b).

The circularised chromosomal sequence mapped with no indication of misassembly (Figure 3.10). As the PacBio RSII generated reads had low coverage of 19.47x, Ion

Torrent PGM reads were mapped onto the chromosomal sequence of *L. fermentum* 3872. The Ion Torrent PGM reads revealed artefactual variations (predominantly SNPs) introduced by the PacBio RSII reads, which were corrected using the Ion Torrent generated data.

To determine if clonal variations were present on both the chromosomal and plasmid (pLF3872) sequence of *L. fermentum* 3872, Ion Torrent reads generated from clonal isolate 2 were mapped onto the finalised genome sequence. This did not detect any clonal variations and thus the genome sequence was representative of *L. fermentum* 3872.

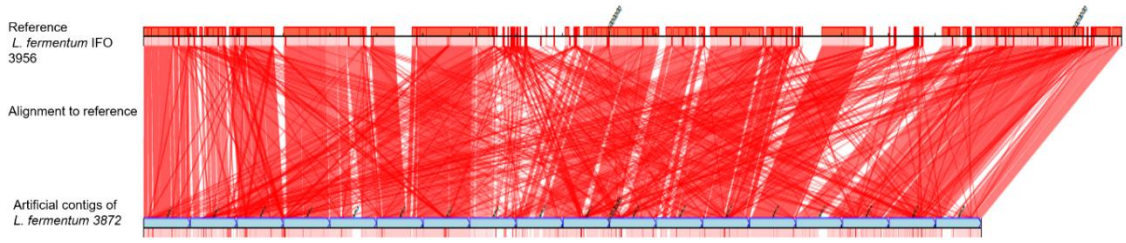


Figure 3.11. Contiguator alignment of correctly assembled *L. fermentum* 3872 genome artificially split at 100,000 bp and aligned to reference sequence *L. fermentum* IFO 3956. BLASTN with default alignment parameters were used. Only contigs (blue) that aligned the reference strains are shown, with red lines indicating regions that match between the two strains. The diagram was obtained from Lehri, Seddon and Karlyshev (2017b).

Purposefully splitting the correctly assembled chromosomal sequence of *L. fermentum* 3872 at 100,000 bp intervals and then aligning to reference sequence *L. fermentum* IFO 3956 using Contiguator revealed that the alignment (Figure 3.11) was better than initially observed in Figure 3.5. Thus, sequence misassembly was the main cause of having multiple regions of the *L. fermentum* 3872 contigs aligning to various areas of the reference *L. fermentum* IFO 3956 (Figure 3.5).

After completion of the *L. fermentum* 3872 genome sequence, the genome was annotated using various annotation tools and manually inspected using Geneious software (Kearse *et al.*, 2012), as certain annotation programmes would classify genes as hypothetical while others would be able to name the genes and assign function. The

annotations were also inspected to identify and correct premature start and stop codons. Once annotation was complete, the genome of *L. fermentum* 3872 was analysed.

3.3 *L. fermentum* 3872 genome sequence analysis

3.3.1 General overview of the *L. fermentum* 3872 genome

The final genome sequence of *L. fermentum* 3872 consists of a circular plasmid (pLF3872) of 32,641 bp, while the chromosomal sequence of *L. fermentum* 3872 is 2, 297,851 bp in length. Together with the plasmid, the genomic size of *L. fermentum* 3872 is 2,330,492 bp. The genome of *L. fermentum* 3872 consists of 2,328 genes, with 2,127 genes encoding proteins and 128 pseudogenes. There are 73 RNA genes, 15 encoding for rRNAs (23S, 16S and 5S) and 58 encoding for tRNAs. The G+C content of the chromosomal and plasmid sequences are 55.6% and 40.1% respectively.

Table 3.6. Plasmid pLF3872 genes, their functions and the closest BLASTX hits in the NCBI non-redundant database. The table was obtained from Lehri, Seddon and Karlyshev (2015).

Loci	Gene	Gene product	Putative function	Reference strain	Identity % (coverage%)
1-1335	*ykgC*	Pyridine nucleotide-disulphide oxidoreductase YkgC	Energy production (oxidoreductase activity)	*Lactobacillus plantarum* CMPG5300	100 (99)
1663-2247	*tnpR*	Putative resolvase	Recombination	*Lactobacillus salivarius* CECT 5713	99 (99)
2571-3197	*hyp4*	Conserved hypothetical protein		*Lactobacillus salivarius*	99 (99)
3409-3693	*dinJ*	DNA damage inducible protein J	Antitoxin	*Lactobacillus brevis* ATCC 367	93 (98)
4083-4535	*hyp5*	Hypothetical protein		Multiple *Lactobacillus* bacteria	99 (99)
5007-5798	*parA*	Plasmid partitioning protein ParA	Cell division, partitioning (replication)	Multiple *Lactobacillus* bacteria	100 (99)
5840-6109	*hyp6*	Hypothetical protein		*Lactobacillus hilgardii*	96 (98)
6670-7833	*repA*	Replication initiator protein RepA	Replication	*Lactobacillus crispatus* EM-LC1	100 (99)
7808-8104	*hyp7*	Hypothetical protein		*Lactobacillus crispatus* EM-LC1	100 (98)
8388-8660	*eatL*	Epsilon anti-toxin	Post-segregation killing system	Multiple *Lactobacillus* bacteria	97 (98)
8663-9493	*zetL*	Zeta toxin	Post-segregation killing system	*Lactobacillus antri*	99 (98)
9577-9855	*hyp10*	Hypothetical protein		*Lactobacillus oris* PB013-T2-3	99 (98)
9878-10120	*hyp11*	Hypothetical protein		*Lactobacillus casei* Zhang	100 (92)
10346-12421	*traA*	Nickase	Conjugation	*Lactobacillus rhamnosus*	99 (99)
12506-12817	*hyp1*	Hypothetical protein		*Lactobacillus plantarum* CMPG5300	99 (99)
12853-13467	*hyp9*	Hypothetical protein		*Lactobacillus plantarum* CMPG5300	99 (99)
13469-13804	*traB*	Transfer complex protein TraB	Conjugation	*Lactobacillus paracasei*	99 (99)
13825-14187	*trsC*	TrsC	Conjugation	*Lactobacillus paracasei*	99 (99)
14156-14815	*trsD*	TrsD	Conjugation	*Lactobacillus plantarum* CMPG5300	99 (99)
14827-16845	*trsE*	TrsE	Conjugation	*Lactobacillus paracasei*	99 (99)
16838-18256	*trsF*	TrsF	Conjugation	*Lactobacillus plantarum* CMPG5300	97 (99)
18257-19414	*tcpG*	Peptidoglycan hydrolase	Hydrolysis of peptidoglycan	*Lactobacillus casei* Zhang	99 (99)
19428-20045	*hyp12*	Hypothetical protein		*Lactobacillus oris* PB013-T2-3	99 (99)
20032-20400	*Trx*	Thioredoxin	Reduction of oxidising compounds	*Lactobacillus oris* PB013-T2-3	94 (99)
20401-20871	*trsJ*	TrsJ	Conjugation	*Lactobacillus helveticus* CIRM-BIA 101	96 (99)
21101-22651	*traG*	Conjugal transfer protein TraG	Conjugation	*Lactobacillus oris* PB013-T2-3	98 (99)
22651-23055	*hyp2*	Hypothetical protein		*Lactobacillus coryniformis* CECT 5711	99 (99)
23074-23913	*trsL*	TrsL	Conjugation	*Lactobacillus paracasei* Lpp189	100 (99)
23928-24338	*hyp3*	Hypothetical protein		Multiple *Lactobacillus* bacteria	100 (99)
24345-26480	*topB*	DNA topoisomerase III	Replication	*Lactobacillus paracasei*	98 (99)
26602-26817	*hyp8*	Hypothetical protein		*Lactobacillus oris* PB013-T2-3	99 (98)
26821-27945	*ltrC*	Low temperature requirement C protien	Phosphatidylglycerophosphatase activity	*Lactobacillus coryniformis*	97 (99)
28656-31823	*cbp*	Collagen binding protein	Adhesion	*Lactobacillus casei, Lactobacillus oris*	94 (99)

Plasmid pLF3872 contains 33 genes (see Table 3.6), 12 of which are hypothetical. There are toxin-antitoxin gene pairs *eatL* and *zetL*, potentially involved in stable maintenance of the plasmid in *L. fermentum* 3872 (Zielenkiewicz *et al.*, 2009), along with an additional antitoxin gene *dinJ* (Hu, Benedik and Wood, 2012). Genes *traA*, *trsC*, *trsD*, *trsE*, *trsF*, *trsG*, *trsJ*, and *trsL* of pLF3872 are potentially involved conjugation. The protein sequence of a peptidoglycan hydrolase (TcpG) encoding gene has a lytic transglucosylase domain and an amidase-5 domain both of which may play a role in conjugation (Scheurwater, Reid and Clarke, 2008). The plasmid contains a putative *cbp* gene potentially involved in promoting bacterial adhesion to collagen (Pietrocola *et al.*, 2007).

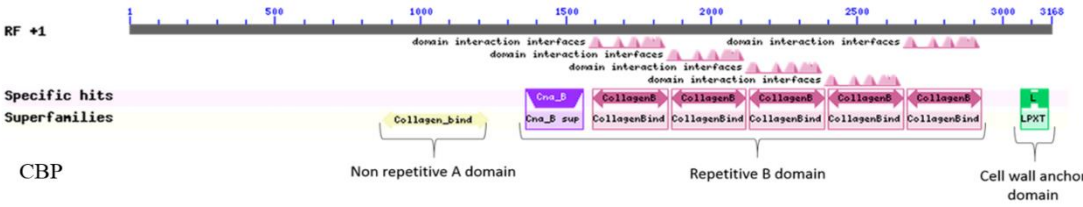


Figure 3.12. pLF3872 CBP protein domains, identified by the NCBI conserved domain (CD) program and the CD database. The diagram was obtained and modified from Lehri, Seddon and Karlyshev (2015).

The CBP protein has an LPXT domain at the C-terminal end, essential for cell wall anchoring (Davies, Svensäter and Herzberg, 2009). The N-terminal region is predicted to form an A-domain, in other CBP proteins the domain is involved in adhesion (Deivanayagam *et al.*, 2000). The A-domain is followed by six repetitive B domains that form stalks presenting the A domain for attachment to host cells (Figure 3.12) (Deivanayagam *et al.*, 2000).

The chromosomal sequence of *L. fermentum* 3872 contains genes that play a role in the production of vitamins B1, B2, B5, B7 and B9. *L. fermentum* 3872 also contains acid symporter genes such as sugar/H+, amino acid/H+, and four copies of Na+/H+. Lactic acid bacteria such as *L. fermentum* 3872 utilise amino acid decarboxylation/deamination and sugar metabolism to sustain pH levels for bacterial survival (Pessione, 2012). *L.*

fermentum 3872 contains 27 genes involved in arginine and proline metabolism and 14 genes involved in glutathione metabolism. The latter is found to participate in acid stress response in *Lactobacillus salivarius* (Lee *et al.*, 2010). The *L. fermentum* 3872 sequence also consists of a gene that encodes dTDP-glucose 4,6-dehydratase (Locus tag: N573_RS00605), a homologue of which is known to be involved in gastric acid tolerance in *Lactobacillus plantarum* (Hamon *et al.*, 2014). There are a number of exopolysaccharide encoding genes with locus tags N573_RS00495, N573_RS00500 and N573_RS00505 located between 117326 – 119653 bp, forming a cluster. *L. fermentum* 3872 also has a undecaprenyl-diphosphatase (EC 3.6.1.27) (Locus tag: N573_RS09665) encoding gene possibly involved in bacitracin resistance (Ghachi *et al.*, 2005). *L. fermentum* 3872 contains genes encoding DnaK (Locus tag: N573_RS04975), and GroEL (Locus tag: N573_RS01895), which may be involved in bacterial survival during heat and hyperosmotic shock (Prasad, McJarrow and Gopal, 2003). Homologues of both genes are also thought to play a role in mucin binding in *Lactobacillus plantarum* (Izquierdo *et al.*, 2009). A gene encoding a translational elongation factor (Locus tag: N573_RS03470) protein is also present. A homologue of the protein is known to promote adhesion to host cells of *Lactobacillus johnsonii* La1 (Granato *et al.*, 2004). *L. fermentum* 3872 contains genes encoding D-Lactate dehydrogenase (Locus tag: N573_R S11010) and 6-phosphogluconate dehydrogenase (Locus tag: N573_R S10960), homologues of which are known to promote bacterial adhesion to mucin in *Lactobacillus plantarum* (Izquierdo *et al.*, 2009). A partial gene potentially encoding for a mucus binding protein (Locus tag: N573_RS03620) is also present for *L. fermentum* 3872. Mucus binding protein encoding genes are also present for *Lactobacillus fermentum* strains CECT 5716, 3956 and F6. A fibronectin binding protein encoded by the *fbpA* (Locus tag: N573_RS07225) gene is likely involved in adhesion to ECM proteins such as fibronectin (Muñoz-Provencio, Pérez-Martínez and

Monedero, 2010). The gene is commonly found in other *L. fermentum* strains. The *L. fermentum* 3872 genome contains common antibiotic resistance genes such as four genes involved in in tetracycline resistance, two genes involved in beta-lactam antibiotic resistance, and four genes involved in fluoroquinolone resistance, along with six multidrug efflux pump-related genes. There are genes involved in heavy metal resistance with five genes involved in arsenate, six in mercury, five in cobalt-zinc-cadmium, and six copper resistance genes.

3.3.2 Comparative genomic analysis

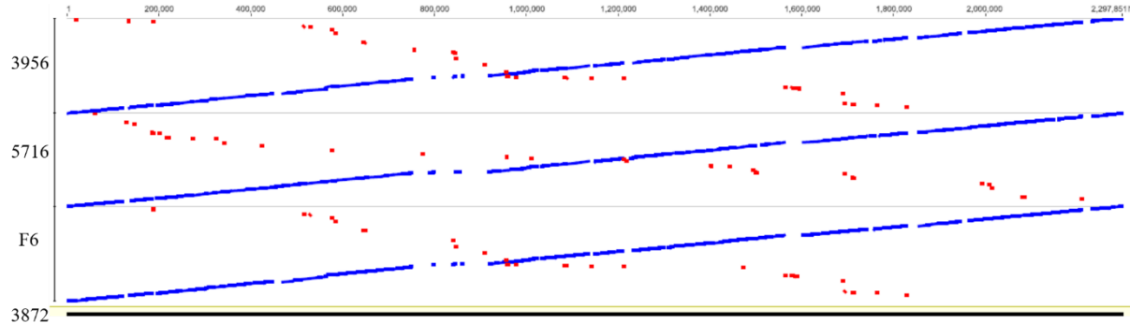


Figure 3.13. LASTZ comparison of the *L. fermentum* 3872 chromosomal sequence to sequences *L. fermentum* strain F6, CECT 5716, and IFO 3956. Default LASTZ parameters were used. Genomic regions that are similar among the strains used for comparative analysis are in blue, while red lines represent inverted areas, non-coloured regions are unique to *L. fermentum* 3872. The diagram was obtained from Lehri, Seddon and Karlyshev (2017a).

LASTZ software (Harris RS, 2007) was used for comparison of the chromosomal sequence of *L. fermentum* 3872 and *L. fermentum* strains F6, IFO 3956, and CECT 5716 (Figure 3.13). The software identified multiple unique genomic regions of *L. fermentum* 3872. The region located between 748,875 and 919,330 bp contained hypothetical protein encoding genes, along with genes that encoded enterolysin A (835, 633–836, 847 bp) and CAAX amino terminal protease self-immunity (838,683–839,366 bp). The latter two genes indicate the presence of a bacteriocin. BAGEL3 software (Heel *et al.*, 2013) confirmed that the region between 830,634 and 840,633 bp encoded a class III bacteriocin (Figure 3.14). A BLASTN search using the non-redundant NCBI database identified no valid matches to the region between 830,634 and 840,633 bp of the *L.*

fermentum 3872 genome. The region between 1,564,375 bp and 1,603,857 bp contained prophage-related genes not found in strains used for comparative analysis, and inversions of certain regions of the strains used for comparative analysis. The region between 1,829,274 and 1,857,186 bp was likely acquired by horizontal DNA transfer, as a match to *Lactobacillus gasseri* ATCC 33323 was identified when using a BLASTN search of the non-redundant NCBI database, with no matches in the genomes of *L. fermentum* strains used for comparative analysis. The region between 2,212,692 and 2,237,160 bp contains conjugation and peptidoglycan hydrolase encoding genes. A BLASTN search of the latter region using the non-redundant NCBI database resulted in multiple matches to plasmid sequences, with the closest to plasmid pPECL-5 (*Pediococcus claussenii* ATCC BAA-344) with an e value of 0.0, and a query coverage of 55%. The area has likely been acquired via horizontal gene transfer and was not identified in other *L. fermentum* strains used for comparative analysis. This region also had a gene that encoded an internalin J-like protein (InIJ) (Locus tag: N573_011130), which contains a mucin-binding domain (MuCBP). The region also has transposase encoding genes.

L. fermentum 3872 also has an aggregation substance precursor (Locus tag: N573_004020) encoding gene not found in other *L. fermentum* strains used for comparative analysis. The protein encoded by the gene has been linked to both adhesion and aggregation (Suessmuth *et al.*, 2000). There is an enolase-encoding gene (Locus tag: N573_002185), the protein encoded by the gene is known to be involved in adhesion to collagen (Salzillo *et al.*, 2015). The gene is also present in other *L. fermentum* genomes. A partial *cbp* gene (Locus tag: N573_000435) was also identified in the chromosomal sequence of *L. fermentum* 3872. The partial CBP protein has a cell wall anchor domain (LPXTG), along with a B domain, but does not have an A domain potentially used by CBP for adhesion (Deivanayagam *et al.*, 2000). Neither the *cbp*

gene of pLF3872 nor the chromosomal sequence of *L. fermentum* 3872 are present in other *L. fermentum* strains used for comparative analysis. An *epsH* (Locus tag: N573_008790) gene involved in biofilm formation (Jones *et al.*, 2014) is also present in the *L. fermentum* 3872 genome and is absent in all the genomes of *L. fermentum* strains used for comparative analysis.

3.3.3 Pan/Core comparative analysis

Pan/Core analysis software (Spine/AGEnt) (Ozer, Allen and Hauser, 2014) was used to identify commonly found genes in *L. fermentum* F6, IFO 3956 and 3872, referred as core genes. *L. fermentum* CECT 5716 was excluded from the analysis as it contained many nucleotide ambiguities. The software was also used to identify unique genes (Pan genes) of *L. fermentum* 3872. As described in Lehri, Seddon and Karlyshev (2017a), the software identified 1650 genes that are common in the strains of *L. fermentum* that were tested. The software also identified 428 genes that are unique to the *L. fermentum* 3872 genome. Of the unique genes present, 148 encoded hypothetical proteins, while nine represented CRISPRs, forty nine mobile elements, five conjugal transfer genes. Other genes were related to ABC transporter, bacteriocin biosynthesis, heavy metal resistance and prophage.

3.3.4 Identified prophages of *L. fermentum* 3872

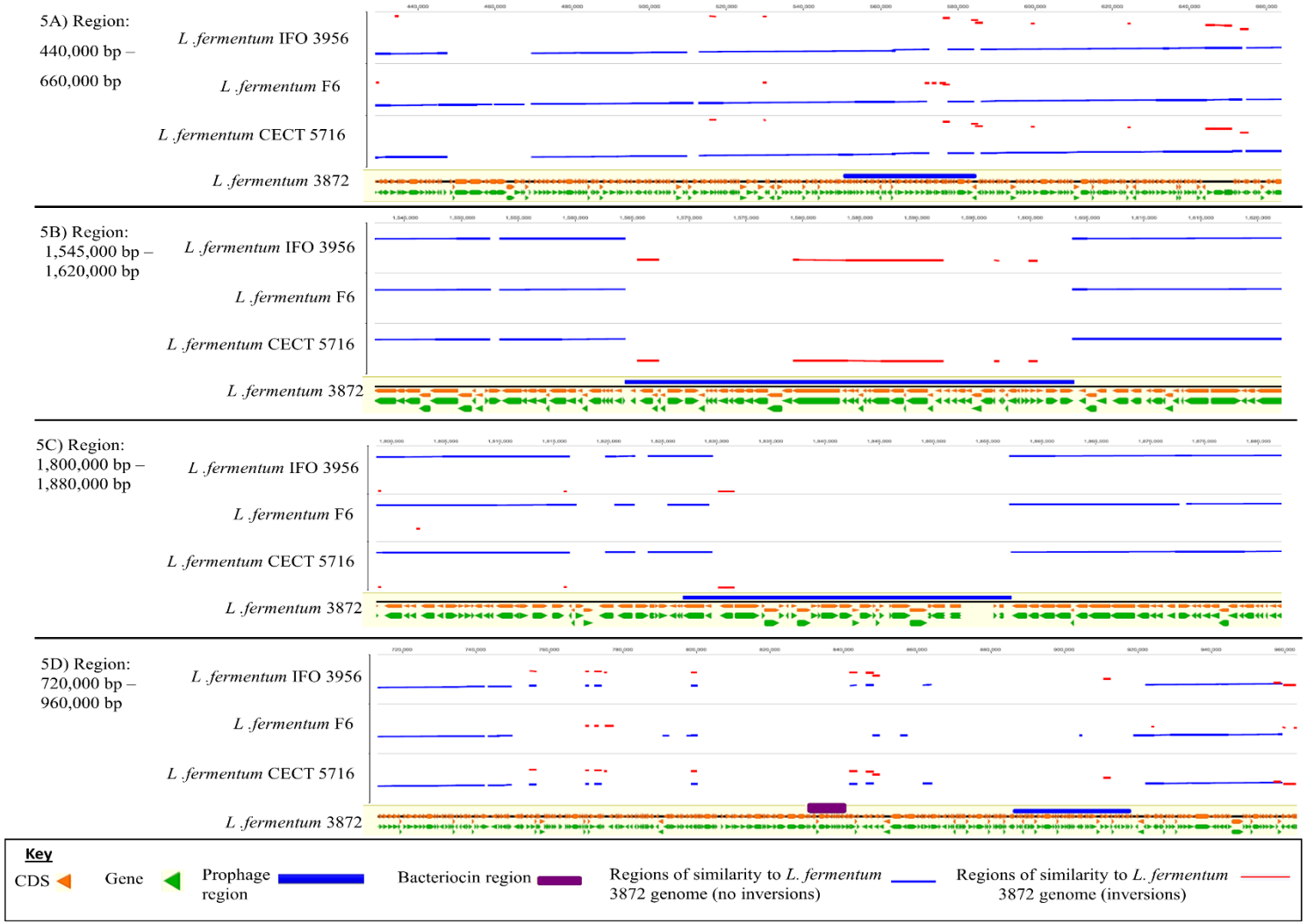


Figure 3.14. LASTZ comparison of the chromosomal sequence of *L. fermentum* 3872 and *L. fermentum* strains F6, CECT 5716, and IFO 3956, with highlighted prophage and bacteriocin encoding regions. Default LASTZ parameters were used. The diagram was obtained from Lehri, Seddon and Karlyshev (2017a).

As Pan/Core analysis found a few prophage related genes, PHAST software (Zhou *et al.*, 2011) was used to identify any other *L. fermentum* 3872 prophages. Four regions that are associated with prophages were recognised, with all having a phage attachment (ATT) site, required for site-specific recombination. An area between 550,236 bp and 584,763 bp of the *L. fermentum* 3872 chromosomal sequence had a 34.5 kb prophage region containing multiple phage tail protein encoding genes. Phage tail proteins are involved in phage attachment to host cells (Auzat *et al.*, 2008). The region also contains transposase and integrase encoding genes. A 32 kb region located between 886,091 bp and 918,126 bp also has many transposase and integrase encoding genes. An area between 1,564,361 bp and 1,603,857 bp is a prophage-related 39.4 kb region. The region has genes that encodes phage tail, head and portal proteins. The region also has transposase, integrase and a protease encoding gene. PHAST software (Zhou *et al.*, 2011) also identified a 30.2 kb prophage region between 1,826,924 bp and 1,857,190 bp, containing genes that encoded portal, head and capsid proteins along with terminase and recombinase encoding genes. This region also has a *mucBP* type gene, the protein of which contains 17 repeats of MuCBP binding domains, although there was no cell wall anchor domain, which is often found in cell surface proteins of Gram positive bacteria (Mazmanian *et al.*, 1999). Multiple prophage-related genes were also identified close to the bacteriocin encoding region (830,634 bp – 840,633 bp) (Figure 3.14), although this region was not classified as prophage-related by PHAST software (Zhou *et al.*, 2011).

As can be seen in Figure 3.14, there were also small similarities between the prophage regions of *L. fermentum* 3872 and the *L. fermentum* strains used for comparative analysis.

3.4 Cloning and expression of the *cbp* gene

Genome sequence analysis of *L. fermentum* 3872 identified a *cbp* gene encoding a collagen binding protein not found in any other strain of the species. The amino acid sequence of the CBP protein of pLF3872 showed that the protein had all the domains essential for being a functional protein. Thus, to determine whether the protein promoted bacterial adhesion to collagen, and consequently contributed to GIT colonisation, the gene was cloned in *E. coli* and purified for future adhesion experiments.

3.4.1 Initial attempt to clone and express the *cbp* gene

PCR was conducted to amplify the *cbp* gene of plasmid pLF3872 using primers CBPREV1: atgagcatgctcaatggtgatggtgatggtgaatagtaaatctacttataactactaaacc and CBPFOR1: aacaatctagaagaaggaggcaacagtatgagtaataaatctaggcatcgtcgac. *Xba*I and *Sph*I restriction enzyme sites were added at the ends of the amplified *cbp* gene, along with a Shine-Dalgarno (S-D) sequence, used to aid protein expression. The *cbp* gene sequence also had an C-terminal hexa histidine tag for purification of recombinant CBP. Once PCR was complete, agarose gel electrophoresis was used to determine sample quality.

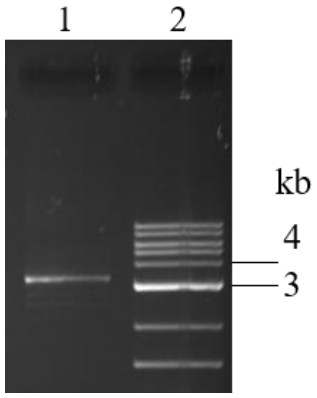


Figure 3.15. Agarose gel electrophoresis result of amplified *cbp* gene with C-terminal hexa histidine tag and a leader peptide sequence. 1, amplified *cbp* gene; 2, 2-log DNA ladder (NEB).

Figure 3.15 shows that the PCR product is of expected size (3.2 kb) and that there are no spurious bands that may affect cloning. The PCR product was used for cloning experiments.

3.4.1.1 Cloning the *cbp* gene

QIAquick PCR purification kit (Qiagen) was used to purify the PCR product. The latter would remove PCR contaminants such as primers that may affect restriction digestion and ligation. The PCR product and plasmid vector pBAD33 were digested using enzymes *Xba*I and *Sph*I, generating sticky ends for ligation. Shrimp alkaline phosphatase (SAP) was not used as the sticky ends were not complementary to each other and the samples were gel purified shortly after restriction digest. After gel extraction, the *cbp* gene insert and the pBAD33 plasmid were ligated and transformed on to NEB express *E. coli*. Clonal isolates of the transformants were grown to confluence, and plasmid extraction was conducted. A restriction digest and agarose gel electrophoresis were performed to determine if the extracted plasmid contained the *cbp* gene insert. DNA concentration readings were taken using a NanoVue spectrophotometer to determine if the sample was concentrated for Sanger sequencing.

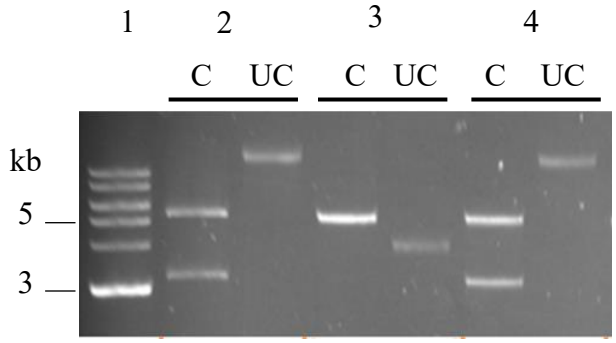


Figure 3.16. Agarose gel electrophoresis result confirming the presence of the *cbp* gene inserted on to the plasmid pBAD33; C represents samples that have been digested while UC are undigested samples (negative controls). The wells are: 1, 2 log-DNA ladder; 2, Sample 1; 3, Sample 2; 4, Sample 3.

Table 3.7. NanoVue spectrophotometer reading of samples 1 and 3 containing pBAD33 plasmid with the *cbp* gene insert.

Sample 1		**Sample 3**	
Concentration:	66 ng/µl	**concentration:**	80 ng/µl
A260/A280	1.826	**A260/A280**	1.785
A260/A230	2.1	**A260/A230**	1.559

Samples 1 and 3 have the *cbp* gene insert of the expected size (3,208 bp), and the digested pBAD33 sample is also of expected size (5.3 kb) (Figure 3.16). Sample 2 does not contain the *cbp* gene insert, as only one band is present. Table 3.7 shows that samples 1 and 3 are concentrated and pure enough to be sent for Sanger sequencing.

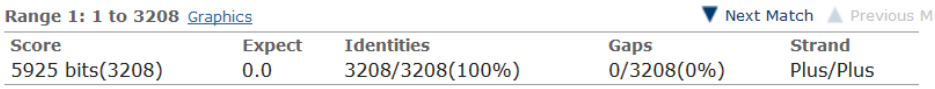


Figure 3.17. NCBI BLAST result of the Sanger sequenced *cbp* gene with a C-terminal hexa histidine tag compared to the *cbp* gene sequence generated from Ion Torrent PGM.

The Sanger sequenced *cbp* gene for both samples had no nucleotide variations, which can be seen in Figure 3.17. There is a 100% match (3208/3208) between the Sanger sequenced *cbp* and that sequenced by Ion Torrent PGM, with no gaps found between sequences.

3.4.1.2 Expression of the CBP protein

The recombinant CBP protein was expressed in NEB express *E. coli* and then purified. An initial small-scale protein purification was conducted using a Magne-His protein purification kit, in order to not waste largescale protein purification columns and samples. Prior to lysis, to prevent protein degradation, a cocktail of protease inhibitors designed for recombinant *E. coli* protein expression was added. At each step of the protein expression and purification process, small volumes of samples were taken to be loaded onto a polyacrylamide gel for Coomassie staining. This would help determine the step at which a possible issue may have occurred during protein expression.

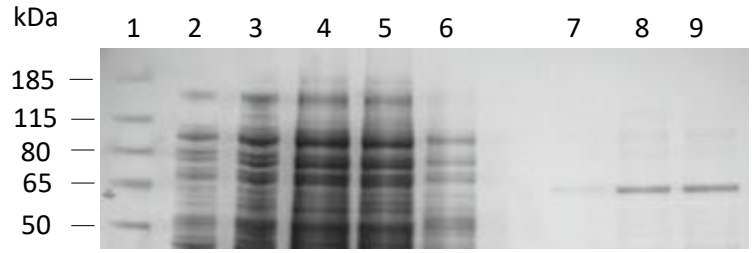


Figure 3.18. Coomassie stain of the recombinant CBP protein with C-terminal hexa histidine tag and a leader peptide. The wells are: 1, pre-strained protein ladder (Page ruler plus); 2, lysate before induction; 3, lysate after induction; 4, flow through of lysate; 5, column wash 1; 6, column wash 2; 7, eluate 1; 8, eluate 2; 9, eluate 3.

Coomassie staining of the gel (Figure 3.18) shows that the CBP protein is approximately 65 kDa, which is considerably smaller than the expected size of 111 kDa. The protein is not useable for adhesion experiments.

3.4.2 Removing the leader peptide sequence of the *cbp* gene and adding an N-terminal hexa histidine tag

The previously cloned *cbp* gene construct resulted in a smaller protein size than predicted, even after using protease inhibitors. Thus, a new *cbp* gene construct was made based on previously published work that expressed an orthologue of the *cbp* gene in *E. coli* (Esmay *et al.*, 2003). The leader peptide sequence was removed from the new *cbp* construct, the hexa histidine tag and S-D sequence were placed at the N-terminal of the CBP. Leader peptide sequences are usually removed from cell surface proteins like CBP by bacteria. The leader peptide sequence of the native CBP protein of *L. fermentum* 3872 may not be recognised by the *E. coli* translocation system, resulting in protein misfolding and a smaller protein size. Deciding which terminal to place the hexa histidine tag is also a matter of trial and error, as sometimes placing the hexa histidine tag in the wrong terminal may lead to protein misfolding. Thus, an N-terminal peptide was chosen, as was done by Esmay *et al*. (2003).

PCR was conducted using primers labelled CBPREV2: atgagcatgctcaaatagtaaatctacttataactactaaacc, and CBPFOR2: atatgcttctagaagaaggaggcaacagtatgcaccatcaccatcaccatgatagcaagacaaatattactcagaacggtacg,

to amplify the *cbp* gene of pLF3872. As mentioned previously the new construct had a S-D sequence, *Xba*I and *Sph*I restriction enzyme sites, and an N-terminal hexa histidine tag. The leader peptide sequence was also removed during the latter step. After amplification of DNA, agarose gel electrophoresis was conducted to determine sample quality.

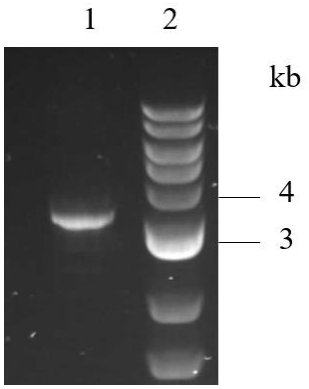


Figure 3.19. Agarose gel electrophoresis result of amplified *cbp* gene encoding for a protein with an N-terminal hexa histidine tag and no leader peptide sequence;1, PCR product; 2, 2-log DNA ladder (NEB).

As can be seen on Figure 3.19, the PCR product is of expected size (3,094 bp). No spurious bands that may affect cloning are present, and thus the product was used for cloning experiments.

3.4.2.1 Cloning experiment

Once amplified, the *cbp* gene was purified to remove PCR contaminants. The PCR product and pBAD33 plasmid vector were digested using enzymes *Xba*I and *Sph*I, after which the samples were gel purified. The insert (*cbp* gene) and vector (pBAD33) were ligated. The ligation mixture was used to transform NEB express *E. coli*, and the clonal isolates were selected and grown to confluence. To determine if transformation had worked, after plasmid extraction, a restriction digestion was carried out before sending the verified pBAD33 plasmid with the *cbp* gene for sample sequencing. Concentration and purity of the DNA were determined using a NanoVue spectrophotometer.

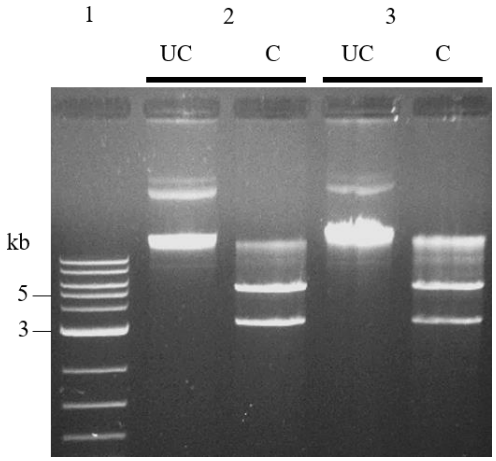


Figure 3.20. Agarose gel electrophoresis image confirming the presence of the *cbp* gene insert with the N-terminal histidine tag; C represents digested sample and UC undigested sample (negative control). The wells are: 1, 2-log DNA ladder (NEB); 2, Sample 1; 3, Sample 2.

Table 3.8. NanoVue spectrophotometer readings of purified pBAD33 plasmid with the *cbp* gene insert used for Sanger sequencing.

Sample 1		**Sample 2**	
Concentration	48 ng/µl	**Concentration:**	60 ng/µl
A260/A280	1.714	**A260/A280**	1.690
A260/A230	2.184	**A260/A230**	1.967

Samples 1 and 2 have the *cbp* gene insert that is of expected size 3094 bp. The digested pBAD33 plasmid is also of the expected size (5.3kb) (Figure 3.20). A band >10 kb can be seen for samples 1 and 2 in Figure 3.20, which is undigested sample caused by incomplete digestion of some of the DNA. The two samples are of good quality and concentrated enough (Table 3.8) for Sanger sequencing, which was used to validate the absence of nucleotide variations.

Range 1: 1 to 3094 Graphics ▼ Next Match ▲ Previous

Score	Expect	Identities	Gaps	Strand
5714 bits(3094)	0.0	3094/3094(100%)	0/3094(0%)	Plus/Plus

Figure 3.21. NCBI BLASTN comparison result of Sanger sequenced *cbp* gene with no leader peptide to that sequenced using Ion Torrent PGM.

No variations were observed for the Sanger sequenced samples, as can be seen in Figure 3.21, with 100 % nucleotide identity to the *cbp* gene sequenced using Ion Torrent PGM.

3.4.3 Cloning and expression of a partial *cbp* gene

According to a study by Deivanayagam *et al* (2000), in other bacteria the A domain of CBP is involved in adhesion. As previous attempts to express recombinant CBP led to degraded protein, a *cbp* gene construct that encoded the A domain only, was made in conjunction with the full *cbp* gene copy. The partial CBP construct lacked a leader peptide sequence and had an N-terminal hexa histidine tag. The gene construct also had *Xba*I and *Sph*I restriction enzyme sites, along with an S-D sequence. PCR was conducted using primers CBPREV3: ctttgcatgctcaatttatagttaaactatgattatatccattag and CBPFOR2:

atatgcttctagaagaaggaggcaacagtatgcaccatcaccatcaccatgatagcaagacaaatattactcagaacggtacg

to amplify the *cbp* gene construct. An agarose gel electrophoresis was conducted to determine sample quality.

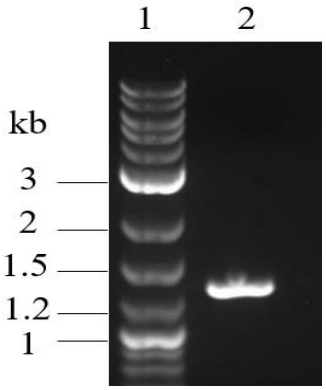


Figure 3.22. Agarose gel electrophoresis result of amplified partial *cbp* gene;1, 2-log DNA ladder (NEB); 2, PCR product.

The PCR product is of expected size (1261 bp) and does not contain any additional bands that may affect cloning. The sample in Figure 3.22 was used for cloning experiments.

3.4.3.1 Cloning the partial *cbp* gene

After PCR, the sample was purified to remove PCR contaminants. The sample and the pBAD33 plasmid were digested by using restriction enzymes *Xba*I and *Sph*I and then gel purified. The purified pBAD33 and *cbp* samples were ligated and used to transform

NEB Express *E. coli*. Transformants were selected and grown to confluence. Plasmid was extracted from the confluent growth and a restriction digestion, along with agarose gel electrophoresis were performed, to check if the extracted sample had the *cbp* gene.

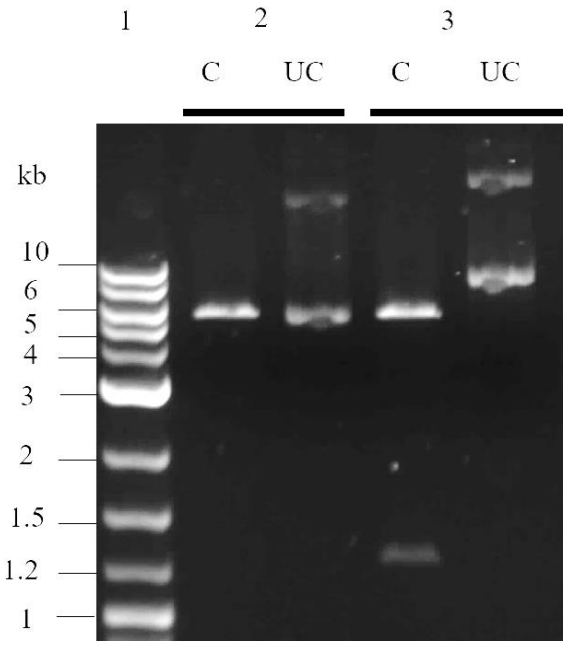


Figure 3.23. Agarose gel electrophoresis result to confirm the presence of partial *cbp* gene insert from extracted plasmid. C represents samples that have been digested, while UC represents undigested samples (negative control). The wells are: 1, 2-Log DNA ladder (NEB); 2, Sample 1; 3, Sample 2.

Table 3.9. NanoVue spectrophotometer reading of purified pBAD33 plasmid with partial *cbp* gene insert.

Sample 2	
Concentration	41.0ng/µl
A260/A280	1.640
A260/A230	2.278

Sample 2 contains the partial *cbp* gene insert of the expected size (1261 bp) (Figure 3.23), and the concentration and purity are sufficient for Sanger sequencing (Table 3.9). Sample 1 does not contain the partial *cbp* gene and therefore was not sequenced.

Range 1: 1 to 1261 Graphics Next Match Previous M

Score	Expect	Identities	Gaps	Strand
2329 bits(1261)	0.0	1261/1261(100%)	0/1261(0%)	Plus/Plus

Figure 3.24. NCBI BLASTN result comparing the Sanger sequenced partial *cbp* gene to the gene sequenced using Ion Torrent PGM.

The NCBI BLASTN result shows that the partial *cbp* gene does not contain any nucleotide variations with a 100 % match to the *cbp* gene sequenced using Ion Torrent PGM (Figure 3.24).

3.4.4 Expression of the *cbp* gene and purification of the CBP protein

Protein expression and purification of the new *cbp* gene constructs that lack a leader peptide sequence and have an N-terminal hexa histidine tag was performed. The latter included expression and purification of the partial *cbp* gene construct, which was to be a backup in case the full copy of the CBP protein was degraded. Initially, a small-scale purification was conducted using the Magne-His Kit. This was done to allow further optimisation of protein expression if CBP was degraded. Small fractions of samples were taken at each of the protein expression and purification steps in case of experiment failure.

3.4.4.1 Small scale protein purification

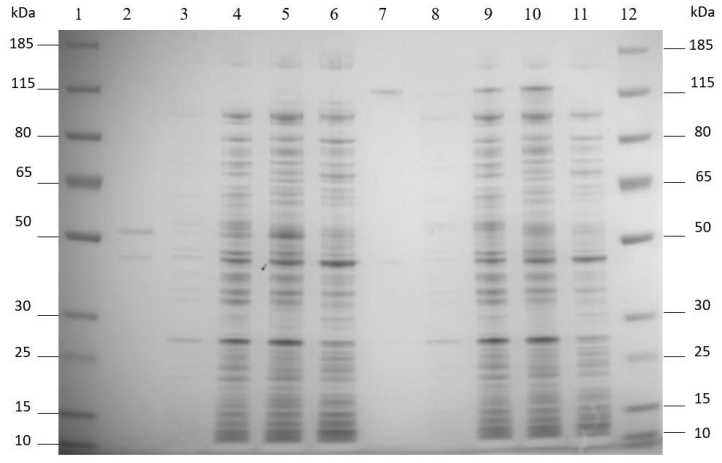


Figure 3.25. Coomassie staining of the recombinant CBP proteins with the leader peptide removed and an N-terminal hexa histidine tag. The wells labelled 2–6 are related to partial CBP purification, while those from 7–11 are for full CBP. The wells are: 1, pre-strained protein ladder (Page ruler plus); 2, eluate partial CBP; 3, column wash 1/2/3 combined; 4, lysate flow through; 5, lysate after induction; 6, lysate before

induction; 7, eluate CBP; 8, column wash 1/2/3; 9, flow through; 10, lysate after induction; 11, lysate before induction; 12, pre-strained protein ladder (Page ruler plus).

Figure 3.25 shows that the full copy of the CBP protein is not degraded and is 115 kDa in size. The protein is larger than the predicted 111 kDa, which is commonly observed among large cell surface proteins due to conformational properties that are not considered when predicting a protein's size. Eluate (well 7) contains the CBP protein, with no observable spurious bands.

The partial CBP protein is around 55 kDa and is also larger than the expected protein size of 47 kDa (Figure 3.25, well 2). There is also a light band below the 55 kDa band in well 2. This may be either non-specific protein or the CBP protein.

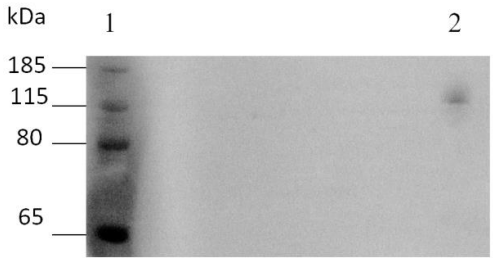


Figure 3.26. Western blot of CBP protein without the N-terminal hexa histidine tag and no leader peptide sequence; 1, pre-stained protein ladder (Page ruler plus); 2, purified CBP protein.

The western blot (Figure 3.26) confirmed that the CBP protein was around 115 kDa in size and that the CBP protein had been eluted by the Magne-His Kit.

3.4.4.2 Large-scale protein purification

Since small scale protein purification using the Magne-His Kit isolated CBP protein without indications of degradation, a large-scale protein purification was conducted. Samples were taken at each step of the protein expression and purification process and loaded into the wells of an SDS polyacrylamide gel for Coomassie staining to determine if the CBP protein had been purified.

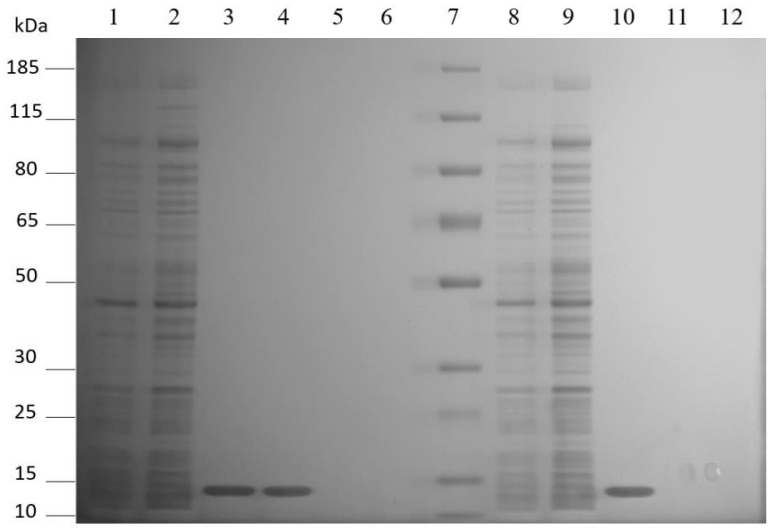


Figure 3.27. Coomassie staining of large-scale CBP expression and purification. The wells labelled from 1–6 represent purification of full CBP, and wells labelled from 8–12 represent partial CBP. The wells are: 1, lysate before induction; 2, lysate after induction; 3, clarified lysate; 4, column flow through; 5, CBP eluate 1; 6, CBP eluate 2; 7, pre-stained protein ladder (Page ruler plus); 8, lysate before induction; 9, lysate after induction; 10, column flow through/wash mix; 11, partial CBP eluate 1; 12, partial CBP eluate 2.

The clarified lysate representing *E. coli* after lysis using a custom lysozyme solution (well 3) showed that all the proteins were degraded (Figure 3.27). Degradation of protein was also observed for well 4 and 10, containing samples, column flow through representing lysate that had passed through the column. The lysates before and after induction had no noticeable degradation, as can be seen in wells 1, 2, 8 and 9 (Figure 3.27). The custom lysis solution may have been contaminated with proteases, leading to fragmentation of the proteins as no degradation was observed before the protein purification step, which used NuPAGE LDS buffer for lysis. No CBP protein was eluted.

3.4.4.3 Large-scale protein purification with Fastbreak Lysis solution from Promega

Previous experiments with purification of recombinant CBP indicated that the lysis solution might have degraded the protein. Thus, the Fastbreak lysis buffer supplied by the Magne-His Kit (Promega) was used for large-scale protein purification. Fractions of the samples were taken at each step of the expression and purification protocol. Equal

amounts of the samples, relating to the number of cells, were loaded to each well of a polyacrylamide gel for Coomassie staining to help determine induction in the lysate.

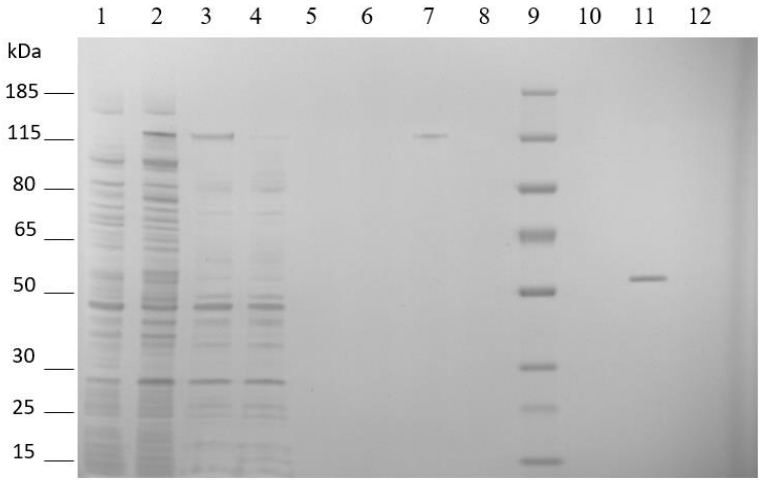


Figure 3.28. Coomassie staining of large scale CBP protein containing an N-terminal hexa histidine tag and no leader peptide. Lysis was conducted using the FastBreak (Promega) lysis solution. The wells labelled from 1–8 are samples from CBP protein expression and purification,10–12 contain partial CBP samples. The wells are: 1, lysate before induction; 2, lysate after induction; 3, clarified lysate; 4, column flow through; 5, column wash 1/2; 6, CBP eluate 1; 7, eluate 2; 8, eluate 3; 9, pre-strained protein ladder (Page ruler plus); 10, partial CBP eluate 1; 11, partial CBP eluate 2; 12, partial CBP eluate 3. The figure was obtained from Lehri, Seddon and Karlyshev (2017c).

Both the partial and full copy of the CBP proteins were successfully purified with no observable spurious bands. The size of the full CBP protein was 115 kDa (well 7, Figure 3.28), which was larger than the predicted 111 kDa, and was similar to that observed during small-scale purification with the Magne-His Kit. The column flow through (well 4) and wash (well 5) did not contain a strong 115 kDa band. The partial CBP protein was also eluted and was slightly larger than 55 kDa, and bigger than that predicted computationally (47 kDa). Both the full CBP and partial CBP proteins were acceptable for downstream adhesion experiments.

3.5 BCA assay to detect protein concentration

After successful protein purification, a Pierce Protein BCA assay kit was used to determine the concentration of both the partial and full copy CBP proteins.

The concentration of the purified CBP was 160 µg/ml, while the concentration of the purified partial CBP protein was 260 µg/ml. Both proteins were concentrated enough for multiple adhesion experiments.

3.6 Enzyme-linked immune absorbance assay results

ELISA was conducted to determine whether the recombinant CBP protein could bind to collagen I.

3.6.1 Optimisation of ELISA

Initially, ELISA was carried out on 96 well UV-treated Nunclon Delta polystyrene plates (ThermoFisher). Test wells were coated with 0.36 µg/well collagen I (Sigma) and control wells were coated with 0.36 µg/well BSA (Acros Organics), 0.36 µg/well was used as the surface area of each well was 0.36 cm^2. Control wells would help determine non-specific interaction between CBP and collagen I. Standard ELISA protocol was followed as described in section 2.15 with each well incubated with 0.36 µg of CBP.

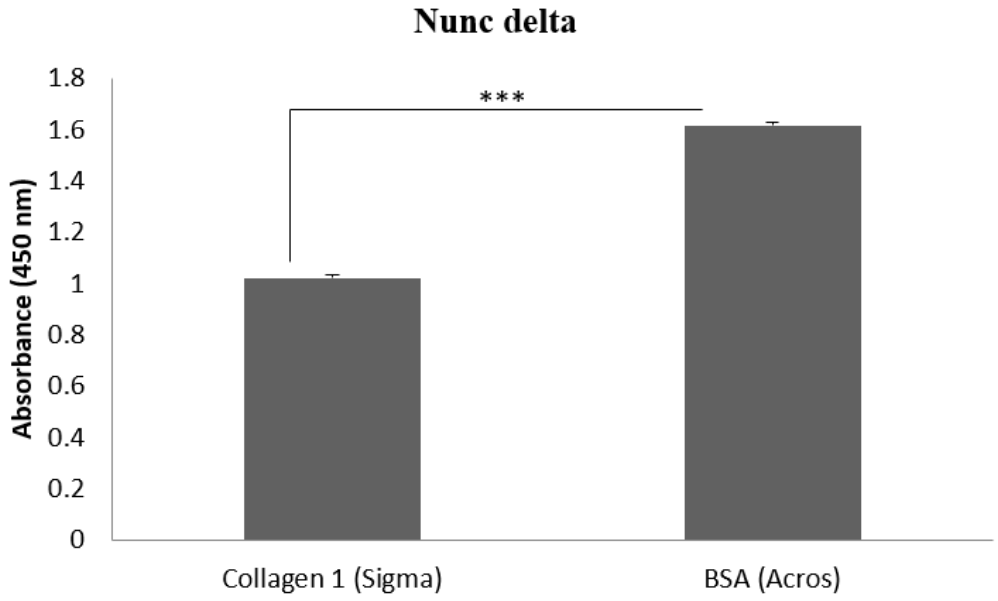


Figure 3.29. CBP adhesion to collagen I, using UV-treated Nunclon Delta plates. The antibodies used were 6x-His Epitope Tag monoclonal primary antibody and HRP-linked secondary polyclonal antibody. A single biological replicate was conducted consisting of three technical replicates (N = 3).

The BSA coated wells had a higher absorbance than that for collagen I coated wells (Figure 3.29). The BSA (Acros Organics) used for control wells may have had

extracellular matrix protein (ECM) contaminants such as various types of collagen, fibrinogen, and fibronectin, and thus a different type of BSA may be required. Another reason for having a higher absorbance for control wells as opposed to test wells could be due to the type of plate used. Since UV treated plates have a higher binding affinity, as there are multiple ways in which a protein can interact with the plate, such as ionic and hydrophobic interactions, collagen I may have been coated in such way that potential collagen binding sites were not exposed on the plate.

3.6.1.1 Further optimisation experiments

Since the absorbance of the negative control wells for the previous experiment was higher than that of the test wells, further optimisation experiments were carried out. Two 96 well polystyrene plates from Corning were used. One was UV treated (High Bind Corning Costar) and the other was untreated (Med Bind Corning Costar). Treated and non-treated plates were used to determine whether the type of plate was causing a lower absorbance for test wells, as protein coating on a non-treated plate is designed for large molecules, such as collagen I. Two types of BSA were tested on the untreated plate, one of which was from Acros Organics (previously used) and the other was of 99% purity from Sigma. For the UV treated plate, only BSA from Sigma was used. This would help determine whether the purity of BSA was causing an increase in absorbance for negative control wells.

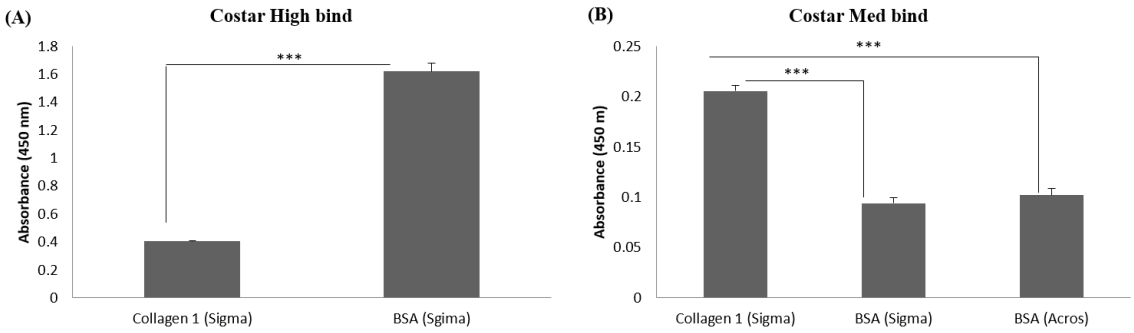


Figure 3.30. Identifying optimal plates to use for future ELISA experiments. (A) UV treated Costar High bind plates, and (B) untreated Costar Med-Bind plate. Each plate was coated with collagen I and BSA, which were incubated with CBP (0.36 µg). The antibodies used were 6x-His Epitope Tag monoclonal primary antibody and HRP-linked secondary polyclonal antibody. For each experiment, one biological replicate, consisting of three technical replicates, was conducted (N = 3).

As can be seen in Figure 3.30 (A), the absorbance of the control wells for the UV treated plate (High Bind Corning Costar) was higher than that of the test wells, even when using a different BSA (Sigma). The absorbance of the negative control for the untreated plate (Med Bind Corning Costar) was lower than that of the test wells (Figure 3.30, B). There was also no difference in absorbance between the two types of BSA used for the untreated plate. Thus, the likely cause of having a higher absorbance for the negative control wells was the type of plate used. For future experiments, untreated plates were used for coating collagen I.

3.6.1.2 Optimisation of diluent for coating wells

As collagen I (Sigma) was initially prepared in 0.1 M acetic acid with a final pH of 3.4, and the recommended ELISA coating buffer has a pH of 9.6, a test ELISA experiment was conducted to determine whether the type of diluent used for coating plates would influence collagen coating and thus binding experiments. To obtain a pH of around 3.4, collagen I was diluted in water, while ELISA coating buffer was used as a diluent for comparative purposes.

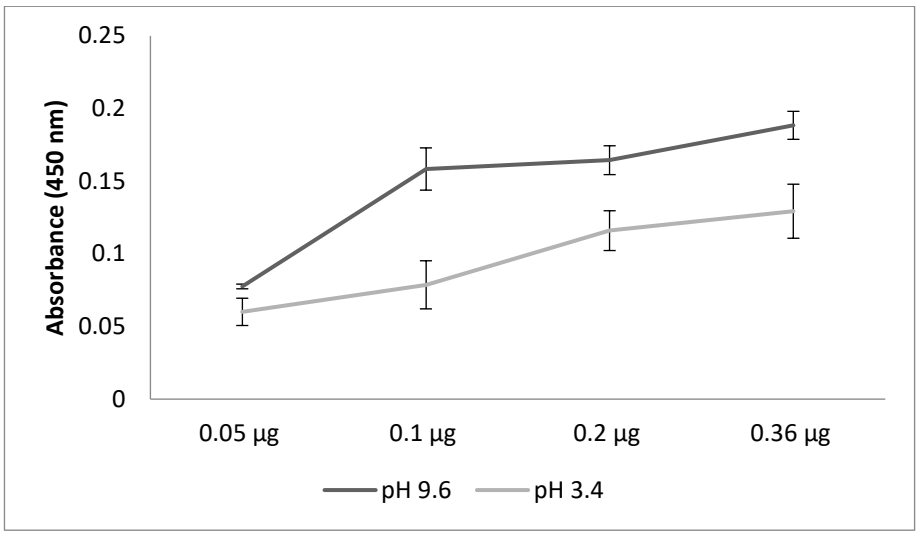


Figure 3.31. ELISA experiment to determine the effect of collagen I coating with varying diluents. The antibodies used were 6x-His Epitope Tag monoclonal primary antibody and HRP-linked secondary polyclonal antibody. A single biological experiment was conducted consisting of three technical repeats (N = 3).

The absorbance values for plates coated with ELISA coating buffer (pH 9.6) was higher than that of distilled water (pH 3.2) (Figure 3.31). For future experiments, ELISA coating buffer was used due to the increased absorbance observed at higher pH.

3.6.2 CBP adhesion to collagen I

After optimisation, to confirm whether the CBP protein binding to collagen I was concentration dependent, an ELISA-based experiment was conducted. Wells were coated with 0.36 µg/well of collagen I and BSA (Sigma). To determine binding to collagen I, a difference in absorbance between varying amounts of CBP (0.05 µg/well, 0.1 µg/well, 0.2 µg/well and 0.36 µg/well) was used. To differentiate between actual and non-specific interactions, BSA (Sigma) coated wells were used as a negative control.

3.6.2.1 CBP binding assay

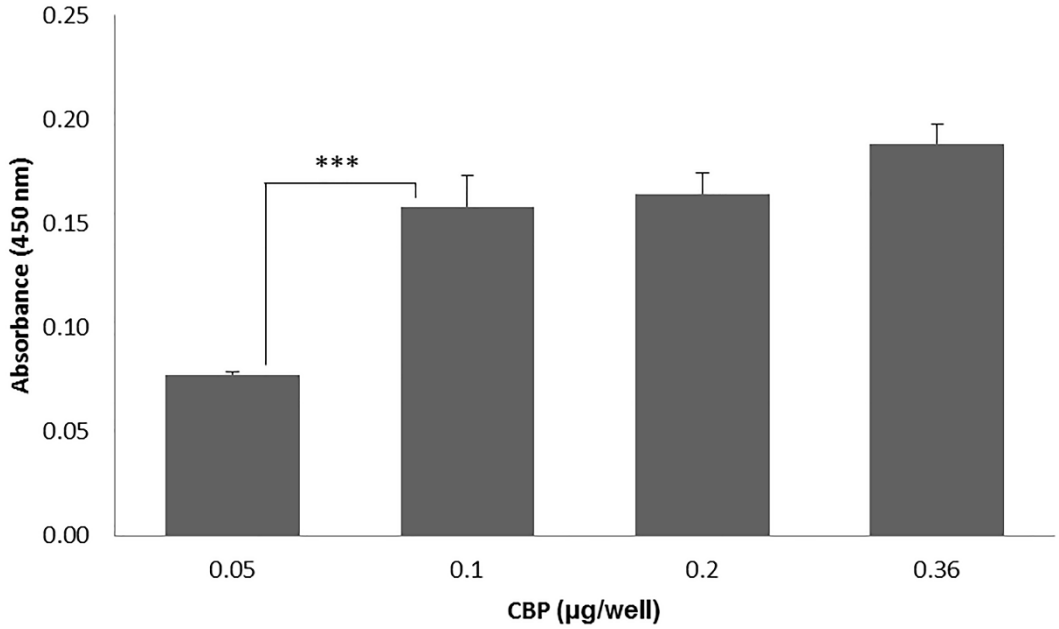


Figure 3.32. Determining CBP concentration dependency in binding to collagen I. The antibodies used were 6x-His Epitope Tag monoclonal primary antibody and HRP-linked secondary polyclonal antibody. The data contains two biological repeats with three technical repeats (N = 6). The figure was obtained from Lehri, Seddon and Karlyshev (2017c).

As mentioned in Lehri, Seddon and Karlyshev's (2017c) article, the CBP protein binds to collagen I in a concentration dependent manner (Figure 3.32). The absorbance when using 0.1 µg/well of CBP is higher than that of 0.05 µg/well. No difference in absorbance was observed between concentrations of 0.1 µg/well to 0.36 µg/well of CBP.

3.6.2.2 *C. jejuni* binding assay

Binding experiments were conducted to discover whether *C. jejuni* 11168H and 81–176 strains attached to collagen I. Varying amounts of *C. jejuni* were added to each well, 2 x 10^8, 2 x 10^7 and 2 x 10^6 CFU/well to see if there was concentration dependency and to find the most useful amount of *C. jejuni* to use for competition experiments.

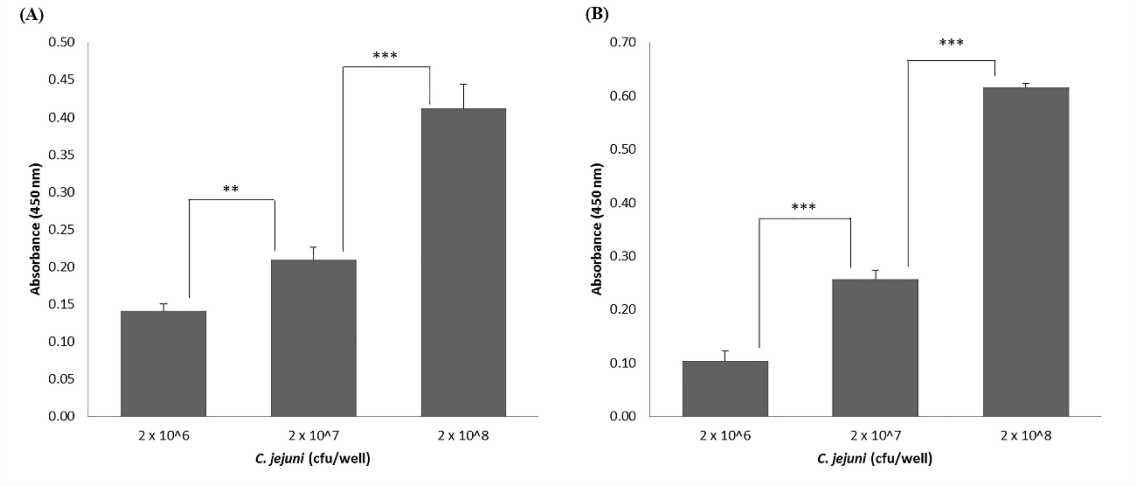


Figure 3.33. Identifying *C. jejuni* strains 81–176 (A) and 11168H (B) concentration dependency in binding to collagen I. The antibodies used were *C. jejuni* monoclonal primary antibody and goat anti-mouse polyclonal secondary antibody. The data consists of two biological repeats, each having three technical repeats (N = 6) except for (A), 2 x 10^8 where N = 5. The figure was obtained from Lehri, Seddon and Karlyshev (2017c).

Figure 3.33 shows that both *C. jejuni* 11168H and 81–176 bind to collagen I in a concentration dependent manner. For future competition experiments, 2 x 10^7 CFU/well of *C. jejuni* were used, as this utilised a minimal number of cells yet had a high enough absorbance to detect changes. Using 2 x 10^7 CFU/well of *C. jejuni* cells would also be useful for obtaining the required ratios of *C. jejuni* to *L. fermentum* 3872 of 2:1, 1:1, 1:5, 1:10 for whole cell competition assay experiments.

3.7 Competition assay

3.7.1 Reduction of *C. jejuni* adhesion to collagen I in the presence of CBP

Binding experiments showed that both the CBP protein and *C. jejuni* 11168H and 81–176 strains bind to collagen I. Thus, competition experiments were carried out to determine if the CBP protein could inhibit *C. jejuni* strains 11168H and 81–76 binding to collagen I.

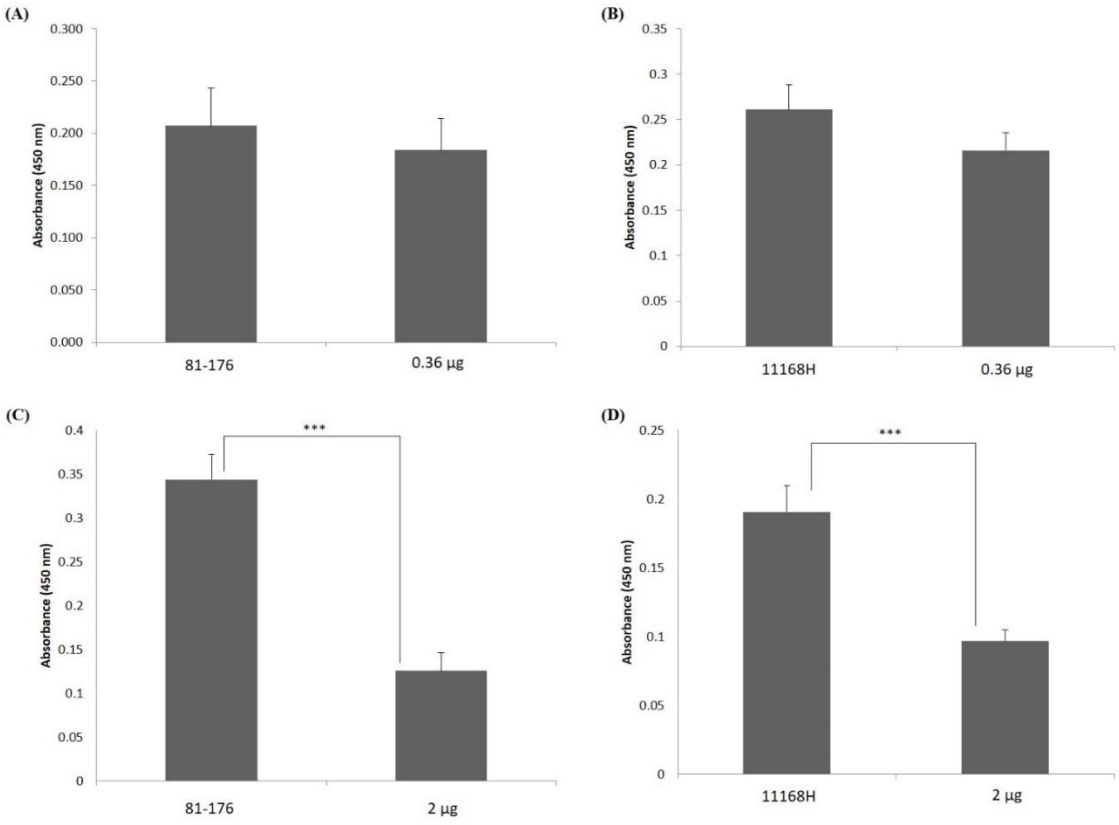


Figure 3.34. ELISA experiment showing the effect of CBP on *C. jejuni* 11168H and 81–176 binding to collagen I. The bars labelled 81–176 and 11168H only contain *C. jejuni* strains 11168H (B and D) and 81–176 (A and C) of 2 X 10^7 CFU/well. The bars labelled 0.36 µg and 2 µg contain a mixture of 0.36 µg or 2 µg CBP and either *C. jejuni* 11168H (B, D) or 81176 (A, C). The antibodies used were *C. jejuni* monoclonal primary antibody and goat anti-mouse polyclonal secondary antibody. Graphs A and B have data from two biological repeats each with three technical repeats (N = 6), while C and D consists of three biological repeats, each with three technical repeats (N = 9). The figure was obtained from Lehri, Seddon and Karlyshev (2017c) with additional graphs (A) and (B) added.

An amount of 0.36 µg/well of CBP does not reduce absorbance of *C. jejuni* 11168H and 81–176 (Figure 3.34; A and B). Increasing the amount of CBP to 2 µg/well causes a decrease in *C. jejuni* 11168H and 81–176 absorbance, inhibiting *C. jejuni* binding to collagen I (Figure 3.34; C and D).

3.7.2 Reduction in *C. jejuni* adhesion to collagen I in the presence of *L. fermentum* 3872

As competition experiment between CBP and *C. jejuni* strains 11168H and 81–176 showed inhibition of *C. jejuni* binding to collagen I, a competition experiment was also carried out to determine whether *L. fermentum* 3872 whole cells could inhibit *C. jejuni* 11168H and 81–176 binding to collagen I.

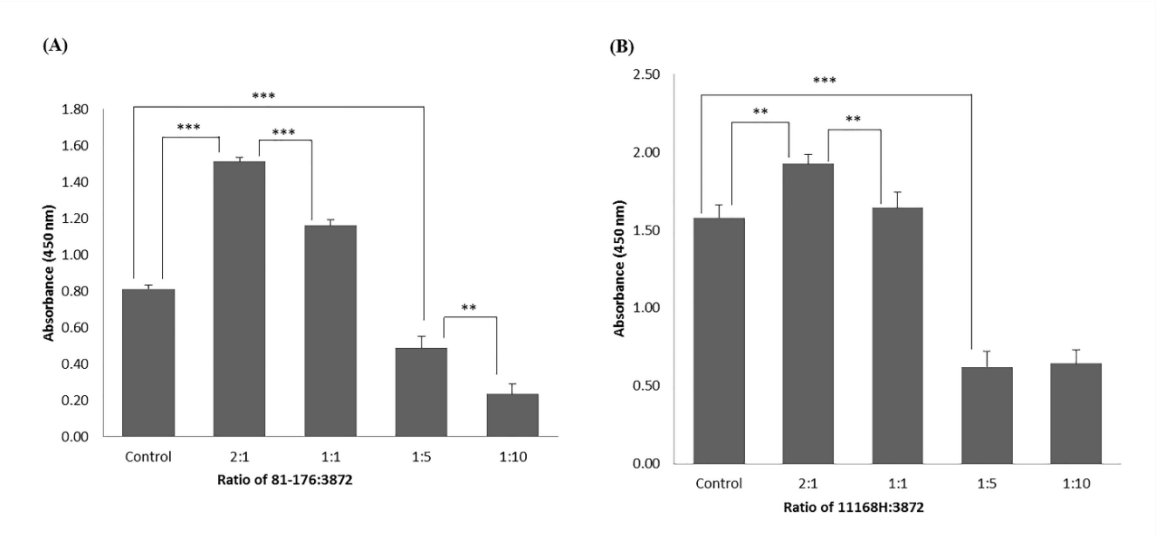


Figure 3.35. Whole cell competition experiments between *L. fermentum* 3872 and *C. jejuni* strains 81-176 (A) and 11168H (B) in binding to collagen I. Bars labelled Control have 2 x 10^7 CFU/well *C. jejuni* strains 11168H or 81–176. The cell to cell ratios shown on the graph are after adding a suspension mixture of bacteria to each well. The antibodies used were *C. jejuni* (PEB1) polyclonal primary antibody and goat-anti-rabbit IgG polyclonal HRP conjugated secondary antibody. Two biological repeats were conducted, each with three technical repeats (N = 6), apart from control (B) and 1:10 (B), for which three biological repeats were conducted, each with three technical repeats (N = 9). The figure was obtained from Lehri, Seddon and Karlyshev (2017c).

The absorbance observed for the *C. jejuni* 11168H and 81–176 wells increases when a ratio of *C. jejuni* to *L. fermentum* 3872 of 2:1 is used (Figure 3.35). The increase in absorbance may be a result of aggregation (further described in the discussion section), although more research is needed to adequately prove the cause of the increased absorbance. When increasing the amount of *L. fermentum* 3872, a decrease in absorbance is seen for *C. jejuni* 11168H and 81–176 wells (Figure 3.35).

3.8 Identifying proteins contributing to collagen I adhesion for *C. jejuni* strains 11168H and 81-176

Genome sequencing identified multiple genes potentially involved in adhesion for *L. fermentum* 3872, including those involved in collagen binding. For *C. jejuni* strains 11168H and 81–176, genome sequence analysis could not identify collagen binding proteins. Therefore, to determine the type of proteins utilised by *C. jejuni* strains 11168H and 81–176 for adhesion to collagen I, a Co-IP was conducted with the lysates of *C. jejuni* strains, followed by mass spectrometry analysis.

3.8.1 Coomassie staining of the Co-IP eluates

Initially, Coomassie was used to stain the eluted proteins of *C. jejuni* strains 11168H and 81–176 that interacted with collagen I coated beads.

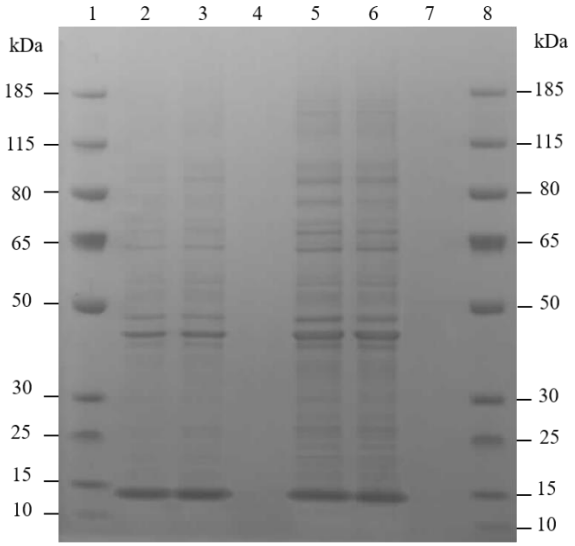


Figure 3.36. Coomassie stained gel picture of *C. jejuni* strains 11168H and 81–176 Co-IP eluates; 1, pre-stained ladder (Page ruler plus); 2, 81–176 column flow through; 3, 81–176 lysate; 4, 81–176 Co-IP eluate; 5, 11168H column flow through; 6, 11168H lysate; 7, 11168H eluate; 9, pre-stained ladder (Page ruler plus).

No bands were detected for the Co-IP eluates of *C. jejuni* strains 11168H and 81–176 (wells 4 and 7; Figure 3.36). The lysate and flow through wells showed that the proteins were not degraded. Either no protein was present in the eluates, or Coomassie staining was not sensitive enough to detect the minute quantity of eluted protein. The Co-IP protocol suggested using a more sensitive protein staining approach.

3.8.2 Silver staining of the Co-IP eluates

Silver staining was used to determine if there was any detectable protein eluted from collagen I coated beads, incubated with the lysates of *C. jejuni* strains 11168H and 81–176. Silver staining is far more sensitive than Coomassie staining in detecting minute quantities of protein.

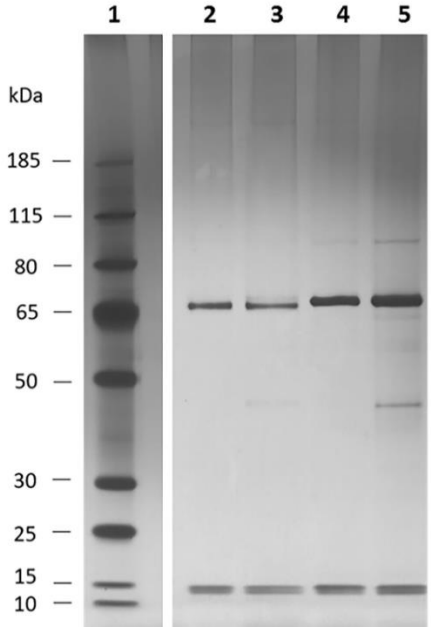


Figure 3.37. Silver stain of *C. jejuni* 11168H and 81–176 Co-IP eluates; 1, pre-stained ladder (Page ruler plus); 2, 11168H eluate after one hour of incubation; 3, 11168H eluate after three hours' incubation; 4, 81–176 eluate after one hour of incubation; 5, 81–176 eluate after three hours' incubation. The figure was obtained from Lehri, Seddon and Karlyshev (2017c).

The silver stain analysis detected two major bands for both *C. jejuni* strains 11168H and 81–176 at around 65 kDa and 15 kDa (Figure 3.37). The Co-IP experiment was repeated twice with similar results. There were also light bands for the *C. jejuni* strains tested, namely after a three-hour incubation, which was likely caused by nonspecific interactions during Co-IP experiment. Nonspecific interactions are common in Co-IP experiments, particularly for long-term incubations, as stated in the manufacturer's protocol. The two most apparent bands, at 65 kDa and 15 kDa, were extracted and sent for mass spectrometry analysis (LC-MS/MS).

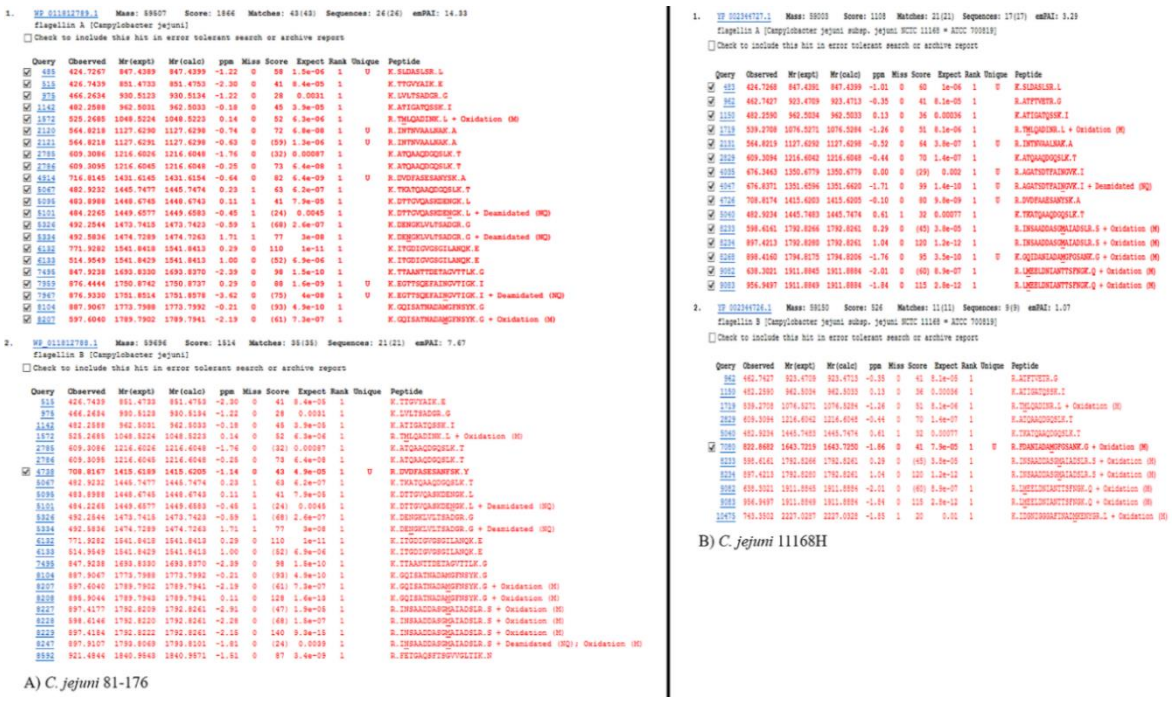


Figure 3.38. Screenshots of Mascot software hits for mass spectrometry (LC-MS/MS) data output. The data shows the major hits for the 65 kDa protein eluted during Co-IP for *C. jejuni* strains 81–176 (A), 11168H (B). The figure was obtained from Lehri, Seddon and Karlyshev (2017c).

Analysis of mass spectrometry data revealed that, for both *C. jejuni* strains 11168H and 81–176, the 65 kDa band was affiliated with flagellin subunits FlaA and FlaB (Figure 3.38). The estimated size of FlaA and FlaB was smaller (60 kDa) than the silver stain result (Figure 3.37). This is likely a result of O-linked glycosylation of the protein in *Campylobacter*, which can alter the mobility of protein in SDS gels (Hitchen *et al.*, 2010). Mass spectrometry data analysis of the 15 kDa protein using Mascot software had partial amino acid hits to FlaA and FlaB proteins of *C. jejuni* strains 11168H and 81–76. The proteins of 15 kDa in size may potentially be fragmented FlaA and FlaB proteins.

3.9 Inhibition of *C. jejuni* 11168H and 81-176 growth by *L. fermentum* 3872 supernatant

An agar well diffusion assay was conducted to determine whether *L. fermentum* 3872 produces antimicrobials that may inhibit *C. jejuni* strains 11168H and 81–176 growth *in vitro*.

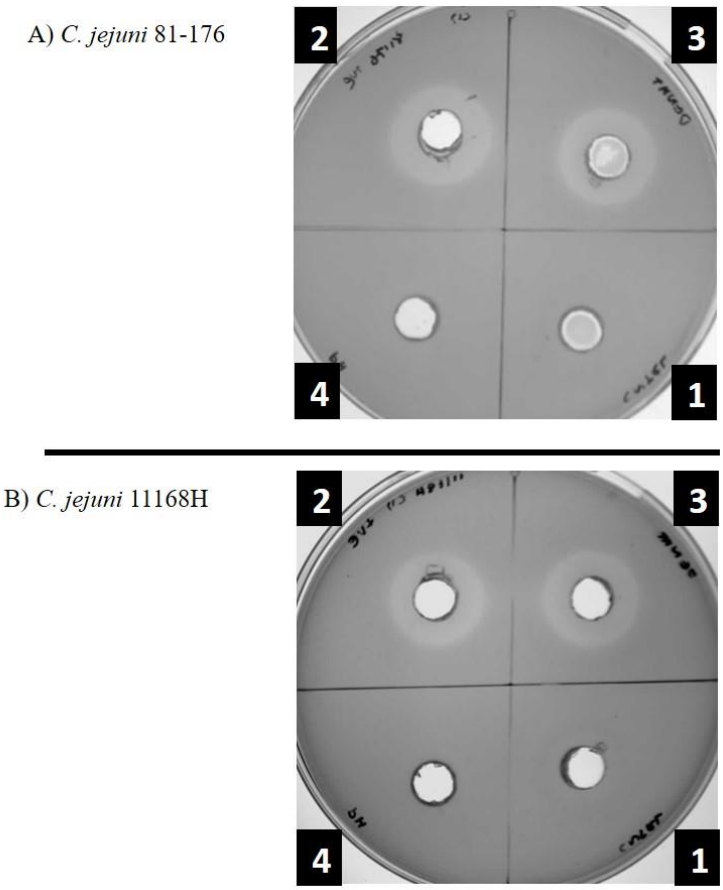


Figure 3.39. Agar well diffusion assay result for *C. jejuni* 11168H and 81–176 grown in the presence of *L. fermentum* 3872 cell free culture supernatant; 1, M.R.S broth (control); 2–4, *L. fermentum* 3872 cell free supernatant; 2, untreated; 3, heat treated; 4, pH adjusted. The figure was obtained from Lehri, Seddon and Karlyshev (2017c).

C. jejuni growth inhibition was observed in wells with low pH (Figure 3.39). The diameter of the zone of inhibition for *C. jejuni* 11168H with *L. fermentum* 3872 cell free supernatant was 21 mm (SD±1), and was 21 mm (SD± 0.5) for *C. jejuni* 81–176. After boiling *L. fermentum* 3872, the zone of inhibition for *C. jejuni* 11168H was 21 mm (SD ±0.5) and for *C. jejuni* 81–176 was 21 (SD± 0.4). No growth inhibition was observed in control wells or in wells where the pH of the *L. fermentum* 3872 supernatant was

adjusted to that of the control wells. There was no difference in the diameter of zones of inhibition between the *C. jejuni* strains tested and between boiled and not boiled *L. fermentum* 3872 supernatants.

Chapter 4: Discussion

4.1 Genome sequencing issues and possible solutions

During the completion of the genome sequence for *L. fermentum* 3872, it was determined that multiple assembly errors could not be successfully identified due to an over reliance on read mapping as a quality indicator. This can be a major issue for many draft genome assemblies deposited in databanks, and for comparative genomic studies that assume good genome quality. For unpaired reads generated by the Ion Torrent PGM, assembly verification can be conducted using sequence scaffolding tools such as Contiguator (Galardini *et al.*, 2011), given that there is minimal genotypic variation between the query and reference sequence. This method is also dependent on the quality of the reference sequence assembly. As an example, during this study, Contiguator (Galardini *et al.*, 2011) detected multiple alignment variations between the query sequence (*L. fermentum* 3872) and that of the reference sequence (*L. fermentum* IFO 3956) thus indicating a misassembly of the query sequence (Figure 3.5). When splitting the correctly assembled *L. fermentum* 3872 chromosomal sequence at 100,000 bp points, Contiguator alignment to the reference sequence had improved, indicating a better assembly (Figure 3.11). Furthermore, the Ion Torrent PGM is able to generate mate pair reads, which are capable of sequencing two ends of a fragmented DNA of 400 bp in length, with a 20kb sequence gap between the sequenced ends. Mate pair read data, in combination with unpaired reads, would enable identification of contig location for scaffolding and genome sequence completion. The drawback of this technique is that the cost of generating mate pair reads is higher than that of a PacBio RSII sequencing run, which can sequence nucleotide fragments of >14kb in length without gaps (Reuter, Spacek and Snyder, 2015). Furthermore, combining mate pair and unpaired read data requires much more computational effort than combining long PacBio RSII generated reads with short reads.

Analysis of the misassembled contigs indicated that assembly errors could be dependent on the type of *de novo* assembly tool used, as certain *de novo* assemblers had more assembly errors than others (Table 3.4). The sequencing region being assembled may also give rise to misassembly as certain sequenced regions were misassembled regardless of the *de novo* assembly tools used. Misassembly predominantly happened within or near repetitive regions (for example, transposons or RNA encoding regions), as shown in Figure 3.7. Many *de novo* assemblers do not extend a sequence beyond repetitive elements, given short reads are being used. Thus, it would be useful not to fill in sequencing gaps unless information on the contig's location/orientation in relation to other contigs is known, along with the information on the gap size between contigs. If such information is known and the gap size is not large, then BLASTN extension, or preferably Sanger sequencing, may be used to fill in sequencing gaps.

An optical map can be a useful sequence assembly validation tool. The optical map of the genome is independent of nucleotide sequence data, which removes sequencing bias and thus provides greater confidence in the assembly quality. As described previously, all *de novo* assemblers had difficulty assembling specific genomic sequence regions, and thus an assembler that utilises optical map data to guide sequence assembly could be useful in accurate genome assembly (Lin *et al.*, 2012). Optical maps are currently helpful in genome sequences of over 30 kb in size, which could be improved by introducing more restriction enzymes and thus improving the resolution of the map generated. Furthermore, as genome sequencing technology is focusing on large, low-quality read data, an optical map would be useful in differentiating between contigs and reads that have artefactually been categorised as contigs by *de novo* assemblers, which was the case with the remaining five sequences of the PacBio RSII output. A major issue with bacterial genome sequencing is that bacterial colonies tend to have genotypic variation, which makes it difficult to determine whether the sequenced clonal isolate is

truly representative of the microbe being administered for probiotic use (Walsh *et al.*, 2016). For this study, two clonal isolates were sequenced at confluence and no variations were found. Nonetheless, re-sequencing after administration of *L. fermentum* 3872 to a host or after prolonged storage would be advisable to determine any genotypic change.

4.2 The chromosomal sequence of *L. fermentum* 3872

As mentioned in the results section, the chromosomal sequence of *L. fermentum* 3872 consisted of multiple vitamin B coding genes. These genes may be useful for infants, as the bacteria was isolated from a healthy human female's milk (Abramov *et al*., 2014). There are several genes and metabolic pathways for pH adjustment and acid tolerance that could play a role in *L. fermentum* 3872 survival within the acidic environment of the GIT. Acid tolerance would enable the probiotic to be ingested and to reach its destined site of action, as well as to colonise the host to provide its desired effect. The *L. fermentum* 3872 genome also contains a group of genes involved in exopolysaccharides (EPS) synthesis. In other lactic acid bacteria, EPS are involved in bacterial protection against toxic compounds and desiccation, thus playing an important role in bacterial survival (Patel, Majumder and Goyal, 2009). EPS also have host cholesterol-lowering properties and have been known to promote bacterial adhesion and biofilm formation (Mozzi *et al.*, 2009).

Another gene located near a cluster of EPS genes is *epsH*, which is not found in other *L. fermentum* strains used for comparative analysis. A study conducted by Jones *et al*. (2014) showed that the protein encoded by *epsH* plays a role in protection against inflammation induced by *Citrobacter rodentium* in mice. The study stated that the EpsH protein may reduce inflammation and help combat colitis induced by enteric pathogens via immunomodulation by affecting TLR4 (Jones *et al.*, 2014). Thus, *L. fermentum* 3872 may be a candidate for studies involving inflammatory bowel diseases.

Other survival genes that encode DnaK and GroEL have also been associated with osmotic and heat shock resistance (Prasad, McJarrow and Gopal, 2003), and have been known to be overly expressed in the cell wall region of bacteria that bind to mucin (Izquierdo *et al.*, 2009). The GroEL encoding gene of *Lactobacilllis johsonii* La1 is known to cause *H. pylori* aggregation, which in turn may aid in the clearance of *H. pylori*, mediated by mucus flushing (Bergonzelli *et al.*, 2006). Genes that encode D-Lactate dehydrogenase, and 6-phosphogluconate dehydrogenase, have also been known to be overexpressed in *Lactobacillus* species that bind to mucin (Izquierdo *et al.*, 2009). The mucus binding protein of *L. fermentum* 3872 has been annotated as partial, this is due to the fact that the protein has a leucin start codon as opposed to the more commonly used methionine, thus verification is needed to determine whether the protein is truly partial. As mentioned in the result section a gene that encodes the internalin J-like precursor (inlJ) has a mucin binding (MucBP) domain and may promote bacterial adhesion to mucin (Sabet *et al.*, 2005; Juge, 2012). A Translational elongation factor gene similar to that found for *L. fermentum* 3872 has also been known to contribute in adhesion to intestinal Caco-2 cells under low pH (Granato *et al.*, 2004). The above mentioned adhesion genes may be useful candidates for future adhesion and competitive exclusion studies. Proteins that bind to mucin are important in that mucin is found in the lining of the GIT and is often one of the first points of contact for many pathogenic bacteria for GIT colonisation of the host (Oelschlaeger, 2010). *L. fermentum* 3872 also has a gene that encodes an aggregation substance precursor protein, which has been associated with bacterial aggregation and adhesion (Suessmuth *et al.*, 2000), and thus may be useful for competitive exclusion studies. An enolase encoding gene is known to promote adhesion to collagen (Salzillo *et al.*, 2015) and may potentially work synergistically with the CBP protein. As mentioned in the results section a gene encoding a fibronectin binding protein was also found. The protein may promote *L.*

fermentum 3872 adhesion to fibronectin (Muñoz-Provencio, Pérez-Martínez and Monedero, 2010), an ECM protein much like collagen. These genes may play a role in *L. fermentum* 3872 host colonisation, survival, and be potential candidates for future competition experiments. There are also genes present for iron acquisition, which may play a role in inhibiting the growth of other bacteria as iron acquisition is important for bacterial growth (Oelschlaeger, 2010). Many of the antibiotic-resistance genes identified in *L. fermentum* 3872 are commonly found in other bacteria, such as genes involved in beta-lactamase resistance, tetracycline resistance, and fluoroquinolone resistance. Thus *L. fermentum* 3872 can be widely used against pathogens with minimal concern about antibiotic-resistance gene transfer.

Comparative genomic analysis revealed that many of the unique regions of *L. fermentum* 3872 have large areas consisting of prophage-related genes, which was confirmed by PHAST software (Zhou *et al.,* 2011) identifying four intact prophage-related regions. The regions seem to be newly incorporated due to the sudden G+C change compared to that of the surrounding genomic area and is a result of horizontal gene transfer. Prophages can be induced under specific conditions to produce bacteriophages that may be utilised to tackle pathogenic bacteria. As mentioned in the introduction, the use of bacteriophages is promising in reducing *C. jejuni* numbers in poultry, but *C. jejuni* can become resistant to bacteriophage therapy (Kaakoush *et al.,* 2015). Thus, a combination of antimicrobials or phages may be useful for combating *C. jejuni* infections.

A class III bacteriocin unique to *L. fermentum* 3872 was also identified. Class III bacteriocins tend to be large, thermolabile and have a broad antimicrobial spectrum (Ben Lagha *et al.*, 2017). They are divided into two subclasses: type IIIa and type IIIb (Ben Lagha *et al.*, 2017), type IIIa is bacteriolytic and catalyses the hydrolysis of

peptidoglycan. Type IIIb bacteriocins are non-lytic but may cause membrane leakage, prevent sugar uptake or inhibit DNA synthesis (Ben Lagha *et al.*, 2017). According to BAGEL3 and BACTIBASE software, the closest match to the bacteriocin of *L. fermentum* 3872 was enterolysin A from *Enterococcus faecalis*, with a query coverage of 40% and identity of 14%. A study has shown that enterolysin A isolated from *E. faecalis* LMG 2333 is bacteriolytic and may inhibit growth of particular *Enterococci*, *Lactococci*, *Lactobacilli* and *Pediococci* bacteria (Nilsen, Nes and Holo, 2003). Based on similarity to enterolysin A, the bacteriocin from *L. fermentum* 3872 may also be a class IIIa bacteriocin.

4.3 Information on the plasmid pLF3872

The G+C content of the plasmid pLF3872 is lower (40.1%) than that of the chromosomal sequence of *L. fermentum* 3872 (50.1%), indicating that pLF3872 has recently been acquired. The presence of multiple conjugation genes mentioned previously suggests that the plasmid has conjugative ability, and that conjugation may be a possible means of plasmid acquisition. Conjugation may pose a risk in transfer of genes such as *cbp* to other microbes. The presence of toxin-antitoxin genes (*eatL-zetL*) may ensure stable maintenance of pLF3872 in *L. fermentum* 3872. To date, there are several plasmids identified in some *L. fermentum* strains (i.e. *L. fermentum* MTCC 8711 contains seven plasmids), none of which are similar to pLF3872. As previously mentioned, there are three closely related plasmids identified in other species of *Lactobacillus*, namely plca36 (*L. casei* Zhang), pWCFS103 (*L. plantarum* WCFS1) and plasmid 1(*L. paracasei subsp*). As mentioned in the results section plasmid pLF3872 consists of a *cbp* gene not found in any other strain of *L. fermentum* sequenced to date. An orthologue of the gene has been identified in *Lactobacillus plantarum* 91. The CBP protein expressed by *L. plantarum* 91 has shown antagonistic properties against

pathogen *E. coli* O157: H7 via competitive exclusion of adhesion ligands (Yadav *et al.*, 2013), which may be of interest for future *L. fermentum* 3872 experiments.

4.4 Competition and inhibition experiments

As genomic analysis identified a putative *cbp* gene, the CBP protein was expressed and analysed for competitive exclusion against enteric pathogen *C. jejuni*. Binding experiments showed that CBP may be utilised by *L. fermentum* 3872 for collagen I adhesion. The study also demonstrated that *C. jejuni* strains 11168H and 81–176 attach to collagen I. Adhesion can be important for bacterial colonisation, the CBP protein may play a role in *L. fermentum* 3872 growth and survival within a host, increasing the possibility of *L. fermentum* 3872 providing its probiotic effect through other mechanisms. Adhesion to collagen I may also increase *C. jejuni* virulence and potential for infection. Co-IP and mass spectrometry data suggested that *C. jejuni* strains 11168H and 81-176 may utilise flagella components (FlaA and FlaB) for binding collagen I. Studies have shown that *C. jejuni* deficient of flagella have a lower adhesion to intestinal epithelia (Haiko and Westerlund-Wikstrom, 2013; Newell, McBride and Dolby, 1985), but this is the first instance that the proteins have been directly affiliated with collagen adhesion.

In vitro experiments showed that the CBP protein prevents *C. jejuni* 11168H and 81–176 attachment to collagen I. Collagen I binding of *C. jejuni* strains 11168H and 81–176 was also inhibited in the presence of *L. fermentum* 3872. This competition may directly be against flagellar components (FlaA and FlaB) of *C. jejuni* 11168H and 81–176. The competition assay results demonstrate a way to use *L. fermentum* 3872 to reduce *C. jejuni* in the host via competitive exclusion. ELISA-based whole cell competition experiments also showed an increase in absorbance when mixing low amounts of *L. fermentum* 3872 to *C. jejuni*. The reasoning behind the increase in absorbance needs to be determined in future studies, a potential theory of the result may

be that either *L. fermentum* 3872 is causing auto-aggregation of *C. jejuni* strains 11168H and 81–176 or there is co-aggregation between the *C. jejuni* strains tested to *L. fermentum* 3872. Previous studies have shown co-aggregation between *Lactobacillus* and *C. jejuni* (Nishiyama *et al.*, 2014), as described in the introduction. The agar well diffusion assay showed that growth of *C. jejuni* strains 11168H and 81-176 was inhibited by the low pH of the *L. fermentum* 3872 cell-free supernatant. The low pH environment, in combination with co-aggregation, could work synergistically in combating *C. jejuni* infection as the distance between the two bacteria would reduce (Tareb *et al.*, 2013). Using numerous ways to combat pathogenic bacteria would also be advantageous in that the pathogen would have to evolve in many ways to survive.

4.5 Conclusion

With increasing interest in OMIC technologies, especially in the field of genomics, and an ever-growing abundance of data generated by the technology, emphasis needs to be on the quality of data generated, particularly for probiotic research, which is becoming heavily reliant on genomic data for characterising probiotics regarding safety, stability, and potential use (Papadimitriou *et al.*, 2015; Johnson and Klaenhammer, 2014; Walsh *et al.*, 2016). As previously mentioned, many genomic studies leave sequences as draft assemblies, making it difficult to build a complete profile of the organism in question (Land *et al.*, 2014). Having a complete genomic profile of a probiotic is important as certain probiotic traits can be overlooked due to their strain-specific nature (Islam, 2015). An example of the importance of a complete genome sequence was demonstrated with *L. fermentum* 3872, where previously, a partial *cbp* gene was identified when *L. fermentum* 3872 was in a draft state. The full copy of the *cbp* gene was only identified during the genome completion project, enabling the determination of essential CBP components, such as its leader peptide sequence. Identifying the leader peptide sequence may have been important for the expression of the recombinant CBP

protein for functional genomic analysis, as before its removal, the recombinant protein was degraded. The genome completion project also demonstrated that the *cbp* gene was present in a newly identified plasmid pLF3872, consisting of many conjugative genes. The conjugative genes may be problematic in that the plasmid may transfer genes to pathogenic bacteria. Nonetheless, pathogens such as *C. jejuni* tend to accept DNA only readily from their own species (Johnson, Shank and Johnson, 2017). Many survival and adhesion genes were identified, which may enable *L. fermentum* 3872 to survive the harsh GIT environment. Both the survival and adhesion genes helped build a profile of the probiotic in that it could potentially be administered orally and be useful in the GIT region. The presence of the adhesion genes may also be the reason for *L. fermentum* 3872 having strong adhesion to HeLa and Buccal cells compared to other *Lactobacillus* strains tested during experiments conducted for patent application (Abramov *et al*., 2014). To obtain a complete profile of the genetic makeup of *L. fermentum* 3872, the hypothetical proteins need to be functionally identified. As an antimicrobial tool, *L. fermentum* 3872 may be useful in combating enteric pathogens via competitive exclusion due to the multiple adhesion genes present. One such gene analysed during this study was a putative *cbp,* which encoded a CBP protein. The protein successfully inhibited *C. jejuni* strains 11168H and 81–176 attachment to collagen I. Furthermore, the combination of low pH and competitive exclusion may further reduce the survivability of *C. jejuni* in the presence of *L. fermentum* 3872, making it difficult for the pathogen to evolve and be resistant to the antagonistic effects of *L. fermentum* 3872. Nonetheless, as discussed in the future work section, *in vivo* experiments need to be conducted to determine whether *L. fermentum* 3872 is effective in combating *C. jejuni* in animal or human models. With *C. jejuni* being an important cause of gastroenteritis (Kaakoush *et al.*, 2015), and travellers' diarrhoea (Kaakoush *et al.*, 2015), *L. fermentum* 3872 may be used as an adjunct to rehydration therapy or as a prophylactic during

travels to high infection risk countries. If found to be effective in reducing the duration of gastroenteritis, the use of *L. fermentum* 3872, and perhaps a selection of probiotics with similar modes of actions, may reduce the incidence of *C. jejuni* affiliated IBD and IBS as the risk of developing the latter two conditions is dependent on the duration of the *C. jejuni* infection (Ericsson, Hatz and DuPont, 2008).

A summary of the major findings of the current study are listed below:

1. The genome sequence of *L. fermentum* 3872 was completed and a plasmid pLF3872 was discovered. Potential genes that may be of interest for future studies contributing to probiotic activity were highlighted.

2. The study demonstrated that read mapping, a commonly used genome sequence quality assessment tool could fail to detect sequence missassemblies.

3. The results showed that putative CBP isolated from *L. fermentum* 3872 and *C. jejuni* strains 11168H and 81-176, bind to collagen I. The *C. jejuni* strains utilise flagellar proteins FlaA and FlaB for collagen I adhesion.

4. The CBP protein inhibits *C. jejuni* 11168H and 81-176 binding to collagen I. *L. fermentum* 3872 also inhibits *C. jejuni* 11168H and 81-176 adhesion to collagen I.

5. *L. fermentum* 3872 is able to inhibit *C. jejuni* 11168H and 81-176 growth by lowering the pH environment.

4.6 Future experiments and limitations

While sequencing the genome of *L. fermentum* 3872, read mapping was unable to detect sequence misassembly. The Ion Torrent PGM used for this project only generates unpaired reads, and thus it would be interesting to determine whether paired end reads would produce similar results to unpaired reads. Paired end reads are similar to mate

pair reads in that two ends of a DNA fragment is sequenced but the region between the sequenced ends are smaller in size e.g. hundreds of bases rather than kilobases, and thus requires less laboratory work. Paired end reads have more software support that can be utilised for sequence gap closure, improvement of nucleotide and assembly errors. Tools such as iterative mapping and assembly for gap elimination (IMAGE) (Tsai, Otto and Berriman, 2010) can be utilised for automated gap closure, iterative correction of reference nucleotides (ICORN) (Otto *et al.*, 2010) can be used to correct small nucleotide errors, and recognition of errors in assemblies using paired end reads (REAPR) (Hunt *et al.*, 2013) can be used for assembly error correction. IMAGE, ICORN and REAPR tools do not require a reference genome, which can mitigate a carryover of potential assembly errors introduced by a reference genome sequence.

Future studies could also determine whether the CBP protein can inhibit adhesion of other pathogenic bacteria that are heavily reliant on collagen I attachment, such as *Staphylococcus aureus* (Ponnuraj *et al.*, 2003). Competitive exclusion studies can also be conducted against *E. coli* O157: H7 as previous *in vitro* experiments have shown a reduction in attachment to collagen I in the presence of the CBP protein expressed by *L. plantarum* 91 (Yadav *et al.*, 2013). *cbp* gene knockout experiments for *L. fermentum* 3872 can also be conducted to determine the extent to which the protein plays a role in *L. fermentum* 3872 adhesion to collagen I. Experiments with various other collagen types can be carried out to determine whether the CBP protein can promote attachment to collagens such as collagen IV, present in the basement membrane of the GIT (Verbeke *et al.*, 2001). The current study did not directly demonstrate *L. fermentum* 3872 attachment to collagen I due to non-availability of *L. fermentum* specific antibodies. The already expressed recombinant CBP protein can be used as an antigen for the development of polyclonal antibodies for confocal microscopy to visually

observe the CBP protein. The antibodies may also be used for ELISA experiments to confirm *L. fermentum* 3872 attachment to collagen I and perhaps other ECM proteins.

There were many other putative genes identified relating to *L. fermentum* 3872 adhesion to the GIT, such as those that promote binding to mucin and intestinal epithelia. It would be interesting to determine whether the genes play a role in the prevention of enteric pathogen attachment to GIT proteins. ELISA experiments between *L. fermentum* 3872 and *C. jejuni* indicated aggregation. Confocal microscopy could be used to observe if *C. jejuni* aggregation or co-aggregation is occurring in the presence of *L. fermentum* 3872. Moreover, previous studies that reported co-aggregation between *Lactobacillus* and *C. jejuni* found that the introduction of proteinase K resulted in no aggregation (Nishiyama *et al.*, 2014), suggesting that a protein was involved in the phenotypic effect. An ELISA could be repeated with the introduction of proteinase K to determine whether a protein is the causative agent of the aggregation effect. The putative bacteriocin of *L. fermentum* 3872 may also be expressed. A study has shown that Enterolysin A production in MRS was optimal at a constant pH of 6.5 but decreased as the pH lowered (Nilsen, Nes and Holo, 2003). Thus, *L. fermentum* 3872 could be grown in MRS broth with a constantly adjusted pH to isolate the bacteriocin, to determine whether the protein has an antagonistic effect on other bacteria.

The current study demonstrated that *L. fermentum* 3872 may utilise competitive exclusion and acidification of the environment to combat *C. jejuni* infection *in vitro*. The *in vitro* experiments need to be followed by *in vivo* experiments to determine whether *L. fermentum* 3872 can truly be used as a tool against *C. jejuni* infections in hosts. This is because certain *in vitro* studies have shown good antagonistic effects on *C. jejuni* but have failed to show similar effects *in vivo* (Johnson, Shank and Johnson, 2017), although many other studies have had positive correlations between *in vitro* and

in vivo models (Kaakoush *et al.*, 2015). As *L. fermentum* 3872 was isolated from a healthy human and utilised in its natural GIT environment, a randomised control trial could be undertaken. The trial may provide *L. fermentum* 3872 as an adjunct to rehydration therapy for individuals suffering from *C. jejuni* induced gastroenteritis. For the study, the time it takes for symptoms of gastroenteritis to be resolved could be measured for both case and control groups. This would provide an indication of whether the introduction of *L. fermentum* 3872 has a beneficial effect in the treatment of *C. jejuni* induced gastroenteritis. A follow-up study may also be conducted to determine future incidences of colitis and other post *C. jejuni* infection-related conditions. A cocktail of probiotics that are known to inhibit *C. jejuni* can also be used to determine whether there is a further reduction in the duration of *C. jejuni* infection symptoms.

Studies on the use of *L. fermentum* 3872 as a prophylactic in individuals travelling abroad could also be conducted to determine if the incidence of travellers' diarrhoea can be reduced. *In vivo* experiments on poultry could be conducted to determine whether *L. fermentum* 3872 can reduce *C. jejuni* in chickens, much like that carried out by Nishiyama *et al*. (2014). This would help identify whether *L. fermentum* 3872 can be used against a major vector for human *C. jejuni* infection (Kaakoush *et al.*, 2015). There may be a benefit in using probiotics isolated from humans in poultry as the risk of the bacterium being pathogenic to a human would be unlikely. The *L. fermentum* 3872 growth condition could also be varied for the agar well diffusion assay to determine whether there are any proteins induced that have an antagonistic effect on *C. jejuni* during varying *L. fermentum* 3872 growth conditions.

Publications

During the course of this study the publications listed below were reported.

Lehri, B., Seddon, A. M. and Karlyshev, A. V. (2015) '*Lactobacillus fermentum* 3872 genome sequencing reveals plasmid and chromosomal genes potentially involved in a probiotic activity', *FEMS Microbiology Letters,* 362 (11), doi: 10.1093/femsle/fnv068.

Lehri, B., Seddon, A. M. and Karlyshev, A. V. (2017) 'Potential probiotic-associated traits revealed from completed high quality genome sequence of *Lactobacillus fermentum* 3872', *Standards in Genomic Sciences,* 12 (1), pp. 19.

Lehri, B., Seddon, A. M. and Karlyshev, A. V. (2017a) 'The hidden perils of read mapping as a quality assessment tool in genome sequencing', *Scientific Reports,* 7, doi: 10.1038/srep43149.

Lehri, B., Seddon, A. M. and Karlyshev, A. V. (2017b) '*Lactobacillus fermentum* 3872 as a potential tool for combatting *Campylobacter jejuni* infections', *Virulence,* pp. 00-00, doi: 10.1080/21505594.2017.1362533.

Lehri, B., Kukreja, K., Vieira, A., Zaremba, M., Bonney, K., & Karlyshev, A. V. (2015). Specific genetic features of *campylobacter jejuni* strain G1 revealed by genome sequencing. *FEMS Microbiology Letters*, *362*(4), 1-3.

References

Abramov VM, Khlebnikov VS, Pchelintsev SJ, Kosarev IV, Karlyshev AV, Vasilenko RN, Melnikov VG. (2014) 'Strain *Lactobacillus fermentum* having broad spectrum of antagonistic activity and probiotic *lactobacterium consortium* for manufacturing bacterial preparations RU 2528862 C1', Russia. Patent. Application Number: 2013118084/10, Application Date: 19.04.2013, Publication Number: 0002528862.

Acheson, D. and Allos, B. M. (2001) '*Campylobacter jejuni* Infections: Update on Emerging Issues and Trends', *Clinical Infectious Diseases*, 32 (8), pp. 1201-1206.

Ajene, A. N., Walker, C. L. F. and Black, R. E. (2013) 'Enteric Pathogens and Reactive Arthritis: A Systematic Review of *Campylobacter*, *Salmonella* and *Shigella*-associated Reactive Arthritis', *Journal of Health, Population, and Nutrition*, 31 (3), pp. 299-307.

Allen, S. J., Martinez, E. G., Gregorio, G. V. and Dans, L. F. (2010) 'Probiotics for treating acute infectious diarrhoea', *The Cochrane Library*, CD003048 (11), doi: 10.1002/14651858.CD003048.pub3.

Alok, A., Singh, I. D., Singh, S., Kishore, M., Jha, P. C. and Iqubal, M. A. (2015) 'Probiotics: A New Era of Biotherapy', *Advanced Biomedical Research*, 6 pp. 31.

Amour, C., Gratz, J., Mduma, E., Svensen, E., Rogawski, E. T., McGrath, M., Seidman, J. C., McCormick, B. J. J., Shrestha, S., Samie, A., Mahfuz, M., Qureshi, S., Hotwani, A., Babji, S., Trigoso, D. R., Lima, A. A. M., Bodhidatta, L., Bessong, P., Ahmed, T., Shakoor, S., Kang, G., Kosek, M., Guerrant, R. L., Lang, D., Gottlieb, M., Houpt, E. R., Platts-Mills, J. (2016) 'Epidemiology and Impact of *Campylobacter* Infection in Children in 8 Low-Resource Settings: Results From the MAL-ED Study', *Clinical Infectious Diseases*, 63 (9), pp. 1171-1179.

Anderson, R. C., Young, W., Clerens, S., Cookson, A. L., McCann, M. J., Armstrong, K. M. and Roy, N. C. (2013) 'Human Oral Isolate *Lactobacillus fermentum* AGR1487 Reduces Intestinal Barrier Integrity by Increasing the Turnover of Microtubules in Caco-2 Cells', *Plos One,* 8 (11), pp. e78774.

Angiuoli, S. V., Gussman, A., Klimke, W., Cochrane, G., Field, D., Garrity, G. M., Kodira, C. D., Kyrpides, N., Madupu, R., Markowitz, V., Tatusova, T., Thomson, N. and White, O. (2008) 'Toward an Online Repository of Standard Operating Procedures (SOPs) for (Meta) genomic Annotation', *OMICS : A Journal of Integrative Biology,* 12 (2), pp. 137-141.

Antipov, D., Korobeynikov, A., McLean, J. S. and Pevzner, P. A. (2016) 'hybridSPAdes: an algorithm for hybrid assembly of short and long reads', *Bioinformatics,* 32 (7), pp. 1009-1015.

Auzat, I., Dröge, A., Weise, F., Lurz, R. and Tavares, P. (2008) 'Origin and function of the two major tail proteins of bacteriophage SPP1', *Molecular Microbiology,* 70 (3), pp. 557-569.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. and Pevzner, P. A. (2012) 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing', *Journal of Computational Biology,* 19 (5), pp. 455-477.

Ben Lagha, A., Haas, B., Gottschalk, M. and Grenier, D. (2017) 'Antimicrobial potential of bacteriocins in poultry and swine production', *Veterinary Research,* 48 (1), pp. 22.

Bergonzelli, G. E., Granato, D., Pridmore, R. D., Marvin-Guy, L. F., Donnicola, D. and Corthésy-Theulaz, I. E. (2006) 'GroEL of *Lactobacillus johnsonii* La1 (NCC 533) Is

Cell Surface Associated: Potential Role in Interactions with the Host and the Gastric Pathogen *Helicobacter pylori*', *Infection and Immunity,* 74 (1), pp. 425-434.

Bermudez-Brito, M., Plaza-Diaz, J., Munoz-Quezada, S., Gomez-Llorente, C. and Gil, A. (2012) 'Probiotic Mechanisms of Action', *Annals of Nutrition and Metabolism,* 61 (2), pp. 160-174.

Berstad, A., Raa, J., Midtvedt, T. and Valeur, J. (2016) 'Probiotic lactic acid bacteria - the fledgling cuckoos of the gut?', *Microbial Ecology in Health and Disease,* 27, doi: 10.3402/mehd.v27.31557.

Bull, M. J., Jolley, K. A., Bray, J. E., Aerts, M., Vandamme, P., Maiden, M. C. J., Marchesi, J. R. and Mahenthiralingam, E. (2014) 'The domestication of the probiotic bacterium *Lactobacillus acidophilus*', 4 (1).

Butel, M.J. (2014) 'Probiotics, gut microbiota and health', *Medicine Et Maladies Infectieuses,* 44 (1), pp. 1-8.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M., Barrell, B. G. and Parkhill, J. (2005) 'ACT: the Artemis comparison tool', *Bioinformatics,* 21 (16), pp. 3422-3423.

Castillo, N. A., Perdigon, G. and de Moreno, d. L. (2011) 'Oral administration of a probiotic *Lactobacillus* modulates cytokine production and TLR expression improving the immune response against *Salmonella enterica serovar Typhimurium* infection in mice', *BMC Microbiology,* 11 (1), pp. 177.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Muller, W. E. G., Wetter, T. and Suhai, S. (2004) 'Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs', *Genome Research,* 14 (6), pp. 1147-1159.

Claesson, M. J., Sinderen, D. V. and O'Toole, P. W. (2007) ‘The genus *Lactobacillus*--a genomic basis for understanding its diversity’, *FEMS Microbiology Letters*, 269(1), pp. 8-22.

Corr, S. C., Li, Y., Riedel, C. U., O'Toole, P. W., Hill, C. and Gahan, C. G. M. (2007) 'Bacteriocin production as a mechanism for the anti-infective activity of *Lactobacillus salivarius* UCC118', *Proceedings of the National Academy of Sciences*, 104 (18), pp. 7617-7621.

Cotter, P. D., Hill, C. and Ross, R. P. (2005) 'Bacteriocins: developing innate immunity for food', *Nature Reviews Microbiology*, 3(10), pp. 777-788.

Cozma-Petrut, A., Loghin, F., Miere, D. and Dumitrascu, D. L. (2017) 'Diet in irritable bowel syndrome: What to recommend, not what to forbid to patients!', *World Journal of Gastroenterology*, 23 (21), pp. 3771-3783.

Davies, J. R., Svensäter, G. and Herzberg, M. C. (2009) 'Identification of novel LPXTG-linked surface proteins from *Streptococcus gordonii*', *Microbiology*, 155 (6), pp. 1977-1988.

De Mattos, B.,Rafael Ramos, Garcia, M. P. G., Nogueira, J. B., Paiatto, L. N., Albuquerque, C. G., Souza, C. L., Fernandes, L. G. R., Tamashiro, W.M., and Simioni, P. U. (2015) 'Inflammatory Bowel Disease: An Overview of Immune Mechanisms and Biological Treatments', *Mediators of Inflammation*, 2015, pp. 1-11.

Deivanayagam, C. C., Rich, R. L., Carson, M., Owens, R. T., Danthuluri, S., Bice, T., Höök, M. and Narayana, S. V. (2000) 'Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein', *Structure*, 8 (1), pp. 67-78.

Dobson, A., Cotter, P. D., Ross, R. P. and Hill, C. (2012) 'Bacteriocin Production: a Probiotic Trait?', *Applied and Environmental Microbiology,* 78 (1), pp. 1-6.

Doron, S. and Snydman, D. R. (2015) 'Risk and Safety of Probiotics', *Clinical Infectious Diseases,* 60 (2) pp. S129-S134.

Drissi, F., Merhej, V., Angelakis, E., El Kaoutari, A., Carriere, F., Henrissat, B. and Raoult, D. (2014) 'Comparative genomics analysis of *Lactobacillus* species associated with weight gain or weight protection', 4 (2), pp. e109.

EFSA Panel on Biological Hazards. (2011) 'Scientific Opinion on *Campylobacter* in broiler meat production: control options and performance objectives and/or targets at different stages of the food chain', *EFSA Journal,* 9 (4), pp. 2105.

Ekblom, R. and Wolf, J. B. W. (2014) 'A field guide to whole-genome sequencing, assembly and annotation', *Evolutionary Applications,* 7 (9), pp. 1026-1042.

Ericsson, C. D., Hatz, C. and DuPont, A. W. (2008) 'Postinfectious Irritable Bowel Syndrome', *Clinical Infectious Diseases,* 46 (4), pp. 594-599.

Esmay, P. A., Billington, S. J., Link, M. A., Songer, J. G. and Jost, B. H. (2003) 'The Arcanobacterium pyogenes Collagen-Binding Protein, CbpA, Promotes Adhesion to Host Cells', *Infection and Immunity,* 71 (8), pp. 4368-4374.

European Food Safety Authority and European Centre for Disease Prevention and Control. (2016) 'The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2014', *EFSA Journal,* 14 (2), pp. 4380.

Evivie, S., Huo, G., Igene, J. and Bian, X. (2017) ‘Some current applications, limitations and future perspectives of lactic acid bacteria as probiotics’, *Food & Nutrition Research,* 61 (1), pp.1318034.

Fiocchi, A., Pawankar, R., Cuello-Garcia, C., Ahn, K., Al-Hammadi, S., Agarwal, A., Beyer, K., Burks, W., Canonica, G. W., Ebisawa, M., Gandhi, S., Kamenwa, R., Lee, B. W., Li, H., Prescott, S., Riva, J. J., Rosenwasser, L., Sampson, H., Spigler, M., Terracciano, L., Vereda-Ortiz, A., Waserman, S., Yepes-Nunez, J. J., Brozek, J. L. and Schunemann, H. J. (2015) 'World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Probiotics', *World Allergy Organization Journal,* 8 (1), pp. 1-13.

Flanagan, R. C., Neal-McKinney, J., Dhillon, A. S., Miller, W. G. and Konkel, M. E. (2009) 'Examination of *Campylobacter jejuni* Putative Adhesins Leads to the Identification of a New Protein, Designated FlpA, Required for Chicken Colonization', *Infection and Immunity,* 77 (6), pp. 2399-2407.

Ford, A. C., Quigley, E. M., Lacy, B. E., Lembo, A. J., Saito, Y. A., Schiller, L. R., Soffer, E. E., Spiegel, B. M. and Moayyedi, P. (2014) 'Efficacy of prebiotics, probiotics, and synbiotics in irritable bowel syndrome and chronic idiopathic constipation: systematic review and meta-analysis', *The American Journal of Gastroenterology,* 109 (10), pp. 1547-1561.

Galardini, M., Biondi, E. G., Bazzicalupo, M. and Mengoni, A. (2011) 'CONTIGuator: a bacterial genome finishing tool for structural insights on draft genomes', *Source Code for Biology and Medicine,* 6 (1) pp. 11.

Gautret, P., Cramer, J. P., Field, V., Caumes, E., Jensenius, M., Gkrania-Klotsas, E., de Vries, P. J., Grobusch, M. P., Lopez-Velez, R., Castelli, F., Schlagenhauf, P., Hervius Askling, H., von Sonnenburg, F., Lalloo, D. G., Loutan, L., Rapp, C., Basto, F., Santos O'Connor, F., Weld, L., Parola, P. and EuroTravNet Network. (2012) 'Infectious diseases among travellers and migrants in Europe, EuroTravNet 2010', *Euro*

Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles, European Communicable Disease Bulletin, 17 (26), pp. 20205.

Ghachi, M. E., Derbise, A., Bouhss, A. and Mengin-Lecreulx, D. (2005) 'Identification of Multiple Genes Encoding Membrane Proteins with Undecaprenyl Pyrophosphate Phosphatase (UppP) Activity in *Escherichia coli*', *Journal of Biological Chemistry,* 280 (19), pp. 18689-18695.

Granato, D., Bergonzelli, G. E., Pridmore, R. D., Marvin, L., Rouvet, M. and Corthésy-Theulaz, I. E. (2004) 'Cell Surface-Associated Elongation Factor Tu Mediates the Attachment of *Lactobacillus johnsonii* NCC533 (La1) to Human Intestinal Cells and Mucins', *Infection and Immunity,* 72 (4), pp. 2160-2169.

Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.', *Journal of Bacteriology,* 177 (14), pp. 4121-4130.

Haiko, J. and Westerlund-Wikstrom, B. (2013) 'The Role of the Bacterial Flagellum in Adhesion and Virulence', *Biology,* 2 (4), pp. 1242-1267.

Hamon, E., Horvatovich, P., Marchioni, E., Aoudé-Werner, D. and Ennahar, S. (2014) 'Investigation of potential markers of acid resistance in *Lactobacillus plantarum* by comparative proteomics', *Journal of Applied Microbiology,* 116 (1), pp. 134-144.

Hampton, T. (2013) 'Report reveals scope of US antibiotic resistance threat', *Jama,* 310 (16), pp. 1661-1663.

Han, K., Jang, S. S., Choo, E., Heu, S. and Ryu, S. (2007) ‘Prevalence, genetic diversity, and antibiotic resistance patterns of *Campylobacter jejuni* from retail raw chickens in Korea. *International Journal of Food Microbiology,* 114 (1), pp. 50-59.

Harris RS. (2007) ‘Improved pairwise alignment of genomic DNA’. *PhD Thesis*. The Pennsylvania State University.

Heel, A. J., Jong, A., Montalban-Lopez, M., Kok, J. and Kuipers, O. P. (2013) 'BAGEL3: automated identification of genes encoding bacteriocins and non-bactericidal post translationally modified peptides', *Nucleic Acids Res,* 41(W1), pp. W448-W453.

Hevia, A., Delgado, S., Margolles, A. and Sánchez, B. (2015) 'Application of density gradient for the isolation of the fecal microbial stool component and the potential use thereof', *Scientific reports,* 5 (1), doi:10.1038/srep16807.

Hitchen, P., Brzostek, J., Panico, M., Butler, J. A., Morris, H. R., Dell, A. and Linton, D. (2010) 'Modification of the *Campylobacter jejuni* flagellin glycan by the product of the Cj1295 homopolymeric-tract-containing gene', *Microbiology,* 156 (7) pp. 1953-1962.

Ho, N. and Prasad, V. (2013) 'Probiotics, prebiotics, synbiotics and naturally fermented foods: why more may be more', *Annals of Gastroenterology : Quarterly Publication of the Hellenic Society of Gastroenterology,* 26 (3), pp. 277-278.

Homan, M. and Orel, R. (2015) 'Are probiotics useful in *Helicobacter pylori* eradication?', *World Journal of Gastroenterology: WJG,* 21 (37), pp. 10644-10653.

Horgan, R. P. and Kenny, L. C. (2011) '‘Omic’ technologies: genomics, transcriptomics, proteomics and metabolomics', *The Obstetrician & Gynaecologist,* 13 (3), pp. 189-195.

Hoveyda, N., Heneghan, C., Mahtani, K. R., Perera, R., Roberts, N. and Glasziou, P. (2009) 'A systematic review and meta-analysis: probiotics in the treatment of irritable bowel syndrome', *BMC Gastroenterology,* 9 (1), pp. 15.

Hsiao, W. W. and Fraser-Liggett, C. M. (2009) 'Human Microbiome Project--paving the way to a better understanding of ourselves and our microbes', *Drug Discovery Today,* 14 (7-8), pp. 331-333.

Hu, L. and Kopecko, D. J. (1999) *'Campylobacter jejuni* 81-176 Associates with Microtubules and Dynein during Invasion of Human Intestinal Cells', *Infection and Immunity,* 67 (8), pp. 4171-4182.

Hu, Y., Benedik, M. J. and Wood, T. K. (2012) 'Antitoxin DinJ influences the general stress response through transcript stabilizer CspE', *Environmental Microbiology,* 14 (3), pp. 669-679.

Huang, X. and Madan, A. (1999) 'CAP3: A DNA Sequence Assembly Program', *Genome Research,* 9 (9), pp. 868-877.

Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M. and Otto, T. D. (2013) 'REAPR: a universal tool for genome assembly evaluation', *Genome Biology,* 14 (5), pp. R47.

Islam, S. U. (2015) 'Clinical Uses of Probiotics', *Medicine,* 95 (5), pp. e2658.

Izquierdo, E., Horvatovich, P., Marchioni, E., Aoude-Werner, D., Sanz, Y. and Ennahar, S. (2009) '2-DE and MS analysis of key proteins in the adhesion of *Lactobacillus plantarum,* a first step toward early selection of probiotics based on bacterial biomarkers', *Electrophoresis,* 30 (6), pp. 949-956.

Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M. and Reddy, D. N. (2015) 'Role of the normal gut microbiota', *World Journal of Gastroenterology : WJG,* 21 (29), pp. 8787-8803.

Johansson, M. A., Sjogren, Y. M., Persson, J., Nilsson, C. and Sverremark-Ekstrom, E. (2011) 'Early Colonization with a Group of *Lactobacilli* Decreases the Risk for Allergy at Five Years of Age Despite Allergic Heredity', *Plos One*, 6 (8), pp. e23031.

Janssen, R., Krogfelt, K. A., Cawthraw, S. A., van Pelt, W., Wagenaar, J. A. and Owen, R. J. (2008) 'Host-Pathogen Interactions in Campylobacter Infections: the Host Perspective', *Clinical Microbiology Reviews*, 21 (3), pp. 505-518.

Johnson, B. R. and Klaenhammer, T. R. (2014) 'Impact of genomics on the field of probiotic research: historical perspectives to modern paradigms', *Antonie Van Leeuwenhoek*, 106 (1), pp. 141-156.

Johnson, T. J., Shank, J. M. and Johnson, J. G. (2017) 'Current and Potential Treatments for Reducing Campylobacter Colonization in Animal Hosts and Disease in Humans*' Frontiers in Microbiology,* 8, doi: org/10.3389/fmicb.2017.00487.

Jones, S. E., Paynich, M. L., Kearns, D. B. and Knight, K. L. (2014) 'Protection from intestinal inflammation by bacterial exopolysaccharides', *Journal of Immunology (Baltimore, Md.: 1950)*, 192 (10), pp. 4813-4820.

Jorgensen, F., Charlett, A., Arnold, E., Swift, C., Madden, B. and C Elviss, N. (2017) 'A microbiological survey of *Campylobacter* contamination in fresh whole UK - produced chilled chickens at retail sale', *Public Health England and UK Food Standards Agency*. Available at: https://www.food.gov.uk/sites/default/files/fsa-project-fs102121-year-2-report.pdf (Accessed: 20 June 2017).

Juge, N. (2012) 'Microbial adhesins to gastrointestinal mucus', *Trends in Microbiology*, 20 (1), pp. 30-39.

Kaakoush, N. O., Castano-Rodriguez, N., Mitchell, H. M. and Man, S. M. (2015) 'Global Epidemiology of *Campylobacter* Infection', *Clinical Microbiology Reviews*, 28 (3), pp. 687-720.

Karlyshev, A. V., Linton, D., Gregson, N. A. and Wren, B. W. (2002) 'A novel paralogous gene family involved in phase-variable flagella-mediated motility in *Campylobacter jejuni*', *Microbiology,* 148 (2), pp. 473-480.

Kawai, F., Paek, S., Choi, K., Prouty, M., Kanipes, M. I., Guerry, P. and Yeo, H. (2012) 'Crystal structure of JlpA, a surface-exposed lipoprotein adhesin of *Campylobacter jejuni*', *Journal of Structural Biology,* 177 (2), pp. 583-588.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P and Drummond, A. (2012) 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28 (12), pp. 1647-1649.

Kline, K. A., Fälker, S., Dahlberg, S., Normark, S. and Henriques-Normark, B. (2009) 'Bacterial Adhesins in Host-Microbe Interactions', *Cell Host & Microbe*, 5 (6), pp. 580-592.

Konkel, M. E., Larson, C. L. and Flanagan, R. C. (2010) '*Campylobacter jejuni* FlpA Binds Fibronectin and Is Required for Maximal Host Cell Adherence', *Journal of Bacteriology,* 192 (1), pp. 68-76.

Kuusela, P., Moran, A. P., Vartio, T. and Kosunen, T. U. (1989) 'Interaction of *Campylobacter jejuni* with extracellular matrix components', *Biochimica Et Biophysica Acta (BBA) - General Subjects*, 993 (2–3), pp. 297-300.

Lahtinen, S. J. (2012) 'Probiotic viability - does it matter?', *Microbial Ecology in Health and Disease*, 23, doi: 10.3402/mehd.v23i0.18567.

Land, M., Hyatt, D., Jun, S., Kora, G., Hauser, L., Lukjancenko, O and Ussery, D. (2014) 'Quality scores for 32,000 genomes', *Standards in Genomic Sciences*, 9 (1), pp. 20.

Lee, K., Pi, K., Kim, E. B., Rho, B., Kang, S. and Lee, H. G. (2010) 'Glutathione-mediated response to acid stress in the probiotic bacterium, *Lactobacillus salivarius*', *Biotechnology Letters*, 32 (7), pp. 969-972.

Lehri, B., Seddon, A. M. and Karlyshev, A. V. (2015) '*Lactobacillus fermentum* 3872 genome sequencing reveals plasmid and chromosomal genes potentially involved in a probiotic activity', *FEMS Microbiology Letters*, 362 (11), doi: 10.1093/femsle/fnv068.

Lehri, B., Seddon, A. M. and Karlyshev, A. V. (2017a) 'Potential probiotic-associated traits revealed from completed high quality genome sequence of *Lactobacillus fermentum* 3872', *Standards in Genomic Sciences*, 12 (1), pp. 19.

Lehri, B., Seddon, A. M. and Karlyshev, A. V. (2017b) 'The hidden perils of read mapping as a quality assessment tool in genome sequencing', *Scientific Reports*, 7, doi: 10.1038/srep43149.

Lehri, B., Seddon, A. M. and Karlyshev, A. V. (2017c) '*Lactobacillus fermentum* 3872 as a potential tool for combatting *Campylobacter jejuni* infections', *Virulence*, pp. 00-00, doi: 10.1080/21505594.2017.1362533.

Lesbros-Pantoflickova, D., Corthesy-Theulaz, I. and Blum, A. L. (2007) '*Helicobacter pylori* and probiotics', *The Journal of Nutrition*, 137 (3), pp. 812S-8S.

Lichtenstein, L., Avni-Biron, I. and Ben-Bassat, O. (2016) ‘Probiotics and prebiotics in Crohn's disease therapies’, *Best Practice & Research Clinical Gastroenterology*, 30 (1), pp. 81-88.

Lilly, D. M. and Stillwell, R. H. (1965) 'Probiotics: Growth-Promoting Factors Produced by Microorganisms', *Science,* 147 (3659), pp. 747-748.

Lin, H. C., Goldstein, S., Mendelowitz, L., Zhou, S., Wetzel, J., Schwartz, D. C. and Pop, M. (2012) ‘AGORA: Assembly Guided by Optical Restriction Alignment’, *BMC Bioinformatics*, 13 (1), pp. 189.

Lin, S. and Liao, Y. (2013) 'CISA: Contig Integrator for Sequence Assembly of Bacterial Genomes', *Plos One,* 8 (3), pp. e60843.

Mack, D. R., Ahrne, S., Hyde, L., Wei, S. and Hollingsworth, M. A. (2003) 'Extracellular MUC3 mucin secretion follows adherence of *Lactobacillus* strains to intestinal epithelial cells in vitro', *Gut,* 52 (6), pp. 827-33.

Mackowiak, P. A. (2013) 'Recycling Metchnikoff: Probiotics, the Intestinal Microbiome and the Quest for Long Life', *Frontiers in Public Health,* 1 (52), doi: 10.3389/fpubh.2013.00052.

Mahdavi, J., Pirinccioglu, N., Oldfield, N. J., Carlsohn, E., Stoof, J., Aslam, A., Self, T., Cawthraw, S. A., Petrovska, L., Colborne, N., Sihlbom, C., Boren, T., Wooldridge, K. G. and Ala'Aldeen, D. A. A. (2014) 'A novel *O*-linked glycan modulates *Campylobacter jejuni* major outer membrane protein-mediated adhesion to human histo-blood group antigens and chicken colonization', *Open Biology,* 4 (1), pp. 130202.

Mazmanian, S. K., Liu, G., Ton-That, H. and Schneewind, O. (1999) '*Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall', 285 (5428), pp. 760-763.

Mendelowitz, L., Pop, M. (2014) ‘Computational methods for optical mapping’, *GigaScience*, 3 (1), doi: https://doi.org/10.1186/2047-217X-3-33.

MetaHIT (2017). Available at: http://www.metahit.eu/index.php?id=410 (Accessed: 21 June 2017).

Moal, V. L. and Servin, A. L. (2014) 'Anti-Infective Activities of *Lactobacillus* Strains in the Human Intestinal Microbiota: from Probiotics to Gastrointestinal Anti-Infectious Biotherapeutic Agents', *Clinical Microbiology Reviews*, 27 (2), pp. 167-199.

Moore, J. E., Wilson, T. S., Wareing, D. R., Humphrey, T. J. and Murphy, P. G. (2002) 'Prevalence of thermophilic *Campylobacter* spp. in ready-to-eat foods and raw poultry in Northern Ireland', *Journal of Food Protection*, 65 (8), pp. 1326-1328.

Mozzi, F., Gerbino, E., Font de Valdez, G. and Torino, M. I. (2009) 'Functionality of exopolysaccharides produced by lactic acid bacteria in an in vitro gastric system', *Journal of Applied Microbiology*, 107 (1), pp. 56-64.

Muñoz-Provencio, D., Pérez-Martínez, G. and Monedero, V. (2010) 'Characterization of a fibronectin-binding protein from *Lactobacillus casei* BL23', *Journal of Applied Microbiology*, 108 (3), pp. 1050-1059.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H., Remington, K. A., Anson, E. L., Bolanos, R. A., Chou, H. H., Jordan, C. M., Halpern, A. L., Lonardi, S., Beasley, E. M., Brandon, R. C., Chen, L., Dunn, P. J., Lai, Z., Liang, Y., Nusskern, D. R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G. M., Adams, M. D. and Venter, J. C. (2000) 'A whole-genome assembly of Drosophila', *Science (New York, N.Y.)*, 287 (5461), pp. 2196-2204.

Mynewgut (2017). Available at: http://www.mynewgut.eu/ (Accessed: 21 June 2017).

Neal-McKinney, J., Lu, X., Duong, T., Larson, C. L., Call, D. R., Shah, D. H. and Konkel, M. E. (2012) 'Production of Organic Acids by Probiotic *Lactobacilli* Can Be Used to Reduce Pathogen Load in Poultry', *Plos One,* 7 (9), pp. e43928.

Newell, D. G., McBride, H. and Dolby, J. M. (1985) 'Investigations on the role of flagella in the colonization of infant mice with *Campylobacter jejuni* and attachment of *Campylobacter jejuni* to human epithelial cell lines', *The Journal of Hygiene,* 95 (2), pp. 217-227.

Nichols, G. L., Richardson, J. F., Sheppard, S. K., Lane, C. and Sarran, C. (2012) '*Campylobacter* epidemiology: a descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011', *BMJ Open,* 2 (4), pp. e001179.

Nielsen, H., Hansen, K. K., Gradel, K. O., Kristensen, B., Ejlertsen, T., Ostergaard, C. and Schonheyder, H. C. (2010) 'Bacteraemia as a result of *Campylobacter* species: a population-based study of epidemiology and clinical risk factors', *Clinical Microbiology and Infection,* 16 (1), pp. 57-61.

Nilsen, T., Nes, I. F. and Holo, H. (2003) 'Enterolysin A, a Cell Wall-Degrading Bacteriocin from *Enterococcus faecalis* LMG 2333', *Applied and Environmental Microbiology,* 69 (5), pp. 2975-2984.

Nishiyama, K., Seto, Y., Yoshioka, K., Kakuda, T., Takai, S., Yamamoto, Y. and Mukai, T. (2014) '*Lactobacillus gasseri* SBT2055 Reduces Infection by and Colonization of *Campylobacter jejuni*', *Plos One,* 9 (9), pp. e108827, doi: 10.1371/journal.pone.0108827.

Ó Cróinín, T. and Backert, S. (2012) 'Host Epithelial Cell Invasion by *Campylobacter jejuni*: Trigger or Zipper Mechanism?', *Frontiers in Cellular and Infection Microbiology,* 2, doi: 10.3389/fcimb.2012.00025.

Oakland, M., Jeon, B., Sahin, O., Shen, Z. and Zhang, Q. (2011) 'Functional Characterization of a Lipoprotein-Encoding Operon in Campylobacter jejuni', *Plos One*, 6 (5), pp. e20084. doi: 10.1371/journal.pone.0020084.

Oelschlaeger, T. A. (2010) 'Mechanisms of probiotic actions – A review', *International Journal of Medical Microbiology*, 300 (1), pp. 57-62.

OpGen (2017). Available at: http://www.opgen.com/sequencing/ (Accessed: 19 Oct. 2017).

Otto, T. D., Sanders, M., Berriman, M. and Newbold, C. (2010) 'Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology', *Bioinformatics*, 26 (14), pp. 1704-1707.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. and Stevens, R. (2014) 'The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)', *Nucleic Acids Research*, 42 (1), pp. D206-D214.

Ozer, E., Allen, J. and Hauser, A. (2014) 'Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt', *BMC Genomics*, 15 (1), pp. 737.

Papadimitriou, K., Zoumpopoulou, G., Foligne, B., Alexandraki, V., Kazou, M., Pot, B. and Tsakalidou, E. (2015) 'Discovering probiotic microorganisms: in vitro, in vivo, genetic and omics approaches', *Frontiers in Microbiology*, 6 pp. 58.

Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A. V., Moule, S., Pallen, M. J., Penn, C. W., Quail, M. A., Rajandream, M. A., Rutherford, K. M., van Vliet, A. H., Whitehead, S. and Barrell, B. G. (2000) 'The genome sequence of

the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences', *Nature*, 403 (6770), pp. 665-668.

Patel, S., Majumder, A. and Goyal, A. (2009) 'Potentials of Exopolysaccharides from Lactic Acid Bacteria', *Indian Journal of Microbiology*, 52 (1), pp. 3-12.

Penner, R. M. and Fedorak, R. N. (2005) 'Probiotics in the Management of Inflammatory Bowel Disease', *Medscape General Medicine*, 7 (3), pp. 19-19.

Pessione, E. (2012) 'Lactic acid bacteria contribution to gut microbiota complexity: lights and shadows', *Frontiers in Cellular and Infection Microbiology*, 2 pp. 86.

Pietrocola, G., Valtulina, V., Rindi, S., Jost, B. H. and Speziale, P. (2007) 'Functional and structural properties of CbpA, a collagen-binding protein from *Arcanobacterium pyogenes*', *Microbiology*, 153 (10), pp. 3380-3389.

Platts-Mills, J. and Kosek, M. (2014) 'Update on the burden of *Campylobacter* in developing countries', *Current Opinion in Infectious Diseases*, 27 (5), pp. 444-450.

Ponnuraj, K., Bowden, M. G., Davis, S., Gurusiddappa, S., Moore, D., Choe, D., Xu, Y., Hook, M. and Narayana, S. V. L. (2003) 'A “dock, lock, and latch” Structural Model for a *Staphylococcal* Adhesin Binding to Fibrinogen', *Cell*, 115 (2), pp. 217-228.

Prasad, J., McJarrow, P. and Gopal, P. (2003) 'Heat and Osmotic Stress Responses of Probiotic *Lactobacillus rhamnosus* HN001 (DR20) in Relation to Viability after Drying', *Applied and Environmental Microbiology*, 69 (2), pp. 917-925.

Public Health England (2017) *Campylobacter data 2006 to 2015*. Public Health England (Campylobacter: guidance, data and analysis), pp. 1-9. Available at: https://www.gov.uk/government/publications/campylobacter-cases-2000-to-2012 (Accessed: 8 August 2017).

Rautava, S., Kalliomaki, M. and Isolauri, E. (2002) 'Probiotics during pregnancy and breast-feeding might confer immunomodulatory protection against atopic disease in the infant', *Journal of Allergy and Clinical Immunology,* 109 (1), pp. 119-121.

Renz-Polster, H., David, M. R., Buist, A. S., Vollmer, W. M., O'Connor, E. A., Frazier, E. A. and Wall, M. A. (2005) 'Caesarean section delivery and the risk of allergic disorders in childhood', *Clinical & Experimental Allergy,* 35 (11), pp. 1466-1472.

Reuter, J. A., Spacek, D. and Snyder, M. P. (2015) 'High-Throughput Sequencing Technologies', *Molecular Cell,* 58 (4), pp. 586-597.

Riddle, M. S., Murray, J. A., Cash, B. D., Pimentel, M. and Porter, C. K. (2013) 'Pathogen-Specific Risk of Celiac Disease Following Bacterial Causes of Foodborne Illness: A Retrospective Cohort Study', *Digestive Diseases and Sciences,* 58 (11), pp. 3242-3245.

Rubinchik, S., Seddon, A. and Karlyshev, A. V. (2012) 'Molecular mechanisms and biological role of *Campylobacter jejuni* attachment to host cells', *European Journal of Microbiology & Immunology,* 2 (1), pp. 32-40.

Ruggiero, P. (2014) 'Use of probiotics in the fight against *Helicobacter pylori*', *World Journal of Gastrointestinal Pathophysiology,* 5 (4), pp. 384-391.

Sánchez , B., Delgado, S., Blanco-Miguez, A., Lourenco, A., Gueimonde, M. and Margolles, A. (2017) 'Probiotics, gut microbiota, and their influence on host health and disease', *Molecular Nutrition & Food Research,* 61 (1), pp. 1613-4133.

Sabet, C., Lecuit, M., Cabanes, D., Cossart, P. and Bierne, H. (2005) 'LPXTG Protein InlJ, a Newly Identified Internalin Involved in *Listeria monocytogenes* Virulence', *Infection and Immunity,* 73 (10), pp. 6912-22.

Salari, P., Nikfar, S. and Abdollahi, M. (2012) 'A meta-analysis and systematic review on the effect of probiotics in acute diarrhea', *Inflammation & Allergy Drug Targets,* 11 (1), pp. 3-14.

Salzillo, M., Vastano, V., Capri, U., Muscariello, L., Sacco, M. and Marasco, R. (2015) 'Identification and characterization of enolase as a collagen-binding protein in *Lactobacillus plantarum*', *Journal of Basic Microbiology*, 55 (7), pp. 890-897.

Sanders, M. E. (2008) 'Probiotics: Definition, Sources, Selection, and Uses', *Clinical Infectious Diseases,* 46 (1) pp. S58-S61.

Scheurwater, E., Reid, C. W. and Clarke, A. J. (2008) 'Lytic transglycosylases: Bacterial space-making autolysins', *The International Journal of Biochemistry & Cell Biology*, 40 (4), pp. 586-591.

Schwille-Kiuntke, J., Enck, P., Zendler, C., Krieg, M., Polster, A. V., Klosterhalfen, S., Autenrieth, I. B., Zipfel, S. and Frick, J. (2011) 'Postinfectious irritable bowel syndrome: follow-up of a patient cohort of confirmed cases of bacterial infection with *Salmonella* or *Campylobacter*', *Neurogastroenterology & Motility,* 23 (11), pp. e479-e488.

Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics,* 30 (14), pp. 2068-9.

Serichantalergs, O., Dalsgaard, A., Bodhidatta, L., Krasaesub, S., Pitarangsi, C., Srijan, A. and Mason, C. (2007) 'Emerging fluoroquinolone and macrolide resistance of *Campylobacter jejuni* and *Campylobacter coli* isolates and their serotypes in Thai children from 1991 to 2000', *Epidemiology and Infection,* 135 (8), pp. 1299-1306.

Sharma, D. and Saharan, B. S. (2016) 'Functional characterization of biomedical potential of biosurfactant produced by *Lactobacillus helveticus*', *Biotechnology Reports*, 11 (2016), pp. 27-35.

Silva, G. G. Z., Dutilh, B. E., Matthews, T. D., Elkins, K., Schmieder, R., Dinsdale, E. A. and Edwards, R. A. (2013) 'Combining de novo and reference-guided assembly with scaffold_builder', *Source Code for Biology and Medicine*, 8 (1), pp. 23. doi:10.1186/1751-0473-8-23.

Singh, P. and Cameotra, S. S. (2004) 'Potential applications of microbial surfactants in biomedical sciences', *Trends in Biotechnology,* 22 (3), pp. 142-6.

SoftBerry - ProtCompB (2017). Available at: http://www.softberry.com/berry.phtml?topic=pcompb&group=programs&subgroup=proloc (Accessed: 7 June 2017).

Sonomoto, K. and Yokota, A. (2011) *Lactic acid bacteria and bifidobacteria: Current Progress in Advanced Research*. Norfolk, UK: Caister Academic Press.

Spiller, R. and Garsed, K. (2009) 'Postinfectious Irritable Bowel Syndrome', *Gastroenterology,* 136 (6), pp. 1979-1988.

Suessmuth, S. D., Muscholl-Silberhorn, A., Wirth, R., Susa, M., Marre, R and Rozdzinski, E. (2000) 'Aggregation Substance Promotes Adherence, Phagocytosis, and Intracellular Survival of *Enterococcus faecalis* within Human Macrophages and Suppresses Respiratory Burst', *Infection and Immunity,* 68(9), pp. 4900-4906.

Svetoch, E. A., Eruslanov, B. V., Levchuk, V. P., Perelygin, V. V., Mitsevich, E. V., Mitsevich, I. P., Stepanshin, J., Dyatlov, I., Seal, B. S. and Stern, N. J. (2011) 'Isolation of *Lactobacillus salivarius* 1077 (NRRL B-50053) and Characterization of Its

Bacteriocin, Including the Antimicrobial Activity Spectrum', *Applied and Environmental Microbiology*, 77 (8), pp. 2749-2754.

Szajewska, H., Guarino, A., Hojsak, I., Indrio, F., Kolacek, S., Shamir, R., Vandenplas, Y., Weizman, Z and European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. (2014) 'Use of probiotics for management of acute gastroenteritis: a position paper by the ESPGHAN Working Group for Probiotics and Prebiotics', *Journal of Pediatric Gastroenterology and Nutrition*, 58 (4), pp. 531-539.

Tajabadi, N., Mardan, M., Saari, N., Mustafa, S., Bahreini, R. and Mohd Yazid, A. M. (2013) 'Identification of *Lactobacillus plantarum, Lactobacillus pentosus* and *Lactobacillus fermentum* from honey stomach of honeybee', *Brazilian Journal of Microbiology*, 44 (3), pp. 717-722.

Takahashi, M., Taguchi, H., Yamaguchi, H., Osaki, T., Komatsu, A. and Kamiya, S. (2004) 'The effect of probiotic treatment with *Clostridium butyricum* on enterohemorrhagic *Escherichia coli* O157:H7 infection in mice', *FEMS Immunology & Medical Microbiology*, 41 (3), pp. 219-226.

Takeda, K., Suzuki, T., Shimada, S., Shida, K., Nanno, M. and Okumura, K. (2006) 'Interleukin-12 is involved in the enhancement of human natural killer cell activity by *Lactobacillus casei* Shirota', *Clinical & Experimental Immunology*, 146 (1), pp. 109-115.

Tareb, R., Bernardeau, M., Gueguen, M. and Vernoux, J. P. (2013) 'In vitro characterization of aggregation and adhesion properties of viable and heat-killed forms of two probiotic *Lactobacillus* strains and interaction with foodborne zoonotic bacteria, especially *Campylobacter jejuni*', *Journal of Medical Microbiology*, 62 (4), pp. 637-649.

Treangen, T. J. and Salzberg, S. L. (2011) 'Repetitive DNA and next-generation sequencing: computational challenges and solutions', *Nature Reviews Genetics*, 13 (1), pp. 36-46.

Tribble, D. R., Sanders, J. W., Pang, L. W., Mason, C., Pitarangsi, C., Baqar, S., Armstrong, A., Hshieh, P., Fox, A., Maley, E. A., Lebron, C., Faix, D. J., Lawler, J. V., Nayak, G., Lewis, M., Bodhidatta, L. and Scott, D. A. (2007) 'Traveler's Diarrhea in Thailand: Randomized, Double-Blind Trial Comparing Single-Dose and 3-Day Azithromycin-Based Regimens with a 3-Day Levofloxacin Regimen', *Clinical Infectious Diseases*, 44 (3), pp. 338-346.

Tsai, I. J., Otto, T. D. and Berriman, M. (2010) 'Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps', *Genome Biology*, 11 (4), pp. R41.

Turner, P. C., Wu, Q. K., Piekkola, S., Gratz, S., Mykkänen, H. and El-Nezami, H. (2008) *'Lactobacillus rhamnosus* strain GG restores alkaline phosphatase activity in differentiating Caco-2 cells dosed with the potent mycotoxin deoxynivalenol', *Food and Chemical Toxicology*, 46 (6), pp. 2118-2123.

Van Domselaar, G. H., Stothard, P., Shrivastava, S., Cruz, J. A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R. and Wishart, D. S. (2005) 'BASys: a web server for automated bacterial genome annotation', *Nucleic Acids Research*, 33 (2), pp. W455-W459.

Van Doorn, P. A., Ruts, L. and Jacobs, B. C. (2008) 'Clinical features, pathogenesis, and treatment of Guillain-Barré syndrome, *Lancet Neurology*, 7 (10), pp. 939-50

Verbeke, S., Gotteland, M., Fernandez, M., Bremer, J., Rios, G. and Brunser, O. (2002 'Basement membrane and connective tissue proteins in intestinal mucosa of patients with coeliac disease', *Journal of Clinical Pathology*, 55 (6), pp. 440-445.

Verdenelli, M. C., Coman, M. M., Cecchini, C., Silvi, S., Orpianesi, C. and Cresci, A. (2014) 'Evaluation of antipathogenic activity and adherence properties of human *Lactobacillus* strains for vaginal formulations', *Journal of Applied Microbiology*, 116 (5), pp. 1297-1307.

Vernocchi, P., Chierico, F. D. and Putignani, L. (2016) 'Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health', *Frontiers in Microbiology*, 7 (2016), pp. 1144.

Vijaya, K. G., Lee, E. M., Philip, J. G. and Mark, A. M. (2013) 'Probiotics: History and Evolution', *Journal of Infectious Diseases & Preventive Medicine*, 1 (2). doi:10.4172/2329-8731.1000107.

Waligora-Dupriet, A. and Butel, M. (2012) "Microbiota and Allergy: From Dysbiosis to Probiotics", *Allergic Diseases - Highlights in the Clinic, Mechanisms and Treatment*. doi: 10.5772/26234.

Walsh, A. M., Crispie, F., Claesson, M. J. and Cotter, P. D. (2016) 'Translating Omics to Food Microbiology', *Annual Review of Food Science and Technology*, 8 (2017), pp. 113-134

Wehkamp, J., Harder, J., Wehkamp, K., Meissner, B. W., Schlee, M., Enders, C., Sonnenborn, U., Nuding, S., Bengmark, S., Fellermann, K., Schroder, J. M. and Stange, E. F. (2004) 'NF-kappaB- and AP-1-Mediated Induction of Human Beta Defensin-2 in Intestinal Epithelial Cells by *Escherichia coli* Nissle 1917: a Novel Effect of a Probiotic Bacterium', *Infection and Immunity*, 72 (10), pp. 5750-5758.

Wine, E., Gareau, M., Johnson-Henry, K. and Sherman, P. M. (2009) 'Strain-specific probiotic (*Lactobacillus helveticus*) inhibition of *Campylobacter jejuni* invasion of human intestinal epithelial cells', *FEMS Microbiology Letters*, 300 (1), pp. 146-152.

World Health Organisation (2017) *Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics*. WHO (Essential medicines and health products), pp. 1-7. Available at: http://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf?ua=1.

World Health Organization and, Food and Agriculture Organization of the United Nations (2001) *Health and Nutritional Properties of Probiotics in Food including Powder Milk with Live Lactic Acid Bacteria'*. WHO/FAO. Available at: http://www.fao.org/3/a-a0512e.pdf.

World Health Organization, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health (2013) *The global view of campylobacteriosis: report of an expert consultation, Utrecht, Netherlands, 9-11 July 2012*. Geneva, Switzerland: WHO. Available at: http://www.who.int/iris/handle/10665/80751 (Accessed: 22 June 2017).

Yadav, A. K., Tyagi, A., Kaushik, J. K., Saklani, A. C., Grover, S. and Batish, V. K. (2013) 'Role of surface layer collagen binding protein from indigenous *Lactobacillus plantarum* 91 in adhesion and its anti-adhesion potential against gut pathogen', *Microbiological Research*, 168 (10), pp. 639-645.

Yang, S., Lin, C., Sung, C. T. and Fang, J. (2014) 'Antibacterial activities of bacteriocins: application in foods and pharmaceuticals', *Frontiers in Microbiology*, 5 pp. 241.

Young, K. T., Davis, L. M. and DiRita, V. J. (2007) '*Campylobacter jejuni*: molecular biology and pathogenesis', *Nature Reviews Microbiology,* 5 (9), pp. 665-679.

Zenner, D. and Gillespie, I. (2011) 'Travel-Associated *Salmonella* and *Campylobacter* Gastroenteritis in England: Estimation of Under-Ascertainment Through National Laboratory Surveillance', *Journal of Travel Medicine,* 18 (6), pp. 414-417.

Zhang, Y., Li, L., Guo, C., Mu, D., Feng, B., Zuo, X. and Li, Y. (2016) 'Effects of probiotic type, dose and treatment duration on irritable bowel syndrome diagnosed by Rome III criteria: a meta-analysis', *BMC Gastroenterology,* 16 (1), pp. 62.

Zhou, C., Ma, F. Z., Deng, X. J., Yuan, H. and Ma, H. S. (2008) '*Lactobacilli* inhibit interleukin-8 production induced by Helicobacter pylori lipopolysaccharide-activated Toll-like receptor 4', *World Journal of Gastroenterology,* 14 (32), pp. 5090-5095.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. and Wishart, D. S. (2011) 'PHAST: A Fast Phage Search Tool', *Nucleic Acids Research,* 39 (2), pp. W347-W352.

Zielenkiewicz, U., Kowalewska, M., Kaczor, C. and Ceglowski, P. (2009) 'In Vivo Interactions between Toxin-Antitoxin Proteins Epsilon and Zeta of *Streptococcal* Plasmid pSM19035 in *Saccharomyces cerevisiae*', *Journal of Bacteriology,* 191 (11), pp. 3677-3684.

Zyrek, A. A., Cichon, C., Helms, S., Enders, C., Sonnenborn, U. and Schmidt, M. A. (2007) 'Molecular mechanisms underlying the probiotic effects of *Escherichia coli* Nissle 1917 involve ZO-2 and PKCzeta redistribution resulting in tight junction and epithelial barrier repair', *Cellular Microbiology,* 9 (3), pp. 804-816.