Characterisation of the *Campylobacter jejuni* PEB3 and GlpT

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the The School of Biological Sciences/Faculty of Biology, Medicine and Health.

2016

Deborah Ruth Cantu

Table of Contents

Final Word Count: 45,811

List of Figures

List of Tables

List of Symbols and Abbreviations

Symbol	Definition
µl	Microliter
µg	Microgram
µm	Micrometre
ml	Millilitre
g	Gram
ng	Nanogram
g	Gravitational force
rpm	Revolutions per minute
kDa	Kilo Dalton
kbp	Kilo base pairs
v/v	Volume per volume
w/v	Weight per volume
° C	Degree Celsius
M	Molar
mM	Millimolar
Mr	Molecular mass
psi	Pounds per square inch
V	Volts
kv	Kilo volt
µf	Micro-farad
Ω	Ohms
OD_{600}	Optical density at 600 nm wavelength
L	Litre
µm	Micromolar
mw	Molecular weight
pmol	Picomole

Abbreviation	Definition
3-PG	3-Phosphoglycerate
APS	Ammonium persulphate
ATP	Adenosine triphosphate
bp	Base pair
CBA	Columbia blood agar
CC-ST	Clonal complex-Sequence Type
CDD	Conserved domain database
CFU	Colony forming units
dH_20	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
ECM	Extracellular matrix
FQ	Fluoroquinolone
G3P	Glycerol-3-phosphate
GAG	Glycosaminoglycan
GBS	Guillain-Barre Syndrome
GST	Glutathione S-transferase
HA	Human influenza hemagglutinin
His	Histidine
IPTG	Isopropyl-β-D-thiogalactopyranoside
LB	Luria Bertani growth media
LOS	Lipooligosaccharide
MCLMAN	Medium cysteine leucine methionine aspartic acid niacinamide
MFS	Miller Fisher Syndrome
MEM	Minimal essential media
MH	Muller Hinton medium
MIC	Minimum inhibitory concentration
MLST	Multi-locus sequence typing
MOPS	3-(n-morpholino) propanesulfonic acid
NCBI	National Center for biotechnology information
NCTC	National Collection of Type
PBS	Phosphate buffered saline
PBST	PBS with Tween
PCR	Polymerase chain reaction
PDB	Protein data bank
pH	Power of hydrogen
pI	Isoelectric point
Pi	Internal phosphate
RGD	Arginine-glycine-aspartic acid region
SDS	Sodium dodecyl sulphate polyacrylamide
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SOC	Super optimal broth with catabolite repression
ST	Sequence type
TAE	Tris-acetate-EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
UPGMA	Un-weighted pair group method with arithmetic
WT	Wild type

Abstract

The University of Manchester

Deborah Ruth Cantu

Characterisation of the *Campylobacter jejuni* PEB3 and GlpT

May 2016

The pathogen *C. jejuni* is now recognised as the leading cause of bacterial foodborne enteritis in the industrial world. The yearly estimate for *Campylobacter* infections in the United States alone is 2.4 million people or 1% of the population. Illness caused by *C. jejuni* is self-limiting, however, some individuals develop complications resulting in autoimmune responses. Despite being a major health burden, the pathogenic process is not fully understood. One aspect of importance is the ability of *C. jejuni* to adhere to glycosaminoglycans (GAGs), such as heparin. GAGs, sulphated carbohydrates expressed on or in host cells, can serve as receptors for bacterial proteins. In the first study, five heparin-binding proteins of *C. jejuni* NCTC 11168H were identified. For PEB3 (Cj0289c), this work showed that native wild-type PEB3 and purified recombinant PEB3 produced in *E. coli* bind heparin. The location of two PEB3 heparin-binding clusters: $_{62}$KAKKD$_{65}$ and $_{122}$NKKVRI$_{127}$, was investigated via site-directed mutagenesis, resulting in impaired heparin-binding. These data suggest GAG-protein-binding may play a role in the pathogenesis of *C. jejuni*. As well as GAG-binding PEB3 binds 3-PG. Though its exact *in vivo* role remains unclear, it may act to deliver 3-PG. Scrutiny of the *C. jejuni* NCTC 11168H genome revealed an uncharacterised gene next to *peb3* encoding *glpT,* or a putative 3-PG transporter. The location of *glpT* adjacent to *peb3* may suggest a related function for the corresponding proteins with PEB3 as the periplasmic binding partner for the transport of 3-PG via GlpT. In this thesis, the roles of *peb3* and *glpT* for two independent phenotypes, 3-PG dependent growth and fosfomycin sensitivity was studied *in vitro*. The findings indicate *glpT* has an effect on both, but not *peb3*. Furthermore, the NCTC 11168H *glpT* pseudogene, despite containing two frameshift mutations, has the capacity to encode a functional protein. Lastly, the NCTC 11168H *peb3/glpT* locus was compared with other *C. jejuni* strains and closely related species *C. coli*, *C. lari* and *C. upsaliensis* genome sequences. The majority of strains *peb3/glpT* locus followed the gene arrangement *lpxB*, *peb3*, *glpT*, *surE*. However, the findings indicate the gene loci between *lpxB/surE* in remaining strains to be hypervariable. Further analysis shows *peb3* to be relatively conserved, whereas, the majority of *glpT* genes display genetic diversity due to interruptions such as indels and deletion. Lastly, I display the organisation of the *peb3/glpT* locus and *glpT* structure in their evolutionary context through MLST. In summary, the findings provide for further characterisation of the PEB3 protein and explores the importance of the uncharacterised GlpT of *C. jejuni*.

Declaration:

The author declares that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I would like to thank my Supervisor Dennis Linton for his guidance and patience he has shown throughout my PhD. I am thankful for the opportunity to work in his lab to complete this research and for his help in the writing process. Without his support I doubt I would have reached this point.

I would like to thank both Dr. Jonathan Butler and my part-time desk partner Dr. Peter Elliott for your advice, moral support and friendship. I would also like to thank other members of the Linton Lab group, Dr. Adrian Jervis and Liz Lord, for your help. Thank you to the University of Manchester Protein Mass Spectrometry department for all your help and guidance.

Finally, thank you to my husband Andy Watson for presentation help, your sense of humour and understanding my commitment to my work. A special thanks to my parents, Dr. George Cantu and Dr. Rosita Cantu, who have supported me throughout my Masters and PhD. Thank you for your emotional and financial support throughout the hardest years of my life. I would not be able to do this without you.

CHAPTER 1

Introduction

1.1 Historical perspectives of the *Campylobacter* genus

German bacteriologist Theodor Escherich first described *Campylobacters* in 1886, whereby he observed, in stool samples of infants with enteric disease, the existence of non-cultivable, spiral-shaped bacteria (Escherich, 1886; Snelling et al., 2005). Almost thirty years later, in 1913, veterinary surgeons Stewart Stockman and John MacFadyean discovered a "Vibrio-like organism" in the foetal tissues of ewes (McFadyean & Stockman, 1913). Then in 1919, Smith and Taylor successfully isolated a similar, spirillum-type, organism from aborted foetuses of cattle which was later termed '*Vibrio fetus*' (Smith & Taylor, 1919). Prior to 1963, classifications of bacteria depended on cell shape. This resulted in the initial assignment of these "Vibrio-like organisms" into the *Vibrio* genus. However, Sebald and Véron showed specific differences in the fermentative metabolism and G+C DNA composition of these "Vibrio-like organisms" and true *Vibrios*. This led to the reclassification, and the genesis of a new genus called *Campylobacter* (Sebald & Véron, 1963). The *Campylobacter* genus includes the order Campylobacterales comprised of Gram-negative generae *Helicobacter* and *Wolinella*. Along with *Campylobacter*, these genera form the epsilon division of Proteobacteria (Figure 1.1). The *Campylobacter* genus comprises twenty species and six subspecies (Fernandez et al., 2008; Silva et al., 2011).

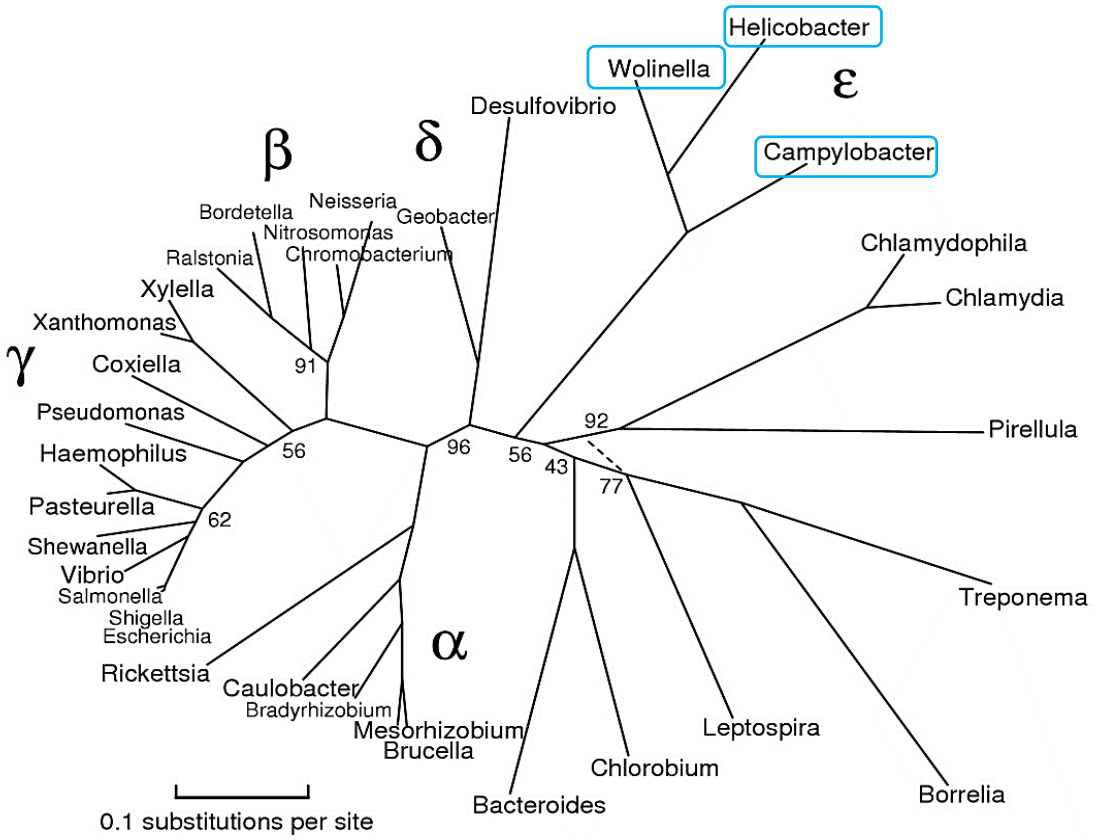


Figure 1.1 The five classes of Proteobacteria. This phylogenetic tree shows the five classes of Proteobacteria based on genomic data. The blue square indicates the epsilon division of Proteobacteria, comprised of *Campylobacter*, *Wolinella* and *Helicobacter*. Image acquired from Bern & Goldberg, 2005.

1.2 *C. jejuni*- the microbe

C. jejuni is a Gram-negative, helical, non-spore forming spiral shaped rod. The cell (0.2 to 0.8 µm x 0.5 to 5 µm) is highly motile due to a single flagella present at one or both ends of the cell (Figure 1.2) (Silva et al., 2011). As a thermophilic organism, *C. jejuni* grows at an optimal temperature range of 37 °C (humans) to 42 °C (Altekruse et al., 1999; Davis & DiRita, 2008). The 42 °C reflects the elevated temperature of the chicken gut, where *C. jejuni* resides as commensal (van Vliet & Ketley, 2001; Snelling et al., 2005). *C. jejuni,* unable to survive aerobically or anaerobically, requires a microaerophilic atmosphere composed of 10 % CO_2, 85 % N_2, and low oxygen content (5 % O_2) (Altekruse et al., 1999; Snelling et al., 2005).

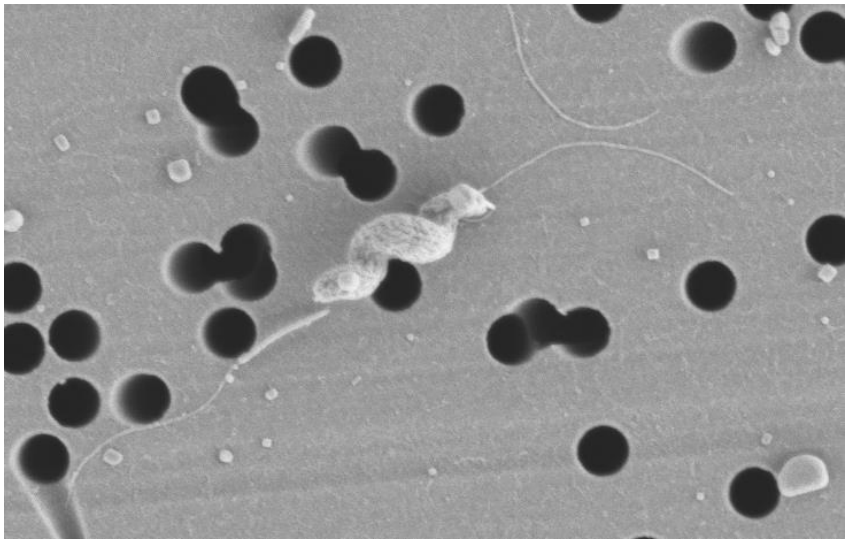


Figure 1.2 Scanning electron micrograph of flagellated *C. jejuni*. The image shows the helical morphology of the *C. jejuni* cell containing a single flagellum at both poles. Image acquired from Kathryn Cross of the Institute of Food Research (http://blogs.ifr.ac.uk/ghfs/2013/04/sgm-poster-prize/#lightbox/2/).

1.3 Epidemiology of Campylobacteriosis

Campylobacter is now recognised as the leading cause of bacterial food poisoning worldwide (Allos, 2001). In the United States, an estimated 2.4 million people or ~1 % of the population contract campylobacteriosis (Allos, 2001; Acheson & Allos, 2001). *C. jejuni* accounts for 90 % of cases with *C. coli* accounting for the remaining 10 % (Gillespie et al., 2002; Keener et al., 2004).

Cases of campylobacteriosis follow seasonal patterns. In developed regions, peaks of infection are in the spring and early summer months (Humphrey et al., 2007). Interestingly, peaks of infection are highest among young adults (15-30) and children. Developing regions show similar infection rates among young adults (Coker et al., 2002). In contrast, the infection rate in children is significantly higher in developing regions (Oberhelman & Taylor, 2000; Coker et al., 2002). This is explained by poor hygiene/sanitation, untreated drinking water/contaminated food and lack of healthcare, surveillance and knowledge of *C. jejuni* infections (Coker et al., 2002; O'Ryan et al., 2005). In developing areas, campylobacteriosis is a paediatric disease associated with children aged <2 (Coker et al., 2002; O'Ryan et al., 2005). Interestingly, breastfed children are protected from infection by maternal antibodies in breast milk (Ruiz-Palacios et al., 1990; Coker et al., 2002).

1.4 *C. jejuni* sources of infection

Poultry is the primary vehicle of campylobacteriosis transmission in man (Whiley et al., 2013). Studies show 50-70 % of cases arise from consumption of contaminated chicken products (Allos, 2001; Keener et al., 2004). This is not surprising as surveillance studies in the United States show at least 90 % of chicken flocks are colonised with *Campylobacter* (Stern et al., 2001; Whiley et al., 2013). The spread of *Campylobacter* from a single bird to a whole flock is rapid, especially among hatchlings, due to high levels of bacteria (Shanker et al., 1990; Whiley et al., 2013). The estimated bacterial counts range from 10^5-10^9 colony-forming units (CFU) per

gram of faeces (Keener et al., 2004). In chicken, *C. jejuni* colonises the mucus layer of the caecal crypts, but epithelial cells lining the gut are not invaded or adhered to (Beery et al., 1988; Meinersmann et al., 1995). Instead, these bacteria exist as commensals with no evidence of inflammation or epithelial cell damage (Beery et al., 1988). This contrasts with human infection whereby bacteria invade and adhere to intestinal epithelial cells, which leads to infection.

There are multiple sources of *C. jejuni* infection besides poultry, which include unpasteurized milk/juices (Van den Brandhof et al., 2003), cattle/raw red meat (Inglis et al., 2004), and sources of water (environmental/drinking) (Moore et al., 2001) (Figure 1.3). Less common vectors include domestic animals (cats/dogs), wild birds, and reptiles (Whiley et al., 2013). Tracking sources of infection is made difficult by the complex *C. jejuni* transmission cycle. However, DNA-based typing methods such as multi-locus sequence typing (MLST) aid in identifying *C. jejuni* isolates association with certain hosts. MLST indexes the variations of seven housekeeping genes, *aspA, glnA, gltA, glyA, pgm, tkt, uncA*, involved in metabolism (Dingle et al., 2001). These highly conserved genes are used to identify genetic lineages within a *C. jejuni* population. The seven locus alleles define the sequence type (ST) which are then grouped into clonal complexes (ST-CC) based on shared identical alleles at four or more loci (Dingle et al., 2001).

1.5 *C. jejuni* pathogenesis and virulence factors

1.5.1 Clinical symptoms, antibiotic resistance and complications of *C. jejuni* infection

Campylobacteriosis requires a low infective dose, ranging from 500-800 cells (Black et al., 1988). After an incubation period of 24-72 hours, abdominal pain, fever, inflammation and diarrhoea develop (Allos, 2001; Zilbauer et al., 2008). The severity of the diarrheal illness varies from mild to bloody, but, is often self-limiting, resolving in 5-7 days (Blaser, 1997; Allos, 2001; van Vliet & Ketley, 2001). Disease outcome depends on the immune status of

the individual and virulence of the infecting strain (Zilbauer et al., 2008). At-risk groups (i.e. elderly and immunocompetent) or those with persistent infection may need antimicrobial treatment (Allos, 2001; Coker et al., 2002; Luangtongkum et al., 2009).

The antibiotic fluoroquinolone (FQ), was the first choice in antimicrobial treatment. However, many isolated *C. jejuni* strains are FQ-resistant, therefore reducing its effectiveness (Alfredson & Korolik, 2007; Wieczorek & Osek, 2013). Resistance to penicillin, tetracycline, metronidazole, and bactericidal β-lactam antibiotics has also been observed (Allos, 2001). The overuse of antibiotics in humans, veterinary medicine, and the farm industry contribute to increased resistance (Endtz et al., 1991; Witte, 2000; Teuber, 2001; Luangtongkum et al., 2009).

Complications of *C. jejuni* infection may result in autoimmune diseases, namely, Guillain-Barre Syndrome (GBS), Miller Fisher Syndrome (MFS) and Reactive Arthritis (ReA) (Willion & O'Hanlon, 1999; Pope et al., 2007; Nyati & Nyati, 2013). GBS and MF, are thought to be caused by molecular mimicry between *C. jejuni* lipo-oligosaccharide (LOS) and human gangliosides (Yuki et al., 1993; Godschalk et al., 2004; Yuki et al., 2004). At least 30-40 % of GBS cases are associated with prior *C. jejuni* infection (Kuwabara, 2004; Vucic et al., 2009). One to three weeks after the diarrhoeal stage GBS develops causing weakness, respiratory distress and sensory loss (Nachamkin et al., 1998; Nyati & Nyati, 2013). In some instances, patients may need ventilation equipment, remain bed bound or lose the ability to walk (Asbury & Cornblath, 1990; Nachamkin et al., 1998; Dimachkie & Barohn, 2013). MFS, a GBS variant (5-10 % of GBS cases), is characterised by paralysis/weakness of eye muscles, loss of tendon reflexes and failure of muscle coordination (Dimachkie & Barohn, 2013). Lastly, ReA is triggered by several bacterial species including *Yersinia, Salmonella* spp. and *C. jejuni* (Braun et al., 1999; Rees et al., 2004; Pope et al., 2007).

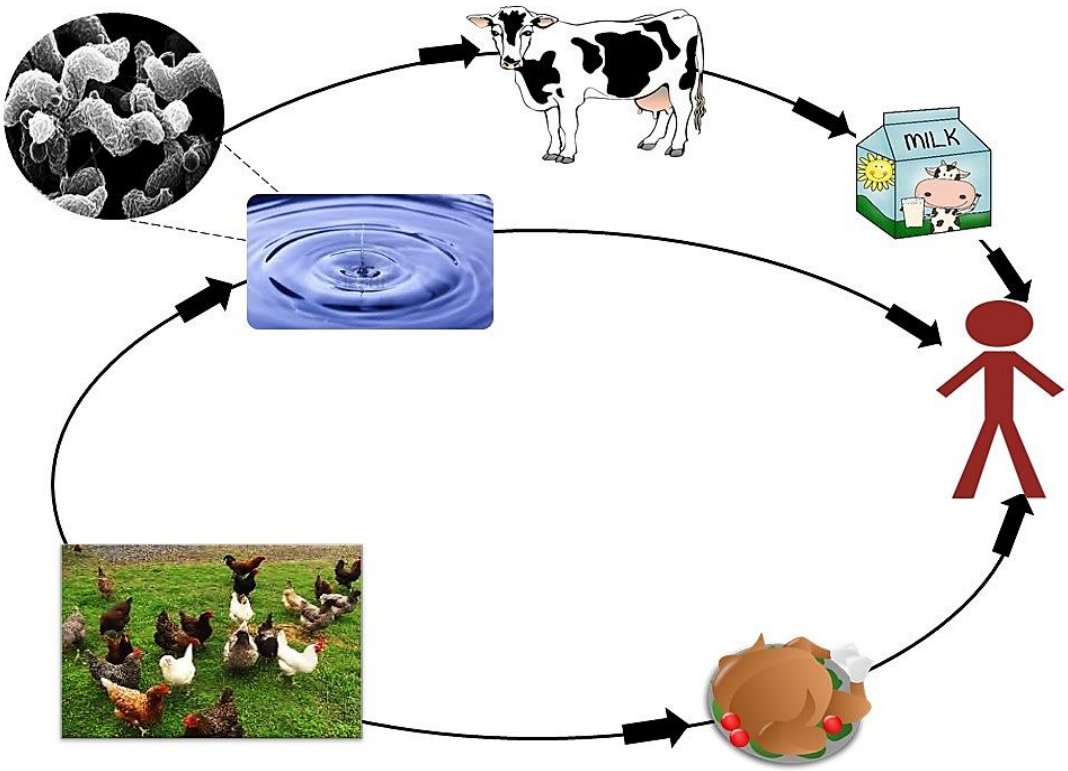


Figure 1.3 The ecological cycle of *C. jejuni* involves animals, food and drinking sources. *C. jejuni* colonises the mucosal layer of the chicken gut where it lives as a commensal. From here *C. jejuni* can be passed throughout the flock rapidly via the faecal-oral route. This also serves as an indirect source of contamination for water sources such as lakes, rivers, and drinking water. *C. jejuni* can be passed on to humans through consumption of contaminated drinking sources, unpasteurized milk, and cattle/raw red meat products. The image was acquired and adapted from Young et al., 2007.

This autoimmune disease is characterised by inflammation of the joints, eyes and the urogenital/gastrointestinal systems, which develop within four weeks of *C. jejuni* infection (Kvien et al., 1994; Pope et al., 2007).

1.5.2 *C. jejuni* virulence factors

In humans, *C. jejuni* must survive the acidic conditions of the stomach to cause infection (Reid et al., 2008). The surviving organisms continue to the gastrointestinal tract aided by motility and chemotaxis. Once *C. jejuni* reaches the gut it penetrates, invades and adheres to the epithelial cells lining the mucus layer via virulence factors (Ketley, 1997; van Vliet et al., 2000; Bereswill & Kist, 2003; Snelling et al., 2005; Guerry, 2007; Zilbauer, 2008; Dasti et al., 2010; Silva et al., 2011; Lertsethtakarn et al., 2011). This section will discuss the virulence factors identified in aiding the survival and progression of *C. jejuni* disease in the human host.

1.5.3 Flagella

Under phase-contrast microscopy, *C. jejuni* has a rapid corkscrew-like motility, which aids *C. jejuni* in penetrating the gut (Ferrow & Lee, 1988; Szymanksi et al., 1995; Snelling et al., 2005). In *C. jejuni*, motility is produced by flagella at one or both ends of the cell. The *C. jejuni* flagellum comprises a membrane-embedded motor, hook-basal body and filament (Lertsethtakarn et al., 2011) composed of two homologous (>95 %) flagellin proteins FlaA and FlaB (Guerry et al., 1991; Nuijten et al., 1990; Guerry, 2007). Flagellin genes encoding FlaA and FlaB are subject to phase and antigenic variation, which masks detection by the immune system (Harris et al., 1987; Young et al., 2007). As well as motility, flagella are also involved in adherence, penetration/colonisation of the gut and secretion of invasive proteins (van Vliet & Ketley, 2001; Snelling et al., 2005; Guerry, 2007).

Campylobacter invasive antigens (Cia)

The flagellar export apparatus (Figure 1.4) also secretes so called invasion proteins, FlaC (Song et al., 2004) and Cia (Konkel et al., 2004). Eight *Campylobacter* invasion antigens (Cia), CiaA-H, are known (Konkel et al., 1999a; Konkel et al., 1999b; Rivera-Amill et al., 2001). Of these CiaB is well characterised and is secreted in the presence of chicken mucus, serum host cell components and induced by bile salts (Konkel et al., 1999b; Konkel et al., 2001; Malik-Kale et al., 2008; Ó Cróinín & Backert, 2012).

FlaC

As well as Cia proteins, the FlaC protein is secreted (Song et al., 2004). FlaC, a 26 kDa protein, possesses sequence similarity with the N- and C-terminal regions of FlaA and FlaB (Song et al., 2004; Guerry, 2007; Young et al., 2007). Unlike FlaA and FlaB, it is not required for the formation or assembly of the flagellum, instead FlaC is required for host cell invasion (Song et al., 2004; Dasti et al., 2010; Ó Cróinín & Backert, 2012).

1.5.4 *C. jejuni* chemotactic behaviour

C. jejuni uses chemotaxis to move up or down chemical gradients via a complex transduction pathway, which controls the direction and rotation of the flagella (Wadhams & Armitage, 2004; Lertsethtakarn et al., 2011). *C. jejuni* is chemotactically attracted to the organic acids pyruvate, succinate, fumarate; carbohydrates such as L-fucose; the amino acids L-aspartate, L-glutamate, L-serine, and L-cysteine and mucin components (Hugdahl et al., 1988; Hendrixson et al., 2001; Vegge et al., 2009; Hartley-Tassell et al., 2010; Korolik et al., 2010; Lertsethtakarn et al., 2011; Hofreuter, 2014). Repellents for *C. jejuni* include bile constituents, namely cholic, deoxycholic, taurocholic and glycocholic acids (Vegge et al., 2009; Hofreuter, 2014).

1.5.5 Lipooligosaccharide (LOS) and Capsular Polysaccharide

C. jejuni expresses LOS, a glycolipid (Figure 1.4) (Karlyshev et al., 2005). In some strains, LOS structure is identical to human gangliosides found on the surface of nerve cells which causes molecular mimicry. In GBS and MFS, auto-reactive antibodies that are produced attack nerve cells (Nachamkin et al., 1998; Godschalk et al., 2004; Yuki et al., 2004). The genes for LOS biosynthesis are located on a hypervariable locus and to date 19 classes of LOS are known (Gilbert et al., 2002; Karlyshev et al., 2005; Parker et al., 2005). The genetic variation of the LOS seems to have arisen through intra-strain variability, homopolymeric runs due to phase variation, point mutations, changes in glycosyltransferases specificity and environmental factors (Parkhill et al., 2000; Linton et al., 2000; Gilbert et al., 2002; Karlyshev et al., 2005; Semchenko et al., 2010). Strains with mutated LOS biosynthesis genes exhibit a reduction in adherence and invasion of host cells along with decreased serum resistance (Young et al., 2007).

C. jejuni produces a capsular polysaccharide, that coats the outside of the cell (Figure 1.4) (Karlyshev et al., 2000; Young et al., 2007). The capsule protects against dehydration, phage infection, and host immune response by serum resistance (Karlyshev et al., 2005). The capsule is composed of varying sugar compositions and linkage units loosely associated with the cell surface (St. Michael et al., 2002; Corcoran et al., 2006; Guerry et al., 2012). The *C. jejuni* NCTC 11168 genome contains a **c**a**ps**ule loci termed *kps*. The *kps* loci genes are necessary for capsule synthesis and translocation to the cell surface (Parkhill et al., 2000; Karlyshev et al., 2005). Like the LOS, variation in the *kps* locus arises through horizontal transfer, gene duplication and presence of homopolymeric G tracts (Bacon et al., 2001; Karlyshev et al., 2005; Guerry et al., 2012).

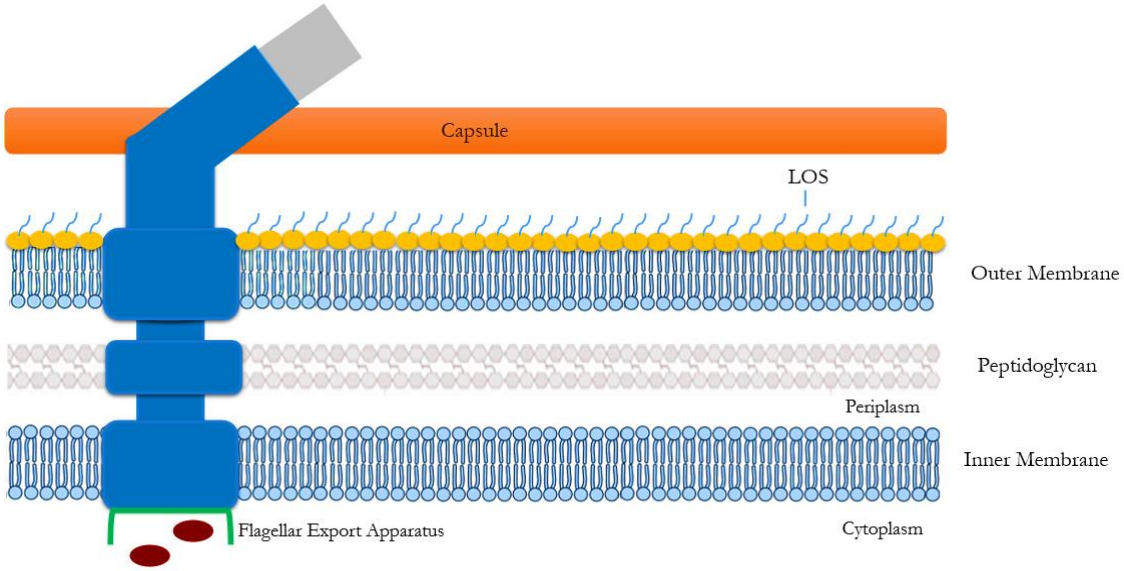


Figure 1.4 *C. jejuni* cell structures. The *C. jejuni* cell expresses several structures including capsule, (LOS), flagellin and secretes proteins via the flagellar export apparatus. These structures play vital roles in the pathogenesis of *C. jejuni*. The flagellum (blue/grey) is required for many roles in colonization, invasion of epithelial cells, chemotaxis and motility. In addition, the flagellum export apparatus (green) secretes the Cia and FlaC proteins (red). The capsule (orange) and LOS (yellow) have been shown to be important for cell adherence, invasion and serum resistance. Image adapted from Young et al. 2007.

The genetic variation of the capsule structure can be used for the Penner scheme of serotyping (Penner & Hennessy, 1980). This scheme uses passive haemagglutination to categorise *Campylobacter* clinical isolates, including those from patients with MFS and GBS. At present, a total of 47 different serotypes for *C. jejuni* have been identified highlighting the extent of capsule structural variation (Preston & Penner, 1989; Karlyshev et al., 2000; Guerry et al., 2012).

1.5.6 *C. jejuni* Adhesins

C. jejuni expresses surface exposed proteins known as adhesins that adhere to epithelial cells, preventing removal by the gut (Rubinchik et al., 2012). Interestingly, the degree of adherence to cells correlates with the severity of clinical symptoms in infected individuals (Fauchère et al., 1986). A number of adhesins with roles in virulence, invasion, colonisation and activating signalling pathways have been identified for *C. jejuni*. Among these are the JlpA, CadF, FlpA and Pei, Ellison & Blaser (PEB) 1-4 proteins.

The *jlpA* gene encodes for the Jejuni LipoProtein A (JlpA), a loosely associated outer membrane lipoprotein of 42 kDa (Jin et al., 2001). JlpA binds the host cell receptor Hsp90a on the surface of HeLa cells signalling pathways for NF-κB factor and p38 MAP Kinase (Jin et al., 2001; Jin et al., 2003).

The *Campylobacter* adhesin to fibronectin (CadF), a 37 kDa outer membrane protein binds fibronectin and contains a fibronectin-binding domain at amino acid positions 134-137 (Konkel et al., 1997; Konkel et al., 1999c; Konkel et al., 2005; Rubinchik et al., 2012). Interestingly, *cadF* mutants are unable to colonise the caecum in baby chicks (Ziprin et al., 1999). Further studies confirm the importance of CadF in attachment and invasion of host epithelial cells (Monteville et al., 2003). A second protein, the fibronectin-like protein A, binds fibronectin through a Type III fibronectin-binding domain (Konkel et al., 2010).

Mutations in the *flpA* gene result in reduced ability to adhere to epithelial cells and colonise the caecum of broiler chickens (Flanagan et al., 2009).

The PEB proteins (1-4), first described as cell binding factors, participate in a multitude of roles. The adhesin PEB1, is a ~25 kDa periplasmic protein that shares homology with proteins involved in amino acid transport (Pei et al., 1993; Del Rocio Leon-Kempis et al., 2006). The PEB1 protein, a component of the (PEB1 system) ABC transporter system, is required for growth on dicarboxylic amino acids (Del Rocio Leon-Kempis et al., 2006; Müller et al., 2007). Deficient *peb1* strains exhibit decreased attachment and invasion of host cells (Pei et al., 1998; Del Rocio Leon-Kempis et al., 2006). Additionally, this protein will be discussed in Chapter 3. The PEB4 protein, an antigenic virulence factor, is involved in adherence, protein export, colonisation and invasion (Kervella et al., 1993; Asakura et al., 2007). The protein PEB2 a major antigenic peptide of *C. jejuni* is a protein of unknown function. However, it bears a similar signal peptide sequence to PEB3 (Pei et al., 1991). The PEB3 protein is the focus of this thesis and will be discussed extensively in the following chapters.

1.6 The presence of PEB3 in *C. jejuni*

PEB3 was first identified alongside PEB1, PEB2 and PEB4 proteins (Pei et al., 1991). Orthologues of PEB3 include the AcfC adhesin of *V. cholerae* (49 % amino acid sequence identity) (Peterson & Mekalanos, 1988; Peterson, 2002) and the Paa adhesin of *E. coli* (52 % amino acid sequence identity) (Batisson et al., 2003). Analysis of the Paa amino acid sequence identified a sulfate binding domain (position 105-115) in the C-terminus which also present in PEB3 and AcfC (Batisson et al., 2003) (Figure 1.5). Interestingly, Paa and AcfC possess residues, Thr138 and Ser139, of the PEB3 ligand binding site (see below), indicating these proteins may bind a similar ligand (Figure 1.5) (Rangarajan et al., 2007).

1.7 Function of the PEB3 protein

Unpublished data identified PEB3 as a putative heparin-binding protein (Linton Lab, unpublished to date). Moreover, a study by Min and co-workers showed PEB3 binds 3-phosphoglycerate (3-PG) (Min et al., 2009). The compound 3-PG, an energy and carbon source, is found in the latter portion of the glycolytic pathway (Figure 1.6).

1.8 *C. jejuni* metabolism

C. jejuni does not metabolise common carbohydrates, such as glucose, due to the lack of enzymes (Parkhill et al., 2000; Stahl et al., 2012). Namely, the 6-phosphofructokinase (*pfk*), needed to catabolise glucose (Figure 1.6) (Parkhill et al., 2000; Velayudhan & Kelly, 2002). Instead, *C. jejuni* metabolises the amino acids serine (Velayudhan et al., 2004), aspartate (Del Rocio Leon-Kempis et al., 2006), asparagine (Vegge et al., 2009; Hofreuter, 2014), glutamate (Del Rocio Leon-Kempis et al., 2006) and proline (Wright et al., 2009). Carbon sources include short fatty acids acetate, lactate and pyruvate (Velayudhan & Kelly, 2002; Wright et al., 2009; Thomas et al., 2011; Stahl et al., 2012; Hofreuter, 2014). Interestingly, *C. jejuni* possess enzymes for the catabolism of molecules present in the later part of the glycolytic pathway such as glycerol-3-phosphate (G3P) (Hofreuter et al., 2006) and 3-PG (Min et al., 2009) (Figure 1.6). Although, PEB3 binds 3-PG the exact transport function was not determined.

However, it was proposed PEB3 may provide ligand to a transport protein for 3-PG (Min et al., 2009). Scrutiny of the *C. jejuni* NCTC 11168H chromosome revealed a completely uncharacterised gene adjacent to *peb3* encoding *glpT*, or a putative phosphoglycerate transporter.

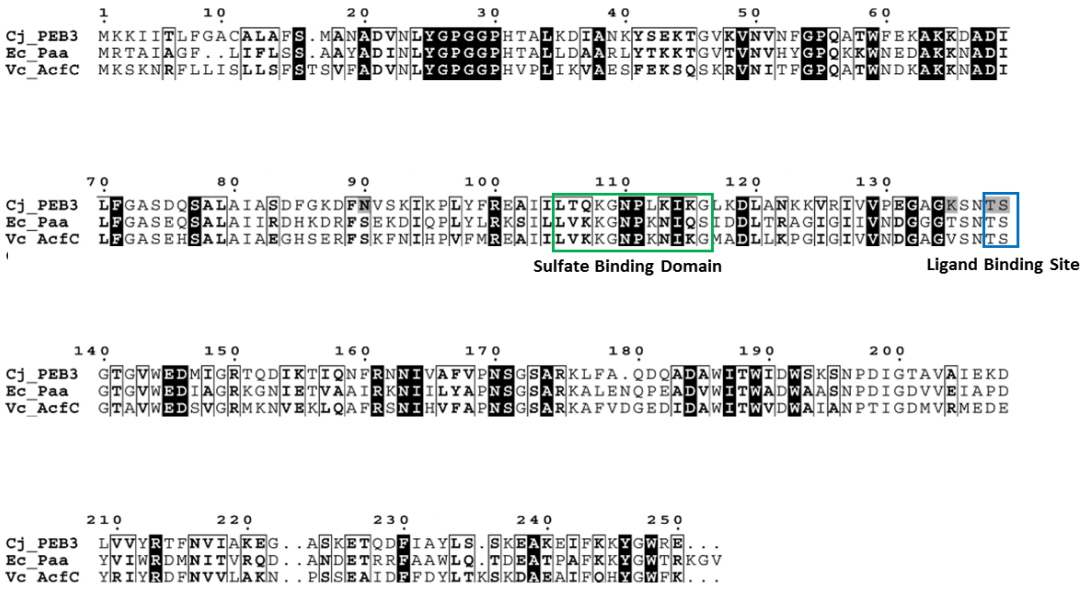


Figure 1.5 Amino acid Sequence alignment of PEB3, Paa and AcfC. Alignment of the amino acid sequences of the Paa adhesin of *E. coli*, the AcfC protein of *V. cholerae* and *C. jejuni* PEB3. The sulfate binding domain of Paa (indicated by the green box), at position 105-116, is also present in AcfC and PEB3. The conserved residues of the PEB3 ligand binding site (indicated by the blue box), at positions 138-139, are also present in AcfC and Paa. The amino acid sequence alignment was acquired and adapted from Min et al., 2009.

1.9 The GlpT transporter of *C. jejuni*

GlpT has been studied in other bacterial species such as *P. aeruginosa* (Castañeda-García et al., 2009), *Haemophilus influenzae* (Song et al., 1998), *Bacillus subtilis* (Nilsson et al., 1994) and *E. coli* (Huang et al., 2003; Lemieux et al., 2004a; Lemieux et al., 2004b; Lemieux et al., 2005). The closest homologue to *C. jejuni* GlpT is the well characterised *E. coli* GlpT. GlpT, an antiporter, exchanging external G3P for internal phosphate (Elvin et al., 1985; Lemieux et al., 2004b) (Figure 1.7). The exchange of G3P influences glycolysis, phospholipid biosynthesis, oxidative phosphorylation and pH (Lemieux et al., 2005).

In *E. coli*, *glpT* expression is induced by G3P and controlled by the *glp* regulon (Cozzarelli et al., 1968; Castañeda-García et al., 2013). In *E. coli* the *glp* regulon, composed of *glp C, B, A, T,* and *Q*, regulates the overall rate of glycerol/G3P uptake (Lemieux et al., 2004b). The expression of *glpT* is also regulated by *glp*R a repressor encoded on the *glpEGR* operon (Yang et al., 1997; Castañeda-García et al., 2013). In *E. coli* G3P, is reduced by an aerobic or anaerobic G3P dehydrogenase (Lemieux et al., 2004b). As well as G3P, *glpT* uptakes the phosphate-containing antibiotic fosfomycin that inhibits the first step in peptidoglycan synthesis leading to cell death (Nilsson et al., 2003; Castañeda-García et al., 2013).

Curiously, *E. coli* also encodes for a second antiporter UhpT that imports glucose-6-phosphate and fosfomycin (Sonna et al., 1988). The expression of *uhpT* is controlled by regulatory genes *uhpA*, *uhpB* and *uhpC* (Kadner & Shattuck-Eidens, 1983; Castañeda-García et al., 2013).

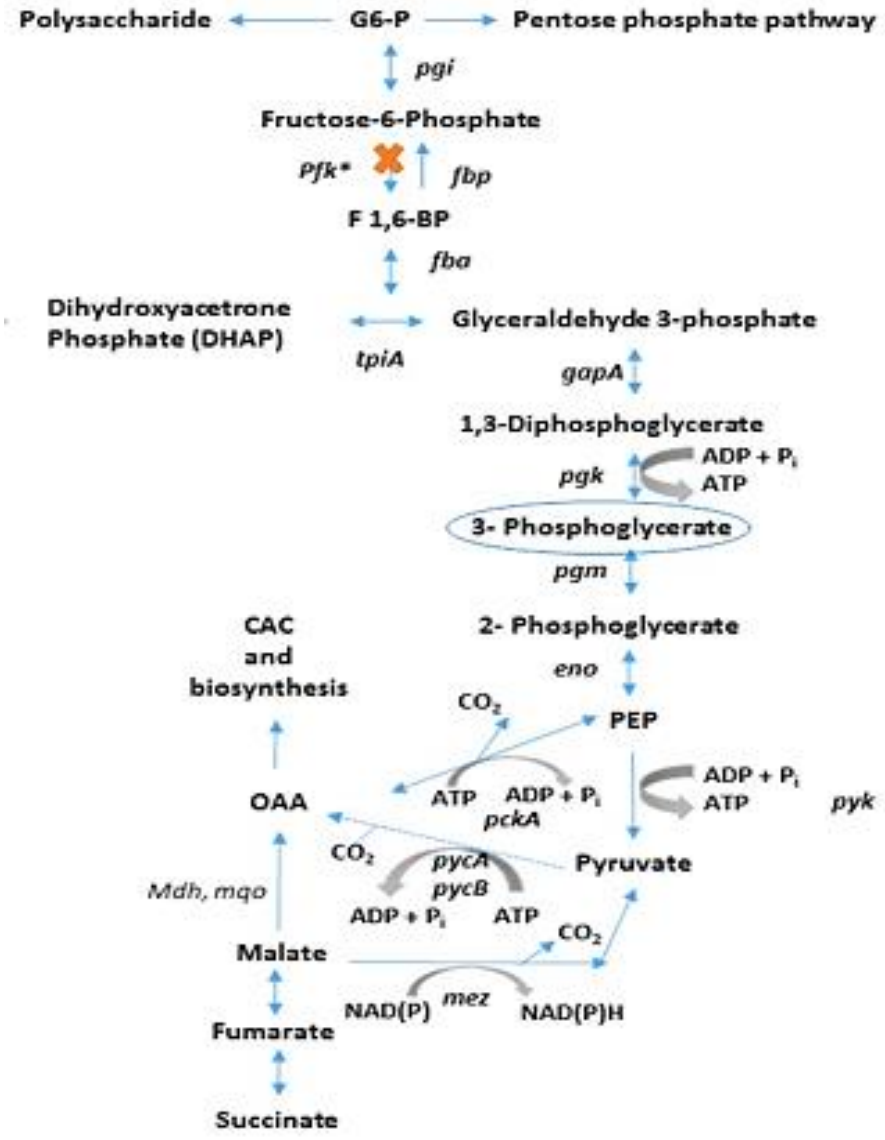


Figure 1.6 Metabolic pathway for 3-PG in *C. jejuni*. The *C. jejuni* genome lacks the 6-phosphofructokinase (*pfk*) (indicated by *), an enzyme needed to catalyse glucose. Additional enzymes of the glycolytic pathway that are present include the glucose-6-phosphate isomerase (*pgi*), fructose-1,6-bisphosphatase (*fbp*), fructose-bisphosphate aldolase (*fba*), triosephosphate isomerase (*tpiA*), glyceraldehyde-3-phosphate dehydrogenase (*gapA*), phosphoglycerate kinase (*pgk*), phosphoglycerate mutase (*pgm*) to catabolise 3-PG, enolase (*eno*), phosphoenolpyruvate carboxykinase (*pckA*), pyruvate carboxylase (*pyc*), malic enzyme (*mez*), malate dehydrogenase (*mdh*) and malate quinone oxidoreductase (*mqo*). The figure was acquired and adapted from Velayudhan & Kelly, 2002.

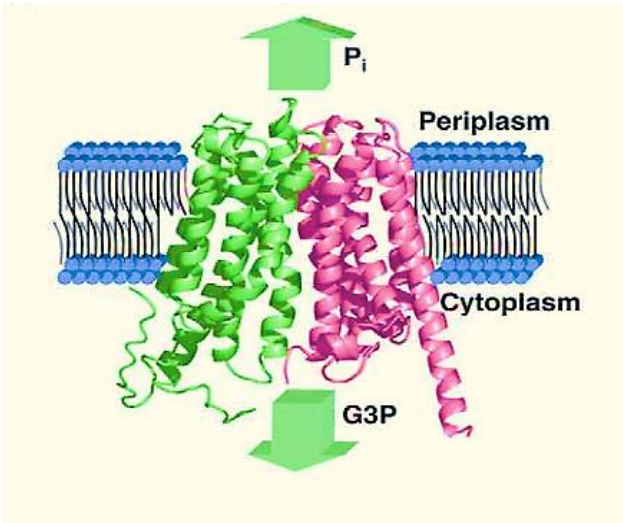


Figure 1.7 The *E. coli* GlpT structure. The GlpT protein is a membrane protein composed of 12 α helices. As an anti-porter, GlpT exchanges internal phosphate (P_i) for G3P. Figure acquired from Lemieux et al., 2004a.

In contrast with *E. coli*, the *C. jejuni* genome does not contain a *glp* regulon, *glp*R repressor or *uhpT* gene. The location of *glpT* adjacent to *peb3* might suggest a related function for the corresponding proteins. Of particular interest is to characterise and identify the role of GlpT in *C. jejuni* metabolism. Furthermore, *glpT* in *C. jejuni* 11168H is designated as pseudogene due to two frameshifting events that results in three ORFs. Interestingly, genes containing frameshifts can be processed via an alternative form of translation called programmed ribosomal frameshifting (Ketteler et al., 2012).

1.10 Programmed Ribosomal Frameshifting

Programmed ribosomal frameshifting (PRF) is a recoding mechanism whereby the translating ribosome switches from the initial reading frame (0) to a -1 or +1 reading frame. PRF occurs at shift-prone sites in signals within the mRNA, which are often downstream of RNA structural elements (Caliskan et al., 2015). Interestingly, these PRF signals are present in viruses, bacteria, higher eukaryotes and even humans (Ketteler et al., 2012). At present, two types of PRF are known, -1PRF and +1PRF.

In viruses and bacteria, -1 PRF is a well-documented mechanism used to overcome the tension from ribosome translocation (Cao & Chen, 2008; Visscher, 2016). The build-up of tension stems from the conformation changes of the anti-codon loop of tRNA and mRNA, whereby both move 9 Å in the 5' direction. The move causes the single-stranded mRNA spacer between the A and P sites and any downstream structure to be pulled towards ribosome tunnel entrance (Cao & Chen, 2008; Visscher, 2016). The pulling of these structures creates tension resulting in the pausing or stalling of the ribosome. In -1 PRF, slippage by one nucleotide in the 5' direction is thought to overcome this tension and help stabilise the system in a new ORF (-1) (Cao & Chen, 2008; Visscher, 2016). Several mRNA stimulatory elements such as a 'slippery sequence' and a secondary RNA structure aid -1PRF (Ketteler et al., 2012). The heptameric slippery sequence is composed of nucleotides X XXY

YYZ (where X: is any nucleotide, Y: is A or U, and Z- is A, U, or C) (Dinman, 2006). This crucial regulatory element guarantees base-pairing between the ribosome bound tRNA anticodon and mRNA codon before and after frameshifting (Chang, 2012; Caliskan et al., 2015). Following the slippery sequence is a spacer region of 1–15 nucleotides (Ketteler et al., 2012). The secondary RNA structure can be a simple stem loop or a pseudoknot, a complex structure composed of two or more stem-loops with intercalated stems which acts as roadblock by pausing or stalling the ribosome (Ketteler et al., 2012). Progression of the ribosome forward is impeded by the pseudoknot instead the ribosome slips back one nucleotide on the slippery sequence to continue translation in the new -1 ORF (Ketteler et al., 2012; Chang, 2012).

For retroviruses, such as HIV1, -1PRF is required for replication and infection. -1PRF is used in the production of the Gag-Pol polyprotein from mRNA, which is composed of two overlapping ORFs encoding for a capsid protein (*gag*) and a DNA polymerase (*pol*) (Biswas et al., 2004; Caliskan et al., 2015). The -1PRF signal present at the site 5'-U UUU UUA-3' upstream of a secondary RNA stem loop modulates the expression levels of the Gag protein relative to the Gag-Pol polyprotein. This ensures that the ratio of Gag/Gag-Pol (20:1) does not exceed its threshold value as this could have detrimental effects on the virus (i.e. disruption of virus assembly and maturation) (Biswas et al., 2004, Chang, 2012, Caliskan et al., 2015). In *E. coli,* the *dnaX* gene encodes for both the tau and gamma subunits of DNA polymerase III. The -1PRF stimulatory signal on *dnaX* is located at the slippery site sequence 5'-A-AAA-AAG-3'. In addition to this site, located 6 nucleotides downstream of the slippery site is a stem loop, which helps to increase -1PRF efficiency. Furthermore, -1PRF in *dnaX* is modulated and stimulated by a Shine-Dalgarno-like sequence upstream of the slippery site sequence A-AAA-AAG (Caliskan et al., 2015).

The second type of PRF, +1PRF is mechanically different from that of -1PRF. In +1PRF the ribosome pauses on a slippery sequence containing an in-frame stop codon (Ketteler et al., 2012). In this case of +1PRF, the slippery sequence with the stop codon is where the mRNA slides within the ribosome complex one nucleotide in the 3' direction (Dinman, 2006; Tsai, 2007). This breaks the tRNA and mRNA pairing in the peptidyl tRNA P site causing a mismatch between tRNA and mRNA. The mismatch causes the stop codon to be bypassed and translation is continued in the new +1 ORF due to pairing with an overlapping region (Tsai, 2007; Caliskan et al., 2015). +1PRF is well documented in *E. coli* specifically for the translation of *prfB,* which encodes for a polypeptide chain release factor release factor 2, RF2 (Dinman, 2006). In *prfB,* the mRNA has a signal for a slippery site, 5'-U CUU UGA-3' which contains an in-frame UGA termination codon recognized by RF2. If RF2 concentration levels are low the stop codon is bypassed, via +1PRF, resulting in the synthesis of full-size RF2. However, high concentrations of RF2 in the cell cause the stop codon to be read ending in the termination of RF2 (Dinman, 2006).

In summary, signals for -1PRF and +1 PRF are present in a variety of organisms (Ketteler et al., 2012; Caliskan et al., 2015). The implications of PRF include alternate mechanisms for the regulation of gene expression, modulation of RNA levels and an alternative way of producing beneficial proteins (Dinman, 2012; Chang, 2012; Ketteler et al., 2012). Further benefits of PRF include increasing the coding and variability of the genome, which may help organisms such as bacteria and viruses, to adapt to changing environments (Caliskan et al., 2015). The PRF signals also present alternative strategies for therapeutic development specifically for treating diseases caused by altered gene expression (Ketteler et al., 2012; Caliskan et al., 2015). This includes combating viruses, where even small changes in PRF efficiency can inhibit viral propagation (Dinman, 2006). Therefore, PRF signals in viruses can serve as potential targets for antiviral therapeutics (Biswas et al., 2004; Caliskan et al., 2015).

1.11 Aims and objectives

C. jejuni is the leading cause of bacterial foodborne illness worldwide. For this reason, mechanisms aiding the pathogenesis of this bacterium require further study. Heparin-binding is associated with adherence, serum resistance, evasion of the immune response and enhanced pathogenesis (Menozzi et al., 2002). Previously identified bacterial heparin-binding proteins are found to reside on the bacterial cell surface in order to interact with heparin. Therefore, I hypothesised that *C. jejuni* cell surface proteins may bind heparin as unpublished data (Linton Lab) showed the PEB3 adhesin, bound heparin. To test this hypothesis, surface adhesins from *C. jejuni* were obtained via an acid-glycine extraction, a common method for identifying cell surface proteins. The heparin affinities of these surface proteins were assessed via heparin-agarose chromatography and their identities confirmed with protein mass spectrometry. The second aim of **Chapter 3** was to determine the importance of putative heparin-binding sites on the protein, PEB3. Amino acid analysis identified two putative heparin-binding sites, $_{62}$KAKKD$_{66}$ and $_{22}$NKKVRI$_{127}$. Interestingly, both sites contained sequence motifs, XBBXB and BXBBX, often associated with heparin-binding proteins. Therefore, it was postulated that these two putative PEB3 binding sites, if surface exposed, may interact with heparin. To test this theory, putative sites were visualised with PyMol molecular docking software and the ClusPro server. Experimental procedures involved the construction of PEB3 site-directed mutants, whereby a lysine or arginine would be replaced with alanine. The heparin affinities of WT PEB3 in comparison with these PEB3 site-directed mutants were assessed with heparin-agarose chromatography.

In **Chapter 4**, the roles of PEB3 and GlpT were investigated. Previous work confirmed PEB3 binds 3-PG (Min et al., 2009). However, the findings of the study suggested PEB3 hands off the ligand to a transporter protein. Genomic analysis of the *C. jejuni* NCTC 11168H, *peb3* gene region led to the discovery of an adjacent gene *glpT*, annotated as pseudogene, encoding for a putative 3-PG transporter. The location of *peb3* adjacent to *glpT*

might suggest a related function for the corresponding proteins. It was theorised that PEB may be the periplasmic binding partner for transport of 3-PG via GlpT. The aims of this chapter sought to investigate GlpT as a 3-PG transporter as it was identified as the only homolog of 3-PG transporters in other bacterial systems and among other *C. jejuni* genomes. The two phenotypes often observed among GlpT transporters are 3-PG dependent growth and sensitivity to fosfomycin. Insertional knockout mutants of *peb3* and *glpT* from the well characterised *C. jejuni* strain 11168H were generated to prevent the expression of these genes by the host cell. This allowed for comparative testing among WT and *peb3* and *glpT* mutant strains using microbiological techniques, such as growth assays and fosfomycin sensitivity assays (Jeong et al., 2001, Min et al., 2009, Ribardo et al., 2010). The reintroduction of both genes into the mutant cell led to complementation and other phenotypes were observed. As *glpT*, annotated as pseudogene with three open reading frames, may encode a full length GlpT protein attempts were made to purify this protein using techniques such as cloning, expression, purification and western blot analysis. Mass spectrometry was also carried out to test this hypothesis.

After the initial bioinformatics analysis of this region in 11168H, it was hypothesised that the *peb3/glpT* locus would be highly conserved among *Campylobacter* species as this region has not be identified as an area of hypervariability (Pittinger et al., 2012). In **Chapter 5**, I focused on comparing the *peb3/glpT* locus in *C. jejuni* and closely related species *C. coli*, *C. lari* and *C. upsaliensis* genome sequences using genome browsers Artemis and WebACT. In 11168H a *glpT* gene composed of three ORFs due to two frameshifting events was observed. A detailed investigation was carried out to determine if other *glpT* genes experience frameshifting events, which may result in structural variation. Lastly, in order to infer evolutionary relationships among the different classes of the *peb3/glpT* locus and *glpT* structure both were placed in evolutionary context through MLST. In **Chapter 6**, a detailed summary of the key findings conveyed in this thesis is provided.

CHAPTER 2

Materials & Methods

2.1 Chemicals and Reagents

All chemicals and reagents used in the experimental procedures of this study were of analytical grade and purchased from, Bioline (London, UK), Bio-Rad (Hempstead, UK), Oxoid (Basingstoke, UK), Sigma-Aldrich (Poole, UK), bioMèrieux (Basingstoke, UK), Thermo Fisher Scientific (Loughborough, UK), Qiagen (Manchester, UK), Invitrogen (California, USA), NEB (Ipswich, MA, USA), SLS (Yorkshire, UK), Expedeon (Cambridge, UK) and GE Healthcare (Hertfordshire, UK).

2.2 Procedure and equipment

All media, glassware and general solutions were sterilised by autoclaving at 121 °C, 15 psi for 20 minutes. All 1.5 ml Eppendorf tubes and pipette tips for volumes ranging from 10 µl to 1 ml were sterilised at 121 °C for 15 minutes.

2.3 Media

All media was mixed to dissolve completely and sterilised by autoclaving at 121 °C, 15 psi for 20 minutes. After cooling below 50 °C antibiotic(s) were added where appropriate prior to inoculation with bacterial cells.

Luria Bertani (LB) broth (Oxoid): 10 g tryptone, 5 g yeast, 10 g NaCl per litre of dH_20.

Mueller Hinton (MH) Broth (Oxoid): 3 g beef, dehydrated infusion, 17.5 g casein hydrolysate 1.5 g starch per litre of dH_20.

SOC Medium: 20 g Bacto-tryptone, 5 g Bacto-yeast extract, 0.5 g NaCl, 2.5 ml of 1 M KCl per litre of dH_20. After autoclaving at 121 °C, at 15 psi for 20 minutes the broth was cooled below 50 °C and supplemented with 10 ml of 1 M $MgCl_2$, 10 ml of 1 M $MgSO_4$, 20 ml of 1 M glucose.

All agar media was mixed to dissolve completely and sterilised by autoclaving at 121 °C, 15 psi for 20 minutes. Once cooled to a temperature below 50 °C antibiotic(s) were added where appropriate and poured into standard 90 mm diameter plastic petri dishes (Sarstedt, Germany).

Columbia Blood Agar (CBA) (Oxoid): 23 g Special peptone, 1 g Starch, 5 g NaCl, 10 g Agar per litre of dH_20. Once cooled to a temperature below 50 °C defibrinated horse blood (TCS Biosciences, Buckingham, UK) at 5 % (v/v) is added.

Mueller Hinton (MH) Agar (Oxoid): 3 g beef, dehydrated infusion, 17.5 g casein hydrolysate, 1.5 g starch, 10 g agar per litre of dH_20.

Luria Bertani (LB) agar (Oxoid): 10 g tryptone, 5 g yeast, 10 g NaCl, 15 g biological agar per litre of dH_20.

MEMα, nucleosides, no phenol red (41061-029) (Thermo Fisher Scientific):

The complete formula comprised of 41 components consisting of amino acids, ribonucleotides, deoxyribonucleotides, vitamins, inorganic salts and sodium pyruvate. The complete formula can be found on the manufacturer's website.

MCLMAN (medium cysteine leucine methionine aspartic acid niacinamide) media:

All chemicals used in this media were purchased from Sigma. The media is comprised of four amino acids: 0.2 mM L-Cysteine HCl, 0.2 mM L-Methionine, 0.8 mM L-Leucine, and 10 mM L-Aspartic acid, seven inorganic salts: 1.8 mM $CaCl_2{\cdot}2H_2O$, 0.25 µM $Fe(NO_3)3{\cdot}9H_2O$, 1.75 mM $MgSO_4$, 5.4 mM KCl, 44 mM $NaHCO_3$, 0.1 M NaCl, 0.9 mM NaH_2PO_4 and one vitamin: 33 µM Niacinamide per litre of dH_20. The components were autoclaved where possible, components not autoclavable were filtered and added after the media reached room temperature. Once combined the media was filtered with a 0.22 µM filter and stored at room temperature for up to two weeks.

2.4 Culturing of bacterial strains

The *C. jejuni* strain (National culture Type Collection, London, UK) 11168H (Appendix Table S3) from a frozen stock was cultured on CBA plates with 5 % (v/v) horse blood and incubated in for 48 hours at 42 °C in a MACS VA500 workstation (Don Whitley, UK) under microaerobic conditions of 85 % N_2, 10 % CO_2 and 5 % O_2. Where appropriate antibiotic(s) kanamycin (50 µg/ml) and chloramphenicol (17 µg/ml) were added.

E. coli strains (Appendix Table S4) were cultured using LB agar plates or LB broth. Antibiotics (Sigma, UK) ampicillin and kanamycin were added to LB media at a final concentration of 100 and 50 µg/ml. Where required, IPTG and X-Gal (Bioline, UK) were added at final concentrations of 50 and 80 µg/ml, respectively.

2.5 PCR Amplification

Oligonucleotide primers were designed to target specific DNA sequences of interest (Appendix Table S1). PCR reaction tubes contained DNA (100 ng), PCR master mix (1X NH_4 buffer, 1.5 mM $MgCl_2$, 0.8 mM dNTPs, 2.5 U Taq polymerase) and primers (MWG, Germany) (20 pmol) at a final volume of 50 µl. The addition of 0.2 µl cloned Pfu DNA polymerase (Aligent) was added for increased fidelity. PCR reaction tubes were loaded in a Thermo Electron Corporation (USA) PxE 0.2 Thermal Cycler PCR machine and DNA was amplified using a defined programme, denaturing at 94 °C for 30 seconds, annealing at 5 °C below predicted melting temperature of primers for 30 seconds and 72 °C extension of 1 minute per kbp length of predicted PCR amplification product.

2.6 Gel Electrophoresis

Following amplification, PCR products were loaded onto a 1 % (v/v) agarose gel for electrophoretic separation of DNA. The gels were prepared by dissolving 4 g of SeaKem LE agarose powder (Lonza, UK) in 400 ml of TAE (Tris-Acetate EDTA) buffer (40 mM Tris,

20 mM acetic acid, 1 mM EDTA at pH 7.0) by heating. After cooling, 3 µl of ethidium bromide (stock-10 mg/ml, Sigma, UK) was added to the gel before setting. A molecular weight marker DNA hyperladder I (Bioline, UK), alongside samples mixed with 5x DNA loading dye (Bioline, UK), were loaded onto the gel submerged in TAE buffer in a Bio-rad sub-cell gel tank. Gel electrophoresis was performed at 100 V for 45 minutes. The gel was visualised with a Syngene GeneFlash Bio imager connected to a Computer H6Z0182 lens camera and pictures of the gel were printed with a Sony UP895MD printer.

2.7 DNA Manipulation

Following amplification, PCR products were ligated, according to manufacturer's instructions, into a pGEM-T easy vector (Promega, UK) with DNA ligase (Promega, UK) and buffer. Plasmids that were restricted were alkaline phosphatised by adding 1 µl of Antarctic Alkaline phosphatase to 9 µl buffer (New England Biolabs, UK) and incubating for 1 hour at 37 °C. Digested plasmids were purified using an Illustra MicroSpin S-300 HR Columns (GE Healthcare) according to the manufacturer's instructions and the concentration was determined by NanoDrop® ND-1000 (NanoDrop technologies Ltd.). The ligation reactions were incubated at 4 °C overnight with T4 DNA ligase (NEB) and 10X ligase buffer (NEB), transformed into chemically competent *E. coli* cells (Appendix Table S4) via heat shock according to the manufacturer's instructions and grown overnight on LB agar plates with appropriate antibiotics.

2.8 Site-Directed mutagenesis

A QuikChange II kit (Strategene) was used to mutate *peb3* (Chapter 3) and *glpT* (Chapter 4) genes according to manufacturer's protocol. Oligonucleotide primers with specific mutations were designed using the Agilent QuikChange Primer design software tool and were synthesised by Eurofins MWG Operon (Ebersberg, Germany) (Appendix Table S1) and mutations were introduced by PCR thermal cycling (Table 2.1). Following PCR, the plasmid

was digested with 0.5 µl DpnI restriction enzyme (NEB Labs, USA) incubated at 37 °C for 1 hour and inactivated at 80 °C for 20 minutes. Site-directed plasmids (Appendix Table S2) were transformed into chemically competent *E. coli* cells.

Segment	Cycles	Temperature	Time
1	1	95 °C	30 s
2	15	95 ºC	30 s
		55 ºC	1 m
		68 ºC	1m kb^{-1}

Table 2.1 Site-Directed mutagenesis PCR cycling parameters. The table lists the temperatures and times used to generate mutations in the heparin-binding sites of *peb3* and to introduce a BamHI site into the *glpT* gene, via PCR.

2.9 *E. coli* transformations

Chemically competent *E. coli* cells DH5α and BL21-AI were generated via the $RbCl_2$ method. The *E. coli* cells grow on LB agar with appropriate antibiotic(s) and incubated overnight at 37 °C. A single fresh *E. coli* colony was used to inoculate 5 ml of LB broth and incubated for 18 hours at 37 ºC with shaking 200 revolutions per minute (rpm). Overnight culture (1 ml) was used to inoculate new flask containing 100 ml of LB broth. The cultures were incubated at 37 °C with shaking, grown to an optical density (OD_{600}) of 0.4 and immediately incubated on ice for 10 minutes. Cultures were placed into chilled 50 ml falcon tubes and centrifuged at 3,500 rpm for 15 minutes, at 4 °C. The cell pellet was re-suspended in 50 ml (25 ml/falcon tube) of cold TfbI buffer (30 mM KOAc, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 $MgCl_2$ in dH_2O with 15 % glycerol (v/v), pH, 5.8 with 0.2 M glacial acetic acid and filter sterilised). The mixture was incubated on ice for 10 minutes and centrifuged at 3,500 rpm at 4 °C for 15 minutes. The cell pellet was re-suspended in 2 ml of TfbII buffer (10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$ in dH_2O followed by the addition of 15 % glycerol (v/v) pH, 6.5 with 1 M KOH, the solution was filter sterilised and incubated on ice for 2 hours. The final

suspension was aliquoted into 1.5 ml Eppendorf tubes at 100 µl and the tubes were snap frozen in liquid N_2 and stored at -80 °C. Plasmid at 1 µl (~25 ng) was added to *E. coli* cells and transformed via heat shocking according to manufacturer's protocols. The suspension was spread onto antibiotic selective LB plates and incubated overnight at 37 °C. A single colony from the plate was inoculated into LB broth with appropriate antibiotics and incubated at overnight 37 °C with shaking. The culture was centrifuged at 3,500 rpm at 4 °C for ten minutes and isolation of plasmids from *E. coli* was carried out with a QIAprep Spin MiniPrep Kit (Qiagen, West Sussex, UK) following the manufacturer's recommendations.

2.10 DNA Sequencing

A BigDye® Terminator v1.1 Cycle sequencing kit according to manufacturer's recommendations (Applied Biosystems) was used to carry out DNA sequencing. Sequencing reactions (10 µl) were subjected to the Applied Biosystems Prism 3100 genetic analyser located in the University of Manchester DNA sequencing facility. DNA sequence data was analysed with Chromas software (Technelysium Pty Ltd. Australia) and sequences were verified using Clustal W, a feature of Biology Workbench (San Diego Supercomputer Centre, USA).

2.11 Acid-glycine extraction of surface-proteins from *C. jejuni*

Acid-glycine extraction of *C. jejuni* NCTC 11168H surface proteins was performed according to the method by Logan and Trust, 1983. *C. jejuni* cells from five CBA plates were harvested in dH_2O, washed twice at 8000 *x g* for 15 minutes and re-suspended in 0.2 M glycine hydrochloride pH 2.2 at 33 ml/1 g of wet cell weight. The suspension was stirred for 15 minutes at room temperature and centrifuged at 12,000 *x g* for 15 minutes prior to neutralisation with sodium hydroxide (NaOH). The supernatant was dialysed at 4 °C against dH_20 with Snakeskin™ Dialysis tubing (Thermo Scientific, UK) according to manufacturer's

recommendations. The final suspension was frozen at -80 °C, lyophilised and then stored in -80 °C for further use or used immediately for heparin-affinity chromatography.

2.12 Heparin-affinity chromatography

Proteins were concentrated using the Bio-Rad protein assay dye reagent protocol according to the manufacturer's instructions. Approximately 0.300 µg of concentrated protein was added to binding buffer A (5 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3). Prior to use a 1 ml HiTrap Heparin HP column (GE Healthcare) attached to an ÄKTA 900 prime system (GE Healthcare) was equilibrated with binding buffers A and elution buffer B (1 M NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3). Once equilibrated the sample was loaded onto the heparin column and washed for 10 column volumes. Proteins were eluted with a 0-1 M linear NaCl gradient using elution buffer B. The protocol was performed in triplicate and error bars based on standard error.

2.13 GST-purification of GST-PEB3 fusion protein from *E. coli*

An overnight culture of *E. coli* BL21-AI cells encoding GST-PEB3 fusion proteins inoculated into 25 ml of LB broth containing antibiotics and incubated at 37 °C with shaking at 100 rpm. The culture was used to inoculate 400 ml of LB broth containing antibiotics incubated at 37 °C with shaking until an OD_{600} of 0.6-1.0 was reached. Once the desired OD_{600} was reached protein production was induced with 1 mM IPTG and 0.1 % arabinose. Induced cells were grown for 4 hours and harvested by centrifugation for 20 minutes at 4,000 *x g* at 4 °C. The cell pellet was re-suspended in cold GST-binding buffer and lysed by two rounds of French pressing at 13,000 psi. The protein extract was clarified by another round of centrifugation 6000 *x g* at 4 °C for 20 minutes and filter sterilised. A 2 ml aliquot of filtered supernatant was applied to a Glutathione Sepharose 4B column (GE Healthcare) and incubated for 30 minutes at room temperature. The GST column was washed with 20 ml of GST-binding buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4,

pH 7.3) to remove unbound protein. The column was incubated at 25 °C for 16 hours with 500 µl of GST-binding buffer containing 20 µl of thrombin (Calbiochem) to release bound protein. The column was washed with 1 ml of GST-binding buffer, an additional three times, to elute PEB3 protein. The concentration of the purified proteins was estimated according to manufacturer's instructions of the Bio-Rad protein assay dye reagent.

2.14 Whole cell lysate preparation

Whole cell protein extracts were prepared by resuspending *C. jejuni* cells in PBS to an OD_{600} of 2.5-3.0. A total of 500 µl of the suspension was centrifuged at 10,000 rpm in a benchtop centrifuge resulting in a pellet and supernatant. The pellet was re-suspended in 50 µl of 1x BugBuster (Novagen) followed by 0.5 µl of DNase I (Thermo Scientific, UK) and incubated at room temperature for 30 minutes. Following incubation, lysates were centrifuged at 10,000 rpm in a benchtop centrifuge and supernatants re-suspended in 1 X SDS-PAGE buffer. For *E. coli*, whole cells were obtained by growing 1 ml of culture to an OD_{600} 2.0. The culture was centrifuged and the remaining pellet resuspending in 50 µl of 1x SDS-PAGE loading buffer.

2.15 Bacterial cell membrane preparation

Membrane preparations of *C. jejuni* cells were prepared by harvesting cells grown on CBA into 50 mM Tris-HCl (pH 7.0), 25 mM NaCl. The cell suspension is centrifuged at 8000 x *g* and pellet retained and weighed. The pellet is resuspended in 50 mM Tris-HCl (pH 7.0), 25 mM NaCl and lysed by a French press cell at 13,000 psi. The unlysed cells were removed by two rounds of centrifugation at 15,000 rpm in a bench top centrifuge for 20 minutes at 4 °C. The supernatant was ultracentrifuged at 100, 000 x *g* (70,000 rpm) in a Beckman OptiMAX for 1 hour at 4 °C. The total membrane fraction pellet was weighed and re-suspended in 50 mM Tris-HCl (pH 7.0), 25 mM NaCl, 1 % Triton X-100, per 1 ml per gram of starting wet cell pellet, and then incubated at room temperature for 1 hour. Solubilised membrane

samples were ultracentrifuged at 100, 000 x *g* (70,000 rpm) for 1 h at 4 °C. The supernatant was resuspended in Triton X-100 and purified via His and HA-Tag purification.

2.16 Purification of GlpT from membrane fractions

Solubilised membrane fractions containing His-Tagged GlpT were purified in single reaction tubes with Ni-NTA Magnetic agarose beads (Qiagen) according to manufactures' instructions with some modifications. Membrane fraction (500 µl) was added to 100 µl of Ni-NTA Magnetic agarose beads, mixed for 1 hour at room temperature and separated magnetically for 1 minute. The magnetic beads were re-suspended in 500 µl of Ni-NTA washing buffer (20 mM sodium phosphate, 500 mM NaCl, 1 % Triton X-100, 40 mM imidazole, pH 7.4), mixed and magnetically separated to removal the washing buffer. This was step was repeated 2 times and protein was eluted in 50-100 µl of Ni-NTA elution buffer (20 mM sodium phosphate, 500 mM NaCl, 1 % Triton X-100, 500 mM imidazole, pH 7.4) and then 1X SDS-PAGE buffer.

Manual immunoprecipitation of GlpT HA-Tagged strains was performed with Pierce Anti-HA magnetic beads (Thermo, UK) per the manufacturer's instruction. The HA-Tagged proteins were eluted into 50 µl of 5X SDS-PAGE sample buffer.

2.17 SDS-PAGE and Immunoblotting

Heparin chromatography fractions were denatured in 2X SDS-PAGE loading buffer (100 mM Tris-HCl (pH 6.8), 2 % SDS, 10 % glycerol, 0.1 % bromophenol blue, 5 % β-mercaptoethanol) by boiling for 5 minutes at 95 °C. A Precision Plus broad range protein standard marker (Bio-Rad Laboratories Ltd.) and samples were separated on a gel a 12% gel. The gel was prepared with a lower 12 % (v/v) resolving gel [3.5 ml dH_2O, 2.5 ml 10 % SDS, 2.5 ml 1.5 M Tris-HCl (pH 8.8), 100 µl 10 % APS, 4 ml Acrylamide/Bis (30 % stock), 20 µl TEMED] and an upper 4 % (v/v) stack [6.1 ml dH_2O, 100 µl 10 % SDS, 2.5 ml 1.5 M Tris-

HCl (pH 8.8), 100 µl 10 % APS, 1.3 ml Acrylamide/Bis (30 % stock), 20 µl TEMED]. Gel electrophoresis was performed in an X-Cell SureLock™ Mini-Cell (Invitrogen) electrophoresis system at 150 V for 1 hour. SDS-PAGE gels were stained with Instant Blue or transferred onto a nitrocellulose membrane for western blotting. For western blotting, proteins were transferred onto a Protran nitrocellulose membrane (Whatman, UK) using an X-Cell SureLock™ Mini-Cell blot module (Invitrogen) at 30 V for 1 hour in transfer buffer (2.4 g Tris, 11.26 g glycine, 20 % (v/v) methanol per litre). Membranes were blocked overnight in PBS (phosphate buffered saline) (Sigma-Aldrich Ltd, Poole UK) containing 10 % (w/v) milk. To detect PEB3 proteins the membrane was incubated with anti-PEB3 antiserum (London School of Hygiene and Tropical Medicine) diluted to 1:5,000 in PBS with 10 % milk for 1 hour at room temperature. The membrane was washed 3 times in PBST (PBS with 0.2 % Tween 20). The secondary antibody, an anti-rabbit IgG (LI-COR, Lincoln, USA), at a dilution of 1: 2,000 in PBS with 10 % milk was incubated with the membrane for 1 hour at room temperature. Following three washes with PBST and one final wash with PBS membranes were visualised on the LI-COR Odyssey® Imaging System at 700-800 nm.

In Chapter 4, whole cell lysates, purified GlpT His-tagged and HA-tagged proteins were denatured in 1X, 1X and 5X SDS-PAGE loading buffer, respectively, by boiling at 95 °C for 10 minutes. Samples (20 µl), briefly centrifuged in a desktop centrifuged, along with a Precision Plus broad range protein standard were loaded onto a NuPAGE® Novex® 4-12 % Bis-Tris Protein Gels (Thermo Scientific, UK) in MOPS buffer (104.6 g MOPS, 60.6 g Tris Base, 10 g SDS, 3 g EDTA). Gel electrophoresis was performed in an X-Cell SureLock™ Mini-Cell (Invitrogen) electrophoresis system at 175 V for 50 minutes. The gel was incubated with Instant Blue (Expedeon) stain overnight or transferred to nitrocellulose membrane. For western blotting, proteins were transferred onto a Protran nitrocellulose membrane (Whatman, Maidstone, UK) using an X-Cell SureLock™ Mini-Cell blot module (Invitrogen) at 30 V for 1 hour in transfer buffer (2.4 g Tris, 11.26 g glycine, 20 % (v/v)

methanol per litre). Membranes were blocked overnight in PBS (phosphate buffered saline) (Sigma-Aldrich Ltd, Poole UK) containing 10 % (w/v) milk. To detect His or HA-Tagged GlpT proteins the membranes were incubated for 1 hour at room temperature with His-antibody (Qiagen) or HA- antibody (Sigma) diluted to 1:1000 in 0.01 % PBS-T with 3 % (w/v) milk. The membrane was washed three times in PBS-T 0.01 %. The secondary antibody, an anti-mouse IgG (LI-COR, Lincoln, USA), diluted to 1: 10,000 in 0.01 % PBS-T with 3 % milk was incubated with the membrane for 1 hour at room temperature. Following three washes with PBST and one final wash with PBS the membranes were visualised on the LI-COR Odyssey® Imaging System at 700-800 nm.

2.18 Protein mass spectrometry

Proteins were separated on a 12 % SDS-PAGE gel and stained with Instant Blue (Expedeon, Cambridge, UK). The gel was placed in a sterile petri dish the band was excised and digested with trypsin overnight by the staff of the Protein Mass Spectrometry Facility at the University of Manchester. Protein identifications from the mass spectrometry were input into Mascot software (version 2.2.06; produced by Matrix Science). The sequence of the identified proteins of interest were found in the UniPROT database (http://www.uniprot.org) (Magrane & Consortium, 2011). The results were also fed into statistical validation software called Scaffold (version 3.0.04; produced by Proteome Software).

2.19 Conserved Domain Database (CDD)

The domain architecture of each *C. jejuni* heparin-binding protein sequence was determined with CDD (Marchler-Bauer et al., 2011). Information regarding putative protein function, specific amino acids or regions involved in binding, other proteins with similar domain structures and a 3-D view of proteins with solved crystal structures was also provided.

2.20 ExPASy pI/M_w Tool

The theoretical pI and M_w of *C. jejuni* heparin-binding proteins was determined by the ExPASy compute pI/M_w (http://www.expasy.org/cgi-bin/pi_tool) (Gasteiger et al., 2005).

2.21 PyMOL: Molecular modelling of PEB3 heparin-binding sites

The structural coordinates for the PEB3 dimer were taken from the Protein Data Bank (PDB) archive (http://www.rcsb.org/pdb) file 2HWX. The heparin-binding sites were mapped to chain A of the PEB3 dimer. Additional changes to the model include deletion of water molecules. Protein structural coordinates of potential heparin-binding sites and docking figures were prepared using The PyMOL Molecular Graphics System (DeLano Scientific LLC, San Carlos, CA, USA).

2.22 ClusPro: Identification of PEB3 heparin-binding Sites

The molecular docking was performed with the ClusPro programme (Mottarella et al., 2014). For this process the PEB3 monomer and heparin were chosen as receptor and ligand, respectively, by selecting the heparin mode of the programme (http://cluspro.bu.edu/). Additionally, the PEB3 dimer and heparin followed the same process, however heparin-binding sites: $_{62}$KAKKD$_{66}$ and $_{122}$NKKVRI$_{127}$ sequences were designated as sites of attraction. The programme ranks and provides the top 15 docking models, these were analysed to identify interactions among regions of the PEB3 in complex with heparin (Mottarella et al., 2014). The PEB3-heparin complexes obtained in this study were visualised and analysed with Pymol Molecular Graphics software 1.7.0.1.

2.23 *In silico* MLST analysis of 64 *C. jejuni* strains

The sequence for each MLST genes: *aspA*, *glnA*, *gltA*, *glyA*, *pgm*, *tkt* and *uncA* was obtained for 64 *C. jejuni* strains. These genes were identified in contig assemblies via homology searching with the UniPROT and the NCBI (http://www.ncbi.nlm.nih.gov/genome).

Comparison of these sequences against allelic sequences in the *C. jejuni* PubMLST database (http://pubmlst.org/campylobacter) and previous published allelic profiles were used to determine sequence type (ST) and clonal complex (ST-CC) for each strain.

2.24 MLST Phylogenetic analyses

The relatedness of the allelic profiles of 64 *C. jejuni* strains (Appendix Table S7) was investigated via phylogentic analysis. In order to infer evolutionary relationships among these 64 strains a dendrogram was constructed with START2 software (package, version 0.5.13) (Jolley et al., 2001). This software is available at the PubMLST *C. jejuni* database (http://www.pubmlst.org).

2.25 Genomic Analysis of the *peb3/glpT* locus

Homology searches of the *glpT* and *peb3* genes were conducted at the amino acid sequence level using a genomic BLAST and UniPROT. The Genome function via the NCBI Genome tool was used to identify the accession number of the contigs, which possessed this gene region (Appendix Table S5 & S6). These numbers were then input into genome browsers Artemis (Rutherford et al., 2000) and webACT (Carver et al., 2005).

2.26 Generation of *C. jejuni* competent cells and electroporation

C. jejuni cells were grown for 24 hours on CBA, harvested into 1 ml of MH broth and pelleted by centrifugation 14,000 rpm for 5 minutes at 4 °C. The cells were re-suspended in 1 ml of ice-cold sucrose buffer (272 mM sucrose and 15 % (v/v) glycerol). After three rounds of centrifugation and washing in sucrose buffer the electrocompetent cells were fractioned in 50 µl to 100 µl increments. These could be used immediately or stored for later use at -80 °C. Plasmid constructs (~10 ng) (Appendix Table S2) were added to 50 µl of competent *C. jejuni* cells. The mixture was transferred into an ice-cold electroporation cuvette (Gene Pulser®/MicroPulser™ Electroporation cuvette, 0.2 cm). Electroporation was executed

using a Bio-rad Gene Pulser II electroporation system set at 2.5 kV, 25 µF and 200 Ω. The cuvette was flushed twice with 100 µl SOC medium, spread onto a non-selective CBA plate and incubated overnight at 42 °C under microaerophilic conditions. After overnight growth cells were cultured onto selective agar and incubated at 42 °C for a further 3-5 days. Individual colonies that formed were subcultured, amplified via colony PCR to verify the genotype and then stored at -80 °C in glycerol (30 % glycerol (v/v) in MH broth).

2.27 Fosfomycin susceptibility assays

1) E-Test

For the E-Test, *C. jejuni* cells were cultured in 30 ml of MH broth without antibiotics to a starting OD_{600} of 0.05. Cultures were grown until mid-log phase (OD_{600} 0.2-0.4) was reached in microaerophilic conditions with shaking at 125 rpm (Figure 4.2 & 4.9). A cell suspension of 0.2 ml was plated on to a dry MH agar plate using a spreader. The plate was allowed to dry for 5-10 minutes and one E-Test (bioMèrieux, UK) strip at room temperature was applied. The plate was incubated face side down at 42 °C for 48 hours. Additional guidance on storage, procedures and interpretations of results strictly followed the E-Test Antimicrobial susceptibility guide (IVD-9302553C-en2012/01). The MICs were read directly from the test strip according to the manufacturer's instructions. Determination of the MIC is where the elliptical zone of inhibition intersected with the MIC scale on the E-Test strip. The compound glucose-6-phosphate was not added to the MH agar due to the lack of the UhpT transporter.

2) Spot-plate Method

The spot plate method was followed as described by Jeong et al. and Ribardo et al. Prior to the agar setting, a range of fosfomycin (Sigma, UK) concentrations, 0, 8, 24, 32, 48, 64, 96, 128 and 192 µg/ml, were added to molten MH agar. The agar was pipetted at a volume of 25 ml into petri dishes. The *C. jejuni* cells were harvested and re-suspended in MH broth to

an OD_{600} of 1.0. The cells were sequentially diluted ten-fold seven times and 10 µl of each dilution was spotted onto MH fosfomycin containing agar. The plates were allowed to dry for 10 minutes and then incubated facedown for 48 hours at 42 °C under microaerobic conditions. The cell viability was calculated via the Miles and Misra technique. The MIC was determined by lowest fosfomycin concentration resulting in a 10-fold decrease in CFU when compared to growth on MH agar without fosfomycin.

2.28 Growth Assays

1) MH Broth

Comparative growth curves were performed in MH broth. In these experiments, 100 µl of *C. jejuni* cells were inoculated into 30 ml MH broth. Flasks were incubated under microaerobic conditions with constant shaking (100 rpm) at 42 °C. After an initial growth period the OD_{600} was measured and a new flask inoculated to an OD_{600} of 0.05 growth was monitored.

2) MEMα

MEMα growth assays were performed as described in by Del Rocio Leon-Kempis et al. and Min et al. Liquid cultures of *C. jejuni* were grown microaerobically at 140 rpm in MH broth without antibiotics to mid-log phase (OD_{600} 0.2-0.4) or 7-10 hours after inoculation (Figure 4.2). Mid-phase cultures were used to inoculate in triplicate, 1 ml samples of MEM-α (Thermo, UK) supplemented with 0 or 5 mM 3-PG (Sigma, UK) and 0.1 mM $FeSO_4$ (Sigma, UK). After growth at 42 °C for 24 hours, the OD_{600} of the cultures was measured.

3) MCLMAN

Growth assays with MCLMAN media followed the protocol of Alazzam et al. with some slight modifications. Bacterial cells were transferred to MH broth and incubated under microaerobic conditions with shaking at 140 rpm. A second culture at an initial OD_{600} of

0.05 was inoculated and incubated at 42 °C under the same conditions. After 7-10 hours, mid-log phase cells (OD_{600} 0.2-0.4) (Figure 4.2 & 4.9) were collected via centrifugation at 10,000 rpm for 20 minutes at 4 °C. The pellet was washed and re-suspended in warm MCLMAN media to an OD_{600} 0.500. The cell suspension was used to inoculate 12 well containing warm MCLMAN media, supplemented with 10 mM pyruvate (Sigma, UK), lactate (Sigma, UK), 3-PG (Sigma, UK) or G3P (Sigma, UK), to an initial OD_{600} of 0.05. The plates were incubated microaerobically at 42 °C for 24 hours with shaking (140 rpm) after, which the OD_{600} was measured. The data was analysed with GraphPad QuickCalcs using the Student's paired t-test ($p < 0.05$).

CHAPTER 3

Identification of *C. jejuni* heparin-binding proteins

3.1 Introduction

The pathogenic bacterium *C. jejuni* is the leading cause of diarrhoeal illness worldwide (Acheson & Allos, 2001). Our current knowledge of the biological systems aiding *C. jejuni* has increased, however, there remains much to be discovered. One area in need of more research is the interaction of this bacterium with glycosaminoglycans (GAGs). GAGs are carbohydrate-bearing structures, present in or on the host cell-surface that act as space-filling molecules for the stabilisation of the extracellular matrix (ECM) (Gandhi & Mancera, 2008). The family of GAG molecules also assists with a multitude of different cell activities - migration, growth, adhesion to the ECM and signalling– as well as binding growth factors, chemokines, proteases and proteins (Jackson et al., 1991; Linhardt & Toida, 2004; Gandhi & Mancera, 2008).

Bacteria produce cell-surface proteins called 'adhesins'. During the first stages of infection, adhesins use GAGs as sites of attachment (Rostand & Esko, 1997). GAGs mediate interactions with pathogens, but can reverse this outcome, by acting as antibacterial molecules (Kamhi et al., 2013). Certain GAGs, such as heparin, are absorbed by adhesins forming a polyanion barrier. This barrier is often referred to as a 'pseudocapsule', due to its clever mimicking of a polysaccharide capsule (Chen et al., 1995; Serruto et al., 2011). These 'pseudocapsules' serve as camouflage and aid in the evasion of host defences (Chen et al., 1995; Serruto et al., 2011). Heparin, the most researched of all GAGs, exhibits several different behaviours. These include forming molecular 'bridges', which link the GAG-binding sites of host molecules and bacterial adhesins (Duensing & Putten, 1998; Menozzi et al., 2002; Blom et al., 2008). It also binds, inhibits and augments molecules, which control the activation of complement pathways (Edens et al., 1993). Most notably, heparin serves as a good prototype for the initial investigations into GAG-protein interactions. This has resulted in the identification of several heparin-binding proteins (Cardin & Weintraub 1989; Menozzi et al., 2002).

Bacterial species with heparin-binding molecules include *Helicobacter pylori* (Asencio et al., 1993; Hirmo et al., 1995; Utt & Wadström, 1997; Ruiz-Bustos et al., 2001; Utt et al., 2001; Dubreuil et al., 2002; Dubreuil et al., 2004; Guzman-Murillo et al., 2008), *Neisseria gonorrhoeae* (Chen et al., 1995), *Neisseria meningitidis* (Serruto et al., 2010), *Legionella pneumophila* (Duncan et al., 2011) and *Bordetella pertussis* (Menozzi et al., 1991; Hannah et al., 1994; Geuijen et al., 1996; Geuijen et al., 1998).

C. jejuni recognises a range of host-cell glycans, namely glycoproteins, mucins and Lewis B-like structures (Hugdahl et al., 1988; McAuley et al., 2007; Day et al., 2009). The *C. jejuni* NCTC 11168-GS and 11168-O strains bind galactose, fucose, mannose and Neu5Ac-containing glycans (Day et al., 2009). However, this is dependent on specific temperature (25 °C, 37 °C or 42 °C) and growth conditions (with or limited oxygen) (Day et al., 2009). Additionally, several *C. jejuni* strains bound heparin and other GAG molecules such as hyaluronan, chondroitin and dermatan (Day et al., 2013). Nevertheless, specific *C. jejuni* GAG-binding proteins are not identified in these studies.

Several adhesins have been identified for *C. jejuni* among these are the PEB proteins. The PEB proteins 1 to 4 were first described as 'cell-binding factors' (Pei et al., 1991). PEB1 is periplasmic, binding protein involved in adherence, transport and invasion of host cells (Pei et al., 1998; Del Rocio Leon-Kempis et al., 2006). Like PEB1, a multi-functional role is suspected for PEB3, an adhesin that binds phosphate molecules (Min et al., 2009). The PEB4 protein, an antigenic virulence factor, has duties in adhesion, protein export, colonisation and invasion (Kervella et al., 1993; Asakura et al., 2007). The protein PEB2, a major antigenic peptide of *C. jejuni,* is a protein of unknown function (Pei et al., 1991). Other adhesins include JlpA, a surface-exposed lipoprotein, aids the attachment of *C. jejuni* to host cells and modulates inflammatory responses (Jin et al., 2001; Jin et al., 2003). The CadF outer membrane protein binds fibronectin, a component of the extracellular matrix (Konkel et al., 1997). CadF stimulates signal transduction pathways and is vital for invasion of host epithelial

cells (Monteville et al., 2003; Konkel et al., 2005). In summary, *C. jejuni* expresses many adhesins with active roles in attachment, transport and invasion; yet, their interactions with GAGs are not documented.

This study aimed to identify *C. jejuni* proteins that interact with the GAG prototype heparin. The specific approach involved isolation of *C. jejuni* surface proteins and investigation of these proteins interaction with heparin by affinity purification and mass spectrometry-based identification.

3.2 Identification and purification of *C. jejuni* heparin-binding proteins

Acid-glycine extraction, outlined by Logan & Trust, enables the isolation of *C. jejuni* cell-surface proteins. This method also isolates flagella, adhesins, PEB antigens and *N*-linked glycosylated proteins (Logan & Trust, 1983; Pei et al., 1991; Scott & Cordwell, 2009).

Applying the same procedure, cell-surface proteins were isolated (1 g in wet weight) from the strain *C. jejuni* NCTC 11168H. The acid-glycine extract was prepared by harvesting cells from two-day blood agar plates, washing the cells twice with distilled water and isolating surface proteins with acid-glycine (see Chapter 2). The resulting extract was lyophilised, equilibrated and applied to a heparin-agarose column. The bound protein was eluted with 0-1 M NaCl gradient via the AKTA purification system (see Chapter 2). The eluted fractions displayed two distinct peaks at 190 (Peak 1) and 335 mM NaCl (Peak 2) (Figure 3.1 A). An SDS-PAGE gel of the peak fractions revealed a single band migrating at ~25 kDa from Peak 1 and four bands migrating at ~28-31 kDa from Peak 2 (Figure 3.1 B).

The bands were excised, digested with trypsin and peptides analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The sequences of individual peptides were derived and protein identifications made using Mascot (version 2.2.06, Matrix Science) (Emma Keevill, FLS Protein Mass Spectrometry Core Facility). The results were imported into Scaffold (version 3.0.04; produced by Proteome Software) and a report generated. Based on peptide scores and protein coverage, five proteins with predicted molecular masses between 25 and 31 kDa were identified. These proteins are PEB1 (Cj0921c), tungstate binding protein, TupA (Cj1540), ATP/GTP-binding protein (Cj1041c), PEB4 (Cj0596) and PEB3 (Cj0289c) (Table 3.1).

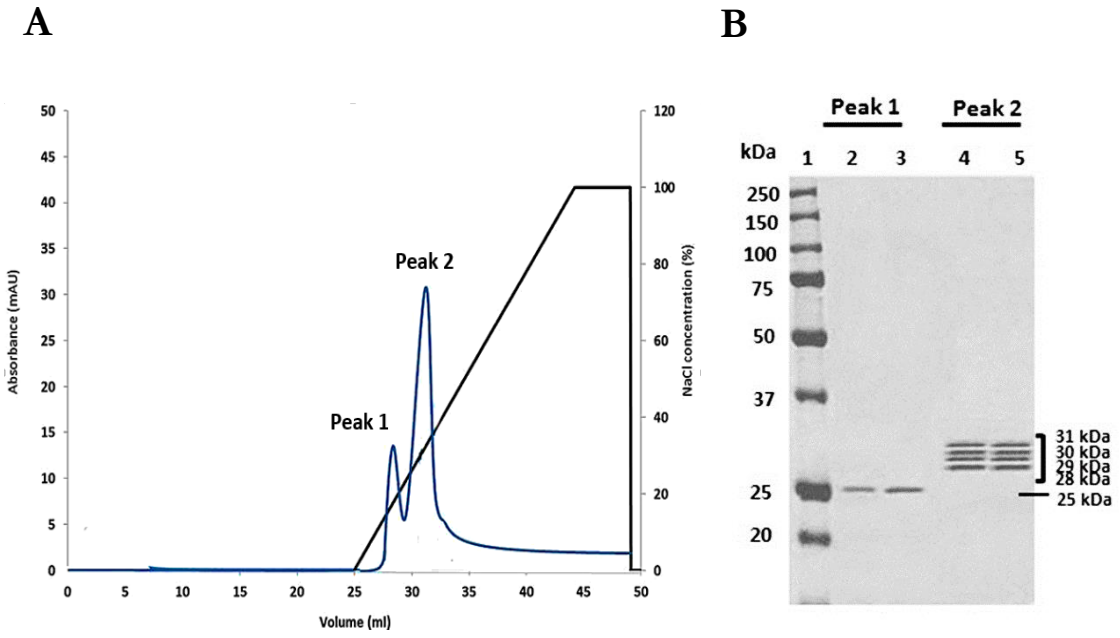


Figure 3.1 Heparin-binding components of a *C. jejuni* NCTC 11168H acid-glycine extract. A) *C. jejuni* proteins bind a heparin-agarose column and are eluted in two peaks. An acid-glycine extract from *C. jejuni* NCTC 11168H was deposited onto a heparin-agarose affinity column. Following extensive washing a 0-1 M NaCl gradient was applied. Two prominent peaks were observed at 190 mM (Peak 1) and 335 mM (Peak 2) NaCl. **B) SDS-PAGE analysis of proteins from a *C. jejuni* acid-glycine extract eluted from a heparin-agarose column.** Proteins were separated by SDS-PAGE and visualised by Coomassie staining. A protein marker was loaded in lane 1. Elution fractions from Peak 1 (lanes 2-3) showed one band migrating at 25 kDa (190 mM NaCl). Elution fractions from Peak 2 (lane 4-5) revealed four bands migrating at 28, 29, 30 and 31 kDa (335 mM NaCl).

Protein ID	Peak	Predicted size (kDa)	Peptide matches	pI	Function	References
PEB1 Cj0921c	1 (190 mM)	25	4	8.4	Adhesin; Amino Acid Transporter	Fauchère et al., 1989 Pei et al., 1991 Kervella et al., 1993 Pei et al., 1993 Pei et al., 1998 Del Rocio Leon-Kempis et al., 2006 Müller et al., 2007
PEB3 Cj0289c	2 (335 mM)	28	9	9.4	Major Antigenic Peptide	Fauchère et al., 1989 Pei et al., 1991 Linton et al., 2002 Rangarajan et al., 2007 Min et al., 2009
Tungstate-binding protein (TupA) Cj1540	2 (335 mM)	29	10	9.2	Tungstate-binding protein	Smart et al., 2009
ATP/GTP -binding protein Cj1041c	2 (335 mM)	31	5	6.4	ATP/GTP-binding Protein	Parkhill et al., 2000
PEB4 Cj0596	2 (335 mM)	30	4	9.9	Major Antigenic Peptide	Fauchère et al., 1989 Pei et al., 1991 Kervella et al., 1993 Burucoa et al.,1995 Asakura et al., 2007 Rathbun & Thompson, 2009 Rathbun et al., 2009 Kale et al., 2011

Table 3.1 Mass spectrometry identification and key features of the *C. jejuni* 11168H heparin-binding proteins. The key features of five heparin-binding proteins identified by mass spectrometry using Mascot (version 2.2.06; produced by Matrix Science). Additional information such as pI was obtained via ExPasy Compute pI/MW (Gasteiger et al., 2005) and references listed.

3.3 Features of *C. jejuni* heparin-binding proteins

The *C. jejuni* heparin-binding proteins were characterised from the annotated *C. jejuni* NCTC 11168H genome sequence (Parkhill et al., 2001; Magrane & Consortium, 2011). The key features included the predicted molecular mass, presence of a signal peptide, cellular location and potential heparin-binding sites. Two groups of heparin-binding proteins were discovered in this study. The first group of proteins comprises the proteins PEB1, PEB3 and PEB4 (Pei et al., 1991). These antigenic proteins have a high density of positive residues and high pI values, 8.4 for PEB1 and <9.3 for proteins PEB3 and PEB4 (Table 3.1) (Pei et al., 1991). The second group, includes a tungstate-binding protein (Cj1540) and an ATP/GTP-binding protein (Cj1041c) (Table 3.1). Both proteins have contrasting pI values, 9.2 for Cj1540 and 6.4 for Cj1041c. All five proteins are predicted to contain a signal peptide suggesting they may be exported (Figure 3.2 A-E).

3.3.1 PEB1 (Cj0921c)

From Peak 1, the PEB1 protein, a 25 kDa periplasmic adhesin was identified (Figure 3.1 & Table 3.1). The PEB1 protein binds epithelial cells and colonises the mucosa of the mouse model (Pei et al., 1998; Del Rocio Leon-Kempis et al., 2006). The PEB1 protein is homologous with periplasmic-binding proteins, comprising amino acid ABC (ATP-binding cassette) transport systems, and is located in an ABC transporter system operon (Pei & Blaser, 1993). As an adhesin, PEB1 binds dicarboxylic amino acids aspartate and glutamate (Del Rocio Leon-Kempis et al., 2006; Müller et al., 2007).

A more in-depth evaluation of the PEB1 protein amino acid content utilised the Conserved Domains Database (CDD) programme. The CDD programme uses multiple amino acid sequence alignments spanning a variety of organisms to classify and infer the functions of proteins (Marchler-Bauer et al., 2011). The CDD classified PEB1 as part of the Type 2 periplasmic binding protein superfamily (PBP2) (Marchler-Bauer et al., 2011). Proteins of

the PBP2 superfamily act as primary receptors, taking part in chemotaxis and uptake of nutrients from the extracellular space (Higgins, 2001; Marchler-Bauer et al., 2011). These proteins bind ligands with increased affinity and interact with a membrane transport complex (MTC). This interaction causes translocation of ligand to the cytoplasmic membrane (Marchler-Bauer et al., 2011).

The PEB1 protein is composed of 11 α-helices and 15 β-strands in two domains (Müller et al., 2007). At the amino acid level, a signal peptide sequence is present in the first twenty-six residues and is rich in basic residues (lysine (K) n=32, histidine (H) n=2 and arginine (R) n=5) (Figure 3.2 A) (Del Rocio Leon-Kempis et al., 2006). Additionally, two putative heparin-binding sites: $_{80}$DKKIKL$_{85}$ on a beta strand and $_{136}$EKKYKS$_{141}$ on a 3_{10} helix, were observed (Table 3.2). These sequences follow a consensus sequence pattern associated with heparin-binding proteins. This pattern is, XBBXBX, where B is a basic amino acid (K, R or H) and X is any other amino acid (Cardin & Weintraub, 1989; Hileman et al., 1998; Capila & Linhardt, 2002). A cluster of positive residues $_{115}$RKR$_{117}$ is present on an α-helix (Table 3.2 & Figure 3.2). A Kyte-Doolittle hydropathy plot provided by the Protein Data Bank (PDB) database (ID: 2V25, Müller et al., 2007) specifies the buried and surface-exposed regions of PEB1. The residues of $_{80}$DKKIKL$_{85}$ and $_{136}$EKKYKS$_{141}$ occupy buried and surface exposed areas of the protein, whereas, $_{115}$RKR$_{117}$ is surface exposed (data not shown). Additionally, 50 % of the PEB1 (pI 8.4) amino acid content consists of hydrophobic amino acids bearing no polar side chains (Pei & Blaser, 1993).

3.3.2 ATP/GTP-Binding protein (Cj1041c)

The Cj1041c gene encodes for a 31 kDa protein, designated as a periplasmic ATP/GTP-binding protein (Figure 3.1 B & Table 3.1). *In silico* analysis via the CD programme failed to categorise this protein (Marchler-Bauer et al., 2011). Nevertheless, examination of the Cj1041c amino acid sequence revealed a signal peptide in the first 20 residues (Figure 3.2 B)

and two potential heparin-binding sites: $_{106}$KTKKE$_{110}$ and $_{174}$RKKHK$_{178}$ (Table 3.2). The sequence, $_{174}$RKKHK$_{178}$, does not follow the consensus sequence pattern, XBBXBX (Cardin & Weintraub, 1989; Hileman et al., 1998; Capila & Linhardt, 2002). Also, $_{106}$KTKKE$_{110}$ does not conform to this pattern, and bears an alternative sequence pattern, BXBBX, also associated with heparin-binding (Fromm et al., 1997). Overall, the amino acid content of Cj1041c contains a low density of positive residues, K (n=33), H (n=3) and R (n=2), making up 14 % of its amino acid content (Figure 3.2 B). As well as a pI of 6.4, the lowest among all the *C. jejuni* heparin-binding proteins (Table 3.1).

3.3.3 PEB4 (Cj0596)

PEB4 (Cj0596), a 30.5 kDa protein, was identified in Peak 2 (Figure 3.1 B). PEB4 is a well-known virulence factor with duties in host-cell adhesion, protein export, colonisation and invasion (Kervella et al., 1993; Asakura et al., 2007). PEB4 also acts as a chaperone exporting proteins to the outer membrane (Rathbun & Thompson, 2009). The *peb4* gene encodes for a 273 protein product that is homologous to export factors, PrsA from *Bacillus subtilis* and PrtM from *Lactococcus lactis* (Burucoa et al., 1995) (Figure 3.2 C). Another PEB4 homologue is recognised in *E. coli* known as SurA. This protein has a cis-trans isomerase domain with duties in outer membrane protein biogenesis, pilus assembly, and *in vivo* persistence (Asakura et al., 2007).

The PEB4 protein is a dimer composed of 8 α-helices and 9 β-strands and is structurally similar to SurA-like chaperones (Kale et al., 2011). The CD programme identified a peptidyl-prolyl cis-trans isomerase domain in PEB4, which is responsible for regulating protein folding (Rathbun et al., 2009; Marchler-Bauer et al., 2011). The PEB4 protein sequence encodes for a signal peptide sequence in the first twenty-one amino acids (Figure 3.2 C). Surprisingly, no consensus sequence motifs or integrin recognition sites were observed in this sequence (Table 3.2). Rather PEB4 (pI of 9.9), has a high density of positive residues, K

(n=39), H (n=2) and R (n=3), accounting for 16 % of its amino acid content (Figure 3.2 C). Likewise, a large number (20 %) of amino acids namely arginine, asparagine, glutamine and tyrosine capable of hydrogen-bonding were identified (Figure 3.2 C). These hydrogen-bond forming amino acids are known to interact with heparin (Hileman et al., 1998).

3.3.4 Tungstate-binding protein: TupA (Cj1540)

A 29 kDa tungstate-binding protein (Cj1540) TupA was identified in Peak 2 (Figure 3.1 B). The TupA, forms part of a TupABC system (Cj1538-Cj1540) for the uptake and transport of tungstate (Smart et al., 2009). The TupABC system consists of: TupA a binding protein located in the periplasm or attached to a lipoprotein on the external face of the cytoplasmic membrane; TupB an integral membrane transport protein, and TupC, a ATPase (Makdessi et al., 2001). The first documented tungstate-specific ABC transporter was observed in the obligate anaerobe *Eubacterium acidaminophilum* (Makdessi et al., 2001). The *E. acidaminophilum* genome codes for the TupABC transporter system, which shares the greatest similarity with transport systems from archaea *Methanobacterium thermoautotrophicum* and *Haloferax volcanii* and gram-negative pathogens *V. cholerae*, and *C. jejuni* (Makdessi et al., 2001). Interestingly, the closest homologues for *E. acidaminophilum* TupA, are in *C. jejuni*, *V. cholerae*, *M. thermoautotrophicum*, and *H. volcanii* and form a separate phylogenetic group of oxyanion ABC transporters (Makdessi et al., 2001).

The *C. jejuni* 11168H TupA protein (Cj1540) is a periplasmic protein capable of binding tungstate and molybdate (Smart et al., 2009). The *cj1540* gene encodes for a 259 amino acid TupA protein (Figure 3.2 D). The CD programme indicates TupA is a member of the phosphonate-bd superfamily, a group of periplasmic proteins that uptake alkylphosphonate (Marchler-Bauer et al., 2011).

A) PEB1 (Cj0921c)

MVFRKSLLKLAVFALGACVAFSNANAAEGKLESIKSKGQLIVGVKNDVPHYALLDQATGE
IKGFEVDVAKLLAKSILGDDKKIKLVAVNAKTRGPLLDNGSVDAVIATFTITPERKRIYN
FSEPYYQDAIGLLVLKEKKYKSLADMKGANIGVAQAATTKKAIGEAAKKIGIDVKFSEFP
DYPSIKAALDAKRVDAFSVDKSILLGYVDDKSEILPDSFEPQSYGIVTKKDDPAFAKYVD
DFVKEHKNEIDALAKKWGL$_{259}$

B) ATP/GTP-binding protein (Cj1041c)

MKKYVLSLALLGSLLGASELKYQEFDGFKSPESIFVDKNYVYVSNVGEKLEPLAKDNDGF
ISKLDKNGKVLEYKFLTHLNAPKGMMEIGKTLYVVDIDVLRGFDLKTKKEIFNLPIKGAI
FLNDIEKLDDNTLLVSDTGTGLILKVDLKTKQYDELLKLDLAKFGGPNGLYLDRKKHKLF
IAGYHPDGVSGGVVMAYDLNTKELSIIKNEKESYDGIVPYKDGLLVSSWGNNLNGYIYNL
DNVKSVKLELPLMKGPADIFIEGNILWIPKMVEGKIFKVELNK$_{283}$

C) PEB4 (Cj0596)

MKKFSLVAATLIAGVVLNVNAATVATVNGKSISDTEVSEFFAPMLRGQDFKTLPDNQKKA
LIQQYIMQDLILQDAKKQNLEKDPLYTKELDRAKDAILVNVYQEKILNTIKIDAAKVKAF
YDQNKDKYVKPARVQAKHILVATEKEAKDIINELKGLKGKELDAKFSELAKEKSIDPGSK
NQGGELGWFDQSTMVKPFTDAAFALKNGTITTTPVKTNFGYHVILKENSQAKGQIKFDEV
KQGIENGLKFEEFKKVINQKGQDLLNSAKVEYK$_{273}$

D) Tungstate-binding protein (Cj1540)

MKKIISLALALALSASAAELKMATTTSTDNTGLLDALKPLYEKESGNTLKWVAVGTGAAL
KMGEDCNADVLFVHSPKAEKEFMKKGFGVDRTPVMYNDFIIIADKSLASKFKGKNLKESL
ELIKNEKLTFIS**RGD**KSGTDNKEKSLWKNLGGVPEKQSWYQQSGQGMLASIKIAEEKKGV
ILTDRGTYIKYEANEKGKPNLVIVNEGDDSLKNFYSVIATNPKHCKNVNYTEASKFIKWV
TSDKTLNFIADFKLLNKPLFVIDAKTRKD$_{269}$

E) PEB3 (Cj0298c)

MKKIITLFGACALAFSMANADVNLYGPGGPHTALKDIANKYSEKTGVKVNVNFGPQATWF
EKAKKDADILFGASDQSALAIASDFGKDFNVSKIKPLYFREAIILTQKGNPLKIKGLKDL
ANKKVRIVVPEGAGKSNTSGTGVWEDMIGRTQDIKTIQNFRNNIVAFVPNSGSARKLFAQ
DQADAWITWIDWSKSNPDIGTAVAIEKDLVVYRTFNVIAKEGASKETQDFIAYLSSKEAK
EIFKKYGWRE$_{250}$

Figure 3.2 Amino acid sequences of five heparin-binding proteins. The protein sequences were obtained from the UniProt database (Magrane & Consortium, 2011). Underlined residues represent signal peptides. Integrin peptide recognition sites are bolded and underlined. Positive residues are shaded grey and potential heparin-binding regions are boxed in black. Boxed potential heparin-binding sites are based on previously identified consensus sequence definitions and basic amino acids clusters: XBBXBX (Cardin & Weintraub, 1989) and XBBXB(Fromm et al., 1997).

Review of the TupA protein sequence identified a signal peptide in the first seventeen residues (Figure 3.2 D) (Smart et al., 2009). Interestingly, an arginine-glycine-aspartic acid (RGD) integrin recognition site was identified at position 133-135 (Table 3.2). These RGD sites are implicated in binding to integrins and syndecans on host cells (D'Souza et al., 1991). The RGD site in TupA resides on the protein surface as suggested by hydropathy plot analysis (data not shown). As well as the RGD site, two potential heparin-binding sites: $_{222}$PKHCKN$_{227}$ and $_{265}$KTRKD$_{269}$ were observed (Table 3.2). These sites follow the sequences XBBXBX and BXBBX, respectively (Cardin & Weintraub, 1989; Fromm et al., 1997; Hileman et al., 1998; Capila & Linhardt, 2002).

3.3.5 PEB3 (Cj0298)

Lastly, the 28 kDa band was confirmed to be PEB3. The PEB3 protein eluted from the heparin column at 335 mM NaCl (Figure 3.1-Peak 2). The PEB3 protein is a dimer, composed of two monomers, consisting of nine α-helices and nine β-strands superimposed onto each other (Rangarajan et al., 2007). Structurally, PEB3 resembles molybdate- and sulphate-binding proteins and is recognised as a member of the periplasmic-binding protein superfamily (PBPb) (Marchler-Bauer et al., 2011). The *peb3* gene encodes for a 252 amino acid product of which the first twenty amino acids code for a signal peptide (Figure 3.2 E). Furthermore, PEB3 (pI 9.3) contains two possible heparin-binding sites: $_{62}$KAKKD$_{66}$ and $_{122}$NKKVRI$_{127}$ (Table 3.2 & Figure 3.2 E).

3.4 Verification of PEB3 protein using specific antiserum

Proteins eluted from the heparin-agarose column in Peaks 1 and 2 were transferred onto nitrocellulose and incubated with anti-PEB3 antiserum. An immuno-reactive band of 28 kDa was detected in the Peak 2 fractions confirming the presence of PEB3 (Figure 3.3).

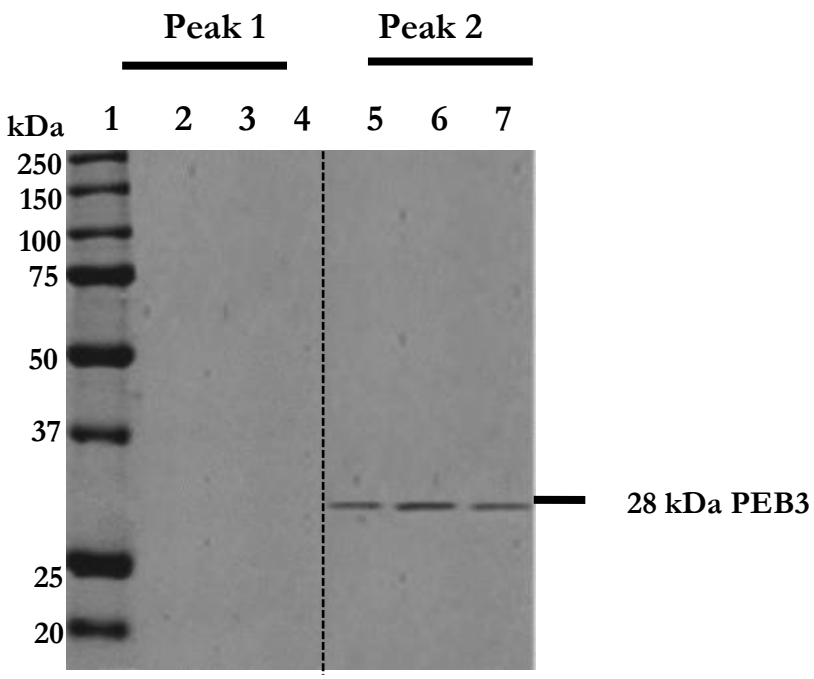


Figure 3.3 Immunoblot of proteins purified from *C. jejuni* 11168H acid-glycine extract by heparin-affinity chromatography. Peak fractions were transferred onto nitrocellulose for western blotting. Peak 1 elution fractions were loaded in lanes 2-4 showed no reactive band with the anti-PEB3 antibody. Elution fractions from Peak 2 (lanes 5-7) reacted with α-PEB3 antiserum revealing a 28 kDa band confirming the PEB3 protein. The dashed lined indicates where a region from this single blot was removed.

Protein ID	Potential Heparin-Binding Region	Location
PEB1 (Cj0921c)	$_{80}$DKKIKL$_{85}$ $_{136}$EKKYKS$_{141}$ $_{115}$RKR$_{117}$	Part of a β-strand End of a 3_{10} helix α-helix
PEB3 (Cj0289c)	$_{62}$KAKKD$_{66}$ $_{122}$NKKVRI$_{127}$	α-helix Loop between α-helix
Tungstate-binding protein TupA (Cj1540)	$_{133}$RGD$_{135}$ $_{222}$PKHCKN$_{227}$ $_{265}$KTRKD$_{269}$	No Crystal Structure
ATP/GTP-Binding Protein (Cj1041c)	$_{106}$KTKKE$_{110}$ $_{174}$RKKHK$_{178}$	No Crystal Structure
PEB4 (Cj0596)	No GAG-binding regions identified	

Table 3.2 Potential heparin-binding regions of *C. jejuni* proteins.

3.5 Further characterisation of the PEB3 heparin-binding protein

3.5.1 Putative PEB3 heparin-binding sites

The remainder of this study will focus on the characterisation of the PEB3 protein. Preliminary data suggest that PEB3 is a heparin-binding adhesin. Regions within the PEB3 protein, namely two putative heparin-binding sites: $_{122}$NKKVRI$_{127}$ and $_{62}$KAKKD$_{66}$ were investigated. The $_{122}$NKKVRI$_{127}$ site, located between a α-helix and β-strand, follows the consensus sequence pattern, XBBXBX (Capila & Linhardt, 2002). The second site, $_{62}$KAKKD$_{66}$, is located on α-helix and follows the alternative sequence, BXBBX (Fromm et al., 1997).

3.5.2 Molecular modelling of the putative PEB3 heparin-binding sites

It was proposed that both heparin-binding sites were present on the protein surface. The PEB3 dimer was downloaded as a (PDB) file: 2HWX and both putative heparin-binding sites were mapped to Chain A and B with PyMOL. Molecular visualisation of the $_{62}$KAKKD$_{66}$ shows the K_{62} residue on the α-helix resides on the external surface of this protein structure. However, the A_{63} residue is oriented inward at the back of the α-helix resulting in a buried residue. Mapping of the remaining residues K_{64}, K_{65}, and D_{66} showed these residues are on the surface and extend from an α-helix (Figure 3.4).

The second site, $_{122}$NKKVRI$_{127}$, side chain views shows that residues N_{122}, K_{123}, K_{124}, and R_{126} are surface exposed facing away from any other secondary structures (Figure 3.5). In contrast, residues V_{125} and I_{127} face inward towards the protein. In summary, the basic amino acids of both putative heparin-binding sites are surface-exposed. However, the interactions of these residues with heparin requires investigation *in silico* and experimentally.

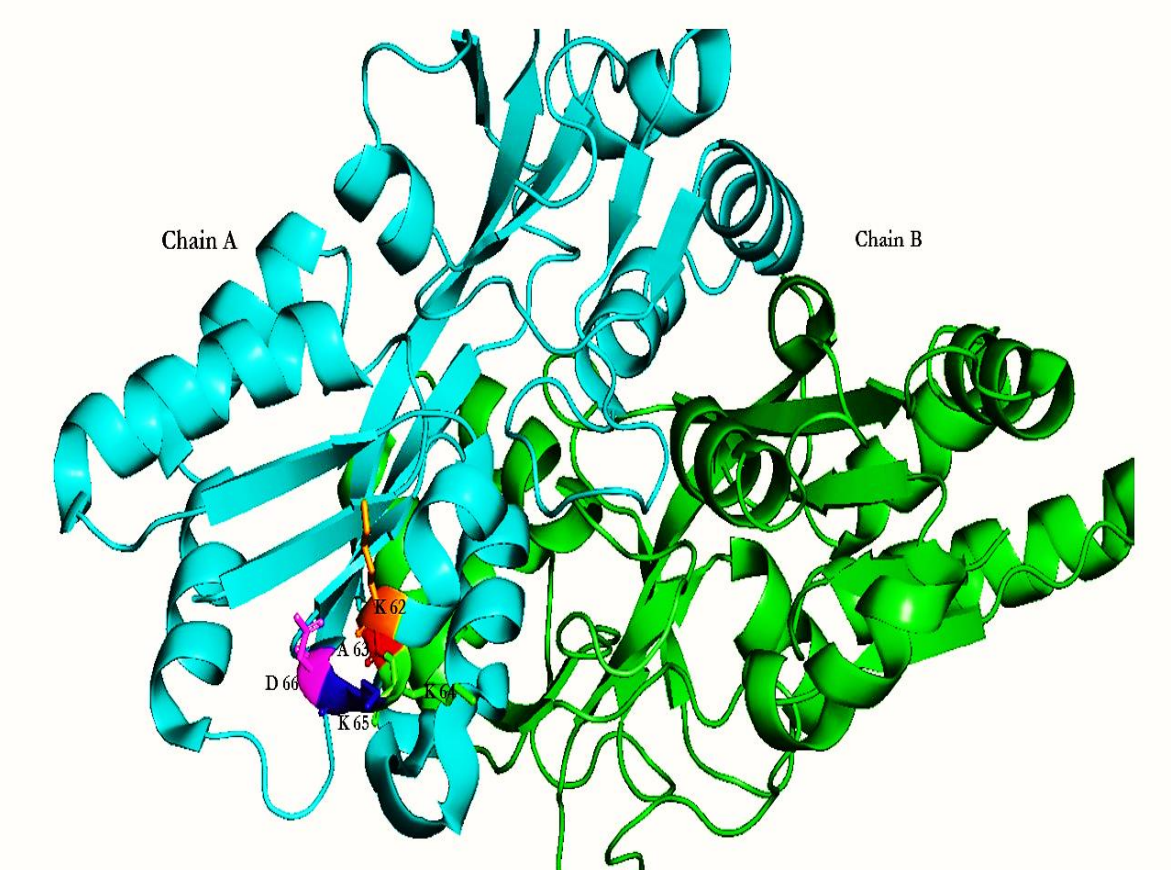


Figure 3.4 Molecular visualisation of the putative heparin-binding site $_{62}$KAKKD$_{66}$ on the PEB3 dimer. The $_{62}$KAKKD$_{66}$ site is located on the second α-helix of both Chains A and B (data not shown). The residues K_{62}: orange, A_{63}: red, K_{64}: green, K_{65}: blue and D_{66}: pink are depicted in the absence of ligand. The protein is centred and close-ups of these sites are shown in zoom mode of 20 angstroms. Images were created with PyMOL visualisation software.

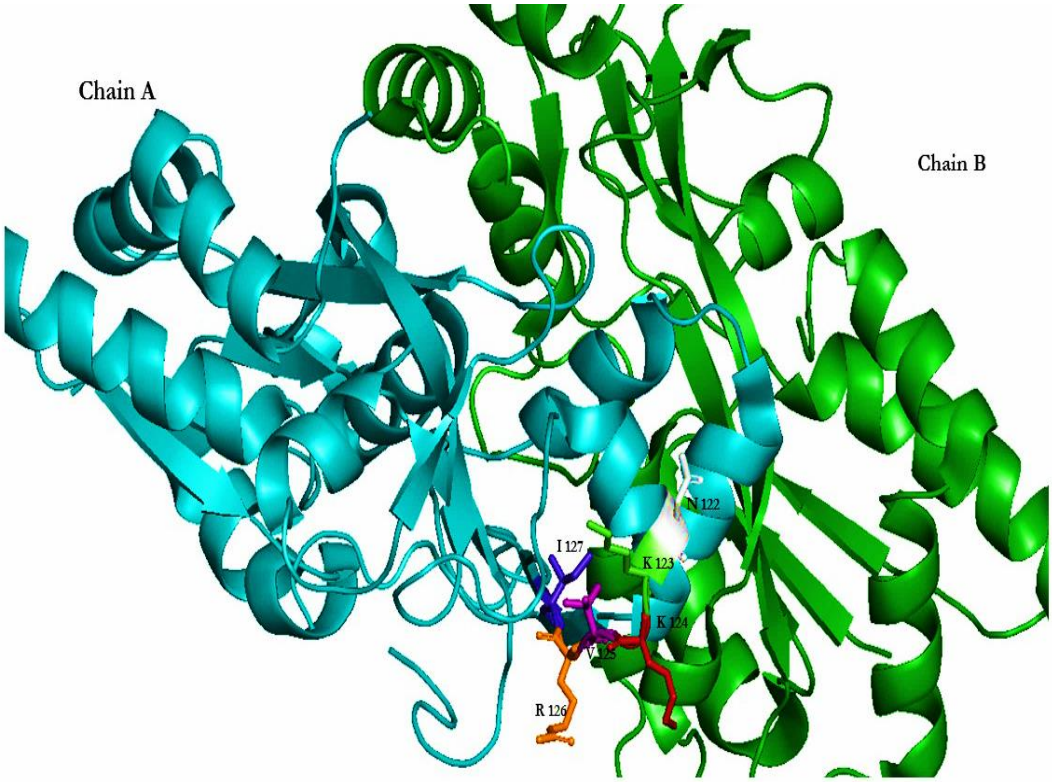


Figure 3.5 Molecular visualisation of the putative heparin-binding site $_{122}$NKKVRI$_{127}$ on the PEB3 dimer. The cluster $_{122}$NKKVRI$_{127}$ is located on a loop of both chains A and B (data not shown). The residues N_{122}: white, K_{123}: green, K_{124}: red, V_{125}: pink, R_{126}: orange and I_{127}: purple are depicted in the absence of ligand. The protein is centred and close-ups of these sites are shown in zoom mode of 20 angstroms. Images were created with PyMOL visualisation software.

3.5.3 *In silico* prediction of the heparin-binding sites of PEB3

The ClusPro server has been modified to locate heparin-binding sites on proteins (Mottarella et al., 2014). One pitfall of the ClusPro programme, concerns target proteins containing multiple chains. Unfortunately, the programme favours the crevices found between these multiple chains defining them as binding sites (Mottarella et al., 2014). A second pitfall of the programme concerns proteins with identical chains such as PEB3. The programme identifies the same site on an identical chain as a binding site, even in the absence of heparin (Mottarella et al., 2014). To avoid these problems, *in silico* modelling was conducted with a single chained PEB3 monomer.

3.5.4 Docking of the heparin tetrasaccharide ligand to the PEB3 monomer

The docking of heparin to the PEB3 monomer used the heparin ligand method, an advanced option on the ClusPro server at http://cluspro.bu.edu/ (Mottarella et al., 2014). First, a contact map identifying the regions of the PEB3 protein in contact with heparin was supplied. The red areas of the contact map represented the largest number of contacts with the docked heparin; white areas had few contacts and dark-grey areas had none. The contact map revealed several heparin-binding residues (data not shown). However, only $_{62}$KAKKD$_{66}$ and $_{122}$NKKVRI$_{127}$ were of concern. The $_{62}$KAKKD$_{66}$ site was located in areas of dark-grey, suggesting it does not participate in heparin-binding (Figure 3.6 A/B). In contrast, the $_{122}$NKKVRI$_{127}$ site namely, N$_{122}$, K$_{123}$, and K$_{124}$ were observed in white areas of the protein, indicative of contact with heparin (Figure 3.7 A/B).

Further to these contact maps, scrutiny of the 15 docking models of the PEB3 monomer: heparin, identified additional heparin-binding residues (Table 3.3). Surprisingly, a large number the heparin-binding residues were located in the PEB3 ligand-binding site, Thr 138 and Ser 139 (Table 3.3-Model 1-4, 12). These residues along with Ser 173, Asn 137, Ser 171 and Asn 170 (Table 3.3 Model 1-4, 12) are known to form hydrogen bonds with the

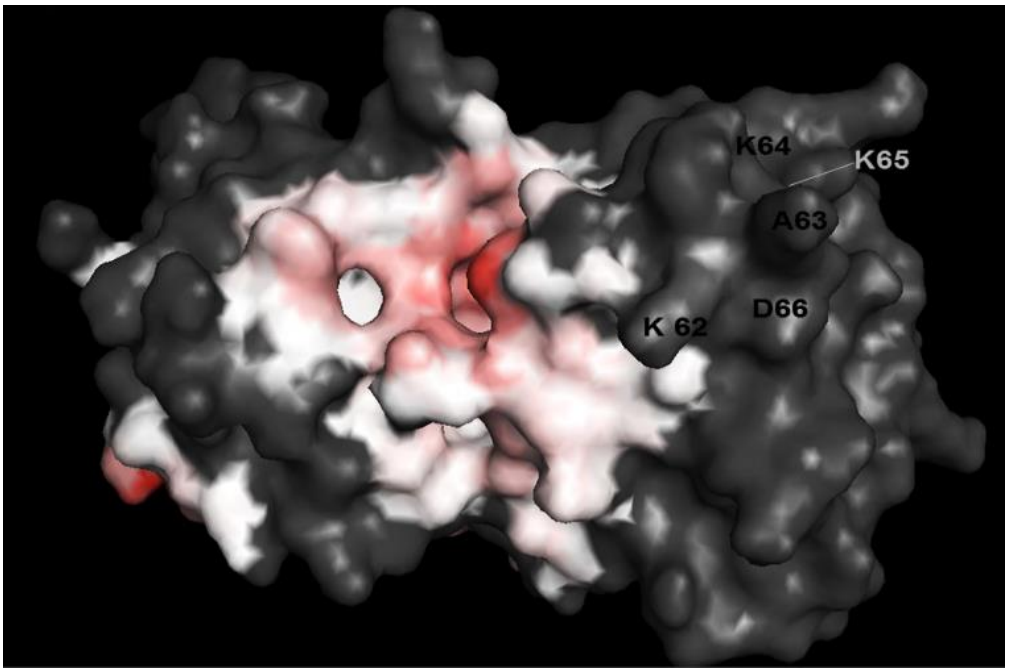
carboxylic groups of citrate (Rangarajan et al., 2009) (Table 3.3-Model 1-4, 12). Additionally, phosphate molecules such as 3-PG, bind this site, forming hydrogen bonds with the previously mentioned residues and Gly 28 (Table 3.3-Model 1-4, 12) (Min et al., 2009). Heparin-binding residues N_{122}, K_{123}, K_{124} of the proposed site $_{122}NKKVRI_{127}$ as well as K_{115} and K_{113} were observed in Model 8 (Table 3.3 & Figure 3.8). Furthermore, K_{123} is identified as a heparin-binding residue in Model 9 alongside K_{113}, K_{115}, K_{207} and N_{110} (Table 3.3).

3.5.5 Docking of the heparin tetrasaccharide ligand to the PEB3 dimer

Docking to specific sites of attraction, $_{62}KAKKD_{66}$ and $_{122}NKKVRI_{127}$, on the PEB3 dimer was carried out. This was to deter the ClusPro programme from favouring the crevice between the chains as binding sites, an issue encountered with previous dockings with the PEB3 dimer (data not shown). Results from this specified docking provided further confirmation that $_{62}KAKKD_{66}$ did not interact with heparin *in silico* (data not shown). In contrast, the second site $_{122}NKKVRI_{127}$ namely residues N_{122}, K_{123}, K_{124} were observed to facilitate the PEB3 dimer/heparin interaction (Figure 3.9). Additional residues K_{113}, K_{115} and N_{159} were also identified (Figure 3.9).

In summary, *in silico* data confirms at least one heparin-binding site is involved in aiding biding of PEB3 to heparin however, these data require experimental confirmation. Furthermore, a number of other residues were found to facilitate heparin-binding. However, the hypothesis is that these two putative sites, $_{62}KAKKD_{66}$ and $_{122}NKKVRI_{127}$, which follow the consensus sequence motifs, XBBXB and BXBBX facilitate heparin-binding (Fromm et al., 1997; Hileman et al., 1998). As these motifs are found in a number of heparin-binding proteins it was therefore decided to conduct further experimental work (i.e. site-directed mutants) in these sites only.

A



B

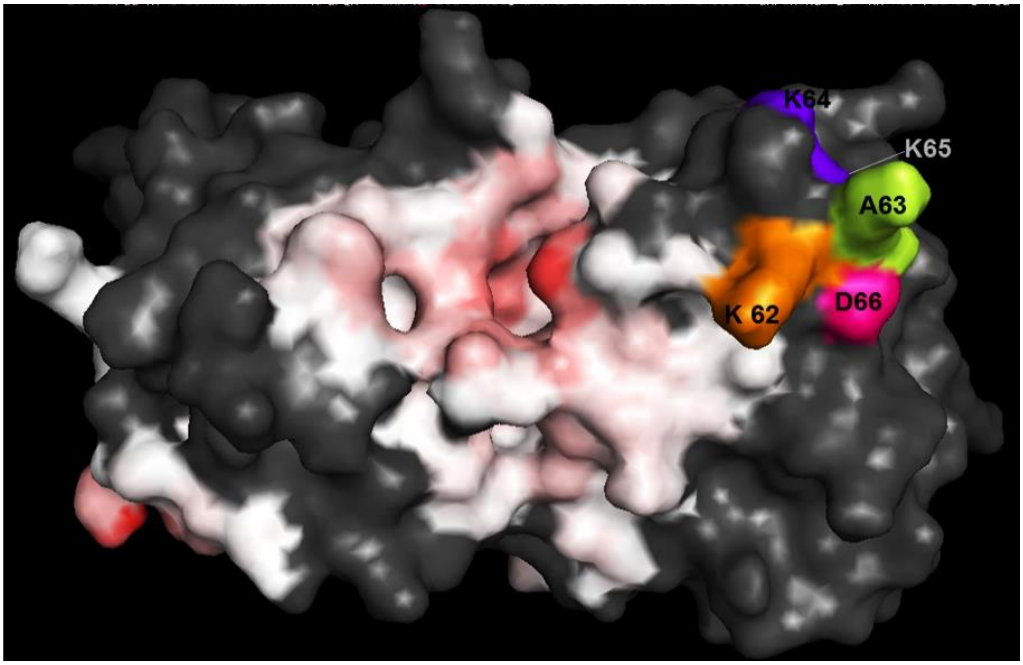
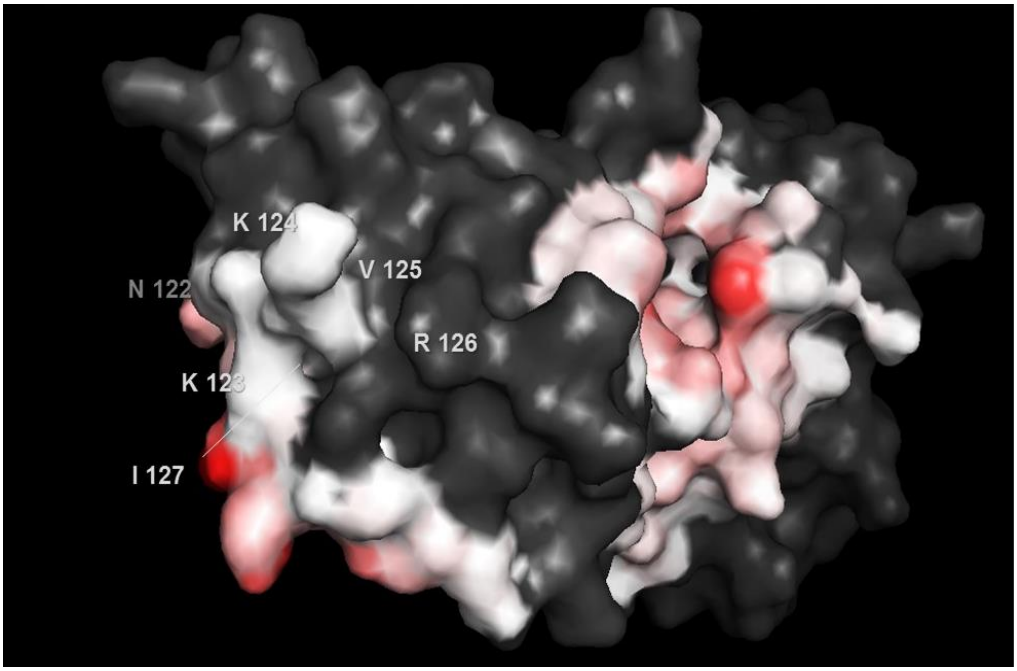


Figure 3.6 The $_{62}$KAKKD$_{66}$ site on the PEB3 monomer is not involved in heparin-binding. A) The red (hot) areas represent a large number of contacts with heparin; the white areas have fewer, and dark areas have none. The figure shows residues from the $_{62}$KAKKD$_{66}$ region of the protein occupying the dark areas. **B)** The contact map in coloured form showing: K_{62}: orange, A_{63}: green, K_{64}: purple, K_{65}: buried and D_{66}: pink, for comparison of regions.

A



B

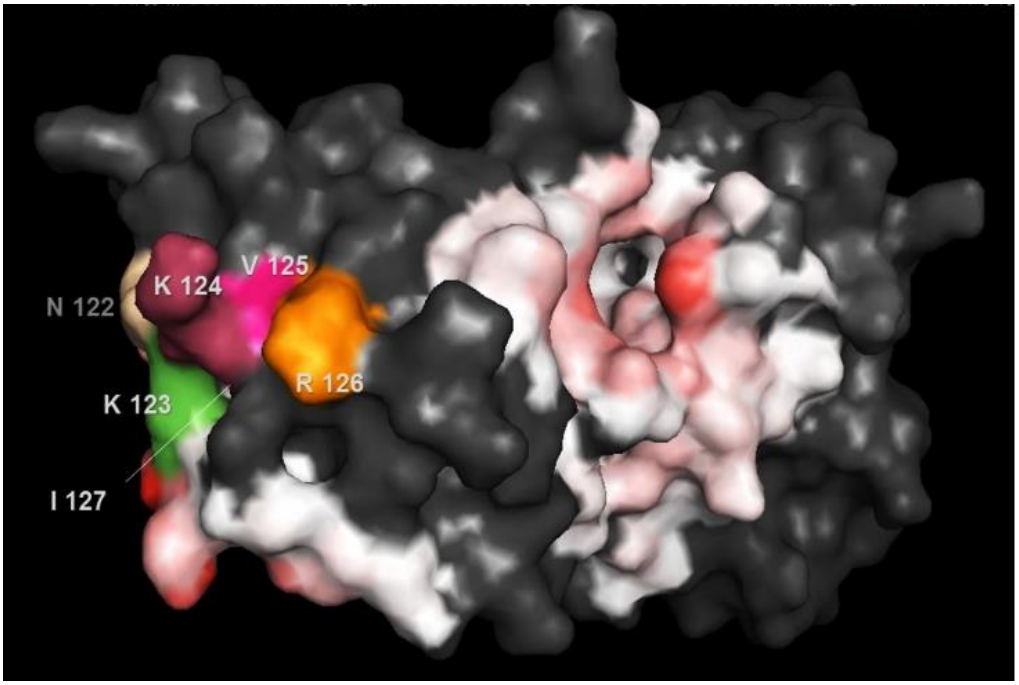


Figure 3.7 The $_{122}$NKKVRI$_{127}$ site on the PEB3 monomer is involved in heparin-binding. A) The red (hot) areas represent a large number of contacts with heparin; the white areas have fewer, and dark areas have none. The figure shows the three residues of $_{122}$NKKVRI$_{127}$ interact with heparin, N_{122}, K_{123} and K_{124}. **B)** The contact map in coloured form showing: N_{122}: beige, K_{123}: green, K_{124}: maroon, V_{125}: hot pink, R_{126}: orange and I_{127}: buried, for comparison of regions.

Model	# of members per/1800 poses (probability)	Predicted Sites
0	339 (0.18)	Lys 113, Lys 115, Lys 118, Lys 207
1	259 (0.14)	Gly 28, Lys 35, Val 51, Phe 53, Ser 136, Thr 138, Thr 139, Asn170, Ser 171, Ser 173
2	245 (0.13)	Gly 28, Lys 35, Val 51, Thr 58, Ser 136, Thr 138, Thr 139, Asn 170, Ser 171, Ser 173
3	214 (0.12)	Gly 28, Phe 53, Arg 100, Ser 136, Asn 137, Thr 138, Ser 139, Asn 170, Ser 171, Gly 172, Ser 173, Asp 191
4	179 (0.10)	Gly 34, Val 129, Lys 135, Asn 137, Thr 138, Val 168, Ser 171, Ser 173, Ala 174, Phe 178, Ala 179, Trp 192
5	152 (0.08)	Lys 194, Ala 204, Lys 207, Tyr 212, Arg 249, His 251
6	141 (0.079)	Gly 109, Asn 110, Lys 113, Lys 115, Thr 201, Ala 202, Asn 204,
7	104 (0.06)	Lys 194, Lys 207, Lys 244, Gly 247, Arg 249, His 251
8	68 (0.04)	Lys 113, Lys 115, **Asn 122**, **Lys 123**, **Lys 124**
9	36 (0.02)	Lys 113, Lys 115, **Lys 123**, Lys 207, Asn 110
10	29 (0.02)	Lys 44, Lys 155, Lys 237, Glu 238, Lys 240, Lys 244, Glu 250, His 251
11	12 (0.01)	Lys155, Lys 237, Lys 240, Lys 244, Glu 250, His 251
12	9 (0.005)	Gly 28, His 31, Lys 35, Gln 56, Ser 136, Ser 137, Thr 138, Ser 139, Gly 140, Asn 170, Gly 172, Ser 173
13	5 (0.02)	Gln 152, Lys 194, Lys 207, Lys 244, Lys 245, Arg 249, Gln 250, His 251
14	3 (0.01)	Lys 194, Lys 207, Lys 244, Arg 249, Glu 250, His 251

Table 3.3 Predicted heparin-binding sites of the PEB3 monomer. The table lists the PEB3 residues that interact with a heparin tetrasaccharide probe for the top fifteen model results (column 3). In addition to the list, the program generates 1800 poses for the fifteen model results. The members (or the number of times the model occurs per 1800 poses) can be ranked from high to low and is correlated with the model number (Column 1:0-15). The experimental probability (# members/1800 poses) is shown in parentheses in column 2. The residues in bold indicate residues from the putative heparin-binding sites: $_{62}$KAKKD$_{66}$ and $_{122}$NKKVRI$_{127}$.

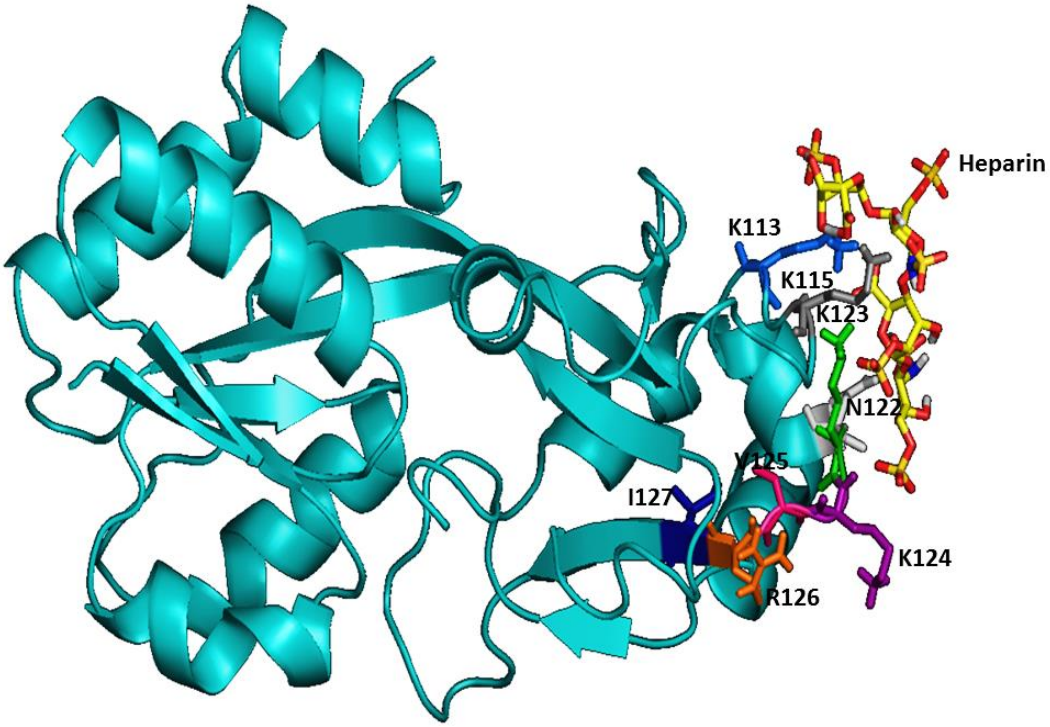


Figure 3.8 Docking of the heparin tetrasaccharide ligand to the $_{122}$NKKVRI$_{127}$ site of the PEB3 monomer. The model (Number 8) shows K_{113}: marine blue, K_{115}: gray and three residues N_{122}: white, K_{123}: green, and K_{124}: purple of $_{122}$NKKVRI$_{127}$ site participating in heparin-binding. Results are generated from the ClusPro server and visualised in PyMOL.

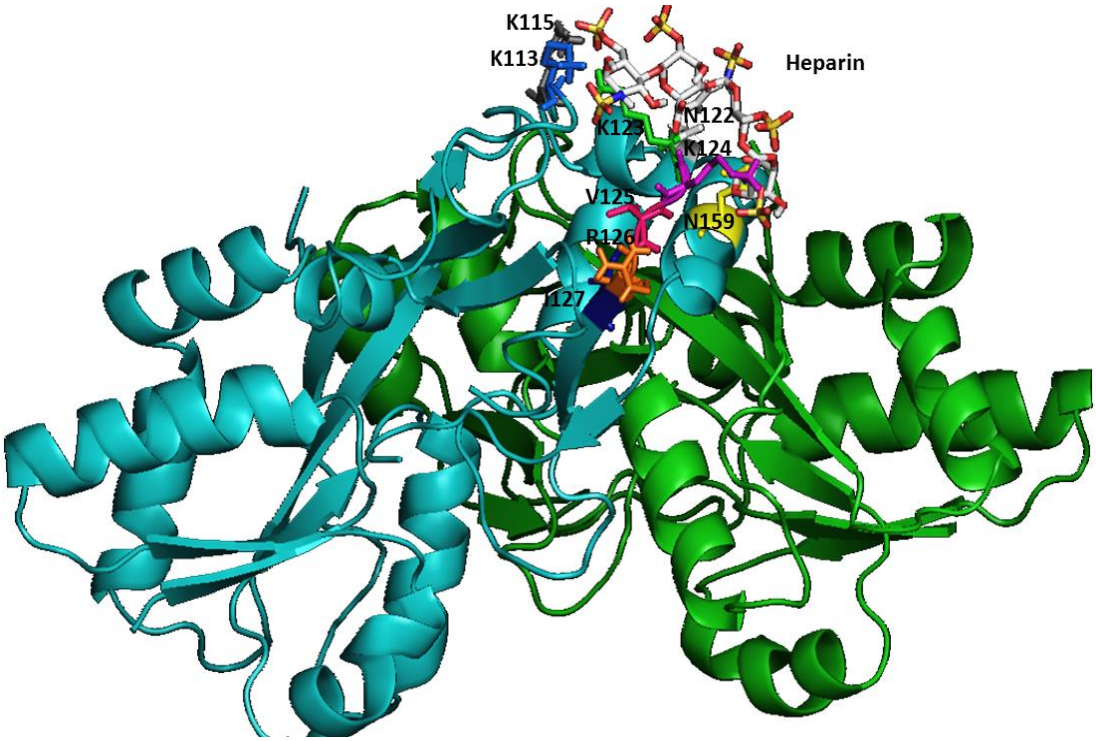


Figure 3.9 Docking of heparin tetrasaccharide ligand to the $_{122}$NKKVRI$_{127}$ site of the PEB3 dimer. The model depicts three residues N_{122}: white, K_{123}: green, and K_{124}: pink of $_{122}$NKKVRI$_{127}$ interacting with heparin. Additional residues include: K_{113}: marine blue, K_{115}: gray and N_{159}: yellow. Results are generated from the ClusPro server and visualised with PyMOL.

3.6 Role of specific clusters of PEB3 positively charged residues in heparin-binding investigated by site-directed mutagenesis

The *in silico* models suggest heparin interacts with $_{122}$NKKVRI$_{127}$, but not $_{62}$KAKKD$_{66}$ of PEB3. Confirmation of these data used site-directed mutants which were constructed by substituting alanine for lysine and arginine residues. Alanine, a non-polar amino acid with a neutral charge, is ideal because it can occupy both hydrophobic and hydrophilic regions without altering protein conformation (Klann et al., 1994).

Prior to my work, the *peb3* gene without the signal peptide sequence was cloned into a pET-41a (+) vector. The vector contains an N-terminal GST-tag, enabling purification of the fusion protein from an *E. coli* expression system. The expression of PEB3 is regulated via the T7 lac promoter in an IPTG-inducible manner. From the template plasmid pET-41a(+) PEB3, two site-directed mutations were introduced via site-directed mutagenesis with primers 321/322 and 299/300 to generate plasmids pET-41a(+)PEB3: K64A and K123A (supplied by the Linton lab). With these template plasmids, further residues K65, K124, R126A were substituted with alanine residues via site-directed mutagenesis with primers, 776/777, 301/302, 774/775 and 801/802 (Appendix Table S1), respectively, creating plasmids pET-41a(+)PEB3: K64A/K65A, K124A, K123A/K124A and K123A/K124A/R126A (Appendix Table S2). Plasmids were transformed into *E. coli* DH5α, sequenced with primers 81/327 (Appendix Table S1) and subcloned into *E. coli* BL-21 (AI) cells for protein expression and purification. Approximately 400 ml of culture was grown to mid-log phase, induced with arabinose and IPTG. After four hours of further incubation, at 37 °C with shaking at 200 rpm, cells were harvested by centrifugation. The pellet was re-suspended into GST-binding buffer, cells lysed by French Press and following centrifugation the supernatant was added to a GST-Sepharose column. After 30 minutes, the column was washed and PEB3 released by thrombin cleavage. Samples were taken throughout the purification process and

analysed by SDS-PAGE and western blot. Prior to purification, the GST-PEB3 fusion protein was detected on a Coomassie stained SDS-PAGE at ~54 kDa, the size of PEB3 (~28 kDa), and the GST-Tag (~26 kDa) combined (Figure 3.10 A). The pure PEB3 protein was visible after thrombin cleavage on a Coomassie stained SDS-PAGE at ~28 kDa and a protein of similar size was detected with an anti-PEB3 antiserum confirming protein identity (Figure 3.10 A & B). This protocol was repeated for all six PEB3 site-directed mutants.

After GST-purification, WT PEB3 and site-directed mutants were applied to a heparin-agarose affinity chromatography to assess their heparin-binding capacities. Purified PEB3 passed through a heparin column and bound protein eluted with a 0-1 M NaCl gradient via the AKTA purification system (see Chapter 2). The recombinant PEB3 WT was eluted from the column at 335 mM NaCl. This affirms that the recombinant version of PEB3 has similar heparin-binding properties as native PEB3 from *C. jejuni* (Figure 3.1). The site-directed mutants of PEB3 in both potential heparin-binding sites were similarly analysed (Figure 3.11 A/B). Mutants K64A and K64A/K65A were eluted from the heparin column with a 6% reduction in NaCl concentration. Mutant K123A required 327 mM NaCl to elute, similar to the WT PEB3. A single mutation at the K124A residue eluted at 296 mM NaCl and the K123A/K124A eluted at 249 mM NaCl, as did the triple mutant K123A/K124A/R126A (Figure 3.11 A/B). Following elution from the heparin column peak fractions was collected. The identity of eluted protein was further confirmed by western blotting with an α-PEB3 antiserum (Figure 3.3).

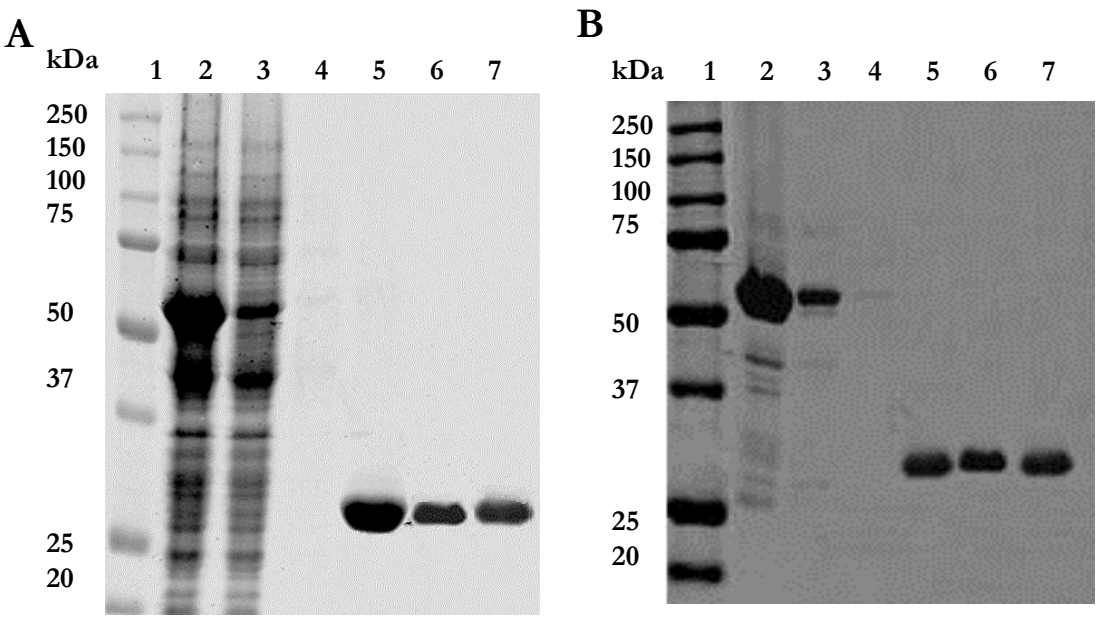
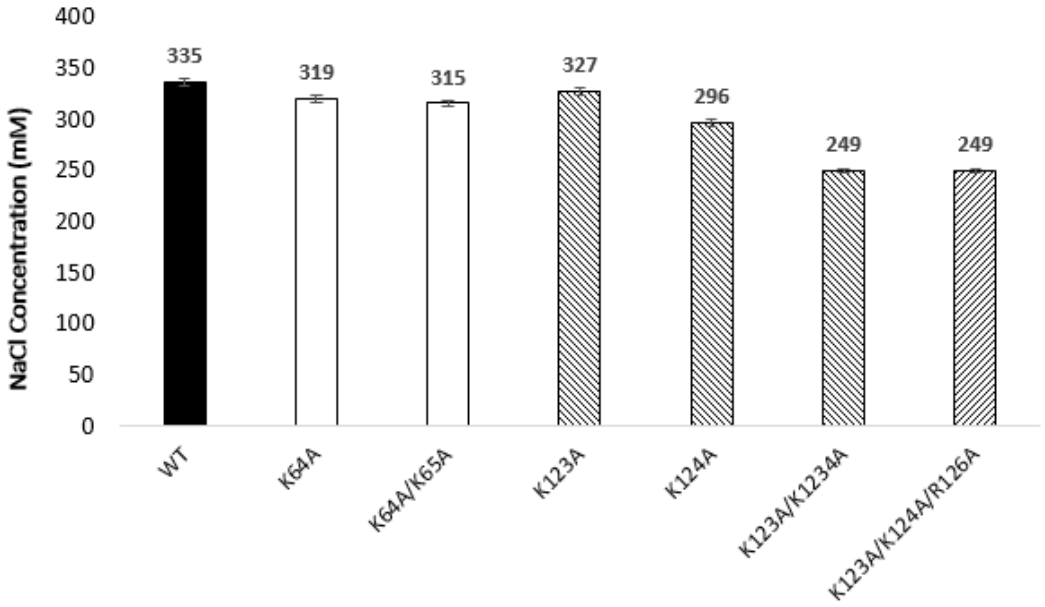


Figure 3.10 SDS-PAGE and Western blot of affinity purification of PEB3 produced in *E. coli*. Proteins were separated by SDS-PAGE and stained with Coomassie Instant blue **(A)** or transferred onto nitrocellulose and probed with anti PEB3 antiserum **(B)**. Lane one contains Precision plus protein standards marker; lane 2: cell extracts from *E. coli* producing PEB3-GST fusion protein; lane 3: extract following a 30-minute incubation on the GST column; lane 4: wash fractions from column; lanes 5, 6, and 7 contain fractions eluted from the column by thrombin to release PEB3 from GST-PEB3 fusion protein bound to column.

A



B

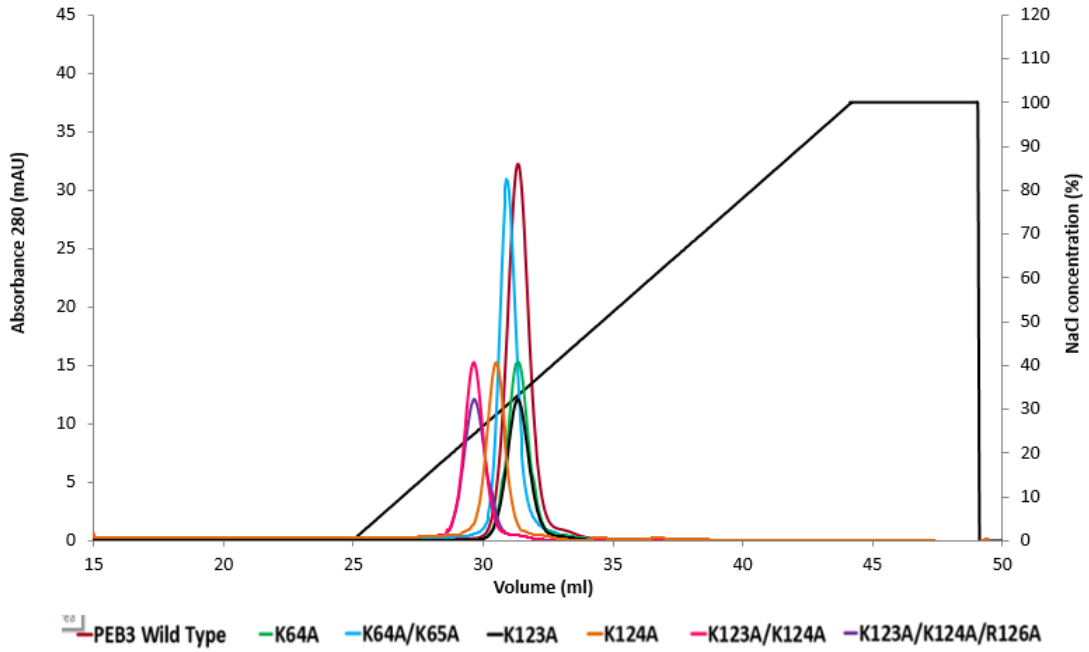


Figure 3.11 Heparin binding capacity of PEB3 WT and site-directed mutants. Wild-type PEB3 eluted from a heparin column at 335 mM NaCl (Figure B-Red Peak). Site directed mutants K64A (Figure B-Green Peak) and K64A/K65A (Figure B-Blue Peak) eluted at 315 and 310 mM NaCl, respectively. The single mutants K123A (Figure B-Black Peak) and K124 (Figure B-Orange Peak) eluted at 327 and 296 mM NaCl, respectively. More significantly site-directed mutants K123A/K124A (Figure B-Pink Peak) and K123A/K124A/R126A (Figure B-Purple Peak) eluted at 249 mM NaCl. Experiments were conducted in triplicate, and error bars indicate standard error, n=3.

3.7 Discussion

A number of bacterial adhesins interact with heparin (Chen et al., 1995; Serruto et al., 2010; Duncan et al., 2011). Binding to heparin is associated with adherence, increased resistance to human sera and evasion of the host immune system (Duensing & Putten, 1998; Menozzi et al., 2002). In this study, I identified five *C. jejuni* heparin-binding proteins (Table 3.1). These proteins were isolated via the acid-glycine method a common approach for isolating surface associated proteins (Logan & Trust, 1983). This approach relies on proteins being soluble in the acid wash solution and so hydrophobic integral membrane proteins are not well represented (Logan & Trust, 1983; Scott & Cordwell, 2009). Thus, it would be premature to state these are the only heparin-binding proteins produced by *C. jejuni*.

The heparin-binding affinities of these five proteins were obtained via heparin-agarose chromatography (Table 3.1). This is a common method that measures the affinity or the strength of attraction between a receptor and its ligand. In heparin-agarose chromatography heparin is covalently linked to a Sepharose column and serves as the ligand (Esko & Lindhardt, 2009). Proteins bound to heparin are displaced and eluted from the column with varying concentrations of NaCl. The concentration required for elution may be generally proportional to the dissociation constant (K_d) a form of measurement that indicates the strength of binding between a receptor and ligand (Esko & Lindhardt, 2009). However, this is based on the assumption that the heparin-protein interaction is strictly ionic, which is not always true as other factors such as hydrogen bonding, hydrophobic, electrostatic interactions, and Van der Wales forces may be involved (Hileman et al., 1998). Thus a limitation of heparin-agarose chromatography is that this method does not provide kinetic data, regarding association and dissociation rate constants, which is important as K_d values indicate that the affinity between ligand and protein is good (protein/ligand tightly bound) (Esko & Lindhardt, 2009). Furthermore, K_d values indicate the strength of binding and

specificity of the protein ligand interaction. Regardless, heparin-agarose chromatography is used as an assessment of relative affinity (Esko & Lindhardt, 2009).

The heparin-binding affinities of these proteins are consistent with heparin-binding proteins from other bacterial species such as *N. meningitidis*, *B. pertussis*, *M. tuberculosis* and *L. pneumophila* (Hannah et al., 1994; Menozzi et al., 1996; Serruto et al., 2010; Duncan et al., 2011). The Neisserial Heparin-Binding Antigen (NHBA) is expressed by *N. meningitidis* (Serruto et al., 2010). Similar to that (190 mM) of the PEB1 adhesin, the full-length NHBA protein is eluted at a low NaCl concentration of 230 mM (Table 3.1). Regardless, the interaction between NHBA-heparin promotes the persistence of *N. meningitis* in the human host by providing protection against human serum (Serruto et al., 2010). The remaining four *C. jejuni* proteins displayed a moderate affinity for heparin with an elution of 335 mM NaCl. This value is comparable to that (380 mM) reported for the Lcl adhesin from *L. pneumophila*, to that (350 mM) of heparin binding to the HBHA of *M. tuberculosis* and to that (300 mM) of heparin binding FHA of *B. pertussis* (Table 3.1) (Hannah et al., 1994; Menozzi et al., 1996; Duncan et al., 2011). These data highlight the significant interactions of these *C. jejuni* proteins with heparin in terms of affinity. However, further work is needed to determine the kinetics, binding strength and specificity of these interactions as this cannot be drawn from these data.

Heparin is a highly sulphated glycosaminoglycan and it was proposed that only basic proteins would elute from the heparin column. However, upon further observation the pIs of the *C. jejuni* heparin-binding proteins ranged from 6.4 to 9.9 (Table 3.1). Furthermore, *in silico* analysis of these proteins identified putative heparin-binding sites with the sequences XBBXBX (Capila & Linhardt, 2002) or BXBBX (Fromm et al., 1997) (Table 3.2), hydrogen-bonding amino acids (PEB4), and a high density of positive residues (Figure 3.2 A-E). However, further work should investigate these *in silico* features and their contribution to the heparin/protein interaction.

The remainder of the study focused identifying the heparin-binding properties of the PEB3 protein. The native version of the PEB3 protein from the acid glycine extract, which is glycosylated, eluted from the heparin column with moderate affinity (335 mM Nacl) (Figure 3.1). Recombinant PEB3 displays identical chromatographic behaviour to that of native PEB3 suggesting that the loss of glycosylation does not influence heparin-binding affinity (Figure 3.11). This is not surprising as the native HBHA form of *Mycobacterium tuberculosis* which is glycosylated binds heparin with same affinity as recombinant HBHA (Menozzi et al., 1998).

In silico analysis of PEB3 via PyMol indicated that residues from the site $_{122}$NKKVRI$_{127}$, $_{62}$KAKKD$_{66}$ are surface-exposed and might participate in heparin-binding (Figure 3.4 & 3.5). These data led me to test the effect of several PEB3 site-directed mutants, via heparin-affinity chromatography. Removal of K_{64} and K_{65} in site $_{62}$KAKKD$_{66,}$ impaired affinity for heparin indicating these residues are involved in PEB3/Heparin interactions (Figure 3.11 A/B). These unexpected findings counter the previous results identified from the ClusPro docking models, which do not indicate $_{62}$KAKKD$_{66}$ as a heparin-binding region (Figure 3.6-3.7 & Table 3.3). One explanation for the data maybe that the ClusPro programme cut-off point excludes these residues as they are not in top fifteen results (Table 3.3) (Mottarella et al., 2014). A second explanation maybe the heparin ligand in the experiment differs from that of the *in silico* modelling programme. The heparin column is coated in a sulphated glucosaminoglycan (5 to 30 kDa) from a porcine mucosa (GE Healthcare), whereas, a heparin tetrasaccharide is used for ClusPro (Mottarella et al., 2014). The heparin ligand on the chromatography column may allow for more flexibility, therefore the side chains of these two residues can bind heparin *in vitro*. Lastly, an attempt to remove K_{62} was unsuccessful.

In the second site $_{122}$NKKVRI$_{127,}$ residues K_{123} and K_{124} were removed resulting in reduced heparin-binding affinity (Figure 3.11). These data were in agreement with the *in silico* docking

models, which indicated these two residues are essential to the PEB/Heparin interaction (Figure 3.8 & Figure 3.9). Prior to these experiments it was hypothesised that $R_{126,}$ an amino acid capable of forming hydrogen bonds with heparin, would be critical for the PEB3/Heparin interaction. But this was not the case as the removal of R_{126} did not alter PEB3/Heparin-binding affinity (Figure 3.11 A & B). These data are consistent with the *in silico* models, whereby R_{126} did not bind heparin (Figure 3.8 & 3.9). The site-directed mutagenesis of the both sites did not abolish heparin-binding suggesting other residues may be involved. *In silico* models indicated the presence of additional residues ($N_{122,}$ K_{113} and K_{115}) on the monomer and on the dimer ($N_{122,}$ $K_{113,}$ K_{115} and N_{159}) (Table 3.3, Figure 3.8 & 3.9) may bind heparin. Further work should focus on experimentally investigating additional residues identified *in silico*.

In summary, five *C. jejuni* heparin-binding proteins were identified and the heparin-binding properties of PEB3 adhesin revealed. It is now known that PEB antigens bind heparin. Furthermore, these findings confirm the PEB3/heparin interaction is not strictly mediated by pI and electrostatic interactions, but involves two sites. Overall, the findings from this study have furthered our understanding of PEB3/heparin interactions, confirmed the function of PEB3 as an adhesin and indicate PEB3 may be involved in *C. jejuni* pathogenesis.

CHAPTER 4

Role of PEB3 and GlpT in 3-PG dependent growth and fosfomycin sensitivity

4.1 Introduction

C. jejuni colonises multiple hosts and acquires nutrients from the environment to enable its survival and growth, but does not metabolise simple sugars such as glucose and galactose (Velayudhan & Kelly, 2002; Stahl et al., 2012). Lack of 6-phosphofructokinase in *C. jejuni* means that the glycolytic pathway for metabolising glucose is not functional, however the enzymes "below" this point are present as discussed in Chapter 1 (Figure 1.6) (Velayudhan & Kelly, 2002; Stahl et al., 2012). Thus, it is possible that *C. jejuni* can metabolise compounds such as 3-PG and G3P (Hofreuter et al., 2006; Min et al., 2009). The periplasmic protein PEB3 was recently shown to bind 3-PG still its exact *in vivo* role remains unclear, though it may act to deliver 3-PG to a membrane transport protein (Min et al., 2009).

Scrutiny of the *C. jejuni* NCTC 11168H chromosome revealed a completely uncharacterised gene adjacent to *peb3* encoding GlpT, a member of the largest class of gradient-driven membrane transporters, the Major Facilitator Superfamily (Law et al., 2008). GlpT has been studied in *Haemophilus influenzae* (Song et al., 1998), *Bacillus subtilis* (Nilsson et al., 1994) and *E. coli* (Lemieux et al., 2004a; Lemieux et al., 2004b) with the *E. coli* GlpT particularly well characterised. This membrane protein is an antiporter exchanging G3P for inorganic phosphate (P_i) via a concentration gradient (Lemieux et al., 2004b). GlpT also imports the phosphate-containing antibiotic fosfomycin into the cell and *glpT* defective strains have decreased antibiotic uptake and enhanced fosfomycin resistance (Nilsson et al., 2003; Castañeda-García et al., 2013).

In addition to GlpT, *E. coli* produces a second transporter, UhpT that imports glucose-6-phosphate as well, as fosfomycin (Kadner & Shattuck-Eidens, 1983; Castañeda-García et al., 2013). This is also an antiporter exchanging internal phosphate (P_i) for glucose-6-phosphate (Sonna et al., 1988). In *P. aeruginosa* and *C. jejuni*, UhpT is absent, but GlpT is present (Castañeda-García et al., 2009; Castañeda-García et al., 2013).

Intriguingly, the *C. jejuni glpT* genes are genetically variable due to indels at various specific locations, which is discussed in detail in Chapter 5. For example, 11168H *glpT* is annotated as a pseudogene, composed of three overlapping ORFs due to two frameshift mutations. The function of *C. jejuni glpT* is not known. The location of *glpT* adjacent to *peb3* might suggest a related function for the corresponding proteins. A plausible hypothesis is that PEB3 is the periplasmic binding partner for transport of 3-PG via GlpT. In this study, I investigated the roles of PEB3 and GlpT in *C. jejuni* NCTC 11168H by constructing a *glpT* insertional knockout mutant and genetic complement of previously constructed *peb3* mutants, these were examined for 3-PG dependent growth and susceptibility to fosfomycin.

Results

4.2 Confirmation that PEB3 is not produced in the insertional knockout mutant

An insertional knockout strain of *peb3* (11168H *peb3*), and a genetic complement (11168H *peb3*$^+$) whereby *peb3* was reintroduced onto the chromosome of 11168H *peb3*, were supplied by the Linton lab. Along with the wild-type 11168H, these strains were tested for PEB3 production. Whole-cell extracts prepared from the harvested cells of overnight plates were standardized to an OD_{600} of 2.5 in PBS, separated on an SDS-PAGE gel, blotted onto a nitrocellulose membrane and probed with an anti-PEB3 antiserum (Linton & Wren, unpublished). A band of ~28 kDa was produced for both NCTC 11168H and 11168H *peb3*$^+$, but not for the 11168H *peb3* strain, confirming the knockout strain does not produce PEB3 (Figure 4.1-Lanes 2 & 3). The blot also shows that although PEB3 production is restored in the genetically complemented strain, it is not to the same level as the WT strain (Figure 4.1-Lanes 2 & 4).

4.3 Growth characteristics of WT and *peb3* mutants

Growth curves of *C. jejuni* WT NCTC 11168H, 11168H *peb3* and 11168H *peb3*$^+$ were generated to determine if PEB3 was important for *in vitro* growth in complex media (Figure 4.2). Cultures were standardised to a starting OD_{600} of 0.05 in MH broth without antibiotics, and growth was monitored periodically by measuring OD_{600} (see Chapter 2). The 11168H *peb3* and 11168H *peb3*$^+$ strains initially grew at a slower rate compared to the WT strain, but at approximately 16 hours, all three had similar OD_{600} values. All three cultures grew to a maximal OD_{600} of about 0.7 by ~24 hrs (Figure 4.2). This was in agreement with previous experiments in the lab and with published data (Davis & DiRita, 2008).

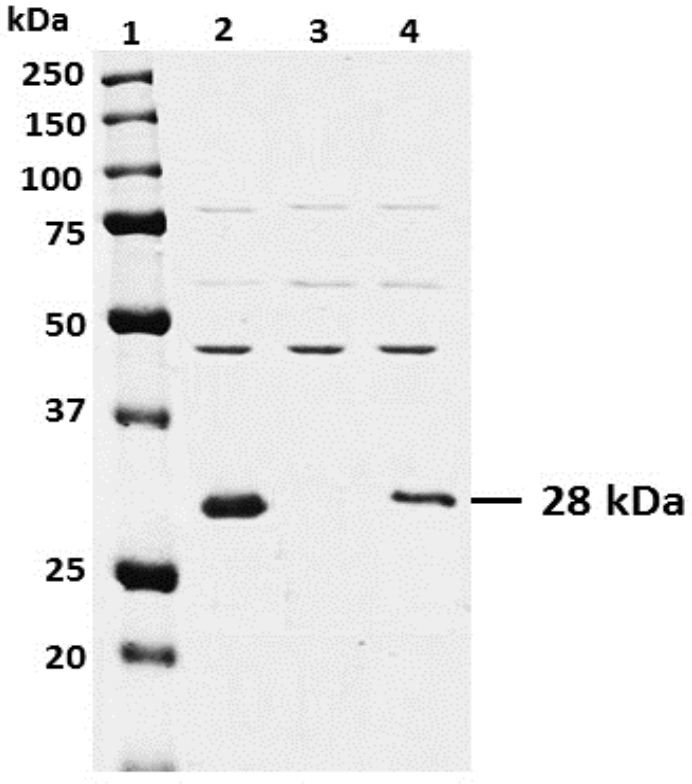


Figure 4.1 Immunoblot of *C. jejuni* WT and *peb3* mutant strains. Whole-cell lysates of 11168H (lane 2), 11168H *peb3* (lane 3) and 11168H *peb3*$^{+}$ (lane 4) were separated by SDS-PAGE and blotted onto a nitrocellulose membrane. The membrane was probed with a rabbit anti-PEB3 antiserum (Linton & Wren, unpublished). Bands consistent with the predicted PEB3 Mr of approximately 28 kDa, were observed in 11168H and 11168H *peb3*$^{+}$ (lanes 2 and 4, respectively), but not in 11168H *peb3* (lane 3).

4.4 Confirmation of the role of PEB3 in 3-PG dependent growth promotion in MEMα

PEB3 binds phosphate-containing compounds more specifically 3-PG. The replacement of WT *peb3* with site-directed mutant versions T138A/S139A and K135E at the 3-PG binding site abolished 3-PG dependent growth promotion (Min et al., 2009). To verify these data, I similarly examined the 3-PG dependent growth promotion in 11168H, 11168H *peb3* and 11168H *peb3*$^+$ (Figure 4.3). Cultures were grown to mid-log phase in MH broth (Figure 4.2) and inoculated in MEM-α with 0.1 mM $FeSO_4$ and either 0, 1 or 5 mM 3-PG to an initial OD_{600} of 0.05 (see Chapter 2). Cultures were incubated under microaerobic conditions at 42 °C for 24 hours with gentle shaking and the OD_{600} measured. The addition of 5 mM 3-PG to the minimal media resulted in significantly enhanced growth of 11168H. This enhanced growth was not observed for 11168H *peb3* and was partially restored for 11168H *peb3* when supplemented with 0 or 1 mM 3-PG. These results confirm work by Min and co-workers that PEB3 is required for 3-PG dependent growth promotion in MEMα medium.

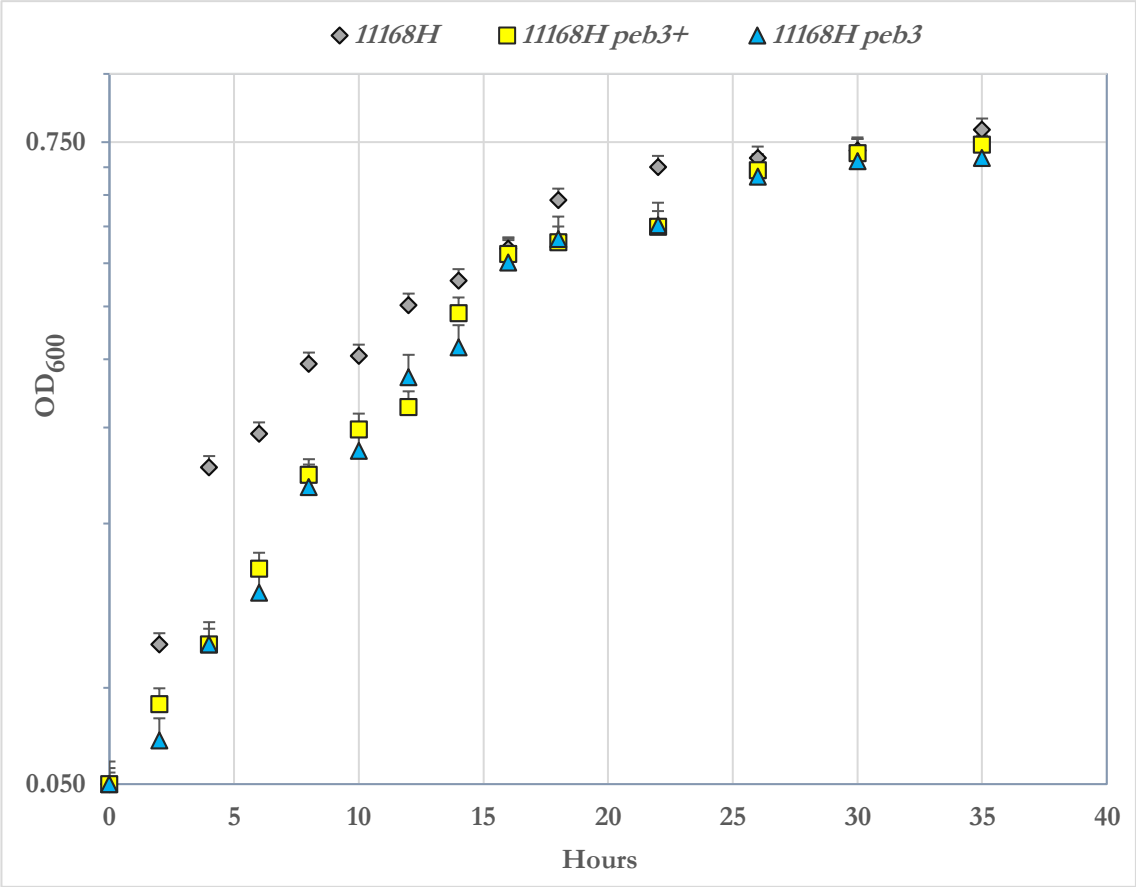


Figure 4.2 Growth of 11168H, 11168H *peb3* and 11168H *peb3*$^{+}$ in MH broth. Growth curves of *C. jejuni* NCTC 11168H (grey diamonds), 11168H *peb3* (blue triangles) and 11168H *peb3*$^{+}$ (yellow squares). Cultures were grown overnight and inoculated into fresh pre-warmed MH broth and growth was monitored by measuring the OD_{600}. Error bars represent standard error where n=3.

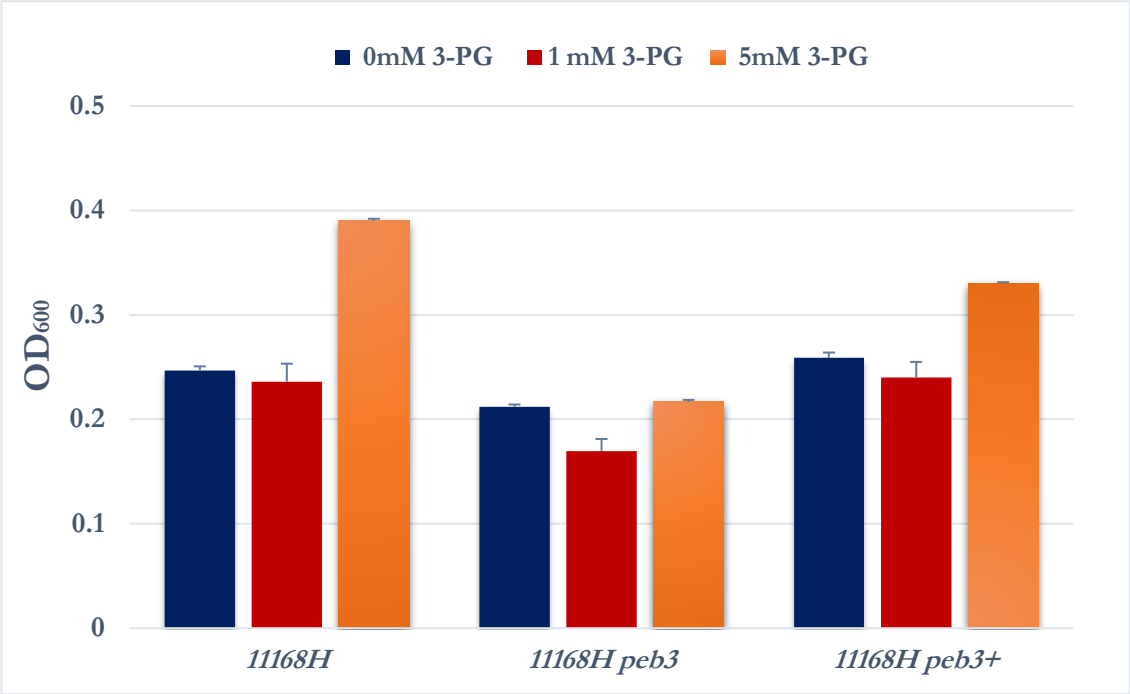


Figure 4.3 PEB3 is required for 3-PG dependent growth promotion of NCTC 11168H. All strains were grown in MEMα media supplemented with 0, 1 and 5 mM 3-PG and for 24 hours at 42 ºC in triplicate. Error bars represent standard error where n=3.

4.5 The gene adjacent to *peb3* encodes a putative 3-PG transporter

The enhanced growth of 11168H in MEMα media supplemented with 3-PG was dependent on PEB3, perhaps due to PEB3 providing 3-PG to a membrane transport protein (Min et al., 2009). The location of *peb3* in the NCTC 11168H genome (Accession: AL111168.1 - Parkhill et al., 2000) was examined using the Artemis genome browser (Rutherford et al., 2000). This identified a *glpT* gene, encoding a putative phosphoglycerate transporter, adjacent to *peb3*. However, this gene is annotated as a pseudogene, with two frameshift mutations resulting in three overlapping ORFs (Figure 4.4 A).

4.6 Verification of the genetic structures of the *glpT* gene from 11168H

The *glpT* gene structure was confirmed in our laboratory strain of 11168H using PCR and sequencing. Genomic DNA was obtained from 11168H, and *glpT* was amplified with primers 1203 and 1204 (Figure 4.4 A) (Appendix Table S1). The amplified product of 1.3 kbp (Figure 4.4 B) was sequenced with primers 1204, 1194, 1195 and 1196 (Appendix Table S1). The sequence data obtained was identical to genome sequence data confirming the presence of two in-frame stop codons.

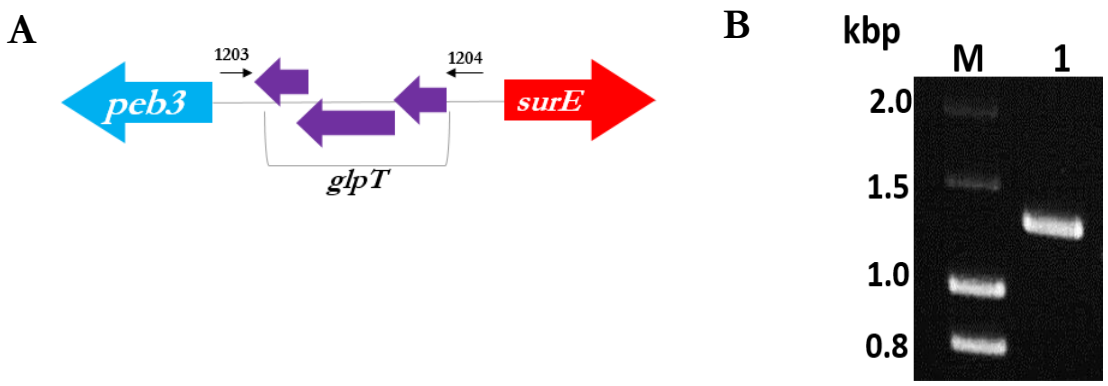


Figure 4.4 Amplification of *glpT* from *C. jejuni* 11168H. A) The *glpT* gene from NCTC 11168H was amplified with primers 1203/1204. **B)** The amplified product was analysed on a 1 % agarose gel and the predicted band size of 1.3 kbp (lane 1) observed.

4.7 Insertional knockout of the 11168H *glpT* gene

In MEMα medium, PEB3 is involved in the 3-PG dependent growth promotion of 11168H (Figure 4.3). I proposed that GlpT is involved in the transport of 3-PG. This hypothesis was examined by generating an 11168H *glpT* insertional knockout. Primers 1190/1191 (Appendix Table S1) were used to amplify *glpT* along with 400 bp of the flanking sequence from the neighbouring gene *peb3*. The 1.7 kbp fragment was ligated into pGEM-T Easy (Promega) vector to produce p*glpT* (Figure 4.5 B-Lane 1). This was modified by site-directed mutagenesis to introduce a BamHI site, within *glpT* with primers 1188/1189, to create p*glpT*B. The p*glpT*B plasmid was digested with BamHI, and a BamHI pre-digested kanamycin cassette ligated into this site to produce p*glpT*B::kan (Appendix Table S2). The presence of the kanamycin resistance cassette was verified by PCR with the primers 1190/1191, which produced a PCR product of 3.2 kbp (Figure 4.5 A & B-Lane 2). The primers 513, 535, 1190 and 1911 (Appendix Table S1) were used to confirm that the kanamycin resistance cassette was in the same transcriptional orientation as *glpT* in p*glpT*B::kan (Figure 4.5 A & B). This plasmid was electroporated (See Chapter 2) into competent cells of *C. jejuni* 11168H resulting in strain 11168H *glpT*. Confirmation that homologous recombination was successful, and that *glpT* onto the chromosome was disrupted by a kanamycin cassette was verified by colony PCR (data not shown).

A

B

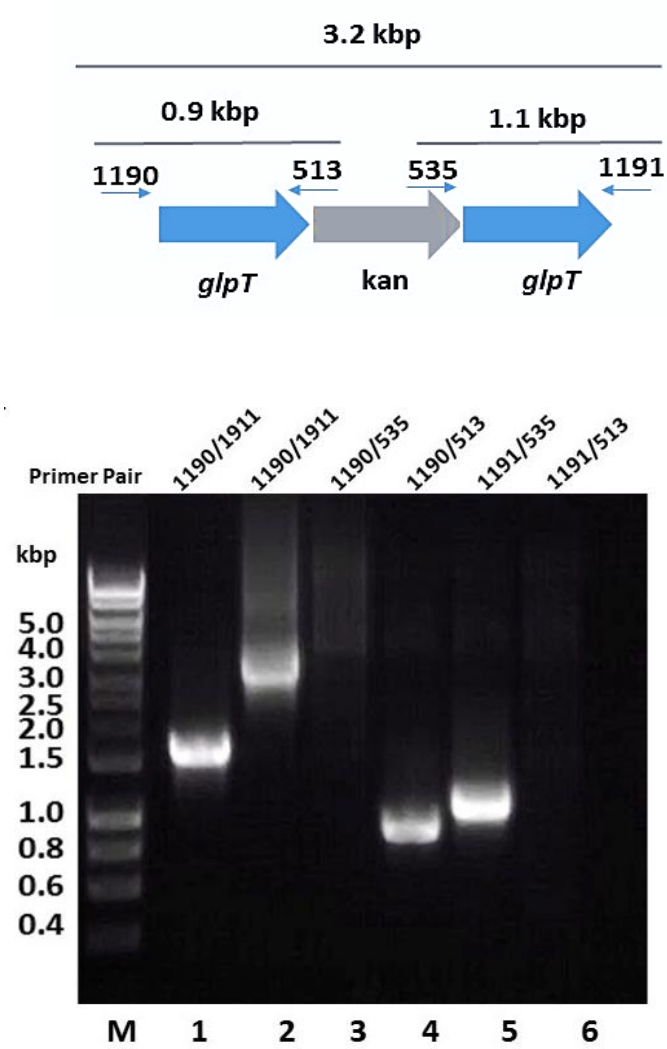


Figure 4.5 PCR verification of p*glpT*B::kan. A) Schematic diagram of *glpT* (blue) with the inserted kanamycin resistance cassette (grey). The orientation of the kanamycin resistance cassette, with respect to the *glpT* gene, was verified using primers 1190, 1191, 513 and 535 (Appendix Table 1). **B) PCR products of p*glpT*B::kan.** Amplified products were analysed on a 1 % agarose gel and band sizes indicated by a DNA Hyper ladder 1 marker (M). The original p*glpT* plasmid without a kanamycin cassette, containing only *glpT* was used as control. The *glpT* gene was amplified with primers 1190/1191 producing a product of 1.7 kbp (lane 1). The p*glpT*B::kan, containing the ligated kanamycin cassette (1.5 bp) into the BamHI site of *glpT* was amplified with 1190/1191 primers revealed a band of 3.2 kbp (lane 2). Further verification of the orientation of the kanamycin cassette was performed with a combination of 1190/1191 and 513/535. The primers 1190/535 yielded no PCR product (lane 3). Amplification with primers 1190/513 and 1191/535 produced a 0.9 kbp (lane 4) and 1.1 kbp (lane 5), respectively. The primer pair 1191/513 yielded no PCR product (lane 6).

4.8 The reintroduction of *glpT* onto the chromosome of the 11168H *glpT*::kan mutant to generate a genetic complement

The *glpT* gene was reintroduced onto the chromosome of the 11168H *glpT* mutant strain using plasmid *pCJ0223*::*cat*, a pUC18 derivative, containing a region of the *C. jejuni* pseudogene cj0223, and a chloramphenicol cassette that has an associated promoter but lacks a terminator (Figure 4.6 A) (Wang & Taylor, 1990). The 11168H *glpT* gene and a 70 base upstream region were amplified using primers 1228/1229 (Appendix Table S1). The primer 1228 contained a PciI restriction site, and primer 1229 contained a SpeI restriction site along with a His_{10} tag. The PCR product was digested with restriction enzymes PciI and SpeI and cloned into *pCJ0223*::*cat* plasmid downstream of the cat cassette to give *pCJ0223*::*cat glpT* (Figure 4.6 B). The arrangement of *glpT* in *pCJ0223*::*cat glpT* plasmid was verified by PCR (Figure 4.7 A & B). The sequence of *glpT* in *pCJ0223*::*cat glpT* was verified by sequencing with primers 311, 66, 1192 and 1193 (Appendix Table S1). The plasmid *pCJ0223*::*cat glpT* was then electroporated (see Chapter 2) into 11168H *glpT* competent cells. Following selection on plates containing chloramphenicol, colonies were subcultured and analysed by PCR using primers 1063, 1064, 1192 and 1193 (Appendix Table S1) to verify successful homologous recombination onto the chromosome (Figure 4.8 A & B).

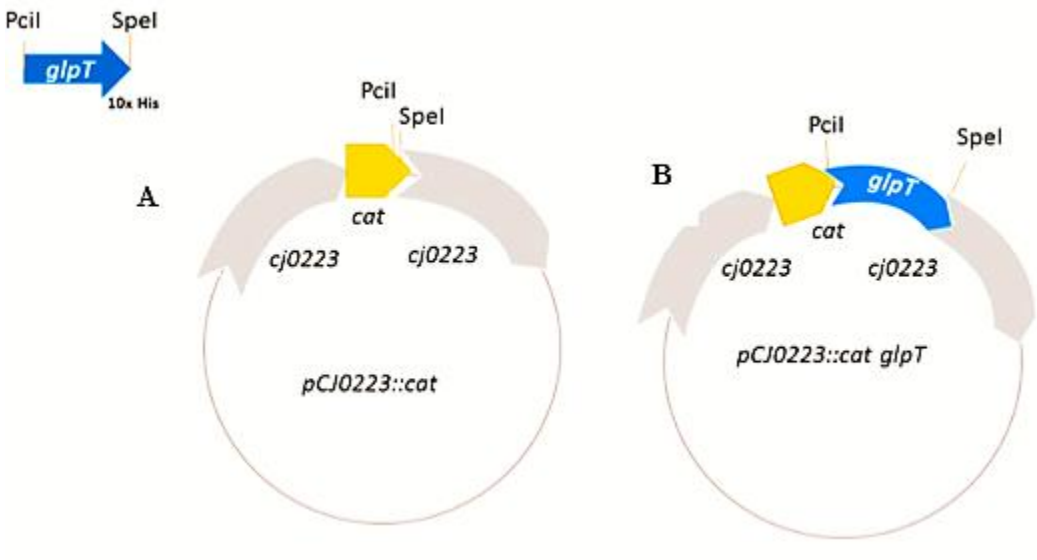


Figure 4.6 Schematic diagram illustrating the cloning strategy used to generate the *pCJ0223*::*cat glpT* plasmid. A) A plasmid map of the plasmid *pCJ0223*::*cat*. The cj0223 gene (grey) flanks the inserted cat resistance cassette (yellow). **B) The *glpT* gene ligated into the pCJ0223 plasmid.** The *glpT* gene (blue) was ligated downstream of the cat cassette (yellow) in the pCJ0223 plasmid, using restriction enzymes PciI and SpeI, to create the *pCJ0223*::*cat glpT* plasmid. The schematic is not drawn to scale.

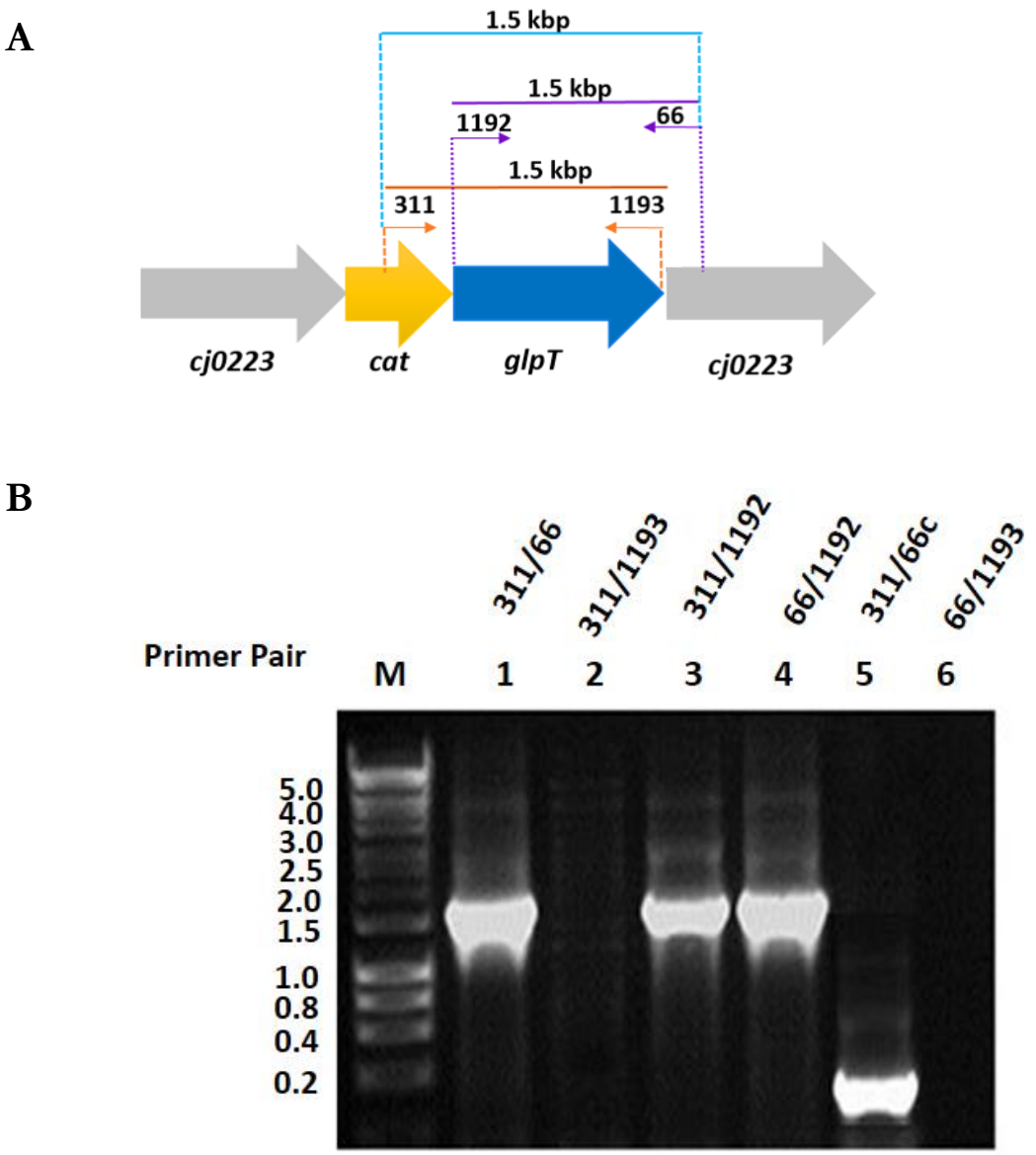


Figure 4.7 PCR verification of *pCJ0223*::*cat glpT*. A) Schematic diagram of *glpT* (blue) in complementation vector *pCJ0223*::*cat*. The pseudogene region cj0223 showing the cassette (yellow) and downstream *glpT* (coloured blue). The orientation of the cat resistance cassette, with respect to the *glpT* gene, was verified using specific primers 311, 66, 1192 and 1193 (Appendix Table 1), as indicated. **B) PCR products from *pCJ0223*::*cat glpT* plasmid.** Amplified products were analysed on a 1 % agarose gel and band sizes indicated by a DNA Hyper ladder 1 marker (M). The *glpT* gene inserted into the *pCJ0223*::*cat* plasmid was amplified with primers 311/66 yielding a PCR product of 1.5 kbp (lane 1). The *glpT* gene amplified with the primer pair 311/1192 showed no PCR band (lane 2). Further verification of the orientation of *glpT* gene was verified with primer pairs, 311/1193 and 66/1192, which produced bands of 1.5 kbp (lane 3/4). Lane 5, shows the amplification of the region between 311 and 66 without *glpT* and produces a PCR product of 0.18 bp. This can be compared with Lane 1, which contains the *glpT* gene in this region. The primer pair 66/1193 yielded no PCR product (lane 6).

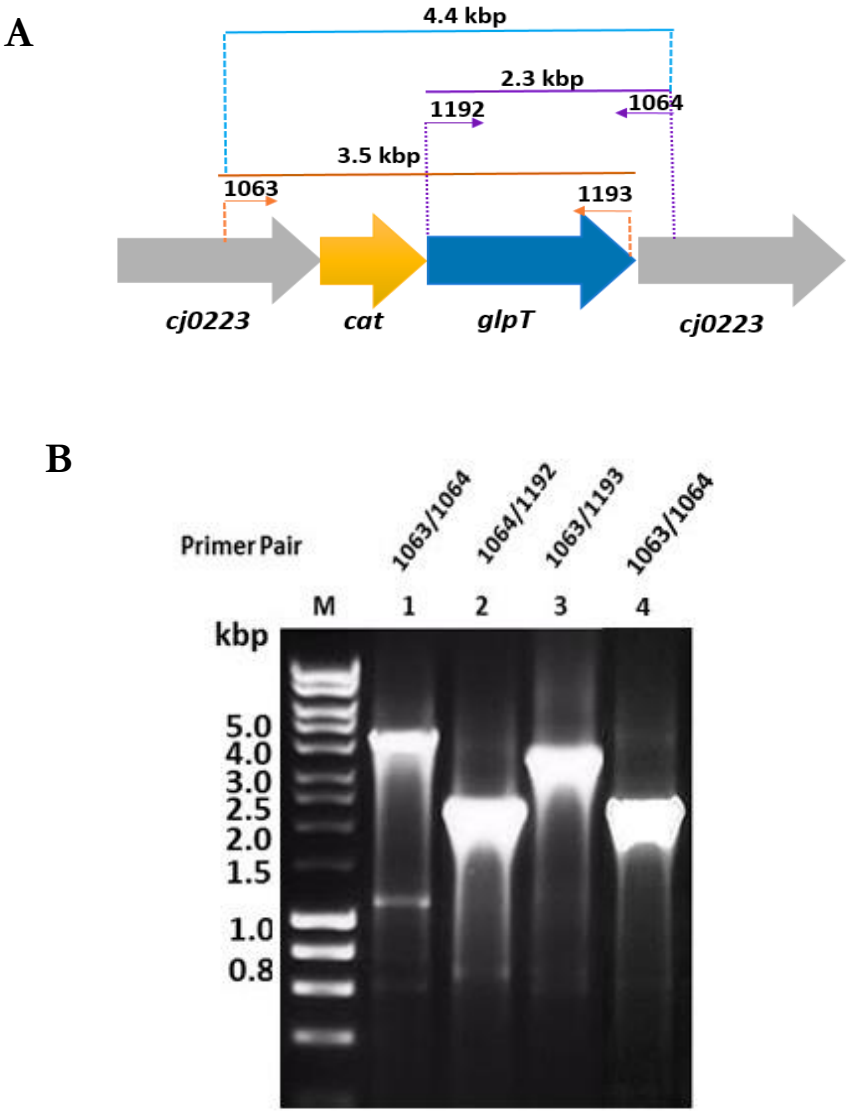


Figure 4.8 PCR verification of 11168H *glpT*⁺ strain. A) Schematic of the predicted structure of the 11168H *glpT*⁺ chromosomal structure in the Cj0223 region following electroporation and selection on chloramphenicol. The pseudogene region cj0223 showing the cassette (yellow) and an upstream *glpT* (coloured blue). The primer combinations are shown 1063/1064 and 1192/1193 along with predicted sizes. **B) PCR products from 11168H *glpT*⁺.** Amplified products were analysed on a 1 % agarose gel and band sizes indicated by a DNA Hyper ladder 1 marker (M). Primers 1063/1064 producing a PCR product of 4.4 kbp (lane 1). Primer pairs 1064/1192 and 1063/1193 produced PCR products of 2.3 (lane 2) and 3.5 kbp (lane 3), respectively. In Lane 4, using gDNA of 11168H, the cj0223 pseudogene (empty) was the amplified with primers 1063/1064, producing a PCR product of 2.2 kbp.

4.9 Growth characteristics of *C. jejuni* 11168H, 11168H *glpT* and 11168H *glpT*$^+$

Growth curves of 11168H, 11168H *glpT* and 11168H *glpT*$^+$ were generated to determine if mutation of *glpT* affected growth in complex media (Figure 4.9). Cultures were standardised to a starting OD_{600} of 0.05 in MH broth without antibiotics and growth was monitored by periodically measuring OD_{600} (see Chapter 2). 11168H *glpT* initially grew at a slower rate, but by approximately 16 hours, OD_{600} values were similar to 11168H. All cultures grew to a maximal OD_{600} of about 0.7 by ~24 hrs. 11168H *glpT*$^+$ grew as well as 11168H (Figure 4.9).

4.10 3-PG dependent growth involves GlpT

The minimal media MEMα used in previous growth assays (Figure 4.3) was composed of 49 components: 21 amino acids, 11 vitamins, six inorganic salts, ribonucleotides, deoxyribonucleotides, and carbon sources. Recently, a much simplified minimal media for *C. jejuni* was described (Alazzam et al., 2011). This minimal medium (MCLMAN) contains only twelve components: four amino acids, one vitamin and seven inorganic salts. MCLMAN medium can support the growth of 11168 (Parkhill et al., 2000) and other *C. jejuni* strains to a similar OD_{600} as MEMα (Alazzam et al., 2011). Previous metabolic studies showed that lactate is used by *C. jejuni* and converted into pyruvate (Alazzam et al., 2011; Thomas et al., 2011; Stahl et al., 2012). Additionally, *C. jejuni* uses pyruvate as a carbon source, however, a transporter for this compound has not been identified (Velayudhan & Kelly, 2002; Stahl et al., 2012). As previously discussed, the ligand for PEB3 is 3-PG a compound that potentially could be fed into the lower part of the glycolytic cycle and hence be used as an energy source by *C. jejuni*.

Growth assays were repeated using MCLMAN to determine compounds capable of supporting growth of 11168H, 11168H *peb3,* and 11168H *glpT* mutant strains along with their corresponding genetically complemented strains 11168H *peb3*$^+$ and 11168H *glpT*$^+$. Cultures grown overnight in MH broth were standardised to a starting OD_{600} 0.05 in fresh warm MH broth. Cells were grown to mid-log phase (OD_{600} 0.2-0.4) (Figure 4.2 & Figure

4.9), harvested by centrifugation, and re-suspended in warm MCLMAN media supplemented with either 10 mM pyruvate, lactate, 3-PG or G3P to a starting OD_{600} of 0.05. The WT 11168H strain grew well in the presence of pyruvate, lactate and 3-PG but not with G3P (Figure 4.10). 11168H *peb3* in the presence of 3-PG grew well in contrast with previous experiments with MEMα (Figure 4.3). 11168H *glpT* grew well on pyruvate, but not with 3-PG nor G3P (Figure 4.10). The observed difference between 11168H and 11168H *glpT* was statistically significant ($p<0.01$). 11168H *glpT*$^+$ grew well in the presence of 3-PG, pyruvate, lactate and some growth was observed for G3P (Figure 4.10). Restoration of the *glpT* gene by complementation rescued this phenotype and the observed difference was statistically significant ($p<0.01$) (Figure 4.10). These data demonstrate that 11168H 3-PG dependent growth in MCLMAN requires GlpT but not PEB3 and this is despite two in-frame stop codons in *glpT*.

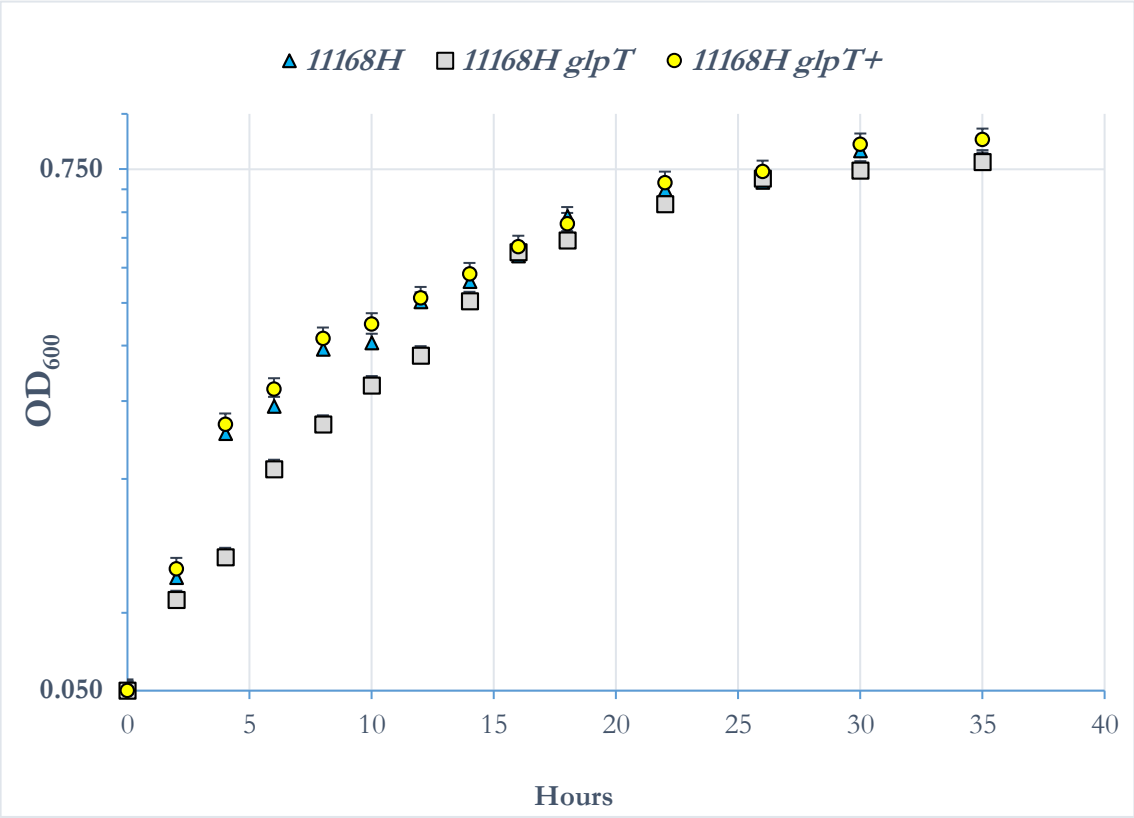


Figure 4.9 Growth of 11168H, 11168H *glpT* and 11168H *glpT*$^{+}$ in MH Broth. Comparative growth curve of 11168H (blue triangles), 11168H *glpT* (grey squares) and 11168H *glpT*$^{+}$ (yellow circles). Cultures were grown overnight and inoculated in to fresh pre-warmed MH broth and growth monitored by measuring the OD_{600}. Error bars represent standard error where n=3.

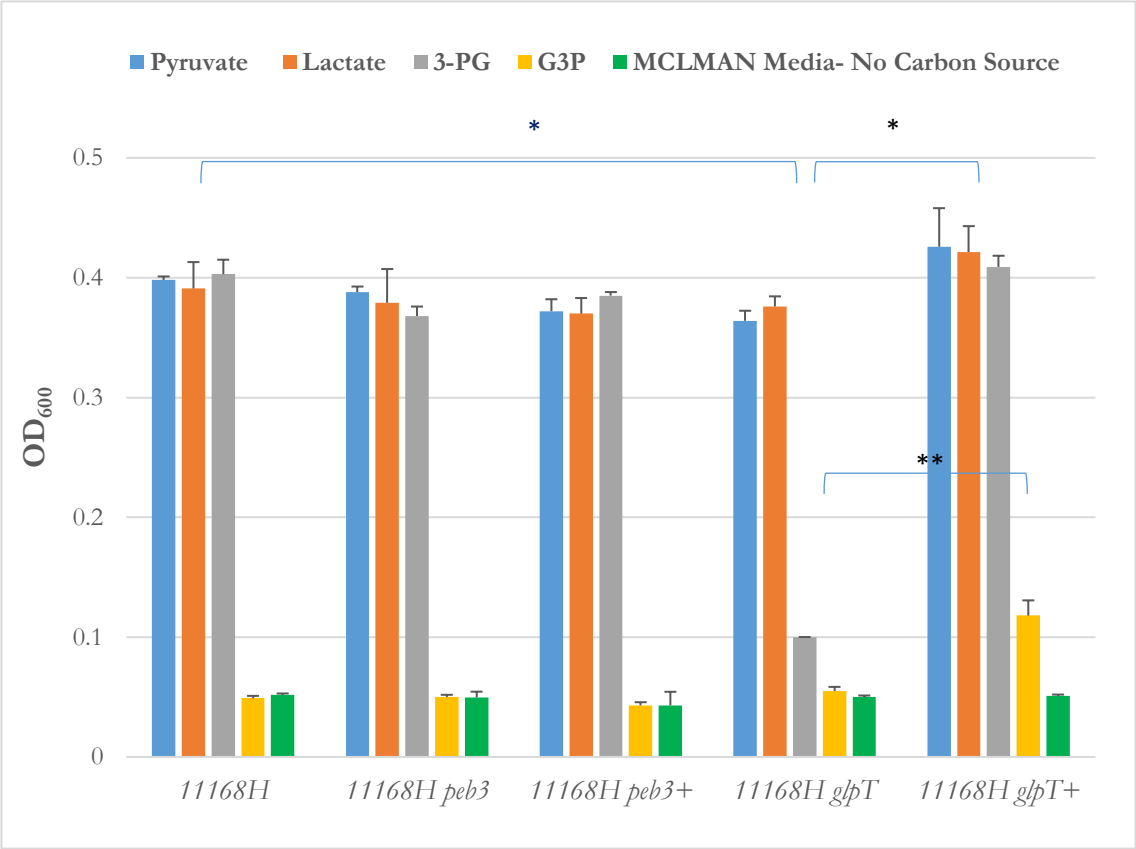


Figure 4.10 11168H GlpT is required for 3-PG dependent growth. All strains were grown in MCLMAN minimal media and supplemented with 10 mM 3-PG, G3P, pyruvate or lactate, as indicated. Cultures with a starting OD_{600} of 0.05 were incubated for 24 hours at 42 °C in triplicate. 11168H *glpT* displayed a reduced ability to use 3-PG compared to WT 11168H and 11168H *glpT*$^{+}$. Error bars represent standard error n=3. * denotes a statistical difference using a paired sample student t test ($P \leq 0.01$) between 11168H and 11168H *glpT* and between 11168H and 11168H *glpT*$^{+}$ in 3-PG dependent growth. **denotes a statistical difference ($P \leq 0.05$) between 11168H *glpT* and 11168H *glpT*$^{+}$ in G3P growth.

4.11 Impaired function of *glpT* confers fosfomycin resistance

The antibiotic fosfomycin inhibits the first step in peptidoglycan biosynthesis by acting as a PEP analogue and binding to MurA, an enzyme essential for peptidoglycan biosynthesis (Castañeda-García et al., 2013). In *E. coli* fosfomycin enters cells via GlpT and UhpT transporters and isolates that have defective GlpT and UhpT transporters are resistant to fosfomycin (Kadner & Shattuck-Eidens, 1983; Castañeda-García et al., 2013). 11168H, 11168H *peb3* and 11168H *glpT* strains were tested for their susceptibility to fosfomycin using E-Test strips (see Chapter 2). The MIC of fosfomycin for 11168H was 24 μg/ml, for 11168H *glpT* 64 μg/ml, and 24 μg/ml for 11168H *glpT*$^{+}$. The MIC of fosfomycin for 11168H *peb3* and the 11168H *peb3*$^{+}$ was 24 and 12 μg/ml, respectively (Table 4.1). To confirm these data a spot plate method for determining fosfomycin MICs was used (see Chapter 2) (Jeong et al., 2001; Ribardo et al., 2010). Using this method, the MIC of fosfomycin for 11168H was 32 μg/ml for 11168H *glpT* 128 μg/ml and for 11168H *glpT*$^{+}$ 32 μg/ml. The MIC of fosfomycin for both 11168H *peb3* and 11168H *peb3*$^{+}$ was 24 μg/ml (Table 4.1).

Strains	Fosfomycin MIC (μg/ml)	
	E-Test	Spot Plate
11168H	24	32
11168H *peb3*	24	24
11168H *glpT*	64	128
11168H *peb3*$^{+}$	12	24
11168H *glpT*$^{+}$	24	32

Table 4.1 *C. jejuni* NCTC 11168H with defective *glpT* gene has increased fosfomycin resistance. For the E-Test, cells were grown to mid-log phase (Figure 4.2 & 4.9) and inoculated onto MH plates. The zone of inhibition was read, and the MIC noted. For the spot plate method, cells were standardised to an OD of 1.0, diluted by tenfold dilutions and plated in triplicate onto plates containing 0, 8, 12, 24, 32, 48, 64, 96 and 128 μg/ml fosfomycin. The MIC was determined by examining the lowest fosfomycin concentration resulting in a 10-fold decrease in CFU compared to growth on MH agar without fosfomycin. All plates were incubated at 42 °C for 48 hours.

4.12 Functionality of NCTC 11168H GlpT suggested the "pseudogene" may encode a functional protein

Given that interrupted 11168H *glpT* produced a phenotype for both 3-PG dependent growth and decreased sensitivity to fosfomycin, I sought to investigate whether a full-length protein was produced from the interrupted gene. This was done by examining GlpT protein production in the 11168H *glpT* $^{+}$that expresses a 10X-histidine (His) tag at the GlpT C-terminus. Experiments were performed in both *E. coli* and *C. jejuni*. A whole-cell lysate of *E. coli* XL1-Blue competent cells containing the *pCJ0223*::*cat glpT* plasmid was separated by SDS-PAGE, blotted onto nitrocellulose and probed with anti-His antibody. The predicted band size of ~51 kDa was not detected (data not shown). A whole cell lysate of *C. jejuni* 11168H *glpT* $^{+}$ was separated by SDS-PAGE, and blotted onto nitrocellulose probed with anti-His antibody. The predicted band of ~51 kDa was not observed (data not shown). A membrane preparation was purified using Ni-NTA beads. Purified protein was probed with an anti-His antibody but again no bands of the expected size of ~51 kDa were detected. In summary, the production of intact GlpT expressed from the upstream chloramphenicol promoter was not observed in *E. coli* nor *C. jejuni*.

4.13 Attempt to increase expression level of 11168H GlpT in *C. jejuni*

In order to increase expression, the 11168H *glpT* gene was cloned into a *pCJ0223* plasmid containing one of two strong promoters, *porA* or *flaA* (Jervis et al., 2015). Genomic DNA was obtained from 11168H, and *glpT* amplified with primers, *glpT*$_{NoTag}$F/*glpT*C-Term$_{His}$R or *glpT*N-Term$_{HA}$F/*glpT*C-Term$_{His}$R, each containing an NdeI restriction site. The amplified PCR product was ligated into pGEM-T Easy (Promega) vector to create p*glpT*$_{His}$ and p$_{HA}$*glpT*$_{His}$ and sequenced with primers 80/81. The insert was digested with restriction enzyme NdeI and cloned into an NdeI pre-digested *pCJ0223porA* or *pCJ022flaA* vector. This produced *pCJporAglpT*$_{His}$, *pCJporA*$_{HA}$*glpT*$_{His}$ and *pCJflaA*$_{HA}$*glpT*$_{His}$ as verified by PCR with primers 65, 546, 1192 and 1193 (data not shown) (Appendix Table S1).

Plasmids *pCJporAglpT*$_{His}$, *pCJporA*$_{HA}$*glpT*$_{His}$ and *pCJflaA*$_{HA}$*glpT*$_{His}$ were electroporated into 11168H *glpT* competent cells resulting in Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$. Following selection on plates containing chloramphenicol, colonies were subcultured and analysed by PCR using a combination of primers 1063/1064 and 1192/1193 (Appendix Table S1) to verify homologous recombination on the chromosome (Figure 4.11 A & B).

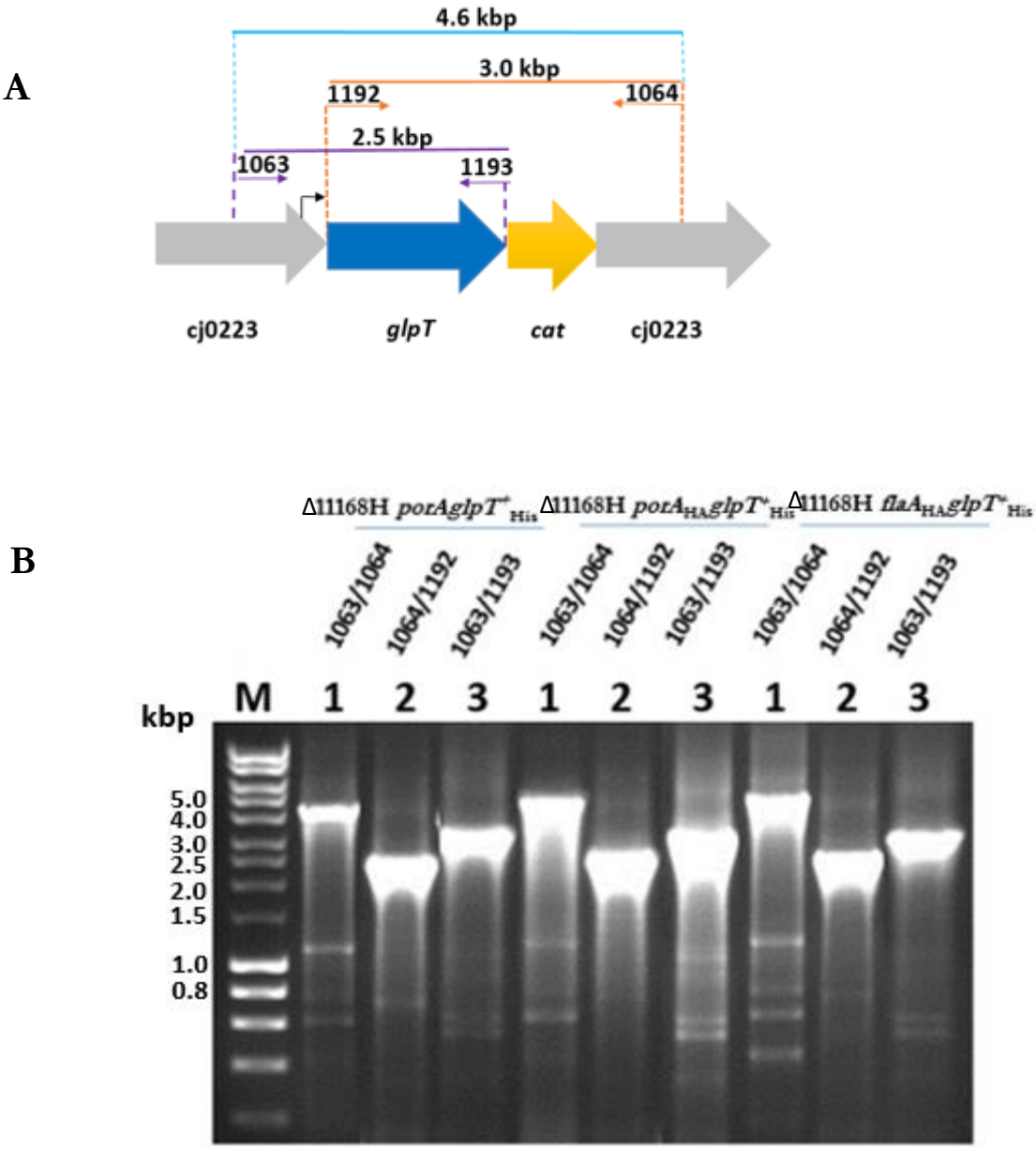


Figure 4.11 PCR verification of Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$. A) A schematic diagram of predicted structure of the 11168H *glpT* chromosomal structure in the Cj0223 region following electroporation and selection on chloramphenicol. The *glpT* is upstream of the cat cassette (yellow) and expression is driven by the *porA* or *flaA* promoter (black arrow) in plasmid *pcj0223*. The primer combinations are shown 1063/1064 and 1192/1193 along with predicted sizes. **B) PCR products from Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$.** Amplified products were analysed on a 1 % agarose gel and band sizes indicated by a DNA Hyper ladder 1 marker (M). Primers 1063/1064, which flank the pseudogene, produced a PCR product of 4.6 kbp (Lane 1). Primers 1064/1192 (lane 2) and 1063/1193 (lane 3) produced PCR products of 3.0 and 2.5 kbp respectively.

4.14 Phenotypic analysis of *C. jejuni* strains expressing *glpT* tagged inserts

The Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ strains were tested for 3-PG dependent growth and fosfomycin sensitivity. Cultures grown overnight in MH broth were standardised to a starting OD of 0.05 in fresh MH broth. Cells were grown to mid-log phase (Figure 4.2 & 4.9), harvested by centrifugation and re-suspended in warm MCLMAN media supplemented with 10 mM pyruvate, lactate, 3-PG or G3P to a starting OD_{600} of 0.05. The WT 11168H strain grew well in the presence of pyruvate, lactate and 3-PG but not with G3P as seen in previous assays (Figure 4.10 & Figure 4.12). 11168H *glpT* grew well on pyruvate, but not with 3-PG nor G3P (Figure 4.10 & Figure 4.12). All three strains, Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$, grew well in the presence of pyruvate, lactate and 3-PG and there was some limited growth in the presence of G3P with *glpT* expressed from *porA* or *flaA* promoters (Figure 4.12). The Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ strains were tested for susceptibility to fosfomycin using E-Test strips (Table 4.2) (see Chapter 2). The MIC of fosfomycin for WT 11168H, Δ11168H *porAglpT*$^{+}_{His}$ and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ was 24 μg/ml and 24/32 μg/ml for Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ (Table 4.2). To confirm these data, a spot plate method for determining the fosfomycin MIC was used (see Chapter 2) (Jeong et al., 2001; Ribardo et al., 2010). The MIC of fosfomycin for all strains was 32 μg/ml (Table 4.2). These data confirm that genetic complementation of the 11168H *glpT* with the highly expressed tagged variants of *glpT* restores ability to grow with 3-PG and sensitivity to fosfomycin.

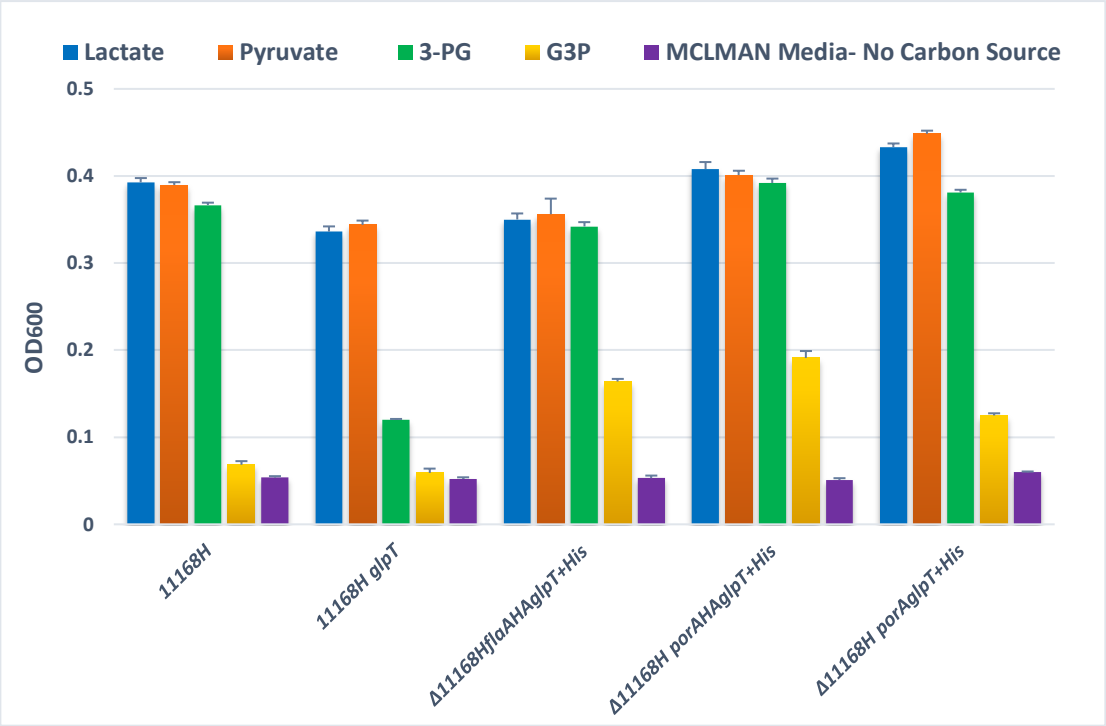


Figure 4.12 Effect of genetic complementation with tagged versions of 11168H *glpT* expressed from *porA* or *flaA* promotes growth with various carbon sources. All strains were grown in MCLMAN minimal media and supplemented with 10 mM 3-PG, G3P, pyruvate and lactate, as indicated. Cultures with a starting OD_{600} of 0.05 were incubated for 24 hours at 42 °C in triplicate. Error bars were based on standard error n=3.

Strain	Fosfomycin MIC (µg/ml) E-Test	Spot Plate
11168H WT	24	32
11168H *glpT*	64	128
Δ11168H *porAglpT*$^{+}_{His}$	24	32
Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$	24	32
Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$	24/32	32

Table 4.2 Effect of genetic complementation with tagged versions of 11168H *glpT* expressed from *porA* or *flaA* are sensitive to fosfomycin. For the E-Test, cells were grown to mid-log phase (Figure 4.2 & 4.9) and inoculated onto MH plates. The zone of inhibition was read and the MIC noted. For the spot plate method cells were standardised to an OD of 1.0, diluted by tenfold dilutions and plated in triplicate onto plates containing 0, 8, 24, 32, 48, 64, 96 and 128 µg/ml fosfomycin. The MIC was determined by examining the lowest fosfomycin concentration resulting in a 10-fold decrease in CFU compared to growth on MH agar without fosfomycin. All plates were incubated at 42 °C for 48 hours.

4.15 Production of GlpT in *C. jejuni* 11168H

Next, I sought to investigate whether a full-length protein was produced from the interrupted 11168H *glpT* gene, when expressed from the *flaA* and *porA* promoters. This was done by purifying the GlpT protein from strains where the expression of GlpT was driven by these promoters. These strains express either a 6x His-Tag at the GlpT C-terminus or both N-Terminus HA-Tag and C-Terminus 6x His-Tag. Experiments were performed in both *E. coli* and *C. jejuni*. Whole-cell lysates from three *E. coli* XL1-Blue competent strains containing the *pCJ0223porAglpT*$_{His}$, *pCJ0223porA*$_{HA}$*glpT*$_{His}$ and *pCJ0223flaA*$_{HA}$*glpT*$_{His}$ was separated by SDS-PAGE, western blotted, and probed with anti-His antibody. The predicted band of ~52 kDa was not detected from these three strains (Figure 4.13).

Whole-cell lysate from Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ were separated on SDS-PAGE (data not shown), blotted onto nitrocellulose, and probed with anti-HA or anti-His antibody (Figure 4.14 A & B). An anti-HA antibody blot showed bands consistent with the predicted Mr of ~52 kDa in lanes containing Δ 11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ whole cell lysates (Figure 4.14 A-lanes 1 and 2). The nitrocellulose probed with the anti-His antibody produced bands of 52 kDa in the lanes containing extracts from Δ11168H *porAglpT*$^{+}_{His,}$ Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ (Figure 4.14 B-lanes 1, 2 and 3). These data demonstrate production of an apparently full length GlpT from 11168H disrupted *glpT* gene.

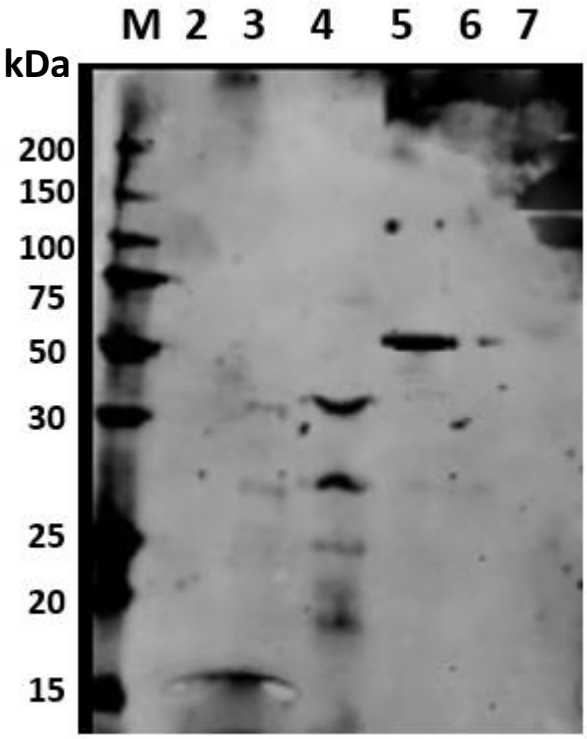


Figure 4.13 Immunoblotting of *E. coli* whole cell lysates confirms that GlpT is asbent. Immunoblot of whole cell lysates probed with anti-His antibody. Whole cell lysates were separated on a SDS-PAGE then blotted onto a nitrocellulose membrane andprobed with an anti-His antibody. Expected bands of approximately 52 kDa consistent with the GlpT Mr were not observed in *E. coli* XL1-Blue*porA*$_{HA}$*glpT*$^{+}_{His}$ (lane 2), *E. coli* XL1-Blue *flaA*$_{HA}$*glpT*$^{+}_{His}$ (lane 3), and *E. coli* XL1-Blue*porAglpT*$^{+}_{His}$ (lane 4). Controls were a PglB His-tagged protein approximately 53 kDa (lane 5), *E. coli* XL1-Blue*pCJ0223porA* (lane 6) and *E. coli* XL1-Blue containing the original complement plasmid *pCJ0223:cat glpT*.

4.16 GlpT production detected from purified membrane fractions

Membrane fractions from harvested cells of 11168H *glpT*$^{+}$, Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ were inoculated into Tris-containing buffer. The cells were disrupted via French press and subjected to two rounds of centrifugation resulting in a supernatant and pellet. The pellet containing the total membrane was retained and resuspended in Tris-containing buffer with Triton. After ultracentrifugation samples were purified with anti-HA immunoprecipitation and Ni-NTA beads in an attempt isolate a full-length GlpT. Purified proteins from Δ11168H *porAglpT*$^{+}_{His}$, Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ and Δ 11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ were separated on SDS-PAGE, blotted onto nitrocellulose, and probed with anti-His antibody. The SDS-PAGE bands of ~52 kDa in lanes containing Δ 11168H *porAglpT*$^{+}_{His}$, Δ 11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ and Δ 11168H *porA*$_{HA}$*glpT*$^{+}_{His}$, but not 11168H *glpT*$^{+}$ (Figure 4.15 A-Lanes 1, 2, 3 and 4). A western blot probed with anti-His antibody reacted with ~52 kDa bands containing Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ (Figure 4.15 B-Lanes 1 and 2).

Purified proteins from HA-immunoprecipitation beads were separated by SDS-PAGE, blotted onto nitrocellulose, and probed with anti-HA antibody. The SDS-PAGE showed a 52 kDa size band in lanes containing Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ (Figure 4.16 A-Lane 2 & 3). A western blot probed with anti-HA antibody reacted with these ~52 kDa bands (Figure 4.16 B-Lane 2 & 3). These data demonstrate production of an apparently full length GlpT protein which can be purified via HA and His-Tag purification. In an attempt to identify the proteins mass spectrometry was used, however, results were unsuccessful.

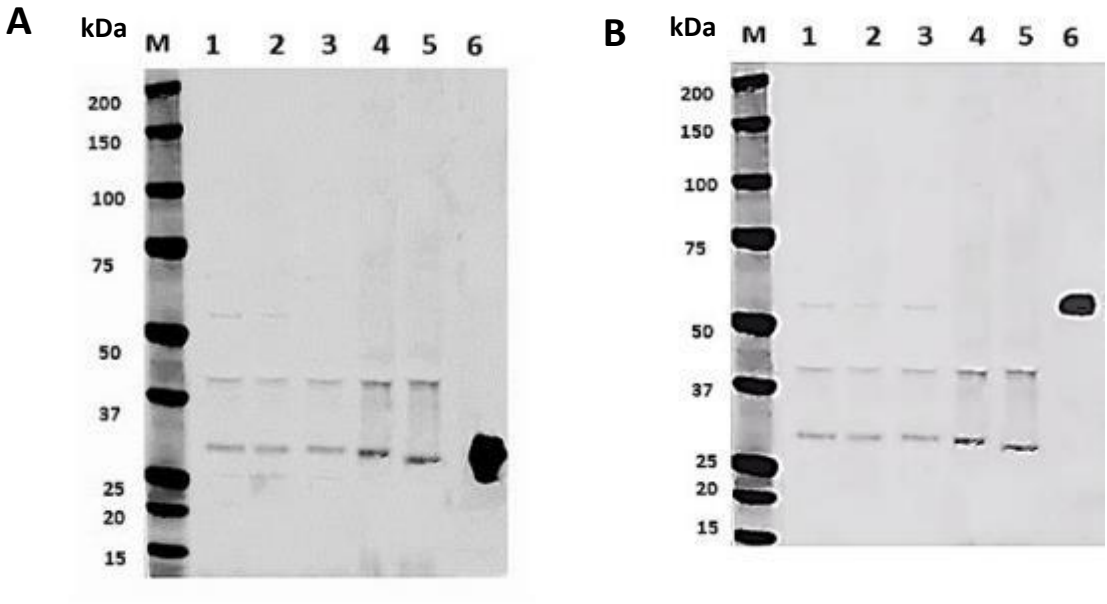


Figure 4.14 Immunoblotting of *C. jejuni* whole cell lysates reveals detectable bands the size of the GlpT Mr. A) Immunoblot of whole cell lysates probed with anti-HA antibody. Whole cell lysates were separated on a SDS-PAGE then blotted onto a nitrocellulose membrane probed with an anti-HA antibody. Detectable bands of approximately 52 kDa consistent with the GlpT Mr were observed in Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ (lane 1) and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ (lane 2). Bands of the size of the GlpT Mr were not detected in Δ11168H *porAglpT*$^{+}_{His}$ (lane 3), 11168H WT (lane 4) and 11168H *glpT* (lane 5). The control (lane 6) is a HA-Tagged W1130 protein that is approximately 30 kDa. **B) Immunoblot of whole cell lysates probed with anti-His antibody.** Whole cell lysates were separated on an SDS-PAGE then transferred to a nitrocellulose membrane probed with an anti-His antibody. Detectable bands of approximately 52 kDa consistent with the GlpT Mr were observed in Δ11168H *porAglpT*$^{+}_{His}$ (lane 1), Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ (lane 2) and in Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ lane (3). No bands of this size could be detected in, 11168H WT (lane 4) and 11168H *glpT* strain (lane 5). The control (lane 6) a PglB His-tagged protein is approximately 53 kDa.

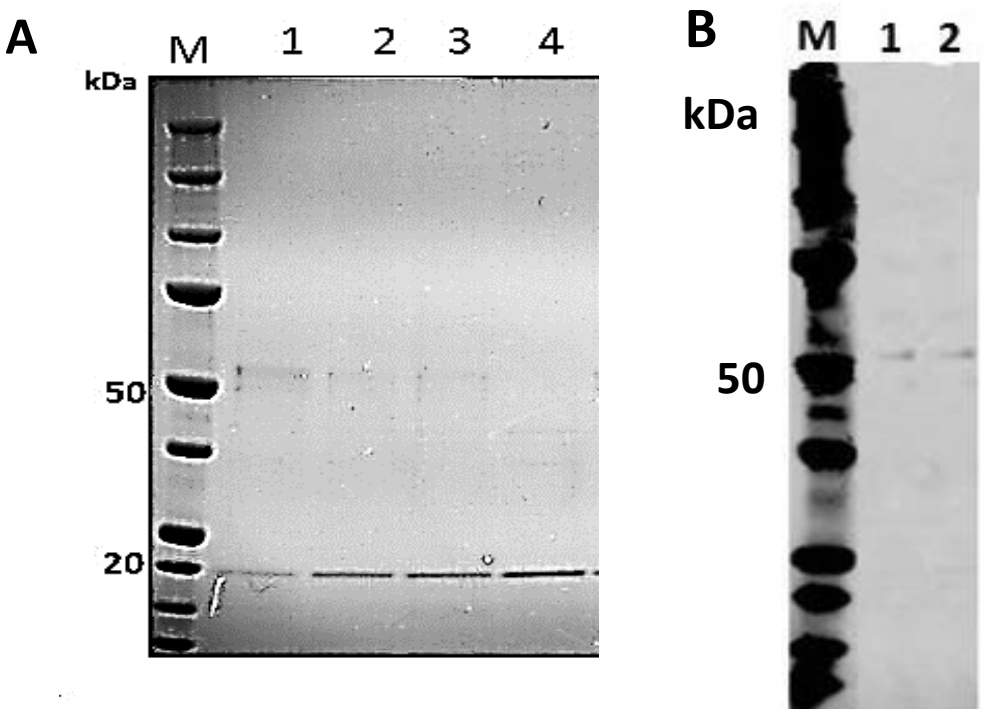


Figure 4.15 Purification of GlpT via Ni-NTA affinity chromatography. A) SDS-PAGE of *C. jejuni* GlpT purification via the His-Tag. A membrane protein preparations was purified using Ni-NTA beads. Following elution in 500 mM imidazole fractions were separated by SDS-PAGE. Bands of 51, 52 and 20 kDa were observed in Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ (lane 1), Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ (lane 2) and Δ11168H *porAglpT*$^{+}_{His}$ (lane 3) strains, but not in 11168H *glpT*$^{+}$ (lane 4). **B) Immunoblotting of purified His-Tagged GlpT protein.** Purified protein from (**A**) was blotted onto a nitrocellulose membrane and probed with anti-His antibody. Bands of approximately 52 kDa were present in both Δ11168H *porA*$_{HA}$*glpT*$^{+}_{His}$ (lane 1) and Δ11168H *flaA*$_{HA}$*glpT*$^{+}_{His}$ (lane 2).

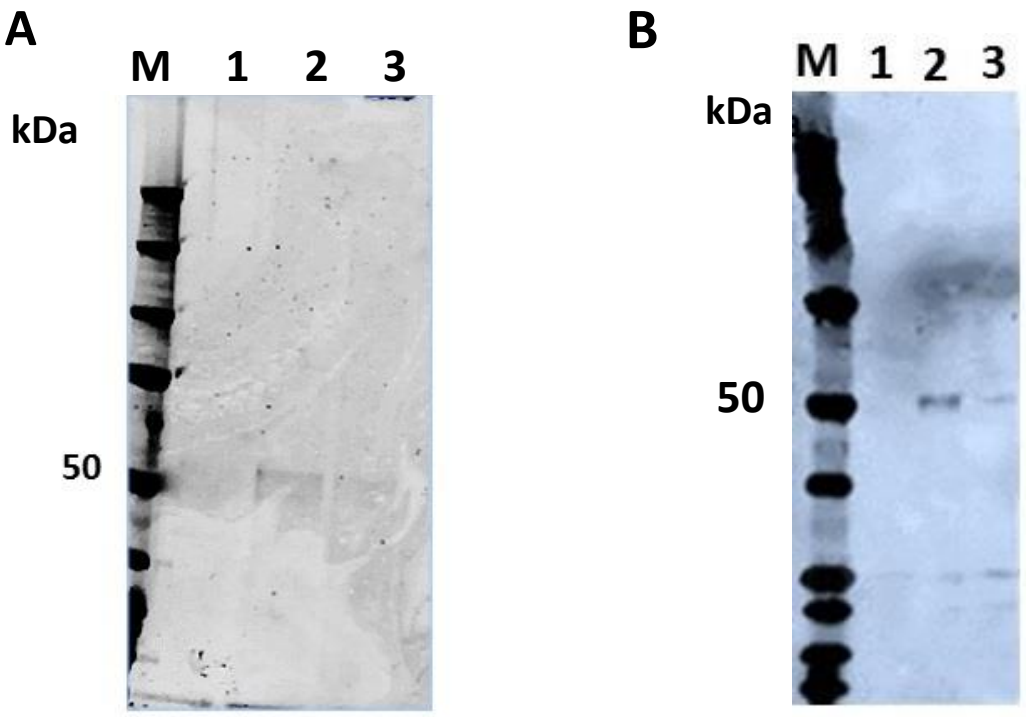


Figure 4.16 Purification of GlpT via HA affinity chromatography. A) SDS-PAGE of *C. jejuni* GlpT purification via the HA-Tag. HA-tagged protein from membrane fractions of strains, Δ11168H *porAglpT*$_{His}$, Δ11168H *porA*$_{HA}$*glpT*$_{His}$ and Δ11168H *flaA*$_{HA}$*glpT*$_{His}$ were purified using HA immunoprecipitation beads. Following a final elution in SDS-PAGE sample buffer fractions were separated by SDS-PAGE. Bands of 52 kDa were observed in Δ11168H *porA*$_{HA}$*glpT*$_{His}$ (lane 2) and Δ11168H *flaA*$_{HA}$*glpT*$_{His}$ strains (lane 3), but not Δ11168H *porAglpT*$_{His}$ (lane 1). **B) Immunoblotting of purified HA-Tagged GlpT protein.** Purified protein from (**A**) was blotted onto nitrocellulose membrane and probed with anti-HA antibody. Bands of approximately 52 kDa were present in both Δ11168H *porA*$_{HA}$*glpT*$_{His}$ (lane 2) and Δ11168H *flaA*$_{HA}$*glpT*$_{His}$ (lane 3) strains, but not Δ11168H *porAglpT*$_{His}$ (lane 1).

4.17 Discussion

PEB3 binds 3-PG, and may deliver it to a transporter protein (Min et al., 2009). Adjacent to the *peb3* gene is *glpT*, encoding a putative phosphoglycerate transporter. In 11168H, *glpT* is annotated as a pseudogene with three overlapping ORFs due to indels at specific locations within the gene. Despite this, evidence presented in this chapter indicates that GlpT, and not PEB3, is required for 3-PG dependent growth in MCLMAN and fosfomycin susceptibility.

Initially, I focused on reproducing experiments performed by Min et al., 2009. Growth assays in this study tested 3-PG dependent growth for a 11168H WT, knockout 11168H *peb3* and complement 11168H *peb3*$^{+}$strain in MEMα media. The knockout and complement strains were not included in the previous Min and co-workers study. Another difference in the experimental design was the incubation temperature for the 3-PG growth assays. In this study, a temperature of 42 °C was used instead of 37 °C. However, this did not have a significant influence on the utilisation of 3-PG as these experiments confirmed the requirement for PEB3 in 3-PG dependent growth in MEMα (Figure 4.3). In comparison to the Min et al. data, identical levels of growth were observed in the 11168H WT. This growth could be partially restored by the genetic complementation, but not to WT levels, perhaps due to the low-level expression of *peb3* from the *cat* gene promoter (Figure 4.1). As expected, the 11168 *peb3* could not utilise 3-PG.

To test the role of *glpT* in 3-PG dependent growth, I switched to a simpler medium known as MCLMAN (Alazzam et al., 2011). This medium was supplemented with various carbon sources, including pyruvate, lactate, G3P and 3-PG. In contrast to experiments in MEMα, PEB3 was not required for 3-PG dependent growth in MCLMAN. Significant growth was observed for the 11168H *peb3* mutant in the presence of 3-PG (Figure 4.10). However, this growth was not displayed in the MEMα growth experiments (Figure 4.3). One explanation may be that the concentration of 3-PG in the MCLMAN experiments was 10 mM compared

to only 5 mM in the MEMα-based growth experiments. This increased 3-PG concentration may overcome the necessity for a periplasmic-binding partner such as PEB3.

The carbon sources pyruvate and lactate promoted growth of 11168H WT, 11168H *glpT* and 11168H *peb3* and respective genetic complements. Pyruvate and lactate are known to be utilised by *C. jejuni* and produced similar OD_{600} values as previously reported (Figure 4.10) (Velayudhan & Kelly, 2002; Velayudhan et al., 2004; Alazzam et al., 2011; Thomas et al., 2011; Stahl et al., 2012). Growth of *C. jejuni* 11168H, but not 11168H *glpT* was promoted by 3-PG in MCLMAN media (Figure 4.10) and this phenotype was genetically complemented with 11168H *glpT*$^+$ (Figure 4.10). This demonstrated that in the MCLMAN system GlpT, and not PEB3, is required for 3-PG dependent growth promotion in spite of the fragmented *glpT* structure. This is consistent with GlpT from other bacterial species in which a periplasmic partner is not needed to gain substrate from the environment (Lemieux et al., 2004a; Lemieux et al., 2004b). Uptake of G3P by GlpT is well documented and might allow G3P to promote *C. jejuni* growth, but this was not the case as *C. jejuni* strains demonstrated limited growth with G3P in MCLMAN (Figure 4.10). However, G3P growth was improved in *glpT*$^+$ strains where *glpT* expression is driven by *flaA* or *porA* promoters (Figure 4.12). This suggests that when expressed at relatively high levels GlpT may allow some limited transport of G3P to promote growth.

In addition, to 3-PG dependent growth promotion, a second phenotype associated with GlpT transporters is susceptibility to the antibiotic, fosfomycin. In *C. jejuni*, inactivation of *glpT* led to fosfomycin resistance and this phenotype was restored by genetic complementation (Table 4.1). These data confirm requirement for the fragmented *glpT* gene in two independent phenotypes namely, 3-PG dependent growth and fosfomycin resistance (Figure 4.10 & Table 4.1).

One explanation for these data is that a full-length functional GlpT protein was produced from the fragmented *glpT* gene. To test this hypothesis, various versions of the 11168H *glpT* gene were constructed with N- and C- terminal His/HA tags. These were recombined onto the *C. jejuni* chromosome downstream of a strong *porA* or *flaA* promoter and I hoped to detect the production of intact GlpT. A protein of the apparently correct Mr of GlpT was detected on SDS-PAGE gels and blots in strains where the expression of *glpT* was driven by a *porA* or *flaA* promoter (Figure 4.14-4.16). It is hypothesised that, *C. jejuni* may produce a full-length GlpT protein from a pseudogene bearing frameshift mutations, possibly via ribosomal frameshifting (Farabaugh, 1996). Ribosomal frameshifting occurs during translation, when the ribosome skips (+1) or slip (-1) back one nucleotide aided by the presence of a 'slippery sequence' and a pseudoknot (Ketteler, 2012).

In order to investigate the process whereby a full-length GlpT protein was produced. I hoped to use mass spectrometry analysis of these purified GlpT preparations. This would enable accurate intact mass determination and sequencing around the frame-shifting regions. These data would help develop hypotheses as to how a full-length GlpT could be produced, namely what mechanism(s) might be employed. Unfortunately, I was unable to obtain any meaningful MS data and future work should focus on this area. Overall the data obtained in this chapter suggested that an apparent pseudogene can produce a functional protein in *C. jejuni*. The work in the next chapter describes in detail analysis of *glpT* gene variation in many *C. jejuni* strains.

CHAPTER 5

In silico characterisation of variation in the *C. jejuni peb3/glpT* locus

5.1 Introduction

The genome sequence of *C. jejuni* NCTC 11168H was published in 2000 (Parkhill et al., 2000), with further *Campylobacter* genome sequences published to identify genes associated with disease severity, metabolism and host specificity (Fouts et al., 2005; Hofreuter et al., 2006; Pearson et al., 2007; Poly et al., 2007). Chapter 4, highlighted the presence of a *glpT* gene adjacent to *peb3* and the functional characterisation of both gene products in *C. jejuni* 11168H. It was shown that NCTC 11168H *glpT,* a pseudogene composed of three overlapping ORFs due to indels, was required for 3-PG dependent growth. Lack of 6-phosphofructokinase in *C. jejuni* means the glycolytic pathway for metabolising glucose is not functional, but enzymes "below" this point are present. Thus, the usage of alternate carbon sources like 3-PG, that could feed into this lower portion of the glycolytic pathway and may be important in certain scenarios e.g. in intracellular environments.

In this study, the *peb3/glpT* locus was compared among *C. jejuni*, *C. coli*, *C. lari* and *C. upsaliensis* genome sequences. Initially, the overall arrangement of the *peb3/glpT* locus was examined and subsequently the detailed structure of *glpT* sequence variation characterised. Strains could be categorised into classes based on variation in *glpT* sequence and structure. These different classes of *peb3/glpT* locus and *glpT* structure were placed in their evolutionary context through MLST.

5.2 *Campylobacter* species exhibit variability in the *peb3/glpT* locus

In *C. jejuni* NCTC 11168H (Accession: AL111168.1 - Parkhill et al., 2001) the *peb3*-containing region consists of the following four genes: *lpxB*, encoding a lipid-A-disaccharide synthetase involved in the condensation of UDP-2,3-diacylglucosamine and 2,3-diacylglucosamine-1-phosphate; *peb3*; *glpT*, encoding a putative 3-PG transporter and *surE* encoding a stationary phase survival protein. Genome sequences of *C. jejuni* (n=64), *C. coli* (n=40), *C. lari* (n=1) and *C. upsaliensis* (n=2) (Appendix Table S5 & S6), were used to compare the structure of this locus using Artemis and WebACT. This region varied in gene content in different *Campylobacter* species (Figure 5.1) and among *C. jejuni* strains (Figure 5.2).

In *C. coli* this region contained *zupT*, encoding a zinc transporter and genes *cdtABC* encoding a cytolethal distending toxin (Figure 5.1). Notably, *peb3* was not present in this locus or elsewhere in the genomes of 40 *C. coli* strains. *C. lari* lacks *glpT*, but does contain *peb3* and six ORFs encoding cytolethal distending toxin (three ORFs), two peptidases and a transporter protein, all located between *lpxB* and *surE* (Figure 5.1). *C. upsaliensis* RM3195 and JV21 lack *peb3/glpT* and contain four ORFs, between *lpxB* and *surE*, encoding a Type III restriction modification (RM) system and a hypothetical protein (Figure 5.1). Interestingly, *lpxB*, *surE* and their flanking genes were conserved in all *Campylobacter* species (Figure 5.1).

These data demonstrate (i) that the structure of this locus varies in closely related *Campylobacter* species, (ii) *peb3* **and** *glpT* are both found in almost all *C. jejuni* strains, but in closely related species only *glpT* (*C. coli*), *peb3* (*C. lari*) or neither (*C. upsaliensis*) are present.

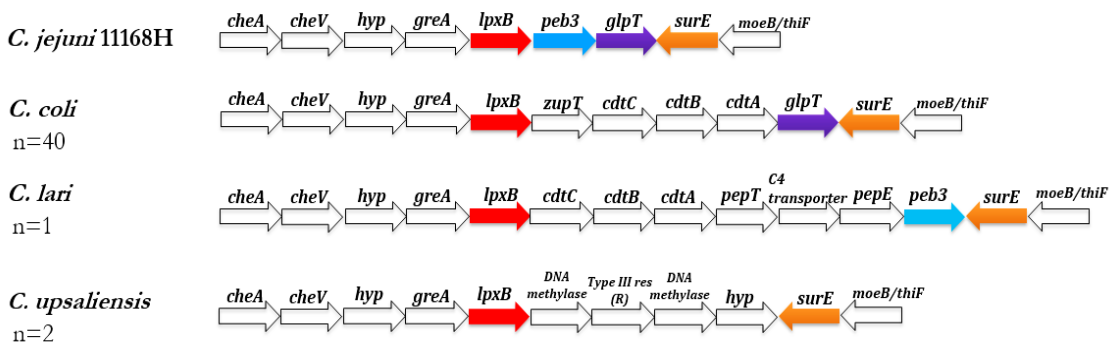


Figure 5.1 Comparison of the *peb3/glpT* locus in *C. jejuni*, *C. coli*, *C. lari* and *C. upsaliensis*. Abbreviations: **che**: chemotaxis protein A and V, **greA**: transcription elongation factor, **lpxB**: lipid-A-disaccharide synthetase, **PEB3**: Pei, Ellison, Blaser protein, **glpT**: phosphoglycerate transporter, **surE**: stationary phase survival protein, **hyp**: hypothetical protein, **DNA Methylase**: DNA methyltransferase, **Type III Res (R)**: Type III restriction enzyme R protein/restriction endonuclease, **cdt**: cytolethal distending toxin subunit A, B and C, **pepT/pepE**: peptidase; **zupT**: zinc transporter, **moeB/ThiF**: molybdopterin biosynthesis protein. Arrows indicate putative coding sequences (not drawn to scale).

5.3 *C. jejuni* strains exhibit variation in gene content in the *peb3/glpT* locus

A collection of 64 *C. jejuni* strains were grouped into thirteen classes (A–M) based on the gene organization of the *peb3/glpT* locus (Figure 5.2). The largest group (Class A) including 11168H consists of 36 strains with a locus structure of *lpxB*→ *peb3*→*glpT*→*surE* (Figure 5.2 A). All remaining classes contained six or fewer strains. Classes B, C and K with six, six and one group member(s) respectively, were very similar in structure to Class A, but also contained one or two ORFs encoding hypothetical proteins (Figure 5.2). Classes D, E, F, G, H, I and J, each with between one and three member(s), contained ORFs encoding Type II or Type III RM systems within this locus (Figure 5.2). Other ORFs found within this locus in one or more strains encoded hypothetical proteins, kinases and ATPases. Two Classes L and M contained single members that lacked *peb3*. In Class L, three ORFs encoding CDT were present between *lpxB* and *surE* (Figure 5.2). Two Classes (J and L) contained single members that lacked *glpT* (Figure 5.2).

The data demonstrate (i) that the majority of *C. jejuni* strains conserved the Class A, *lpxB*→ *peb3*→*glpT*→*surE,* locus arrangement, (ii) for remaining Classes B-M the location between *lpxB* and *surE* was clearly variable in gene content, (iii) both *peb3* **and** *glpT* are conserved in almost all *C. jejuni* strains, but in atypical strains *peb3* (*C. jejuni* 1336 and LMG23216) and *glpT* (*C. jejuni* 1336 and 414) are absent.

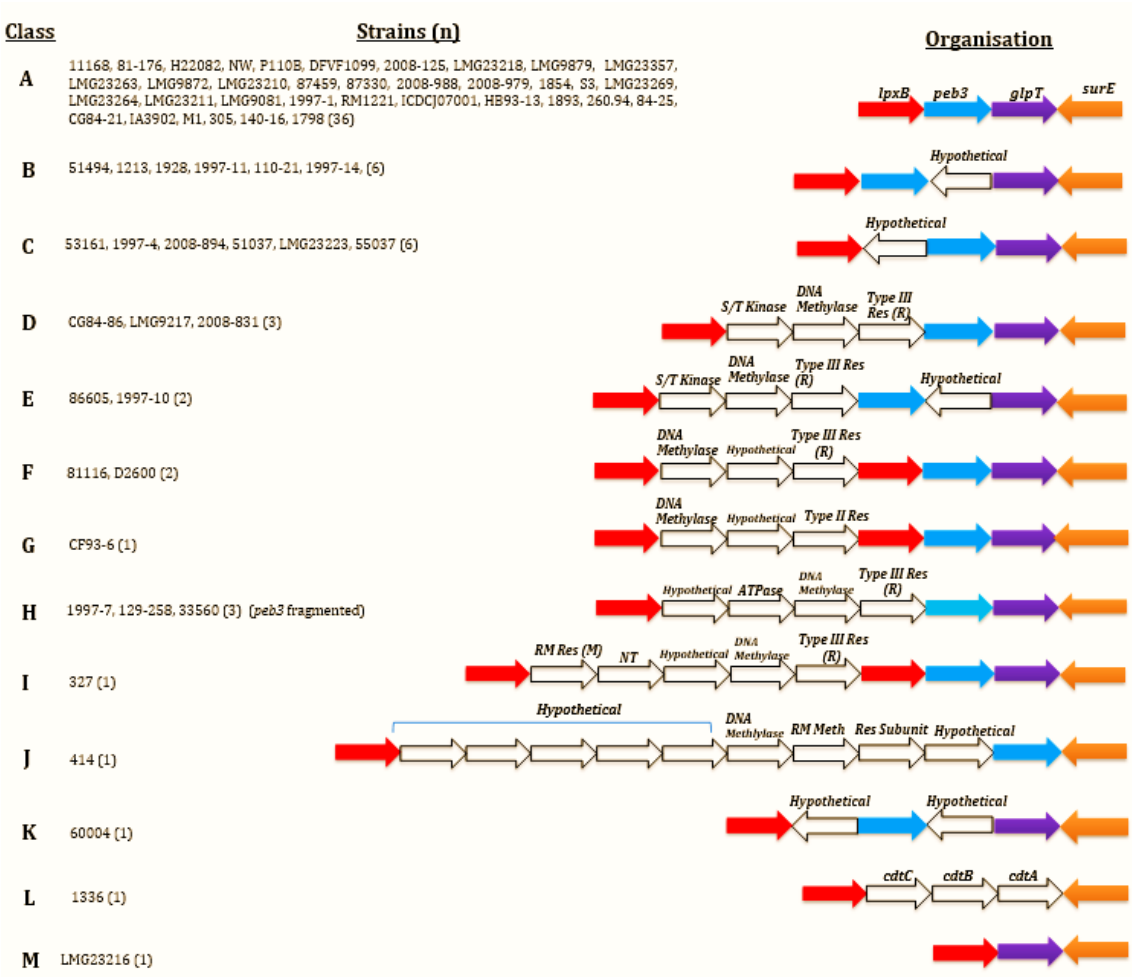


Figure 5.2 Gene arrangement of *C. jejuni* locus bounded by lpxB and surE. Abbreviations: **GreA**: transcription elongation factor, **lpxB**: lipid-A-disaccharide synthetase, **PEB3**: Pei, Ellison, Blaser protein, **GlpT**: phosphoglycerate transporter, **SurE**: stationary phase survival protein, **NT**: nucleotidyltransferase, **DNA Methylase**: DNA methyltransferase, **S/T Kinase**: serine/threonine protein kinase, **Type II/III Res (R)**: Type III restriction enzyme R protein/restriction endonuclease, **RM Res (M)**: Type III restriction enzyme M protein, **RM Meth**: Restriction modification methyltransferase, **Res subunit:** Type III restriction-modification system res enzyme, **CDT**: cytolethal distending toxin subunit A,B and C. Arrows indicate putative coding sequences (not drawn to scale).

5.4 *peb3* is generally conserved in structure among *C. jejuni* strains

Among 64 *C. jejuni* isolates, *peb3* was present in all strains apart from 1336 and LMG23216. For 59 strains *peb3* was present as a single ORF of 753 bases whilst for three strains (129-258, 1997-7 and 33560, forming Class H), the *peb3* coding sequence was fragmented into 5 ORFs due to single nucleotide indel events and two in-frame stop codons (Figure 5.3 & Figure 5.4).

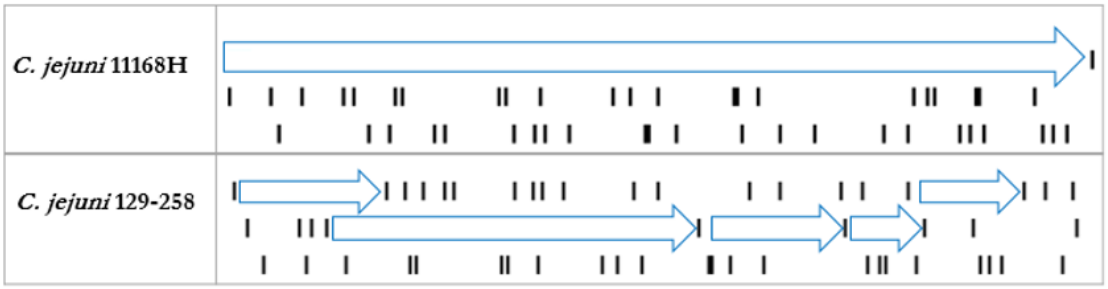


Figure 5.3 Diagrammatic representations of the *peb3* gene coding sequence. The *peb3* of 11168H is full-length consisting of 1 ORF. The *peb3* coding sequences of *C. jejuni* 129-258, 1997-7 and 33560 (Class H) strains consisted of 5 ORFs due to indel events and in-frame stop codons. The 129-258 *glpT* structure shown is identical in strains 1997-7 and 33560. The black vertical lines represent stop codons.

```
11168H          ---------------ATGAAAAAAATTATTACTTTATTTGGTGCATGTGCCTTAGCTTTT
129-258         ATGAAAAGAAACACTATTAAAAAAATTATTACTTTATTTGGTGCATGCGCCTTAGCTTTT
                                **.******************************.************60

11168H          AGTATGGCAAATGCAGATGTAAACCTTTACGGACCAGGTGGCCCACACACGGCCTTA-AA
129-258         AGTATAGCAAATGCGGATATAAATCTTTACGGACGAGGTGACCCACACACTGCCTTACAA
                *****.********.***.****.**********.*****.*********.******.**120

11168H          AGATATAGCAAACAAATATAGCGAAAAAACAGGCGTTAAAGTAAATGTAAATTTTGGCCC
129-258         AGATATGGCAAGCAAATATAGCGAAAAAACAGGCGTTAAAGTAAATGTAAATTTTGACTC
                ******.****.*********************************************.*.*180

11168H          TCAAGCGACTTGGTTTGAAAAGGCTAAAAAAGATGCAGATATTTTATTTGGCGCTTCAGA
129-258         TCAAGCGACTTGGTTTGAAAAGGCTAAAAAAGATGCAGATATTTTATTTGGCGCTTCAGA
                ************************************************************240

11168H          TCAATCGGCTTTAGCTATAGCGAGTGATTTTGGAAAAGATTTTAATGTGAGTAAAATCAA
129-258         TCAATCGGCTTTAGCTATAGCGAGTGATTTTGGAAAAGATTTTAATGTGAGTAAAATCAA
                ************************************************************300

11168H          GCCTTTATATTTTAGAGAAGCCATCATACTTACTCAAAAAGGCAATCCTTTAAAAATCAA
129-258         GCCTTTATATTTTAGAGAAGCTATCATACTTACTCAAAAAGGCAATCCTTTAAAAATCAA
                *********************.**************************************360

11168H          AGGTTTAAAAGATTTGGCTAATAAAAAAGTAAGAATCGTTGTGCCTGAAGGTGCTGGAAA
129-258         AGGTTTAAAAGATTTGGCCAATAAAAAAGTAAGAATCGTTGTGCCTGAAGGTGCTCGAAA
                ******************.***********************************.****420

11168H          GAGCAATACTTCTGGAACTGGAGTTTGGGAAGATATGATAGGTAGAACTCAAGATATAAA
129-258         GAGCAATACTTCTGGAACTGGAGTTTAGGAAGATATGATAGGCAGAACCCAAGATATAAA
                **************************.***************.*****.***********480

11168H          AACCATACAAAATTTTAGAAACAATATCGTGGCCTTTGTTCCAAATAGTGGAAGTGCAAG
129-258         AACCATACAAAATTTTAGAAACAATATCGTAGCTTTTGTTCCAAATAGTGGAAGTGCAAG
                ******************************.**.***************************540

11168H          AAAGCTTTTCGCACAAGATCAAGCCGATGCTTGGATCACTTGGATTGACTGGTCAAAAAG
129-258         AAAGCTTTTCGCACAAGATCAAGCCGATGCTTGGATCACTTGGATTGACTGATCAAAAAG
                ****************************************************.********600

11168H          CAATCCTGACATAGGAACTGCCGTAGCTATAGAAAAAGATTTGGTTGTTTATAGAACTTT
129-258         CAATTCTGACATAGGAACTGTCGTAGCTATAGAAAAAGACTTGGTTGTTTATAGAAC-TT
                ****.***************.******************.***************** **660

11168H          TAATGTGATAGCTAAAGAAGGTGCGAGCAAAGAAACACAAGATTTTATTGCTTATTTAAG
129-258         TAATGTGGTAGCCAAGGAAGGTGCGAGCAAAGAAACACAAGATTTTATTGCTTATTTAAG
                *******.*****.**.**********************************************720

11168H          TTCTAAGGAAGCTAAAGAAATTTTTAAAAAATACGGCTGGAGAGAATAA
129-258         TTCTAAGGAAGCTAAAGAAATTTTTAAAAAATACGGCTGGAGAGAATAA769
                **************************************************
```

Figure 5.4 Nucleotide alignment of *peb3* genes from 11168H and 129-258. The *peb3* sequences of 129-258, 1997-7 and 33560 were identical with the exception of 33560, which differs by single base (A→G) at position 307 (bolded and underlined in blue). An insertion was observed at position 118 and a deletion at 658 (boxed and shaded). A total of four premature stop codons were observed in 129-258 (underlined in red). The single strain 129-258 is a representative of Class H strains.

5.5 The diversity of the *glpT* genetic structure in *C. jejuni*

When *glpT* gene sequences from 64 *C. jejuni* strains were aligned using Clustal O significant diversity in gene sequence and structure was observed (Figure 5.5 & Appendix Figure S1). In only five strains (81116, 327, 2008-894, LMG9872 and LMG23216), *glpT* is present as a single ORF that lacked frameshift mutations. Two strains, *C. jejuni* 1336 and 414, lacked *glpT*, but in 57 strains *glpT* was disrupted by in-frame stop codons due to, in the majority of cases, small indel (1/2 nt) mutations. In a smaller number of strains (n=7) larger (>10 nt) deletions were present. This variation is highly unusual suggesting atypical selection pressure on the *glpT* gene.

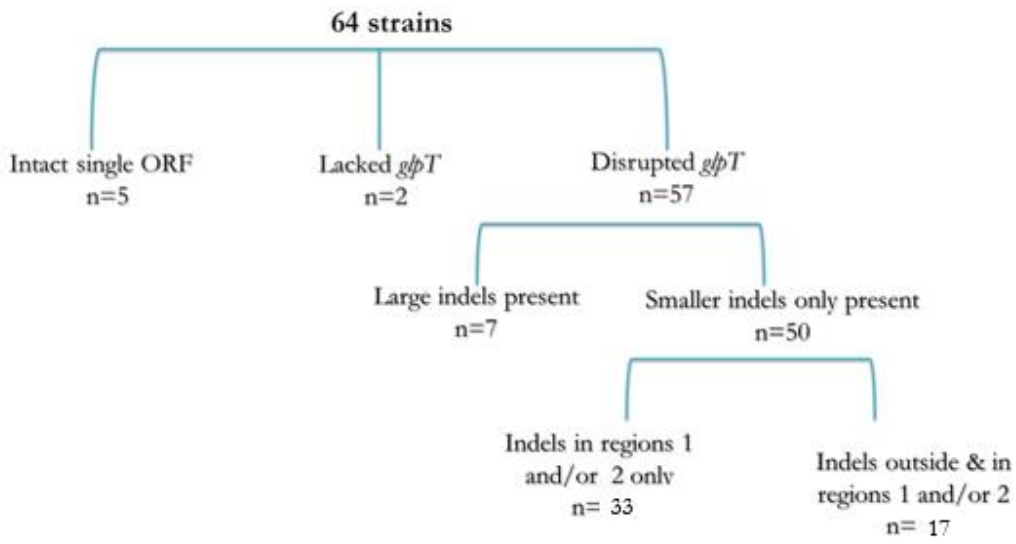


Figure 5.5 Overview of *glpT* variations in *C. jejuni*

Five strains (81116, 327, LMG23216, 2008-894, and LMG9872) possessed an intact *glpT* single ORF. These *glpT* genes of 1371 bp encoded GlpT proteins of 456 aa that had high levels of sequence identity at the nucleotide (98-100 %) (data not shown) and amino acid (97-100 %) level (Figure 5.6).

Among the 57 strains that contained a disrupted *glpT*, seven strains (81-176, 2008-831, M1, 1997-7, 129-258, 33560 and LMG9879) contained single large deletions of >10 nucleotides (Figure 5.7 A & Appendix Figure S2). Strains 81-176 and 2008-831 had identical 277 nt deletions, following base 140 relative to 81116 *glpT*. Strain M1 had a 50 nt deletion following nucleotide 335. Strains 1997-7, 129-258 and 33560 contained a 45 nt deletion following nucleotide 1315. Finally, LMG9879 had a 13 nt deletion following base 519. All 7 *glpT* genes with >10 nt deletions did not encode a single ORF due to these large deletion events and/or other 1 or 2 nucleotide indel events. The various arrangements of the *glpT* ORFs are depicted in Figure 5.7 B.

5.6 Small indels clustered in Indel region 1 and/or 2

Among the remaining 50 strains with only small indels two groups were evident, 33 strains with indels in specific regions 1 and/or 2 only (see below) and a second group of 17 strains with indels outside of and within region 1 and/or 2 (Figure 5.8). Of these 17 strains, five (1893, 1798, 140-16, 129-258 and 1997-7) had a single nucleotide deletion at nt 929, whilst four strains (1213, 110-21, 1928 and 84-25) had two nt deletions at residues 668, 669. In a further eight strains a variety of further small indel events outside of region 1 and 2 were present (Figure 5.8).

The majority of *glpT* small indel events were present in two regions of the gene (Figure 5.9). These were located around nucleotides 278-290 and 1243-1255, termed indel region 1 and 2, respectively (Figure 5.9). In the five strains with in-frame *glpT* sequences these regions had

```
LMG23216        MFDFFKPKAKAIKFPKEEILPRYYKMRTWSLSGVFIGYMGYYLVRNNITLSTPFIQNQLD
LMG9872         MFDFFKPKAKAIKFPKKEILPRYYKMRTWSLSGVFIGYMGYYLVRNNITLSTPFIQNQLD
81116           MFDFFKPKAKAIKFPKEEILPRYYKMRTWSLSGVFIGYMGYYLVRNNITLSTPFIQNQLD
327             MFDFFKPKAKAIKFPKEEILPRYYKMRTWSLSGVFIGYMGYYLVRNNITLSTPFIQNQLD
2008-894        MFDFFKPKAKAMKFPKEEILPRYYKMRTWSLSGVFIGYMGYYLVRNNITLSTPFIQNQLD
               1***********:****:*******************************************

LMG23216        LSKSDIGTITGSMLIAYGISKGAMSVISDKADPKKYMALGLILCALVNVLLGFSNSFYAY
LMG9872         LSKSDIGTITGSMLIAYGISKGAMSVISDKADPKKYMALGLILCALVNVLLGFSNSFYAY
81116           LSKSDIGTITGSMLIAYGISKGAMSVISDKADPKKYMALGLILCALVNVLLGFSNSFYAY
327             LSKSDIGTITGSMLIAYGISKGAMSVISDKADPKKYMALGLILCALVNVLLGFSNSFYAY
2008-894        LSKSDIGTITGSMLIAYGISKGAMSVISDKADPKKYMALGLILCALVNVLLGFSNSFYAY
                ************************************************************

LMG23216        VGFVIALGVFQGMGVGPSFITLANWYPKKERGIYTAIWNISHNIGGGIVAPIVSLSGFAL
LMG9872         VGFVIALGVFQGMGVGPSFITLANWYPKKERGIYTAVWNISHNIGGGIVAPIVSLSGFAL
81116           VGFVIALGVFQGMGVGPSFITLANWYPKKERGIYTAVWNISHNIGGGIVAPIVSLSGFAL
327             VGFVIALGVFQGMGVGPSFITLANWYPKKERGIYTAVWNISHNIGGGIVAPIVSLSGFAL
2008-894        VGFVIALGVFQGMGVGPSFITLANWYPKKERGIYTAVWNISHNIGGGIVAPIVSLSGFAL
                ************************************:***********************

LMG23216        AALLGVSMADFNETYWHMNHFYAPAACAVIISLYVLYAVKGSPKNEGLVDITEINEMRGI
LMG9872         AALLGVSMADFNETYWHMNHFYVPAACAVIISLYVLYAVKGSPKNEGLVDITEINEMRGI
81116           AALLGVSMADFNETYWHMNHFYVPAACAVIISLYVLYAVKGSPKNEGLIDITEINEMRGI
327             AALLGVSMADFNETYWHMNHFYVPAACAVIISLYVLYAVKGSPKNEGLIDITEINEMRGI
2008-894        AALLGVSMADFNETYWHMNHFYAPAACAVIISLYVLYAVKGSPKNEGLVDITEINEMRGI
                **********************.**************************:***********

LMG23216        KTEEIKAVETPNLSSFEIFYRYVLKNKNAWYVAWMDTFVYMVRFGLISWLPIYLLETKGF
LMG9872         KTEEIKAVESPNLSSFEIFYRYVLKNKNAWYVAWMDTFVYMVRFGLISWLPIYLLETKGF
81116           KTEEIKAIETPNLSSFEIFYRYVLKNKNAWYVAWMDTFVYMVRFGLISWLPIYLLETKGF
327             KTEEIKAVETPNLSSFEIFYRYVLKNKNAWYVAWMDTFVYMVRFGLISWLPIYLLETKGF
2008-894        KTEEIKAIFTPNLSSFEIFYRYVLKNKNAWYVAWMDTFVYMVRFGLISWLPIYLLETKGF
                *******:*:**************************************************

LMG23216        NKEQMGIAFWLFEWAAIPSTLLAGYISDKIFKGYRMPPAIGAMVIIFFMIIGYFTSNNLY
LMG9872         NKEQMGIAFWLFEWAAIPSTLLAGYISDKIFKGYRMPPAIGAMVIIFFMIIGYFTSNNLY
81116           NKEQMGIAFWLFEWAAIPSTLLAGYISDKIFKGYRMPPAIGAMVIIFFMIIGYFTSNNLY
327             NKEQMGIAFWLFEWAAIPSTLLAGYISDKIFKGYRMPPAIGAMVIIFFMIIGYFTSNNLY
2008-894        NKEQMGIAFWLFEWAAIPSTLLAGYISDKIFKGYRMPPAIGAMVIIFFMIIGYFTSNNLY
                ************************************************************

LMG23216        MVIFFAAMAGCLVYIPQFLASVQTMEVVPAFAVGSCVGLRGFMSYVVGASLGTKAIGWAV
LMG9872         MVIFFAAMAGCLVYIPQFLASVQTMEVVPAFAVGSCVGLRGFMSYVVGASLGTKAIGWAV
81116           MVIFFAAMAGCLVYIPQFLASVQTMEVVPAFAVGSCVGLRGFMSYVVGASLGTKAIGWAV
327             MVIFFAAMAGCLVYIPQFLASVQTMEVVPAFAVGSCVGLRGFMSYVVGASLGTKAIGWAV
2008-894        MVIFFAAMAGCLVYIPQFLASVQTMEVVPAFAVGSCVGLRGFMSYVVGASLGTKAIGWAV
                ************************************************************

LMG23216        DYYGSWNAGLIMLLSACILCILCSTLCHFSAKKKYA
LMG9872         DYYGSWNAGLIMLLSACILCILCSTLCHFSAKKKYQ
81116           DYYGSWNAGLIMLLSACILCILCSTLCHFSAKKKYQ
327             DYYGSWNAGLIMLLSACILCILCSTLCHFSAKKKYQ
2008-894        DYYGSWNAGLIMLLSACILCILCSTLCHFSAKKKYQ
                ************************************ 456
```

Figure 5.6 Alignment of the full length GlpT proteins in *C. jejuni* strains 81116, LMG23216, LMG9872, 327 and 2008-894. The amino acid alignment of the five *C. jejuni* strains, which possess a full-length GlpT protein. The grey boxes indicate changes in the amino acid sequence.

sequences, AAAAAATACAT (Region 1) and AAA(A/G)GCAT (Region 2). Fourteen further *glpT* sequences that contained these sequences at Region 1 and 2 do not produce a single ORF, as they contain indels of various types that result in failure to produce a single ORF.

Further characterisation of this area revealed additional in-frame and out-of-frame sequences. Strains (n=7) with in-frame sequence AAAAAATACAT (region 1), but out-of-frame sequence AAAAGCAT in Indel region 2, were observed. A large number of strains (n=26) contained an out-of-frame Indel Region 1 sequence: AAAAAAATAT, AAAAAATCAT, AAAAAACAT, AAAAATCAT and AAAAATACAT and an out-of-frame Indel Region 2 sequence: AAAAGCATAG and AAGGCTTAT. Lastly, strains with an out-of-frame Indel region 1 sequence: AAAAAAATAT, AAAAAATCAT, AAAAAACAT, AAAAATCAT and AAAAATACAT and in-frame Indel region 2 AAAAGCTATAG was observed among 12 strains. Strains 81-176 and 2008-831 lack an indel region 1, however contain an Indel region 2 (Figure 5.9). In 81-176 the Indel region 2 is out-of-frame, but in-frame in 2008-831.

In summary, *glpT* is diverse and this diversity is characterised by indel events. These events are also diverse but two regions are particular hot spots for small indel events.

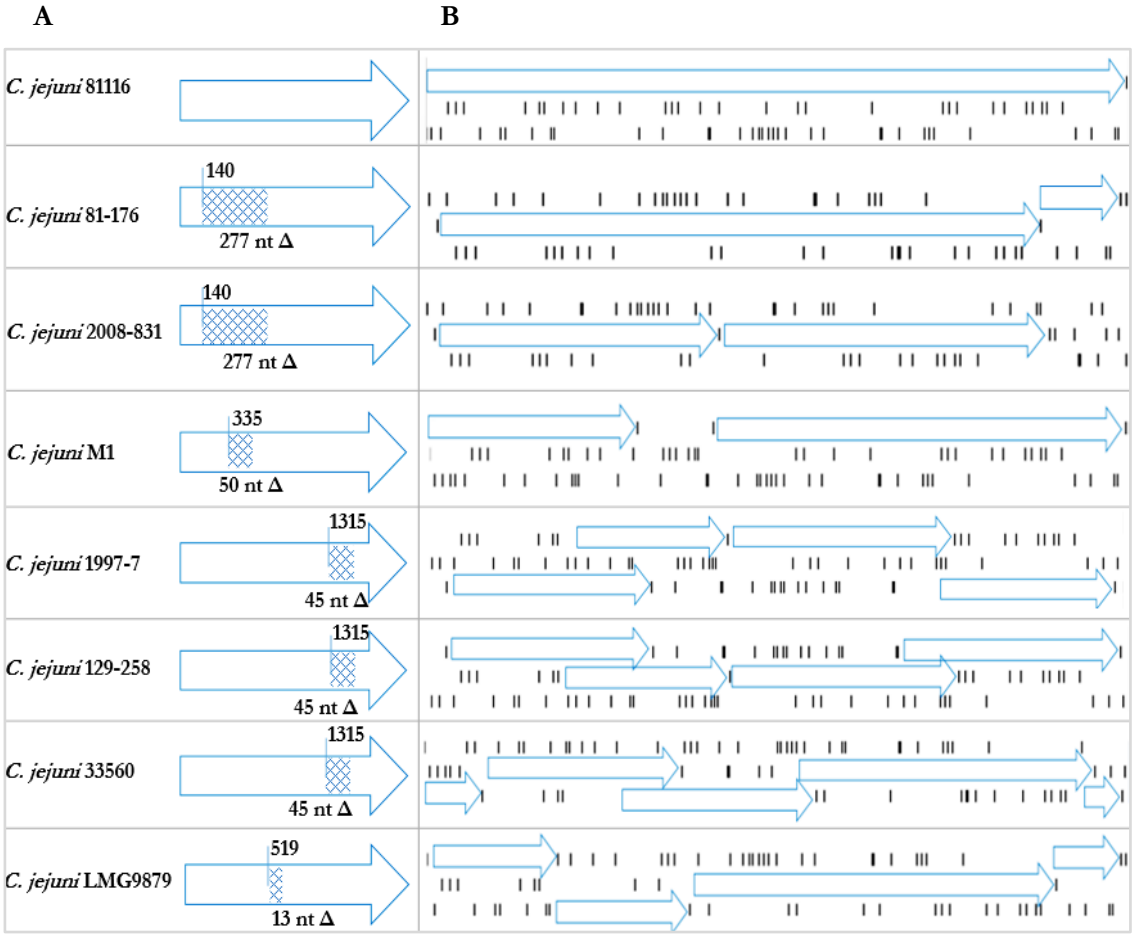


Figure 5.7 Diagrammatic representations of the *C. jejuni glpT* genes that contain large deletions (A) and their corresponding ORFs (B). Arrows indicate the coding sequence of *glpT* genes from strains, 81-176, 2008-831, M1, LMG9879, 33560, 1997-7 and 129-258. Deletions (**Δ**) of 277, 50, 45 and 13 nt, respectively, compared with 81116 *glpT* are shown below the arrow and starting positions of the deletion above the arrow. **B) The *glpT* gene structure viewed with the Artemis genome browser**. A number of *glpT* gene structures contained multiple overlapping and non-overlapping ORFs (2-5). The black vertical lines represent stop codons.

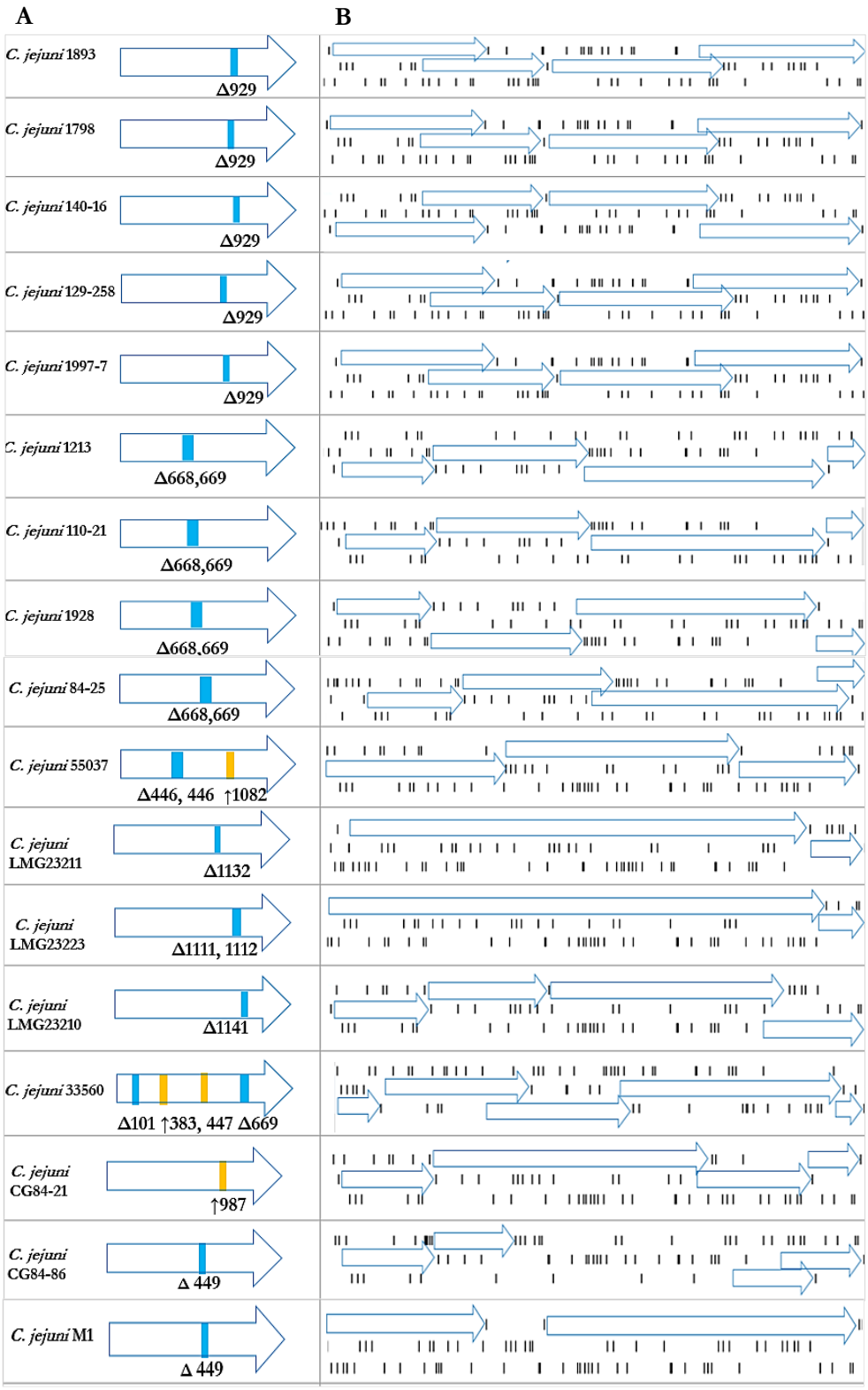


Figure 5.8 Schematic diagram of *glpT* genes containing indels in scattered locations. A) The arrows indicate the coding sequence of 17 *glpT* genes (not drawn to scale) that contain indels outside of Indel regions 1 and 2. Deletions and insertions, are represented by Δ (blue lines) and ↑ (yellow lines), respectively, along with nt position. **B) Structure of *glpT* gene coding sequence with indel events outside of regions 1 and 2.** The Artemis genome browser view of 17 *glpT* genes were shown to contain multiple ORFs (2-5). Black vertical lines represent stop codons.

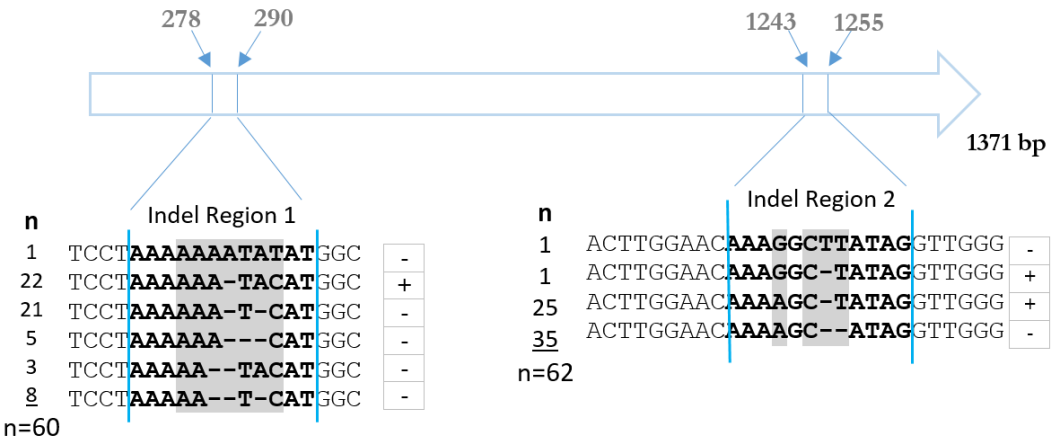


Figure 5.9 Sequence patterns associated with *glpT* Indel Region 1 and 2. A schematic diagram of the full length *glpT* of 81116 is shown with arrows indicating the positions of Indel regions 1 & 2. In the remaining *C. jejuni* strains an alignment of the various sequence patterns in these regions and strain number (n) is provided. It should be noted strains 81-176 and 2008-831 lack indel region 1 due to a 277 nt deletion hence n=60, rather than 62. These sequences could be categorised as in-frame (+) or out-of-frame (-). Highlighted regions indicate deletions and substitutions in the sequences.

5.7 *In silico* MLST of 64 *C. jejuni* strains

In order to compare the evolutionary history of *C. jejuni* strains with both the variation in gene arrangement in the *lpxB/surE* region and the sequence variation in *glpT* itself I carried out *in silico* MLST for all 64 *C. jejuni* strains analysed (Appendix Table S7). MLST utilises the sequences of seven genes associated with intermediary metabolism to categorise isolates by sequence type (ST) (not shown) and clonal complex (ST-CC) (Dingle et al., 2001).

5.7.1 Sequence types (ST) and clonal complexes (ST-CC) of *C. jejuni* population

A total of 54 sequence types (STs) were identified among the 64 strains (Appendix Table S7). All strains were assigned to clonal complexes (ST-CCs) apart from strains 414, LMG23216, LMG23210, 1854, 2008-979, 1997-10, LMG23223, 305, 60004, and 2008-894 (Figure 5.10). The allelic profiles (Appendix Table S7) of these strains were used to construct a UPGMA tree (bootstrap with 500 replications) (Figure 5.10) using START2 software (see Methods). The six most prevalent ST-CCs (all with 3 or more strains) comprised 33 (52%) of all isolates and were (ST-21 [n=11], ST-353 [n=7], ST-48 CC [n=6], ST-354 CC [n=3], ST-61 CC [n=3] and ST-45 [n=3]). Five ST-CC's (ST-206 ST-42, ST-362, ST-22 and ST-52) contained two strains each. Eleven ST-CC's (ST-508, ST-283, ST-1275, ST-841, ST-403, ST-677, ST-179, ST-446, ST-574, ST-658 and ST-677) contained only one isolate each.

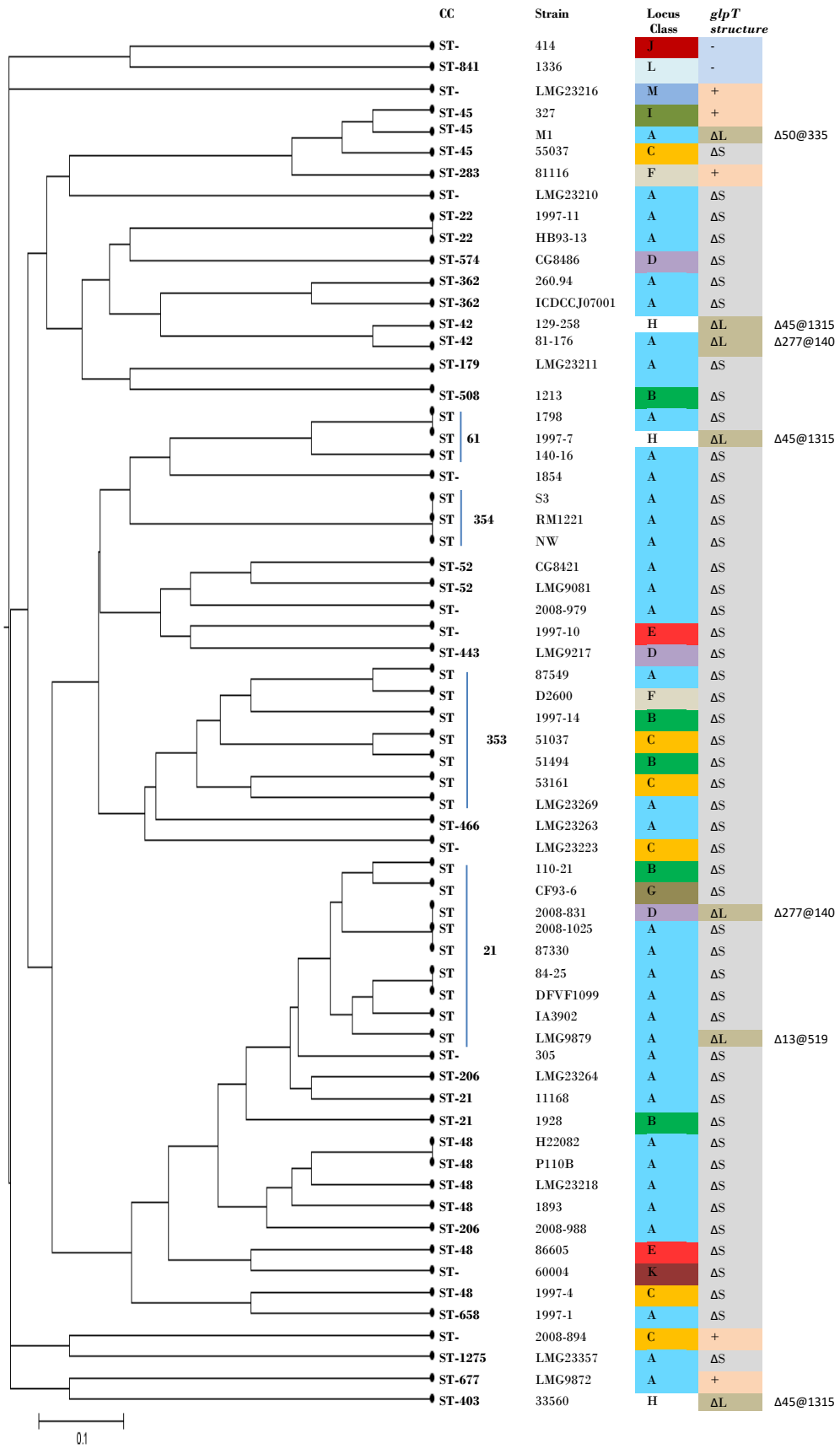


Figure 5.10 MLST UPGMA tree showing the evolutionary relation of 64 strains. UPGMA-based tree constructed with MLST allelic profiles (Appendix Table S7). The clonal complex (ST-CC), *peb3/glpT* locus arrangement class (A-M) and type of *glpT* structure: full-length (+), large deletion (ΔL), small deletion (ΔS) and absent (-) are provided.

5.7.2 Association of *C. jejuni glpT/peb3* locus gene arrangement with MLST clonal complexes

Gene arrangements, A, B, C, D, E, F, H, comprised of 36, 6, 6, 3, 2, 2 and 3 members respectively, were distributed throughout the phylogenetic tree and did not cluster with specific clonal complexes (Figure 5.10). Of the gene arrangements with single members (G, J, K, L and M) classes J, L and M formed another group, whilst G, I and K clustered with other strains.

5.7.3 Association of *glpT* structure with MLST clonal complexes

Two strains (414 and 1336) lacked *glpT* (designated as (–) in the *glpT* structure column of Figure 5.10) and these were outliers of the MLST tree (Figure 5.10). Strains with larger deletions (n=7) were found throughout the tree and those with large identical deletions at the same positions were not closely related. For example, 81-176 is located in a completely different lineage (ST-42) to strain 2008-831 (ST-21) yet both have a 277 nt deletion at base 140. These data on *glpT* locus arrangement and *glpT* structure indicate that these variations do not correlate with strain evolutionary history as inferred by MLST.

5.8 Discussion

In this chapter it was demonstrated that the genetic locus between *lpxB* and *surE* contains both inter and intra-species diversity among *Campylobacter*. This diversity was particularly prominent in the *C. jejuni glpT* gene.

In 36 *C. jejuni* strains, this locus consisted of *lpxB*, *peb3*, *glpT* and *surE* (Figure 5.2), but in *C. coli*, *C. lari* and *C. upsaliensis*, and 28 *C. jejuni* strains, additional genes were found within this locus (Figure 5.1 & 5.2). These findings suggest this is a hypervariable locus. Such hypervariable loci identified by microarray analysis include the capsule, LOS, and flagellum (Dorrell et al., 2001; Taboada et al., 2004; Karlyshev et al., 2005). These loci encode surface features needed for the survival, transmission, and pathogenesis of *C. jejuni*. This contrasts with the *lpxB*/*surE* locus that does not encode such surface features, which might be under strong selection for structural diversification. The sequencing of the 11168 strain revealed the genetic diversity of these hypervariable regions do not result from transposons, phage association or insertion sequence elements (Dorrell et al., 2001). However, other mechanisms such as phase variation due to homopolymeric tracts, horizontal gene transfer, recombination and intragenomic variation are known to contribute to genetic diversity (Wassenaar et al., 1995; Karlyshev et al., 2005).

One of the genes in the *C. jejuni lpxB*/*surE* locus is *peb3*. This gene was relatively conserved in structure/sequence (Figure 5.3) with 59/64 strains having an uninterrupted gene. The *glpT* gene is much more diverse with the majority of strains having an interrupted gene. This suggests differing evolution pressures on these physically linked genes. In the previous chapter, I provide evidence that a disrupted *glpT* gene in 11168H can still encode a functional protein. So some of these other disrupted genes may similarly encode functional proteins. In genome sequence annotations, these disrupted genes are termed pseudogenes and in individual *C. jejuni* strains there are a number of pseudogenes. For example, in reference

strains *C. jejuni* 11168, 81116 and 81-176 there are 19, 19, and 17 pseudogenes, respectively (Parkhill et al., 2000; Hofreuter et al., 2006; Pearson et al., 2007; Gundogdu et al., 2007). It may therefore also be the case that some of these pseudogenes encode functional proteins. Further examination of *glpT* genes revealed that they were interrupted due to diverse genetic events including both larger deletions of between 13 and 277 bases (Figure 5.7) and a variety of smaller indels (Figure 5.8). Larger deletions would likely mean that a functional protein could not be produced. The majority of strains contained smaller indels that were in many cases clustered into two regions of *glpT* and it is unclear why this has occurred (Figure 5.9).

To relate these various genetic events to the evolutionary history of *C. jejuni* strains an MLST based tree was used to superimpose both arrangement/composition of the *lpxB*/*surE* locus and *glpT* structure on the tree (Figure 5.10). This revealed that neither correlated with evolutionary relationships. Previous studies used microarrays to identify regions of genetic diversity in *C. jejuni* (Dorrell et al., 2001; Pittenger et al., 2012) however, *glpT* has not been identified as a gene with sequence diversity. This is presumably due to the low overall level of sequence diversity. Furthermore, areas containing small deletions, gene rearrangements, intergenic regions and mutations are typically excluded from DNA microarrays and comparative genome indexing (Pittenger et al., 2012). Nevertheless, the presence of even small indel events have a profound effect on gene structure and function.

Atypical strains, 414 and 1336, lacking *glpT* are distinct from the main group of *C. jejuni* strains (Hepworth et al., 2011) and belong to a Water/Wildlife associated lineage. However, all typical *C. jejuni* strains contain *glpT*.

Further work should address the reason *glpT* has diverse structures in different *C. jejuni* strains. Given the MLST data it would seem that these variants are not vertically inherited. Instead data suggest variation in genetic structure may arise very frequently/independently

or are passed among strains via horizontal gene transfer. If these mutations render *glpT* non-functional then it would seem that *C. jejuni* is losing the *glpT* gene and does not require this gene product for survival. This may be associated with some shift in niche/habitat. If these mutations, at least in some cases, do not disrupt function as suggested by data in Chapter 4, then what is their role or advantage over conserving *glpT* sequence? Furthermore, *glpT* is not found in water/wildlife strains suggesting that it may not be involved in interactions with some host species. These questions will require further study of both *glpT* function and diversity in *C. jejuni* strains.

CHAPTER 6

Conclusions

6.1 Introduction

The initial focus of this thesis was to characterise the PEB3 protein, present in the genome of pathogenic bacterium, *C. jejuni* 11168H. PEB3 was first identified alongside PEB1, PEB2 and PEB4 proteins (Pei et al., 1991). Prior to this study, PEB3 was shown to bind GAGs, like heparin (Linton unpublished). PEB3 also binds 3-PG (Min et al., 2009). However, the role of PEB3 *in vivo* remained unclear as it may deliver 3-PG to a transporter protein (Min et al., 2009). Scrutiny of the 11168H genome revealed an uncharacterised gene adjacent to *peb3,* a *glpT* gene, encoding for a putative phosphoglycerate transporter. This discovery might suggest a related function for the corresponding proteins. Regardless, the location of these genes was not sufficient for the assignment of the gene product function. As such, an investigation into the functionality of both PEB3 and GlpT was required.

6.2 Identification of *C. jejuni* heparin-binding proteins

In Chapter 3, results confirmed that *C. jejuni* binds heparin. Five heparin-binding adhesins were identified in this study. However, additional experiments are needed to confirm the heparin-binding features of these proteins identified *in silico*. Experimental work via modelling where possible, site-directed mutagenesis, and heparin-affinity chromatography would provide more detailed information and establish how the *C. jejuni* protein/Heparin interaction is facilitated. For one of these adhesins, PEB3, further work was carried out as a prior study indicated that PEB3 binds heparin (Linton Lab unpublished). Further to this study, the role of two putative PEB3 heparin-binding sites: $_{62}$KKAKD$_{66}$ and $_{122}$NKKVRI$_{127}$ was investigated. This was approached with molecular visualisation (PyMol) and docking software (ClusPro), site-directed mutagenesis and heparin-affinity chromatography. Molecular visualisation models showed these sites to be surface exposed. A series of docking models via the ClusPro server were carried out utilising the published PEB3 and heparin tetrasaccharide structures. *In silico* models indicated heparin interacted with three residues from PEB3 site $_{122}$NKKVRI$_{127}$, but no residues from $_{62}$KKAKD$_{66}$. These results were tested

with site-directed mutagenesis and heparin-affinity chromatography. The WT and site-directed PEB3 mutants were suitable for testing the validity of the modelling predictions, experimentally. In these mutants, two basic residues K123 and K124 were removed, leaving only R126 unchanged. The removal of K123 and K124 in single and double mutant mutants resulted in decreased heparin-binding. Lastly, a triple mutant resulted in the removal of R126. It was expected when this mutant was subjected to heparin-affinity chromatography, it would either not bind or its affinity would be reduced. However, the deletion of R126 in a triple mutant failed to disrupt the heparin-binding affinity of PEB3. These data indicate that R126 does not play a critical part in PEB3/Heparin interactions. Furthermore, the experimental data is consistent with the results from the docking models, which shows that residue R126 is not orientated in a way conducive to participating in heparin-binding. Even so, at this stage, no final conclusions can be made regarding this matter as heparin-binding is not wholly abolished in these experiments. As it was explained in Chapter 3, *in silico* models identified additional residues N122, K113, K115 and N159, interacting with heparin.

In regards to the $_{62}$KKAKD$_{66}$ binding site, residues K64 and K65 were replaced with alanine residues experimentally. *In silico* results showed no participation of these residues in heparin-binding. Nonetheless, the removal of these residues resulted in reduced heparin-binding. This shows the *in silico* predictions for this area are incorrect. One explanation for this difference is the ClusPro programme treats the heparin ligand as rigid, with minimal flexibility in structure, which may not allow for binding to this region of PEB3 (Mottarella et al., 2014). However, in the heparin-chromatography the heparin ligand may provide flexibility allowing for interaction with this PEB3 site. Lastly, an attempt to remove K62 was unsuccessful. This may be due to the position/combinations of substituted residue(s) or possibly the size of the site-directed residue. Even so, this would necessitate further investigation and perhaps an alternative approach to introduce this mutation.

In summary, PEB3 binds heparin. *In silico* findings suggest additional residues in addition those proposed in this work maybe involved. However, generating and testing the effect that all the substitutions have on the heparin-binding ability of the PEB3 protein is likely to be a laborious task. Furthermore, issues arose in the attempted to generate of a K62/K64/K65 mutant. Alternative approaches to the site-directed technique could employ a peptide-based approach. This would involve identifying key heparin-binding residues of PEB3, synthesising peptides and then assessing heparin-binding.

Further research should focus on the bigger picture (i.e. the immunological advantages) of heparin-binding. Fascinatingly, heparin is shown to protect *Neisseria* and *Helicobacter* species from host serum defences (Chen et al., 1995; Dubreuil et al., 2004; Serruto et al., 2010). In *Neisseria* species heparin is bound/adsorbed by surface adhesins (i.e. OPA+ proteins-high pI, adhesin GNA2132) (Chen et al., 1995; Serruto et al., 2010). This forms a negatively charged 'pseudocapsule' which serves as camouflage and aids in the evasion of host defences. Future work beyond this thesis may involve (i) investigating how heparin might confer protection *C. jejuni* from human responses and (ii) does PEB3 absorb heparin in the same manner as adhesins in *Neisseria* species.

In conclusion, this work identifies (a) *C. jejuni* heparin-binding proteins and (b) two sites important for the PEB3/Heparin interaction.

6.3 The role of *C. jejuni* 11168H PEB3 and GlpT in 3-PG dependent growth and fosfomycin sensitivity

In Chapter 4, I build upon the work by Min et al., 2009 whereby PEB3 is shown to bind 3-PG. However, the role of PEB3 *in vivo* remained unclear as it may deliver 3-PG to a transporter protein (Min et al., 2009). Genomic analysis showed that PEB3, is located next to *glpT* gene, encoding for a putative phosphoglycerate transporter. However, this gene is

annotated as a pseudogene, due to frameshift mutations resulting in three ORFs. The results from phenotypic analysis of both PEB3 and GlpT proteins, namely for 3-PG dependent growth and fosfomycin susceptibility, were investigated. The role of both PEB3 was studied, which required the disruption of this gene and then reintroduction of a functional copy on to the bacterial chromosome. This was successfully achieved in *C. jejuni* 11168H for PEB3, whereby a PEB3-specific antibody is available (Linton and Wren unpublished). The PEB3 band was present in WT and complemented strains, but absent in the *peb3* insertional knockout. This result clearly confirmed the insertion of the kanamycin cassette in the *peb3* gene resulting in the loss of gene product. The 11168H *peb3* deletion mutant grew on agar plates and in complex media. Furthermore, differences in the observed phenotype between the WT and PEB3 knockout indicate a functional role for PEB3. Specifically, in MEMα supplemented with 3-PG where the *peb3* knockout is unable to demonstrate 3-PG dependent growth, but growth is observed in the WT and partially in the complement.

The same techniques were employed to investigate the role of GlpT. Like PEB3, growth of *glpT* mutant strains was assessed in complex media, whereby the knockout showed a slower rate of growth, however, this was not detrimental. Further to this, growth in MCLMAN medium, a simplified media for *C. jejuni*, was assessed for WT, *peb3* and *glpT* mutants. One of the most striking findings of this study is that the deletion of *glpT* showed a significant decrease in 3-PG dependent growth. However, this was not observed with the *peb3* knockout. These findings confirm that (i) PEB3 is not required for 3-PG dependent growth in MCLMAN, (ii) GlpT does not require a periplasmic binding partner and (iii) GlpT despite being a pseudogene is functional and necessary for 3-PG dependent growth.

A second phenotype, tested the sensitivity to the antibiotic fosfomycin in the form of the E-Test. These were applied to all strains, in order to assess the effect of the inactivation of both *peb3* and *glpT*. The WT and *peb3* mutants displayed identical MIC values. Surprisingly, the

glpT knockout exhibited fosfomycin resistance, which was restored to WT levels with a genetic complement. A second test, in the form of a spot plate confirmed these data. In summary, these data further strengthened the hypothesis that GlpT is functional despite being a pseudogene. This is the first report of a functional pseudogene in *C. jejuni*.

Lastly, I attempted to purify the GlpT protein. Successful purification of this GlpT protein followed after optimisation in *C. jejuni* whereby expression of *glpT* was driven promoters, *flaA* and *porA*. Upon SDS-PAGE analysis a dominate band of 52 kDa was present in *porA*/*flaA* samples processed from whole cell lysates and membrane fractions, purified with His/HA-Tag chromatography. However, when analysed with mass spectrometry, the identity of these bands could not be confirmed.

Future research should focus on renewed efforts in identifying the GlpT protein via mass spectrometry. Although, this study provides phenotypic data indicating GlpT is functional more direct evidence is needed. One issue with GlpT is the minimal tryptic cleavage sites present. To effectively cleave proteins for identification with trypsin, lysine and arginine residues must be plentiful and accessible. Proteins with few tryptic sites make obtaining the appropriate size peptides within the target mass range for MS analysis very difficult. Further work beyond this study should focus on employing other strategies for effective cleavage of bands suggested to be GlpT. This maybe approached by combining trypsin digestion with alternatives such as cyanogen bromide or chymotrypsin (Griffin & Schnitzer, 2011), which have been proven to increase protein cleavage of membrane proteins. Further suggestions for effective cleavage of membrane proteins include the use of proteases such as Proteinase K, elastase and pepsin (Griffin & Schnitzer, 2011).

This work may entail obtaining GlpT RNA-sequence data and checking for slippery sequences or pseudoknots, which are associated with this mechanism. In conclusion, these

data indicate that (a) GlpT, not PEB3 is required for 3-PG dependent growth and (b) GlpT is a functional pseudogene involved in 3-PG dependent growth and fosfomycin resistance.

6.4 *In silico* characterisation of variation in the *C. jejuni peb3/glpT* locus

In Chapter 4, genomic analysis of the *C. jejuni* 11168H revealed a possible transporter adjacent to the *peb3* gene, a *glpT gene*, encoding for a 3-PG transporter. This discovery led to further analysis of the *peb3/glpT* locus in 11168H, whereby these genes are bound by *lpxB* and *surE*. This is the first report on the *peb3/glpT* locus.

The majority of strains contained loci arrangement *lpxB-peb3-glpT-surE*. However, further comparison of this region in other *C. jejuni* and closely relate species *C. coli*, *C. lari* and *C. upsaliensis* showed genes between *lpxB* and *surE* to be hypervariable with addition and deletion of other genes. Other hypervariable loci identified in *C. jejuni* include LOS, capsule and flagella which encode for surface structure integral to *C. jejuni* pathogenesis. In contrast with other hypervariable loci of *C. jejuni* the *peb3/glpT* locus does not encode for such surface features (i.e. flagella, capsule). Interestingly, this locus has not been identified in previous genetic diversity studies. This may be because regions which code for metabolic, biosynthetic, cellular, and regulatory processes, are conserved due to limited sequence diversity (Dorrell et al., 2001). This is evident with regions encoding the CDT toxin, the PEB proteins and other proteins such as CiaB and CadF (Dorrell et al., 2001).

Peb3 was present in the majority, but not all strains. In three strains, *peb3* had multiple ORFS due to in-frame stop codons. Interestingly, the atypical *C. jejuni* 1336 strain, representative of the WW group and *C. jejuni* LMG23216, isolated from a chicken source, lacked *peb3*. Surprisingly, was *glpT* was not present in atypical strains, 414 and 1336, from the WW group (Hepworth et al., 2011). Both *peb3* and *glpT* are found in almost all *C. jejuni* strains, but in closely related species only *glpT* (*C. coli*), *peb3* (*C. lari*) or neither (*C. upsaliensis*) are present.

The absence of the *peb3* and *glpT* genes may indicate evidence of evolution of the genome leading to niche/habitat specialisation. Therefore, it would be of significant importance to examine other atypical WW group strains for the presence/absence of *peb3* and *glpT* in order and what disadvantage/advantage this may confer.

One of the most interesting findings was the genetic diversity and structure observed among *glpT* from 64 *C. jejuni* strains. Analysis of the *glpT* gene sequences showed indels located in two hotspots termed Indel regions 1 and 2. It is unclear why these mutations occur at these two distinct sites. It is possible that these regions facilitate ribosomal frameshifting (i.e. slippery sequence), which may explain how *glpT* from 11168H is produced. However, this will require further investigation of the two frameshifting sites in the 11168H *glpT*, which is beyond the scope of this study. In addition to these hot spots several strains contained larger deletions or smaller indel events. Furthermore, a majority of these *glpT* genes were shown to be disrupted, containing indels, which suggests selective pressure on *glpT*. In Chapter 4, I provided evidence that the 11168H *glpT* pseudogene was functional. Interestingly, a number of strains possess a *glpT* gene with identical sequence and structure to the 11168H *glpT* pseudogene, therefore, it would be of considerable interest to test these genes for 3-PG dependent growth and fosfomycin sensitivity. As it may be plausible that these pseudogenes are also functional. As with pseudogenes, it is not always apparent when pseudogenes are no longer transcribed or translated. Further work may evaluate the *glpT* gene region in the form of a RNA-sequence analysis, which would provide evidence of transcription and indicate if the inactivation of these *glpT* genes is recent.

Further work should investigate possible mechanisms in the production of GlpT. These include transcriptional slippage and ribosomal frameshifting. The process of transcriptional slippage occurs during the transcribing of homopolymeric-like nucleotide patterns such as poly A or T tracts (8-11 nucleotide runs) (Baranov et al., 2005). This results in the insertion

or deletion of one or more nucleotides and the synthesis of more than one protein from the gene product (Baranov et al., 2005). Due to this insertion or deletion the mRNA product will be shorter or longer than the DNA sequence template. The purpose of transcriptional slippage is to correct the frameshift and restore the open reading frame (Baranov et al., 2005). A second mechanism, known as programmed ribosomal frameshifting (PRF) occurs during translation whereby the ribosome shifts from the initial frame (0) to the -1 or +1 frame (Ketteler et al., 2012). This process is aided by the presence of signals within the mRNA such as a slippery sequence, spacer region and a secondary RNA structure (Ketteler et al., 2012; Caliskan et al., 2015). The resulting protein may be truncated, or contain two frameshift events that produce a sequence of three different frames such as seen in GlpT.

In transcriptional slippage differences between the DNA and mRNA due to the insertion or deletion of nucleotides would be present (Baranov et al., 2005). In order to define which mechanism aids the production of GlpT an RNA-sequence analysis would need to be performed. This sequence could also be used to determine if +1 or -1PRF is responsible as the FDSB and FSCAN can be used to search the mRNA sequence for the slippery sequence, spacer region and secondary RNA pseudoknot structures or if -1PRF, for the presence of a shifty stop sequence (Moon et al., 2007; Liao et al., 2009; Ketteler et al., 2012). It is possible that there is a -1PRF slippery sequence signal (AAAAAA T(U)), in Indel Region 1 and AAAG in Indel Region 2 (Figure 5.9) (Ketteler et al., 2012). However, this would require further work as the mRNA sequence may differ from DNA. If the mRNA mirrors the DNA sequence the presence and efficiency of these frameshifting sites can be studied. A common approach that experimentally tests these signals would involve fusing a *glpT* DNA fragment that contains these two putative frameshift motifs to a *lacZ* gene. If the *glpT* gene is expressed by -1PRF the frameshift can then be measured by β-galactosidase activity thus establishing whether or not -1PRF is involved (Miller, 1992; Chen & Hu 2006).

In genome sequence annotations, disrupted genes are termed pseudogenes. In reference strains *C. jejuni* 11168, 81116 and 81-176 there are 19, 19, and 17 pseudogenes, respectively (Parkhill et al., 2000; Hofreuter et al., 2006; Pearson et al., 2007; Gundogdu et al., 2007). This number of pseudogenes is small compared to the 337 pseudogenes present in *Yersinia pestis* CO92 and 1133 pseudogenes present in *M. leprae*, whose genome is decaying (Lerat & Ochman, 2005; Muro et al., 2013). As +1 and -1 PRF signals are present in viruses, bacteria and humans there is potential for this mechanism to apply to other bacterial pseudogenes (Ketteler et al., 2012).

Lastly, the *peb3/glpT* locus arrangement and *glpT* structure was correlated with *in silico* MLST data. The relative position of the strains on the MLST tree indicated that these genes were not inherited by vertical transmission. For example, strains with large deletions, of the same *peb3* size and location, in the *glpT* gene were observed in multiple CC-ST. This suggests that these deletions may have arisen independently, maybe more than once, or are horizontally transferred. In summary, these data show a) the region bound by *lpxB* and *surE* is hypervariable, b) is quite conserved, c) *glpT* displays genetic diversity and d) the *peb3/glpT* locus and *glpT* is not acquired through vertical transmission.

6.5 Concluding Remarks

The focus of this thesis was to improve our understanding of the PEB3 and GlpT Protein in the foodborne pathogen *C. jejuni*. From the assays and experiments conducted in this study it would seem that PEB3 is a heparin-binding protein and is likely to be associated with *C. jejuni* pathogenesis. Interestingly, the location of the gene *peb3* in 11168H is adjacent to a *glpT* gene encoding for a putative phosphoglycerate transporter. This discovery may suggest a related function for the corresponding proteins in 3-PG dependent growth. However, the findings of this study show that GlpT, not PEB3, is essential for this phenotype. Additionally, GlpT is shown to be involved in fosfomycin sensitivity. These data indicate that this protein

is important for *C. jejuni* metabolism and antibiotic resistance, which may influence pathogenesis of this microbe. By conducting further work suggested throughout this discussion it is hoped that a better understanding of *C. jejuni* and PEB3/heparin interactions and the production/expression of the GlpT pseudogene can be gained.

References

Acheson, D. & Allos, B. M. (2001). *Campylobacter jejuni* Infections: Update on emerging issues and trends, *Clinical Infectious Diseases*. **32**, 1201-1206

Alazzam, B., Bonnassie-Rouxin, S., Dufour, V. & Ermel, G. (2011). MCLMAN, a new minimal medium for *Campylobacter jejuni* NCTC 11168, *Research in Microbiology*. **162**, 173-179

Alfredson, D. A. & Korolik, V. (2007). Antibiotic resistance and resistance mechanisms in *Campylobacter jejuni* and *Campylobacter coli, FEMS Microbiology Letters*. **277**, 123-132

Allos, B. M. (2001). *Campylobacter jejuni* infections: update on emerging issues and trends, *Clinical Infectious Diseases*. **32**, 1201-1206

Altekruse, S. F., Stern, N. J., Fields, P. I. & Swerdlow, D. L. (1999). *Campylobacter jejuni* – an emerging foodborne pathogen, *Emerging Infectious Diseases*. **5**, 28-35

Asakura, H., Yamasaki, M., Yamamoto, S. & Igimi, S. (2007). Deletion of *peb4* gene impairs cell adhesion and biofilm formation in *Campylobacter jejuni*, *FEMS Microbiology Letters*. **275**, 278-285

Asbury, A.K. & Cornblath, D. R. (1990). Assessment of current diagnostic criteria for Guillain-Barré syndrome, *Annals of Neurology*. **27**, S21-S24

Ascencio, F., Fransson, L. A. & Wadström, T. (1993). Affinity of the gastric pathogen *Helicobacter pylori* for the N-sulphated glycosaminoglycan heparan sulphate, *Journal of Medical Microbiology*. **38**, 240-244

Bacon, D. J., Szymanski, C. M., Burr, D. B., Silver, R. P., Alm, R. A. & Guerry, P. (2001). A phase variable capsule is involved in virulence of *Campylobacter jejuni* 81-176, *Molecular Microbiology*. **40**, 769-777

Baranov, P. V., Hammer, A. W., Zhou, J., Gesteland, R. F. & Atkins, J. F. (2005). Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression, *Genome biology*. **6**, 1

Batisson, I., Guimond, M. P., Girard, F., An, H., Zhu, C., Oswald, E., Fairbrother, J. M., Jacques, M. & Harel, J. (2003). Characterization of the novel factor paa involved in the early steps of the adhesion mechanism of attaching and effacing *Escherichia coli*, *Infection and Immunity*. **71**, 4516-4525

Beery, J. T., Hugdahl, M. B. & Doyle, M. P. (1988). Colonization of gastrointestinal tracts of chicks by *Campylobacter jejuni*, *Applied and Environmental Microbiology*. **54**, 2365-2370

Bereswill, S. & Kist, M. (2003). Recent developments in *Campylobacter* pathogenesis, *Current opinion in Infectious Diseases*. **16**, 487-491

Bern, M. & Goldberg, D. (2005). Automatic selection of representative proteins for bacterial phylogeny, *BMC Evolutionary Biology*. **5**, 34

Black, R. E., Levine, M. M., Clements, M. L., Hughes, T. P. & Blaser, M. J. (1988). Experimental *Campylobacter jejuni* infection in humans, *Journal of Infectious Diseases*. **157**, 472-479

Blaser, M. J. (1997). Epidemiologic and clinical features of *Campylobacter jejuni* infections, *Journal of Infectious Diseases*. **176**, S103-S105

Blom, A. M., Hallström, T. & Riesbeck, K. (2008). Complement evasion strategies of pathogens-acquisition of inhibitors and beyond, *Molecular Immunology*. **46**, 2808-2817

Biswas, P., Jiang, X., Pacchia, A. L., Dougherty, J. P. & Peltz, S. W. (2004). The human Immunodeficiency Virus Type 1 ribosomal frameshifting rite is an invariant sequence determinant and an important target for antiviral therapy, *Journal of Virology*. **78**, 2082-2087

Braun, J., Kingsley, G., van der Heijde, D. & Sieper J. (2000). On the difficulties of establishing a consensus on the definition of and diagnostic investigations for reactive arthritis. Results and discussion of a questionnaire prepared for the 4th International Workshop on Reactive Arthritis, Berlin, Germany, July 3–6, 1999, *The Journal of Rheumatology*. **27**, 2185-2192

Brown, E. D., Vivas, E. I., Walsh, C. T. & Kolter, R. (1995). MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in *Escherichia coli*, *Journal of Bacteriology*. **177**, 4194-4197

Burucoa, C., Frémaux, C., Pei, Z., Tummuru, M., Blaser, M. J., Cenatiempo, Y. & Fauchère, J. L. (1995). Nucleotide sequence and characterization of *peb4A* encoding an antigenic protein in *Campylobacter jejuni*, *Research in Microbiology*. **146**, 467-476

Caliskan, N., Peske, F. & Rodnina, M.V. (2015). Changed in translation: mRNA recoding by −1 programmed ribosomal frameshifting, *Trends in Biochemical Sciences*. **40**, 265-274

Cao, S. & Chen, S-J. (2008). Predicting ribosomal frameshifting efficiency, *Physical biology*. **5**, 16002.

Capila, I. & Linhardt, R. J. (2002). Heparin-protein Interactions, *Angewandte Chemie International Edition*. **41**, 390-412

Cardin, A. D. & Weintraub H. J. (1989). Molecular modeling of protein-glycosaminoglycan interactions, *Arteriosclerosis*. **9**, 21-32

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. (2005). ACT: the Artemis comparison tool, *Bioinformatics*. **21**, 3422-3423

Castañeda-García, A., Rodríguez-Rojas, A., Guelfo, J. R. & Blázquez, J. (2009). The Glycerol-3-Phosphate permease GlpT is the only fosfomycin transporter in *Pseudomonas aeruginosa*, *Journal of Bacteriology*. **191**, 6968-6974

Castañeda-García, A., Blázquez, J. & Rodríguez-Rojas, A. (2013). Molecular mechanisms and clinical impact of acquired and intrinsic fosfomycin resistance, *Antibiotics*. **2**, 217-236

Chang, K-C. (2012). Revealing −1 programmed ribosomal frameshifting mechanisms by single-molecule techniques and computational methods, *Computational and Mathematical Methods in Medicine*. **2012**, 569870

Chen, T., Swanson, J., Wilson, J. & Belland, R. J. (1995). Heparin protects Opa+ *Neisseria gonorrhoeae* from the bactericidal action of normal human serum, *Infection and Immunity*. **63**,

1790-1795

Chen, C.C. and Hu, S.T. (2006). Two frameshift products involved in the transposition of bacterial insertion sequence IS629, *Journal of Biological Chemistry*, **281**, 21617-21628

Coker, A. O., Isokpehi, R. D., Thomas, B. N., Amisu, K. O. & Obi, C. L. (2002). Human Campylobacteriosis in developing countries1, *Emerging Infectious Diseases*. **8**, 237-243

Corcoran, A. T., Annuk, H. & Moran, A. P. (2006). The structure of the lipid anchor of *Campylobacter jejuni* polysaccharide, *FEMS Microbiology Letters*. **257**, 228-235

Cozzarelli, N. R., Freedberg, W. B. & Lin, E.C. (1968). Genetic control of L-alpha-glycerophosphate system in *Escherichia coli*, *Journal of Molecular Biology*. **31**, 371-387

Dasti, J. I., Tareen, A. M., Lugert, R., Zautner, A. E. & Gross, U. (2010). *Campylobacter jejuni*: a brief overview on pathogenicity-associated factors and disease-mediating mechanisms, *International Journal of Medical Microbiology*. **300**, 205-211

Davis, L. & DiRita, V. (2008). Growth and laboratory maintenance of *Campylobacter jejuni*, *Current protocols in microbiology*. 10:A:8A.1:8A.1.1–8A.1.7

Day, C. J., Tiralongo, J., Hartnell, R. D., Logue, C. A., Wilson, J. C., von Itzstein, M. & Korolik, V. (2009). Differential carbohydrate recognition by *Campylobacter jejuni* strain 11168: Influences of temperature and growth conditions, *PLoS ONE*. **4**, e4927

Day, C. J., Tram, G., Hartley-Tassell, L. E., Tiralongo, J. & Korolik, V. (2013). Assessment of glycan interactions of clinical and avian isolates of *Campylobacter jejuni*, *BMC Microbiology*. **13**, 1

Del Rocio Leon-Kempis, M., Guccione, E., Mulholland, F., Williamson, M. P. & Kelly, D. J. (2006). The *Campylobacter jejuni* PEB1a adhesion is an aspartate/glutamate-binding protein of an ABC transporter essential for microaerobic growth on dicarboxylic acids, *Molecular Microbiology*. **60**, 1262-1275

Dimachkie, M. M. & Barohn, R. J. (2013). Guillain-Barré Syndrome and variants, *Neurologic Clinics*. **31**, 491-510

Dinman, J. D. (2006). Programmed ribosomal frameshifting beyond viruses: Organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift, *Microbe* (Washington, DC). **1**, 521-527

Dinman, J. D. (2012). Mechanisms and implications of programmed translational frameshifting, *Wiley Interdisciplinary Reviews*. **5**, 661-673

Dingle, K. E., Colles, F. M., Wareing, D. R., Ure, R., Fox, A. J., Bolton, F. E., Bootsma, H. J., Willems R. J., Urwin, R. & Maiden, M. C. (2001). Multi-locus sequence typing system for *Campylobacter jejuni*, *Journal of Clinical Microbiology*. **39**, 14-23

Dorrell, N., Mangan, J. A., Laing, K. G., Hinds, J., Linton, D., Al-Ghusein, H., Barrell, B. G., Parkhill, J., Stoker, N. G., Karlyshev, A. V. & Butcher, P. D. (2001). Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity, *Genome research*. **11**, 1706-1715

D'Souza, S. E., Ginsberg, M. H. & Plow, E. F. (1991). Argniyl-glycyl-aspartic acid (RGD): a cell adhesion motif, *Trends in biochemical sciences*. **16**, 246-250

Dubreuil, J. D., Del Giudice, G. & Rappuoli, R. (2002). *Helicobacter pylori* interactions with host serum and extracellular matrix proteins: Potential role in the infectious process, *Microbiology and Molecular Biology Reviews*. **66**, 617-629

Dubreuil, J. D., Ruggiero, P., Rappuoli, R. & Del Giudice, G. (2004). Effect of heparin binding on *Helicobacter pylori* resistance to serum, *Journal of Medical Microbiology*. **53**, 9-12

Duensing, T. D. & van Putten, J. P. M. (1998). Vitronectin binds to the gonococcal adhesin OpaA through a glycosaminoglycan molecular bridge*, Biochemical Journal*. **334**, 133-139

Duncan, C., Prashar, A., So, J., Tang, P., Low, D. E., Terebiznik, M. & Guyard, C. (2011). Lcl of *Legionella pneumophila* is an immunogenic GAG binding adhesin that promotes interactions with lung epithelial cells and plays a crucial role in biofilm formation, *Infection and Immunity*. **79**, 2168-2181

Edens, R. E., Linhardt, R. J. & Weiler, J. M. (1993). Heparin is not just an anticoagulant anymore: Six and one-half decades of studies on the ability of heparin to regulate complement activation, *Complement Profiles*. **1**, 96-120

Elvin, C. M., Hardy, C. M. & Rosenberg, H. (1985). Pi exchange mediated by the *glpT*-dependent sn-glycerol-3-phosphate transport system in *Escherichia coli, Journal of Bacteriology*. **161**, 1054-1058

Endtz, H. P., Ruijs, G. J., Van Klingeren, B., Jansen, W. H., Van der Reyden, T. & Mouton, R. P. (1991). Quinolone resistance in *Campylobacter* isolated from man and poultry following the introduction of fluoroquinolones in veterinary medicine, *Journal of Antimicrobial Chemotherapy*. **27**, 199-208

Escherich T. (1886). Articles adding to the knowledge of intestinal bacteria. III. (On the existence of vibrios in the intestines and feces of babies.) *Mu¨nchener Med Wochenschrift* **33**, 815-817

Esko, J. D., Linhardt, R. J. (2009). Proteins that Bind Sulfated Glycosaminoglycans. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. Chapter 35.

Farabaugh, P. J. (1996). Programmed translational frameshifting, *Microbiological reviews*. **60**, 103

Fauchère, J. L., Rosenau, A., Veron, M., Moyen, E. N., Richard, S. & Pfister, A. (1986). Association with HeLa cells of *Campylobacter jejuni* and *Campylobacter coli* isolated from human feces, *Infection and Immunity*. **54**, 283-287

Fernández, H., Vera, F., Villanueva, M. P. & García, A. (2008). Occurrence of *Campylobacter* species in healthy well-nourished and malnourished children, *Brazilian Journal of Microbiology*. **39**, 56-58

Ferrow, R. L. & Lee, A. (1988). Motility of *Campylobacter jejuni* in a viscous environment: comparison with conventional rod-shaped bacterial, *Journal of General Microbiology*. **134**, 53-59

Flanagan, R. C., Neal-McKinney, J. M., Dhillon, A. S., Miller, W. G. & 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**, 2399-2407

Fouts, D. E., Mongodin, E. F., Mandrell, R. E., Miller, W. G., Rasko, D. A., Ravel, J., Brinkac, L. M., DeBoy, R. T., Parker, C. T., Daugherty, S. C., Dodson, R. J., Durkin, A. S., Madupu, R., Sullivan, S. A., Shetty, J. U., Ayodeji, M. A., Shvartsbeyn, A., Schatz, M. C., Badger, J. H., Fraser, C. M. & Nelson, K. E. (2005). Major structural differences and novel potential virulence mechanisms from the genomes of multiple campylobacter species, *PLoS Biology*. **3**, 0072-0085

Fromm, J. R., Hileman, R. E., Caldwell, E. E. O., Weiler, J. M. & Linhardt, R. J. (1997). Pattern and spacing of basic amino acids in heparin binding sites, *Archives of Biochemistry Biophysics*. **343**, 92-100

Gandhi, N. S. & Mancera, R. L. (2008). The structure of glycosaminoglycans and their interactions with proteins, *Chemical biology & drug design*. **72**, 455-482

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D. & Bairoch A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*. In: John M. Walker, editor. The Proteomics Protocols Handbook, Humana Press, pp. 571-607

Gilbert, M., Karwaski, M. F., Bernatchez, S., Young, N. M., Taboada, E., Michniewicz, J., Cunningham, A. M. & Wakarchuk, W. W. (2002). The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*. Biosynthesis of sialylated ganglioside mimics in the core oligosaccharide, *The Journal of Biological Chemistry*. **277**, 327-337

Gillespie, I. A., O'Brien, S. J., Frost, J. A., Adak, G. K., Horby, P., Swan, A.V., Painter, M. J. & Neal, K. R. (2002). A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: a tool for generating hypotheses, *Emerging Infectious Diseases*. **8**, 937-942

Godschalk, P. C., Heikema, A. P., Gilbert, M., Komagamin, T., Ang, C. W., Glerum, J., Broch, D., Li, J., Yuki, N., Jacobs, B. C., van Belkum, A. & Endtz, H. P. (2004). The crucial role of *Campylobacter jejuni* genes in anti-ganglioside antibody induction in Guillain-Barré syndrome, *Journal of Clinical Investigation*. **114**, 1659-1665

Geuijen, C. A., Willems, R. J. & Mooi, F. R. (1996). The major fimbrial subunit of *Bordetella pertussis* binds to sulfated sugars, *Infection and Immunity*. **64**, 2657-2665

Geuijen, C. A., Willems, R. J., Hoogerhout, P., Puijk, W. C., Meloen, R. H. & Mooi, F. R. (1998). Identification and characterization of heparin binding regions of the Fim2 subunit of *Bordetella pertussis*, *Infection and Immunity*. **66**, 2256-2263

Griffin, N. M. & Schnitzer, J. E. (2011). Overcoming key technological challenges in using mass spectrometry for mapping cell surfaces in tissues, *Molecular & Cellular Proteomics*. **10**, R110.000935

Guerry, P., Alm, R. A., Power, M. E., Logan, S. M. & Trust, T. J. (1991). Role of two flagellin genes in *Campylobacter* motility, *Journal of Bacteriology*. **173**, 4757-4764

Guerry, P. (2007). *Campylobacter* flagella: not just for motility, *Trends in Microbiology*. **15**, 456-461

Guerry, P., Poly, F., Riddle, M., Maue, A. C., Chen, Y-H. & Monteiro M. A. (2012). *Campylobacter* polysaccharide capsules: Virulence and vaccines, *Frontiers in Cellular and Infection Microbiology*. **2**, 7

Gundogdu, O., Bentley, S. D., Holden, M. T., Parkhill, J., Dorrell, N. & Wren, B. W. (2007). Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC 11168 genome sequence, *BMC genomics*. **8**, 1

Guzman-Murillo, M. A., Ruiz-Bustos, E., Ho, B. & Ascencio, F. (2001). Involvement of the heparan sulphate-binding proteins of *Helicobacter pylori* in its adherence to HeLa S3 and Kato III cell lines, *Journal of Medical Microbiology*. **50**, 320-329

Hannah, J. H., Menozzi, F. D., Renauld, G., Locht, C. & Brennan, M. J. (1994). Sulfated glycoconjugate receptors for the *Bordetella pertussis* adhesin filamentous hemagglutinin (FHA) and mapping of the heparin-binding domain on FHA, *Infection and Immunity*. **62**, 5010-5019

Harris, L. A., Logan, S. M., Guerry, P. & Trust, T. J. (1987). Antigenic variation in *Campylobacter* flagella, *Journal of Bacteriology*. **169**, 5066-5071

Hartley-Tassell, L. E., Shewell, L. K., Day, C. J., Wilson, J. C., Sandhu, R., Ketley, J. M. & Korolik, V. (2010). Identification and characterization of the aspartate chemosensory receptor of *Campylobacter jejuni*, *Molecular Microbiology*. **75**, 710-730

Hendrixson, D. R., Akerley, B. J. & DiRita, V. J. (2001). Transposon mutagenesis of *Campylobacter jejuni* identifies a bipartite energy taxis system required for motility, *Molecular Microbiology*. **40**, 214-224

Hepworth, P. J., Ashelford, K. E., Hinds, J., Gould, K. A., Witney, A. A., Williams, N. J., Leatherbarrow, H., French, N. P., Birtles, R. J., Mendonca, C. & Dorrell, N. (2011). Genomic variations define divergence of water/wildlife-associated *Campylobacter jejuni* niche specialists from common clonal complexes, *Environmental Microbiology*. **13**, 1549-1560

Higgins, C. (2001). ABC Transporters: Physiology, structure, and mechanism-an overview, *Research in Microbiology*. **152**, 201-210

Hileman, R. E., Fromm, J. R., Weiler, J. M. & Linhardt, R. J. (1998). Glycosaminoglycan-Protein interactions: definition of consensus sites in glycosaminoglycan binding proteins, *Bioessays*. **20**, 156-167

Hirmo, S., Utt, M., Ringner, M. & Wadström, T. (1995). Inhibition of heparan sulphate and other glycosaminoglycans binding to *Helicobacter pylori* by various polysulphated carbohydrates, *FEMS Immunology and Medical Microbiology*. **10**, 301-306

Hofreuter, D., Tsai, J., Watson, R. O., Novik, V., Altman, B., Benitez, M., Clark, C., Perbost, C., Jarvie, T., Du, L. & Galán J. E. (2006). Unique features of a highly pathogenic *Campylobacter jejuni* strain, *Infection and Immunity*. **74**, 4694-4707

Hofreuter, D. (2014). Defining the metabolic requirements for the growth and colonization capacity of *Campylobacter jejuni*, *Frontiers in Cellular and Infection Microbiology*. **4**, 137

Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang D. N. (2003). Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli, Science*. **301**, 616-620

Hugdahl, M. B., Beery, J. T. & Doyle, M. P. (1988). Chemotactic behavior of *Campylobacter jejuni*, *Infection and Immunity*. **56**, 1560-1566

Humphrey, T., O'Brien, S. & Madsen, M. (2007). *Campylobacters* as zoonotic pathogens: a food production perspective, *International Journal of Food Microbiology*. **117**, 237-257

Inglis, G. D., Kalischuk, L. D. & Busz, H. W. (2004). Chronic shedding of *Campylobacter* species in beef cattle, *Journal of Applied Microbiology*. **97**, 410-420

Jackson, R. L., Busch, S. J. & Cardin, A. C. (1991). Glycosaminoglycans: Molecular properties, protein interactions, and role in physiological processes, *Physiological Reviews*. **71**, 481-539

Jeong, J. Y., Mukhopadhyay, A. K., Akada, J. K., Dailidiene, D., Hoffman, P. S. & Berg, D. E. (2001). Roles of FrxA and RdxA nitroreductases of *Helicobacter pylori* in susceptibility and resistance to metronidazole, *Journal of Bacteriology*. **183**, 5155-5162

Jervis, A. J., Butler, J. A., Wren, B. W. & Linton, D. (2015). Chromosomal integration vectors allowing flexible expression of foreign genes in *Campylobacter jejuni*, *BMC Microbiology*. **15**, 230

Jin, S., Joe, A., Lynett, J., Hani, E. K., Sherman, P. & Chan, V. L. (2001). JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells, *Molecular Microbiology*. **39**, 1225-1236

Jin, S., Song, Y. C., Emili, A., Sherman, P. M. & Chan, V. L. (2003). JlpA of *Campylobacter jejuni* interacts with surface-exposed heat shock protein 90α and triggers signalling pathways leading to the activation of NF-κB and p38 MAP kinase in epithelial cells, *Cellular Microbiology*. **5**, 165-174

Jolley, K. A., Feil, E. J., Chan, M. S. & Maiden, M. C. J. (2001). Sequence type analysis and recombinational tests (START), *Bioinformatics*. **17**, 1230-1231

Kadner, R. J. & Shattuck-Eidens, D. M. (1983). Genetic control of the hexose phosphate transport system of *Escherichia coli*: mapping of deletion and insertion mutations in the *uhp* region, *Journal of Bacteriology*. **155**, 1052-1061

Kale, A., Phansopa, C., Suwannachart, C., Craven, C. J., Rafferty, J. B. & Kelly, D. J. (2011). The virulence factor PEB4 (Cj0596) and the periplasmic protein Cj1289 are two structurally related SurA-like chaperones in the human pathogen *Campylobacter jejuni, The Journal of Biological Chemistry*. **286**, 21254-21256

Kamhi, E., Joo, E. J., Dordick, J. S. & Linhardt, R. J. (2013). Glycosaminoglycans in infectious disease, *Biological Reviews*. **88**, 928-943

Karlyshev, A. V., Linton, D., Gregson, N. A., Lastovica, A. J. & Wren, B. W. (2000). Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity, *Molecular Microbiology*. **35**, 529-541

Karlyshev, A. V., Ketley, J. M. & Wren, B. W. (2005). The *Campylobacter jejuni* glycome, *FEMS Microbiology Reviews*. **29**, 377-390

Keener, K. M., Bashor, M. P., Curtis, P. A., Sheldon, B. W. & Kathariou, S. (2004). Comprehensive review of *Campylobacter* and poultry processing, *Comprehensive Reviews in Food Science and Food Safety*. **4**, 105-116

Kervella, M., Pagès, J. M., Pei, Z., Grollier, G., Blaser, M. J. & Fauchère, J. L. (1993). Isolation and characterization of two *Campylobacter* glycine-extracted proteins that bind to HeLa cell membranes, *Infection and Immunity*. **61**, 3440-3448

Ketley, J. M. (1997). Pathogenesis of enteric infection by *Campylobacter*, *Microbiology*. **143**, 5-21

Ketteler, R. (2012). On programmed ribosomal frameshifting: the alternative proteomes, *Frontiers in Genetics*. **3**, 10-3389

Klann A. G., Hull R. A., Palzkill T. & Hull S. I. (1994). Alanine-scanning mutagenesis reveals residues involved in binding of pap-3-encoded pili, *Journal of Bacteriology*. **176**, 2312-2317

Konkel, M. E., Garvis, S. D., Tipton, S., Anderson, D.E. & Jr. Cieplak, W. (1997). Identification and molecular cloning of a gene encoding a fibronectin binding protein (CadF) from *Campylobacter jejuni*, *Molecular Microbiology*. **24**, 953-963

Konkel, M. E., Kim, B. J., Rivera-Amill, V. & Garvis, S. G. (1999a). Identification of proteins required for the internalization of *Campylobacter jejuni* into cultured mammalian cells, *Advances in Experimental Medicine and Biology*. **473**, 215-224

Konkel, M. E., Kim, B. J., Rivera-Amill, V. & Garvis, S. G. (1999b). Bacterial secreted proteins are required for the internalization of *Campylobacter jejuni* into cultured mammalian cells, *Molecular Microbiology*. **32**, 691-670

Konkel, M. E., Gray, S. A., Kim, B. J., Garvis, S. D. & Yoon, J. (1999c). Identification of the enteropathogens *Campylobacter jejuni* and *Campylobacter coli* based on the *cadF* virulence gene and its product, *Journal of Clinical Microbiology*. **37**, 510-517

Konkel, M. E., Klena, J. D., Rivera-Amill, V., Monteville, M. R., Biswas, D., Raphael, B. & Mickelson, J. (2004). Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus, *Journal of Bacteriology*. **186**, 3296-3303

Konkel, M. E., Christensen, J. E., Keech, A. M., Monteville, M. R., Klena, J. D. & Garvis, S. D. (2005). Identification of a fibronectin-binding domain within the *Campylobacter jejuni* CadF protein, *Molecular Microbiology*. **57**, 1022-1035

Konkel, M. E., Larson, C. L. & Flanagan, R. C. (2010). *Campylobacter jejuni* FlpA binds fibronectin and is required for maximal host cell adherence, *Journal of Bacteriology*. **192**, 68-76

Korolik, V. (2010). Aspartate chemosensory receptor signalling in *Campylobacter jejuni*, *Virulence*. **5**, 414-417

Kuwabara, S. (2004). Guillain-Barré syndrome: epidemiology, pathophysiology and management, *Drugs*. **64**, 597-610

Kvien, T. K., Glennås, A., Melby, K., Granfors, K., Andrup, O., Karstensen, B. & Thoen, J. E. (1994). Reactive arthritis: incidence, triggering agents and clinical presentation, *The Journal of Rheumatology*. **21**, 115-122

Law, C. J., Maloney, P. C. & Wang, D.N. (2008). Ins and outs of major facilitator superfamily antiporters, *Annual review of Microbiology*. **62**, 289-305

Lemieux, M. J., Huang, Y. & Wang, D.N. (2004a). The structural basis of substrate translocation by the *Escherichia coli* glycerol-3-phosphate transporter: A member of the major facilitator superfamily, *Current Opinion in Structural Biology*. **14**, 405-412

Lemieux, M. J., Huang, Y. & Wang, D. N. (2004b). Glycerol-3-phosphate transporter of *Escherichia coli*: Structure, function and regulation, *Research in Microbiology*. **155**, 623-629

Lemieux, M. J., Huang, Y. & Wang, D. N. (2005). Crystal structure and mechanism of GlpT, the glycerol-3-phosphate transporter from *E. coli*, *Journal of Electron Microscopy*. **54**, i43-i46

Lerat, E. & Ochman, H. (2005). Recognizing the pseudogenes in bacterial genomes, *Nucleic Acids Research*. **33**,3125-3132

Lertsethtakarn, P., Ottemann, K. M. & Hendrixson, D. (2011). Motility and chemotaxis in *Campylobacter* and *Helicobacter*, *Annual Review of Microbiology*. **65**, 389-410

Liao P.-Y., Choi Y. S. & Lee K. H. (2009). FSscan: a mechanism-based program to identify +1 ribosomal frameshift hotspots, *Nucleic Acids Research*. **37**, 7302-7311

Linhardt, R. J. & Toida, T. (2004). Role of glycosaminoglycans in cellular communication, *Accounts of Chemical Research*. **37**, 431-438

Linton, D., Gilbert, M., Hitchen, P. G., Dell, A., Morris, H. R., Wakarchuk, W. W., Gregson, N. A. & Wren, B. W. (2000). Phase variation of a beta-1,3 galactosyltransferase involved in generation of the ganglioside GM1-like lipo-oligosaccharide of *Campylobacter jejuni*, *Molecular Microbiology*. **37**, 501-514

Linton, D., Allan, E., Karlyshev, A. V., Cronshaw, A. D. & Wren, B. W. (2002). Identification of *N*-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*, *Molecular Microbiology*. **43**, 497-508

Logan, S. M. & Trust, T. J. (1983). Molecular identification of surface protein antigens of *Campylobacter jejuni*, *Infection and Immunology*. **42**, 675-682

Luangtongkum, T., Jeon, B., Han, J., Plummer, P., Logue, C. M. & Zhang, Q. (2009). Antibiotic resistance in *Campylobacter*: emergence, transmission and persistence, *Future Microbiology*. **4**, 189-200

Magrane, M. & Consortium, U. (2011). UniProt Knowledgebase: hub of integrated protein data, *Database: The Journal of Biological Databases and Curation*. **2011**, bar0029

Malik-Kale, P., Parker, C. T. & Konkel, M. E. (2008). Culture of *Campylobacter jejuni* with sodium deoxycholate induces virulence gene expression, *Journal of Bacteriology*. **190**, 2286-229

Makdessi, K., Andreesen, J. R. & Pich, A. (2001). Tungstate uptake by a highly specific ABC transporter in *Eubacterium acidaminophilum*, *The Journal of Biological Chemistry*. **276**, 24557-24564

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng C. & Bryant, S. H. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins, *Nucleic Acids Research*. **39**, D225-D229

McAuley, J. L., Linden, S. K., Png, C. W., King, R. M., Pennington, H. L., Gendler, S. J., Florin, T. H., Hill, G. R., Korolik, V. & McGuckin, M. A. (2007). MUC1 cell surface mucin is a critical element of the mucosal barrier to infection, *Journal of Clinical Investigation*. **117**, 2313-2324

McFadyean, F. & Stockman, S. (1913). Report of the Departmental Committee on Epizootic Abortion, *Board of Agriculture and Fisheries*. London.

Meinersmann, R. J., Rigsby, W. E., Stern, N. J., Kelley, L. C., Hill, J. E. & Doyle, M. P. (1991). Comparative study of colonizing and non-colonizing *Campylobacter jejuni*, *American Journal of Veterinary Research*. **52**, 1518-1522

Menozzi, F. D., Gantiez, C. & Locht, C. (1991). Interaction of the *Bordetella pertussis* filamentous hemagglutinin with heparin, *FEMS Microbiology Letters*. **62**, 59-64

Menozzi, F. D., Rouse, J. H., Alavi, M., Laude-Sharp, M., Muller, J., Bischoff, R., Brennan, M. J. & Locht, C. (1996). Identification of a heparin-binding hemagglutinin present in mycobacteria, *The Journal of Experimental Medicine*. **184**, 993-1001

Menozzi, F. D., Pethe, K., Bifani, P., Soncin, F., Brennan, M. J. & Locht, C. (2002). Enhanced bacterial virulence through exploitation of host glycosaminoglycans, *Molecular Microbiology*. **43**, 1379-1386

Miller, J. H. (1992). A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Min, T., Vedadi, M., Watson, D. C., Wasney, G. A., Munger, C., Cygler, M., Matte, A. & Young, N. M. (2009). Specificity of *Campylobacter jejuni* adhesin PEB3 for phosphates and structural differences among its ligand complexes, *Biochemistry*. **48**, 3057-3067

Monteville, M. R., Yoon, J. E. & Konkel, M. E. (2003). Maximal adherence and invasion of INT 407 cells by *Campylobacter jejuni* requires the CadF outer-membrane protein and microfilament reorganization, *Microbiology*. **149**, 153-165

Moon S., Byun Y. & Han K. (2007). FSDB: a frameshift signal database. *Computational Biology & Chemistry*. **31**, 298-302

Moore, J., Caldwell, P. & Millar, B. (2001). Molecular detection of *Campylobacter* spp. in drinking, recreational and environmental water supplies, *International Journal of Hygiene and Environmental Health*. **204**, 185-189

Mottarella, S. E., Beglov, D., Beglova, N., Nugent, M. A., Kozakov, D. & Vajda, S. (2014). Docking server for the identification of heparin binding sites on proteins, *Journal of Chemical Information and Modeling*. **54**, 2068-2078

Müller, A., Del Rocio Leon-Kempis, M., Dodson, E., Wilson, K. S., Wilkinson, A. J. & Kelly, D. J. (2007). A bacterial virulence factor with a dual role as an adhesin and a solute-binding protein: the crystal structure at 1.5 Å resolution of the PEB1a protein from the foodborne human pathogen *Campylobacter jejuni*, *Journal of Molecular Biology*. **372**, 160-171

Muro, E. M., Mah, N., Moreno-Hagelsieb, G., & Andrade-Navarro, M. A. (2011). The pseudogenes of *Mycobacterium leprae* reveal the functional relevance of gene order within operons, *Nucleic Acids Research*. **39**, 1732-1738.

Nachamkin, I., Allos, B. M. & Ho, T. W. (1998). *Campylobacter* species and Guillain-Barre syndrome, *Clinical Microbiology Reviews*. **11**, 555-567

Nilsson, R. P., Beijer, L. & Rutberg, B. (1994). The *glpT* and *glpQ* genes of the glycerol regulon in *Bacillus subtilis*, *Microbiology*. **140**, 723-730

Nilsson, A. I., Berg, O. G., Aspevall, O., Kahlmeter, G. & Andersson, D. I. (2003). Biological costs and mechanisms of fosfomycin resistance in *Escherichia coli*, *Antimicrobial agents and Chemotherapy*. **47**, 2850-2858

N**uijten, P. J., van Asten, F. J., Gaastra, W. & van der Zeijst, B. A.** (1990). Structural and functional analysis of two *Campylobacter jejuni* flagellin genes, *The Journal of Biological Chemistry*. **265**, 17798-17804

Nyati, K. K. & Nyati, R. (2013). Role of *Campylobacter jejuni* Infection in the Pathogenesis of Guillain-Barré Syndrome: An Update, *BioMed Research International*. **2013**, 852195

Ó Cróinín, T. & Backert, S. (2012). Host epithelial cell invasion by *Campylobacter jejuni*: Trigger or Zipper Mechanism?, *Frontiers in Cellular and Infection Microbiology*. **2**, 25

Oberhelman, R. A. & Taylor, D. N. (2000). *Campylobacter* infections in developing countries. In: I. Nachamkin & M. Blaser, editors. Campylobacter. Washington DC: ASM, pp. 139-153

O'Ryan, M., Prado, V. & Pickering, L. K. (2005). A millennium update on pediatric diarrheal illness in the developing world, *Seminars in Pediatric Infectious Diseases*. **16**, 125-136

Parker, C. T., Horn, S. T., Gilbert, M., Miller, W. G., Woodward, D. L. & Mandrell, R. E. (2005). Comparison of *Campylobacter jejuni* lipooligosaccharide biosynthesis loci from a variety of sources, *Journal of Clinical Microbiology*. **43**, 2771-2781

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. M., Whitehead, S. & Barrell, B. G. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences, *Nature*. **403**, 665-668

Pearson, B. M., Gaskin, D. J., Segers, R. P., Wells, J. M., Nuijten, P. J. & van Vliet, A. H. (2007). The complete genome sequence of *Campylobacter jejuni* strain 81116 (NCTC11828), *Journal of Bacteriology*. **189**, 8402-8403

Pei, Z. H., Ellison, R. T. & Blaser, M. J. (1991). Identification, purification, and characterization of major antigenic proteins of *Campylobacter jejuni*, *The Journal of Biological Chemistry*. **266**, 16363-16369

Pei, Z. & Blaser, M. J. (1993). PEB1, the major cell-binding factor of *Campylobacter jejuni*, is a homolog of the binding component in gram-negative nutrient transport systems, *The Journal of Biological Chemistry*. **268**, 18717-18725

Pei, Z. Burucoa, C., Grignon, B., Baqar, S., Huang, X. Z., Kopecko, D. J., Bourgeois, A. L., Fauchère, J. L. & Blaser, M. J. (1998). Mutation in the peb1A locus of *Campylobacter jejuni* reduces interactions with epithelial cells and intestinal colonization of mice, *Infection and Immunity*. **66**, 938-943

Penner, J. L. & Hennessy, J. N. (1980). Passive hemagglutination technique for serotyping *Campylobacter fetus* subsp. *jejuni* on the basis of soluble heat-stable antigens, *Journal of Clinical Microbiology*. **12**, 732-737

Peterson, K. M., & Mekalanos, J. J. (1988). Characterization of the *Vibrio cholerae* ToxR regulon: identification of novel genes involved in intestinal colonization, *Infection and Immunity*. **56**, 2822-2829

Peterson, K. M. (2002). Expression of *Vibrio cholerae* virulence genes in response to environmental signals, *Current Issues in Intestinal Microbiology*. **3**, 29-38

Pittenger, L. G., Frye, J. G., McNerney, V., Reeves, J., Haro, J., Fedorka-Cray, P. J., Harrison, M. A. & Englen, M. D. (2012). Analysis of *Campylobacter jejuni* whole-genome DNA microarrays: Significance of prophage and hypervariable regions for discriminating isolates, *Foodborne Pathogens and Disease*. **9**, 473-479

Preston, M. A. & Penner, J. L. (1989). Characterization of cross-reacting serotypes of *Campylobacter jejuni*, *Canadian Journal of Microbiology*. **35**, 265-273

Poly, F., Read, T., Tribble, D. R., Baqar, S., Lorenzo, M. & Guerry, P. (2007). Genome sequence of a clinical isolate of *Campylobacter jejuni* from Thailand, *Infection and Immunity*. **75**, 3425-3433

Pope, J. E., Krizova, A., Garg, A. X., Thiessen-Philbrook, H. & Ouimet, J. M. (2007). *Campylobacter* Reactive Arthritis: A systematic review, *Seminars in Arthritis and Rheumatism*. **37**, 48-55

Rangarajan, E. S., Bhatia, S., Watson, D. C., Munger, C., Cygler, M., Matte, A. & Young, N. M. (2007). Structural context for protein *N*-glycosylation in bacteria: The structure of PEB3, an adhesin from *Campylobacter jejuni*, *Protein Science: A Publication of the Protein Society*. **16**, 990-995

Rathbun, K. M. & Thompson, S. A. (2009). Mutation of PEB4 alters the outer membrane protein profile of *Campylobacter jejuni*, *FEMS Microbiology Letters*. **300**, 188-194

Rathbun, K. M., Hall, J. E. & Thompson, S. A. (2009). Cj0596 is a periplasmic peptidyl prolyl *cis-trans* isomerase involved in *Campylobacter jejuni* motility, invasion, and colonization, *BMC Microbiology*. **9**, 160

Rees, J. R., Pannier, M. A., McNees, A., Shallow, S., Angulo, F. J. & Vujia, D. J. (2004). Persistent diarrhea, arthritis, and other complications of enteric infections: a pilot survey based on California FoodNet surveillance, 1998–1999, *Clinical Infectious Diseases*. **38**, S311-S317

Reid, A.N., Pandey, R., Palyada, K., Naikare, H. & Stintzi, A. (2008). Identification of *Campylobacter jejuni* genes involved in the response to acidic pH and stomach transit, *Applied and Environmental Microbiology*. **74**, 1583-1597

Ribardo, D. A., Bingham-Ramos, L. K. & Hedrixson, D. R. (2010). Functional analysis of the RdxA and RdxB nitroreductases of *Campylobacter jejuni* reveals that mutations in *rdxA* confer metronidazole resistance, *Journal of Bacteriology*. **192**, 1890-1901

Rivera-Amill, V., Kim, B. J., Seshu, J. & Konkel, M. E. (2001). Secretion of the virulence-associated *Campylobacter* invasion antigens from *Campylobacter jejuni* requires a stimulatory signal, *Journal of Infection Diseases*. **183**, 1607-1616

Rostand, K. S. & Esko, J. D. (1997). Microbial adherence to and invasion through proteoglycans, *Infection and Immunity*. **65**, 1-8

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

Ruiz-Bustos, E., Ochoa, J. L., Wadström, T. & Ascencio, F. (2001). Isolation and characterisation of putative adhesins from *Helicobacter pylori* with affinity for heparan sulphate proteoglycan, *Journal of Medical Microbiology*. **50**, 215-222

Ruiz-Palacios, G. M., Calva, J. J., Pickering, L. K., Lopez-Vidal, Y., Volkow, P., Pezzarossi, H. & West, M. S. (1990). Protection of breast-fed infants against Campylobacter diarrhea by antibodies in human milk, *The Journal of paediatrics*. **116**, 707-713

Rutherford, K., Parkhill, J., Crook, J., Hornsell, T., Rice, P. Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation, *Bioinformatics (Oxford, England)*. **16**, 944-945

Scott, N. E. & Cordwell, S. J. (2009). *Campylobacter* proteomics: guidelines, challenges and future perspectives, *Expert Review of Proteomics*. **6**, 61-74

Sebald, M. & Véron, M. (1963). Base DNA content and classification of vibrios, *Annales de l'Institut Pasteur*. **105**, 897-910

Semchenko, E. A., Day, C. J., Wilson, J. C., Grice, I. D., Moran, A. P. & Korolik, V. (2010). Temperature-dependent phenotypic variation of *Campylobacter jejuni* lipooligosaccharides, *BMC Microbiology*. **10**, 305

Serruto, D., Spadafina, T., Ciucchi, L., Lewis, L. A., Ram, S., Tontini, M., Santini, L., Biolchi, A., Seib, K. L., Giuliani, M. M., Donnelly, J. J., Berti, F., Savino, S., Scarselli, M., Costantino, P., Kroll, J. S., O'Dwyer, C., Qiu, J., Plaut, A. G., Moxon, R.,

Rappuoli, R., Pizza, M. & Aricò, B. (2010). *Neisseria meningitidis* GNA2132, a heparin-binding protein that induces protective immunity in humans, *Proceedings of the National Academy of Sciences*. **107**, 3770-3775

Shanker, S., Lee, A. & Sorrell, T. C. (1990). Horizontal transmission of *Campylobacter jejuni* amongst broiler chicks: Experimental studies, *Epidemiology and Infection*. **104**, 101-110

Silva, J., Leite, D., Fernandes, M., Mena, C., Gibbs, P. A. & Teixeira, P. (2011). *Campylobacter* spp. as a foodborne pathogen: A Review, *Frontiers in Microbiology*. **2**, 200

Smart, J. P., Cliff, M. J. & Kelly, D. J. (2009). A role for tungsten in the biology of *Campylobacter jejuni:* tungstate stimulates formate dehydrogenase activity and is transported via an ultra-high affinity ABC system distinct from the molybdate transporter, *Molecular Microbiology*. **74**, 742-757

Smith, T. & Taylor, M. S. (1919). Some morphological and biological characters of Spirilla (*Vibrio fetus N.SP.)* associated with disease of fetal membranes in cattle, *Journal of Experimental Medicine*. **30**, 299-311

Snelling, W. J., Matsuda, M., Moore, J. E. & Dooley, J. S. G. (2005). *Campylobacter jejuni*, *Letters in Applied Microbiology*. **41**, 297-302

Sonna, L.A., Ambudkar, S. V. & Maloney, P. C. (1988). The mechanism of glucose 6-phosphate transport by *Escherichia coli*, *The Journal of Biological Chemistry*. **263**, 6625-6630

Song, X. M., Forsgren, A. & Janson, H. (1998). Glycerol-3-phosphate transport in *Haemophilus influenzae*: cloning, sequencing, and transcription analysis of the *glpT* gene, *Gene*. **215**, 381-388

Song, Y. C., Jin, S., Louie, H., Ng, D., Lau, R., Zhang, Y., Weerasekera, R., Al Rashid, S., Ward, L. A., Der, S. D. & Chan V. L. (2004). FlaC, a protein of *Campylobacter jejuni* TGH9011 (ATCC43431) secreted through the flagellar apparatus, binds epithelial cells and influences cell invasion, *Molecular Microbiology*. **53**, 541-553

St. Michael, F., Szymanski, C. M., Li, J., Chan, K. H., Khieu, N. H., Larocque, S., Wakarchuck, W. W., Brisson, J. R. & Monteiro, M. A. (2002). The structures of the lipooligosaccharide and capsule polysaccharide of *Campylobacter jejuni* genome sequenced strain NCTC 11168, *European Journal of Biochemistry*. **269**, 5119-5136

Stahl, M., Friis, L. M., Nothaft, H., Liu, X., Li, J., Szymanski, C. M. & Stintzi, A. (2011). L-Fucose utilization provides *Campylobacter jejuni* with a competitive advantage, *Proceedings of the National Academy of Sciences of the United States of America*. **108**, 7194-7199

Stahl, M., Butcher, J. & Stintzi, A. (2012). Nutrient acquisition and metabolism by *Campylobacter jejuni*, *Frontiers in Cellular and Infection Microbiology*. **2**, 5

Stern, N. J., Fedorka-Cray, P., Bailey, J. S., Cox, N.A., Craven, S. E., Hiett, K. L., Musgrove, M. T., Ladley, S., Cosby, D. & Mead, G. C. (2001). Distribution of *Campylobacter* spp. in selected U.S. poultry production and processing operations, *Journal of Food Production*. **64**, 1705-1710

Szymanski, C. M., King, M., Haardt, M. & Armstrong, G. D. (1995). *Campylobacter jejuni* motility and invasion of Caco-2 cells, *Infection and Immunity*. **63**, 4295-4300

Teuber, M. (2001). Veterinary use and antibiotic resistance, *Current Opinion in Microbiology*. **4**, 493-499

Thomas, M. T., Shepherd, M., Poole, R. K., van Vliet A. H., Kelly, D. J. & Pearson B. M. (2011). Two respiratory enzyme systems in *Campylobacter jejuni* NCTC 11168 contribute to growth on L-lactate, *Environmental Microbiology*. **13**, 48-61

Taboada, E. N., Acedillo, R. R., Carrillo, C. D., Findlay, W. A., Medeiros, D. T., Mykytczuk, O. L., Roberts, M. J., Valencia, C. A., Farber, J. M. & Nash, J. H. (2004). Large-scale comparative genomics meta-analysis of *Campylobacter jejuni* isolates reveals low level of genome plasticity, *Journal of Clinical Microbiology*. **42**, 4566-4576.

Tsai, C. S. (2007). Translation and protein biosynthesis In: C.S Tsai editor. Biomacromolecules: Introduction to structure, function and Informatics. University of Michigan: Wiley, pp. 461-490

Utt, M. & Wadström, T. (1997). Identification of heparan sulphate binding surface proteins of *Helicobacter pylori*: inhibition of heparan sulphate binding with sulphated carbohydrate polymers, *Journal of Medical Microbiology*. **46**, 541-546

Utt, M., Danielsson, B. & Wadström, T. (2001). *Helicobacter pylori* vacuolating cytotoxin binding to a putative cell surface receptor, heparan sulfate, studied by surface plasmon resonance, *FEMS Immunology and Medical Microbiology*. **30**, 109-113

van Vliet, A. H. & Ketley, J. M. (2001). Pathogenesis of enteric *Campylobacter* infection, *Symposium Series (Society for Applied Microbiology)*. **30**, 45S-56S

Van den Brandhof, W., Wagenaar, J.A. & Van den Kerkhof, H. (2003). An outbreak of campylobacteriosis after drinking unpasteurized milk, 2002, the Netherlands, *International Journal of Medical Microbiology*. **293**, 548-549

Vegge, C. S., Brondsted, L., Li, Y. P., Bang, D. D. & Ingmer, H. (2009). Energy taxis drives *Campylobacter jejuni* toward the most favorable conditions for growth, *Applied and Environmental Microbiology*. **75**, 5308-5314

Velayudhan, J. & Kelly, D. J. (2002). Analysis of gluconeogenic and anaplerotic enzymes in *Campylobacter jejuni*: an essential role for phosphoenolpyruvate carboxykinase, *Microbiology*. **148**, 685-694

Velayudhan, J., Jones, M. A., Barrow, P. A. & Kelly, D. J. (2004). L-serine catabolism via an oxygen-labile L-serine dehydratase is essential for colonization of the avian gut by *Campylobacter jejuni*, *Infection and Immunity*. **72**, 260-268

Visscher, K. (2016). Chapter Two – −1 Programmed Ribosomal Frameshifting as a Force-Dependent Process, *Progress in Molecular Biology and Translational Science*. **139**, 45-72

Vucic, S., Kiernan, M. C. & Cornblath, D. R. (2009). Guillain-Barré syndrome: an update, *Journal of Clinical Neuroscience*. **16**, 733-741

Wadhams, G. H. & Armitage, J. P. (2004). Making sense of it all: bacterial chemotaxis, *National Reviews Molecular Cell Biology*. **5**, 1024-1037

Wang, Y. & Taylor, D. E. (1990). Chloramphenicol resistance in *Campylobacter coli*: nucleotide sequence, expression, and cloning vector construction, *Gene*. **94**, 23-28

Wassenaar, T. M., Fry, B. N. & Van der Zeijst, B. A. M. (1995). Variation of the flagellin gene locus of *Campylobacter jejuni* by recombination and horizontal gene transfer, *Microbiology*. **141**, 95-101

Whiley, H., van den Akker, B., Giglio, S. & Bentham, R. (2013). The role of environmental reservoirs in human campylobacteriosis, *International Journal of Environmental Research and Public Health*. **10**, 5886-5907

Wieczorek, K. & Osek, J. (2013). Antimicrobial resistance mechanisms among *Campylobacter*, *BioMed Research International*. **2013**, 340605

Willison, H. J. & O' Hanlon, G. M. (1999). The immunopathogenesis of Miller Fisher syndrome, *Journal of Neuroimmunology*. **100**, 3-12

Witte, W. (2000). Ecological impact of antibiotic use in animals on different complex microflora: environment, *International Journal of Antimicrobial Agents*. **14**, 321-325

Wright, J. A., Grant, A. J., Hurd, D., Harrison, M., Guccione, E. J., Kelly, D. J. & Maskell, D. J. (2009). Metabolite and transcriptome analysis of *Campylobacter jejuni in vitro* growth reveals a stationary-phase physiological switch, *Microbiology*. **155**, 80-94

Yang, B., Gerhardt, S. G. & Larson, T. J. (1997). Action at a distance for *glp* repressor control of *glpTQ* transcription in *Escherichia coli* K-12, *Molecular Microbiology*. **24**, 511-521

Young, K. T., Davis, L. & Dirita, V. J. (2007). *Campylobacter jejuni*: molecular biology and pathogenesis, *Nature*. **5**, 665-679

Yuki, N., Taki, T., Inagaki, F., Kasama, T., Takahashi, M., Saito, K., Handa, S. & Miyatake, T. (1993). A bacterium lipopolysaccharide that elicits Guillain-Barre syndrome has a GM1 ganglioside structure, *The Journal of Experimental Medicine*. **178**, 1771-1775

Yuki, N., Susuki, K., Koga, M., Nishimoto, Y., Odaka, M., Hirata, K., Taguchi, K., Miyatake, T., Furukawa, K., Kobata, T. & Yamada, M. (2004). Carbohydrate mimicry between human ganglioside GM1 and *Campylobacter jejuni* lipooligosaccharide causes Guillain-Barré syndrome, *Proceedings of the National Academy of Sciences of the United States of America*. **101**, 11404-11409

Zilbauer, M., Dorrell, N., Wren, B. W. & Bajaj-Elliott, M. (2008). *Campylobacter jejuni*-mediated pathogenesis: an update, *Transactions of the Royal Society of Tropical Medicine and Hygiene*. **120**, 123-129

Ziprin, R. L., Young, C. R., Hume, M. E. & Konkel, M. E. (1999). The absence of cecal colonization of chicks by a mutant of *Campylobacter jejuni* not expressing bacterial fibronectin-binding protein, *Avian Diseases*. **43**, 586-589

Appendix Table S1 Primers used for PCR amplification. Primers were used for the amplification and site-directed mutagenesis of *Campylobacter jejuni* genes and to confirm the correct construction of plasmids and knock-out mutants. Primers are all shown in the 5' to 3' orientations.

#	Primer Name	Sequence (5'-3')
321	peb3SDMK64AF	TGG TTT GAA AAG GCT GCA AAA GAT GCA GAT ATT
322	peb3SDMK64AR	ATA TCT GCA TCT TTT GCA GCC TTT TCA AAC CA
776	peb3SDMK65A/K64AF	TGG TTT GAA AAG GCTG CAG CAG ATG CAG ATA TT
777	peb3SDMK65A/K64AR	AAT ATC TGC ATC TGC TGC AGC CTT TTC AAA CC
299	peb3SDMK123AF	GAT TTG GCT AAT GCA AAA GTA AGA ATC GTT
300	peb3SDMK123AR	AAC GAT TCT TAC TTT TGC ATT AGC CAA ATC
301	peb3SDMK124AF	GAT TTG GCT AAT AAA GCA GTA AGA ATC GTT
302	peb3SDMK124AR	AAC GAT TCT TAC TGC TTT ATT AGC CAA ATC
774	peb3SDMK123A/K124AF	GAT TTG GCT AAT GCA GCA GTA AGA ATC GTT
775	peb3SDMK123A/K124AR	AAC GAT TCT TAC TGC TGC ATT AGC CAA ATC
801	peb3DMK123A/K124A/R126AF	GAT TTG GCT AAT GCA GCA GTA GCA ATC GTT
802	peb3SDMK123A/K124A/R126AR	AAC GAT TGC TAC TGC TGC_ATT AGC CAA ATC
327	S-TAGR	ATT TCG GTG ACA CTA TAG AAT
81	T7F	TAA TAC GAC TCA CTA TAG
80	SP6	ATT AGG TGA CAC TAT AG
1190	glpTF	ATG AGA ACT TGG AGT TTA AG
1191	glpTR	CCA AAA TCA CTC GCT ATA GCT AAA G
1188	glpTSDMBamHIF	GTG GCT TGG ATG GAT CCT TTT GTT TAT ATG
1189	glpTSDMBamHIR	CAT ATA AAC AAA AGG ATC CAT CCA AGC CAC
1203	11168outF	ATT TTA CCT AGA TAT TAC AAA
1204	11168outR	CAA ATT TTA CTT TTG CTT GCA A
1192	11168glpTF	AAA TTT TAC CTA GAT ATT ACA AA
1193	11168glpTR	TTG CAA GCA AAA GTT AAA ATT G
1194	11168glpTSeqF1	TGC AAA CTG GTA TCC TAA AAA
1195	11168glpTSeqF2	TTT ATA TGG TGC GTT TTG GG
1196	11168glpTSeqF3	AAC AAA AGC ATA GGT TGG GC
513	KanFout	AAA TAT GGC GCT TCA TAG AC
535	KanRout	ATC AAG CCT GAT TGG GAG AA
311	Cat EndSeqF	GTG ACG GCT TTC ATG TTT GCC G
66	Cam148SeqR	GAT CAA AGC ATA ATC AAC CCT GC
1063	Pcam148F Flank	GTG AAT TTG AAG TGA TGC AAA AAG G
1064	Pcam148R Flank	CAT AAG CGT GTA TTC TAG CAT G
1228	glpTPci1F	GCT ACA TGT TTT GAT TTT TTC AAA CCT AAG GCA AAA
1229	glpTSpeIR	GAC ACT AGT TTA GTG ATG GTG ATG GTG ATG GTG ATG GTG ATG TTG ATG CTT CTT TTT AGC GCT AAA ATG
1230	*glpT*N-TermHAF	CAAA CATATG TAC CCG TAC GAC GTT CCC GAC TAC GCT AGA ACT TGG AGT TTA AGT GGG
1231	*glpT*NoTagF	CAAA CAT ATG AGA ACT TGG AGT TTA AGT GGG
1232	*glpT*C-TermHisR	CAC CAT CAC CAT CAC TAA ACT AGC ACT AGT AAT TGT ACA CAT ATG GCG
65	Cam148F	TCA ATA TCC GGG GAT TCT GG
546	Cam148R	CTT GGA AAG GAA CAC GGC CG

Appendix Table S2 Plasmid List A * denotes a site-directed plasmid.

Plasmid	Description
pETpeb3GST	pET41a (+) containing *peb3* without signal peptide and with a N-terminal His-tag. Generated prior to my PhD.
* pET-41a(+)PEB3$_{K}$64$_{A}$	Mutant pET-41a(+)PEB3 plasmid expressing alanine instead of lysine at position 64. Generated prior to my PhD.
* pET-41a(+)PEB3 $_{K}$64$_{A}$/$_{K}$65$_{A}$	Mutant pET-41a(+)PEB3 plasmid expressing alanine instead of lysine at position 64 and 65.
* pET-41a(+)PEB3$_{K}$123$_{A}$	Mutant pET-41a(+)PEB3 plasmid expressing alanine instead of lysine at position 123. Generated prior to my PhD.
* pET-41a(+)PEB3$_{K}$124$_{A}$	Mutant pET-41a(+)PEB3 plasmid expressing alanine instead of lysine at position 124.
* pET-41a(+)PEB3$_{K}$123$_{A}$ / $_{K}$124$_{A}$	Mutant pET-41a(+)PEB3 plasmid expressing alanine instead of lysine at position 123 and 124.
* pET-41a(+)PEB3 $_{K}$123$_{A}$ / $_{K}$124$_{A}$/ $_{R}$126$_{A}$	Mutant pET-41a(+)PEB3 plasmid expressing alanine instead of lysine at position 123 and 124 and alanine instead of arginine at position 126.
p*glpT*	pGEM-T easy plasmid containing the *glpT* gene
*p*glpT*B	pGEM-T easy plasmid containing the *glpT* gene with a site-directed BamHI restriction site.
p*glpT*B::kan	pGEM-T easy plasmid containing the *glpT* gene interrupted with a kanamycin resistance cassette.
pCJ0223::cat glpT	Complementation vector expressing *glpT* with an 10x C-Terminal His-Tag whereby expression is driven by a chloramphenicol promoter.
p*glpT*$_{His}$	pGEM-T easy plasmid containing the *glpT* gene C-Terminal 6x His-Tag
p$_{HA}$*glpT*$_{His}$	pGEM-T easy plasmid containing the *glpT* gene N-Terminal HA-Tag and C-Terminal 6x His-Tag
pCJporAglpT$_{His}$	Complementation vector expressing *glpT* with an C-Terminal 6x His-Tag whereby expression is driven by a *porA* promoter
pCJporA$_{HA}$*glpT*$_{His}$	Complementation vector expressing *glpT* with an N-Terminal HA-Tag and C-Terminal 6x His-Tag whereby expression is driven by a *porA* promoter
pCJflaA$_{HA}$*glpT*$_{His}$	Complementation vector expressing *glpT* with an N-Terminal HA-Tag and C-Terminal 6x His-Tag whereby expression is driven by a *flaA* promoter

Appendix Table S3 *C. jejuni* strains. The genetic manipulation of the *peb3* and *glpT* (insertional knockout and genetic complement) was carried out in *C. jejuni* 11168H. The strains were created in or provided by* the Linton laboratory for this thesis.

***C. jejuni* Strain**	**Genotype Description**
*NCTC 11168H	Wild type genome sequence strain
*NCTC 11168H *peb3*	Knockout mutant
*NTCT 11168H *peb3*$^{+}$	11168H *peb3*::*kan* complemented with *peb3*
NTCT 11168H *glpT*	Knockout mutant
NTCT 11168H *glpT*$^{+}$	11168H *glpT*::*kan* complemented with *glpT-His*$_{10x}$
ΔNCTC 11168H *porAglpT*$^{+}_{His}$	11168H *glpT*::*kan* complemented with *glpT-His*$_{6X}$
ΔNCTC11168H *porA*$_{HA}$*glpT*$^{+}_{His}$	11168H *glpT*::*kan* complemented with *HAglpT-His*$_{6x}$
ΔNCTC11168H*flaA*$_{HA}$*glpT*$^{+}_{His}$	11168H *glpT*::*kan* complemented with *HAglpT-His*$_{6x}$

Appendix Table S4 *E. coli* strains. Listed are *E. coli* strains used for molecular biology, gene expression and protein production, along with a genotype description.

***E. coli* Strain**	**Genotype Description**
DH5α (Linton Lab)	*F' Phi80dlacZ DeltaM15 Delta(lacZYA-argF)U169 deoR recA1 endA1 hsdR17(rK-mK+)phoA supE44 lambda- thi-1*
BL21-AI (Invitrogen)	*F*$^{-}$ *ompT gal dcm lon hsdS*$_{B}$*(r*$_{B}^{-}$ *m*$_{B}^{-}$*) araB::T7RNAP-tetA*
XL1-Blue (Stratagene)	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZ⊿M15 Tn10 (Tetr)]* (Note: Prior to use 1.7 ul β-mercaptoethanol was added to cells and incubated on ice for 10 minutes with gentle mixing)

Appendix Table S5: *C. jejuni* strain accession numbers for *peb3/glpT* locus

C. jejuni strain	*peb3/glpT* locus	Accession	Total Contigs
D2600	AGTF01000007	AGTF00000000.1	AGTF01000001-AGTF01000056
IA3902	CP001876	CP001876.1	CP001876.1
RM1221	NC_003912	NC_003912	NC_003912
11168	NC_002163	NC_002163	NC_002163
81116	NC_009839	NC_009839	NC_009839
81-176	NC_008787	NC_008787	NC_008787
33560	AIOL01000013	AIOL00000000	AIOL01000001-AIOL01000220
c414	ADGM01000031	NZ_ADGM00000000	ADGM01000001-ADGM01000035
1336	ADGL01000005	ADGL00000000	ADGL01000001-ADGL01000035
S3	CP001960	CP001960.1	CP001960.1
NW	AGTE01000012	AGTE00000000	ADHK01000001-ADHK0100007
DFVF1099	ADHK01000039	ADHK00000000.1	ADHK01000001-ADHK01000071
305	ADHL01000162	ADHL00000000.1	ADHL01000001-ADHL01000333
ICDCCJ07001	NC_014802	NC_014802	NC_014802
129-258	AINY01000001	AINY00000000	AINY01000001-AINY01000102
51494	AINZ01000014	AINZ00000000	AINZ01000001-AINZ01000215
84-25	AANT02000003	NZ_AANT00000000	AANT02000001-AANT02000005
260.94	AANK01000002	NZ_AANK00000000	AANK01000001-AANK01000010
CG8486	AASY01000013	NZ_AASY00000000	AASY01000001-AASY01000019
HB93-13	AANQ01000003	NZ_AANQ00000000	AANQ01000001-AANQ01000035
M1	CP001900	CP001900.1	CP001900.1
CF93-6	AANJ01000001	NZ_AANJ00000000	AANJ01000001-AANJ01000014
327	ADHM01000004	ADHM00000000.1	ADHM01000001-ADHM01000048
P110B	AEIO01000010	AEIO00000000	AEIO01000001-AEIO01000029
H22082	AEIP01000001	AEIP00000000	EIP01000001-AEIP01000028
LMG23216	AIOA01000011	AIOA00000000	AIOA01000001-AIOA01000121
LMG23218	AIOB01000025	AIOB00000000	AIOB01000001-AIOB01000106
LMG 23223	AIOC01000011	AIOC00000000	AIOC01000001-AIOC01000105
LMG23263	AIOD01000020	AIOD00000000	AIOD01000001-AIOD01000172
60004	AIOE01000021	AIOE00000000	AIOE01000001-AIOE01000143
LMG23264	AIOF01000046	AIOF00000000	AIOF01000001-AIOF01000147
LMG23269	AIOG01000038	AIOG00000000	AIOG01000001-AIOG01000124
55037	AIOH01000005	AIOH00000000	AIOH01000001-AIOH01000104
LMG9879	AIOI01000021	AIOI00000000	AIOI01000001-AIOI01000129
86605	AIOJ01000014	AIOJ00000000	AIOJ01000001-AIOJ01000120
LMG23357	AIOK01000007	AIOK00000000	AIOK01000001-AIOK01000126
LMG9081	AIOM01000014	AIOM00000000	AIOM01000001-AIOM01000108
53161	AION01000013	AION00000000	AION01000001-AION01000123
LMG 9217	AIOO01000007	AIOO00000000	AIOO01000001-AIOO01000118
2008-1025	AIOP01000003	AIOP00000000	AIOP01000001-AIOP01000139
2008-894	AIOQ01000054	AIOQ00000000	AIOQ01000001-AIOQ01000116
2008-988	AIOS01000033	AIOS00000000	AIOS01000001-AIOS01000206
1997-1	AIOT01000012	AIOT00000000	AIOT01000001-AIOT01000112
2008-979	AIOU01000013	AIOU00000000	AIOU01000001-AIOU01000217
2008-831	AIOV01000005	AIOV00000000	AIOV01000001-AIOV01000120
1997-4	AIOW01000041	AIOW00000000	AIOW01000001-AIOW01000129
1997-7	AIOX01000035	AIOX00000000	AIOX01000001-AIOX01000076
1997-10	AIOY01000011	AIOY00000000	AIOY01000001-AIOY01000193
1997-11	AIOZ01000033	AIOZ00000000	AIOZ01000001-AIOZ01000115
1997-14	AIPA01000006	AIPA00000000	AIPA01000001-AIPA01000196
51037	AIPB01000032	AIPB00000000	AIPB01000001-AIPB01000207
110-21	AIPC01000014	AIPC00000000	AIPC01000001-AIPC01000113

87330	AIPD01000012	AIPD00000000	AIPD01000001-AIPD01000107
87459	AIPE01000032	AIPE00000000	AIPE01000001-AIPE01000201
140-16	AIPF01000026	AIPF00000000.	AIPF01000001-AIPF01000123
1213	AIPG01000033	AIPG00000000	AIPG01000001-AIPG01000125
1798	AIPIO1000018	AIPI00000000	AIPI01000001-AIPI01000106
1854	AIPJ01000006	AIPJ00000000	AIPJ01000001-AIPJ01000110
1893	AIPK01000044	AIPK00000000	AIPK01000001-AIPK01000101
1928	AIPL01000001	AIPL00000000	AIPL01000001-AIPL01000127.
LMG9872	AIPM01000030	AIPM00000000	AIPM01000001-AIPM01000091
LMG23210	AIPN01000024	AIPN00000000	AIPN01000001-AIPN01000164
LMG23211	AIPO01000006	AIPO00000000	AIPO01000001-AIPO01000112
CG8421	ABGQ01000004	ABGQ00000000	ABGQ01000001-ABGQ01000020

Appendix Table S6: *Campylobacter species* accession numbers for *peb3/glpT* locus

C. lari	***peb3/glpT* locus**	**Accession**	**Total Contigs**
RM1200	NC_012039	NC_012039	NC_012039-NC_012041
C. upsaliensis	***peb3/glpT* locus**	**Accession**	**Total Contigs**
JV21	AEPU01000011	AEPU00000000	AEPU01000001-AEPU01000049
RM3195	AAFJ01000001	AAFJ00000000	AAFJ01000001- AAFJ01000020
***C. coli* strain**	***peb3/glpT* locus**	**Accession**	**Total Contigs**
RM2282	AAFL01000003	AAFL00000000	AAFL01000001-AAFL01000038
JV20	AEER01000027	AEER0000000	AEER01000001-AEER01000034
86119	AIMU01000004	AIMU00000000	AIMU01000001-AIMU01000120
LMG23336	AINM01000033	AINM00000000	AINM01000001-AINM01000120
LMG23341	AINN01000004	AINN00000000	AINN01000001-AINN01000118
132-6	AINA01000029	AINA00000000	AINA01000001-AINA01000148
H56	AINW01000004	AINW00000000	AINW01000001-AINW01000134
202/204	AINH01000006	AINH00000000	AINH01000001-AINH01000104
2548	AIML01000019	AIML00000000	AIML01000001-AIML01000162
LMG23344	AINP01000028	AINP00000000	AINP01000001-AINP01000192
LMG9853	AINR01000007	AINR00000000	AINR01000001-AINR01000085
LMG9854	AINL01000001	AINL00000000	AINL01000001-AINL01000118
LMG9860	AINS010000039	AINS000000000	AINS01000001-AINS01000187
2680	AIMN01000006	AIMN00000000	AIMN01000001-AIMN01000128
2685	AIMO01000028	AIMO00000000	AIMO01000001-AIMO01000103
2688	AIMP01000017	AIMP00000000	AIMP01000001-AIMP01000209
2698	AIMR01000005	AIMR0000000	AIMR01000001-AIMR01000141
317/04	AINJ01000002	AINJ00000000	AINJ01000001-AINJ01000133
37/05	AINK01000040	AINK00000000	AINK01000001-AINK01000122
151-9	AINQ01000001	AINQ00000000	AINQ01000001-AINQ01000107
1909	AINC01000035	AINC00000000	AINC01000001-AINC01000133
1948	AINE01000046	AINE00000000	AINE00000001-AINE01000105
1957	AINF01000009	AINF00000000	AINF01000001-AINF01000123
1961	AING01000009	AING00000000	AING01000001- AING01000128
59-2	AIND01000004	AIND00000000	AIND01000001-AIND01000140
67-8	AINI01000007	AINI000000000	AINI01000001-AINI01000137
7-1	AIMZ01000004	AIMZ00000000	AIMZ01000001-AIMZ01000099
80352	AIMT01000064	AIMT00000000	AIMT01000001-AIMT01000260
90-3	AIMJ01000007	AIMJ00000000	AIMJ01000001- AIMJ01000121
H6	AINT01000001	AINT00000000	AINT01000001 AINT01000122
H8	AINU01000020	AINU00000000	AINU01000001- AINU01000191
H9	AINV01000003	AINV00000000	AINV01000001- AINV01000091
1091	AIMV01000025	AIMV00000000	AIMV01000001- AIMV01000132
1098	AIMW0100025	AIMW00000000	AIMW01000001- AIMW01000126
111-3	AIMI01000023	AIMI00000000	AIMI01000001- AIMI01000122
1417	AIMY01000021	AIMY00000000	AIMY01000001- AIMY01000097
1148	AIMX01000024	AIMX00000000	AIMX01000001- AIMX01000113
Z156	AINX01000002	AINX00000000	AINX00000001- AINX00000087
Z163	AIMK01000002	AIMK0000000	AIMK01000001- AIMK01000105
84-2	AIMS01000028	AIMS00000000	AIMS01000001- AIMS01000099

Figure S1 Nucleotide alignment of 64 *C. jejuni glpT* genes. The sequence in highlighted light grey areas indicate (-) deletions, highlighted/**bolded** letters indicate insertions and dark grey highlighting indicates nucleotide substitutions. Indel Regions 1 and 2 are boxed in blue. * indicate fully conserved residues; . indicate weak conservation; blank space indicates no conservation. The alignment was compiled using MAFFT in Clustal format and modified in Jalview. The conserved consensus sequence is given below the alignment.

```
2008-831        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
81-176          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
RM1221          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTTCTCTTTAGGAAATTTTA
S3              ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTTCTCTTTAGGAAATTTTA
NW              ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTTCTCTTTAGGAAATTTTA
2008-979        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
260.94          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
ICDCJ07001      ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
87459           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
LMG23269        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
51494           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
1997-14         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
33560           ATGTTTGATTTTTTCAAGCCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
1997-7          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
129-258         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
CG84-86         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
140-16          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
LMG23263        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
1798            ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
1893            ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
LMG23216        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
M1              ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
2008-894        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATGAAATTTCCTAAAGAGGAAATTTTA
LMG23357        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
LMG23211        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
55037           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
1997-1          ATGTTTGATTTTTTCAAACCTAAAGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
1854            ATGTTTGATTTTTTCAAACCTAAAGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
LMG23223        ATGTTTGATTTTTTCAAACCTAAAGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
LMG9872         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAAAGGAAATTTTA
81116           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
327             ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
LMG23210        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
CG84-21         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
P110B           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTTCTCTTTAGGAAATTTTA
LMG9081         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
1213            ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
110-21          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
1928            ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
84-25           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
86605           ATGTTTGATTTTTTCAAACCTAAGACAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
LMG23218        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
1997-10         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
LMG9217         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
LMG9879         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
DFVF1099        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
HB93-13         ACGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
87330           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
2008-1025       ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
11168H          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
LMG23264        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
60004           ATGTTTGATTTTTTCAAACCTAAGACAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
305             ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
IA3902          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
H22082          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
1997-11         ACGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
53161           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
2008-988        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
51037           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
D2600           ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
CF93-6          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
1997-4          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
                *.***************.*****..********** ******.**....***********60
Consensus       ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
```

```
2008-831        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
81-176          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
RM1221          CCTAGATATTACAAAATGAGAAATTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
S3              CCTAGATATTACAAAATGAGAAATTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
NW              CCTAGATATTACAAAATGAGAAATTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
2008-979        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
260.94          CCTAGATATTACAAAATGAGAAATTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
ICDCJ07001      CCTAGATATTACAAAATGAGAAATTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
87459           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23269        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
51494           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1997-14         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
33560           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGG-TTTTATAGGATATATGGGA
1997-7          CTTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
129-258         CTTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
CG84-86         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGAGTTTTTATAGGATATATGGGA
140-16          CTTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23263        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1798            CTTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1893            CTTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23216        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGAGTTTTTATAGGATATATGGGA
M1              CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
2008-894        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23357        CCTAGATATTATAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23211        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
55037           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1997-1          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1854            CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23223        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG9872         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
81116           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
327             CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23210        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
CG84-21         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
P110B           CCTAGATATTACAAAATGAGAAATTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG9081         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1213            CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
110-21          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1928            CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
84-25           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
86605           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23218        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1997-10         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG9217         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGATTTTTATAGGATATATGGGA
LMG9879         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
DFVF1099        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
HB93-13         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
87330           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
2008-1025       CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
11168H          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG23264        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
60004           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
305             CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
IA3902          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
H22082          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1997-11         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
53161           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
2008-988        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
51037           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
D2600           CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
CF93-6          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
1997-4          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
                *.*********.**********.***************..********************120

Consensus       CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
```

```
2008-831        TATTATCTTGTGCGAAATAA----------------------------------------
81-176          TATTATCTTGTGCGAAATAA----------------------------------------
RM1221          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
S3              TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
NW              TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
2008-979        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
260.94          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
ICDCJ07001      TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
87459           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
LMG23269        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
51494           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
1997-14         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
33560           TATTATCTTGTACGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
1997-7          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
129-258         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
CG84-86         TATTATCTTGTACGAAACAATATCACTCTTTCAACTCCATTTATACAAAATCAACTCGAA
140-16          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
LMG23263        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
1798            TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
1893            TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
LMG23216        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
M1              TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
2008-894        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
LMG23357        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
LMG23211        TATTATCTTGTACGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
55037           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
1997-1          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
1854            TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
LMG23223        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
LMG9872         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
81116           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
327             TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
LMG23210        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
CG84-21         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
P110B           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
LMG9081         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
1213            TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
110-21          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
1928            TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
84-25           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
86605           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
LMG23218        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
1997-10         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
LMG9217         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
LMG9879         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
DFVF1099        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
HB93-13         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
87330           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
2008-1025       TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
11168H          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
LMG23264        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
60004           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
305             TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
IA3902          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
H22082          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
1997-11         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
53161           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
2008-988        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
51037           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
D2600           TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
CF93-6          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
1997-4          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT 180
                ***********.*****.*************************************..**
Consensus       TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
```

```
2008-831        ------------------------------------------------------------
81-176          ------------------------------------------------------------
RM1221          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
S3              CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
NW              CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
2008-979        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
260.94          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
ICDCJ07001      CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
87459           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23269        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
51494           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1997-14         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
33560           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1997-7          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
129-258         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
CG84-86         CTTAGCAAGTCAGATATCGGGACAATCACAGGTTCCATGCTTATAGCCTATGGAATTAGC
140-16          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23263        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1798            CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1893            CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23216        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
M1              CTCAGCAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
2008-894        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23357        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23211        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
55037           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1997-1          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1854            CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23223        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG9872         CTCAGTAAATCGGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
81116           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
327             CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23210        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
CG84-21         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
P110B           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG9081         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1213            CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
110-21          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1928            CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
84-25           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
86605           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23218        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1997-10         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG9217         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG9879         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
DFVF1099        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
HB93-13         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
87330           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
2008-1025       CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
11168H          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG23264        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
60004           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
305             CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
IA3902          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
H22082          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1997-11         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
53161           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
2008-988        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
51037           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
D2600           CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
CF93-6          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1997-4          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC 240
                **.**.**.*.*********.***********.**.**************.*********

Consensus       CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
```

Indel Region 1

```
2008-831        ------------------------------------------------------------
81-176          ------------------------------------------------------------
RM1221          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
S3              AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
NW              AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
2008-979        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
260.94          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
ICDCJ07001      AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
87459           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
LMG23269        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
51494           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
1997-14         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
33560           AAAGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCCAAAAAAATATATGGCTTTAGG
1997-7          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA---AACATGGCTTTAGG
129-258         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA---AACATGGCTTTAGG
CG84-86         AAAGGCGCAATGAGCGTACTTAGTGATAAAGCTGATCCTAAAA-A-TACATGGCTTTAGG
140-16          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA---AACATGGCTTTAGG
LMG23263        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
1798            AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA---AACATGGCTTTAGG
1893            AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA---AACATGGCTTTAGG
LMG23216        AAAGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
M1              AAGGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
2008-894        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
LMG23357        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAA--AATACATGGCTTTAGG
LMG23211        AAGGGTGCAATGAGCGTTATAAGTGACAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
55037           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
1997-1          AAAGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
1854            AAAGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
LMG23223        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
LMG9872         AAGGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
81116           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
327             AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-AATACATGGCTTTAGG
LMG23210        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAA--AATACATGGCTTTAGG
CG84-21         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-A-TACATGGCTTTAGG
P110B           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAA--A-ATCATGGCTTTAGG
LMG9081         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-A-TACATGGCTTTAGG
1213            AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
110-21          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
1928            AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
84-25           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
86605           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
LMG23218        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAA--A-ATCATGGCTTTAGG
1997-10         AAGGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCTAAAA-A-TACATGGCTTTAGG
LMG9217         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-A-TACATGGCTTTAGG
LMG9879         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
DFVF1099        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
HB93-13         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
87330           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
2008-1025       AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
11168H          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
LMG23264        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
60004           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
305             AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
IA3902          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
H22082          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
1997-11         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
53161           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA---ATCATGGCTTTAGG
2008-988        AAGGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCTAAAA-A-TACATGGCTTTAGG
51037           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAA-AATACATGGCTTTAGG
D2600           AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
CF93-6          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG
1997-4          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAA-A-ATCATGGCTTTAGG 300
                *****.***********..*.**.**.*******.***.***......************

Consensus       AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAAAAAATCATGGCTTTAGG
```

```
2008-831        ------------------------------------------------------------
81-176          ------------------------------------------------------------
RM1221          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
S3              TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
NW              TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
2008-979        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
260.94          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
ICDCJ07001      TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
87459           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG23269        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
51494           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1997-14         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
33560           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1997-7          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
129-258         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
CG84-86         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
140-16          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG23263        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1798            TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1893            TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG23216        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
M1              TCTTATTTTATGTGCTCTTGTAAATGTTTTACTT--------------------------
2008-894        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG23357        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG23211        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
55037           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1997-1          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1854            TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG23223        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG9872         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
81116           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
327             TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGCTTTTCAAATTCTTTTTACGCTTA
LMG23210        TCTTATTTTATGTGCTCTTGCAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
CG84-21         TCTTATTTTATGCGCTCTTGCAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
P110B           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG9081         TCTTATTTTATGTGCTCTTGCAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1213            TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
110-21          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1928            TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
84-25           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
86605           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG23218        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1997-10         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG9217         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG9879         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
DFVF1099        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
HB93-13         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
87330           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
2008-1025       TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
11168H          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG23264        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
60004           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
305             TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
IA3902          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
H22082          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1997-11         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
53161           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
2008-988        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
51037           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
D2600           TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
CF93-6          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1997-4          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA 360
                ************.*******.***************.***********************
Consensus       TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
```

```
2008-831        ---------------------------------------------------------TAT
81-176          ---------------------------------------------------------TAT
RM1221          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
S3              TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
NW              TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
2008-979        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
260.94          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
ICDCJ07001      TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
87459           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG23269        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
51494           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
1997-14         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
33560           TGTAGGTTTTGTTATCTCGCTT-GGGTTTTTCAAGGTATGGGAGTGGATCCTTCTTTTAT
1997-7          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGATATGAGGGTAGGTCCTTCTTTTAT
129-258         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGATATGAGGGTAGGTCCTTCTTTTAT
CG84-86         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
140-16          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGATATGAGGGTAGGTCCTTCTTTTAT
LMG23263        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
1798            TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGATATGAGGGTAGGTCCTTCTTTTAT
1893            TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGATATGAGGGTAGGTCCTTCTTTTAT
LMG23216        TGTAGGTTTTGTTATCGCACTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
M1              ------------------------GGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
2008-894        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG23357        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG23211        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
55037           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
1997-1          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
1854            TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG23223        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG9872         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
81116           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
327             TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG23210        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
CG84-21         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
P110B           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG9081         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
1213            TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
110-21          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
1928            TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
84-25           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
86605           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG23218        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTTAAGGTATGGGGGTAGGTCCTTCTTTTAT
1997-10         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG9217         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG9879         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
DFVF1099        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
HB93-13         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
87330           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
2008-1025       TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
11168H          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG23264        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
60004           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
305             TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
IA3902          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
H22082          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
1997-11         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
53161           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
2008-988        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
51037           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
D2600           TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
CF93-6          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
1997-4          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
                ...............................................************420

Consensus       TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
```

```
2008-831        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
81-176          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
RM1221          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
S3              CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
NW              CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
2008-979        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
260.94          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
ICDCJ07001      CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
87459           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG23269        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
51494           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
1997-14         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
33560           CACTCTTGCAAACTGGTATCCTAAAAAAAGAGCGGGGAATTTACACGGCTGTTTGGAATA
1997-7          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGAGGGGAATTTACACGGCCGTTTGGAATA
129-258         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGAGGGGAATTTACACGGCCGTTTGGAATA
CG84-86         CACTCTTGCAAACTGGTATCCTAAAA-A-GAGCGGGGAATTTATACAGCTATTTGGAATA
140-16          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGAGGGGAATTTACACGGCCGTTTGGAATA
LMG23263        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGAGGGGAATTTACACGGCCGTTTGGAATA
1798            CACTCTTGCAAACTGGTATCCTAAAA-AAGAGAGGGGAATTTACACGGCCGTTTGGAATA
1893            CACTCTTGCAAACTGGTATCCTAAAA-AAGAGAGGGGAATTTACACGGCCGTTTGGAATA
LMG23216        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCATTTGGAATA
M1              CACTCTTGCAAACTAGTATCCTAAAA-A-GAGCGAGGAATTTACACGGCCGTTTGGAATA
2008-894        CACTCTTGCAAACTGGTATCCCAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG23357        CACTCTTGCAAACTGGTATCCCAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG23211        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
55037           CACTCTTGCAAACTGGTATCCTAAA--AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
1997-1          CACTCTTACAAACTGATATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
1854            CACTCTTACAAACTGATATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG23223        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG9872         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
81116           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
327             CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG23210        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
CG84-21         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
P110B           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG9081         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
1213            CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
110-21          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
1928            CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
84-25           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
86605           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG23218        CACTCTTGCAAACCGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
1997-10         CACTCTTGCGAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG9217         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG9879         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
DFVF1099        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
HB93-13         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
87330           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
2008-1025       CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
11168H          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG23264        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
60004           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
305             CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
IA3902          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
H22082          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
1997-11         CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCATTTGGAATA
53161           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
2008-988        CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
51037           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
D2600           CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
CF93-6          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
1997-4          CACTCTTGCAAACTGGTATCCTAAAA-AAGAGCGGGGAATTTACACGGCCGTTTGGAATA
                *******.*.***...*****......*.***..*.********.****...******** 480
Consensus       CACTCTTGCAAACTGGTATCCTAAAAAAAGAGCGGGGAATTTACACGGCCGTTTGGAATA
```

```
2008-831        TCTCACATAATATAGGAGGTGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
81-176          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
RM1221          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
S3              TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
NW              TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
2008-979        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
260.94          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
ICDCJ07001      TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
87459           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG23269        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
51494           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
1997-14         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
33560           TCTCACATAATATAGGAGGCGAGATAGTTGCTCCTATAGTTTCACTTTCAGGTTTTGCTT
1997-7          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTGTCACTTTCTGGATTTGCTT
129-258         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTGTCACTTTCTGGATTTGCTT
CG84-86         TCTCGCACAATTTAGGAGGAGGTATAATCGCTCCTATAGTGTCGCTTTCTGGATTTGCTT
140-16          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTGTCACTTTCTGGATTTGCTT
LMG23263        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTGTCACTTTCTGGATTTGCTT
1798            TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTGTCACTTTCTGGATTTGCTT
1893            TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTGTCACTTTCTGGATTTGCTT
LMG23216        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
M1              TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCTGGATTTGCTT
2008-894        TCTCACATAATATAGGAGGAGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG23357        TCTCACATAATATAGGAGGAGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG23211        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
55037           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
1997-1          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
1854            TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG23223        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG9872         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
81116           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
327             TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG23210        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
CG84-21         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
P110B           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG9081         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
1213            TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
110-21          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
1928            TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
84-25           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
86605           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG23218        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
1997-10         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG9217         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG9879         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATA-------------GTTTTGCTT
DFVF1099        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
HB93-13         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
87330           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
2008-1025       TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
11168H          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG23264        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
60004           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
305             TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
IA3902          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
H22082          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
1997-11         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
53161           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
2008-988        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
51037           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
D2600           TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
CF93-6          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
1997-4          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
                ****.**.***.******* *. ***.*.*********.............* ******* 540

Consensus       TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
```

```
2008-831        TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
81-176          TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
RM1221          TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
S3              TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
NW              TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
2008-979        TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
260.94          TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
ICDCJ07001      TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
87459           TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
LMG23269        TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
51494           TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
1997-14         TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
33560           TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
1997-7          TAGCAGCATTTTTAGGCGTAAGCTTAGCAGATTTTAATGAAACACATTAGCATATCAATC
129-258         TAGCAGCATTTTTAGGCGTAAGCTTAGCAGATTTTAATGAAACACATTAGCATATCAATC
CG84-86         TAGCAGCATTTTTAGGCGTAAGCTTGGCAGATTTTAATGAAACACATTAGCATATCAATC
140-16          TAGCAGCATTTTTAGGCGTAAGCTTAGCAGATTTTAATGAAACACATTAGCATATCAATC
LMG23263        TAGCAGCATTTTTAGGCGTAAGCTTAGCAGATTTTAATGAAACACATTAGCATATCAATC
1798            TAGCAGCATTTTTAGGCGTAAGCTTAGCAGATTTTAATGAAACACATTAGCATATCAATC
1893            TAGCAGCATTTTTAGGCGTAAGCTTAGCAGATTTTAATGAAACACATTAGCATATCAATC
LMG23216        TAGCAGCATTATTGGGTGTAAGTATGGCAGATTTTAACGAAACATATTGGCACATGAATC
M1              TAGCAGCATTTTTAGGCGTAAGCTTGGCAGATTTTAATGAAACACATTAGCATATCAATC
2008-894        TAGCAGCATTATTGGGTGTGAGTATGGCAGATTTTAACGAAACATATTGGCACATGAATC
LMG23357        TAGCAGCATTATTGGGTGTAAGTATGGCAGATTTTAACGAAACATATTGGCACATGAATC
LMG23211        TAGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
55037           TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
1997-1          TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
1854            TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
LMG23223        TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
LMG9872         TGGCAGCGTTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
81116           TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
327             TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
LMG23210        TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACAAAACATATTGACACATGAATC
CG84-21         TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
P110B           TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
LMG9081         TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGACACATGAATC
1213            TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
110-21          TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
1928            TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
84-25           TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
86605           TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
LMG23218        TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
1997-10         TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
LMG9217         TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
LMG9879         TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
DFVF1099        TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
HB93-13         TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
87330           TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
2008-1025       TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
11168H          TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
LMG23264        TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
60004           TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
305             TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
IA3902          TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
H22082          TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
1997-11         TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
53161           TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAATGAAACATATTGGCACATGAATC
2008-988        TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
51037           TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
D2600           TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
CF93-6          TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
1997-4          TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
                *.*****.**.**.**.**.**..*.**..*******..*****.***..**.**.**** 600
Consensus       TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
```

```
2008-831        ATGTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
81-176          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
RM1221          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
S3              ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
NW              ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
2008-979        ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
260.94          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
ICDCJ07001      ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
87459           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
LMG23269        ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
51494           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
1997-14         ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
33560           ATTTTTATGTTCCTGCTGTTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
1997-7          ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
129-258         ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
CG84-86         ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
140-16          ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
LMG23263        ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
1798            ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
1893            ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
LMG23216        ATTTTTATGCTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
M1              ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
2008-894        ATTTTTATGCTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCGGTTA
LMG23357        ATTTTTATGCTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCGGTTA
LMG23211        ATTTTTATGTTCCTGCTGTTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
55037           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
1997-1          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
1854            ATTTTTACACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
LMG23223        ATTTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
LMG9872         ATTTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
81116           ATTTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
327             ATTTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
LMG23210        ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
CG84-21         ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
P110B           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
LMG9081         ATTTTTATACTCCTGCTACTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
1213            ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
110-21          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
1928            ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
84-25           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
86605           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
LMG23218        ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
1997-10         ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
LMG9217         ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
LMG9879         ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
DFVF1099        ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
HB93-13         ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
87330           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
2008-1025       ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
11168H          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
LMG23264        ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
60004           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
305             ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
IA3902          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
H22082          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
1997-11         ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
53161           ATTTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
2008-988        ATGTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
51037           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
D2600           ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
CF93-6          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
1997-4          ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
                **.*****..*******..*****.****.*******.*******.*********.****660
Consensus       ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
```

```
2008-831        AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
81-176          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
RM1221          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
S3              AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
NW              AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
2008-979        AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
260.94          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
ICDCJ07001      AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
87459           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG23269        AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
51494           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1997-14         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
33560           AAGGAAGT-CTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1997-7          AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTAGCGAAATCAACGAAATGAGAGGTA
129-258         AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTAGCGAAATCAACGAAATGAGAGGTA
CG84-86         AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
140-16          AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTAGCGAAATCAACGAAATGAGAGGTA
LMG23263        AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTAGCGAAATCAACGAAATGAGAGGTA
1798            AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTAGCGAAATCAACGAAATGAGAGGTA
1893            AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTAGCGAAATCAACGAAATGAGAGGTA
LMG23216        AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
M1              AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
2008-894        AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG23357        AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG23211        AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
55037           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1997-1          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1854            AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG23223        AAGGAAGTCCTAAAAATGAAGGTTTGATTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG9872         AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
81116           AAGGAAGTCCTAAAAATGAAGGTTTGATTGATATTACTGAAATCAATGAAATGCGAGGGA
327             AAGGAAGTCCTAAAAATGAAGGTTTGATTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG23210        AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
CG84-21         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
P110B           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG9081         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1213            AAGGAAA--CTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
110-21          AAGGAAA--CTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1928            AAGGAAA--CTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
84-25           AAGGAAA--CTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
86605           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG23218        AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1997-10         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG9217         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG9879         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
DFVF1099        AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
HB93-13         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
87330           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
2008-1025       AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
11168H          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG23264        AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
60004           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
305             AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
IA3902          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
H22082          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1997-11         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
53161           AAGGAAGTCCTAAAAATGAAGGTTTGATTAATATTACTGAAATCAATGAAATGCGAGGGA
2008-988        AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
51037           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
D2600           AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
CF93-6          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1997-4          AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
                ******...*****************.**.******..********.******.****.* 720
Consensus       AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
```

```
2008-831        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAACAGTTTTGAAATTTTTT
81-176          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
RM1221          TTAAAACAGAAGAAATCAAAACTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
S3              TTAAAACAGAAGAAATCAAAACTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
NW              TTAAAACAGAAGAAATCAAAACTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
2008-979        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
260.94          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
ICDCJ07001      TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
87459           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG23269        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
51494           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
1997-14         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
33560           TTAAAACAGAAGAAATCAAAACTGTAGAAAGTCCAAATTTAAGCAGTTTTGAAATTTTTT
1997-7          TTAAAACCGAAGAAATTAAAGCTGTAGAAAGTCCAAATTTAAGCGGTCTTGAAATTTTTT
129-258         TTAAAACCGAAGAAATTAAAGCTGTAGAAAGTCCAAATTTAAGCGGTCTTGAAATTTTTT
CG84-86         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
140-16          TTAAAACCGAAGAAATTAAAGCTGTAGAAAGTCCAAATTTAAGCGGTCTTGAAATTTTTT
LMG23263        TTAAAACCGAAGAAATTAAAGCTGTAGAAAGTCCAAATTTAAGCGGTCTTGAAATTTTTT
1798            TTAAAACCGAAGAAATTAAAGCTGTAGAAAGTCCAAATTTAAGCGGTCTTGAAATTTTTT
1893            TTAAAACCGAAGAAATTAAAGCTGTAGAAAGTCCAAATTTAAGCGGTCTTGAAATTTTTT
LMG23216        TTAAAACAGAAGAAATCAAAGCTGTAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
M1              TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
2008-894        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG23357        TTAAAACAGAAGAAATCAAAGCTGTAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG23211        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
55037           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
1997-1          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
1854            TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG23223        TTAAAACAGAAGAAATCAAAGCTGTAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG9872         TTAAAACAGAAGAAATCAAAGCTGTAGAAAGTCCAAATTTAAGCAGTTTTGAAATTTTTT
81116           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
327             TTAAAACAGAAGAAATCAAAGCTGTAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG23210        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
CG84-21         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
P110B           TTAAAACAGAAGAAATCAAAACTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG9081         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
1213            TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
110-21          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
1928            TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
84-25           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
86605           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAACAGTTTTGAAATTTTTT
LMG23218        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
1997-10         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCGGTTTTGAAATTTTTT
LMG9217         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG9879         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
DFVF1099        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
HB93-13         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
87330           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
2008-1025       TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
11168H          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG23264        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
60004           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
305             TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
IA3902          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
H22082          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
1997-11         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
53161           TTAAAACAGAAGAAATCAAAGCTGTAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
2008-988        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTCT
51037           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
D2600           TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
CF93-6          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
1997-4          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
                *******.*******.***.***.******.***********.*.**.**********.* 780
Consensus       TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
```

```
2008-831        ATCATTATGTACTTAAAAACAAAAATGCTTAGCATGTGGCTTGGATGGATACTTTTGTTT
81-176          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
RM1221          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTCGCTTGGATGGATACTTTTGTTT
S3              ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTCGCTTGGATGGATACTTTTGTTT
NW              ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTCGCTTGGATGGATACTTTTGTTT
2008-979        ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
260.94          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
ICDCJ07001      ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
87459           ATCATTATGTACTTAAAAACAAAAATGCTTGGCATGTGGCTTGAATGGATACTTTTGTTT
LMG23269        ATCATTATGTACTTAAAAACAAAAATGCTTGGCATGTGGCTTGAATGGATACTTTTGTTT
51494           ATCATTATGTACTTAAAAACAAAAATACTTGGCATGTGGCTTGAATGGATACTTTTGTTT
1997-14         ATCATTATGTACTTAAAAACAAAAATACTTGGCATGTGGCTTGAATGGATACTTTTGTTT
33560           ATCGTTATGTGCTTAAAAATAAAAATGCTTGGTATGTGGATTGGATGGATACTTTTGTTT
1997-7          ATCATTATGTCCTTAAAAATAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
129-258         ATCATTATGTCCTTAAAAATAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
CG84-86         ATCATTATGTACTTAAAAACAAAAATGCTTAGTATGTGGCTTGGATGGATACTTTTGTTT
140-16          ATCATTATGTCCTTAAAAATAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
LMG23263        ATCATTATGTCCTTAAAAATAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
1798            ATCATTATGTCCTTAAAAATAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
1893            ATCATTATGTCCTTAAAAATAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
LMG23216        ATCGTTATGTGCTTAAAAATAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
M1              ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
2008-894        ATCGTTATGTGCTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
LMG23357        ATCGTTATGTGCTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
LMG23211        ATCGTTATGTGCTTAAAAATAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
55037           ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
1997-1          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
1854            ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTATTT
LMG23223        ATCGTTATGTGCTTAAACACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
LMG9872         ATCGTTATGTGCTTAAAAATAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
81116           ATCGTTATGTGCTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
327             ATCGTTATGTGCTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
LMG23210        ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
CG84-21         ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
P110B           ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTCGCTTGGATGGATACTTTTGTTT
LMG9081         ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
1213            ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
110-21          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
1928            ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
84-25           ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
86605           ATCATTATGTACTTAAAAACAAAAATGCTTAGTATGTGGCTTGGATGGATACTTTTGTTT
LMG23218        ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
1997-10         ATCATTATGTACTTAAAAACAAAAATGCTTGGCATGTGGCTTGGATGGATACTTTTGTTT
LMG9217         ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
LMG9879         ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
DFVF1099        ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
HB93-13         ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
87330           ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
2008-1025       ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
11168H          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
LMG23264        ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
60004           ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
305             ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
IA3902          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
H22082          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
1997-11         ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
53161           ATCGTTATGTGCTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
2008-988        ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
51037           ATCATTATGTACTTAAAAACAAAAATGCTTAGTATGTGGCTTGGATGGATACTTTTGTTT
D2600           ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
CF93-6          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
1997-4          ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
                ***.******.********.******.***.*.***.**.****************.*** 840
Consensus       ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
```

```
2008-831        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
81-176          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
RM1221          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
S3              ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
NW              ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
2008-979        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
260.94          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
ICDCJ07001      ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
87459           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23269        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
51494           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1997-14         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
33560           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTACTAGAAACAAAAGGCT
1997-7          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
129-258         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
CG84-86         ATATGGTGCGTTTTAGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
140-16          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23263        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1798            ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1893            ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23216        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
M1              ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
2008-894        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23357        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23211        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
55037           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1997-1          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1854            ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23223        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG9872         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
81116           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
327             ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23210        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
CG84-21         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
P110B           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG9081         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1213            ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
110-21          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1928            ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
84-25           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
86605           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23218        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1997-10         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG9217         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG9879         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
DFVF1099        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
HB93-13         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
87330           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
2008-1025       ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
11168H          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG23264        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
60004           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
305             ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
IA3902          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
H22082          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1997-11         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
53161           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
2008-988        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
51037           ATATGGTGCATTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
D2600           ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
CF93-6          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
1997-4          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
                *********.****.****************************.**************** 900

Consensus       ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
```

```
2008-831        TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
81-176          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
RM1221          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
S3              TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
NW              TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
2008-979        TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
260.94          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
ICDCJ07001      TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
87459           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG23269        TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
51494           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1997-14         TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
33560           TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
1997-7          TTAATAAAGAACAAATAGGTATTGCCTT-TGGCTTTTTGAATGGGCTGCTATACCTTCTA
129-258         TTAATAAAGAACAAATAGGTATTGCCTT-TGGCTTTTTGAATGGGCTGCTATACCTTCTA
CG84-86         TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATAGGCTGCTATACCTTCTA
140-16          TTAATAAAGAACAAATAGGTATTGCCTT-TGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG23263        TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1798            TTAATAAAGAACAAATAGGTATTGCCTT-TGGCTTTTTGAATGGGCTGCTATACCTTCTA
1893            TTAATAAAGAACAAATAGGTATTGCCTT-TGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG23216        TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
M1              TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
2008-894        TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
LMG23357        TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
LMG23211        TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
55037           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1997-1          TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1854            TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG23223        TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
LMG9872         TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
81116           TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
327             TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
LMG23210        TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
CG84-21         TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
P110B           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG9081         TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1213            TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
110-21          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1928            TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
84-25           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
86605           TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG23218        TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1997-10         TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG9217         TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG9879         TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
DFVF1099        TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
HB93-13         TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
87330           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
2008-1025       TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
11168H          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
LMG23264        TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
60004           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
305             TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
IA3902          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
H22082          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1997-11         TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
53161           TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATAGGCTGCTATACCTTCTA
2008-988        TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATAGGCTGCTATACCTTCTA
51037           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
D2600           TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
CF93-6          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
1997-4          TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
                ****************.***********.*************.******.********** 960
Consensus       TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
```

```
2008-831        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
81-176          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
RM1221          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
S3              CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
NW              CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
2008-979        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
260.94          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
ICDCJ07001      CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
87459           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23269        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
51494           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1997-14         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
33560           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1997-7          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
129-258         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
CG84-86         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
140-16          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23263        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1798            CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1893            CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23216        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
M1              CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
2008-894        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23357        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23211        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
55037           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1997-1          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1854            CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23223        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG9872         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
81116           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
327             CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23210        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
CG84-21         CTTTACTTGCAGGATATATTTCAGATAAAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
P110B           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG9081         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1213            CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
110-21          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1928            CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
84-25           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
86605           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23218        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1997-10         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG9217         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG9879         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
DFVF1099        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
HB93-13         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
87330           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
2008-1025       CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
11168H          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
LMG23264        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
60004           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
305             CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
IA3902          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
H22082          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1997-11         CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAAAATGCCTCCTGCT
53161           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
2008-988        CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
51037           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
D2600           CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
CF93-6          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
1997-4          CTTTACTTGCAGGATATATTTCAGAT-AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
                ************************** ******************.************* 1020
Consensus       CTTTACTTGCAGGATATATTTCAGAT AAAATTTTTAAAGGCTATAGAATGCCTCCTGCT
```

```
2008-831        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
81-176          ATAGGCGCCATAGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
RM1221          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
S3              ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
NW              ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
2008-979        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
260.94          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
ICDCJ07001      ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
87459           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG23269        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
51494           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1997-14         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
33560           ATAGGTGCTATAGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1997-7          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
129-258         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
CG84-86         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
140-16          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG23263        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1798            ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1893            ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG23216        ATAGGTGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
M1              ATAGGTGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
2008-894        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG23357        ATAGGTGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG23211        ATAGGTGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
55037           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1997-1          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1854            ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG23223        ATAGGTGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG9872         ATAGGCGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
81116           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
327             ATAGGTGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG23210        ATAGGCGCCATGGTGATTATTTTTTTTATGATTATTGGATATTTTACTTCAAATAATCTT
CG84-21         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
P110B           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG9081         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1213            ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
110-21          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1928            ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
84-25           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
86605           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGAATATTTTACTTCAAATAATCTT
LMG23218        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1997-10         ATAGGCGCCATAGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG9217         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG9879         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
DFVF1099        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
HB93-13         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
87330           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
2008-1025       ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
11168H          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
LMG23264        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
60004           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
305             ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
IA3902          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
H22082          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1997-11         ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
53161           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
2008-988        ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
51037           ATAGGTGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
D2600           ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
CF93-6          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
1997-4          ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
                *****.**.**.***********.*************.********************** 1080

Consensus       ATAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTT
```

```
2008-831        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
81-176          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
RM1221          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
S3              T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
NW              T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
2008-979        T-ATATGGTCATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
260.94          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
ICDCJ07001      T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
87459           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23269        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
51494           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
1997-14         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
33560           T-ATATGATTACTTTCTTTGCAGCTATGGCGGGGTATTTGGTCTATATACCACAATTTTT
1997-7          T-ATATGGTTATTTTCTTTGCAGCTATGGCAGGGTGTTTAGTCTATATACCACAATTTTT
129-258         T-ATATGGTTATTTTCTTTGCAGCTATGGCAGGGTGTTTAGTCTATATACCACAATTTTT
CG84-86         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
140-16          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23263        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
1798            T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
1893            T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23216        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
M1              T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
2008-894        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23357        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23211        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTGGTCTATATACCA-AATTTTT
55037           TGATATGGTTATTTTCTTTGCAGCTATGGCAGGGTGTTTAGTCTATATACCACAATTTTT
1997-1          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
1854            T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23223        T-ATATGGTTATTTTCTTTGCAGCTATGGC--GGTGTTTAGTCTATATACCACAATTTTT
LMG9872         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTGGTCTATATACCACAATTTTT
81116           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
327             T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23210        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
CG84-21         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
P110B           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG9081         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
1213            T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
110-21          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
1928            T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
84-25           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
86605           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23218        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
1997-10         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG9217         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG9879         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
DFVF1099        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATATCACAATTTTT
HB93-13         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
87330           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
2008-1025       T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
11168H          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
LMG23264        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
60004           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
305             T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
IA3902          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
H22082          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
1997-11         T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
53161           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
2008-988        T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
51037           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
D2600           T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGATGTTTAGTCTATATACCACAATTTTT
CF93-6          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGATGTTTAGTCTATATACCACAATTTTT
1997-4          T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGATGTTTAGTCTATATACCACAATTTTT
                * *****.*.*.******************..*.*.***.*********.**.******* 1140
Consensus       T-ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTT
```

```
2008-831        AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
81-176          AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGAATCTTGTGTAGGACT
RM1221          AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
S3              AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
NW              AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
2008-979        AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
260.94          AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
ICDCJ07001      AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
87459           AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG23269        AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
51494           AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
1997-14         AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
33560           AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
1997-7          AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATT
129-258         AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATT
CG84-86         AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
140-16          AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATT
LMG23263        AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGATT
1798            AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATT
1893            AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATT
LMG23216        AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
M1              AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATT
2008-894        AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG23357        AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG23211        AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
55037           AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
1997-1          AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
1854            AGCAAGTGTGTAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG23223        AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
LMG9872         AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
81116           AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
327             AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
LMG23210        -GCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
CG84-21         AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
P110B           AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG9081         AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
1213            AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
110-21          AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
1928            AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
84-25           AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
86605           AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG23218        AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
1997-10         AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG9217         AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG9879         AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
DFVF1099        AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
HB93-13         AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
87330           AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
2008-1025       AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
11168H          AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
LMG23264        AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
60004           AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
305             AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
IA3902          AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
H22082          AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
1997-11         AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
53161           AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
2008-988        AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATT
51037           AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
D2600           AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
CF93-6          AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
1997-4          AGCAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGCGTAGGACT
                .******.**.**********************************.******.******.* 1200

Consensus       AGCAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACT
```

Indel Region 2

```
2008-831        TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
81-176          TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
RM1221          TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCGCTTGGAACAAAAG--CATAGGTTGGG
S3              TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCGCTTGGAACAAAAG--CATAGGTTGGG
NW              TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCGCTTGGAACAAAAG--CATAGGTTGGG
2008-979        TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
260.94          TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
ICDCJ07001      TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
87459           TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
LMG23269        TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
51494           TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
1997-14         TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
33560           TCGTGGTTTTATGAGTTATATAGTCGGCGCTTCACTTGGAACAAAGGCTTATAGGTTGGG
1997-7          TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
129-258         TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
CG84-86         TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
140-16          TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
LMG23263        TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
1798            TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
1893            TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
LMG23216        TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAGGC-TATAGGTTGGG
M1              TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
2008-894        TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
LMG23357        TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
LMG23211        TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
55037           TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
1997-1          TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
1854            TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
LMG23223        TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
LMG9872         TCGTGGCTTTATGAGCTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
81116           TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
327             TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
LMG23210        TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGA
CG84-21         TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
P110B           TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCGCTTGGAACAAAAG--CATAGGTTGGG
LMG9081         TCGTGGCTTTATGAGTTATGTAGTTGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
1213            TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
110-21          TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
1928            TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
84-25           TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
86605           TCGTGGCTTTATGAGTTATGTAGTTGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
LMG23218        TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
1997-10         TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
LMG9217         TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
LMG9879         TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
DFVF1099        TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
HB93-13         TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
87330           TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
2008-1025       TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
11168H          TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
LMG23264        TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
60004           TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
305             TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
IA3902          TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
H22082          TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
1997-11         TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
53161           TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
2008-988        TCGTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGG
51037           TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
D2600           TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
CF93-6          TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
1997-4          TCGTGGCTTTATGAGTTATGTAGTCGGTGCTTCACTTGGAACAAAAGC-TATAGGCTGGG
                **.***.********.***.****.**.*****.***********.**-.*****.***. 1260
Consensus       TCATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAG--CATAGGTTGGG
```

```
2008-831        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
81-176          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
RM1221          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
S3              CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
NW              CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
2008-979        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
260.94          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATAC
ICDCJ07001      CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
87459           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23269        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
51494           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1997-14         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
33560           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTC------
1997-7          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTC------
129-258         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTC------
CG84-86         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
140-16          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23263        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1798            CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1893            CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23216        CTGTGGATTACTATGGTAGCTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
M1              CTGTGGATTACTATGGTAGTTGGAATACTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
2008-894        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23357        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23211        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
55037           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1997-1          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1854            CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23223        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG9872         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGCGCTTGTATTC
81116           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
327             CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23210        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
CG84-21         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
P110B           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG9081         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1213            CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
110-21          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1928            CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
84-25           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
86605           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23218        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1997-10         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG9217         CTGTGGATTGCTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG9879         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
DFVF1099        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
HB93-13         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
87330           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
2008-1025       CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
11168H          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
LMG23264        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
60004           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
305             CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
IA3902          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
H22082          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
1997-11         CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
53161           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
2008-988        CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
51037           CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
D2600           CTGTGGATTACTATGGTAGTTAGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
CF93-6          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATAC
1997-4          CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATAC
                *********.*********.*.****.**********************.***...*...1320

Consensus       CTGTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTC
```

```
2008-831        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
81-176          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
RM1221          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCACTAAAAAGAAGCATCAATAA
S3              TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCACTAAAAAGAAGCATCAATAA
NW              TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCACTAAAAAGAAGCATCAATAA
2008-979        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
260.94          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
ICDCJ07001      TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
87459           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
LMG23269        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
51494           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1997-14         TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
33560           ----------------------------------------AAAAGAAGTATCAATAA
1997-7          ----------------------------------------AAAAGAAGTATCAATAA
129-258         ----------------------------------------AAAAGAAGTATCAATAA
CG84-86         TTTGTATACTTTATTCTACTTTATGTCATTTTAGCACTAAAAAGAAGCATCAATAA
140-16          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
LMG23263        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1798            TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1893            TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
LMG23216        TTTGCATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAGAAAAAATACGCTTAA
M1              TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
2008-894        TTTGTATACTTTGCTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
LMG23357        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
LMG23211        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCCAAAAAGAAGTATCAATAA
55037           TTTGTATACTTTGTTCTAGTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1997-1          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1854            TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
LMG23223        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
LMG9872         TTTGCATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAGAAGAAGTATCAATAA
81116           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
327             TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
LMG23210        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
CG84-21         TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCACTAAAAAGAAGCATCAATAA
P110B           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCACTAAAAAGAAGCATCAATAA
LMG9081         TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1213            TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
110-21          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1928            TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
84-25           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
86605           TTTGTATACTTTGTTCTACTTTATGTTATTTTAGCGCTAAAAAGAAGCATCAATAA
LMG23218        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1997-10         TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
LMG9217         TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCACTAAAAAGAAGCATCAATAA
LMG9879         TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
DFVF1099        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
HB93-13         TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
87330           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
2008-1025       TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
11168H          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
LMG23264        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
60004           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
305             TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
IA3902          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
H22082          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
1997-11         TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
53161           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
2008-988        TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
51037           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
D2600           TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
CF93-6          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
1997-4          TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
                ..........................................................*.......*** 1376

Consensus       TTTGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
```

Figure S2 Nucleotide alignment of the deletion containing *glpT* genes of LMG9879, M1, 33560, 1997-7, 129-258, 81-176 and 2008-831 with the 81116 *glpT*. Highlighted light grey areas indicate (-) deletions, bolded nucleotide is a insert, dark grey indicates nucleotide subsitutions. * indicate fully conserved residues, . indicate weak conservations, and blank indicates no conservation.

```
81116          ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
LMG9879        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
M1             ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
33560          ATGTTTGATTTTTTCAAGCCTAAGGCAAAAGCTATTAAATTTCCTAAAGAGGAAATTTTA
1997-7         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
129-258        ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
81-176         ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
2008-831       ATGTTTGATTTTTTCAAACCTAAGGCAAAAGCTATTAAATTTCCTCTTTAGGAAATTTTA
               *****************.****************************....***********60

81116          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
LMG9879        CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
M1             CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
33560          CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGG-TTTTATAGGATATATGGGA
1997-7         CTTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
129-258        CTTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
81-176         CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
2008-831       CCTAGATATTACAAAATGAGAACTTGGAGTTTAAGTGGGGTTTTTATAGGATATATGGGA
               *.*******************************************************120

81116          TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
LMG9879        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
M1             TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
33560          TATTATCTTGTACGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTGAT
1997-7         TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
129-258        TATTATCTTGTGCGAAATAATATCACTCTTTCAACTCCATTTATACAAAATCAACTTAAT
81-176         TATTATCTTGTGCGAAATAA---------------------------------------
2008-831       TATTATCTTGTGCGAAATAA---------------------------------------180
               ***********.********.......................................

81116          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
LMG9879        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
M1             CTCAGCAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
33560          CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
1997-7         CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
129-258        CTCAGTAAATCAGACATCGGAACAATCACAGGCTCAATGCTTATAGCTTACGGGATTAGC
81-176         ------------------------------------------------------------
2008-831       ------------------------------------------------------------240
               ............................................................

81116          AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAAAA-TACATGGCTTTAGG
LMG9879        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGATCCTAAAAAA-T-CATGGCTTTAGG
M1             AAGGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCTAAAAAA-TACATGGCTTTAGG
33560          AAAGGTGCAATGAGCGTTATAAGCGATAAAGCTGATCCCAAAAAAATATATGGCTTTAGG
1997-7         AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAAAA---CATGGCTTTAGG
129-258        AAGGGTGCAATGAGCGTTATAAGCGACAAAGCTGGTCCTAAAAAA---CATGGCTTTAGG
81-176         ------------------------------------------------------------
2008-831       ------------------------------------------------------------300
               ..................... ......................................

81116          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
LMG9879        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
M1             TCTTATTTTATGTGCTCTTGTAAATGTTTTACTT---------------------------
33560          TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
1997-7         TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
129-258        TCTTATTTTATGTGCTCTTGTAAATGTTTTACTTGGTTTTTCAAATTCTTTTTACGCTTA
81-176         ------------------------------------------------------------
2008-831       ------------------------------------------------------------360

81116          TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
LMG9879        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
M1             ------------------------GGTTTTTCAAGGTATGGGGGTAGGTCCTTCTTTTAT
33560          TGTAGGTTTTGTTATCTCGCTT-GGGTTTTTCAAGGTATGGGAGTGGATCCTTCTTTTAT
1997-7         TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGATATGAGGGTAGGTCCTTCTTTTAT
129-258        TGTAGGTTTTGTTATCGCGCTTGGGGTTTTTCAAGATATGAGGGTAGGTCCTTCTTTTAT
81-176         ------------------------------------------------------------TAT
2008-831       ------------------------------------------------------------TAT
               ..................... ....................................***420
```

```
81116          CACTCTTGCAAACTGGTATCCT-AAAAAAGAGCGGGGAATTTACACGGCCGTTTGGAATA
LMG9879        CACTCTTGCAAACTGGTATCCT-AAAAAAGAGCGGGGAATTTACACGGCCGTTTGGAATA
M1             CACTCTTGCAAACTAGTATCCT--AAAAAGAGCGAGGAATTTACACGGCCGTTTGGAATA
33560          CACTCTTGCAAACTGGTATCCTAAAAAAAGAGCGGGGAATTTACACGGCTGTTTGGAATA
1997-7         CACTCTTGCAAACTGGTATCCT-AAAAAAGAGAGGGGAATTTACACGGCCGTTTGGAATA
129-258        CACTCTTGCAAACTGGTATCCT-AAAAAAGAGAGGGGAATTTACACGGCCGTTTGGAATA
81-176         CACTCTTGCAAACTGGTATCCT-AAAAAAGAGCGGGGAATTTACACGGCCGTTTGGAATA
2008-831       CACTCTTGCAAACTGGTATCCT-AAAAAAGAGCGGGGAATTTACACGGCCGTTTGGAATA
               **************.*******  ********.*.**************.**********480

81116          TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
LMG9879        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATA--------------GTTTTGCTT
M1             TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCTGGATTTGCTT
33560          TCTCACATAATATAGGAGGCGAGATAGTTGCTCCTATAGTTTCACTTTCAGGTTTTGCTT
1997-7         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTGTCACTTTCTGGATTTGCTT
129-258        TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTGTCACTTTCTGGATTTGCTT
81-176         TCTCACATAATATAGGAGGCGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT
2008-831       TCTCACATAATATAGGAGGTGGGATAGTTGCTCCTATAGTTTCGCTTTCAGGTTTTGCTT540
               *******************.*.****************.............*.*******

81116          TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
LMG9879        TGGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
M1             TAGCAGCATTTTTAGGCGTAAGCTTGGCAGATTTTAATGAAACACATTAGCATATCAATC
33560          TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
1997-7         TAGCAGCATTTTTAGGCGTAAGCTTAGCAGATTTTAATGAAACACATTAGCATATCAATC
129-258        TAGCAGCATTTTTAGGCGTAAGCTTAGCAGATTTTAATGAAACACATTAGCATATCAATC
81-176         TAGCAGCATTATTGGGTGTGAGTATGGCAAATTTTAACGAAACATATTGGCACATGAATC
2008-831       TGGCAGCATTATTGGGTGTGAGTATGGCGGATTTTAATGAAACATATTGGCACATGAATC
               *.********.**.**.**.**..*.**..*******.******.***.***.**.****600

81116          ATTTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
LMG9879        ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
M1             ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
33560          ATTTTTATGTTCCTGCTGTTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
1997-7         ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
129-258        ATTTTTATGTTCCTGCTTTTTGTGCTGTGCTTATTAGCCTTTATGTGCTTTATGCAGTCA
81-176         ATTTTTATACTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGCGCTTTATGCAGTTA
2008-831       ATGTTTATGTTCCTGCTGCTTGTGCAGTTATCATTAGTCTTTATGTGCTTTATGCAGTTA
               ** *****..*******..******.**..*.*****.*******.************.*660

81116          AAGGAAGTCCTAAAAATGAAGGTTTGATTGATATTACTGAAATCAATGAAATGCGAGGGA
LMG9879        AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
M1             AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
33560          AAGGAAGT-CTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
1997-7         AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTAGCGAAATCAACGAAATGAGAGGTA
129-258        AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTAGCGAAATCAACGAAATGAGAGGTA
81-176         AAGGAAATCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA
2008-831       AAGGAAGTCCTAAAAATGAAGGTTTGGTTGATATTACTGAAATCAATGAAATGCGAGGGA720
               ******.* *****************.*********..********.******.****.*

81116          TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
LMG9879        TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
M1             TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
33560          TTAAAACAGAAGAAATCAAAACTGTAGAAAGTCCAAATTTAAGCAGTTTTGAAATTTTTT
1997-7         TTAAAACCGAAGAAATTAAAGCTGTAGAAAGTCCAAATTTAAGCGGTCTTGAAATTTTTT
129-258        TTAAAACCGAAGAAATTAAAGCTGTAGAAAGTCCAAATTTAAGCGGTCTTGAAATTTTTT
81-176         TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAGCAGTTTTGAAATTTTTT
2008-831       TTAAAACAGAAGAAATCAAAGCTATAGAAACTCCAAATTTAAACAGTTTTGAAATTTTTT
               *******.********.***.**.******.***********.*.**.************780

81116          ATCGTTATGTGCTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
LMG9879        ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
M1             ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
33560          ATCGTTATGTGCTTAAAAATAAAAATGCTTGGTATGTGGATTGGATGGATACTTTTGTTT
1997-7         ATCATTATGTCCTTAAAAATAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
129-258        ATCATTATGTCCTTAAAAATAAAAATGCTTGGTATGTAGCTTGGATGGATACTTTTGTTT
81-176         ATCATTATGTACTTAAAAACAAAAATGCTTGGTATGTGGCTTGGATGGATACTTTTGTTT
2008-831       ATCATTATGTACTTAAAAACAAAAATGCTTAGCATGTGGCTTGGATGGATACTTTTGTTT
               ***.****** ********.**********.*.****.*.********************840
```

```
81116          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
LMG9879        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
M1             ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
33560          ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTACTAGAAACAAAAGGCT
1997-7         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
129-258        ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
81-176         ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
2008-831       ATATGGTGCGTTTTGGGCTTATTTCTTGGCTTCCTATTTACTTGCTAGAAACAAAAGGCT
               ******************************************.****************900

81116          TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
LMG9879        TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
M1             TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
33560          TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCCATACCTTCTA
1997-7         TTAATAAAGAACAAATAGGTATTGCC-TTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
129-258        TTAATAAAGAACAAATAGGTATTGCC-TTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
81-176         TTAATAAAGAACAAATAGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
2008-831       TTAATAAAGAACAAATGGGTATTGCCTTTTGGCTTTTTGAATGGGCTGCTATACCTTCTA
               ****************.********* **********************.**********960

81116          CTTTACTTGCAGGATATATTTCAGATAAAATTTTTAAAGGCTATAGAATGCCTCCTGCTA
LMG9879        CTTTACTTGCAGGATATATTTCAGATAAAATTTTTAAAGGCTATAGAATGCCTCCTGCTA
M1             CTTTACTTGCAGGATATATTTCAGATAAAATTTTTAAAGGCTATAGAATGCCTCCTGCTA
33560          CTTTACTTGCAGGATATATTTCAGATAAAATTTTTAAAGGCTATAGAATGCCTCCTGCTA
1997-7         CTTTACTTGCAGGATATATTTCAGATAAAATTTTTAAAGGCTATAGAATGCCTCCTGCTA
129-258        CTTTACTTGCAGGATATATTTCAGATAAAATTTTTAAAGGCTATAGAATGCCTCCTGCTA
81-176         CTTTACTTGCAGGATATATTTCAGATAAAATTTTTAAAGGCTATAGAATGCCTCCTGCTA
2008-831       CTTTACTTGCAGGATATATTTCAGATAAAATTTTTAAAGGCTATAGAATGCCTCCTGCTA
               ************************************************************1020

81116          TAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTTT
LMG9879        TAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTTT
M1             TAGGTGCTATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTTT
33560          TAGGTGCTATAGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTTT
1997-7         TAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTTT
129-258        TAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTTT
81-176         TAGGCGCCATAGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTTT
2008-831       TAGGCGCCATGGTGATTATTTTCTTTATGATTATTGGATATTTTACTTCAAATAATCTTT
               ****.**.**.**************************************************1080

81116          ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTTAG
LMG9879        ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTTAG
M1             ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTTAG
33560          ATATGATTACTTTCTTTGCAGCTATGGCGGGGTATTTGGTCTATATACCACAATTTTTAG
1997-7         ATATGGTTATTTTCTTTGCAGCTATGGCAGGGTGTTTAGTCTATATACCACAATTTTTAG
129-258        ATATGGTTATTTTCTTTGCAGCTATGGCAGGGTGTTTAGTCTATATACCACAATTTTTAG
81-176         ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTTAG
2008-831       ATATGGTTATTTTCTTTGCAGCTATGGCGGGGTGTTTAGTCTATATACCACAATTTTTAG
               *****.***.******************.****.***.**********************1140

81116          CAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACTTC
LMG9879        CAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACTTC
M1             CAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATTTC
33560          CAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACTTC
1997-7         CAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATTTC
129-258        CAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGATTTC
81-176         CAAGTATGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGAATCTTGTGTAGGACTTC
2008-831       CAAGTGTGCAAACTATGGAAGTCGTTCCTGCTTTTGCAGTAGGATCTTGTGTAGGACTTC
               *****.************************************.*************.***1200

81116          GTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGGCT
LMG9879        ATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC--ATAGGTTGGGCT
M1             GTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGCTGGGCT
33560          GTGGTTTTATGAGTTATATAGTCGGCGCTTCACTTGGAACAAAGGCTTATAGGTTGGGCT
1997-7         GTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGGCT
129-258        GTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGGCT
81-176         ATGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC--ATAGGTTGGGCT
2008-831       GTGGCTTTATGAGTTATGTAGTCGGCGCTTCACTTGGAACAAAAGC-TATAGGTTGGGCT
               .***.************.************************.**..*****.******1260
```

```
81116           GTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTCTT
LMG9879         GTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTCTT
M1              GTGGATTACTATGGTAGTTGGAATACTGGTCTTATCATGCTCCTTAGTGCTTGTATTCTT
33560           GTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTCAAA-----
1997-7          GTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTCAAA-----
129-258         GTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTCAAA-----
81-176          GTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTCTT
2008-831        GTGGATTACTATGGTAGTTGGAATGCTGGTCTTATCATGCTCCTTAGTGCTTGTATTCTT 1320
                ************************.************************** *

81116           TGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGTATCAATAA
LMG9879         TGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
M1              TGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
33560           ---------------------------------------AGAAGTATCAATAA
1997-7          ---------------------------------------AGAAGTATCAATAA
129-258         ---------------------------------------AGAAGTATCAATAA
81-176          TGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA
2008-831        TGTATACTTTGTTCTACTTTATGTCATTTTAGCGCTAAAAAGAAGCATCAATAA 1374
                                                 *****.********
```

Table S7 Allelic profiles of 64 *C. jejuni* strains. Sequence types (ST) and clonal complexes (ST-CC) among *C. jejuni* isolates in this study.

Strain	(ST- CC)	ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
1854	NA	922	1	1	2	83	2	3	6
81-176	ST-42	604	1	2	3	27	5	9	3
ICDCJ07001	ST-362	2993	1	2	42	4	11	9	8
129-258	ST-42	459	1	2	3	3	5	9	3
260.94	ST-362	362	1	2	49	4	11	66	8
HB93-13	ST-22	22	1	3	6	4	3	3	3
1798	ST-61	61	1	4	2	2	6	3	17
140-16	ST-61	5161	1	4	2	2	225	1	17
1213	ST-508	132	1	6	22	24	12	28	1
1997-7	ST-61	61	1	4	2	2	6	3	17
1997-11	ST-22	22	1	3	6	4	3	3	3
LMG23211	ST-179	220	1	6	29	2	40	32	3
IA3902	ST-21	8	2	1	1	3	2	1	6
84-25	ST-21	21	2	1	1	3	2	1	5
DFVF1099	ST-21	21	2	1	1	3	2	1	5
305	NA	NA	2	1	nd	nd	2	1	5
11168	ST-21	43	2	1	5	3	4	1	5
CF93-6	ST-21	883	2	17	2	3	2	1	5
LMG9879	ST-21	47	2	1	1	5	2	1	5
60004	NA	4836	2	378	27	2	11	3	5
LMG23264	ST-206	46	2	21	5	3	2	1	5
LMG23218	ST-48	48	2	319	1	2	7	1	5
P110B	ST-48	474	2	4	1	2	2	1	5
H22082	ST-48	474	2	4	1	2	2	1	5
2008-1025	ST-21	50	2	1	12	3	2	1	5
86605	ST-48	4840	2	4	27	122	11	1	5
1997-1	ST-658	658	2	4	2	4	19	3	6
2008-831	ST-21	50	2	1	12	3	2	1	5
1997-4	ST-48	475	2	4	1	4	19	62	5
110-21	ST-21	982	2	1	2	3	2	1	5
87330	ST-21	50	2	1	12	3	2	1	5
1893	ST-48	38	2	4	2	2	6	1	5
1928	ST-21	806	2	1	1	3	140	3	5
M1	ST-45	137	4	7	10	4	42	7	1
81116	ST-283	267	4	7	40	4	42	51	1
327	ST-45	230	4	7	41	4	42	7	1
55037	ST-45	45	4	7	10	4	1	7	1
2008-988	ST-206	572	62	4	5	2	2	1	5
D2600	ST-353	452	7	17	12	2	10	3	6
CG8486	ST-574	2943	7	53	27	15	11	3	3
LMG 23223	NA	791	7	97	5	2	135	68	26
LMG 23263	ST-446	3504	7	55	5	10	11	68	6
LMG 23210	NA	380	7	2	6	10	78	37	1
53161	ST-353	4838	7	17	5	68	11	3	6
1997-14	ST-353	5159	7	17	5	2	167	457	6
51037	ST-353	939	7	2	5	2	156	3	6
87459	ST-353	452	7	17	12	2	10	3	6
RM1221	ST-354	354	8	10	2	2	11	12	6
S3	ST-354	354	8	10	2	2	11	12	6
NW	ST-354	354	8	10	2	2	11	12	6
LMG23269	ST-353	4837	8	17	5	3	10	59	6
1997-10	NA	4839	9	17	2	2	86	3	309
2008-979	NA	2274	9	17	5	10	350	3	3
CG8421	ST-52	1919	9	2	2	10	10	3	5
LMG9872	ST-677	677	10	81	50	99	120	76	52
LMG9081	ST-52	52	9	25	2	10	22	3	6
33560	ST-403	403	10	27	16	19	10	5	7

414	NA	3704	227	297	253	338	424	337	250
1336	ST-841	841	64	99	78	104	139	108	16
51494	ST-353	4834	103	2	5	2	156	3	6
LMG23216	NA	4835	64	89	319	100	94	103	16
LMG23357	ST-1275	4883	27	33	22	49	101	9	31
LMG 9217	ST-443	443	24	17	2	15	23	3	12
2008-894	NA	1962	55	172	21	49	125	83	51