Genetic and functional characterisation of pTet-like plasmids of *Campylobacter jejuni*

Submitted by Andrea Gori to the University of Exeter
as a thesis for the degree of
PHD Biological Sciences (SWDTP)
In November 2016

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……

Abstract

With more than 400 million cases a year, *Campylobacter jejuni* is a leading cause of gastroenteritis worldwide. Farm animals, such as cattle poultry and pigs, act as an environmental reservoir for the bacterium. Although many studies have been performed in order to gain a better understanding of the disease and of the environmental lifestyle of the bacterium, much remains to be uncovered. This work aim is to understand the role of the plasmid pTet in the bacterium lifestyle: firstly, analysing which genes are most represented in the plasmids pTet-like and if a difference in genetic set-up between pTet-like plasmids exists and what is its significance. Secondly, assessing whether the plasmids pTet-like are associated with bacterial characteristics, such as virulence, survival or adaptation to different ecological niches.

Several plasmids have previously been described in *C. jejuni*: pTet is the most widely distributed plasmid amongst the known *C. jejuni* strains. This plasmid is about 45 kilobase pairs (kbp) long and 52 open reading frames (ORFs) can be identified in its sequence. Before this work, it was known to be found in about 20 % of *C. jejuni* strains, to encode for tetracycline resistance and a conjugative type four secretion system and to possess several genes involved in horizontal gene transfer. However, the function of this plasmid and the reason why it is so broadly distributed remain largely unknown.

In the first part of this work I screened 4005 genome sequenced strains of *C. jejuni* for the presence of the plasmid. I investigated the genetic make-up of the pTet plasmids, sequencing 19 new plasmids and described a core and accessory sets of genes in this plasmid family using an innovative approach. I found that the core set of genes is predicted to be primarily involved with the maintenance and transfer of

the plasmid pTet, while the accessory set of genes are predicted to be associated with metabolism, resistance and regulative functions.

In the second part of this work I produced 11 *C. jejuni* deletion mutants. I tested these mutants for phenotypes associated with environmental survival and host invasion such as biofilm production, motility and infection of the model organism *Galleria mellonella* larvae. Two of the mutant strains showed an impaired phenotype, when compared to the wild type strain, in motility, production of biofilm and growth in MH broth.

These findings indicate that the pTet plasmids have functions other than tetracycline resistance and horizontal transfer, and may help to explain the broad distribution of the plasmids in *C. jejuni*.

Table of Contents

List of Figures

List of Tables

Declaration

Unless otherwise stated, the results and data presented in this thesis were solely my work.

The perl scripts “pangenome_parser.pl” and “presence_absence_bowtie2.pl” were originally developed by Jamie Harrison (University of Exeter) and modified in this work.

The perl scripts “compare_coverage.pl” and “generate_reads.pl” were developed by Prof. David J. Studholme (University of Exeter).

The PCRs and antibiotic testing described in Chapter 3.1, were completed with the aid of Katie Lukies in 2014, during her final year rotation project under my supervision.

The mutants *DepsG* and *Dunk9* described in Chapter 6 and 7 were produced with the aid of Emily Holton in 2015, during her final year rotation project under my supervision.

List of Abbreviations

AAG	Auto-agglutination
ABC	ATP-binding cassette
Amp100	Ampicyllin 100 µg/ml
bp	Base pair
CAD	Cytolethal distending toxin
Cas	CRISPR associated protein
CBA	Columbia blood agar
CBA+	Columbia blood agar + Skirrow's supplement
CD	Conserved domain
CFU	Colony-forming unit
CME	*Campylobacter* multidrug efflux pump
CRISPR	Clustered regularly interspaced short palindromic repeats
GWAS	Genome-wide association study
IAA	Indole-3-acetic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan50	Kanamycin 50 µg/ml
kbp	1000 base pairs
kDa	kilodaltons
LB	Lysogeny broth
LBA	Lysogeny broth agar
LOS	Lipooligosaccharide
MATE	Multi-antimicrobial extrusion protein
MH	Mueller-Hinton growth medium
MHA	Mueller-Hinton agar
MHB	Mueller-Hinton broth
MIC	minimum inhibitory concentration
MLST	Multi-locus sequence tag
OD_{600}	Optical density measured at 600nm
ORF	Open reading frame
PBS	Phospate-saline buffer
PCR	Polymerase chain reaction

PSK	Post-segregational killing
ROS	Reactive oxygen species
SGF	Synthetic gastric fluid
ST	Sequence type
T4SS	Type four secretion system
T6SS	Type six secretion system
TAE	Tris-acetate EDTA buffer
VBNC	Viable bur nonculturable
WT	Wild-type
w/v	weight/volume
x-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Supplementary Material

S1 - Strains from Molecular Microbiology Laboratory (Geoffrey Pope building, laboratory 401, EX44QD) used in this study. Columns show (from right to left): strain name, origin of isolation of the strain, ability of growing on CBA+ at 42º C, ability of growing on MH plates with tetracycline 10µg/ml, presence of *virD4* marker, presence of *tet(O)* marker, presence of Helicase marker, presence of h*icA/B* marker, species (identified via pcr screening Wang *et al.,* 2002), plasmid setup 1 to 4 according to the presence of the different markers. Y = present, N = absent.

S2 - Plasmids sequenced in this study. Plasmids maps were obtained with BRIG software. Green and Red arrows represent the genes annotated on the plasmid, as reported by RAST automated annotation. Inner circles represent C+G % and GC skew (C+G % only for pCj3 and pK2).

S3 - Annotated genes in the 19 newly sequenced plasmids and in the plasmid pTet (NC_008790.1). Plasmids have been linearised at the start of the *virB2* gene. Each plasmid annotation shows the start nucleotide, the end nucleotide, the gene orientation (+/- = sense or antisense), gene name (as reported by RAST automated annotation pipeline).

S4 – Conserved domains found in the putative proteins of the pTet-like plasmid pan-genome.

Electronic Supplementary Material

All the Electronic supplementary material is available for consultation at

http://bit.ly/2g2c8AJ

E0 - Perl Scripts. (zip) used in this work.

E1 - Genome sequences (fasta) and annotation (gff) files obtained in this study (zip).

E2 - Plasmid sequences (fasta) and annotation (gff) files obtained in this study (zip).

E3 - Table 3.11, Breadth of overage of each ORF of pTet in strain 81-176 (NC008790.1) and metadata of the 4005 *C. jejuni* genome sequenced strains used in Chapter 3.2 (xlsx).

E4 - Figure 3.12 - Cluster analysis of the plasmid pTet genes (1) (pdf).

E5 - Figure 3.13 - Cluster analysis of the plasmid pTet genes (2) (pdf).

E6 - Figure 3.20 - Cluster analysis of the *C. jejuni* plasmid pan genome (pdf).

E7 - Figure 3.22 - Mauve alignment of the plasmids used in the *C. jejuni* plasmid pan genome analysis (jpg).

E8 - Figure 3.23 (A) - Cluster analysis of the pTet-like plasmid pan-genome (pdf).

E9 - Figure 3.23 (B) - Cluster analysis of the pTet-like plasmid pan-genome (pdf).

Chapter 1 - Introduction

Campylobacter spp. are the causal agents of *Campylobacter* enteritis, one of the leading causes of gastroenteritis worldwide (Tauxe, 2002). It is responsible for 400-500 million cases of food poisoning worldwide every year (Ruiz-Palacios, 2007), with an estimated cost for England and Wales of £583 million in 2008, according to the "UK Research and Innovation Strategy for *Campylobacter* in the food chain, Global Food Security, 2010-2015". Overall, more than 80% of human cases are caused by *C. jejuni,* around 10% by *C. coli*, and the remainder by other *Campylobacter* species (e.g. *foetus*). The infectious dose of *Campylobacter* is as low as 500-800 cells (Young *et al.,* 2007), and common symptoms of the disease include diarrhoea containing mucus and blood, acute abdominal pain, vomit, fever, and general malaise (Van Vliet and Ketley, 2001). The disease is usually limited to 5-8 days, but immunocompromised patients can face a persistent infection and, in rare cases, death (less than 1 case in 1000). Complications following *Campylobacter* infections are rare but involve severe pancreatitis (De bois, 1989) and in some cases Guillain-Barrè syndrome (Nachamkin *et al.,* 1998).

From a phenotypical and morphological point of view, the bacterium is a Gram negative, spirally curved rod, and it was described for the first time by Jones *et al.,* in 1931. The *Campylobacter* species are fastidious microorganisms, obligate micro-aerophilic and thermophilic, requiring complex growth media, a concentration of oxygen between 3% and 15%, and a temperature range for growth of 34 to 44 °C with an optimum of 42 °C. Moreover, they are unable to ferment carbohydrates. *Campylobacter* cells usually show a polar flagellum at both ends and are highly motile (Van Vliet and Ketley 2001).

The genomes of *Campylobacter* species are relatively small compared to other enteric pathogens species such as *Escherichia* or *Salmonella*, between 1.6 and 1.7

megabases, and the G+C content is relatively low, around 30% (Van Vliet and Ketley, 2001; Chang and Taylor, 1990). The first complete genome sequencing of *C. jejuni* was published in 2000 (Parkhill *et al.,* 2000); since then several studies have aimed to understand the genome diversity of the bacterial species, particularly the differences between isolates deriving from different sources. In a DNA microarray study performed on 111 *C. jejuni* isolates, six genes belonging to a flagellin glycosylation locus were associated with the chicken-associated strains (Champion *et al.,* 2005); that study also suggested that the majority of *C. jejuni* infection may derive from non-livestock sources. A comparison amongst *C. jejuni* isolated from wild birds and farm animals highlighted that the genetic differences between strains isolated from wild birds was much more pronounced than the strains isolated from a "domestic" environment, concluding that the flow between different hosts makes the differentiation of strains isolated from farm animals much more complex (Sheppard *et al.,* 2011). However, the same study associated the presence of different Sequence Types (STs, identified by MLST) in different hosts. In 2013, the first genome-wide association study (GWAS) was performed on *Campylobacter*, and it identified a statistical association between genes for the biosynthesis of vitamin B5 and strains isolated from cattle, rather than other farm animals (Sheppard *et al.,* 2013). Finally, a study on the entire *Campylobacter* genus, published in 2014 by Iraola *et al.,* showed how different *Campylobacter* species have emerged, originating from a putative non-pathogenic *Campylobacter* ancestor, which acquired a repertoire of virulence genes from other bacteria in order to establish success in a different ecological niche (Iraola *et al.,* 2014, Figure 1.1).

Farm animals such as cattle, poultry, and pigs are known to act as reservoirs for *Campylobacter* species (Young *et al.,* 2007), and they are adapted to survive in

the intestine and cecum of warm-blooded birds. Cross contamination usually occurs during the slaughter process, during which faeces containing bacteria can contaminate meat. Consumption of inadequately cooked meat is one of the major sources of infection (Van Vliet and Ketley, 2001). Other sources include infected milk, other animal products, and water, where the bacterium has been shown to be capable of surviving for prolonged periods of time (Rollins and Colwell, 1986; Sacks *et al.,* 1986; Young *et al.,* 2007; Figure 1.2).

During the infection process, *Campylobacter* makes use of several different virulence factors: an ability to move into the mucus layer that protects the intestine is crucial for colonisation (Szymanski *et al.,* 1995), as well as binding and invasion of the host cell with the aid of membrane oligo and poly-saccharides (Woolbridge and Ketley, 1997) and producing toxins once the invasion is started (Wassenaar, 1997). For a successful colonisation, the bacterium needs to acquire iron from the host (Field, 1986; Pickett, 1992); it needs to be able to deal with the oxidative stress during transmission or the host response, as well as heat stresses due to the different temperatures found in human or avian guts (37°C and 42°C respectively) or during the transmission via water, meat, or milk.

This introduction will firstly describe the virulence and survival factors identified in *C. jejuni* to date. Secondly, it will describe the plasmids identified in *C. jejuni*, with a particular focus on pTet and pVir plasmids. Finally, it details the aims and research questions of this thesis.

Figure 1.1 - Main evolutionary processes in *Campylobacter*. Species are highlighted in blue (genital), red (gastrointestinal), and green (oral). Adapted from Iraola *et al.,* 2014.

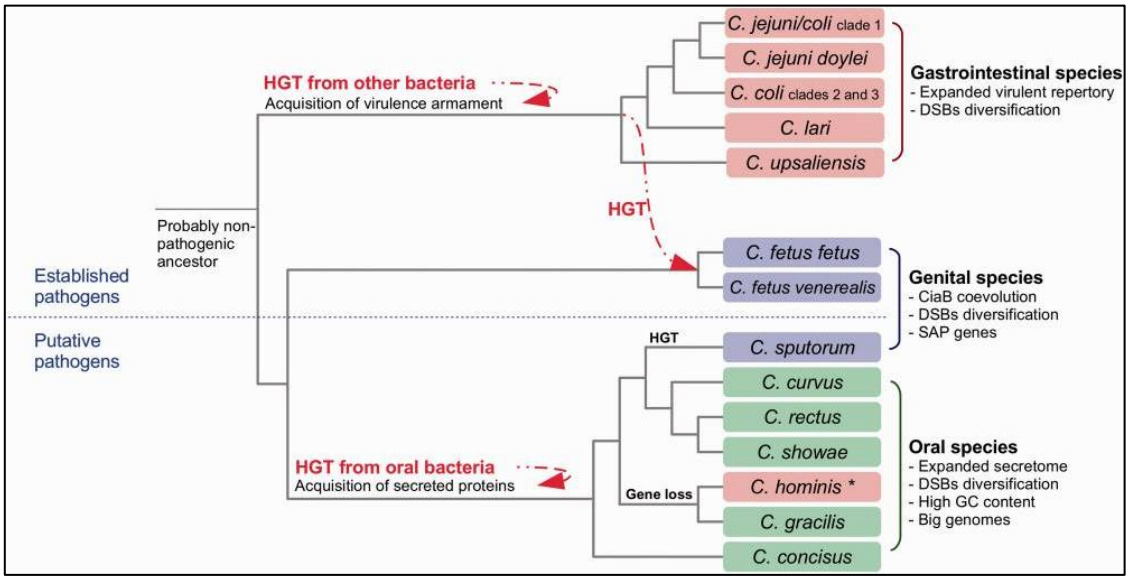
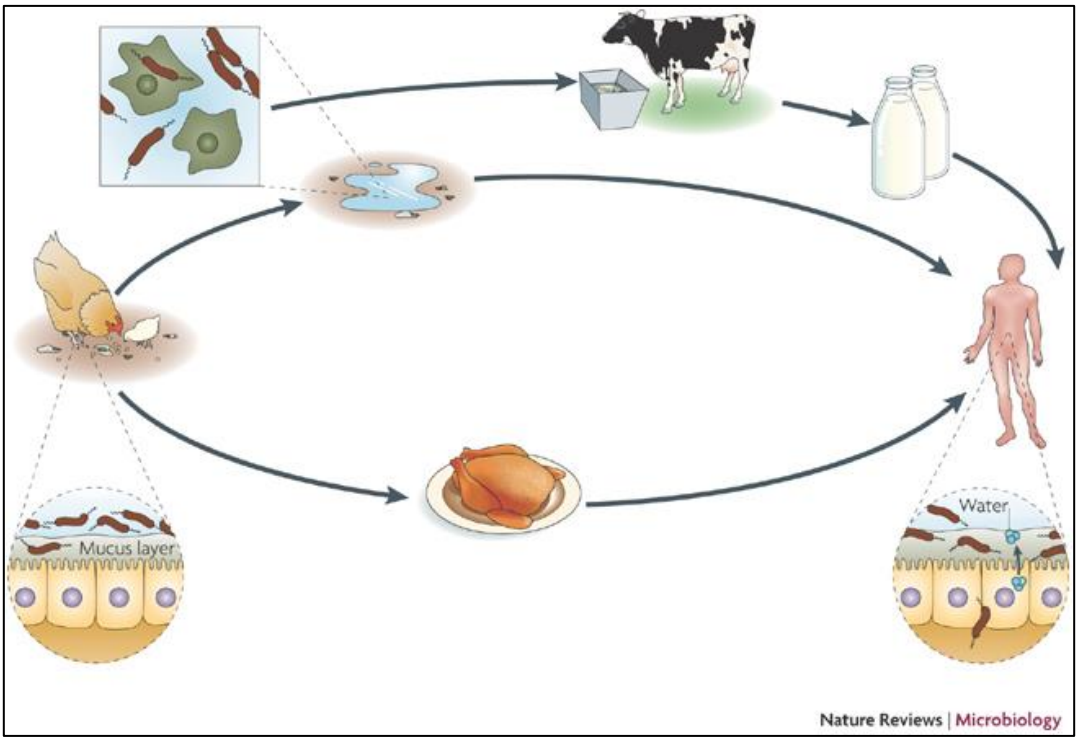


Figure 1.2 - The sources and outcomes of *C. jejuni* infection. Environmental reservoirs that can lead to human infection by *C. jejuni*. It colonizes the chicken gastrointestinal tract and is passed between chicks through the faecal-oral route. *C. jejuni* can be associate with protozoans in water supplies, such as freshwater amoebae. Unpasteurised milk, contaminated meat and drinking water are the primary routes of infection. In humans, *C. jejuni* invades the intestinal epithelial layer provoking inflammation and diarrhoea. Adapted from Young *et al.,* 2007.



1.1 Virulence and survival factors of *C. jejuni*

1.1.1 Motility

C. jejuni motility is mediated by flagella composed of a major and a minor flagellin (FlaA and FlaB, Nuijten *et al.,* 1990), which are highly conserved proteins and about 59 kDa in size. *Campylobacter* species are characterised by a fast, darting motility, and flagella have been recognised as essential elements for pathogenesis, as early studies observed the necessity of functional motility for the successful colonisation of the intestine of animals and humans (Guerry, 2007; Morooka *et al.,* 1985; Nachamkin *et al.,* 1993, Pavlovskis *et al.,* 1991). Gene expression for *flaA* and *flaB* is mediated by a σ^{28} and a σ^{54} dependent promoter respectively (Hendrixson and DiRita, 2003). Mutations in the gene *flaA* result in the production of a truncated flagellum, while the mutation of the gene *flaB* does not appear to have an effect on motility and the flagellar structure (Wassenaar *et al.,* 1991). A kinase/response regulator system FlgSR has been highlighted as a primary regulatory system controlling the expression of flagella in *C. jejuni* (Hendrixson and DiRita, 2003). Moreover, flagellar gene expression and biosynthesis are also controlled by a phase variation process (Lertsethtakarn *et al.,* 2011). Phase variation is a fast and reversible genetic process used by bacteria to influence quickly the transcription or the translation of a gene in response to a rapid change in the environment (van der Woude and Bäumler, 2004). For instance, a mechanism of phase variation may involve the modification of the number of nucleotide repeats in a homopolymer present in the gene or in the promoter, leading to a frameshift of the coding sequence or in the loss of the promoter “power”. An example of phase variation regulation in the flagellar genes has been identified in *C. jejuni*, where

homopolymeric tracts of adenines or thymines are present in the genes *flaS* and *flaR*, and whose alteration truncate the genes (Hendrixson, 2006; Hendrixson, 2008).

Flagella in *C. jejuni* have been shown to be involved in a series of cellular functions important for host colonisation and environment survival other than motility. For instance, a functional flagellar system is essential for the export of secreted proteins important in pathogenesis (Konkel *et al.,* 2004), such as the Cia proteins, a protein family typical of *Campylobacter* essential for successful cell invasion but that is not secreted in the absence of flagellar structures (Konkel *et al.,* 1999). Moreover, flagellar structure is considered essential for autoagglutination (AGG), a process shared by many bacterial species in which cells aggregate as micro-colonies: it is often shown as an initiation of biofilm formation (Misawa *et al*, 2000; Cole *et al.,* 2004; Moreira *et al.,* 2006; Golden and Acheson, 2002). AAG has been associated with virulence in *C. jejuni* strain 81-176 (Misawa *et al*, 2000).

1.1.2 Biofilm formation

C. jejuni has a strict microaerobic requirement, and it is typically hard to grow and maintain in a laboratory environment. It is susceptible to a variety of environmental stresses, food processing procedures, and antimicrobial treatments. Nevertheless, it is extremely widespread in the environment and its high incidence is nearly solely due to the ingestion of living bacterial cells, as the human-to-human transmission is extremely rare. This fact was described as the "*Campylobacter* paradox" in 1999 (Solomon and Hoover, 1999). A hypothesis commonly embraced is that the bacterium makes the best use of biofilm structures to survive in its known environmental reservoirs (Buswell *et al.,* 1998). Biofilms are described as bacterial

populations adherent to each other and/or to surfaces and interfaces (Costerton *et al.,* 1995).

C. jejuni was shown to form three different types of biofilms in a monoculture: attached to a glass surface, as a pellicle at the gas-liquid interface in a stationary culture, and as free cellular aggregate in a liquid culture (Joshua *et al.,* 2006). *C. jejuni* 11168 was shown to be more resistant to environmental stresses when in a biofilm structure, compared to planktonic bacteria (Joshua *et al.,* 2006). Environmental and growth conditions influence the ability of *C. jejuni* to form biofilm: high nutrient conditions and high osmolarity were shown to inhibit biofilm formation (Reeser *et al.,* 2007), while growth in atmospheric oxygen concentration enhanced the formation of biofilm (Reuter *et al.,* 2010). Both flagellar structure and quorum sensing play a central role in biofilm formation in *C. jejuni*, as mutants in *flaAB* and *luxS* genes are significantly impaired during biofilm formation (Reeser *et al.,* 2007).

1.1.3 Adhesion and invasion of host cells

Campylobacter's ability to adhere to gastrointestinal epithelial cells is essential for host colonisation. The process is believed to be mediated by a series of adhesins (Jin *et al.,* 2001). The adhesion to a glycoprotein commonly found in the membrane of epithelial cells also triggers the signalling cascade, mediated by Rac1 and Cdc24 GTPases, which promotes the internalisation of the bacterial cells (Monteville *et al.,* 2003). The protein CadF mediates the adhesion of *Campylobacter* cells to fibronectin. It is reported that an inactivation of *cadF* gene results in a reduced ability of *Campylobacter* to adhere and invade INT407 human intestinal epithelial cells (Krause-Gruszczynska *et al.,* 2007). Other factors controlling the ability of *C. jejuni* cells to adhere to epithelial cells include: the *Campylobacter* adhesion protein A

(CapA), mutation of which results in a reduced capacity to adhere to and invade human epithelial cells (Flanagan *et al.,* 2009); FlpA, a protein that is capable of binding fibronectin and is thought to work together with CadF in the adhesion/invasion process (Konkel *et al.,* 2010; Eucker and Konkel, 2012); and JlpA, which has been suggested to be involved in the adhesion to Hep-2 cells via mutation analysis (Jin *et al.,* 2003). The plasmid pVir has also been reported to be involved in adherence and invasion, as mutants lacking in the *virB11* gene of the type four secretion system encoded by that plasmid are significantly impaired both in adherence and invasion of INT 407 cells (Bacon *et al.,* 2000). However, subsequent studies failed to replicate the result (Dasti *et al.,* 2010).

The surface lipooligopolysaccharide (LOS) molecule of *C. jejuni* is composed of a lipid A and a core oligosaccharide and was reported to serve in immune system evasion, host cells adhesion, and invasion. The sialylation of the LOS is a marker of increased invasive potential and immune system evasion (Guerry *et al.,* 2000; Louwen *et al.,* 2008). The oligosaccharide encapsulation of *C. jejuni* cells also facilitates adherence and evasion from the host immune system (Karlyshev *et al.,* 2000). Non-capsulated mutants (e.g. mutants in the gene that encodes the capsular polysaccharide transport protein, KspM) show reduced ability to colonise INT407 human cells and poultry intestine (Bacon *et al.,* 2001, Jones *et al.,* 2004 and Bachtiar *et al.,* 2007). Other glycosylation systems encoded by the bacterial genome, such as the N-linked glycosylation system encoded by the *pgl* locus, facilitate the immune system evasion and promote the ability to colonise the chicken gastrointestinal tract (Hendrixson and DiRita, 2004, Jones *et al.,* 2004 and Karlyshev *et al.,* 2004).

Finally, flagellar structures are known to be involved in host invasion. The secretion function of the flagellar apparatus is pivotal in this process (Poly and

Guerry 2008). Mutants in *fla* and *flg* genes are reported to have impaired ability to invade host cells (Konkel *et al.,* 2004). The products of *cia* genes (CiaB, CiaC and CiaI) are delivered in to the cytoplasm of the host cells through the flagellar secretion system, and they have an essential role in the invasion process and in the intracellular survival of the bacterium (Konkel *et al.,* 2004; Konkel *et al.,* 1999; Buelow *et al.,* 2011).

1.1.4 Production of toxins

When colonising the host, *Campylobacter* is known to produce an array of different toxins (McFarland and Neill, 1992 and Schulze *et al.,* 1998). The cytolethal distending toxin (CDT), produced by many Gram-negative bacteria, is the only one studied in detail (Pickett and Whitehouse, 1999). The CDT is composed of three subunits (encoded by *cdtA*, *cdtB* and *cdtC* genes). It is delivered in the cytosol of the host's cell where it blocks the kinase cascade essential for the transition to the G2 phase of mitosis (Asakura *et al.,* 2008, Pickett and Whitehouse, 1999). The LOS molecules trigger complications of the *Campylobacter* infection, such as the neuropathies Guillain–Barré and Miller–Fisher Syndrome (Nachamkin *et al.*, 2002).

1.1.5 Iron uptake

The significance of iron acquisition and metabolism for successful microbial proliferation is displayed by numerous examples from medical and environmental biology (Ratledge and Dover, 2000). *Campylobacter* acquires iron from the host in order to establish a successful infection, and it is believed that in order to achieve this, the bacterium uses siderophores derived from the existing microflora (Palayada *et al.,* 2004). Mutants in genes encoding for membrane ferric receptors (Cj0178, *cfrA*

and *cfrB*) and ferric uptake regulator (*fur* gene) all show impaired chicken colonisation ability (Palayada *et al.,* 2004; Xu et al, 2010), and in the chicken cecum an increased transcription level of a gene encoding for a haemoglobin membrane receptor have been observed (Woodall *et al.,* 2005).

1.1.6 Resistance to antibiotics and other antimicrobial molecules

Campylobacter is able to resist to a high concentration of bile salts, heavy metals, and several other antimicrobial agents thanks to the *Campylobacter* multi efflux pump (CME, Lin *et al.,* 2002). This pump is made of a periplasmic protein, an efflux transporter, and an outer membrane protein, encoded by the *cmeABC* operon (Lin *et al.,* 2002). The expression of the operon *cmeABC* is controlled by the transcriptional repressor CmeR: mutation of the *cmeR* gene impaires the ability of *C. jejuni* to colonise chicken (Guo *et al.,* 2008).

Antibiotic resistance in *Campylobacter* is often mediated by specific resistance genes located on the chromosome or on plasmids (Pratt and Korolik, 2004; Chatzipanagiotou *et al.,* 2005; Zhao *et al.,* 2015). The problem of antibiotic resistance is recognised as an issue of public health importance (Takkinen *et al.,* 2003; McDermott *et al.,* 2002; Moore *et al.,* 2006). Even if *Campylobacter* infections are usually self-limiting and not treated in humans other than replacing lost fluids and electrolytes, antibiotic treatments are used to treat severe systemic infections of *Campylobacter* spp. (Aarestrup and Engberg, 2001). Several studies comparing the evolution of antibiotic resistance in time observed a rapid development of multiple antibiotic resistant strains both in the food chain and in clinical isolates (Gallay *et al.,* 2003; Mazi *et al.,* 2008; Smole Možina *et al.,* 2009). For decades, animal production was supplemented with an indiscriminate use of antibiotics (EFSA, 2008; Igimi *et al.,*

2008), and strong evidence supports the conclusion that this led to the emergence and spread of antibiotic resistance in *Campylobacter* spp. (Takkinen *et al.,* 2003; Smith and Fratamico, 2010). In countries or farming areas where the use of antibiotics in animal production is limited, the incidence of antibiotics-resistant *Campylobacter* strains is reduced (Norström *et al.,* 2007; Luangtongkum *et al.,* 2006). It was shown that a limitation in the use of antimicrobial drugs in animals food resulted in a reduction in fluoroquinolone resistance (Gallay *et al.,* 2007, Han *et al.,* 2009): however Nelson *et al.,* in 2007 and Price *et al.,* 2007, have suggested that the antimicrobial resistance may persist in the environment for long periods of time.

1.1.7 Stress tolerance and response

Following the first genome sequence of *C. jejuni* (Parkhill *et al.,* 2000), it was observed that a homologue of the *E. coli* stationary phase regulator *rpoS* (which encode for the global regulator RpoS, σ^{38}) was missing, and therefore a range of adaptive responses present in other bacteria are not present in *C. jejuni*. Other σ factors are present (FliA - σ^{28}, RpoN - σ^{54} and RpoD - σ^{70}), and they control the expression of a range of genes involved in survival and virulence (Jagannathan *et al.,* 2001, Carrillo *et al.,* 2004, Wösten *et al.,* 2004 and Hendrixson, 2006). The protein SpoT is extremely important in controlling the cell response in conditions of stress: SpoT is an effector molecule of stringent response (guanosine tetraphosphate - ppGpp) that negatively regulates the rRNA transcription (acting with the RNA polymerase-associated protein DksA), in favour of amino acid biosynthesis, moving cellular resources away from growth and division, in favour of cellular survival (Gaynor *et al.,* 2005). *spoT* deletion impairs oxygen tolerance, stationary phase survival, invasion and survival in intestinal cells, as well as rifampicin

resistance (Gaynor *et al.,* 2005). The DksA protein has also been shown to be involved in iron metabolism and host cell invasion (Yun *et al.,* 2008)

Campylobacter is primarily a foodborne pathogen, and in the food chain it encounters several stresses that it must endure. In the slaughter chain, *Campylobacter* is subject to desiccation and oxidative stress. The bacterium is particularly sensitive to desiccation and its presence is reduced considerably on beef and pork meat, which are air chilled in the slaughtering process, while its survival is facilitated on chicken where the wet surface is maintained by the water chilling procedure (Humphrey *et al.,* 1995; Oosterom *et al.,* 1983; Butzler and Oosterom, 1991). *C. jejuni* endures different oxygen conditions during its life cycle: aerobic respiration and the immune response of hosts generate ROSs (Reactive oxygen species) including the superoxide anion, hydrogen peroxide, and the hydroxyl radical (Atack and Kelly, 2009). *Campylobacter* can respond to high concentration of ROS with an antioxidant defence system which includes a wide range of enzymes and antioxidant molecules, such as glutathione, catalase (KatA, Atack and Kelly, 2009), alkyl hydroperoxide reductase (AhpC), and other peroxiredoxins (such as the thiol peroxidise Tpx and bacterioferritin co-migratory protein Bcp), cytochrome c peroxidises (*Cj0358* and *Cj0020c*; Atack and Kelly, 2009), and superoxide dismutase (Storz and Imlay, 1999): the increased production of these enzymes results in a long-term aerobic adaptation (Klancnik *et al.,* 2009). Superoxide dismutase proteins (SOD) protect against the superoxide anions (Palyada *et al.,* 2009): the survival of mutants deficient in SOD is impaired in the avian gut due to the presence of oxygen and to the ability to invade Caco-2 cells (Garénaux *et al.,* 2008; Purdy *et al.,* 1999; Mihaljevic *et al.,* 2007).

In the food chain, *Campylobacter* also encounter starvation stress and may be exposed to low pH. In these conditions, the *Campylobacter* cells were shown to enter a viable but non-culturable state (VBNC; Chaveerach *et al.,* 2003), which is characterised by a decreased metabolic activity and change in cell shape and in which the cell survives to stress condition for a prolonged period of time (Moore, 2001; Mihaljevic *et al.,* 2007). Different abilities of retaining virulence were observed in *Campylobacter* VBNC cells, with studies that demonstrated a loss of virulence and others that demonstrated maintenance of the ability to invade intestinal epithelial cells (Byrne *et al.,* 2007, Klancnik *et al.,* 2009).

C. jejuni has an optimum growth temperature of 41.5 °C to 42 °C, and cells are unable to grow below 30 °C (Bolton, 2015). During food processing, bacterial cells are exposed to temperatures close to 4 °C and over 42 °C. Chilled cells remain viable and metabolically active for a prolonged period (Hazeleger *et al.,* 1998 and Dasti *et al.,* 2010). A heat-shock response has been previously described in *Campylobacter* (Konkel *et al.,* 1998). The differential expression of at least 15 genes, the majority of which encode for regulative proteins, was reported when comparing the growth of *C. jejuni* at 37 °C and 42 °C—the body temperatures of humans and chickens respectively (Zhang *et al.,* 2009). The RacR/RacS signal transduction system may also have a role in growth at 42 °C: it regulates a heat-shock protein DnaJ. Mutation of the gene encoding DnaJ severely impairs the ability to colonise chicken (Hermans *et al.,* 2011; Konkel *et al.,* 2004; Ziprin *et al.,* 2001).

C. jejuni is characterised by an extremely high frequency of intragenomic recombination, which can alter its phenotypic characteristics, virulence, and ability to survive in an array of different adverse conditions (Park, 2005; Ridley *et al.,* 2008).

1.2 Plasmids

Plasmids are circular or linear DNA molecules able to replicate autonomously in the host cell (Jackson *et al.,* 2011). They appear in all domains of life and they carry genes encoding for a broad variety of functions (Lipps, 2008; Sundin 2007). In bacteria, plasmids can spread horizontally via conjugation, and indeed plasmid transfer has been defined as “sex in bacteria” (Lederberg and Tatum, 1953). Plasmids play a central role in the adaptation and evolution of prokaryotes (Gogarten and Townsend, 2005).

Classically plasmids were classified into five major groups, according to the functions they encode (Hardy, 1986):

(i) Fertility plasmids (F-plasmids) encode *tra* genes and are responsible for the expression of pili and conjugation structures. These plasmids are capable of horizontal transfer between bacteria via conjugation. Many of the *tra* genes are necessary for the formation of the conjugative pili, and almost all of the genes necessary for the DNA transfer are organised into a single operon. Further genes carried by an F-plasmid may encode for production of bacteriocins or metabolism of carbon compounds. Resistance to antibiotics may be included in an F-plasmid (Hardy, 1986);

(ii) Resistance plasmids (R-plasmids) possess genes conferring antibiotic resistance. The most notable and studied examples are tetracycline and chloramphenicol, but genes encoding for many different antibiotic resistances are known to be carried by plasmids. More broadly, the resistance properties conferred by these plasmids also extend to different classes of compounds such as heavy metals or other toxic

molecules, to physical stresses such as ultraviolet radiations, and to bacteriophages (Martins *et al*, 2001; Thomas and Summers 2008);

(iii) Col plasmids encode for genes responsible for the synthesis of bacteriocins. These proteins are toxic to other bacteria and are used to gain competitive advantage in a particular ecological niche (Hardy 1986);

(iv) Degradative plasmids carry those genes that confer on the bacterium the ability to metabolise particular substances. The classes of compounds that are degraded this way are various and may be simple carbohydrates, complex carbon compounds (such as toluene, nicotine or camphor), halogenated carbon compounds, or proteins (Schmidt *et al.,* 2001);

(v) Virulence plasmids carry “virulence genes”, usually organised in contiguous pathogenicity islands. Virulence plasmids may encode for: toxins, such as the enterotoxins of *Escherichia coli* or the neurotoxins of *Clostridium tetani*; cell structures to evade host defence, such as the capsule of *Bacillus anthracis* or colonisation antigens produced by *E. coli*; or effectors to obtain nutrients from the host, such as an iron transport system in *E. coli* (Thomas and Summers 2008; Hardy 1986).

Another way of classifying plasmids is through incompatibility groups. Replication of plasmids is typically controlled by a region of about 1-3 kbp, known as the basic replicon (which is the shortest piece of the plasmid able to replicate independently). The replicon contains the functions to initiate and control the replication (such as the origin of replication), and the inhibitor of the replication (which can be, a protein factor derived from a bacteriophage or an RNA antisense) which controls the copy number per cell typical of every plasmid (Del Solar and

Espinoza, 2002). If two plasmids share the same replicon, they belong to the same incompatibility group. Two plasmids belonging to different incompatibility groups replicate independently and use independent mechanism to control their copy number; two plasmids that belong to the same incompatibility group share those mechanisms, and also share the total number of copy per cells. They are selected randomly for replication inside the cell and, consequently, one of the incompatible plasmids ends up diluted in subsequent generations, leaving only one of the two incompatible plasmids in the bacterial cell (Bergquist, 1987, Ebersbach *et al.,* 2005). Some plasmids can have more than one replicon and belong to more than one incompatibility group. Closely related plasmids tend to be incompatible, while plasmids that are distantly related tend to be compatible (Thomas, 2014).

Several examples of the roles of plasmids in pathogenic bacteria have been published over the last 50 years. Plant pathogenic bacteria, for example, are known often to rely on several different plasmids to express secretion systems for delivering toxin molecules, which are themselves expressed by several different plasmids (Vivian *et al.,* 2001). For example, *Erwinia arboicola* pv. *gypsophilae* carries the *hrp* genes for the expression of the Type Three Secretion System on a large pPATH plasmid of about 150 kbp in size (Nizan *et al.,* 1997), and in *Ralstonia solanacearum* these genes are on a pVir megaplasmid (Boucher *et al.,* 1986). Other functions specifically linked to virulence are often encoded by plasmids in phytopathogenic bacteria. One example is the production of phytohormones, such as the indole-3-acetic acid (IAA), to affect plant cell proliferation at the infection site (Vivian *et al.,* 2001); several pathovars of *Pseudomonas savastanoi* and *Erwinia herbicola* are known to carry the *iaaLMH* genes for IAA biosynthesis on plasmids, such as pIAA1

and pIAA2 (*P. savastanoi*, Yamada *et al.,* 1986) and pPATH (*E. herbicola*, Clark *et al.,* 1993).

What is described for plant pathogenic bacteria is also true of bacterial pathogens of animals and humans. *Clostridium perfringens* is known to be dependent on its toxin-producing ability to be successful in causing disease (Li *et al*, 2013): many of the 16 known toxins produced by the bacterium are encoded by large plasmids (Miyamoto *et al.,* 2006, Lepp *et al.,* 2010). In *Salmonella enterica*, a group of virulence plasmids with size between 50 and 100 kbp was described (Rychlik *et al.,* 2006). These plasmids are all characterised by the presence of the genes spvRABCD (*Salmonella* virulence genes), which are known to be involved in the systemic phase of *Salmonella* infection and with the destabilisation of the macrophage cells cytoskeleton (Wallis *et al.,* 1995; Lesnick *et al.,* 2001). Moreover, different serovars of *S. enterica* carry plasmids encoding additional virulence-associated genes, such as *rck* (resistance to complement killing, Hackett *et al.,* 1987) for evading the host immune response, or pef (plasmid encoding fimbriae, Baumler *et al.,* 1996), which contributes to *S. typhimurium* adhesion to the mouse intestinal epithelium.

Plasmids encode for a variety of genes and functions, but there is an important similarity among them that is represented by the circumstances in which they are useful for the bacterial host (Top *et al.,* 2000). A part of each plasmid is devoted to the “survival” of the plasmid itself: it encodes for genes that promote replication, maintenance, and conjugation of the plasmid—genes typically referred as “plasmid-selfish” or backbone genes. A part of the plasmid encodes for genes that are useful to the bacterial host only in particular circumstances, such as genes for the adaptation to a particular environment or genes used to initiate the disease in a host.

These genes are concentrated on plasmids as a direct consequence of natural selection: in conditions where a gene would increase the host fitness, the plasmid localisation would allow a more rapid spread both through horizontal and vertical transmission, and the chromosomal localisation would be more limited in conferring advantage only through vertical transmission (Top *et al.,* 2000; Thomas and Summers, 2008).

1.3 Plasmids in *C. jejuni*

The presence of several different plasmids is reported in *Campylobacter* spp. With the exception of a few cases—describing the presence and the structure of small “cryptic plasmids” in regional isolates (Jesse *et al.,* 2006, Hiett *et al.,* 2013), and a handful of mega-plasmids between 80 and 150 kbp long (Gunther *et al.,* 2016; Marasini and Fakir, 2016)—the literature is focused on two plasmids: namely, pTet and pVir (Bacon *et al.,* 2000, Bacon *et al.,* 2002, Batchelor *et al.,* 2004).

Both plasmids were described for the first time in *C. jejuni* strain 81-176, one of the better-characterised strains of *C. jejuni*. It was isolated from a diarrheal outbreak associated with raw-milk consumption (Korlath *et al.,* 1985); it causes inflammatory diarrhoea in humans and disease in several different experimental models (Black *et al.,* 1988, Russell *et al.,* 1989). Moreover, *C. jejuni* strain 81-176 invades INT407 cells at particularly high levels (Hu *et al.,* 1999, Bacon *et al.,* 2000, Bacon *et al.,* 2001, Yao *et al.,* 1994).

1.3.1 Plasmid pVir

The plasmid pVir in the strain 81-176 is reported to be 37,468 bp and to encode for 54 putative ORFs. The G+C content of this plasmid is 26% and just the 83% of

the pVir sequence is predicted to be coding: both these numbers are significantly lower than the chromosome of the strain 81-176 where the G+C content is 30.6% and the percentage of coding sequence is 94% (Bacon *et al.,* 2002).

Amongst the 54 ORFs, the function of the vast majority is unknown. Seven genes encode for homologues of the type four secretion system (T4SS): namely *virB4, virB7, virB8, virB9, virB10, virB11, virD4*, all grouped in an 8.9 kbp long area. The type four secretion system is described in a wide number of bacterial species, and it is involved in a variety of bacterial cell processes including conjugation, DNA translocation, and protein delivery (Bacon *et al.,* 2000).

Although, the plasmid pVir alone does not confer the ability to invade INT407 cells, as demonstrated transforming the strain NCTC 11168 with the full pVir plasmid, some of the pVir genes were associated with the bacterium's ability to move and to invade INT407 cells *in vitro* (Bacon *et al.,* 2002). Mutants in genes encoding for putative components of the T4SS, genes with orthologs in *H. pylori* or genes that appear to be unique to the campylobacter plasmid show a broad range of reduction in INT407 cells invasion, which is difficult to interpret and might be due to the presence of a second plasmid in the strain where this mutant were produced (strain 81-176, which harbours both plasmid pTet and pVir; Bacon *et al.,* 2002). A second study tried to replicate this result unsuccessfully (Dasti *et al.,* 2010). More recently, the presence of the plasmid pVir was also associated with symptoms in humans: patients infected with pVir positive strains were more likely to develop bloody diarrhoea (Tracz *et al.,* 2005). However, the distribution of pVir amongst clinical strains is exceptionally low; a survey conducted in 2005 reported that just one strain out of 56 isolated from symptomatic humans harbouring plasmids contains a plasmid related to pVir (Schmidt-Ott *et al.,* 2005). This observation challenges this plasmid

family's importance in *C. jejuni* virulence, as in that case a higher occurrence in clinical isolates would be expected (Friis *et al.,* 2007). Moreover, the higher virulence of the strains carrying the pVir plasmid have been strongly criticised: Louwen *et al.* in 2006 was not able to replicate the results shown by shown by Tracz *et al.*, in 2005.

1.3.2 Plasmid pTet

A second plasmid was identified in the strain 81-176 of *C. jejuni* (Bacon *et al.,* 2000). Plasmid pTet in the strain 81-176 was determined to be 45,205 bp, approximately 20% bigger than pVir, and to encode 49 ORFs. The G+C content in this case is 29.1%, just slightly lower than the G+C content of the chromosome. It is worth noting that a region with a particularly high G+C content of 40.4% incorporates the tetracycline resistance gene, suggesting the incorporation of this region via horizontal transfer (Batchelor *et al.,* 2004).

The majority of the predicted ORFs are of unknown function. A *tetO* gene was identified, known to confer tetracycline resistance, along with several hypothetical proteins sharing a high percent of similarity with *Helicobacter pylori*. Ten genes grouped in a region that spans across approximately 12.6 kbp encode for proteins that share a high amino acid similarity with the T4SS of some *Brucella* and *Actinobacillus* species. These genes were demonstrated to be involved in the conjugative transfer of the plasmid pTet, which happens to a rate of 10^{-4} to 10^{-6} transconjugant per donor cell (Batchelor *et al.,* 2004). Genes homologous to T4SS have been designated as *Campylobacter* mating genes (*cmg*). Other genes with predicted known functions encoded by pTet are *tolA* (uptake of *coli*cin A), a virulence-associated protein (sharing 42% similarity with *Helicobacter pilori*), and several genes involved in conjugative transfer (Batchelor *et al.,* 2004).

Schmidt-Ott *et al.,* in 2005 reported the presence of several pTet-like plasmids, which were identified as the “mob” family, in a study involving 56 patients with bloody or watery diarrohea. This class of conjugative plasmids conferring tetracycline resistance was identified in 53% of clinical isolates. Their study shows that it is impossible to cure the “mob” plasmid, except from 81-176 (the only described strain that carries both pTet and pVir), in which the loss of pTet does not affect the ability to invade INT407 cells. Finally, the study reveals the lack of correlation between invasiveness and the presence of a “mob” plasmid: if a strain that is unable to invade caco-2 cells acquires a “mob” plasmid, it does not acquire the ability of invading caco-2 cells, and vice versa.

In both the studies described here (Batchelor *et al.,* 2004; Schmidt-Ott *et al.,* 2005), the conjugation experiments involved the use of DNAse in the suspension buffers used to harvest the bacteria and on the plates used to induce the conjugation. The DNAse was used to avoid the possibility of natural transformation of extracellular DNA.

A second survey published in 2007 by Friis *et al.,* where sequences of the plasmids were considered, reported the high diffusion of pTet-like plasmid across 29 tetracycline-resistance or plasmid-carrying strains isolated in various geographical areas and sources (UK, U.S., and Canada; both from environmental and clinical samples). The conservation and diffusion of the pTet-like sequence and structure is remarkable: 83% of the analysed strains showed the presence of a full or partial set of genes previously annotated on the plasmid pTet. Moreover, not one of the analysed strains in that case showed the presence of pVir. The study concludes that pTet-like plasmids are stably maintained and broadly diffused in *Campylobacter*, and

its prevalence in *Campylobacter* population may play an important role in genomic plasticity of *Campylobacter* (De friis *et al.,* 2007).

Other than the plasmid pTet isolated from the strain 81-176 in 2000 (Bacon *et al.,* 2000), plasmids similar to pTet have been isolated from several strains: one example is the strain ICDCCJ07001, harbouring the plasmid 07001_pTet (Zhang *et al.,* 2010). A mosaic structure is described for the plasmid 07001_pTet in this work. It shares homology with 26 pTet genes, and harbours 5 unique genes previously undescribed in the plasmid (Zhang *et al.,* 2010). Mosaic plasmids share homologous genes with a variety of different bacteria (Batchelor *et al.,* 2004), and it is suggested that their origin derives from the acquisition of different DNA regions from different bacterial sources (Boyd *et al.,* 1996).

1.3.3 Other plasmids in *C. jejuni*: small "cryptic" plasmids and mega-plasmids.

Plasmids are defined as cryptic when they do not have genes that appear to contribute to the phenotype of the bacterium. They are typically very small, carry the genes necessary for self-repliction and can be used to develop plasmid vectors for molecular manipulation of specific bacteria (Burian et al., 1997; Li-Zhao et al., 2011).

The first plasmid defined as "cryptic" in *C. jejuni* was a short plasmid described in 2000 (Luo and Zhang): the plasmid pCJ01 was 3212 bp long and exhibited a G+C content of 33.5%. Four open reading frames were identified in that plasmid, and at the time they did not show any sequence similarity with any of the *C. jejuni* DNA sequences available. It was suggested that the plasmid could be used to produce molecular tools for the genetic manipulation of *C. jejuni*, such as *E. coli* - *C. jejuni* shuttle vectors (Luo and Zhang, 2000). In 2003, a second cryptic plasmid was

described in a human clinical isolate of *C. jejuni*: the plasmid pCJ419 (Alfredson and Korolik, 2003). It was 4013 bp long and its G+C content was lower pCJ01 and the chromosome of *C. jejuni* with a value of 27.1%. The plasmid pCJ419 encodes for four ORFs, three of which involved with DNA mobilization and replication (encoding for the proteins Mob, RepA and RepB), and one sharing a higly similar with a gene of unknown function (*cjp23*). The plasmid was successfully used to produce a shuttle vector pGU0202 for the transformation of both *C. jejuni* and *E. coli*. In 2006, the small plasmid pCJ1170 was described in *C. jejuni*: it was 4381 bp long and shared a 99.7% identity with the plasmid pCC2228 isolated in the same study from *Campylobacter coli*. It was identified to be a member of a novel incompatibility group for plasmids in *Campylobacter* and it was suggested that it could be used as a base for the design of vectors to modify plasmid carrying strains of *C. jejuni* and *C. coli* (Miller *et al.,* 2006). Finally, in 2013 the plasmid pTIW94, purified from *C. jejuni* strains isolated from wild birds in south-east US, was described (Hiett *et al.,* 2013). It was a short 3860 bp plasmid, encoding for five ORFs: two ORFs encoded for the *repA* and *repB* found on the pCC2228 plasmid (100% and 95%, similarity respectively), one ORF was 99% similar to a mob gene identified in *C. coli* previously, and the remaining two ORFs exhibited no significant similarity with any of the other DNA sequences present in the GenBank database (Hiett *et al.,* 2013).

With the advent of novel sequencing and assembly techniques, it has been possible to identify new and larger plasmids in several strains of *C. jejuni*. The first example is an 81.08 kbp plasmid isolated from a clinical bacterial isolate from 1994 showing an increased resistance to UV light (strain RM3194; Gunther *et al.,* 2016). The plasmid was identified to encode for an UV-damage repair protein (ImpB), a full type six secretion system, and several proteins involved in the plasmid conjugation.

These characteristics could associate the plasmid with an increased fitness in particular environmental conditions and in virulence. Secondly, a plasmid conferring kanamycin and tetracycline resistance was described in the strain 11601MD isolated from turkey. The length of this plasmid and its organization showed a high similarity with the plasmid pTet (44095 bp, Crespo *et al.,* 2016), and it is the first fully characterized plasmid carrying both the gene *tet(O)* and the *aphA-3* for resistance to tetracycline and kanamycin. The presence of the two resistance genes on plasmids was first shown in 2004 in a partially characterized plasmid (Nirdnoy *et al.,* 2004) and further addressed in the study from Zhao *et al.,* 2016, concluding that the two resistance genes are often associated. Finally, the genome sequencing of the strains OD267 and WP2202, isolated from the liver and the digestive tract of chicken, showed the presence of two mega-plasmids and a shorter plasmid (pCJDM67S) homologous to pVir (Marasini and Fakir, 2016). In particular, the plasmid pCJDM202L, isolated from the strain WP2202 is 119 kbp, encodes for 116 genes and has a C+G content of about 30.5%. It was shown to encode for tetracycline resistance, *cag* pathogenicity island genes, and the 14 core genes for the conjugative type four secretion system.

1.4 Conclusions

The plasmid pTet in *C. jejuni* is found in a high number of known strains. If studies that take into account the tetracycline resistance phenotype are considered, the number of strains potentially carrying this phenotypic marker is even higher. Although the plasmid is reported to be extremely common, the function of the majority of pTet genes is unknown. Moreover, no study of the genetic differences between the several plasmids similar to pTet that were isolated and sequenced was

ever performed. In fact, the literature about the plasmid pTet stops in 2007, when de Friis *et al.,* were surprised by the level of conservation and maintenance of the plasmids pTet-pCC31-like in *Campylobacter* populations (pCC31 is a pTet homologous plasmid isolated from *C. coli*, Batchelor *et al.,* 2004). The question that concludes that paper from 2007 is indeed still open: why and how are plasmids pTet-like maintained?

In this study, a thorough genomic survey for the presence of the plasmid pTet will be performed: several plasmid markers were tested via PCR and phenotypically on the strain collection of the University of Exeter's Molecular Pathogenesis Laboratory for the identification of new strains carrying the plasmid pTet; the presence of the plasmid pTet was assessed in all the *C. jejuni* strains sequenced to date (June 2016). The pTet-like plasmid pan-genome was defined in order to identify the genes present in all the pTet-like plasmids sequenced. The core and the accessory genes for the pTet-like plasmids will be defined and the putative function of those genes will be defined. In order to study the function of a set of genes found with different frequencies on the pTet-like plasmids, a series of 12 mutants will be produced in one model strain of *C. jejuni* and their phenotypic difference when compared to the wild-type will be tested in conditions linked to environmental survival or virulence.

1.5 Aim of this work

The aim of this work is to identify the importance of the plasmid pTet-like for *C. jejuni*: which genes are most represented in the plasmids pTet-like, whether a difference in genetic set-up between pTet-like plasmids exists, and what is its significance? This study will also assess whether or not the plasmids pTet-like are

associated with different functions, such as virulence, survival in particular conditions, or adaptation to different ecological niches.

Chapter 2 - Materials and Methods

2.1 Bacterial Strains and Growth conditions

Campylobacter jejuni strains used in this study are held by the *Campylobacter* collection of the Molecular Pathogenesis Laboratory of University of Exeter (Geoffrey Pope building, laboratory 401, EX44QD) and have been isolated in Asia (Thailand and Pakistan), USA, and the United Kingdom. Table S1 lists the bacterial strains present in this collection. Bacterial strains were stored in glycerol (30% v/v) or on Cryobank™ ceramic beads (Copan Diagnostics Inc.) and maintained at -80 °C. *Campylobacter* strains were cultured under the conditions detailed here or in the relevant section of Materials and Methods for each phenotypic analysis. All strains were streaked in Colombia Blood Agar plates (CBA, CBA Base from Oxoid LTD, supplemented with 5% defribinated horse blood, Oxoid LTD) with selective growth supplement for *Campylobacter* (CBA+, Skirrow's supplement, Oxoid LTD) and incubated for 48 hours in microaerophilic conditions at 42 °C (83% Nitrogen, 8% Oxygen, 5% Carbon Dioxide). Colonies were selected from these plates and streaked on CBA+ and incubated at 37 °C for 24 hours in microaerophilic conditions prior to any further analysis. For counting colony forming units (CFU), *C. jejuni* strains were spotted on Mueller-Hinton agar (MHA, Oxoid LTD) and the protocol described by Miles *et al.* 1938 was followed. For growth in liquid media, *C. jejuni* was grown on a CBA+ plate as a lawn for 24 hours in microaerophilic conditions at 37 °C; bacterial cells were removed from this plate with a sterile swab to produce a starter bacterial suspension in phosphate buffered saline (PBS, pH 7.4, Sigma-Aldrich®) or Mueller-Hinton broth (MHB, Oxoid LTD); the starter suspension was used to inoculate a volume of MHB to a starting concentration of bacterial cells equal to 0.1 optical density measured as turbidity at 600 nm with a spectrophotometer in a cuvette with 1 cm of path length, OD_{600}.

E. coli strain TOP 10 was used for molecular cloning experiments. *E. coli* was stored in glycerol (30% v/v) or on Cryobank™ ceramic beads (Copan Diagnostics Inc.) and maintained at -80 °C. For growing *E. coli*, plates of lysogeny broth agar (LBA) or lysogeny broth (LB) were used, with incubation at 37 °C and supplementary antibiotics as required.

2.2 DNA amplification (PCR)

All primers in this study were designed with Clone Manager Professional 9 software (Scientific & Educational Software) and synthesised by Eurofins Scientific UK (HPSF quality, scale 0.01). Standard DNA amplification was achieved using the following protocol, unless otherwise stated: Buffer Q5 (1X, New England Biolabs); $MgCl_2$ (2mM to 4 mM if needed); dNTP (0.2 µM, Thermo Fisher Scientific Inc.), primers (1 µM each); Q5® High-Fidelity DNA Polymerase (0.5 units/reaction, New England Biolabs); bacterial DNA (10-50 ng); MilliQ water to a final volume of 25 µl. For each reaction in this work Q5® High-Fidelity DNA Polymerase (New England Biolabs) was used, with the following thermal protocol, unless otherwise stated.

1- 30 seconds at 95 °C;

2- 10 seconds at 95 °C;

3- 10 seconds at 45-55 °C, according to the annealing temperature of each primer pair;

4- 30 seconds per kbp at 72 °C;

5- Repeat steps 2 to 4 for 35 times;

6- 2 minutes at 72 °C;

7- Hold 12 °C.

Primers used in this work are listed in the appropriate result section. PCR products were run electrophoretically on agarose gels for analysis. Unless otherwise stated, 1% w/v agarose/Tris-EDTA acetate buffer (TAE) gels were made according to manufacturer's directions (Hi-Res standard agarose, ATGC bioproducts). Agarose gels were stained with 5 µL/100 ml of Midori green advanced gel stain (Geneflow LTD.) according to manufacturer's directions. 20 µL of PCR product was mixed with an appropriate volume of DNA Gel Loading Dye (6X, Thermo Fisher Scientific Inc.) and loaded on agarose gel for electrophoresis with a current of 120 V/cm for a variable time (between 30 and 90 minutes, according to the specific experiment requirements). Gel imaging was performed with a Gel Doc™ Gel Documentation System (BioRad Laboratories, inc.).

2.3 DNA purification

Whole genome bacterial DNA extractions were performed using Wizard® Genomic DNA Purification Kit (Promega corp.), according to the specific protocol for Gram negative bacteria. When DNA was extracted from *C. jejuni*, cells were lawn-plated on a CBA+ plate and left to grow overnight at 37 °C. Cells were then dislodged with a swab and suspended in 5 ml of PBS buffer. 1 ml of bacterial suspension was used for DNA extraction (OD_{600} between 0.8 and 1.5).

For the PCR reactions described in Chapter 3.1, a thermal lysis extraction method was used. One colony of *C. jejuni* was picked from a plate with a sterile toothpick and suspended in 100 µL of MilliQ water. The bacterial suspension was incubated at 95 °C for 15 minutes in a thermal block, then transferred immediately to ice for 5 minutes. The suspension was then spun for 5 minutes at 13000 x g. 2 µl of supernatant were used for template in a PCR reaction.

DNA was extracted from agarose gels following electrophoresis using the GeneJET® Gel Extraction Kit (Thermo Fisher Scientific Inc.), according to the manufacturer's directions.

Plasmid DNA was extracted from *E. coli* using GeneJET® Plasmid Miniprep Kit (Thermo Fisher Scientific Inc.), according to the manufacturer's directions.

Plasmid DNA was extracted from *C. jejuni* strains using either QIAprep® Spin Miniprep Kit (QIAGEN Ltd.) following the protocol modification advised for low copy-number plasmids; QIAGEN Plasmid Midi Kit (QIAGEN Ltd.); or PureYield™ Plasmid Midiprep System (Promega corp.), according to the manufacturer's directions. For QIAprep® Spin Miniprep Kit and PureYield™ Plasmid Midiprep System, *C. jejuni* was grown overnight in 25 ml of MHB at 37 °C, and 10 ml or 25 ml of bacterial culture was used as starting material. For QIAGEN Plasmid Midi Kit 50 ml of bacterial culture was used.

DNA was stored at 4 °C up to 1 month or at -20 °C for longer periods. DNA concentration was evaluated with Nanodrop® ND-1000 (Thermo Fisher Scientific Inc.), when needed, according to the manufacturer's directions.

2.4 Molecular cloning

DNA digestions were performed using restriction enzymes from Thermo Fisher Scientific, following manufacturer's directions.

pGEM®-T Easy Vector System II (Promega corp.) was used to build the constructs used for *Campylobacter* mutagenesis, according to the TA cloning protocol described by the manufacturer. A-tailing of PCR products was used to achieve successful cloning according to pGEM®-T Easy Vector System II protocol, using Qiagen HotstarTaq® (QIAGEN Ltd.).

Ligation of digested DNA fragments in digested vectors was achieved using DNA ligase (Thermo Fisher Scientific Inc.), following manufacturer's directions for sticky-ends. Insert to vector molar ratio was calculated to be between 3:1 and 5:1 using the following formula:

[(mass of vector (ng)* size of insert (bp))/size of vector (bp)] * insert/vector molar ratio = mass of insert (ng)

Ligations were transformed in 100 µL of chemically competent *E. coli* strain TOP 10.

2.5 Transformation of chemically competent *E. coli*

Chemically competent *E. coli* strain TOP 10 was produced according to the following protocol: 5 ml of *E. coli* were grown overnight in LB medium at 37 °C; 1 ml of overnight growth was used to inoculate 99 ml of LB medium; once the bacterial culture reached to 0.4-0.5 OD_{600} (1.5-3 hours), the culture was divided in two 50 ml falcon tubes and centrifuged at 4000 x g for 10 minutes at 4 °C; the supernatant was removed and the pellet was resuspended in 10 ml of ice cold 0.1 M $CaCl_2$; the bacterial suspension was incubated on ice for 1 hour, then centrifuged at 4000 x g for 10 minutes at 4 °C; the supernatant was discarded and the bacterial cell pellet was carefully resuspended in 1.5 ml of ice-cold 0.085 M $CaCl_2$ and 15% (v/v) glycerol. After 1 hour's incubation in ice, the chemically competent cells were divided in 100 µL aliquots and used immediately for chemical transformation of DNA or snap-frozen in liquid nitrogen and stored at -80 °C for up to 1 month.

100 µL of chemically competent *E. coli* strain TOP 10 were inoculated with 10 µL of ligation mixture or 100 ng of purified plasmid DNA in an Eppendorf tube and incubated on ice for 1 hour. The mixture was incubated for 2 minutes at 42 °C, and transferred immediately back to ice for at least 2 minutes. 1 ml of LB broth was added to the Eppendorf tube and incubated in orbital agitation (200 rpm) at 37 °C for 1 hour, in order to allow antibiotic resistance to be expressed. After 1 hour, 100 µL of bacterial suspension was spread-plated on an appropriate LBA plate supplemented with antibiotic/screening molecule.

LB plates supplemented with ampicillin (100 µL/ml), IPTG (0.1 mM) and X-gal (40 µL/ml) (LB-Amp100-IPTG-Xgal) or kanamycin (50 µL/ml) (LB-Kan50) were used when required.

2.6 Transformation of electro-competent *C. jejuni*

The protocol for producing electro-competent *C. jejuni* and subsequent electroporation was described previously (Van Vliet *et al.,* 1998). Briefly, *C. jejuni* was grown overnight as lawns on 4 CBA+ plates at 37 °C; cells were collected with a sterile swab and resuspended in 2 ml of MH broth. Bacterial cells were then washed three times: the suspension was centrifuged for 5 minutes at 10000 x g at 4°C, the supernatant was removed, and cells were carefully resuspended in 2 ml of ice-cold wash buffer (272 mM sucrose, 15% glycerol). Finally, the cells were resuspended in 1 ml of ice-cold wash buffer, and 100 µL aliquots were used immediately for electroporation.

For electroporation 1 to 5 µg of plasmid DNA in a maximum volume of 10 µL was added to 100 µL of electro-competent cells. The mixture was added to an ice-cold electroporation cuvette (1mm chamber, BioRad Laboratories, Inc.) and

electroporated immediately at 2.5 kV, 200 Ω, 25 pF; 1 ml of MHB was used to wash the cuvette and transfer the bacteria to a fresh 1.5 ml Eppendorf tube; the bacterial suspension was spotted on an MHA plate (15 µL spots), and the plate was incubated at 37 °C in microaerophilic conditions for 5 hours. Cells were then collected with a sterile swab and resuspended in 500 µl of MHB. The bacterial suspension was then centrifuged at 6000 x g for 2 minutes, the supernatant was removed and the cell pellet was resuspended in 100 µl of MHB, which was then spread on the appropriate MHA screening plate.

MHA plates supplemented with kanamycin (50 µL/ml) (MHA-Kan50) or chloramphenicol (30 µg/ml) (MHA-Cat30) were used in this study.

2.7 Mutagenesis of *Campylobacter jejuni*

The allelic exchange protocol used in this work to produce mutant strains of *Campylobacter jejuni* is detailed in Chapter 5. Section 5.4 lists the primers and the plasmid vectors used for the mutant design.

2.8 Complementation of the mutant strains of *Campylobacter jejuni*

Mutant strains of *Campylobacter jejuni* that exhibited a phenotypic difference when compared to the wild-type have been complemented re-introducing the deleted gene in the genome of the mutant strain.

The strategy used here to complement the mutant strains was described by Jervis *et al*. in 2015. The study describes the construction of a series of suicide expression vectors for the integration of heterologous genes in *C. jejuni*: these plasmids were named pCJC1, pCJC2, pCJC3, and pCJC4. Each one of them was characterised by a selection marker conferring resistance to chloramphenicol and

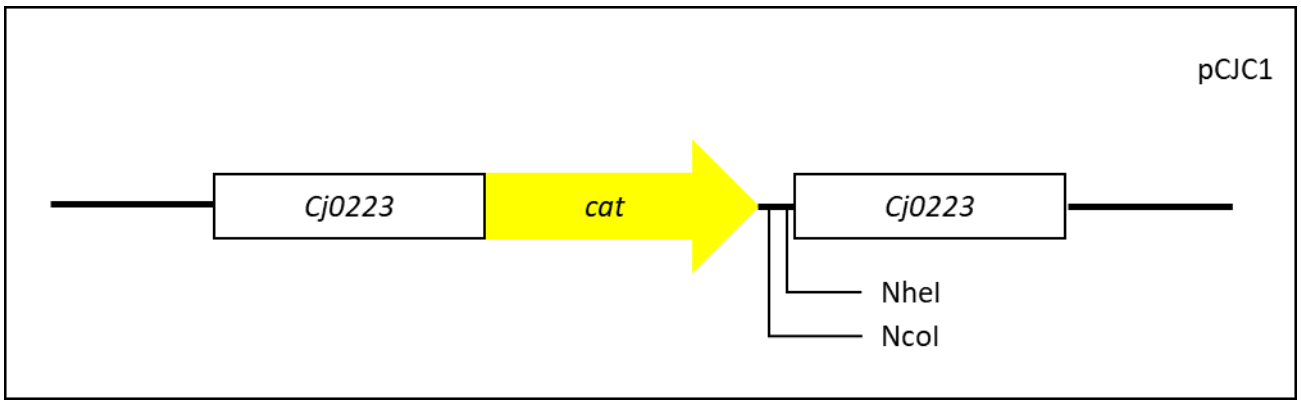
one of four different promoters followed by the cloning site for the gene of interest. These elements were included between the upstream and the downstream sections of the gene *Cj0223*, promoting the integration of the chloramphenicol resistance and the gene of interest in the chromosome of the target strain. To produce the complemented mutants in this study, the vector pCJC1 was used (kindly provided by Dr Dennis Linton), in which the expression of the integrated gene is controlled by the chloramphenicol (cat) cassette promoter. The section of the plasmid vector showing the orientation of the *cat* cassette, the *Cj0023* genes and the position of the restriction sites is represented in Figure 2.1.

The genes of interest for each complementation (namely *hicA, unk9* and *tetO*) were amplified using the primers described in Table 6.3. Each primer pair introduced a restriction site for the enzymes *NcoI* and *NheI* in the 5' and the 3' end of the PCR product respectively, which were then used to clone the gene in the complementation vector, as described in paragraph 2.4. The vectors built for the complementation of the mutants are reported in Table 6.4.

Each mutant was transformed with the corresponding complementation vector via electroporation (*DhicA* with pCJC1::*hicA*, *Dunk9* with pCJC1::*unk9*, *DtetO* with pCJC1::*tetO*), as described in paragraph 2.5. Successful transformation was screened via chloramphenicol resistance, using MHA plates supplemented with 30 µg/ml of chloramphenicol.

The correct insertion of each wild-type gene was checked by PCR, using primers annealing outside the integration site. Sanger sequencing confirmed the correct integration in each of the complemented mutants.

Figure 2.1 – Representation of the integration cassette of pCJC1 expression vector. The vector is described in Jervis et al. in 2015. A fragment of Cj0223 was cloned into pUC18 and interrupted by a chloramphenicol resistance cassette (cat). NheI and NcoI restriction sites were used to insert the gene to be complemented in each mutant.



2.9 Phenotypic assays

2.9.1 Growth curves

A starter culture of *C. jejuni* was prepared suspending bacteria grown as a lawn on CBA+ plates overnight in MHB. The initial OD_{600} of 25 ml of MHB was adjusted to 0.1. Bacterial cultures (in a volume of 25 ml) were incubated for up to 30 hours in microaerophilic conditions at 37 °C.

Bacterial growth was monitored removing 1 ml for OD_{600} measurement when needed.

2.9.2 Identification of minimal inhibitory concentration (MIC) of antibiotics

MIC of different antibiotics for the different *C. jejuni* strains used in this study was calculated in different ways, as described in different results sections (e.g. Chapter 3.1 or 6.5).

MIC of tetracycline was calculated on plate, preparing MH plates supplemented with increasing quantity of tetracycline (1, 2, 5, 10, 20 and 50 mg/l). The tested strain of *C. jejuni* was streaked on the different plates and after 48 hours incubation in microaerophilic conditions at 37 °C, the lowest concentration of antibiotic inhibiting the growth of the bacterium was identified as MIC.

MIC for tetracycline was also calculated with etest® strips (bioMérieux UK Ltd.), following manufacturer specifications for *C. jejuni* (a bacterial suspension of 2 McFarland standards in PBS was spread onto MHA plates supplemented with 5% of horse blood and incubated in microaerophilic conditions for 48 hours).

MIC for phenotypic testing of mutants (Chapter 6.5) was calculated with the plate dilution method, unless otherwise stated. Each well of the first column of a 96-wells plate was filled with 100 µl of MHB supplemented with twice the maximum

concentration of antibiotic to be tested. All remaining wells were filled with 50 µL of MHB. Each column was subsequently topped-up with 50 µL of MHB with antibiotic taken from the previous column (e.g.: column 2 was topped-up with 50 µL of MHB+antibiotic taken from column 1, then column 3 was topped-up with 50 µL of MHB+antibiotic taken from column 2 and so on). In the final layout, each well was filled with 50 µL of MHB and each column had half the concentration of the antibiotic present in the previous column. The last column was filled with MHB and no antibiotic (growth positive control). *C. jejuni* grown overnight as a lawn on CBA+ in microaerophilic conditions at 37 °C were suspended in MHB, and the cell density was adjusted to 0.2 OD_{600}. Each well was topped-up with 50 µL of bacterial suspension and the plate was finally incubated for 24 hours in microaerophilic conditions at 37 °C. The last row was filled with fresh MHB instead of bacterial suspension as a sterility control. After incubation, the OD_{600} in each well was measured with a microplate reader (Infinite® 200 PRO, Tecan Trading AG, Switzerland): the lowest concentration of antibiotic inhibiting the growth of the bacterium was identified as MIC.

2.9.3 Acid tolerance assay

Strains of *C. jejuni* were harvested in MHB after growing overnight on CBA+ plates at 37 °C in microaerophilic condition.

MHB pH 3, pH 4 or pH 5 (Test conditions) were prepared by adding drops of 2 M HCl to normal MHB (pH 7.2, Control condition). Bacteria were collected with a sterile swab from solid medium and resuspended in MH broth. Each suspension was diluted to an OD_{600} of 0.5, then bacteria were centrifuged @ 13000 x g for 5 minutes and resuspended in standard MHA or in modified MHA at pH 3.

Suspensions were incubated for 20 minutes at 37°C in microaerophilic conditions, centrifuged for 5 minutes at 13000 x g, resuspended in standard MH medium, and serially diluted to 10^{-7} of the control test.

Finally, three 10 µl drops for each dilution point were spotted on separate MHA plate. After 48 hours of incubation in microaerophilic condition at 37 °C colonies in each spot showing a number of colonies between 30 and 300 were counted.

Acid tolerance data are here expressed in % of survival following the equation % survival = (test cfu/ml ÷ control cfu/ml) x100.

2.9.4 Survival in Synthetic Gastric Fluid (SGF)

The SGF test medium was produced according to the “Simulated stomach environment” as reported by Beumer *et al.,* 1992. The pH of the SGF was buffered to 4.0, as shown by Reid et al, 2008. The SGF was composed of proteose-peptone (8.3 g/liter), D-Glc (3.5 g/liter), NaCl (2.05 g/liter), KH2PO4 (0.6 g/liter), CaCl2 (0.11 g/liter), KCl (0.37 g/liter), bile (bovine; 0.05 g/liter), lysozyme (0.1 g/liter), and pepsin (13.3 mg/liter) (15). All of the components except the enzymes were dissolved in distilled water, the pH of the solution was adjusted to 4.0 with 1 M HCl, and the solution was filter sterilized and stored at 4 °C. Prior to its use, lysozyme and pepsin were added from fresh stock solutions.

Bacterial suspensions to a concentration of 1.0 OD_{600} in PBS were prepared by collecting bacteria grown overnight as a lawn on CBA+ in microaerophilic conditions at 37 °C with a sterile swab.

1 ml of bacterial suspension in PBS was spun down and resuspended in 200 µl of sterile MilliQ water, then quickly transferred to 4.8 ml of SGF or MHB (Control condition). After 20 minutes, 3 hours or 6 hours of incubation at 37 °C in

microaerophilic conditions, 100 µL of bacterial suspension in SGF or MBH were transferred to a microtiter plate (96 wells) and serially diluted for cell counting.

Cell counting and % survival was estimated as reported in “acid tolerance assay”.

2.9.5 Motility assays

The ability of *C. jejuni* to move in soft agar was evaluated following a modification of the protocol described in Reuter *et al.,* 2013.

Cells grown overnight on CBA+ in microaerophilic conditions at 37 °C as a lawn, were suspended in PBS with a sterile swab and cell concentration was adjusted to OD_{600} 1.0. 5 µL of bacterial suspension was spotted onto each plate for testing the motility: two or three strains were tested at the same time on each plate.

MH motility plates were prepared starting from MHB powder (Oxoid LTD.) and adding agarose (Agar n. 2 bacteriological, Lab M, Ltd.) separately to a final concentration of 0.4 % w/v. The quantity of MHB powder was also adjusted to half the advised quantity when needed.

The two tests conditions were: MH plates with full quantity of nutrients (21 g/l of MHB powder) and 0.4% w/v agarose; MH plates with half quantity of nutrients (10.5 g/l of MHB powder) and 0.4% w/v agarose.

Media were autoclaved at 121 °C, 2 atm for 20 minutes and allowed to cool down to 50 °C before pouring in the petri dishes. Plates were incubated at room temperature for, at the least, over night before being inoculated. After inoculation, spots were left to dry on the bench for 30 minutes and plates were incubated in microaerophilic conditions at 37 °C upside-down (lid up); pictures were taken every 24 hours. Measurements of the diameter of the bacterial movement area were

performed analysing each picture with imageJ software (version 1.48, https://imagej.nih.gov/ij/).

2.9.6 Biofilm production

MH plates with half quantity of nutrients (10.5 g/l of MHB powder) were prepared as described in “Motility Assay” section (2.7.5) with two adjustments: the quantity of agarose used was 0.8% w/v and before autoclaving 40 µg/ml of Congo red (Certified biological stain, Thermo Fisher Scientific Inc.) and 20 µg/ml of Coomassie Brilliant Blue G-250 (Thermo Fisher Scientific Inc.) were added to the medium.

5 µL of bacterial suspension prepared as described in motility assay section (2.9.5) were spotted on each plate: two or three strains were tested at the same time on each plate.

After inoculation, spots were left to dry on the bench for 30 minutes and plates were incubated in microaerophilic conditions at 37 °C for 48 hours upside-down. Plates were then incubated in aerobic conditions for 48 hours upside-down (lid up) and before observing the results of the test and taking pictures.

2.9.7 Persister cells formation

Persister cells were induced incubating a bacterial culture in MHB supplemented with a quantity of antibiotic up to 100 times the MIC. 1 ml of bacterial culture supplemented with the appropriate quantity of antibiotic was incubated for 24 hours in microaerophilic conditions at 37 °C in a 12 wells plate (Corning Inc.), together with the appropriate controls (bacterial culture without antibiotic, uninoculated broth). 100 µL of bacterial suspension were transferred to a microtiter

plate for serial dilutions and the number of CFU per ml was enumerated as described in 2.1 (Miles *et al.,* 1938).

2.9.8 Survival in water

Tap water was collected from Geoffrey Pope building (Exeter, EX4 4QD, lab. 401), and autoclaved for 20 minutes at 121 °C, 2 atm. *C. jejuni* grown overnight as a lawn on CBA+ in microaerophilic conditions at 37 C were suspended in autoclaved tap water and cells concentration was adjusted to 0.5 OD_{600}. 5 mL of bacterial suspension in water was incubated at 4 °C for 8 days. 100 µL samples were collected at the start of the experiment and after 8 days of incubation. Cell counting and % survival was estimated as reported in “acid tolerance assay”.

2.9.9 *Galleria mellonella* larvae injection

Larvae of *G. mellonella* TruLarv™ were provided by Biosystems Technology (biosystemstechnology.com). Larvae were maintained in plastic pots for up to 2 weeks at 15 °C and were weighted before each experiment: only larvae with a body weight between 0.2 and 0.3 grams were challenged with *C. jejuni*.

The injection of *G. mellonella* larvae was performed as reported by Champion *et al.,* 2010, with the following adjustments.

C. jejuni grown overnight as a lawn on CBA+ in microaerophilic conditions at 37 °C were suspended in PBS buffer with a sterile swab; cell concentration was adjusted to 2.0 OD_{600}. 10 µL of bacterial suspension was injected with the aid of a microsyringe (Hamilton) in the first right foreleg of the larvae. 10 larvae per strain were injected. Each experiment included 10 larvae injected with sterile PBS buffer and 10 non-injected control larvae. Larvae were incubated in microaerophilic

conditions at 37 °C for up to 48 hours in empty petri dishes, lined at the bottom with a disc of filter paper.

2.10 Bioinformatics techniques and Genome sequencing

2.10.1 Genome Sequencing

All the strains analysed in this study, were sequenced using the in-house Illumina MiSeq facility available at the University of Exeter.

Paired End Library preparation for each genome was carried out following the procedure reported in the Illumina protocols (Preparing Samples for Paired-End Sequencing, Rev. A, June 2008) on 1 to 5 µg of DNA, shearing the DNA to obtain a final length of <800 bp per fragment.

Illumina MiSeq V3 reagent kit was used, allowing a read length of 300 bp per fragment in each paired read.

2.10.2 Bioinformatics tools and analysis

In this work, for the scripting and programming activity, the following programming languages were used: Perl (version 5.10.1, www.perl.org) with BioPerl module, when necessary (ver 1.0069, www.bioperl.org); R (version 3.3.0, R Core Team (2016), https://www.R-project.org/) and GNU/Linux Bash (under Ubuntu server environment, version 14.04). Scripts used in this work are available in the online supplementary material.

2.10.3 Quality check and trimming of sequence reads

In this study FastQC (version 0.10.0, www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) package was used for the quality

check step of all the raw data. The report that FastQC outputs highlights the following: (i) Basic statistics, a report of the number, length, overall GC% and the sequencing pipeline used; (ii) Per base sequence quality; (iii) Per base quality score; (iv) Per base sequence content; (v) Per base GC content; (vi) Per sequence GC content; (vii) Per base N content; (viii) Sequence length distribution; (ix) Duplicate sequences; (x) Overrepresented sequences; (xi) Overrepresented k-mers.

Fastq-mcf (from package ea-utils, version 1.1.2-537. Aronesky, 2013) software was used for the adaptors-removal process. It was run with the following options: (i) an adapter file, listing the overrepresented sequences in the input file; (ii) an initial sub-sampling of 100000 of reads; (iii) an average quality of 20 and the presence of more than 1% of N to discard a read; (iv) a minimum sequence length of 45 bp remaining after the cleaning; (v) Illumina PF filtering was enabled. Fastx_trimmer (from FASTX-toolkit package, version 0.0.7, http://hannonlab.cshl.edu/fastx_toolkit/) software was used to trim the beginning and/or the end of each read when a drop under 28 points in quality or a GC% bias was identified. After this, Fastq-mcf was run again with the only purpose of eliminating all the reads shorter than 35 bp, in order to avoid issues in the assembly phase due to the presence of too short reads.

2.10.4 Genome assembly and annotation

Once raw reads were cleaned and trimmed a series of software was used to produce a draft assembly. SPAdes genome assembler software (version 3.9, Nurk, Bankevich *et al.,* 2013, http://bioinf.spbau.ru/SPAdes) was used to assemble each genome. This software is composed of a series of algorithms designed to apply the de Brujin graph to a shotgun sequencing composed of very short reads. Briefly, the de Brujin graph is preferred to classical and “greedy” assembly methods with very

short reads: it is based on the identity between a series of k-mers in which each read is fragmented and the sequence of the assembly is represented by a "path" that connects k-mers that differs by just one base (Miller *et al.,* 2010).

The optimisation of the k-mer length was performed as advised by the software developers for assembly of reads obtained with Illumina MiSeq technology: after running the assembler with a k-mer value of 21, 33, 55, 77, 99 and 127, the output was processed via QUAST software (version 3.2, http://bioinf.spbau.ru/quast, Gurevich *et al.,* 2013) which estimates all the typical statistics for the assembled genomes, such as N50, longest contig, total length of the assembly.

Genome assemblies obtained this way have been annotated with RAST ("Rapid Annotation using Subsystem Technology", Overbeek *et al.,* 2014, http://rast.nmpdr.org/), an automated on-line pipeline for the annotation of bacterial genomes.

Final genome assembly graphs were visualised with Bandage software (Wick, *et al.,* 2015); the software displays the connections between contigs that are not present in the final assembly file, also highlighting circular contigs that can be identified as plasmids.

2.10.5 Alignment and coverage analysis

The approach used to evaluate the presence and absence of plasmid genes from each genome analysed in this study was composed of two main steps: (i) the alignment of the raw reads of each sample to a reference sample, and (ii) the evaluation of the coverage of each annotated feature in the reference in each alignment.

The plasmid pTet or the plasmid pan genome genes (described in section 2.8.7) were used as a reference for the reads alignment. For the genomes retrieved as scaffolds or completed (e.g the genomes retrieved from the NCBI database http://www.ncbi.nlm.nih.gov/) a set of raw reads for each one was produced with an ad-hoc perl script (generate_reads.pl). These artificial read sets were composed of 300000 reads 100 bp long in fasta format, randomly obtained from the whole genome.

For the strains of *C. jejuni* that were sequenced in this study or for which the sequencing reads were available, a subset of 300000 reads was used for the alignment of reads obtained with Hi-seq technology (up to 100 bp long, for the sequencing reads used in this study), and a subset of 100000 reads was used for reads obtained with Miseq technology (up to 300 bp long, for the sequencing reads used in this study). Fastq reads subsets were randomly sub-sampled with an ad-hoc perl script (reads_subset.pl).

The random sub sampling of sequencing reads provided the same theoretical coverage for each alignment, which was on average ~20X considering the length of *C. jejuni* strain 81-176 genome, according to the formula:

Expected coverage = (Read Length * Number of Reads) / Genome length

In order to ascertain whether the depth of coverage fell within 10% difference of the expected coverage, the alignments were inspected using and the “depth” module of Samtools (http://samtools.sourceforge.net/), which provides the depth of coverage per site for the sequence alignment and visually inspected using Integrated Genome

Viewer software (version 2.3.5, Thorvaldsdóttir *et al.,* 2013, http://www.broadinstitute.org/igv/).

For the sequencing read alignments, Bowtie 2 software was used (version 2.10.0, Langmead and Salzmed 2012, http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). The software is specifically designed to align a set of short reads to a genome sequence or an assembly and outputs an alignment file in sam format (detailed information about the format can be found at http://samtools.sourceforge.net/SAMv1.pdf). In order to evaluate the coverage of each feature, the sam file was converted to bam format (http://samtools.sourceforge.net/SAMv1.pdf), ordered and indexed with three softwares belonging to the samtools suite (Samtools view, Samtools order and Samtools index respectively). CoverageBed was finally used (software included in Bedtools suite, version 2.17.0, Quinlan and Hall, 2010, https://code.google.com/p/bedtools) in order to assign a point scale from 0 to 1 to each annotated feature, where 0 is completely non-covered and 1 is fully covered.

An ad-hoc perl script was used to produce a table containing the percentage of coverage of each feature in each strain (compare_coverages.pl). Alignments were visually inspected with the aid of Integrated Genome Viewer software. R software (version 3.0.1, R core team, 2013) was then used to produce a graphical analysis as well as a cluster analysis of the obtained data.

In particular, the aheatmap function contained in the NMF package (version 0.17.6, Gaujoux R and Seoighe C, 2010) was used to produce the heatmap showing the presence and absence of each plasmid feature, in order to graphically identify clusters of gene presence or absence in different strains.

Hierarchical clustering analysis was performed with the Rfunction “hclust” on a Jaccard distance matrix calculated using the with R functions “vegdist” contained in the “vegan” pagkage (version 2.4.3, https://cran.r-project.org/web/packages/vegan/index.html). Hierarchical clustering trees were converted into Newick tree format when needed (e.g. for the annotation on Evolview software, as described in results chapter 3), with the R function “as.phylo” (“ape” package, version 3.5, http://ape-package.ird.fr/).

2.10.6 Sequence retrieving, pairwise alignments and phylogenetic analysis

To retrieve the sequences by similarity from genome assemblies (e.g in the MLST analysis in Chapter 3.2.3.2), a BLAST alignment approach was used. The sequence of each fragment in a reference strain (strain *C. jejuni* NTCC 11168 in the MLST analysis) was used as template for command line BLAST (blastall version 2.2.21, Camacho *et al.,* 2009, ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.21/).

Each genome file in FASTA format (contigs or the completed genome sequence) was formatted as a database with the format-db command, then blastall command was run with the following input arguments and options: (i) the sequences in fasta format we wanted to find in each genome as input file, (ii) the database formatted sequence of the genome as database to perform the alignment, (iii) “blastn” as the BLAST algorithm used for the research, (iv) 1×10^{-6} as the e-value used for the reliability of the alignment.

The result of the BLAST alignment was parsed with an ad-hoc perl script using the bio-perl module to facilitate the input-output of the sequences, in order to obtain a fasta file showing the hits of the BLAST alignment.

For pairwise alignment, MUSCLE (version 3.7, Edgar, 2004) algorithm was used. The alignment was then visually inspected and used as input for phyML (version 3.0, Guindon *et al.,* 2010) in order to produce a maximum likelihood phylogenetic tree. Option used for PhyML were the following, unless otherwise stated: (i) GTR (General Time Reversible) as tree model, (ii) aLTR (SH-like) as branch support algorithm (Anisimova and Gascuel, 2006), (iii) Optimized across site variation rate, (iv) NNI (Nearest-Neighbor-Interchange) as tree searching model, and (v) BioNj with optimized tree topology as starting tree.

Alignments and phylogenetic tree building was performed with of SEAVIEW software (version 4.2, Gouy *et al*., 2010, http://pbil.univ-lyon1.fr/software/seaview.html) that provides a graphical user interface for MUSCLE and PhyML. Trees were explored and formatted for presentation using Mega 5 (version 5.2, Tamura *et al.,* 2011, http://www.megasoftware.net/) or CLC sequence viewer (version 7, QIAGEN Inc.), and annotated with Evolview online suite (version 2, https://evolview.codeplex.com/, He et al, 2016). The branching layout of each tree as inferred by PhyML was not modified.

2.10.7 Plasmid pan genome construction and analysis

Plasmid pan-genomes (described in Chapter 3.3) were built pooling all the orthologous proteins in all the plasmids annotations then extracting a representative gene sequence for each orthologous protein.

The list of proteins and genes in fasta format annotated on each plasmid was retrieved from the output of the RAST annotation pipeline. The list of proteins of each plasmid was used as input for ProteinOrtho software (Version 5.15, Lechner *et al.,* 2011). The output of this software was parsed with an ad-hoc perl script (pangenome_parser.pl) in order to extract a representative gene sequence from the annotation of a representative plasmid.

The final output of the plasmid-pan genome shows one representative gene sequence for each of the orthologous proteins identified by the ProteinOrtho software. The list of genes in FASTA format was used as a reference for the alignment of the sequencing reads of each *C. jejuni* strain analysed in this study using an approach similar to the one described previously (Section 2.10.5 - Alignment and coverage analysis). In contrast to Section 2.10.5, an ad-hoc perl script (presence_absence_bowtie2.pl) was used to align the sequencing reads to the reference and produce a pileup file using the “mpileup” module of SamTools (Handsaker *et al.,* 2009). The pileup file indicates the coverage of each base in the reference—in this case the gene lists: the breadth of coverage for each gene was calculated as number of covered bases divided by length of the gene.

Chapter 3 - Conservation and distribution of plasmid pTet in *Campylobacter jejuni*

The plasmid pTet is reported to be present in a large number of strains of *C. jejuni* (Schmidt-Ott *et al.,* 2005), but its importance has yet to be elucidated.

The first section of this chapter will assess the presence of a plasmid pTet in the collection of *C. jejuni* strains hosted by the Molecular Pathogenesis Lab in Exeter. In each strain I will check for the presence of the plasmid using PCR markers on genes that have been described in pTet, and determine the tetracycline resistance phenotype (Batchelor *et al.,* 2004). The strains exhibiting the presence of one of the molecular of phenotypic markers will be sequenced in order to assess the variability of the plasmid pTet-like family.

Secondly, I will assess the presence of the plasmid pTet in all the available sequenced *C. jejuni* strains (corrected in May 2016). I will profile the frequencies of the genes found in the plasmid pTet in all the sequenced *Campylobacter* strains and assess where the greatest conservation and greatest variability resides. I will also assess the association of plasmid presence with the strain origin, year of isolation and phylogeny.

Finally, I will describe the gene composition of the plasmid pan-genome of *C. jejuni* and the frequency of those genes amongst the sequenced strains of *C. jejuni*. I took a pan-genome approach in order to describe the full picture of the plasmid genes in the *C. jejuni* species instead of describing every plasmid singularly.

The aims of this chapter are to assess the presence of the plasmid pTet and the plasmids pTet-like amongst all the sequenced *C. jejuni* strains, to describe the variablility of this plasmid family and the genes composing it and, ultimately, studying the frequencies of the different genes annotated in the plasmids pTet-like to ascertain whether this plasmid appears to be maintained solely for the purpose of antibiotics resistance, a thesis already challenged by Friis *et al.,* 2006.

3.1 In vitro testing for the presence of the plasmid pTet

To assess the distribution of the plasmid pTet in *C. jejuni,* the collection of 121 *C. jejuni* strains hosted by the Molecular Pathogenesis Lab in Exeter (Table S1) was first tested with PCR markers specific for the plasmid pTet and for the tetracycline resistance phenotype. Moreover, all the strains positive for at least one marker were sequenced using MiSeq sequencing technology and a 300+300 bp paired end library. The first step was to identify markers for the plasmid pTet. Four primer pairs were designed, covering four different regions of the plasmid pTet in the strain 81-176 (Table 3.1). These primers were designed to amplify a 300 to 400 bp section of the helicase gene, the *virD4* gene of the T4SS cluster, the *tet(O)* gene, and the *hicA/B* cluster, described for the first time on the plasmid pTet in this thesis (and its presence is detailed in Chapter 4). Secondly, I measured the MIC of tetracycline on plate for the strain 81-176 (10 ug/ml, as described in Section 2.7.2), and I used the same concentration for all the other 121 strains of the collection.

Each strain was grown on Columbia Blood agar plates with Skirrow supplement (CBA+) for 2 days at 42°C in microaerophilic conditions, then re-streaked and grown at 37 °C in microaerophilic conditions. Single colonies were then tested for growth on tetracyline (Muller-Hinton agar plates w 10 ug/ml of antibiotic), and for presence of virD4 PCR marker (DNA was obtained with the thermal lysis method described in Section 2.3). In case of growth or positive PCR, the strain was tested for the remaining PCR markers, tested with a PCR assay to identify the *Campylobacter* species (*C. jejuni* or *C. coli,* Wang *et al.,* 2002), and grown for plasmid and genomic DNA extraction.

Table 3.1 - Sequences of PCR primers for the plasmid genes marker. Table columns (from right to left): Primer name, primer sequence, product length for the primer pair, description of the PCR product.

Primer Name	Sequence	Product Length	Description
hicAB_markF	GATCAGGATGGATATTTTGCAC	342	Plasmid marker for the *hicBA* operon
hicAB_markR	TTAGGGTGCAATATTTCTCC		
tet_markF	TACGGGTCTGTGCCTGTATG	303	Plasmid marker for the *tet(O)* gene
tet_markR	CTCACGTTGACGCAGGAAAG		
virD4_markF	TATCTACCGCCAGCACAAAG	399	Plasmid marker for the *virD4* gene
virD4_markR	TATCCGCCCACGCATTAATC		
hel_markF	TGGATTGCCTGTTGCTTCTG	329	Plasmid marker for the helicase gene
hel_markR	TGGGTGCTGGAACAAATGTG		

3.1.1 Molecular and phenotypic test overview

A total of 23 strains out of 121 were found to be positives for at least one marker between *virD2*, helicase gene, or *hicA/B*. The PCR products were visualised on agarose gel, as reported in Material and Methods (section 2.2). Figure 3.1 shows the result PCR for each of the markers on selected strains. 20 strains out of 121 were able to grow on MH plates supplemented 10 ug/ml of Tetracycline. All the strains positive for the *tet(O)* gene were also resistant to tetracycline. One of the MH plates is shown in Figure 3.2. Table 3.2 summarises the results of the PCR and the tetracycline screening for the strains that were identified as *C. jejuni*. The full results of this analysis are reported in Supplementary Materials Table S1. Four different makeups were identified according to the presence of the PCR markers and the phenotypic resistance to tetracycline: (i) three strains are positive for 3/5 markers, and negative for both tetracycline resistance on plate and the tetracycline PCR marker; (ii) three strains are positive for both tetracycline resistance on plate and the tetracycline PCR marker and negative for the *virD4*, the helicase and the *hicA/B* PCR marker; (iii) two strains are negative for the *hicA/B* marker and positive to the other four makers, (iv) fifteen strains are positive for all the PCR markers tested and for tetracycline resistance.

Figure 3.1 - PCR gels for the 23 strains positive to at least one marker. (A) Marker *virD4*, (B) Marker *tet(O)*, (C) Marker *hlcAB*, (D) Marker Helicase. Each image is composed of 2 gels; the lanes are ordered as follows: 1kb Plus marker (Thermo scientific®, Lane 1), 100 bp Plus marker (Thermo scientific®, lane 14 for the first gel and lane 13 for the second gel), strain 93/372, strain MB3, strain MB12, strain Hi81006, strain Hi81214, strain MB8, strain Goose_222, strain K2, strain 11818, strain Hi40620300, strain PS623, strain PS762 (lanes 2-13 of the first gel), strain A1.CF.12, strain A6.T2.15, strain A8/35/15A, strain C1/C/2, strain C3/T/25col3, strain C5/T2/8, strain MB9, strain MB18, strain Cj1, strain Cj2, strain Cj3 (lanes 2-12 of the second gel), positive control (strain 81-176, lane 14 of the second gel), negative control (no template, lane 15 of the second gel). Agarose gel 1% in TAE, 30 minutes run, 120 V.

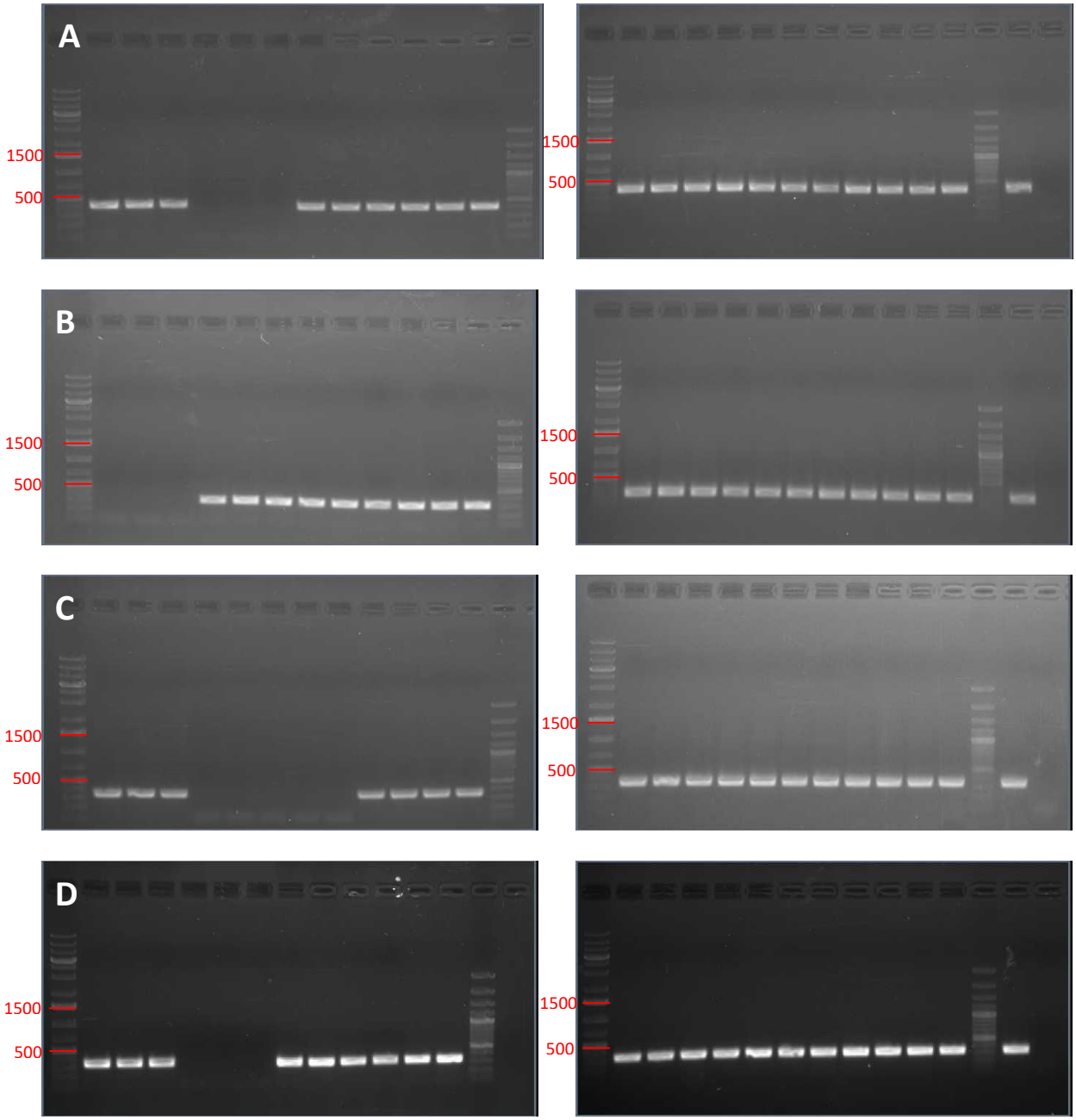


Figure 3.2 - Example of one plate for tetracycline resistance screening. The plate is divided into three sectors for strains EX1286, MB9 and PS304. The strain present in each sector is recorded on the plate. Each strain is streaked in triplicate with an "X" shape.

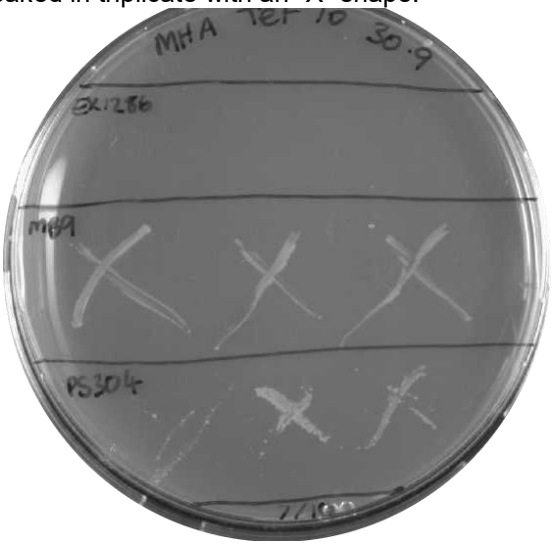


Table 3.2 - Summary of the PCR and tetracycline resistance screening for the 23 positive strains. Table columns (from right to left): strain name, origin of isolation of the strain, ability of growing on MH plates with tetracycline 10µg/ml, presence of *virD4* marker, presence of *tet(O)* marker, presence of Helicase marker, presence of *hicA/B* marker, species (identified via pcr screening Wang *et al*, 2002), plasmid setup 1 to 4 according to the presence of the different markers. Y = present, N = absent.

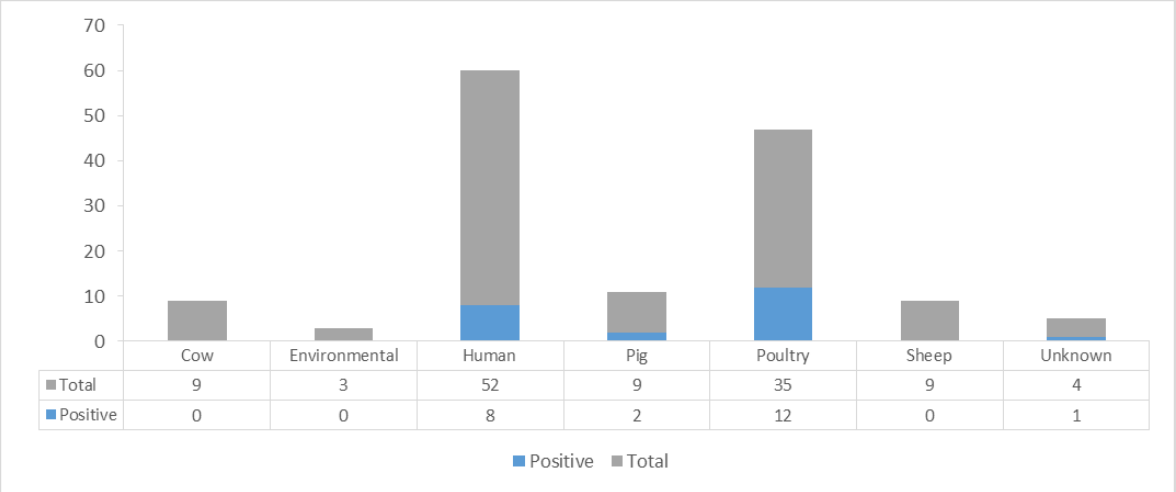
Strain	Origin	Tet R	virD4 mark	tet(O) mark	hel mark	hicA/B mark	Species	Setup
93/372	Unknown (Pet)	N	Y	N	Y	Y	jejuni	1
MB3	Poultry	N	Y	N	Y	Y	jejuni	1
MB12	Poultry	N	Y	N	Y	Y	jejuni	1
Hi81006	Human	Y	N	Y	N	N	jejuni	2
Hi81214	Human	Y	N	Y	N	N	jejuni	2
MB8	Poultry	Y	N	Y	N	N	jejuni	2
222	Goose	Y	Y	Y	Y	N	jejuni	3
K2	Human	Y	Y	Y	Y	N	jejuni	3
11818	Human	Y	Y	Y	Y	Y	jejuni	4
Hi40620300	Human	Y	Y	Y	Y	Y	jejuni	4
PS623	Pig	Y	Y	Y	Y	Y	jejuni	4
PS762	Pig	Y	Y	Y	Y	Y	jejuni	4
A1.CF.12	Poultry	Y	Y	Y	Y	Y	jejuni	4
A6.T2.15	Poultry	Y	Y	Y	Y	Y	jejuni	4
A8/35/15A	Poultry	Y	Y	Y	Y	Y	jejuni	4
C1/C/2	Poultry	Y	Y	Y	Y	Y	jejuni	4
C3/T/25col3	Poultry	Y	Y	Y	Y	Y	jejuni	4
C5/T2/8	Poultry	Y	Y	Y	Y	Y	jejuni	4
MB9	Poultry	Y	Y	Y	Y	Y	jejuni	4
MB18	Poultry	Y	Y	Y	Y	Y	jejuni	4
Cj1	Human	Y	Y	Y	Y	Y	jejuni	4
Cj2	Human	Y	Y	Y	Y	Y	jejuni	4
Cj3	Human	Y	Y	Y	Y	Y	jejuni	4

The barplot in Figure 3.3 includes the number of strains that gave a positive result for at least one plasmid pTet marker and the origin of the strain. A higher number of strains were isolated from farm animal/environmental sources than from human patients. 8 out of 52 strains (15.4%) isolated from humans were positive for least one pTet marker, while out of 65 strains isolated from animal or environmental sources 14 were positive (21.5%).

The percentages of strains showing the presence of at least 1 pTet marker are reported in Figure 3.4. Of positive strains 52% were isolated from poultry including goose, 35% from symptomatic humans (clinical strains), 9% from pigs, and 4% had unknown origins.

The gel in Figure 3.5 shows an example of the plasmid extraction result for three strains found positive for at least one marker: strains Cj1, Cj2, and Cj3. The strain Cj1 did not show the presence of plasmid when subjected to extraction using the Promega PureYield™ Plasmid Midiprep System, the QIAprep Spin Miniprep Kit or the QIAprep HiSpeed® Plasmid Midi kit, designed to extract high molecular weight plasmid and equal the CsCl gradient results (HiSpeed® plasmid purification handbook, May 2012). The genomic set-up of Cj1 will be detailed in Chapter 5.

Figure 3.3 - Number of strains positive for at least one marker, divided by origin. The grey bar represents the total number of strain analysed; the blue part represents the number of strains identified as positive to at least one marker.



	Cow	Environmental	Human	Pig	Poultry	Sheep	Unknown
Total	9	3	52	9	35	9	4
Positive	0	0	8	2	12	0	1

Figure 3.4 - Percentage of analysed strains positive for at least one marker and their origin.

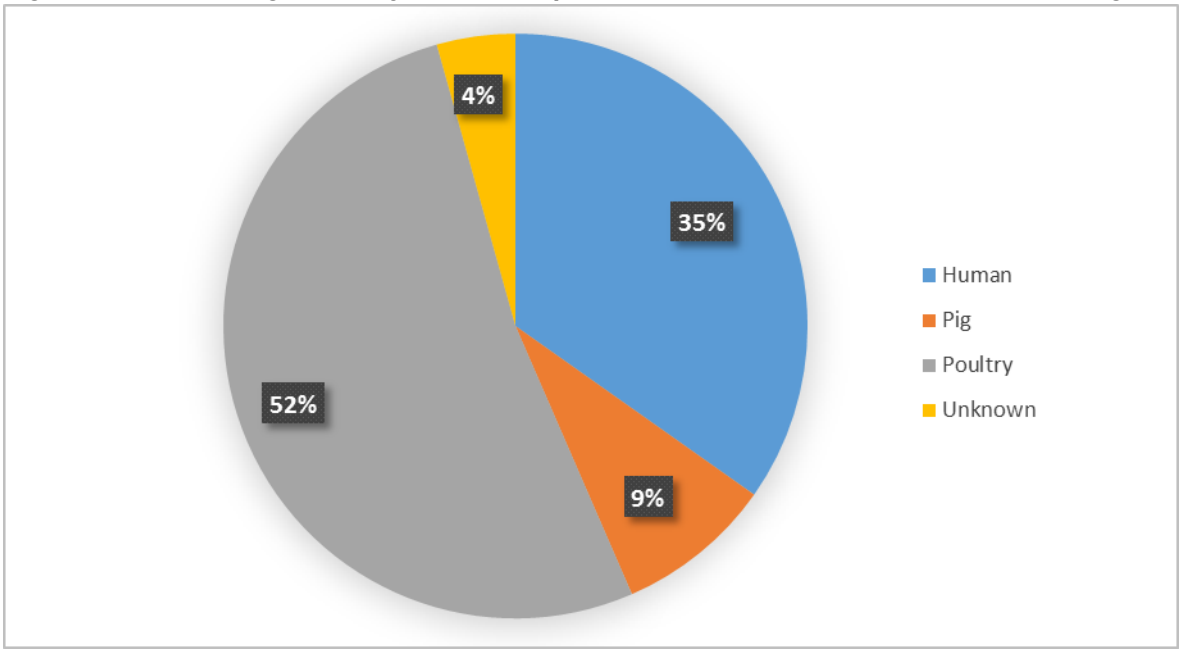
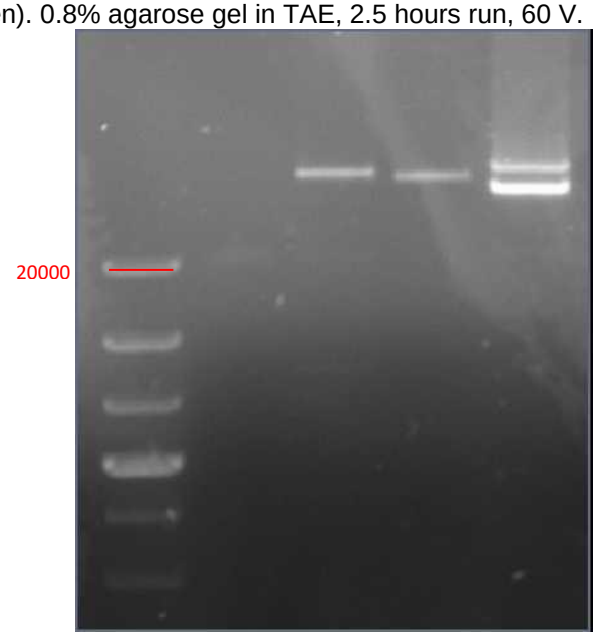


Figure 3.5 - Plasmid extraction from selected *C. jejuni* strains. Lane 1, 1Kb plus ladder (Thermo scientific); Lanes 2-5, strain Cj1, Cj2, Cj3 and 81-176. Extraction performed with QIAprep Hi-Speed plasmid midi kit (Qiagen). 0.8% agarose gel in TAE, 2.5 hours run, 60 V.



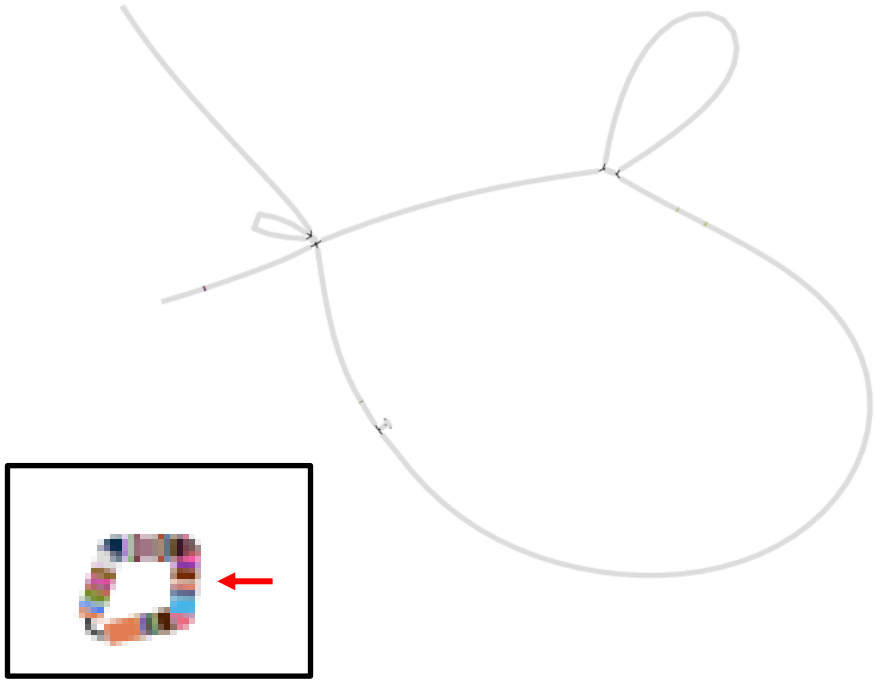
3.1.2 Plasmid sequence determination via whole genome sequencing

The 23 strains positive for at least one molecular or phenotypic marker listed in Table 3.1.1.1 have been sequenced to identify the full sequence of the plasmid. Full assembly statistics for each sequenced genome are reported in Table 3.3; the sequences of each assembly can be retrieved as Electronic Supplementary Material (E1). After assembly, the sequence of the plasmid was identified with the help of Bandage (https://rrwick.github.io/Bandage/, Wick *et al.,* 2015). This software shows a graphical representation of the de-novo assembly graph, showing connections that are not present in the draft genome contigs because the assembler was not able to resolve them. These connections derive from the physical distance between the paired end reads, whose sequence is determined by the sequencing process. This is a powerful tool for plasmid study, because it allows the identification of contigs deriving from a circular plasmid by a physical DNA connection between the ends. Another important feature of this software for this application is the possibility of using BLAST internally in order to identify the location of particular sequences in the assembly: I used this feature to identify the location of the PCR markers (*tet(O), virD2,* helicase gene, *hicA/B*) in the assembly, confirming that they were in the circular plasmid contig or locating them in the chromosomal contigs set. Figure 3.6 shows an example output of Bandage, highlighting the presence of a small circular plasmid structure separated from the large chromosomic contigs set. For the rest of the work, for the strains isolated from animal hosts, the strain name and the host of isolation shown in Table 3.2 will be merged into a new strain name (e.g.: strain 222 isolated from goose will be named strain Goose_222).

Table 3.3 - Assembly statistics for the sequenced strains. Table columns (from right to left): Strain name, number of contigs longer than 500 bp, largest contig in the assembly (bp), total length of the assembly (bp), percent of G+C, N50 (bp), L50 (bp), number of contigs of any length, number of contigs longer than 1000 bp. Unless otherwise specified, every statistic shown in this table is based on the number of contigs longer than 500 bp. The N50 is the weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value. The L50 is the number of contigs longer than or equal to the N50 value.

Strain name	# contigs >=500	Largest contig	Total length	GC (%)	N50	L50	# contigs >= 0 bp	# contigs >= 1000 bp
11818	24	728271	1767836	30.45	191357	2	49	16
Poultry_C3_T_25	21	563591	1751552	30.24	331004	2	52	13
Cj1	22	543624	1763494	30.22	215989	3	24	20
Cj2	19	656205	1696696	30.43	297362	2	61	12
Cj3	22	547640	1746094	30.37	211082	3	46	20
Goose_222	24	559741	1674799	30.49	213026	3	78	16
Hi81006	19	688355	1722779	30.44	231706	2	32	14
Hi81214	28	553931	1636265	30.48	282176	2	69	12
Hi40620300	85	548118	1810698	30.36	272673	3	167	18
K2	33	301612	1786320	30.19	176277	4	50	30
Poultry_MB8	27	612691	1690067	30.31	190197	3	52	18
93/327	28	544865	1665129	30.48	189514	3	113	19
Pig_PS623	91	320151	1863679	30.23	170709	5	198	31
Pig_PS762	63	319777	1808539	30.2	175289	4	134	24
Poultry_A1_CF_12	31	563460	1757893	30.25	331004	2	82	15
Poultry_A6_T2_15	18	563464	1748612	30.25	331006	2	49	13
Poultry_A8_35_15A	33	563467	1758955	30.25	331006	2	101	13
Poultry_C1_C_2	19	563463	1749593	30.25	331002	2	50	13
Poultry_C5_T2_8	18	563589	1750271	30.25	331006	2	46	13
Poultry_MB12	21	776946	1676902	30.44	245358	2	48	14
Poultry_MB18	49	777734	1734473	30.39	189506	2	112	15
Poultry_MB3	29	658226	1695267	30.44	297333	2	58	14
Poultry_MB9	22	563964	1753260	30.27	331004	2	72	13

Figure 3.6 - Graphical representation of the assembly of the strain Cj2 using Bandage software. Each segment of the figure represents a contig of the genome assembly (grey lines), linked to each other using the “graph” information provided by the software SPAdes. The red arrow indicates the plasmid pCj2 (zoomed in the image inset), which self-circularises in the assembly procedure thanks to the information deriving to the paired end. Colours on the contig indicated by the red arrow represent Blastn hits of the genes annotated in the plasmid pTet (NC008790.1), indicating the presence of the pTet genes on this contig, rather than on the rest of the assembly.



3.1.3 Overview of the plasmid sequences in the strains analysed

Each of the sequenced plasmids from the previous analysis was graphically represented with the aid of BLAST Ring Image Manipulator (BRIG, Alikhan *et al.,* 2011) and is reported in Supplementary Material S2. Full plasmid DNA sequences and annotations are available as Electronic Supplementary Material E2. Table 3.4 shows the general characteristics of each plasmid.

Table 3.4 - Summary of the characteristics of the sequenced plasmids. Table columns (from right to left): Strain name, name of the plasmid, Size in kilobases, G+C percentage, number of putative proteins, number of predicted genes.

Organism	Plasmid name	Size (Kb)	GC(%)	Protein	Gene
Campylobacter jejuni 11818	p11818	44.95	29.06	52	52
Campylobacter jejuni poultry C3_T_25	pPoultry_C3_T_25	47.66	28.04	56	56
Campylobacter jejuni Cj2	pCj2	43.67	27.57	48	48
Campylobacter jejuni Cj3	pCj3	42.01	27.84	46	46
Campylobacter jejuni goose 222	pGoose222	46.08	28.13	52	52
Campylobacter jejuni Hi40620300	pHi4062300	44.83	29.27	50	50
Campylobacter jejuni K2	pK2	47.59	28.77	54	54
Campylobacter jejuni Pet_93_327	pPet93_327	37.73	27.19	43	43
Campylobacter jejuni Pig_PS623	pPigPS623	45.38	28.72	50	50
Campylobacter jejuni Pig_PS762	pPigPS762	45.59	28.99	51	51
Campylobacter jejuni Poultry_A1_CF_12	pPoultry_A1_CF_12	47.64	28.05	54	54
Campylobacter jejuni Poultry_A6_T2_15	pPoultry_A6_T2_15	47.64	28.03	54	54
Campylobacter jejuni Poultry_A8_35_15A	pPoultry_A8_35_15A	47.64	28.05	55	55
Campylobacter jejuni Poultry_C1_C_2	pPoultry_C1_C_2	47.64	28.05	54	54
Campylobacter jejuni Poultry_C5_T2_8	pPoultry_C5_T2_8	47.64	28.05	55	55
Campylobacter jejuni Poultry_MB12	pPoultry_MB12	37.53	27.23	45	45
Campylobacter jejuni Poultry_MB18	pPoultry_MB18	44.76	28.99	52	52
Campylobacter jejuni Poultry_MB3	pPoultry_MB3	38.15	27.17	43	43
Campylobacter jejuni Poultry_MB9	pPoultry_MB9	47.66	28.04	56	56

3.1.3.1 Comparison of newly sequenced strains against plasmid pTet from strain 81-176

The graphical representation of the BLASTn alignment between the genome assembly of the newly sequenced strains against the pTet reference sequence is shown in Figure 3.7. In this analysis, the sequence of pTet (NC008790.1) was re-annotated with RAST automated annotation pipeline (Overbeek *et al.*, 2014). This structural comparison was obtained using BRIG. Similar data are reported in Table 3.5, where the raw reads of each sequenced strain were mapped to the reference plasmid with using Bowtie2 and assessing the breadth of coverage with the tool CoverageBed, as described in materials and methods 2.10.5.

The columns in Table 3.5 are ordered from the lowest to the highest difference compared with the reference: the method used to quantify the difference was as follows. Given the value of 1 as 100% breadth of coverage and 0 as 0% coverage of each feature, considering 52 as the number of features in the plasmid pTet, the sum of the breadths of coverage in each strain would be a maximum of 52. The rows of Table 3.5 have been ordered from the start of the plasmid pTet as retrieved from the NCBI database, accession number NC008790.1.

The presence of conserved domains and further hypothetical functions has been assessed with the CDS engine from the ncbi (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?). Each CDS result reported from here on is associated to an e-value lower than 10^{-3}.

The greatest variability occurred around specific regions of the reference plasmid; the gene encoding for “replication protein” was present in 9 of the strains. The plasmids from other strains are missing the gene encoding the replication

protein of the reference plasmid pTet. Interestingly, this CDS is missing from nearly all the strains isolated from poultry, except from the strain Poultry_MB18.

The BRIG plot in figure 3.7 highlights several conserved gaps when the sequences of the newly isolated plasmids are compared to the pTet plasmid NC008790.1: in particular, in the regions around the replication protein and the surrounding hypothetical proteins, and the hypothetical proteins FIG00469861 and FIG00471987.

Figure 3.7 - Graphical representation of the BLASTn alignment of each genome sequence of the 23 strains in Table 1.2.1 and the plasmid pTet (NC008790.1). Coloured line shows >70% identity. Each ring represents (starting from the inner ring): strain HI81214, strain HI81106, strain MB8, strain MB12, strain 93/327, strain MB3, strain goose222, strain K2, strain pig726, strain Hi40620300, strain 11818, strain Pig623, strain A6/T2/15, strain A1/CF/12, strain C1/C/2, strain A8/35/15A, strain T3/C/25, strain MB9, strain Cj3, strain Cj1, strain Cj2, gene annotation of pTet (NC008790.1) produced with RAST automated annotation pipeline (Overbeek *et al.,* 2014).

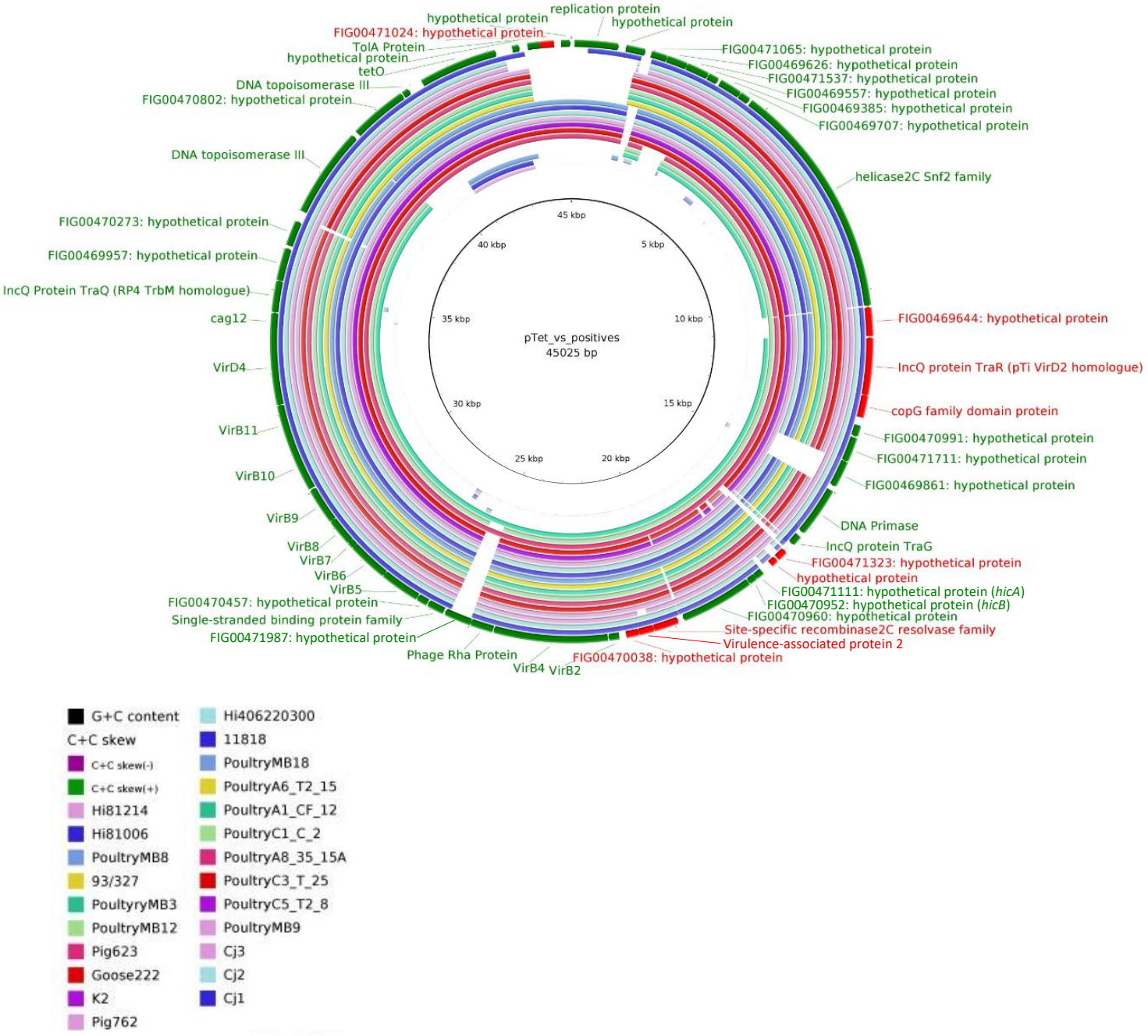


Table 3.5 - Breadth of overage of each ORF of pTet in strain 81-176 (NC008790.1) in the newly sequenced strains of *C. jejuni* (Table 3.4). Columns are (from left to right): gene name, position in NC008790.1, and each of the newly sequenced strain. Each cell shows the value of coverage percentage in each strain, and is coloured in dark grey if it is higher than 90%, in light grey if it is lower than 10%, and in shades of grey for intermediate values.

Gene	Location	81-176	Pig_PS623	Goose_222	K2	Pig_PS762	11818	Po_MB18	Hi40620300	Cj1	Po_A1_CF_12	Po_C5_T2_8	Po_C1_C_2	Cj3	Po_MB9	Po_A8_35_15A	Po_C3_T_25	Cj2	Po_A6_T2_15	Pet_93_327	Po_MB3	Po_MB12	Po_MB8	Hi81214	Hi81006
replication protein	NC008790.1:66..1139	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.70	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FIG00471069: hypothetical protein	NC008790.1:1434..1796	1.00	0.91	0.87	0.84	0.88	0.88	0.90	0.87	0.85	0.56	0.56	0.53	0.00	0.41	0.41	0.40	0.00	0.62	0.00	0.00	0.00	0.84	0.00	0.00
FIG00471065: hypothetical protein	NC008790.1:1938..2336	1.00	0.99	1.00	0.98	1.00	0.98	0.99	1.00	0.99	1.00	1.00	1.00	0.96	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00469626: hypothetical protein	NC008790.1:2333..2845	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.39	0.37	0.38	0.00	0.00	0.00
FIG00471537: hypothetical protein	NC008790.1:2920..3399	1.00	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.14	0.14	0.04	0.00	0.00	0.00
FIG00469557: hypothetical protein	NC008790.1:3404..3670	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00469385: hypothetical protein	NC008790.1:3729..4289	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00469707: hypothetical protein	NC008790.1:4292..4558	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
helicase2C Snf2 family	NC008790.1:4575..10403	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.05	0.00	0.00
FIG00469644: hypothetical protein	NC008790.1:10425..11129	1.00	0.10	1.00	0.12	0.11	0.08	0.09	0.12	0.11	1.00	1.00	1.00	0.10	1.00	1.00	1.00	0.06	1.00	0.74	0.04	0.72	0.00	0.00	0.00
IncQ protein TraR (pTi VirD2 homolog)	NC008790.1:11155..12543	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Ribbon-helix-helix protein2C copG family domain protein	NC008790.1:12543..13094	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00470991: hypothetical protein	NC008790.1:13295..13576	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00471711: hypothetical protein	NC008790.1:13605..14216	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00	0.99	1.00	0.97	1.00	0.96	1.00	1.00	1.00	0.00	0.00	0.00
FIG00469861: hypothetical protein	NC008790.1:14220..14873	1.00	1.00	0.84	0.86	0.85	0.03	0.03	0.02	1.00	0.02	0.00	0.02	1.00	0.00	0.02	0.00	1.00	0.00	1.00	1.00	0.85	0.00	0.00	0.00
DNA primase	NC008790.1:14992..16218	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.76	1.00	1.00	1.00	1.00	0.00	0.00	0.00
IncQ plasmid conjugative transfer protein TraG	NC008790.1:16284..16589	1.00	1.00	1.00	0.85	0.96	0.96	0.76	0.88	1.00	0.93	0.96	0.97	0.91	0.90	0.94	0.89	1.00	0.94	1.00	1.00	1.00	0.00	0.00	0.00
FIG00471323: hypothetical protein	NC008790.1:16811..17068	1.00	1.00	0.00	0.00	0.91	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.89	1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
hypothetical protein	NC008790.1:17102..17287	1.00	1.00	0.95	0.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00470952: hypothetical protein	NC008790.1:17560..17778	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00471111: hypothetical protein	NC008790.1:17771..17974	1.00	1.00	0.64	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00470960: hypothetical protein	NC008790.1:17978..19771	1.00	1.00	0.94	0.97	1.00	1.00	1.00	0.99	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Site-specific recombinase2C resolvase family	NC008790.1:19894..20508	1.00	1.00	1.00	1.00	0.68	0.68	0.69	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Virulence-associated protein 2	NC008790.1:20505..20882	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.54	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00470038: hypothetical protein	NC008790.1:20873..21196	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
VirB2	NC008790.1:21367..21630	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
hypothetical protein	NC008790.1:21615..21737	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
VirB4	NC008790.1:21703..24411	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
Phage Rha protein	NC008790.1:24422..24988	1.00	1.00	0.89	0.90	0.94	0.92	0.92	0.91	0.73	0.92	0.90	0.87	0.93	0.83	0.86	0.87	0.74	0.92	1.00	1.00	0.88	0.00	0.00	0.00
FIG00471987: hypothetical protein	NC008790.1:24985..25650	1.00	1.00	0.30	0.33	0.27	0.31	0.30	0.30	0.28	0.31	0.27	0.29	0.28	0.29	0.26	0.27	0.29	0.26	1.00	1.00	0.27	0.00	0.00	0.00
Single-stranded DNA-binding protein	NC008790.1:25684..26109	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00470457: hypothetical protein	NC008790.1:26125..26400	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
VirB5	NC008790.1:26404..27375	1.00	1.00	1.00	1.00	0.97	0.98	0.98	0.97	1.00	0.98	0.97	0.98	0.97	0.97	0.97	0.97	1.00	0.97	1.00	1.00	1.00	0.00	0.00	0.00
VirB6	NC008790.1:27372..28370	1.00	1.00	1.00	1.00	0.64	0.63	0.64	0.66	1.00	0.66	0.68	0.66	0.67	0.68	0.64	0.64	1.00	0.63	1.00	1.00	1.00	0.00	0.00	0.00
VirB7	NC008790.1:28345..28527	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
VirB8	NC008790.1:28517..29179	1.00	1.00	1.00	1.00	0.68	0.69	0.68	0.68	1.00	0.70	0.70	0.66	1.00	0.69	0.67	0.68	1.00	0.68	1.00	1.00	1.00	0.00	0.00	0.00
VirB9	NC008790.1:29176..30063	1.00	1.00	1.00	1.00	0.66	0.65	0.64	0.78	1.00	0.66	0.63	0.80	1.00	0.81	0.59	0.65	1.00	0.63	1.00	1.00	1.00	0.00	0.00	0.00
VirB10	NC008790.1:30069..31265	1.00	1.00	1.00	1.00	0.61	0.47	0.60	0.59	1.00	0.59	0.60	0.58	1.00	0.60	0.60	0.58	1.00	0.61	1.00	1.00	1.00	0.00	0.00	0.00
VirB11	NC008790.1:31246..32238	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
VirD42C ATPase required for T-DNA transfer	NC008790.1:32239..34050	1.00	1.00	1.00	1.00	0.96	0.98	0.99	0.99	1.00	0.99	1.00	0.98	1.00	0.98	1.00	0.99	1.00	0.99	1.00	1.00	0.99	0.16	0.00	0.29
cag pathogenicity island protein (cag12)	NC008790.1:34064..34468	1.00	1.00	1.00	1.00	0.73	1.00	1.00	0.81	1.00	0.78	0.69	0.75	1.00	0.67	0.70	0.77	1.00	0.03	1.00	1.00	1.00	0.00	0.00	0.00
IncQ protein TraQ (RP4 TrbM homolog)	NC008790.1:34480..35244	1.00	1.00	1.00	1.00	0.49	1.00	1.00	0.45	1.00	0.00	0.33	0.00	1.00	0.29	0.00	0.00	1.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00469957: hypothetical protein	NC008790.1:35247..36044	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.79	1.00	0.74	0.77	0.79	1.00	0.82	0.80	0.74	1.00	0.81	1.00	1.00	1.00	0.00	0.08	0.00
FIG00470273: hypothetical protein	NC008790.1:36115..36735	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.89	0.00
DNA topoisomerase III	NC008790.1:36999..39191	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
FIG00470802: hypothetical protein	NC008790.1:39297..40718	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
DNA topoisomerase III	NC008790.1:40737..40916	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.39	0.31	0.43
Tetracydine resistance protein TetO	NC008790.1:41275..43194	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00
hypothetical protein	NC008790.1:43432..43545	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00
TolA protein	NC008790.1:43617..44285	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.42	0.51	0.52	0.52	0.00	0.41	0.53	0.52	0.02	0.50	0.00	0.00	0.00	0.25	0.00	0.00
FIG00471024: hypothetical protein	NC008790.1:44244..44627	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
hypothetical protein	NC008790.1:44580..45002	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

A series of genes encoding for hypothetical proteins annotated close to the replication protein were missing or showed up to 50% coverage in plasmids that lacked the pTet replication protein (Table 3.6). These proteins carry a partial CRISPR/Cas domain (FIG00471069_hypothetical_protein), a partial Ras/GTPase enhancer domain (FIG00471024_hypothetical_protein), and a partial TolA domain involved in several cellular processes, from the uptake of *coli*cins and filamentous DNA to pathogenesis (TolA_protein).

Two other hypothetical proteins from the reference plasmid pTet were absent in the pTet-like plasmids from a large number of the strains analysed (Table 3.7): FIG00469644 hypothetical protein, present in 11 of the plasmids and which encodes for a partial acetyl transferase or sialyl tranferase largely found in *C. jejuni* (Gilbert *et al.,* 2000) and FIG00469861 hypothetical protein, which carries a conserved domain shared with bacteriophage resistance proteins and is present in 11 analysed strains.

The pTet plasmid from 3 strains, Pet 93 327, Poultry MB3 and Poultry MB12 lacked the tetracycline resistance gene. Moreover, all these plasmids were missing genes flanking the *tet(O)* gene. In particular: the replication protein and the proteins associated with it (described above in this section), the DNA topoisomerase III encoded by a gene upstream *tet(O),* and two hypothetical proteins—FIG00470802 encoding for a conserved domain of unknown function and a hypothetical protein encoding for no putative conserved domains.

The strains Poultry MB8, Hi81214 and Hi81006 possessed the *tet(O)* gene and a gene encoding a single hypothetical protein flanking the *tet(O)* gene but lacked any other genes from the reference pTet plasmid. The *tet(O)* in these strains is integrated in the chromosome.

Table 3.6- Conserved domains identified in the hypothetical protein sequences of FIG00471069_hypothetical_protein FIG00471024_hypothetical_protein and TolA_protein in the strain 81-176. For each hypothetical protein, a screenshot of the results retrieved from NCBI's CD-Search search engine, the name of each conserved domains, the accession number of the conserved domain, the region showing the predicted domain, and the e-value is reported.

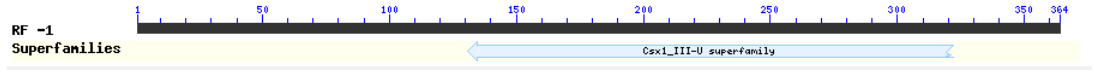
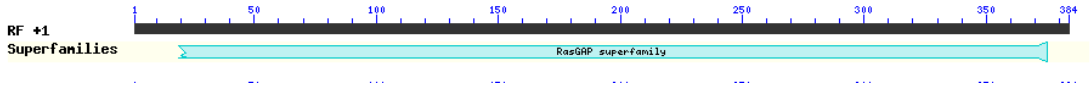
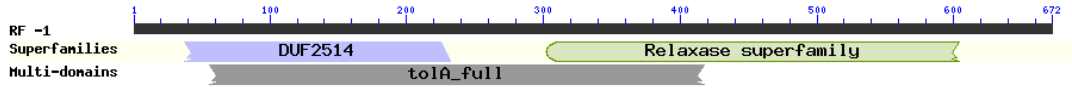
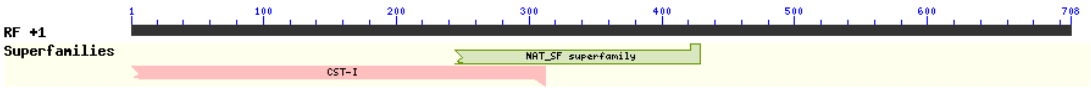
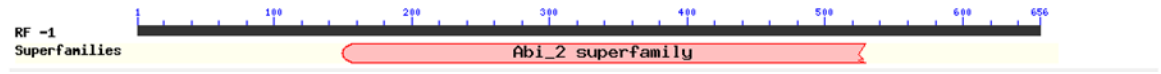
FIG00471069_hypothetical_protein		
		
CRISPR/Cac system associated domain Csx1		
cl21516	130-321	6.28e-6
FIG00471024_hypothetical_protein		
		
Ras GTPase Activating domain, enhancer of the hydrolysis of GTP bound to Ras-GTPases		
cl02569	19-375	7.29e-5
TolA_protein		
		
TolA protein, part of the Tol/Pal complex involved in maintaining outer membrane integrity. Also implicated in transport of *coli*cins, filamentous dna and pathogenesis Relaxase/mobilisation nuclease domain		
TIGR02794 cl21589	55-417 301-603	2.62e-6 6.65e-7

Table 3.7 - Conserved domains identified in the hypothetical protein sequences of FIG00469644_hypothetical_protein, FIG00469861_hypothetical_protein. For each hypothetical protein a screenshot of the results retrieved from NCBI's CD-Search search engine, the name of each conserved domain, the accession number of the conserved domain, the region showing the predicted domain, and the e-value is reported.

FIG00469644_hypothetical_protein		
		
N-Acetyl transferase superfamily NAT_SF Alpha-2,3-Sialyltransferase CST-I		
cl17187 cl05511	244-429 1-312	1.98e-3 8.06e-3
FIG00469861_hypothetical_protein		
		
Abi-2 superfamily, Abi-like protein found in various bacterial species involved in phage resistance		
cl01988	145-528	7.35e-5

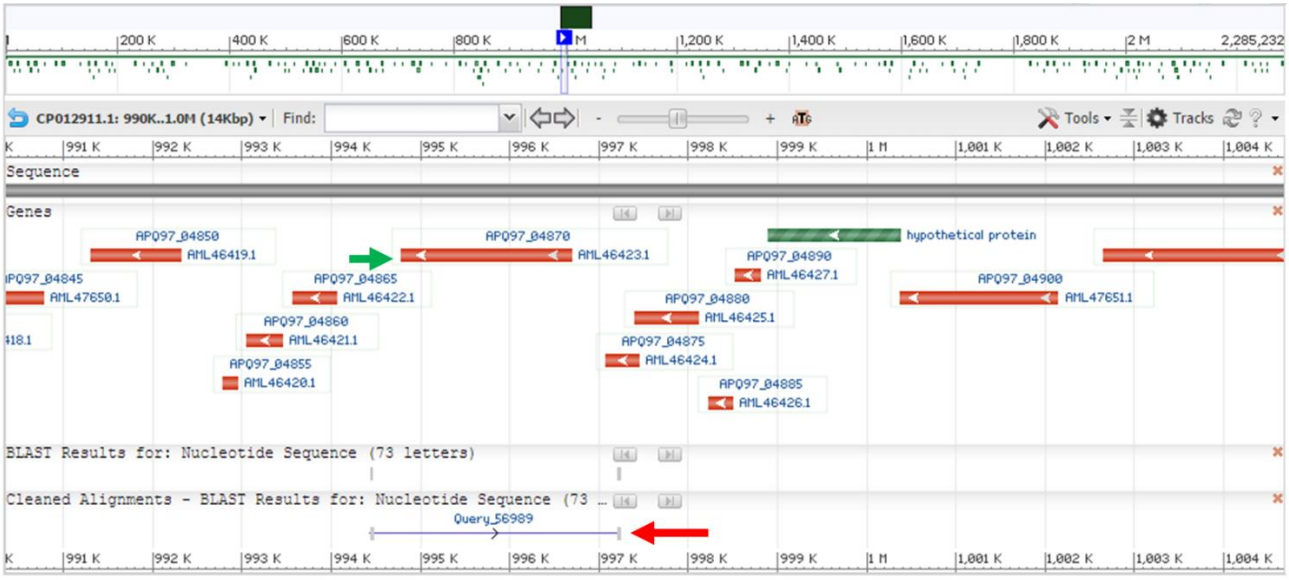
The integration point of the *tet(O)* gene from strains Poultry MB8 and Hi81214 showed a repeated sequence, highlighted in the following sequence:

5’_ATTTAGGGTATAACAAAATAACCCACCCGAATATC<*tet(O)*___HypotheticalProtein>CTATGAGCCTTTCAAAGCGTTCCTGTGCCTGTCTGTTG_3’

A BLASTn alignment of this sequence against the bacterial genomes database highlights how the same sequence is present in several strains of *C. jejuni* and in different bacterial species, such as *Streptococcus* spp. (Figure 3.8), when the gene *tet(O)* is present chromosomically.

The plasmid of about 37 kbp that can be identified in the Poultry MB8 appears to belong to a different plasmid family compared to the other families described in this chapter: it shows >90% similarity to the plasmid pVir (Bacon *et al.,* 2000), and for this reason it will not be discussed further (data not shown).

Figure 3.8 Localisation of *tet(O)* gene in the genome of *Streptococcus* species. Graphic representation of a BLASTn alignment of the 73 bp surrounding the *tet(O)* gene in the strain MB8 against the complete genome of *Streptococcus suis* NSUI060. The red arrow indicates the query sequence; the green arrow indicates the gene *tet(O)* in *S. suis* genome.



3.1.3.2 Global plasmid comparison

To understand the differences and the similarities amongst all the plasmids, a global comparison of all the plasmids sequenced from *C. jejuni* strains (19 plasmids) and the original pTet (NC008790.1) was next carried out.

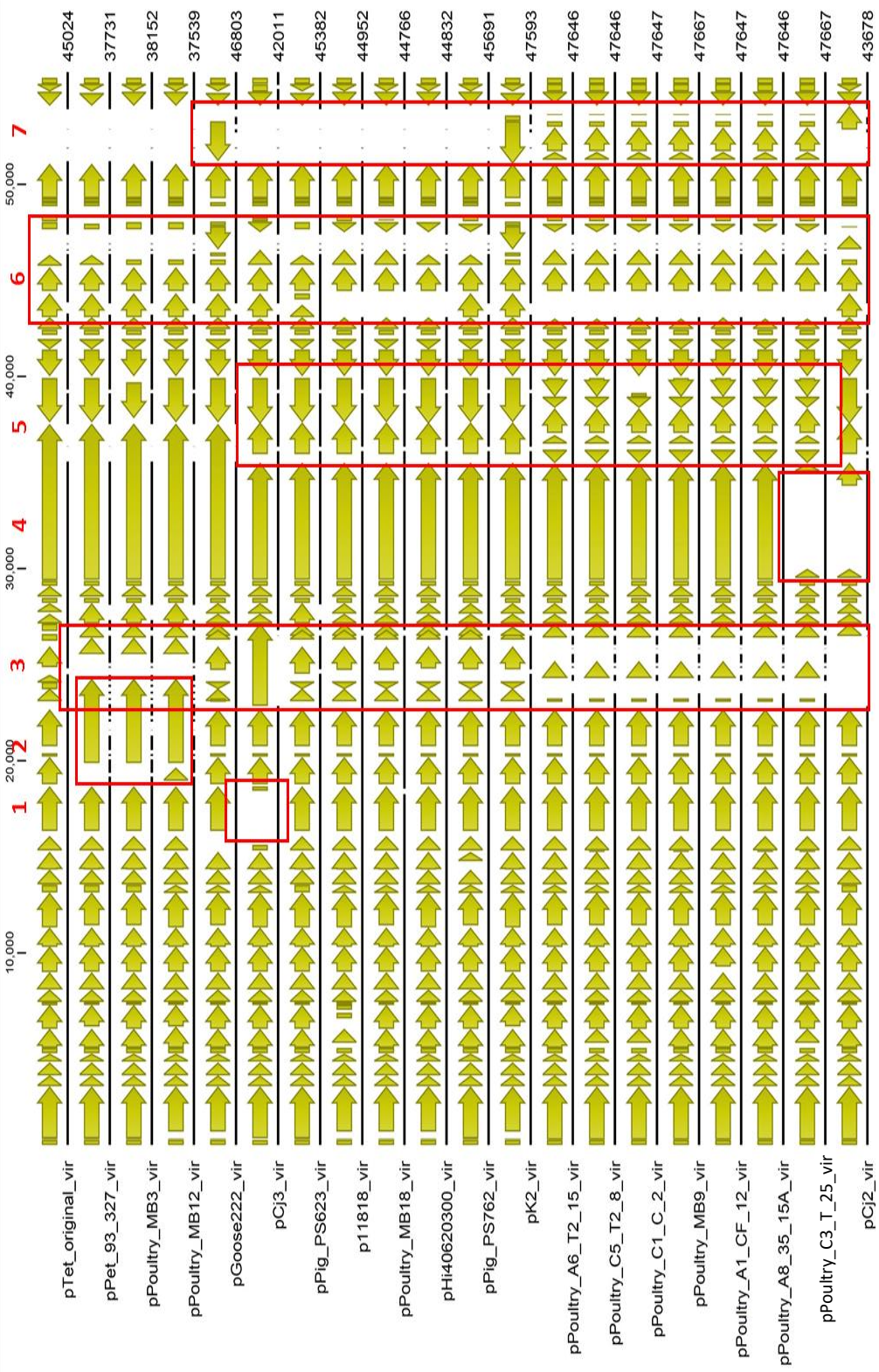
In order to do so, the plasmids were linearised starting from the gene annotated as *virB2*. This was chosen because it was present in each of the plasmids. The plasmids were aligned using the internal aligner of CLC sequence viewer (standard parameters) and visualised with the same software (ClustalW2 algorithm, Thompson *et al.*, 1994). This software was chosen because it was relatively simple to retain the annotation in the alignment and this allowed a better visualisation. The output is shown in Figure 3.9, and Table S3 (Supplementary material) shows the annotated genes identified in each plasmid.

This alignment revealed that the most conserved parts of the alignment correspond to the type four secretion system cluster, a helicase gene (involved in the ATP-dependent unwinding of DNA), a *virD2* homologous gene involved in the DNA transfer, a DNA primase gene 40000 bp from the beginning of the alignment, and a hypothetical protein (FIG00470960) possessing a domain associated with DNA repair (Table 3.8).

This analysis also highlights several differences amongst the plasmids, unnoticeable according to the analysis performed in the previous section. This is because the approach previously adopted did not take into account features absent from the original plasmid pTet.

Several hypothetical proteins were completely absent from the plasmid pTet of strain 81-176 and are highlighted for the first time in this analysis.

Figure 3.9 - Representation of the global alignment between 19 pTet like plasmids and the plasmid pTet (NC008790.1). pTet_original = sequence of pTet accession NC008790.1. Each black line represents the sequence of a plasmid linearised at the start of the annotated gene *virB2.* Numbers are related to the length of each plasmid. Red squares highlight sections of the alignment described in paragraph 3.1.3.2; red numbers on top of the alignment are referred in the text when the relevant part of the alignment id described.



Starting from the beginning of the alignment, (1) the CDS encoding the DNA topoisomerase in pCj3 was absent despite the fact that it appears to be present in Table 3.5. It is possible to notice an insertion of 127 bp that seems to disrupt the CDS is present in this strain. (2) where 17 plasmids show the presence of the *tet(O)* gene, the three plasmids (pPet_93_327, pPoultry_MB3 and pPoultry_MB12) lacking the tetracycline resistance gene show a protein annotated as hypothetical protein pVir0015, conserved in the *C. jejuni* pVir plasmid.

(3) CDS encoding for several different hypothetical proteins are present in the region of the alignment around the “replication protein” gene of pTet. This region was already described as highly variable in Figure 3.7. It is possible to identify a gene encoding for a small hypothetical protein in the three strains lacking the tetracycline resistance gene, a gene encoding for a hypothetical protein FIG00471065 in the strain Cj3, and a gene encoding for a hypothetical protein FIG00471069 in most of the strains isolated from poultry. None of these hypothetical proteins shows any putative conserved domain.

(4) Strains Poultry C3_T_25 and Cj2 lack the complete helicase (snf2 family) gene because of a 127 bp insertion at the 3’ end of it. Downstream of the Helicase (snf2 family) gene, plasmid pTet NC008790.1 has a CDS encoding a hypothetical protein (FIG00469644) carrying a conserved acetyl transferase domain (Table 3.1.3.2.3). (5) Upstream the gene encoding for the hypothetical protein FIG00469644, seven plasmids (p11818, pPoultryMB18, pHi40620300, pPigPS762, pCj2, pK2 and pCj3) show the presence of a gene encoding for a putative protein FIG00472625 carrying a LabA-like conserved domain, involved with the regulation of several cellular process (from circadian rhythm in cyanobacteria to degradation of toxic substance, Taniguchi *et al.,* 2007; Tang *et al.,* 2008). In the same region of the

plasmid, most of strains isolated from poultry (Poultry_A6_T2_15, Poultry_C5_T2_8, Poultry_C1_C_2, Poultry_MB9, Poultry_A1_CF_12, Poultry_A8_35_15_A and Poultry_C3_T_25), show a different gene cluster. It is formed by 5 CDSs, all encoding for hypothetical proteins, with conserved domains associated with acetyl transferase, a porin, type 2 secretion system associated proteins, and with a peptidase for bacteriocin processing (Table 3.9).

(6) Moreover, between 40000 and 50000 bp from the beginning of the alignment, a CDS encoding for a hypothetical protein FIG00469861 (described in section 3.1.3.1) present in pTet was missing from 10 plasmids (p11919, pPoultry_MB18, pHi40620300, pPoultry_A6_T2_15, pPoultry_C5_T2_8, pPoultry_C1_C_2, pPoultry_MB9, pPoultry_A1_CF_12, pPoultry_A8_35_15_A and pPoultry_C3_T_25). In the same area of the alignment, a CDS encoding for a serine/threonine kinase signal transductor was only found in 2 plasmids (pK2 and pGoose222).

(7) Towards the 3’ end of the alignment, a gene cluster is present that is associated with strains isolated from poultry. It contains four CDSs, with domains associated with signal transduction and regulation (TIR_2 bacterial toll like receptor), virulence (VapD), and protein metabolism (metallo-beta-lactamase) (Table 3.1.3.2.3). Plasmid pCj2 and pGoose222 possessed a single gene encoding for a hypothetical protein with no CD identifiable in the same region, while a membrane aromatic cluster protein (lipoprotein associated with *Mycoplasma*) is identifiable in the plasmid pK2.

Table 3.8 - Conserved domains identified in the hypothetical protein sequence of FIG00470960_hypothetical_protein. A screenshot of the results retrieved from NCBI's CD-Search search engine: the name of each conserved domain, the accession number of the conserved domain, the region showing the predicted domain, and the e-value is reported.

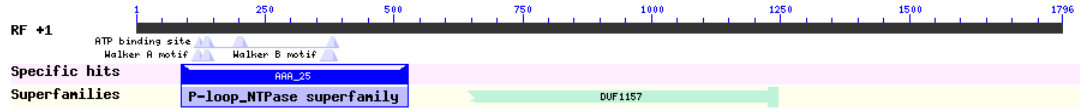
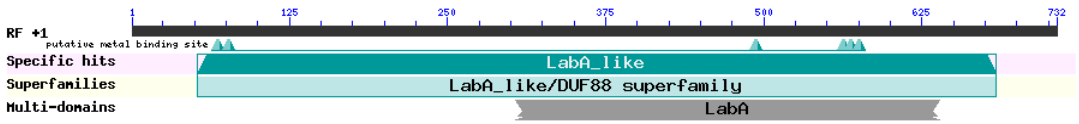
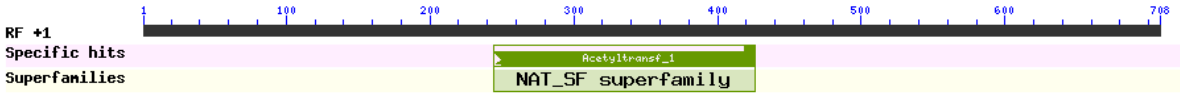
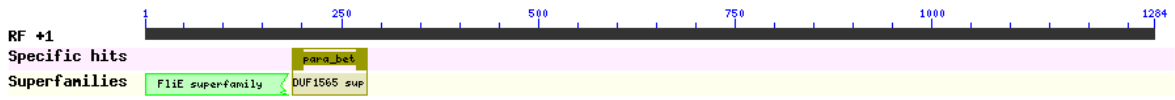
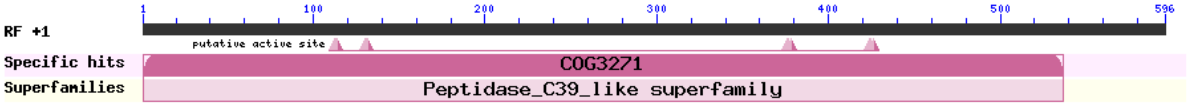
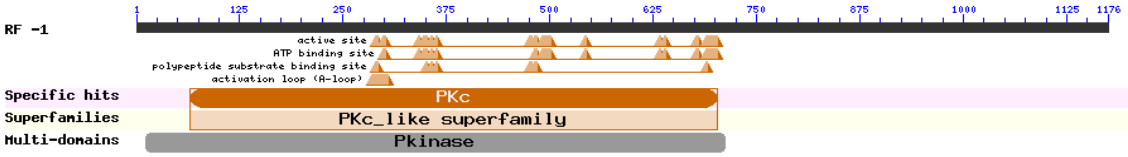
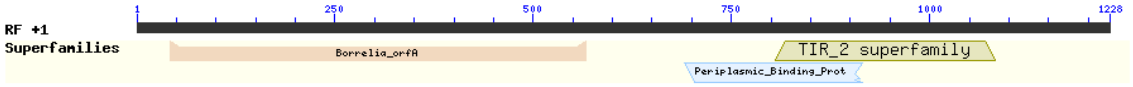
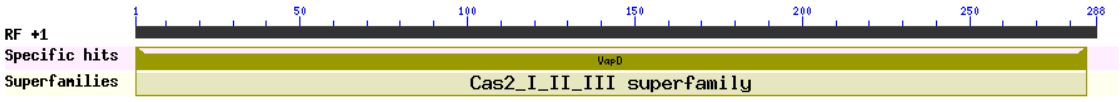
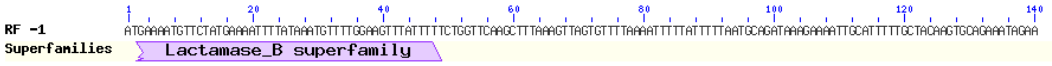
FIG00470960_hypothetical_protein		
		
AAA_25, AAA domain, N-loop NTPase presumably involved in DNA repair		
pfam13841	88-528	1.72e-15

Table 3.9 - Conserved domains identified in the hypothetical protein sequences of FIG00472625_hypothetical_protein, FIG00469644_hypothetical_protein, FIG00470281_hypothetical_protein, Hypothetical protein (FIG|354242.88.peg.1849 on the annotation of strain A6/T2/15), FIG00628667_hypothetical_protein, Serine threonine protein kinase PrkC, Hypothetical protein (FIG|354242.88.peg.1854 on the annotation of strain A6/T2/15), Virulence associated protein D (VapD), Hypothetical protein (FIG|354242.88.peg.1851 on the annotation of strain A6/T2/15). For each hypothetical protein, a screenshot of the results retrieved from NCBI's CD-Search search engine, the name of each conserved domain, the accession number of the conserved domain, the region showing the predicted domain, and the e-value is reported.

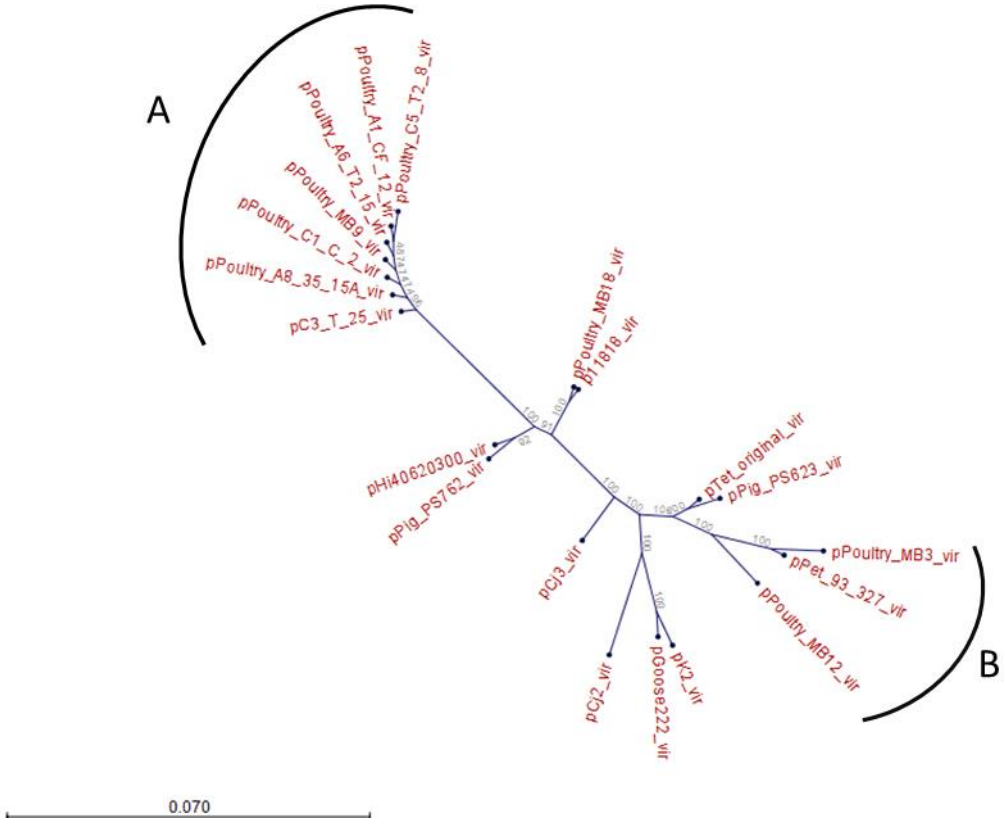
FIG00472625 hypothetical protein		
		
LabA_like; conserved domain involved in several cellular processes, no defined function.		
cd06167	52-684	6.40e-20
FIG00469644 hypothetical protein		
		
Acetyltransferase (GNAT) family		
pfam00583	244-426	2.21e-03
FIG00470281 hypothetical protein		
		
Porin superfamily, form non-specific channels for small hydrophillic molecules		
cl21487	7-354	2.96e-03
Hypothetical protein (FIG|354242.88.peg.1849)		
		
FliE super family Parallel beta-helix repeat, found in many carbohydrate-binding proteins and sugar hydrolases.		

cl09139 TIGR03804	1-183 187-282	3.76e-03 2.11e-03
FIG00628667 hypothetical protein		
		
Predicted double-glycine peptidase		
COG3271	1-537	4.53e-40
Serine threonine protein kinase PrkC		
		
PKc_superfamily, Catalytic domain of protein kinase		
cd00180	64-702	1.99e-44
Hypothetical protein (FIG|354242.88.peg.1854)		
		
Borrellia_orfA, Plasmid-associated dna-repeats in Borrellia spp. TIR_2 superfamily domain, bacterial toll-like receptor		
cl20231 cl22869	43-567 805-1083	2.83e-6 2.67e-7
Virulence associated protein D (VapD)		
		
VapD, Virulence associated protein (unknown function)		
COG3309	1-285	7.63e-42
Hypothetical protein (FIG|354242.88.peg.1851)		
		
Metallo beta lactamase superfamily		
cl00446	1-48	3.45e-3

The alignment shown in Figure 3.9 was used to infer the phylogeny of the plasmids. A tree was produced using the software CLC Sequence Viewer (Neighbor-joining algorithm, Jukes-Cantor distance measuring method, 100 bootstrap replicates). The result of this analysis is shown in Figure 3.10.

The tree has several characteristics worth highlighting: the plasmids isolated from the same source do not always cluster together (e.g.: the two plasmids isolated from pig, the plasmids isolated from a human host, or the plasmids isolated form poultry do not cluster with each other). Moreover, there is not a clear clustering related to the length of the plasmid. Plasmids in cluster A exhibit a higher average length than the plasmids in cluster B. When compared to the plasmid pTet NC008790.1 (isolated from 81-176), the strains identified in cluster A appear to have acquired several gene—for instance, the gene cluster towards the 3’ end of the alignment described in above (area highlighted with number 7 in Figure 3.9). The strains identified in the cluster B on the other hand appear to have lost several genes, in particular the gene cluster around the tetracycline resistance gene. None of the other strains in this analysis follows any pattern associating phylogeny and length of the plasmid.

Figure 3.10 - Phylogenetic tree obtained with the alignment of the linearised plasmid sequences (shown in Figure 3.9). Neighbor-joining algorithm, Jukes-Cantor distance measuring method, 100 bootstrap replicates. (pC3_T_25 = pPoultry_C3_T_25). Plasmids highlighted by A show a higher average length than plasmids highlighted by B.



3.2 - Survey of all available *Campylobacter jejuni* genome sequences for the presence of pTet

To understand the distribution of the plasmid pTet amongst *C. jejuni*, I built a database of all available *C. jejuni* genome sequences (updated in May 2016). This included 3211 strains from a study conducted by the Sanger Institute between 2008 and 2012, monitoring an Oxford Hospital's patients (http://www.ebi.ac.uk/ena/data/view/ERP000129, "The genomic speciation of *Campylobacter*", unpublished, study accession number PRJEB2075). It also included 78 strains from a Genome-Wide Association Study (GWAS) reported by Sheppard *et al.,* 2013 (http://datadryad.org/resource/doi:10.5061/dryad.28n35) and 679 strains sequenced in the last 15 years and retrieved from the NCBI database (completed and draft stage, retrieved in May 2016 - http://www.ncbi.nlm.nih.gov/genome/). Finally, I included 37 strains sequenced at Exeter University, before or during this work.

The final database included a total number of 4005 strains and is summarised in Table 3.10. The complete table with the meta-data associated for each strain is available in electronic format (Electronic Supplementary Material E3).

In this analysis, the presence of each gene of the plasmid pTet in the database of 4005 *C. jejuni* has been assessed. Briefly, raw reads were aligned to the pTet sequence, using Bowtie2 as described at 2.10.5. If the raw reads were not available, the draft genome or the complete genome was treated with an ad-hoc perl script to obtain a set of artificial raw reads (generate_reads.pl). All of the alignments were then analysed using CoverageBed. Each file reporting the "breadth of coverage" for each alignment was parsed with a perl script (compare_coverages.pl) to obtain a single table showing the breadth of coverage of each single gene in each strain.

Table 3.11 shows a subset of the full analysis, reporting the breadth of coverage of the 50 strains showing the highest number of genes covered and the 50 strains showing the lowest number of genes covered. The full result of this analysis is available in electronic format (Electronic Supplementary Material E3).

Table 3.10 - Number of strains of *C. jejuni* genome sequences used for surveying the presence of the plasmid pTet. Columns are (from left to right): number used for each group of strains (i to iv), number of strain in each group, study from which the group was retrieved, status of the genome sequences in each group.

	# of strains	Study	Sequencing status
(i)	3211	"The genomic speciation of campylobacter" (Unpublished)	Raw reads
(ii)	78	Sheppard et al. 2013	Contigs (draft)
(iii)	679	Various - retrieved from http://www.ncbi.nlm.nih.gov/genome/	Complete - Contigs (draft)
(iv)	37	Seqeunced in-house	Contigs (draft)
Total	4005		

Table 3.11 - Breadth of coverage of each ORF of pTet in strain 81-176 (NC008790.1) in the 50 strains showing the highest number of genes covered and the 50 strains showng the lowest number of gene covered. One different strain per row. Columns are (from left to right): strain name, study name (according to Table 3.10: SANGER = i, GWAS = ii, NCBI = iii, IN-HOUSE = iv), origin of isolation (where unavailable = N/A), year of isolation (where unavailable = N/A), breadth of coverage of each gene annotated in pTet (NC008790.1). Last column is the sum of the breadths of coverage. Each cell shows the value of coverage percentage in each strain, and the dot present in each cell is a graphical representation of each number (completely black = 80-100%, three quarters black = 60-80%, half black = 40-60%, one quarter black = 20-40%, blank = 0-20%). The full table is available in xlsx format (Electronic Supplementary Material E3).

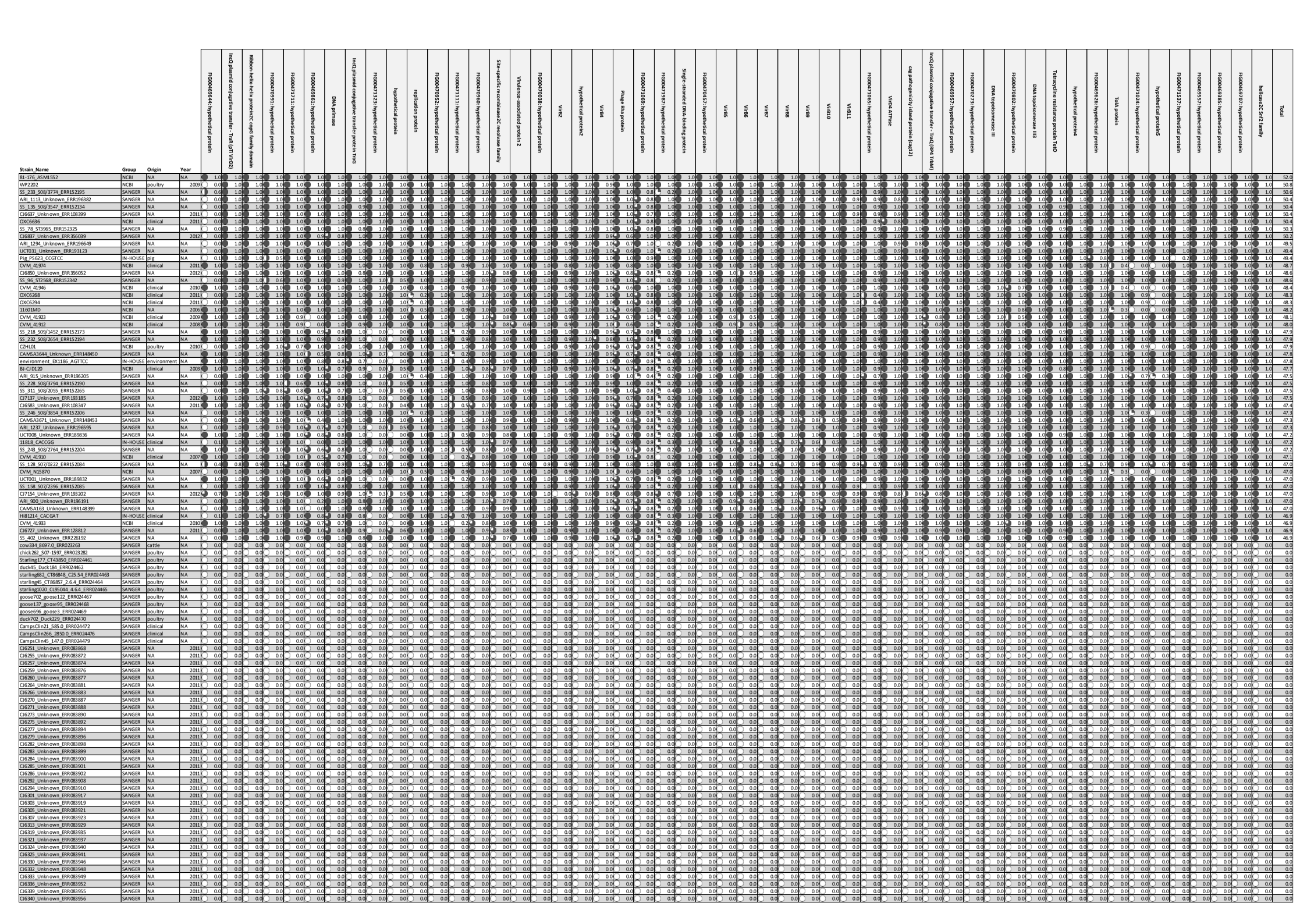
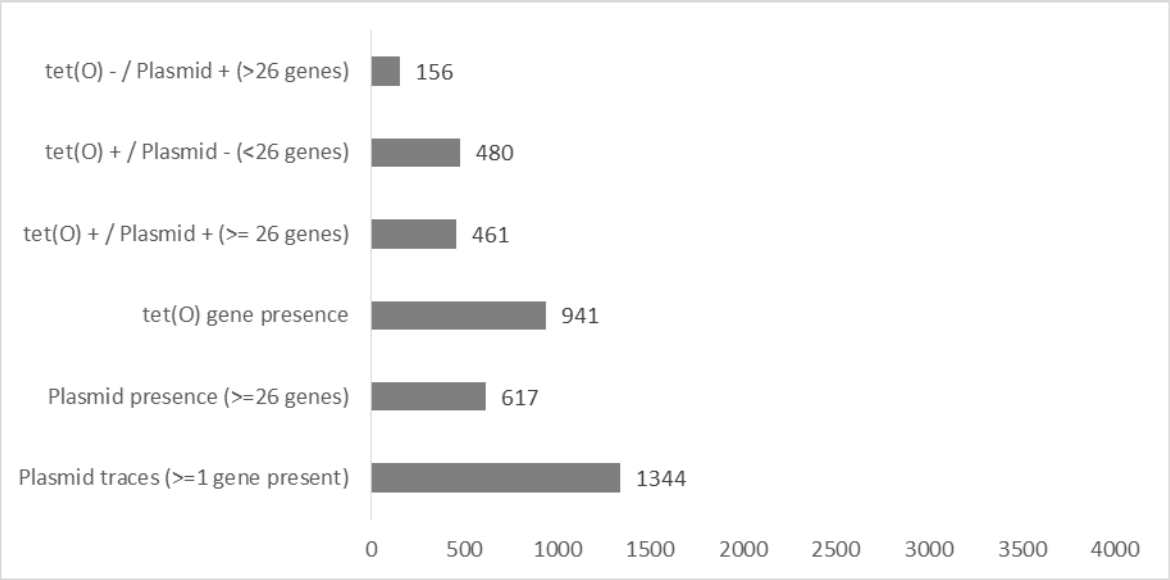


FIG004656644: hypothetical protein
IncQ plasmid conjugative transfer - TraR (pT1 VirG2)
Ribbon-helix-helix protein,CC cog5 family domain
FIG004709991: hypothetical protein
FIG004717111: hypothetical protein
FIG004698611: hypothetical protein
DNA primase
IncQ plasmid conjugative transfer protein TraG
FIG004713123: hypothetical protein
hypothetical protein
replication protein
FIG004709552: hypothetical protein
FIG004711111: hypothetical protein
FIG004709560: hypothetical protein
Site-specific recombinase2C resolvase family
Virulence-associated protein 2
FIG004700838: hypothetical protein
VirB2
hypothetical protein2
VirB4
Phage Rha protein
FIG004710169: hypothetical protein
FIG004721987: hypothetical protein
Single-stranded DNA-binding protein
FIG004704537: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11
FIG004710655: hypothetical protein
VirD4 ATPase
cag pathogenicity island protein (cagT2)
IncQ plasmid conjugative transfer - TraQ (RP4 TrbM)
FIG004699357: hypothetical protein
FIG004702723: hypothetical protein
DNA topoisomerase III
FIG004708602: hypothetical protein
DNA topoisomerase III3
Tetracycline resistance protein TetO
hypothetical protein4
FIG004696226: hypothetical protein
TolA protein
FIG004721024: hypothetical protein
hypothetical protein5
FIG004721537: hypothetical protein
FIG004695557: hypothetical protein
FIG004693385: hypothetical protein
FIG004697707: hypothetical protein
helicase2C SWI2 family
Total
Strain_Name
Group
Origin
Year

3.2.1 Overview of the pTet presence survey

A general analysis of the number of strains possessing pTet-like plasmids is reported in this paragraph. As shown in Figure 3.11, a total of 1344 (33.6%) strains show the presence of at least 1 gene annotated in the 81-176 pTet plasmid. To establish the presence of a gene, I defined a cut-off of 80% in breadth of coverage in the coverage table described in Section 3.2. Of this set of strains, 617 had more than 50% (n>26) of the total 81-176 pTet genes. This indicates that 15.4% of the total database analysed has more than half of the annotated genes for the plasmid pTet (NC008790.1) identified in this study with RAST (Overbeek *et al.*, 2014), and 45.9% of the strains showed evidence of possessing a pTet-like plasmid. Of the 1344 strains with at least 1 pTet gene in their genome, 941 strains show the presence of the tetracycline resistance gene *tet(O)*, while 403 lacked this gene sequence. From the analysis of the group showing the presence of at least 50% of the plasmid pTet 81-176, it emerges that a total of 461 strains out of 617 (74.7%) showed the presence of the *tet(O)* gene, while 156 (25.3%) strains show a complete lack of *tet(O)* sequence. One of the main reasons why pTet is believed to be conserved is because it encodes resistance to the tetracycline (Batchelor *et al.*, 2004). This analysis highlights a possible absence of this association, as the 25.3% of *C. jejuni* sequenced strains showing the presence of the majority of the pTet plasmid from 81-176 do not appear to possess the *tet(O)* gene.

Figure 3.11 - Overview of presence of pTet genes in the 4005 genome sequences of *C. jejuni* analysed. A gene is defined as present in a genome if it is covered for at least 80%, when the breadth of coverage is assessed as described in Section 2.1. tet(O) + or - = Tetracycline resistance gene is present or absent, plasmid + or - = at least 26 genes of the plasmid pTet are present, plasmid "traces" = at least one gene of the plasmid pTet is present.



The heatmap in Figure 3.12 A (also reported in full electronically in Electronic Supplementary Material E4) is obtained using the data matrix of breadth of coverage of the 52 pTet NC008790.1 genes (columns) in the 617 strains showing more than 26 genes present (rows). The clustering tree of the column dimension is also reported in Figure 3.12 B. Cluster one, corresponding to the rightmost side of the heatmap, shows genes conserved in the majority of the strains having the pTet-like plasmid. A second cluster shows less conserved genes characterised by a higher variability (including some of the *vir* genes of the type four secretion system), and four clusters alternatively present in different groups of strains, including the *tet(O)* gene and the replication protein.

If the same type of analysis is expanded to the strains exhibiting the presence of at least one plasmid gene, the result obtained is shown in Figure 3.13 A and B (heatmap in Figure 3.13 A is available as Electronic Supplementary Material E5). In this case, the data matrix used as input for the heatmap clustering shows the breadth of coverage of the 52 pTet NC008790.1 genes (columns) in the 1344 strains showing at least 1 of the genes present (rows). The plasmid genes are clustered in three main clusters in this case: this is because the small frequency differences between the 617 strains showing the presence of the majority of the plasmid genes are masked by the high number of *C. jejuni* strains showing the lack of the genes in the cluster II, and the presence of the tetracycline resistance gene chromosomally (cluster III).

Figure 3.12 - Cluster analysis of the plasmid pTet genes (1). (A) Heatmap built with the breadth of coverage table of each gene annotated in the plasmid pTet (NC008790.1) in each of the 617 strain showing at least the presence of 26 genes of the plasmid pTet. Clustering distance for columns and rows = euclidean, clustering method = complete. The 52 genes annotated on the plasmid pTet are in the column, the 617 strains are in the rows. Each cell is coloured from white (0% coverage) to dark green (80-100%) as reported in the legend. (B) Clustering tree for the genes (columns). The rightmost element is on top (rotated 90 degrees left). The clusters identified on the tree (numbered from I to IV) are also highlighted on the heatmap. The full heatmap is available in pdf format (Electronic Supplementary Material E4).

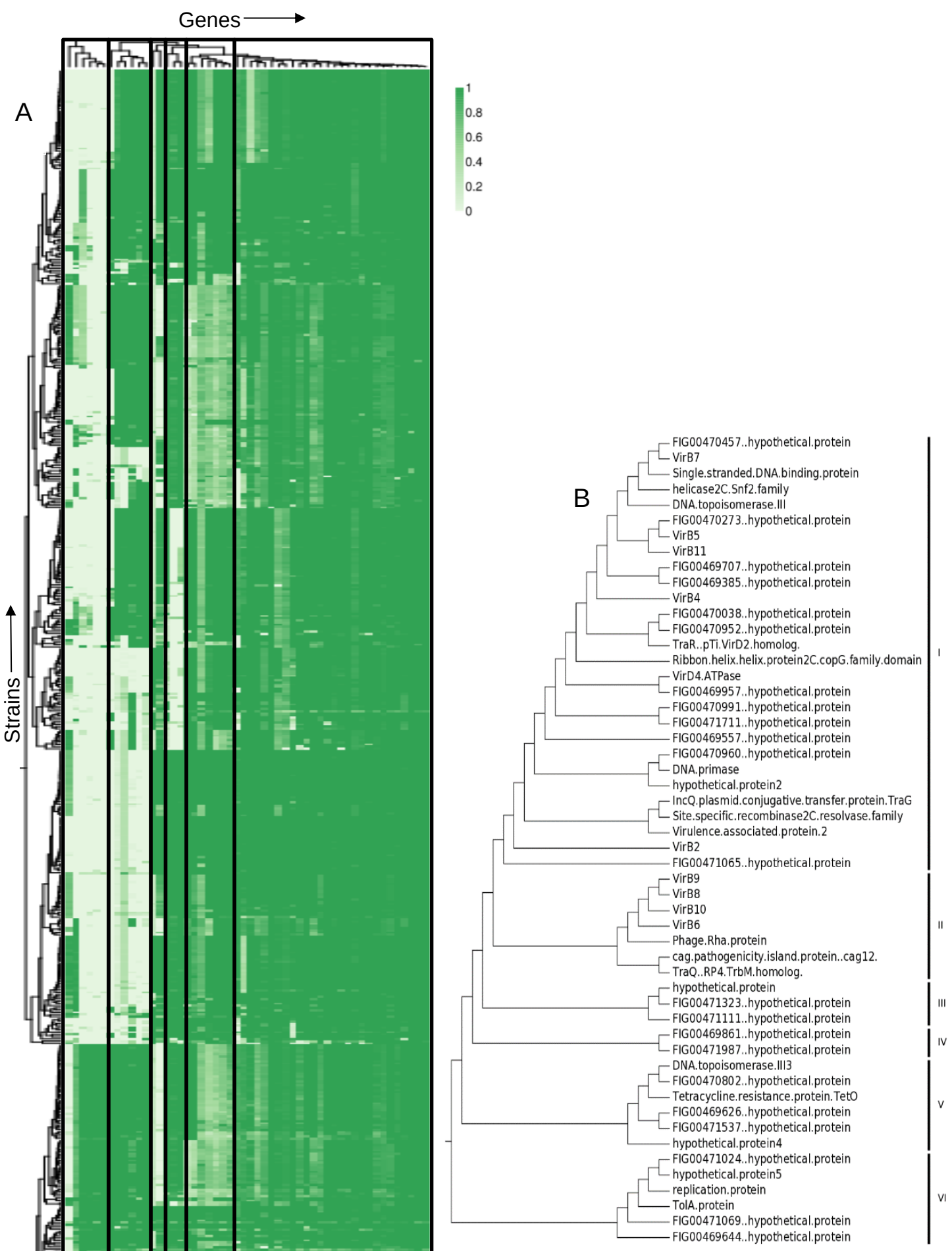
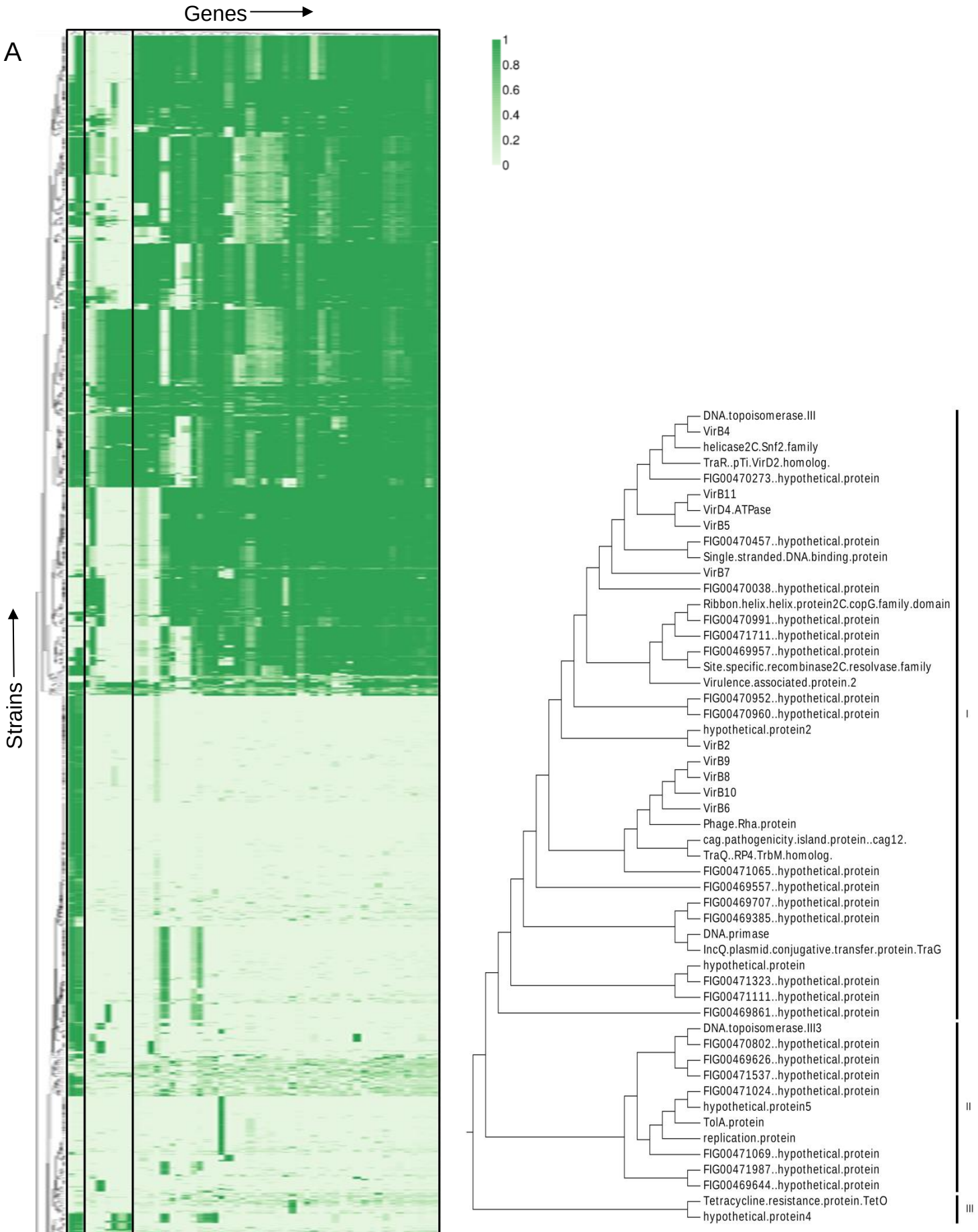

A
Genes
Strains
1
0.8
0.6
0.4
0.2
0
B
FIG00470457..hypothetical.protein
VirB7
Single.stranded.DNA.binding.protein
helicase2C.Snf2.family
DNA.topoisomerase.III
FIG00470273..hypothetical.protein
VirB5
VirB11
FIG00469707..hypothetical.protein
FIG00469385..hypothetical.protein
VirB4
FIG00470038..hypothetical.protein
FIG00470952..hypothetical.protein
TraR..pTi.VirD2.homolog.
Ribbon.helix.helix.protein2C.copG.family.domain
VirD4.ATPase
FIG00469957..hypothetical.protein
FIG00470991..hypothetical.protein
FIG00471711..hypothetical.protein
FIG00469557..hypothetical.protein
FIG00470960..hypothetical.protein
DNA.primase
hypothetical.protein2
IncQ.plasmid.conjugative.transfer.protein.TraG
Site.specific.recombinase2C.resolvase.family
Virulence.associated.protein.2
VirB2
FIG00471065..hypothetical.protein
VirB9
VirB8
VirB10
VirB6
Phage.Rha.protein
cag.pathogenicity.island.protein..cag12.
TraQ..RP4.TrbM.homolog.
hypothetical.protein
FIG00471323..hypothetical.protein
FIG00471111..hypothetical.protein
FIG00469861..hypothetical.protein
FIG00471987..hypothetical.protein
DNA.topoisomerase.III3
FIG00470802..hypothetical.protein
Tetracycline.resistance.protein.TetO
FIG00469626..hypothetical.protein
FIG00471537..hypothetical.protein
hypothetical.protein4
FIG00471024..hypothetical.protein
hypothetical.protein5
replication.protein
TolA.protein
FIG00471069..hypothetical.protein
FIG00469644..hypothetical.protein
I
II
III
IV
V
VI

Figure 3.13 - Cluster analysis of the plasmid pTet genes (2). (A) Heatmap built with the breadth of coverage table of each gene annotated in the plasmid pTet (NC008790.1) in each of the 1344 strains showing at least the presence of 1 gene of the plasmid pTet. Clustering distance for columns and rows = euclidean, clustering method = complete. The 52 genes annotated on the plasmid pTet are in the column, the 617 strains are in the rows. Each cell is coloured from white (0% coverage) to dark green (80-100%) as reported in the legend. (B) Clustering tree for the genes (columns). The rightmost element is on top (rotated 90 degrees left). The clusters identified on the tree (numbered from I to IV) are also highlighted on the heatmap. The full heatmap is available in pdf format (Electronic Supplementary Material E5).


A
Genes
Strains
1
0.8
0.6
0.4
0.2
0
DNA.topoisomerase.III
VirB4
helicase2C.Snf2.family
TraR..pTi.VirD2.homolog.
FIG00470273..hypothetical.protein
VirB11
VirD4.ATPase
VirB5
FIG00470457..hypothetical.protein
Single.stranded.DNA.binding.protein
VirB7
FIG00470038..hypothetical.protein
Ribbon.helix.helix.protein2C.copG.family.domain
FIG00470991..hypothetical.protein
FIG00471711..hypothetical.protein
FIG00469957..hypothetical.protein
Site.specific.recombinase2C.resolvase.family
Virulence.associated.protein.2
FIG00470952..hypothetical.protein
FIG00470960..hypothetical.protein
hypothetical.protein2
VirB2
VirB9
VirB8
VirB10
VirB6
Phage.Rha.protein
cag.pathogenicity.island.protein..cag12.
TraQ..RP4.TrbM.homolog.
FIG00471065..hypothetical.protein
FIG00469557..hypothetical.protein
FIG00469707..hypothetical.protein
FIG00469385..hypothetical.protein
DNA.primase
IncQ.plasmid.conjugative.transfer.protein.TraG
hypothetical.protein
FIG00471323..hypothetical.protein
FIG00471111..hypothetical.protein
FIG00469861..hypothetical.protein
DNA.topoisomerase.III3
FIG00470802..hypothetical.protein
FIG00469626..hypothetical.protein
FIG00471537..hypothetical.protein
FIG00471024..hypothetical.protein
hypothetical.protein5
TolA.protein
replication.protein
FIG00471069..hypothetical.protein
FIG00471987..hypothetical.protein
FIG00469644..hypothetical.protein
Tetracycline.resistance.protein.TetO
hypothetical.protein4
I
II
III

3.2.3 Association with the presence of the plasmid pTet and the time of isolation, the origin of the strain and phylogeny

The previous analysis was performed on a set of 4005 *C. jejuni* isolates, composed of strains isolated at different times, by different labs, following different standards for recording the metadata of each strain. For this reason, each of the following Sections describes just a subset of the entire dataset. For instance, strains with unknown year of isolation are not included in Section 3.2.3.1.

3.2.3.1 Year of isolation

For this analysis, a subset of strains from the Sanger Institute collection and from the NCBI database was used. Most of the strains from the Sanger Institute collection were isolated from symptomatic patients between 2009 and 2012, and the genome sequencing deposited on the NCBI database were obtained from strains isolated between 1980 and 2014: the metadata were retrieved from the project page at the MLST database of *Campylobacter jejuni/coli* (http://pubmlst.org/perl/bigsdb/, Jolley and Maiden, 2010) and from the BioSample database entry associated with each genome assembly (http://www.ncbi.nlm.nih.gov/biosample). Unfortunately, the number of strains isolated each year is not consistent and not all the strains included in the collection have a timestamp. The numbers of strains isolated in each year is reported in Table 3.12.

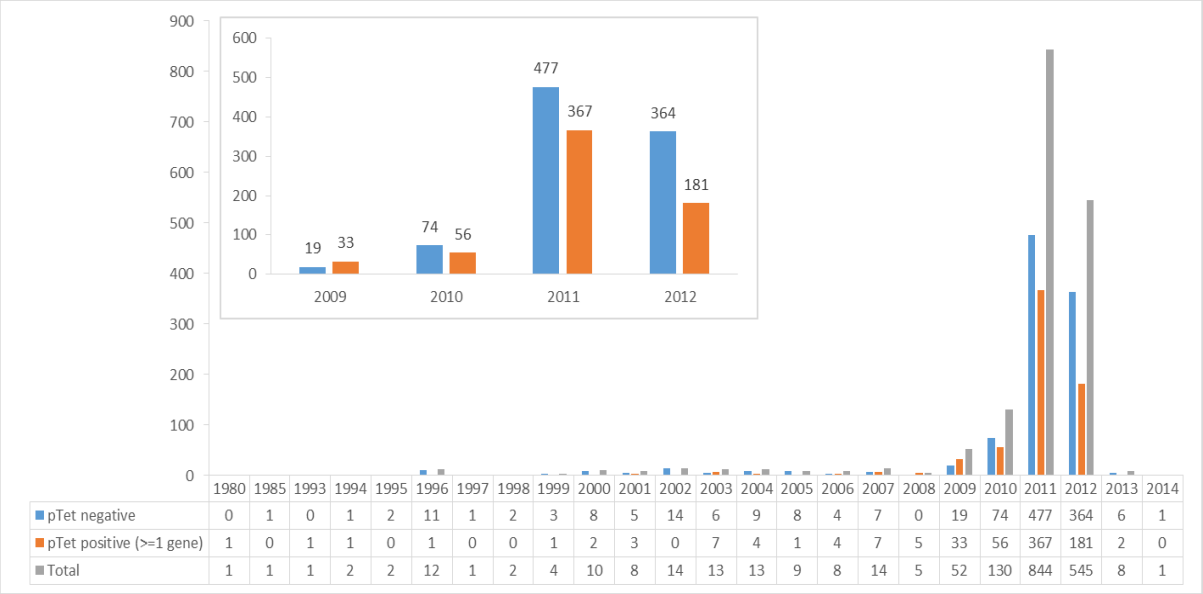
The number of strains possessing a pTet-like plasmid is reported in the barplot in Figure 3.13 for strains showing the presence of at least 1 gene annotated in pTet and in Figure 3.14 for strains showing the presence of at least half of the genes annotated in the pTet.

From this analysis, I identified strains retaining the pTet-like plasmid every year between 2006 and 2013—particularly, between year 2009 and 2012, when the sample size is larger, the number of strains positives for the plasmid presence was 34.5% (2009), 13.1% (2010), 20.1% (2011), 13.6% (2012).

Table 3.12 - Number of strains isolated per year included in the analysis described in 3.2.3.1.

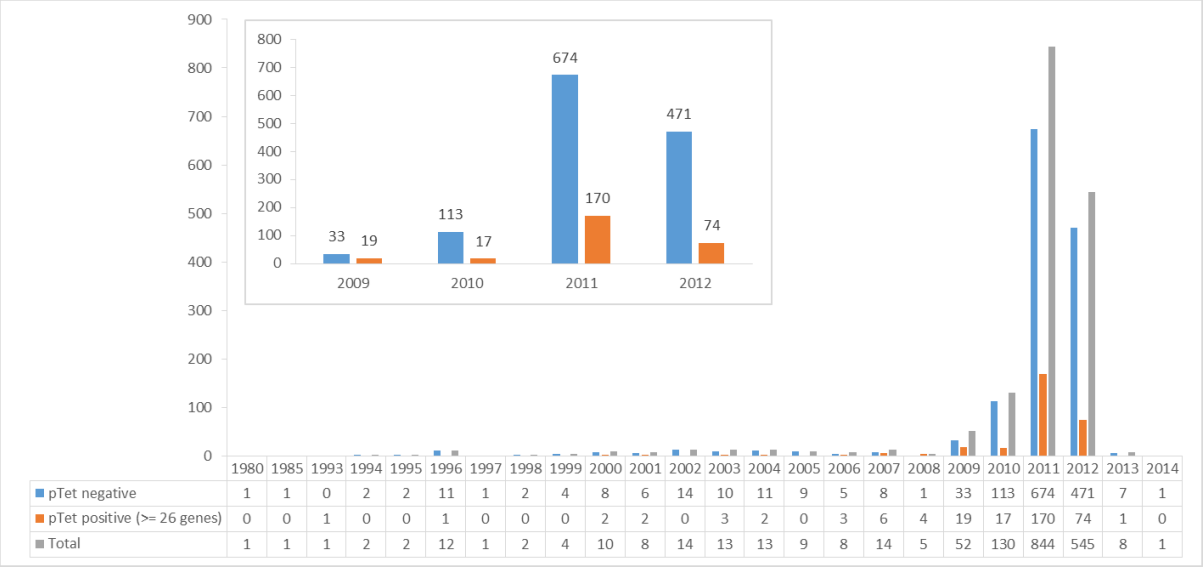
Year	Number of CJ strains
1980	1
1985	1
1993	1
1994	2
1995	2
1996	12
1997	1
1998	2
1999	4
2000	10
2001	8
2002	14
2003	13
2004	13
2005	9
2006	8
2007	14
2008	5
2009	52
2010	130
2011	844
2012	545
2013	8
2014	1

Figure 3.14 - Number of strains per year showing the presence of at least 1 gene annotated on the plasmid pTet. The inset shows a zoom for the years 2009 to 2012.



	1980	1985	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
pTet negative	0	1	0	1	2	11	1	2	3	8	5	14	6	9	8	4	7	0	19	74	477	364	6	1
pTet positive (>=1 gene)	1	0	1	1	0	1	0	0	1	2	3	0	7	4	1	4	7	5	33	56	367	181	2	0
Total	1	1	1	2	2	12	1	2	4	10	8	14	13	13	9	8	14	5	52	130	844	545	8	1

Figure 3.15 - Number of strains per year showing the presence of at least 26 genes annotated on the plasmid pTet. The inset shows a zoom for the years 2009 to 2012.



	1980	1985	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
pTet negative	1	1	0	2	2	11	1	2	4	8	6	14	10	11	9	5	8	1	33	113	674	471	7	1
pTet positive (>= 26 genes)	0	0	1	0	0	1	0	0	0	2	2	0	3	2	0	3	6	4	19	17	170	74	1	0
Total	1	1	1	2	2	12	1	2	4	10	8	14	13	13	9	8	14	5	52	130	844	545	8	1

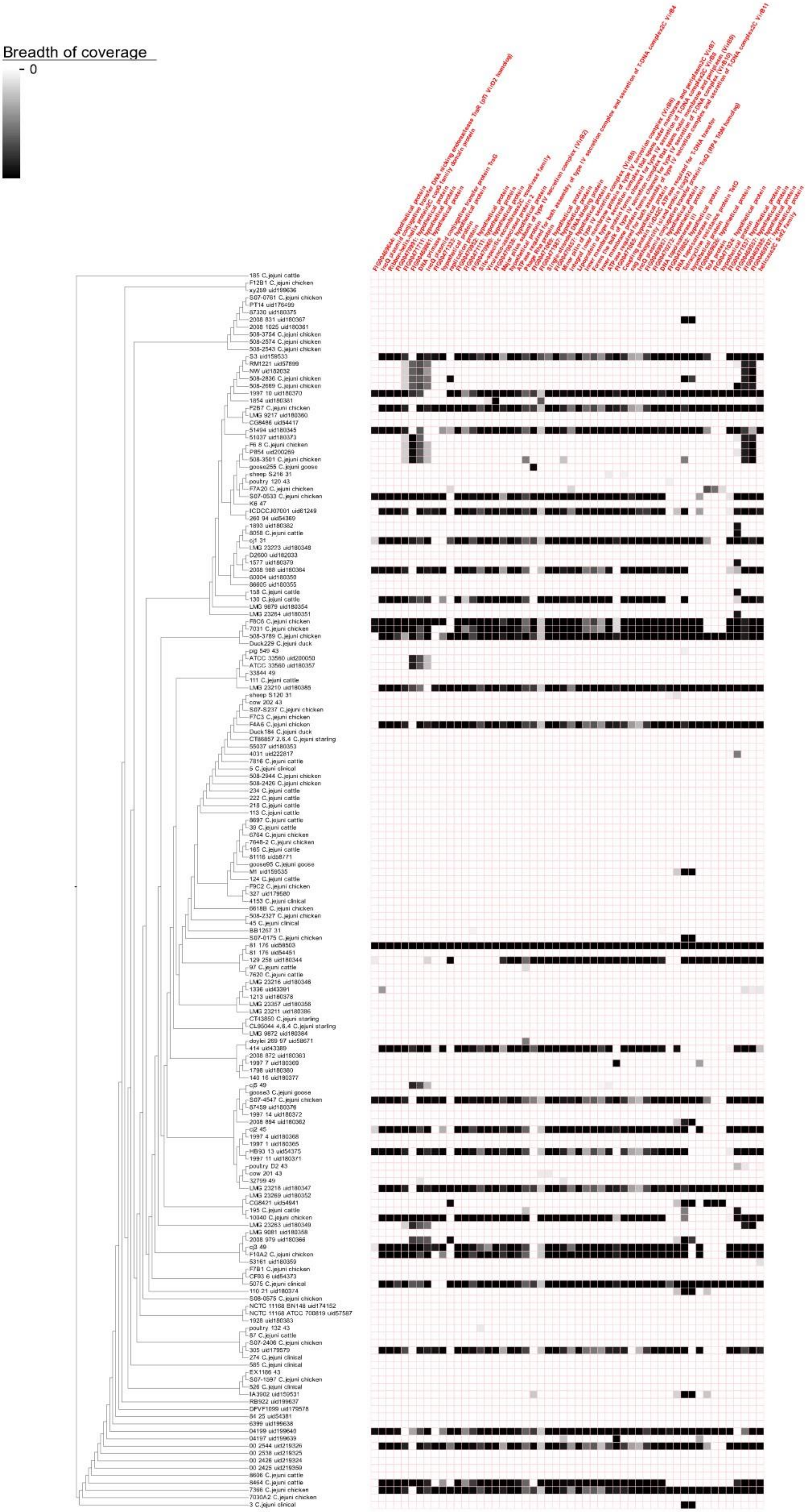
3.2.3.2 Phylogeny

In order to understand whether the possession of the pTet-like plasmids is associated with the phylogeny of *C. jejuni*, the presence of each gene was associated to the phylogenetic analysis of the Multi-Locus Sequence Tag of *C. jejuni.* A 400 to 600 bp fragment of 7 core genes have been extracted from the genome sequence of a subset of 168 strains. This included the strains in the GWAS study (Sheppard *et al.,* 2013), the completed genomes retrieved from the NCBI database (NCBI bacteria), and 20 strains sequenced in-house up to 2013. The MLST of *C. jejuni* is described in "Multilocus Sequence Typing System for *C. jejuni*" (Dingle, 2001), curated by http://pubmlst.org/*Campylobacter*/, which includes the sequence of internal fragments of 7 housekeeping genes: *aspA* (aspatase), *glnA* (glutamine synthetase), *gltA* (citrate synthase), *glyA* (serine hydroxy methyl transferase), *pgm* (phospho glucomutase), *tkt* (transketolase), *uncA* (ATP synthase alpha subunit). The sequence of these fragments in the reference strain 11168 were retrieved from the NCBI database (NC_002163.1), aligned against the entire draft or complete genome of each strain included in the analysis using BLASTn, and the matching sequence extracted.

The sequences of the 7 fragments were concatenated, pairwise-aligned with MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) and the phylogenetic tree was built as described in 2.8.6. The phylogenetic tree was then annotated with EvolView (version 2, https://evolview.codeplex.com/, He *et al.*, 2016): the presence or absence of each gene is reported as breadth of coverage in the alignment of raw reads or artificial reads (as described in Section 2.8.5 of materials and methods).

Figure 3.16 shows the result of the analysis. It is clear that none of the clusters are associated with the presence of one or more plasmid genes.

Figure 3.16 - MLST analysis tree for a subset of 168 strains and presence of pTet genes. The table shows the presence of each gene annotated on the plasmid pTet in each strain as percent of coverage as reported in the legend (white = 0% coverage, black = 100% coverage). Strains are listed on the rows of the table, corresponding to the leaves of the phylogenetic tree, genes are listed on the columns.


Breadth of coverage
0

3.2.3.3 Origin

For this analysis a subset of strains of known origin of isolation was used. This included the NCBI sequence assembly whose origin could be retrieved from the "Biosample" database (http://www.ncbi.nlm.nih.gov/biosample), 36 of the strains sequenced in-house, the strains from the GWAS analysis and part of the strains retrieved from the Sanger Institute Database. The total number of strains with recorded origin used in this section was 821, and they were divided in 7 groups as reported in Table 3.13. The poultry group includes every strain isolated from chicken, goose, duck, and turkey; the clinical group includes the strains isolated from humans in a clinical environment; the environmental group includes strains that were not isolated from animals or stool (e.g. from soil or water).

The results are reported in Figures 3.17 and 3.18 and are associated with strains showing at least one gene or half of the genes of pTet plasmid.

The analysis reveals that 206 strains of known origin of isolation show at least half the 52 genes of the 81-176 pTet covered. Of these, 64 are found in the 234 strains isolated from poultry (27.3% of poultry strains), 131 in the 503 clinical isolates (26.0%), 9 in the 61 strains isolated from cattle (14.7%), and 2 out of 3 strains isolated from pig. None of the environmental strains or the strains isolated from sheep and rabbit exhibit the presence of the majority of the plasmid pTet genes.

Almost the same percentage of clinical isolates and strains isolated from poultry shows more than 26 genes associated with the plasmid pTet (more than 25% of the isolates in both groups), while 14.7% of isolates from cattle possessed a pTet-like plasmid.

Table 3.13 - Number of strains per different origin included in the analysis described in Section 3.2.3.3.

Origin	Number of CJ strains
cattle	70
clinical	503
environment	7
pig	3
poultry	234
rabbit	1
sheep	3

Figure 3.17 - Number of strains per different origin of isolation showing the presence of at least 1 gene annotated on the plasmid pTet. The inset shows a zoom for the strains isolated from cattle, poultry and clinical environment (symptomatic humans).

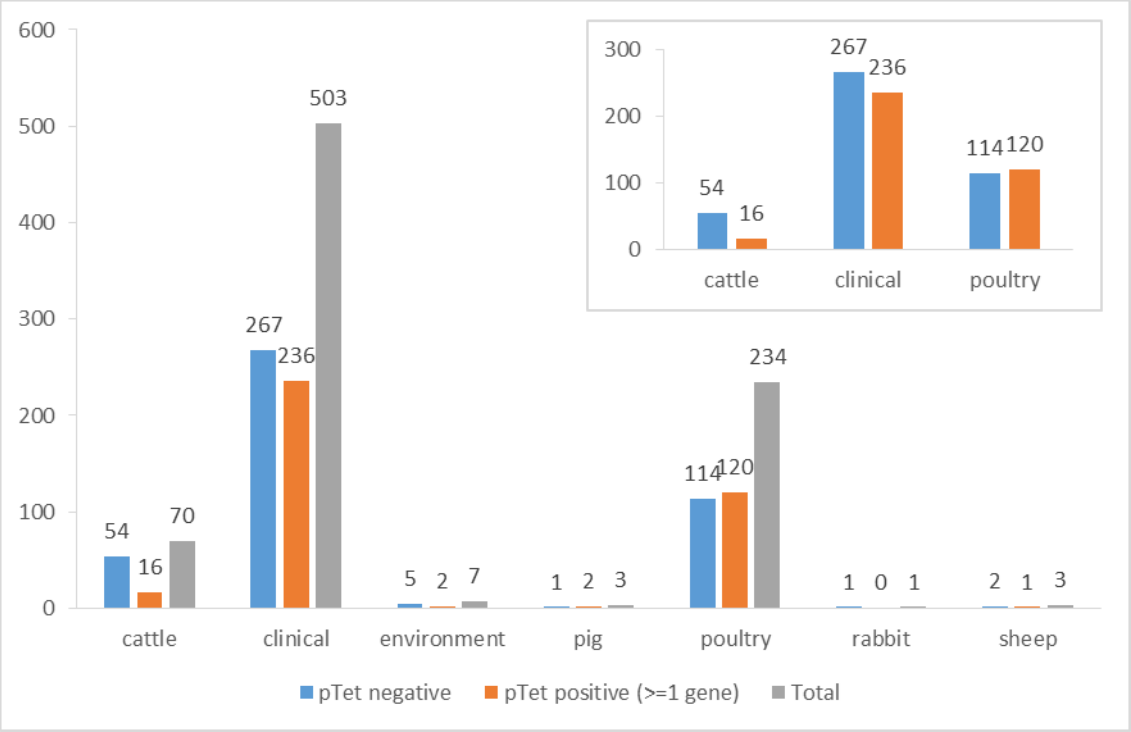
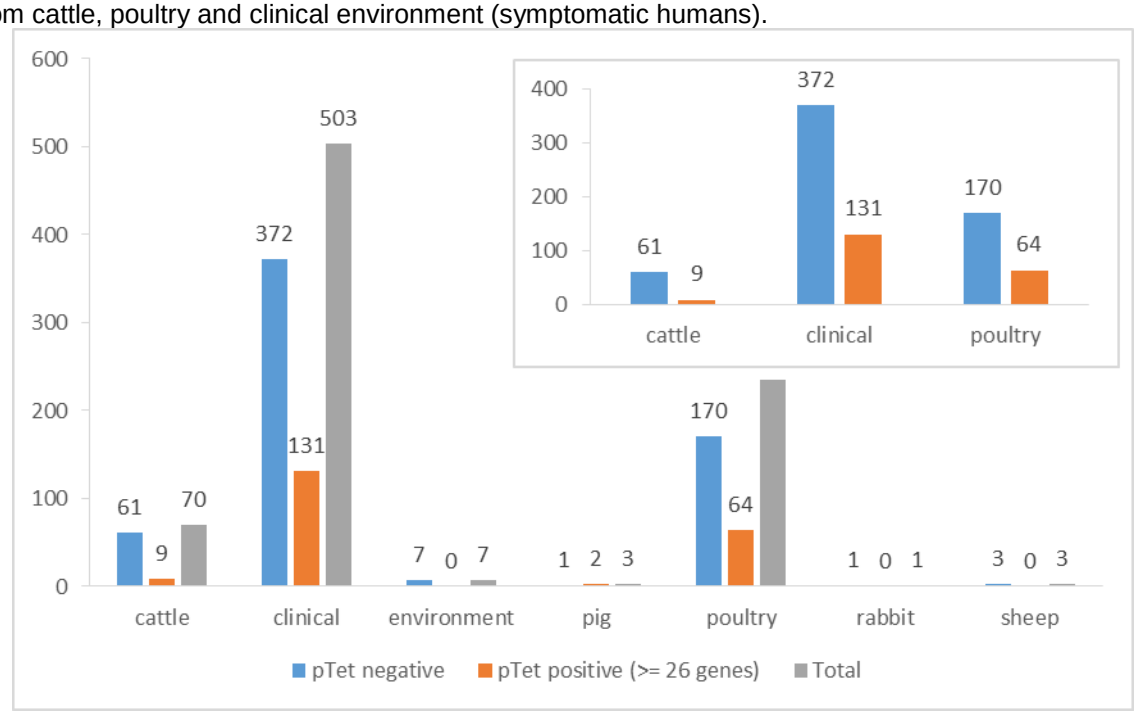
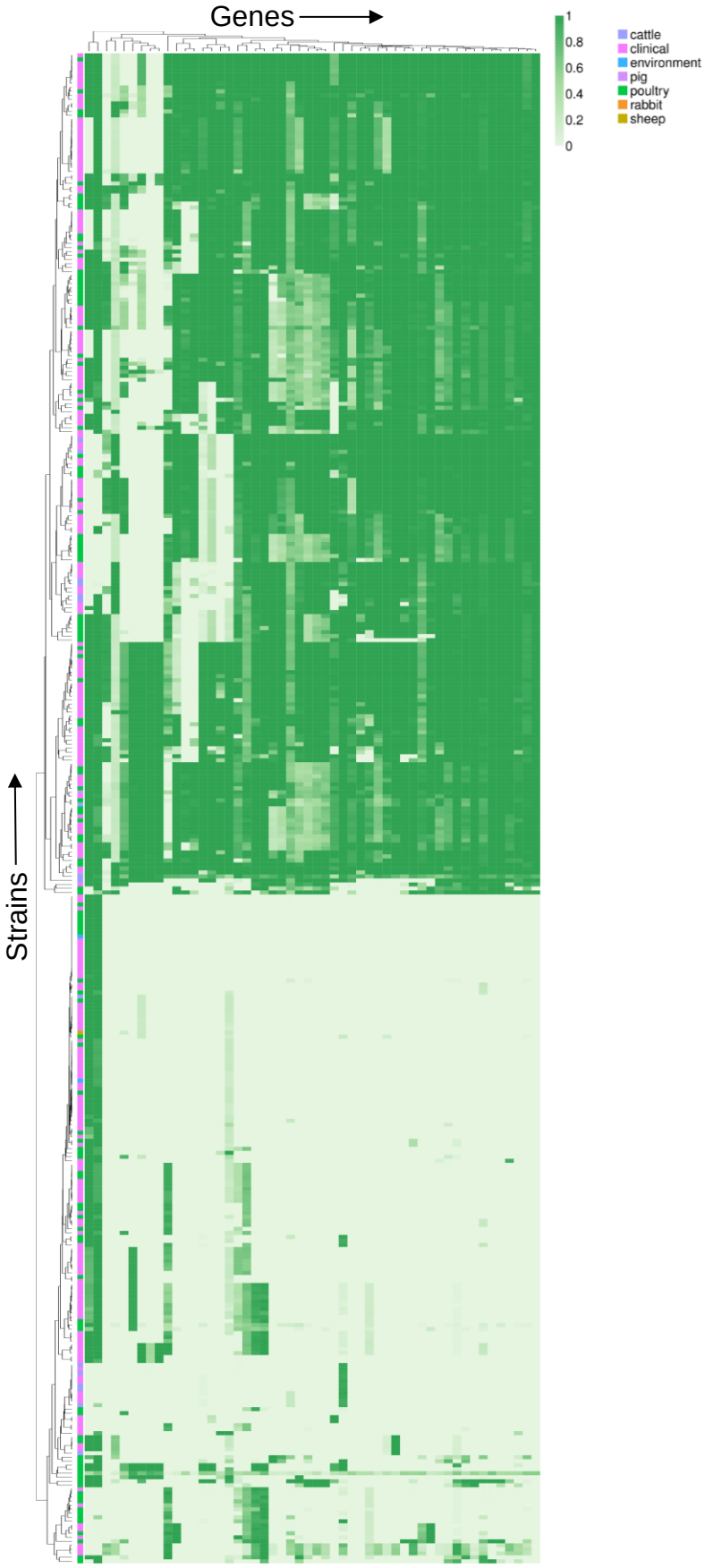


Figure 3.18 - Number of strains per different origin of isolation showing the presence of at least 26 genes annotated on the plasmid pTet. The inset shows a zoom for the the strains isolated from cattle, poultry and clinical environment (symptomatic humans).



The heat-map reported in Figure 3.19 shows clustering according to gene presence and the rows are annotated with the origin of each strain (for the 377 strains showing the presence of at least 1 gene annotated on the plasmid pTet), showing that the presence of different pTet genes is not linked to a particular origin of the strain.

Figure 3.19 - Cluster analysis of the plasmid pTet genes and the strains of known origin. Heatmap built with the breadth of coverage table of each gene annotated in the plasmid pTet (NC008790.1) in each of the 377 strains of known origin showing at least the presence of 1 gene of the plasmid pTet. Clustering distance for columns and rows = euclidean, clustering method = complete. The 52 genes annotated on the plasmid pTet are on the column, the 617 strains are on the rows. Each cell is coloured from white (0% coverage) to dark green (80-100%) as reported in the legend. The leftmost row shows the annotation related to the origin of the strain (as shown in legend).



3.3 Plasmid pan-genome analysis

Pan-genome studies are described in the literature (Rouli *et al.,* 2015) and are used to define the set of core and accessory genes of a bacterial species. A pan-genome study aims to identify the core and the accessory genome of a species or a set of strain belonging to a species: every gene annotated in every genome sequence that is part of the study is compared, in order to identify every orthologous sequence present in at least one genome. The list these unique orthologous genes is the pan-genome of the species. Usually, if the genes are present in every strain, they are defined as “core genes”, essential for the bacterium survival or lifestyle; otherwise, they are defined as “accessory genes” (Ozer *et al.*, 2014). Several pan-genome studies have been published in recent years, including studies on *C. jejuni* (Meric *et al.,* 2014; Lefeubre *et al.,* 2010). I treated the plasmids as genomes to try to define groups of core and accessory plasmid genes, focusing particularly on the pTet-like plasmids. With the exception of two works in *Acinetobacter* that used a completely different approach to the one used here (Fondi *et al.*, 2010, where the focus was the plasmids rearrangement and the exchange of information between plasmids and chromosomes), this is the first time that plasmids isolated from different strains of a bacterial species have been treated as a plasmid pan-genome to assess the presence and infer the importance of each gene for the make up of a plasmid family.

The study of 20 plasmids isolated and sequenced highlighted a variety of different plasmid organisations. The analysis of the entire database of *C. jejuni* strains described in Section 3.2 shows a remarkably high number of strains with traces of plasmid sequences. In order to separate the comparison of these plasmids from the single reference pTet, I decided to pool all the sequenced plasmids in *C.*

jejuni in a pan-genome-like study in order to assess the variability in gene presence. At the end of this section, I will narrow down the plasmid pan-genome analysis only to the pTet-like plasmids, and define a core and an accessory set of genes for this plasmid family.

3.3.1 Overview of the genes distribution of *C. jejuni* plasmid pan-genome

To assess the full list of plasmid genes, the full list of sequenced *C. jejuni* plasmids was retrieved (corrected in February 2016, Table 3.14). The first analysis included pTet and pTet-like plasmids, along with other sequenced plasmids including pVir plasmids, small “cryptic plasmids” (Jesse *et al.*, 2006; Hiett *et al.*, 2013; pCJ1170-pCJ01-pTIW94- pCJ419) and large plasmids longer than 100 kbp. This allowed the relationships of these plasmids to be explored.

The plasmid pan-genome was built according to the protocol detailed in Materials and Methods, Section 2.10.7. Briefly, in this analysis I assessed the orthologous proteins between the plasmids and extracted one gene sequence for these orthologous proteins. This resulted in a list of 178 genes (the plasmid pan-genome of *C. jejuni*).

In order to assess the presence of each of these 178 genes in the 4005 strains that were described in Section 3.2, sequencing reads were aligned to their DNA sequence as described in Section 2.10.5.

The breadth of coverage of each gene, for each strain, compared to the reference, was used to produce the heatmap in Figure 3.20 (an electronic version is available as Electronic Supplementary Material E6) and the representation of the hierarchical clustering of the genes according to the presence in the database analysed in Figure 3.21. For these analyses, the breath of coverage value was

turned into presence or absence, selecting a cut-off of 0.8 (80% of sequence covered) for a gene to be present. In other words, if the breadth of coverage were greater than 80% the gene was considered present and the coverage value was converted into a 1; if the breadth of coverage were lower than 80% the gene was considered absent and the coverage value was converted into a 0. The heatmap shows how different genes clusters are associated with different strains, which is not surprising considering that this analysis is obtained using several different plasmid families of *C. jejuni*: the different genes clusters are associated to these different plasmids.

The hierarchical clustering is reported in Figure 3.21 in the same order as the heat-map columns, and it is annotated with the frequency of each gene. The dendrogram was produced on the basis of a Jaccard distance matrix built on the presence and absence of each gene in each strain as described in Material and Methods section 2.10.6. The clusters are highlighted by different colours. They are associated with different genes frequencies, and contain genes deriving from different plasmids. In particular we can observe: cluster V in green, containing genes present in pTet-like plasmids (present also in pcjDM T1-T2, as it contains ~40000 bp homologous to pTet); cluster II in red, associated with plasmid pCJDM202, containing, amongst others, genes for the type 6 secretion system; cluster III in light blue, associated with the genes of the plasmid pVir; clusters I and IV in yellow, containing genes deriving from several different pTet-like plasmids, showing an intermediate frequency between pTet-like cluster (V) and pVir cluster (III).

Table 3.14 - Summary of the characteristics of the plasmids included in the *C. jejuni* plasmid pan-genome analysis. Table columns (from right to left): Strain name, name of the plasmid, Accession number (where available), Size in kilobases, G+C percentage, number of putative proteins, number of predicted genes, number of predicted pseudogenes.

Organism	Plasmid name	RefSeq	INSDC	Size (Kb)	GC(%)	Protein	Gene	Pseudogene
Campylobacter jejuni 11818	p11818	This study	This study	44.95	29.06	52	52	-
Campylobacter jejuni poultry C3_T_25	pPoultry_C3_T_25	This study	This study	47.66	28.04	56	56	-
Campylobacter jejuni Cj2	pCj2	This study	This study	43.67	27.57	48	48	-
Campylobacter jejuni Cj3	pCj3	This study	This study	42.01	27.84	46	46	-
Campylobacter jejuni goose 222	pGoose222	This study	This study	46.08	28.13	52	52	-
Campylobacter jejuni Hi40620300	pHi4062300	This study	This study	44.83	29.27	50	50	-
Campylobacter jejuni K2	pK2	This study	This study	47.59	28.77	54	54	-
Campylobacter jejuni Pet_93_327	pPet93_327	This study	This study	37.73	27.19	43	43	-
Campylobacter jejuni Pig_PS623	pPigPS623	This study	This study	45.38	28.72	50	50	-
Campylobacter jejuni Pig_PS762	pPigPS762	This study	This study	45.59	28.99	51	51	-
Campylobacter jejuni Poultry_A1_CF_12	pPoultry_A1_CF_12	This study	This study	47.64	28.05	54	54	-
Campylobacter jejuni Poultry_A6_T2_15	pPoultry_A6_T2_15	This study	This study	47.64	28.03	54	54	-
Campylobacter jejuni Poultry_A8_35_15A	pPoultry_A8_35_15A	This study	This study	47.64	28.05	55	55	-
Campylobacter jejuni Poultry_C1_C_2	pPoultry_C1_C_2	This study	This study	47.64	28.05	54	54	-
Campylobacter jejuni Poultry_C5_T2_8	pPoultry_C5_T2_8	This study	This study	47.64	28.05	55	55	-
Campylobacter jejuni Poultry_MB12	pPoultry_MB12	This study	This study	37.53	27.23	45	45	-
Campylobacter jejuni Poultry_MB18	pPoultry_MB18	This study	This study	44.76	28.99	52	52	-
Campylobacter jejuni Poultry_MB3	pPoultry_MB3	This study	This study	38.15	27.17	43	43	-
Campylobacter jejuni Poultry_MB9	pPoultry_MB9	This study	This study	47.66	28.04	56	56	-
Campylobacter jejuni	pCJ419	NC_004997.1	AY256846	4.01	30.15	4	4	-
Campylobacter jejuni	pVir	NC_005012.1	AF226280	37.47	25.88	52	54	2
Campylobacter jejuni	pTet	NC_006135.1	AY394561	45.21	29.14	49	49	-
Campylobacter jejuni	pTet	NC_007141.1	AY714214	45.21	29.14	44	44	-
Campylobacter jejuni	pCJ1170	NC_008052.1	DQ518173	4.38	30.79	3	3	-
Campylobacter jejuni	pCJ01	NC_008438.1	AF301164	3.21	33.50	4	4	-
Campylobacter jejuni	pTIW94	NC_021493.1	KF192842	3.86	31.11	5	5	-
Campylobacter jejuni	unnamed	NZ_CP013117.1	CP013117	82.73	29.82	103	106	3
Campylobacter jejuni	unnamed	NZ_CP014345.1	CP014345	81.08	25.99	70	80	10
Campylobacter jejuni	pCJDM67 L	NZ_CP014745.1	CP014745	116.88	26.91	97	108	11
Campylobacter jejuni	pCJDM67 S	NZ_CP014746.1	CP014746	36.60	26.07	45	46	1
Campylobacter jejuni	pCJDM202	NZ_CP014743.1	CP014743	119.54	27.24	104	116	12
Campylobacter jejuni subsp. jejuni 01-1512	pCj1	NZ_CP010073.1	CP010073	48.87	29.01	51	54	3
Campylobacter jejuni subsp. Jejuni 01-1512	pCj2	NZ_CP010074.1	CP010074	36.60	25.95	46	47	1
Campylobacter jejuni subsp. Jejuni 00-0949	pTet	NZ_CP010302.1	CP010302	48.87	29.01	52	55	3
Campylobacter jejuni subsp. Jejuni 00-0949	pVir	NZ_CP010303.1	CP010303	36.60	25.95	46	47	1
Campylobacter jejuni subsp. jejuni 00-2544	unnamed	NC_022354.1	CP006710	46.90	29.41	51	52	1
Campylobacter jejuni subsp. jejuni 81-176	pVir	NC_008770.1	CP000550	37.47	25.89	46	47	1
Campylobacter jejuni subsp. jejuni 81-176	pTet	NC_008790.1	CP000549	45.03	29.09	50	50	-
Campylobacter jejuni subsp. jejuni 81-176-UMCW7	pVir	NZ_AZNS01000034.1	AZNS01000034	37.57	25.85	46	47	1
Campylobacter jejuni subsp. jejuni IA3902	pVir	NC_017284.1	CP001877	37.17	25.91	48	48	-
Campylobacter jejuni subsp. jejuni ICDCCJ07001	pTet	-	CP002030.1	44.08	28.69	37	37	-
Campylobacter jejuni subsp. jejuni S3	pTet	NC_017282.1	CP001961	43.22	28.99	42	45	3

Figure 3.20 - Cluster analysis of the *C. jejuni* plasmid pan genome. Heatmap built with the breadth of coverage table of each gene identified in the plasmid pan-genome of *C. jejuni* in each of the strains showing at least the presence of 9 genes (5%) of genes identified in the plasmid pan-genome of *C. jejuni*. Clustering distance was calculated with Jaccard distance matrix. The 178 genes are in the column; the strains are in the rows. Each cell is coloured in light green if the percentage of coverage is smaller than 80% or dark green if it is greater than or equal to 80%. The full heatmap is available in pdf format (Electronic Supplementary Material E6).

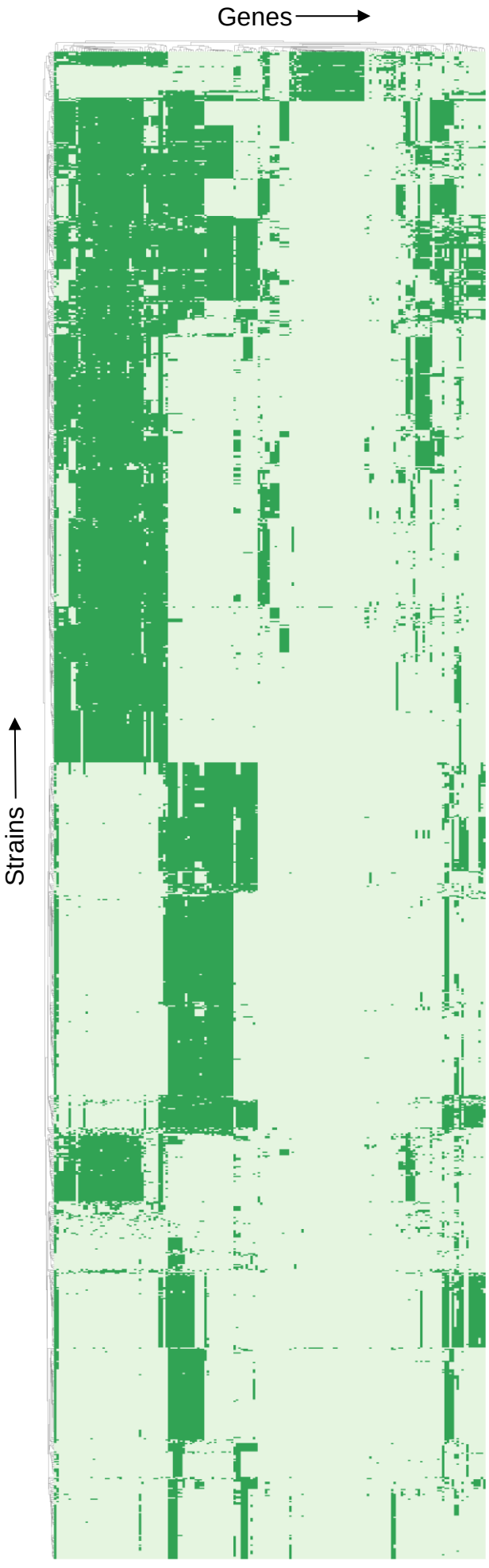
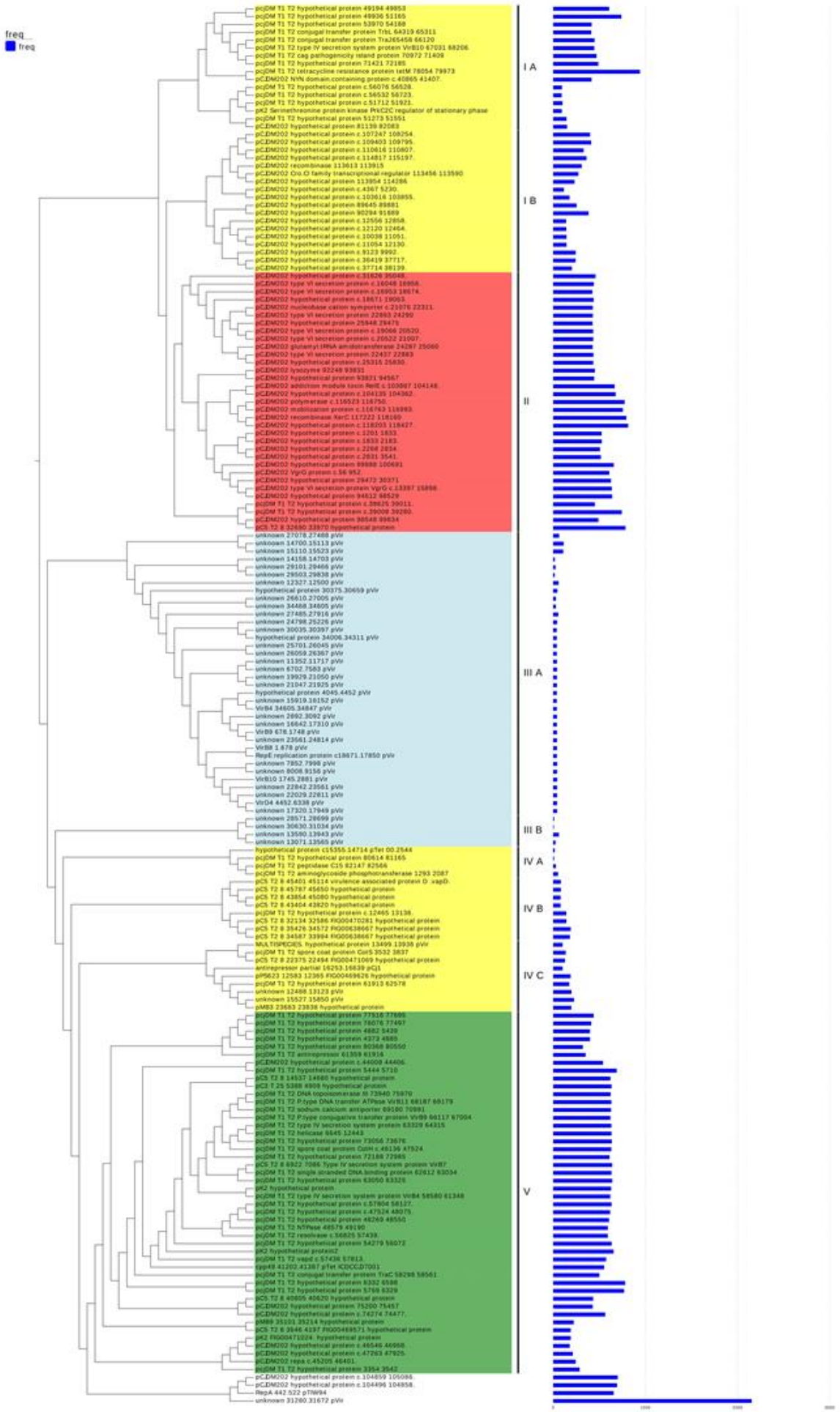

Genes
Strains

Figure 3.21 - Hierarchical clustering tree of the genes identified in the *C. jejuni* plasmid pan genome. The clustering tree was built according to the presence/absence data used for the heatmap in Figure 3.2.3. Clustering distance was calculated using Jaccard distance matrix. The colours (yellow, red, green and light blue) highlight different clusters. The tree annotation (roman numbers I to V) follows the description in the text, the blue bar next to each leaf of the tree represent the frequency of each gene in the subset of strains used for the heatmap in Figure 3.3.1.3.


freq
freq
I A
I B
II
III A
III B
IV A
IV B
IV C
V

The analysis was then narrowed down to just the pTet-like plasmids. In order to identify which plasmids to include in this final analysis, I aligned every plasmid isolated from *C. jejuni* using the software Mauve (http://darlinglab.org/mauve/mauve.html, Darling *et al.,* 2004) and obtained the result in Figure 3.22 (available as electronic supplementary material E7). For clarity, in the image I masked the pVir-like plasmids and the small plasmids pCJ1170-pCJ01-pTIW94- pCJ419, and I did not include the plasmids pTet NC006135.1 and NC007141.1. The plasmids included in the following analysis are the ones indicated by the arrows: they require a clear similarity with pTet, in length and gene content.

Figure 3.22 - Mauve alignment of the plasmids used in the *C. jejuni* plasmid pan genome analysis. The name of each plasmid is indicated in the figure, if the coloured bars above the plasmid name (representing the plasmid sequence) are missing, that plasmid sequence has been masked for clarity. Red arrows indicate the plasmids picked for the pTet-like plasmid pan-genome analysis. The image is available in jpeg format (Electronic Supplementary Material E7).

Table 3.15 - Summary of the characteristics of the plasmids included in the pTet-like plasmid pan-genome analysis. Table columns (from right to left): Strain name, name of the plasmid, Accession number (where available), Size in kilobases, G+C percentage, number of putative proteins, number of predicted genes, number of predicted pseudogenes.

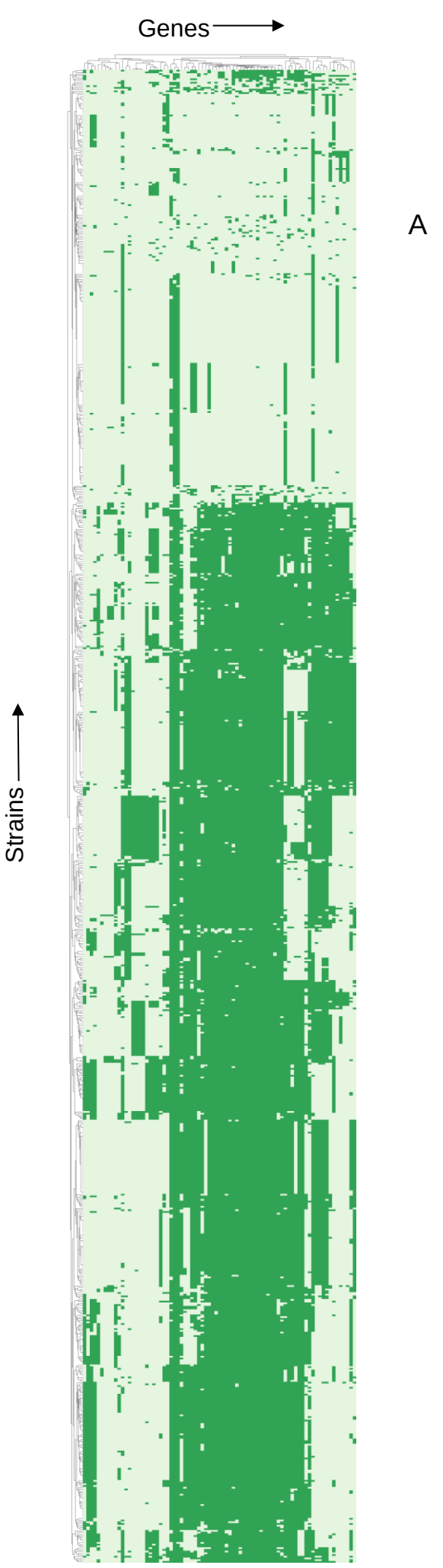
Organism	Plasmid name	RefSeq	INSDC	Size (Kb)	GC(%)	Protein	Gene	Pseudogene
Campylobacter jejuni 11818	p11818	This study	This study	44.95	29.06	52	52	-
Campylobacter jejuni poultry C3_T_25	pPoultry_C3_T_25	This study	This study	47.66	28.04	56	56	-
Campylobacter jejuni Cj2	pCj2	This study	This study	43.67	27.57	48	48	-
Campylobacter jejuni Cj3	pCj3	This study	This study	42.01	27.84	46	46	-
Campylobacter jejuni goose 222	pGoose222	This study	This study	46.08	28.13	52	52	-
Campylobacter jejuni Hi40620300	pHi4062300	This study	This study	44.83	29.27	50	50	-
Campylobacter jejuni K2	pK2	This study	This study	47.59	28.77	54	54	-
Campylobacter jejuni Pet_93_327	pPet93_327	This study	This study	37.73	27.19	43	43	-
Campylobacter jejuni Pig_PS623	pPigPS623	This study	This study	45.38	28.72	50	50	-
Campylobacter jejuni Pig_PS762	pPigPS762	This study	This study	45.59	28.99	51	51	-
Campylobacter jejuni Poultry_A1_CF_12	pPoultry_A1_CF_12	This study	This study	47.64	28.05	54	54	-
Campylobacter jejuni Poultry_A6_T2_15	pPoultry_A6_T2_15	This study	This study	47.64	28.03	54	54	-
Campylobacter jejuni Poultry_A8_35_15A	pPoultry_A8_35_15A	This study	This study	47.64	28.05	55	55	-
Campylobacter jejuni Poultry_C1_C_2	pPoultry_C1_C_2	This study	This study	47.64	28.05	54	54	-
Campylobacter jejuni Poultry_C5_T2_8	pPoultry_C5_T2_8	This study	This study	47.64	28.05	55	55	-
Campylobacter jejuni Poultry_MB12	pPoultry_MB12	This study	This study	37.53	27.23	45	45	-
Campylobacter jejuni Poultry_MB18	pPoultry_MB18	This study	This study	44.76	28.99	52	52	-
Campylobacter jejuni Poultry_MB3	pPoultry_MB3	This study	This study	38.15	27.17	43	43	-
Campylobacter jejuni Poultry_MB9	pPoultry_MB9	This study	This study	47.66	28.04	56	56	-
Campylobacter jejuni	pTet	NC_006135.1	AY394561	45.21	29.14	49	49	-
Campylobacter jejuni	pTet	NC_007141.1	AY714214	45.21	29.14	44	44	-
Campylobacter jejuni subsp. jejuni 01-1512	pCj1	NZ_CP010073.1	CP010073	48.87	29.01	51	54	3
Campylobacter jejuni subsp. jejuni 00-0949	pTet	NZ_CP010302.1	CP010302	48.87	29.01	52	55	3
Campylobacter jejuni subsp. jejuni 00-2544	unnamed	NC_022354.1	CP006710	46.90	29.41	51	52	1
Campylobacter jejuni subsp. jejuni 81-176	pTet	NC_008790.1	CP000549	45.03	29.09	50	50	-
Campylobacter jejuni subsp. jejuni ICDCCJ07001	pTet	-	CP002030.1	44.08	28.69	37	37	-
Campylobacter jejuni subsp. jejuni S3	pTet	NC_017282.1	CP001961	43.22	28.99	42	45	3

This second analysis the 27 pTet-like plasmids reported in Table 3.15 were analysed with the same technique described above and in section 2.10.7. The plasmid pan-genome narrowed down to just the pTet-like plasmids (the pTet-like plasmid pan-genome) was composed of 79 genes. The sequencing reads deriving from the 4005 *C. jejuni* strains analysed here were aligned to the DNA sequence of these 79 genes, as described above in this section.

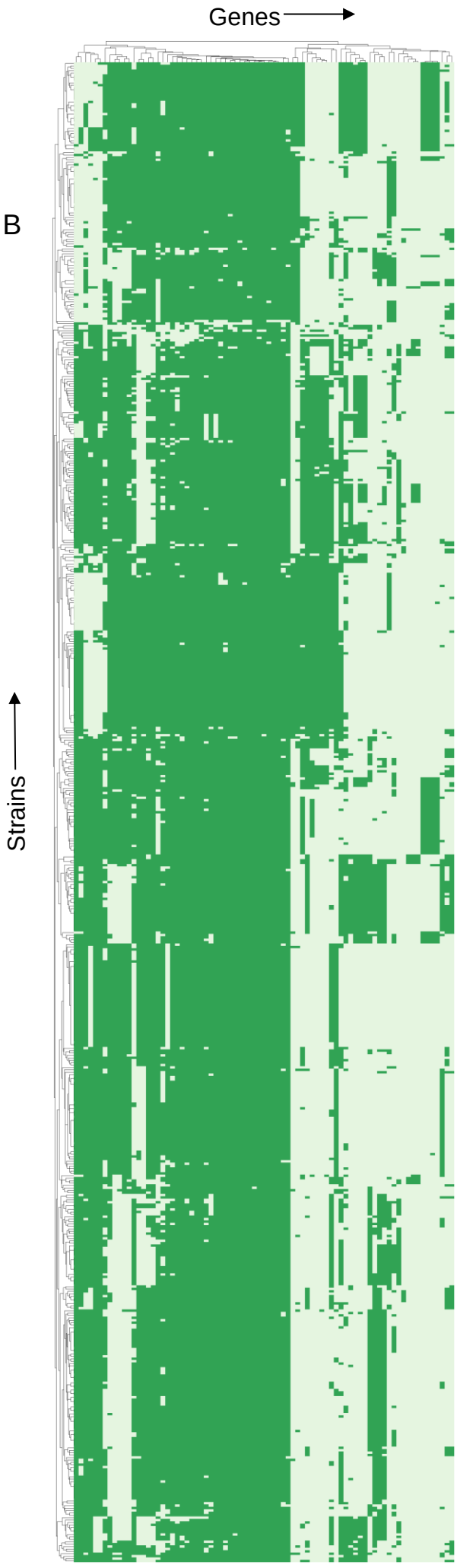
The same clustering performed with the *C. jejuni* plasmid pan-genome was performed with the pTet-like plasmid pan-genome (Figures 3.23 A and B, Available as Electronic Supplementary Material E8 and E9). The heatmap shown in Figure 3.23 A was produced with the frequency of the 79 genes of pTet-like plasmid pan-genome in the strains exhibiting at least the presence of 5% of the genes (n=4). This heatmap highlights the presence of a high number of strains showing just a few genes, for instance, the tetracycline resistance gene. The second heatmap, Figure 3.23 B, was produced with the strains showing at least 26 genes (half of the genes present in the original plasmid pTet). The second heatmap is essentially a close-up of the bottom section of Figure 3.23 A, which highlights the presence of a cluster present in most of the strains and several genes clusters present in just some groups of strains.

The gene names shown in the heatmaps (Electronic Supplementary Material E8 and E9) and in the hierarchical clustering described in section 3.3.2 have been simplified for clarity. The sequence of an orthologous gene with the accession number and the putative function of each gene will be discussed in Chapter 4.

Figure 3.23 - Cluster analysis of the pTet-like plasmid pan-genome. (A) Heatmap built with the breadth of coverage table of each gene identified in pTet-like the plasmid pan-genome in each of the strains showing at least the presence of 4 genes (5%) of genes identified in the plasmid pan-genome of *C. jejuni*. (B) Heatmap built with the breadth of coverage table of each gene identified in pTet-like the plasmid pan-genome in each of the strains showing at least the presence of 26 genes (half of the plasmid pTet NC008790.1) of genes identified in the plasmid pan-genome of *C. jejuni*. Clustering distance was calculated using Jaccard distance matrix. The 178 are on the column; the strains are on the rows. Each cell is coloured in light green if the percent of coverage is smaller than 80% or dark green if it is greater than or equal to 80%. The full heatmaps are available in pdf format (Electronic Supplementary Material E8 and E9).



A

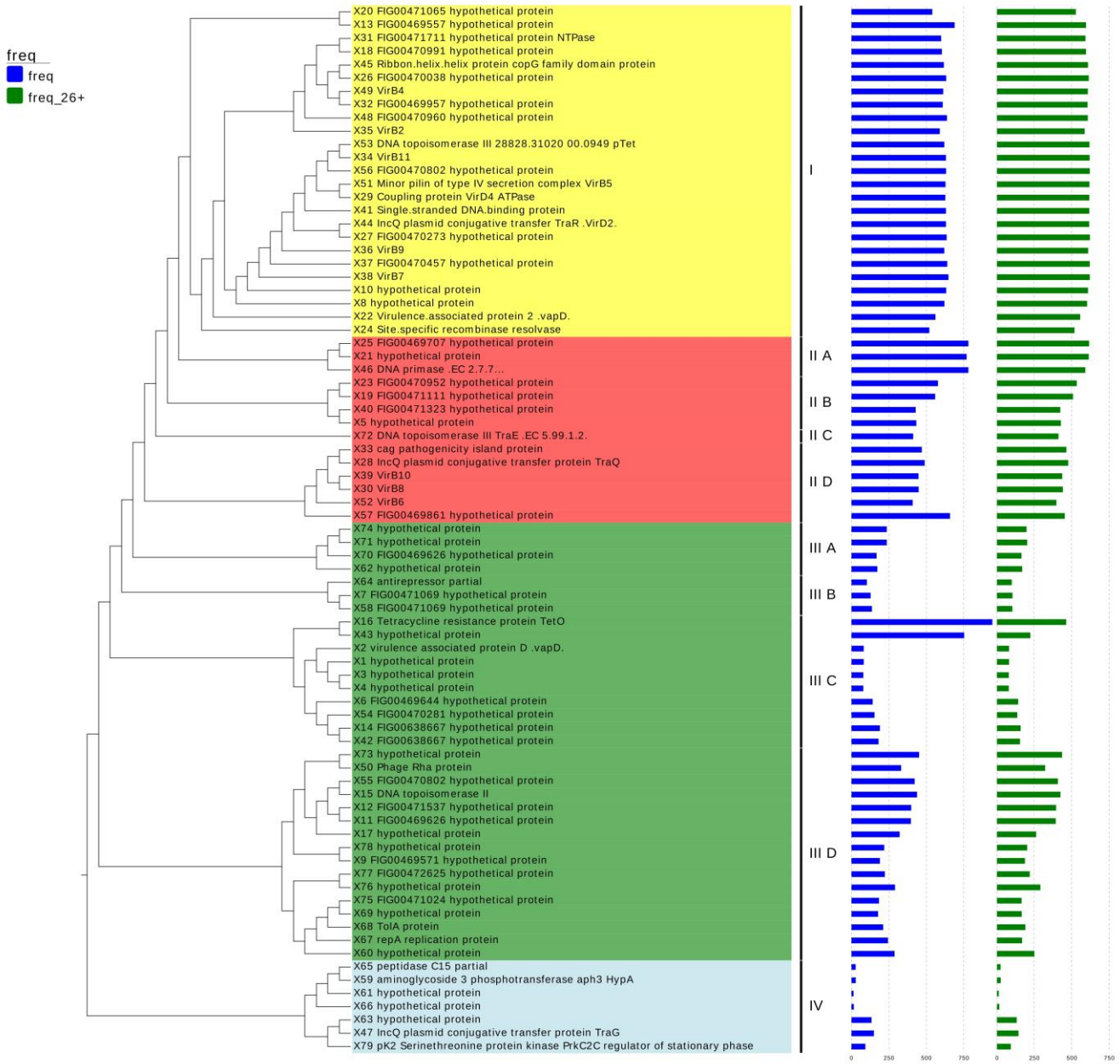


B

3.3.2 Core genes and accessory genes of the plasmid pan-genome associated with pTet-like plasmids

Figure 3.24 shows the clustering dendrogram associated with the column of the heat-map in Figure 3.23 A (at least 5% of pTet-like plasmid pan genome). The hierarchical clustering tree is annotated with two bar-plots: the green plot shows the total frequency of the genes in all 4005 strains, the blue barplot shows the frequency of the genes in the strains showing at least the presence of 26 genes associated with the pTet-like plasmid pan-genome. It is important to notice the CDs encoding for X43 hypothetical protein, Tet(O), and X57 FIG00469861 hypothetical protein showing a remarkable difference between the two barplots. This highlights the presence of these genes on the chromosomes of several strains regardless of the presence of the plasmid pTet or on plasmids different from pTet.

Figure 3.24 - Hierarchical clustering tree of the genes identified in the pTet-like plasmid pan genome. The clustering tree was built according to the presence/absence data used for the heatmap in Figure 3.3.1.8 (A). Clustering distance was calculated using Jaccard distance matrix. The colours (yellow, red, green, and light blue) highlight different clusters. The tree annotation (I to IV) follows the description in the text; the bars next to each leaf of the tree represent the frequency of each gene in the subsets of strains used to produce the heatmaps in Figure 3.3.1.8 A (green bar) and B (blue bar), respectively.



Moreover, the clusters that are identifiable in the heatmaps and in this hierarchical clustering analysis are associated with different frequencies of the genes included in the cluster. In the clustering tree, four main clusters can be identified. The average frequency relative to each cluster is reported in Figure 3.25.

Clusters I and II are present in most of the strains exhibiting at least 26 genes associated with the plasmid pTet. The genes that are easily identifiable from the annotation encode for all the type four secretion system proteins, for proteins associated with conjugations and DNA repair, and for proteins associated with virulence and pathogenicity. Clusters III and IV are associated with a lower frequency, with the exception of the cluster IIIB, which includes the *tet(O)* gene and other genes that have been frequently identified in chromosomes. These clusters include, for instance, genes for antibiotic resistance and genes encoding a virulence associated protein, a Tol protein (associated with translocation of *coli*cins in *E. coli*, Lazzaroni *et al.,* 2002), and other hypothetical proteins.

The core and the accessory pTet-like plasmid pan genome were defined combining the gene frequencies of clusters I and II and the frequencies of clusters III and IV. Genes present in cluster I and II combined form the “core” part of the pTet-like plasmid pan genome; genes present in the clusters III and IV form the “accessory” part pTet-like plasmid pan genome. The average gene frequency of the core and the accessory pTet-like plasmid pan genome is shown in Figure 3.26.

Figure 3.25 - Average gene frequency per cluster identified in Figure 3.24. Error bars indicate standard deviation.

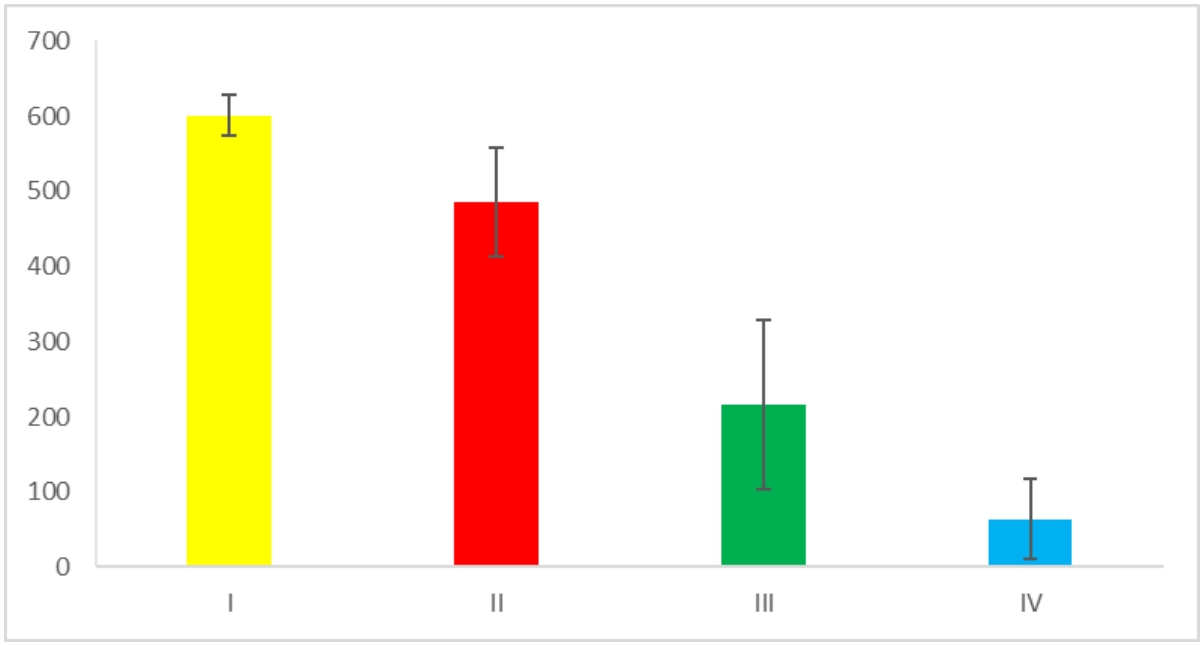
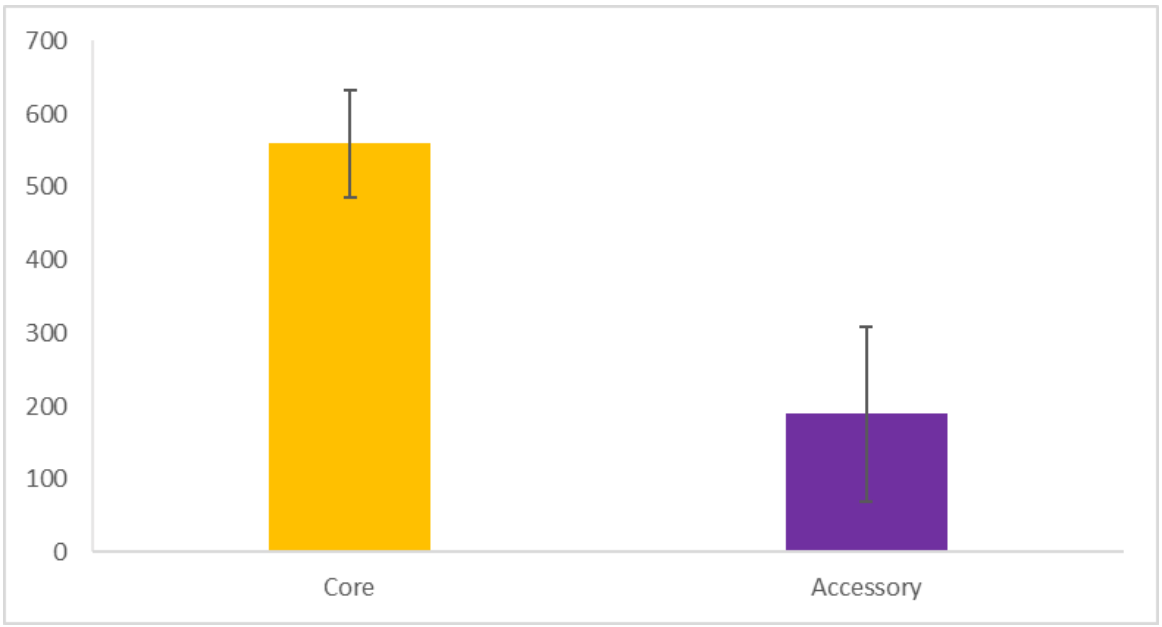


Figure 3.26 Average gene frequency per core and accessory gene set of the pTet-like plasmid pan genome. Error bars indicate standard deviation.



3.4 Discussion

In this chapter, I analysed the distribution of pTet-like plasmids in genome sequenced strains of *C. jejuni*. I assessed the sequence diversity of the plasmid pTet and sequenced 19 strains carrying the plasmid. I described the plasmid pan-genome of *C. jejuni*, focusing on the pTet-like plasmid pan genome and defining a core and an accessory set of genes.

The method used in this chapter to assess the presence, absence, and variability of the genes annotated on the plasmid pTet or identified on the pTet-like plasmid pan-genome involved the alignment of raw sequencing reads to the plasmid or pan-genome reference. This method was chosen because for the majority of the 4005 strains analysed the sequencing reads were available, and to avoid the bias which may be introduced by the genome assembly step. For 679 strains for which the sequencing reads were not available, it was necessary to transform the draft or the completed genome in a set of simulated reads. Although this method may also introduce bias in the analysis, this extra step was necessary to apply the same pipeline to every sample analysed. The ability of correctly identifying gene presence and absence in the set of strains in which raw sequencing reads was not available, was confirmed manually in a subset of those strains (via BLAST alignment of the annotated genes of the plasmid pTet against the draft or the completed genome assembly, data not shown). If this analysis was to be repeaded in the future with a larger subset of genomes for which sequencing reads is not available, those strains could be treated separately.

A further source of bias in this anaylsis may originate from the alignment technique. The default parameters used with Bowtie2 in this study are stringent (corresponding to the “sensitive” paramenters set described in the Bowtie2 manual,

http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml), and were chosen to allow a balance between computational cost and precision of the alignment. With these paramenters, the alignment of a sequencing read needs a score of at least -0.6 + (0.6 * Length of the read) to be valid. Considering the length of the read to be 150, the score threshold for an alignment to be valid is -90. The scoring system for the default “sensitive” parameters set in Bowtie2 is: mismatch penality of -6, a gap penality of -5 for gap opening, and -3 for gap extension. This means that 16 bases mismatch, or 10 bases mismatch and 10 bases gap, are enough to consider the alignment of the read non-valid. Both these examples show how the cut-off for the read alignment is below 20% mismatch. Although this value could be improved in future analyses, having used sequencing reads that have been cleaned as described in section 2.10.3 (quality check and trimming of sequence reads), and having used the scoring system described here, artefacts such as those with very short alignments, which may skew the analysis, were ruled out.

The findings described in this chapter regarding the distribution and the diversity of pTet are in accordance with previous literature. For instance, in the study from Smidt-Ott *et al.,* 2007, 29% of strains analysed are reported to have a plasmid up to 66 kbp long of which 2% was pVir plasmid. Here, I show that a plasmid pTet-like is likely to be present in 15.4 to 33.6 % of the 4005 *C. jejuni* strains analysed (where 15.4 % of strains show half of the genes annotated in pTet and 33.6 % of strains show the presence at least one feature annotated in pTet). It has been reported previously that the tetracycline resistance is not always associated with the presence of pTet and in some strains the gene is located chromosomally (Pratt *et al.,* 2005). It is interesting to note a group of strains that has the majority of the genes for the plasmid pTet but lacks the tetracycline resistance gene. This characteristic has

never been suggested at a sequence level before and reinforces the hypothesis that the function of the plasmid pTet is not limited to a tetracycline resistance phenotype.

A study of the presence of the pTet plasmid in association with the phylogeny of the different strains, the year of isolation, and the ecological niche of isolation did not show any clear association. This plasmid has been shown to be transmitted via conjugation (Avrain *et al.,* 2004; Smitt-Ott *et al.,* 2007), and the broad distribution of the plasmid is in accordance with this observation.

A molecular and phenotypic survey was carried on 120 *C. jejuni* strains. This showed 100% association between the presence of tetracycline resistance gene *tet(O)* and the tetracycline resistance phenotype confirming what recently shown in Zhao *et al.,* 2016. It also showed the presence of 4 different plasmid makeups. The 20 strains positive for at least one molecular marker have been sequenced, in order to obtain the full sequence of 19 completely new pTet-like plasmids.

Three of these plasmids lack the tetracycline resistance gene: this is the first complete sequence of a plasmid showing homology with part of the plasmid pTet but lacking the tetracycline resistance gene. These plasmids have also lost genes flanking the tetracycline resistance gene. The loss of the replication protein gene *repA* may influence the compatibility group of the plasmid and the possible distribution of the plasmid. Also, the loss of a Topoisomerase III encoding gene suggests the involvement of this protein in a recombination event (Roja, 1995). Plasmids pPoultry_MB3, pPet93_372 and pPoultyry_MB12 may have lost the tetracycline resistance or they may have yet to acquire it. What is possible is that these plasmids belong to a lineage that either did not ancestrally acquire the gene, or they belong to a lineage that lost the gene. In either of these cases, the plasmids would be closely related, as they are. If they have yet to acquire the gene for the

tetracycline resistance, it would be reasonable to expect to find the gene cluster that they are missing (*tet(O)*, with genes for a DNA topoisomerase II, replication protein RepA, a protein TolA and a further 3 hypothetical proteins, as reported in table 3.1.3.1.1) in any of the strains analysed that show the chromosomal presence of the *tet(O)* gene. Such a gene cluster is missing from the strains sequenced in this study: in fact, in the three sequenced strains that encode for Tet(O) chromosomally (PoultryMB8, Hi81214, and Hi81006), *tet(O)* is associated with just one single hypothetical protein. Furthermore, when the plasmid pTet genes distribution is analysed in the 4005 sequenced strains database (Figure 1.6.2.1), we can identify a single hypothetical protein clustering together with *tet(O)*, suggesting that the gene cluster missing from the plasmids pMB3, pPet93, and pMB12 is not present in any of the strains exhibiting a chromosomal integration of *tet(O)*. It is reasonable to expect that the plasmids pPoultry_MB3, pPet93_327, and pPoultry_MB12 have lost the gene cluster containing *tet(O)*; however, the possibility that these plasmids belong to a lineage that has not yet acquired the tetracycline resistance cannot be dismissed.

The reason why the plasmids lack the tetracycline resistance can only be conjectured. The reduction in use of tetracycline in the farming industry might have played a role in the loss of selective advantage in carrying the resistance gene. The fact that the plasmid is still conserved in the largest part suggests further selective advantage of the remaining genes.

The plasmids lacking the *tet(O)* gene show the presence of a gene originally annotated in pVir but with no predicted function. This could indicate the exchange of material between the plasmids pTet and pVir (Bacon *et al.,* 2004).

Two of the three strains that possess the *tet(O)* gene chromosomally also have a repeated sequence at each end of the gene cluster that includes *tet(O)*. This

repeated sequence might represent a previously undescribed integration point, or part of it. It has been reported that tetracycline resistance is acquired via HGT in the intestine of chicken (Avrain *et al.,* 2004), and this event may also happen in other ecological niches. I described the existence of the same *tet(O)* chromosomal cluster in the three strains MB8, Hi81006, and Hi81214 and in *Streptococcus spp.*, which include the gene *tet(O)* and the a small CDS upstream. This observation, together with a marginally different G+C content of the *tet(O)* cluster compared to the rest of the plasmid, reinforces the theory that the gene cluster was acquired horizontally (Batchelor *et al.,* 2004; Luangdonkum *et al.,* 2009).

A global comparison of the pTet-like plasmid sequenced in this study highlights the presence of areas that are more conserved than others. The type four secretion system (T4SS) gene cluster appears is present in all the plasmids sequenced and is highly conserved. The T4SS is involved in the conjugal transfer of the genetic material (Wallden *et al.,* 2010). Also highly conserved between the plasmids are 2 hypothetical proteins involved with DNA repair (FIG00470960) and a DNA primase. Other small differences are identifiable within the sequenced plasmids; the most striking one is the presence of a cluster of 5 different genes in seven strains isolated from poultry. This cluster encodes proteins whose functions signal transduction and regulation, virulence, and protein metabolism, which have not previously been described in plasmids of the pTet-like family. The literature focuses on the presence of different antibiotic-resistance genes on the *C. jejuni* plasmids: recently, papers showing the association of *tet(O)* with several different antibiotics resistances have been published, including very recently the sequence of a plasmid encoding for tetracycline and kanamycin resistance, in which the maintenance of the structure of the plasmid pTet was described (Crespo *et al.,* 2016). The study of the plasmid pan-

genome of *C. jejuni* in section 3 also highlighted that the genes *tet(O)* and *aph3* (aminoglycoside resistance) appear at remarkably different frequencies, and they appear in two different clusters when their presence in the sequenced *C. jejuni* genomes is assessed. Several papers reported an association between these 2 resistances (Gibreel *et al.,* 2004; Zhao *et al.,* 2016), but the data reported in this chapter are in contrast with this observation. They are present together in a fraction of sequenced plasmid as reported by Zhao *et al.,* 2016, but this association is not universally present in *C. jejuni.*

Focusing on the plasmid pan-genome of *C. jejuni* showed in Section 3.3.1, strong differences are noticeable in frequencies between the genes present in the different clusters highlighted by the analysis. For instance, the cluster III of the Figure 3.21 shows the frequency of several genes associated with the plasmid pVir. When the frequency of the genes belonging to this cluster is compared to the genes belonging to cluster V, which is composed of genes associated with plasmids pTet-like, it appears clear that the plasmid pVir appears less frequently in *C. jejuni* (as reported by Friis *et al.,* 2006), and this may be an indication of the lack of phenotypic advantage brought by the plasmid pVir (Friis *et al.,* 2007; Schmitd-Ott *et al.,* 2005).

Another gene cluster present in several strains reported in the global plasmid pan-genome is cluster II, which includes the genes encoding for the type six secretion system (T6SS). The T6SS have been reported to be present chromosomally in several *Campylobacter* strains (Harrison *et al.,* 2014; Lertpiriyapong *et al.,* 2012), and it is known to be resent on a plasmid (pCJDM202). It is associated with virulence (Lertpiriyapong *et al.,* 2012), but it lacks association with the plasmid pTet.

When the plasmid pan-genome analysis is narrowed down to the plasmids pTet-like only, it appears that there are different genes clusters associated with different plasmids set-ups. The plasmid pTet-like, have been identified via a global alignment with the software Mauve, and 79 genes are picked up when these plasmids are parsed into a plasmid pan-genome pipeline.

As shown in the Figure 3.3.1.8 and highlighted in the clustering analysis 3.3.2.1, two clusters are present in most of the strains exhibiting a plasmid pTet, and two clusters are present to a much lower extent. As highlighted in the heatmap, these gene clusters can be associated to a CORE and an ACCESSORY plasmid pan genome.

This distribution suggests that the plasmid pTet has a mosaic structure, built on a foundation that is generally conserved, and different genes originated from different bacterial sources (which may be strains of *C. jejuni* or other commensal bacteria) are added on to it. One might speculate, in this regard, that the accessory genome might add a phenotypic advantage in particular ecological niches and that different strains might conserve the genes useful in the particular environment that they are colonising.

Simply, one remark can be addressed to the antibiotics resistance carried by the pTet-like plasmids. According to the gene distribution of *tet(O)* and *aph3*, it appears clear that there are other functions associated with the plasmids. Two scenarios can be imagined if the function of this plasmid family were only linked to the presence of antibiotic resistance: either they were much more frequent or the plasmid would have been lost by *Campylobacter*, leaving the resistance at a chromosomal level. None of these scenarios appear to have occurred: plasmids

pTet-like are very common and hundreds of strains exhibit the presence of the plasmid and no antibiotic resistance, three of which are fully sequenced in this study

In the next chapter, the putative function of each of these hypothetical proteins will be inferred, searching for conserved domains in the putative protein sequence.

Chapter 4 - Function prediction of the genes in the p-Tet like plasmid pan-genome

Approximately 60 percent of genes in the pTet-like plasmid pan genome are predicted to encode for “hypothetical proteins”, accorded to the RAST automatic annotation pipeline. In order to try to understand the putative function of the encoded proteins, I searched for conserved domains using the NCBI Conserved Domains Search tool (CD-Search, Marchler-Bauer *et al.,* 2015).

NCBI’s CD-search tool allows searching in the Conserved Domain Database using a protein or a nucleotide query (which is translated in 6 possible reading frames). The search algorithm used by this tool is RPS-BLAST (Reverse Position-Specific Blast), which uses the query sequence to scan quickly a database of pre-calculated Position-Specific Score Matrices, corresponding to conserved domain models retrieved from several different databases sources: NCBI-curated domains, SMART database (Simple Modular Architecture Research Tool - Letunic *et al.,* 2015), Pfam (Protein families - Finn *et al.,* 2016), COGs database (Clusters of Orthologous Groups of proteins - Tatusov *et al.,* 1997), TIGRFAM (The Institute for Genomic Research's database of protein families) and PRK (NCBI’s collection of Protein Clusters).

The CD-Search result includes hits at various confidence levels, such as specific hits (top ranked RPS-BLAST hits) or non-specific hits (RPS-BLAST hits that exceed the confidence threshold for statistical significance), Multi-domains hits, and the Superfamily to which the domains hits belong.

Each ORF identified in the pTet-like plasmid pan genome was used as input in the CD Search interface. The sequence was searched against the CDDs database v3.15 (Marchler-Bauer *et al.,* 2015), with an e-value threshold of 10e-3, and the output reported shows only domain models identified in the reading frame +1.

This chapter opens with a section summarising the conserved domains identified. The full results are reported in Section 4.3.

4.1 Summary of CD-search results

Table 4.1 was produced in order to summarise the results of the domains identified in the hypothetical proteins of pTet.

In Supplementary material section S4, the details of each conserved domains identified in each gene are described: this includes the results retrieved from NCBI's CD-Search search engine for each hypothetical protein, together with the name, the accession number and the e-value of each conserved domain, and a description of the putative function of each conserved domain identified.

Table 4.1 shows the gene names used in Chapter 3.3.1 for the genes identified in the pTet-like plasmid pan-genome.

Several key functions emerge from the identified conserved domains. VirB/D4 genes have been identified, together with a group of proteins involved in DNA transfer (e.g. TrbM). Horizontal transfer of DNA is one of the main functions described for this plasmid family (Batchelor *et al.,* 2004), and the identification of the expected conserved domains in the plasmid pan-genome seems to confirm this function.

A second function associated with this plasmid family was antibiotic resistance: the sequenced pTet-like plasmids are primarily associated with tetracycline resistance (*tet(O)*) and aminoglycoside resistance (*aph3*) (Smitt-Ott *et al.,* 2007, Zhao *et al.,* 2016). In this analysis, conserved domains for resistance to tetracycline, aminoglycoside, and glycopeptide antibiotics have been identified, together with a

series of peptidase involved in bacteriocins resistance and multidrugs toxin extrusion (MATE-like) proteins.

A group of hypothetical proteins are associated with CRISPR/Cas domain (e.g. 18_FIG00470991_hypothetical_protein, 58_FIG00471069_hypothetical_protein, 76_hypothetical_protein). This function was not identified in any of the pTet-like plasmids before and is associated with bacterial immunity from bacteriophages.

Other key functions that emerge from this analysis include an HicA/B toxin-antitoxin system for plasmid stabilisation (19_FIG00471111_hypothetical_protein, 23_FIG00470952_hypothetical_protein); several proteins that are associated with regulation or signal transduction, such as the ATPase domain in 12_FIG00471537_hypothetical_protein 32_FIG00469957_hypothetical_protein and 45_Ribbon-helix-helix_protein_copG_family_domain_protein, the acetyl transferase domain in 6_FIG00469644_hypothetical_protein, the regulatory domain in 77_FIG00472625_hypothetical_protein; and several hypothetical proteins carrying conserved domains for membrane transporters and receptors, like the ABC transporter domains in 54_FIG00470281_hypothetical_protein or 20_FIG00471065_hypothetical_protein, the carbohydrate transporter in 70_FIG00469626_hypothetical_protein, the *coli*cin uptake domain in 68_TolA_protein, the C4 carbohydrate transporter in 7_FIG00471069_hypothetical_protein.

Table 4.1 - Results summary of the conserved domains search in the hypothetical proteins encoded by the pTet-like plasmid pan genome. Columns are (from left to right): hypothetical protein name (as reported in the pTet-like plasmid pan genome analysis in chapter 1), predicted function according to the conserved domains found, gene length, +/- = conserved domains found or not found, accession numbers for the conserved domains hits (56_FIG00470802_hypothetical_protein shows 12 conserved domains whose accession number is not reported in the table for lack of space).

Name (Chapter 3.3)	Predicted function from associated domains	Gene (bp)	Hits (RF +1)													
1_hypothetical_protein	---	138	-													
2_virulence_associated_protein_D_(vapD)	CRISPR/Cas2	288	+	COG3309	pfam09827											
3_hypothetical_protein	TIR_2 Bacterial toll-like receptor - Lipoprotein [membrane]	1227	+	pfam13676	TIGR04313											
4_hypothetical_protein	Cell-surface hemolisin	417	+	cl11371												
5_hypothetical_protein	---	186	-													
6_FIG00469644_hypothetical_protein	Acetyltransferase	705	+	pfam00583	cl24092											
7_FIG00471069_hypothetical_protein	7TM domains - Membrane carbohidrate transporter	120	+	cl22924	cd09321											
8_hypothetical_protein	---	144	-													
9_FIG00469571_hypothetical_protein	Unknown - Chromosome structure [Eukariotic]	552	+	pfam05483												
10_hypothetical_protein	Guanilate binding protein [Eukariotic]	480	+	cd16269												
11_FIG00469626_hypothetical_protein	MATE-like protein [Multidrug toxin extrusion]	513	+	cd13125	pfam07666	MTH00095	TIGR04370	COG2244	pfam00335							
12_FIG00471537_hypothetical_protein	AAA_23 ATPase - EpsG domain [Biofilm/EPS]	558	+	pfam14897	cd12823	pfam13476	PHA02590	COG4897	PRK03918							
13_FIG00469557_hypothetical_protein	LIM protein interaction domain - Zinc finger domain [DNA interaction]	267	+	cl02475	cl25271											
14_FIG00638667_hypothetical_protein	---	855	-													
15_DNA_topoisomerase_II	Viral endonuclease	180	+	PHA02546												
16_Tetracycline_resistance_protein_TetO	Tetracycline resistance	1920	+	cd04168	cd01684	cd16258	cd03690	cd03711	COG0480							
17_hypothetical_protein	---	183	-													
18_FIG00470991_hypothetical_protein	CRISPR/Cas8 - Viral protein domain - Varous cellular functions	282	+	COG2604	pfam06878	cd09730	PRK14149	PRK03918	COG0419	pfam04111						
19_FIG00471111_hypothetical_protein	HicA toxin	204	+	COG1724	pfam07927											
20_FIG00471065_hypothetical_protein	ABC transporter - Glycopeptide antibiotics resistance	399	+	pfam13346	TIGR00912	cd10336	COG4767	PRK09609	TIGR04370	pfam01757	COG2244	MTH00091				
21_hypothetical_protein	ParA atpase replication protein	561	+	cd02042	COG1192											
22_Virulence-associated_protein_2_(vapD)	CRISPR/Cas2	408	+	COG3309	cl11442											
23_FIG00470952_hypothetical_protein	HicB antitoxin	219	+	COG1598	pfam15919											
24_Site-specific_recombinase_resolvase	Resolvase/Invertase	615	+	cd03768	pfam00239	smart00857	PRK13413	COG1961	TIGR04523							
25_FIG00469707_hypothetical_protein	---	267	-													
26_FIG00470038_hypothetical_protein	Blood coagulase (S. aureus)	324	+	pfam08764												
27_FIG00470273_hypothetical_protein	Alpha helices rich domain	621	+	TIGR04523												
28_IncQ_plasmid_conjugative_transfer_protein_TraQ	Conjugal transfer (TrbM)	765	+	pfam07424	PRK13893											
29_Coupling_protein_VirD4_ATPase	VirD4	1812	+	cd01126	pfam12696	COG3202	pfam02534	COG3505	PRK13897	TIGR02767						
30_VirB8	VirB8	663	+	pfam04335	COG3736	PRK13865										
31_FIG00471711_hypothetical_protein_NTPase	Unknown function (Borrellia bugdoferi) - DNA mobilisation/recombination	612	+	pfam02414	COG1372	PRK13909										
32_FIG00469957_hypothetical_protein	ParB-like domain - Ligand-binding of fucose receptor - AAA15 ATPase	798	+	COG1479	pfam02414	TIGR04527	cd06276	PRK08474	TIGR04523	pfam13175	PRK07133	COG4487	smart00787			
33_cag_pathogenicity_island_protein	Membrane protein (Helicobacter pylori)	438	+	pfam13117												
34_VirB11	VirB11	993	+	TIGR02788	COG0630	PRK13900	cd01130	pfam00437	COG2804	TIGR02525	smart00382	pfam12846	PRK11131			
35_VirB2	TrbC/VirB2	264	+	pfam04956	TIGR00814	cd13148	pfam13347									
36_VirB9	VirB9	888	+	TIGR02781	pfam03524	cd06911	PRK13861									
37_FIG00470457_hypothetical_protein	Translation initiation factor	276	+	pfam04760	PTZ00446											
38_VirB7	---	183	-													
39_VirB10	VirB10	1173	+	pfam03743	COG2948	PRK13855	pfam07423									
40_FIG00471323_hypothetical_protein	Membrane protein	276	+	cd02434	pfam03798	MTH00093	TIGR01218	TIGR01770	MTH00095	COG4984						
41_Single-stranded_DNA-binding_protein	Single strand DNA binding protein [Transcription, replication, recombination]	426	+	cd04496	pfam00436	TIGR00621	PRK08763	PRK08486								
42_FIG00638667_hypothetical_protein	C39G peptidase [Bacteriocins resistance]	594	+	COG3271	cd02423	pfam03412	COG2274	TIGR01193								
43_hypothetical_protein	---	1245	-													
44_IncQ_plasmid_conjugative_transfer_TraR_(VirD2)	VirD2 - Relaxase	1389	+	pfam03432	COG3843											
45_Ribbon-helix-helix_protein_copG_family_domain_protein	CRISPR/Cas9 - ATPase	552	+	PRK13436	pfam16595	COG2604	cd12794	TIGR04313	TIGR04523	pfam13175	PTZ00440	COG1106				
46_DNA_primase_(EC_2.7.7.-)	Topoisomerase/Primase	1227	+	smart00493	pfam13362	cd01029	COG4643	TIGR04523	pfam02463							
47_IncQ_plasmid_conjugative_transfer_protein_TraG	Unkown function	279	+	PRK15396	pfam13978											
48_FIG00470960_hypothetical_protein	AAA_25 ATPase - RepA replication	1794	+	pfam13481	cd01125	TIGR02237	COG2874	COG3421	PRK07133	COG3598	TIGR01612					
49_VirB4	VirB3 - VirB4	2769	+	pfam03135	pfam05101	COG3702	PRK13899	pfam13476	TIGR00929	COG3451	PRK13898	pfam12846				
50_Phage_Rha_protein	Phage regulatory protein	558	+	pfam09669	TIGR02681	cd02774	COG3646									
51_Minor_pilin_of_type_IV_secretion_complex_VirB5	VirB5	987	+	pfam07996	cd14262	TIGR02791	pfam10473	COG1344								
52_VirB6	VirB6	993	+	pfam04610	COG3704	cd06261	TIGR02783	PRK13852	COG0697	PRK05846	pfam06808	COG1757				
53_DNA_topoisomerase_III_28828-31020_00-0949_pTet	DNA Topoisomerase	2193	+	cd00186	smart00437	cd03362	pfam01751	pfam01396	pfam02414	PRK07727	TIGR01056	COG0550	pfam01131	COG0551	PRK07219	
54_FIG00470281_hypothetical_protein	Porin [membrane channel] - ABC transporter	444	+	TIGR04219	COG3637	TIGR03431										
55_FIG00470802_hypothetical_protein	---	1137	+	TIGR04313	cd02774	pfam02414	COG1479	PHA03016	PRK07143	pfam13175	TIGR04499	COG3711	TIGR00912			
56_FIG00470802_hypothetical_protein	DNA protection and repair	5799	+	COG4646	COG4646	pfam13659	smart00490	pfam04851	pfam00271	cd00046	cd00079	pfam02414	PHA03412	COG0553	+ 12 CDs	
57_FIG00469861_hypothetical_protein	Unknown function (Borrellia bugdoferi)	654	+	pfam02414	pfam13175	TIGR04523										
58_FIG00471069_hypothetical_protein	CRISPR/Cas Csx1 - Membrane signal transductor	306	+	cd09728	pfam16224											
59_aminoglycoside_3_phosphotransferase_aph3_HypA	Aminoglycoside 3'-phosphotransferase	918	+	cd05120	COG3173	pfam01636	COG2334									
60_hypothetical_protein	Nichel chaperone - DNA interaction	189	+	pfam01155	COG0375	COG1241	PRK09263									
61_hypothetical_protein	AAA_18 ATPase - Protein kinase	552	+	pfam13238	cd00227	COG1102	pfam13671	PRK12338	TIGR02173	PRK14961						
62_hypothetical_protein	Membrane protein	666	+	pfam07851	TIGR04523	COG3883										
63_hypothetical_protein	Helix-Turn-Helix DNA interaction protein	945	+	cd00093	smart00530	pfam01381	COG1476	cd00093	pfam12844	COG1813						
64_antirepressor_partial	Phage regulatory protein	387	+	pfam09669	TIGR02681	COG3646										
65_peptidase_C15_partial	Peptidase C15	216	+	PRK13197	COG2039	cd00501	pfam01470	TIGR00504								
66_hypothetical_protein	RNA polymerase subunit	642	+	pfam03874	TIGR04523	PTZ00440	pfam07218									
67_repA_replication_protein	---	1140	-													
68_TolA_protein	TolA colicins uptake - DNA mobilisation relaxase	663	+	pfam03432	pfam00841	TIGR02926	TIGR02794	pfam05672	PTZ00121	COG3064						
69_hypothetical_protein	---	423	-													
70_FIG00469626_hypothetical_protein	Carbohydrate transporter	198	+	pfam07457	COG4095	cd08554	TIGR02876									
71_hypothetical_protein	---	483	-													
72_DNA_topoisomerase_III_TraE_(EC_5.99.1.2)	DNA topoisomerase I	186	+	PRK06319												
73_hypothetical_protein	---	120	-													
74_hypothetical_protein	---	156	-													
75_FIG00471024_hypothetical_protein	GTPase [Regulatory - Eukariotic organisms]	414	+	cd05392												
76_hypothetical_protein	CRISPR/Cas6	129	+	COG1583												
77_FIG00472625_hypothetical_protein	LabA regulatory domain family	729	+	cd06167	pfam01936	COG1432										
78_hypothetical_protein	---	114	-													
79_pK2_Serinethreonine_protein_kinase_PrkC2C	Protein kinase	1173	+	cd00180	TIGR03724	PRK14879	COG3642	pfam02414	pfam00069	smart00220	COG0515	PTZ00024	TIGR04523	pfam14093		

4.2 Association between cluster analysis and predicted functions

Table 4.2 shows the genes displayed in Table 4.1 ordered and coloured according to the clustering analysis shown in Section 3.3.2.

The table shows how the functions predicted for the genes appearing in the pTet-like plasmid pan genome can be clustered according to how frequently they are present in a bacterial strain.

In particular, the analysis highlights how the majority of the core genes of the plasmid pan genome (yellow cluster and red cluster) are associated with DNA transfer, with a few exceptions, such as the ABC Transporter 20_FIG00471065_hypothetical_protein or the 33_cag_pathogenicity_island_protein. These genes encode for the *virB/D2* cluster, topoisomerases, and conjugative proteins. Moreover, the *hicAB* system for plasmid stabilisation is present in the yellow cluster, which is associated with a higher frequency in the *C. jejuni* database analysed. The proteins of the plasmid pan genome associated with lower frequencies (green and light blue clusters) seem to be associated with functions presumably associated with adaptation, like membrane transporters and nutrients uptake, antibiotic resistances, regulation, and signal transduction.

Table 4.2 - Results summary of the conserved domains search in the hypothetical proteins encoded by the pTet-like plasmid pan genome. Columns are analogous to Table 4.1. The second column is coloured according to the different clusters identified in the pTet-like plasmid pan genome analysis (Chapter 3, Section 3.3.2).

Name (Chapter 3.3)	Predicted function from associated domains	Gene (bp)		Hits (RF +1)												
20_FIG00471065_hypothetical_protein	ABC transporter - Glycopeptide antibiotics resistance	399	+	pfam13346	TIGR00912	cd10336	COG4767	PRK09609	TIGR04370	pfam01757	COG2244	MTH00091				
13_FIG00469557_hypothetical_protein	LIM protein interaction domain - Zinc finger domain [DNA interaction]	267	+	cl02475	cl25271											
31_FIG00471711_hypothetical_protein_NTPase	Unknown function (Borrellia bugdoferi) - DNA mobilisation/recombination	612	+	pfam02414	COG1372	PRK13909										
18_FIG00470991_hypothetical_protein	CRISPR/Cas8 - Viral protein domain - Varous cellular functions	282	+	COG2604	pfam06878	cd09730	PRK14149	PRK03918	COG0419	pfam04111						
45_Ribbon-helix-helix_protein_copG_family_domain_protein	CRISPR/Cas9 - ATPase	552	+	PRK13436	pfam16595	COG2604	cd12794	TIGR04313	TIGR04523	pfam13175	PTZ00440	COG1106				
26_FIG00470038_hypothetical_protein	Blood coagulase (S. aureus)	324	+	pfam08764												
49_VirB4	VirB3 - VirB4	2769	+	pfam03135	pfam05101	COG3702	PRK13899	pfam13476	TIGR00929	COG3451	PRK13898	pfam12846				
32_FIG00469957_hypothetical_protein	ParB-like domain - Ligand-binding of fucose receptor - AAA15 ATPase	798	+	COG1479	pfam02414	TIGR04527	cd06276	PRK08474	TIGR04523	pfam13175	PRK07133	COG4487	smart00787			
48_FIG00470960_hypothetical_protein	AAA_25 ATPase - RepA replication	1794	+	pfam13481	cd01125	TIGR02237	COG2874	COG3421	PRK07133	COG3598	TIGR01612					
35_VirB2	TrbC/VirB2	264	+	pfam04956	TIGR00814	cd13148	pfam13347									
53_DNA_topoisomerase_III_28828-31020_00-0949_pTet	DNA Topoisomerase	2193	+	cd00186	smart00437	cd03362	pfam01751	pfam01396	pfam02414	PRK07727	TIGR01056	COG0550	pfam01131	COG0551	PRK07219	
34_VirB11	VirB11	993	+	TIGR02788	COG0630	PRK13900	cd01130	pfam00437	COG2804	TIGR02525	smart00382	pfam12846	PRK11131			
56_FIG00470802_hypothetical_protein	DNA protection and repair	5799	+	COG4646	COG4646	pfam13659	smart00490	pfam04851	pfam00271	cd00046	cd00079	pfam02414	PHA03412	COG0553	+ 12 CDs	
51_Minor_pilin_of_type_IV_secretion_complex_VirB5	VirB5	987	+	pfam07996	cd14262	TIGR02791	pfam10473	COG1344								
29_Coupling_protein_VirD4_ATPase	VirD4	1812	+	cd01126	pfam12696	COG3202	pfam02534	COG3505	PRK13897	TIGR02767						
41_Single-stranded_DNA-binding_protein	Single strand DNA binding protein [Transcription, replication, recombination]	426	+	cd04496	pfam00436	TIGR00621	PRK08763	PRK08486								
44_IncQ_plasmid_conjugative_transfer_TraR_(VirD2)	VirD2 - Relaxase	1389	+	pfam03432	COG3843											
27_FIG00470273_hypothetical_protein	Alpha helices rich domain	621	+	TIGR04523												
36_VirB9	VirB9	888	+	TIGR02781	pfam03524	cd06911	PRK13861									
37_FIG00470457_hypothetical_protein	Translation initiation factor	276	+	pfam04760	PTZ00446											
38_VirB7	---	183	-													
10_hypothetical_protein	Guanilate binding protein [Eukariotic]	480	+	cd16269												
8_hypothetical_protein	---	144	-													
22_Virulence-associated_protein_2_(vapD)	CRISPR/Cas2	408	+	COG3309	cl11442											
24_Site-specific_recombinase_resolvase	Resolvase/Invertase	615	+	cd03768	pfam00239	smart00857	PRK13413	COG1961	TIGR04523							
25_FIG00469707_hypothetical_protein	---	267	-													
21_hypothetical_protein	ParA atpase replication protein	561	+	cd02042	COG1192											
46_DNA_primase_(EC_2.7.7.-)	Topoisomerase/Primase	1227	+	smart00493	pfam13362	cd01029	COG4643	TIGR04523	pfam02463							
23_FIG00470952_hypothetical_protein	HicB antitoxin	219	+	COG1598	pfam15919											
19_FIG00471111_hypothetical_protein	HicA toxin	204	+	COG1724	pfam07927											
40_FIG00471323_hypothetical_protein	Membrane protein	276	+	cd02434	pfam03798	MTH00093	TIGR01218	TIGR01770	MTH00095	COG4984						
5_hypothetical_protein	---	186	-													
72_DNA_topoisomerase_III_TraE_(EC_5.99.1.2)	DNA topoisomerase I	186	+	PRK06319												
33_cag_pathogenicity_island_protein	Membrane protein (Helicobacter pylori)	438	+	pfam13117												
28_IncQ_plasmid_conjugative_transfer_protein_TraQ	Conjugal transfer (TrbM)	765	+	pfam07424	PRK13893											
39_VirB10	VirB10	1173	+	pfam03743	COG2948	PRK13855	pfam07423									
30_VirB8	VirB8	663	+	pfam04335	COG3736	PRK13865										
52_VirB6	VirB6	993	+	pfam04610	COG3704	cd06261	TIGR02783	PRK13852	COG0697	PRK05846	pfam06808	COG1757				
57_FIG00469861_hypothetical_protein	Unknown function (Borrellia bugdoferi)	654	+	pfam02414	pfam13175	TIGR04523										
74_hypothetical_protein	---	156	-													
71_hypothetical_protein	---	483	-													
70_FIG00469626_hypothetical_protein	Carbohydrate transporter	198	+	pfam07457	COG4095	cd08554	TIGR02876									
62_hypothetical_protein	Membrane protein	666	+	pfam07851	TIGR04523	COG3883										
64_antirepressor_partial	Phage regulatory protein	387	+	pfam09669	TIGR02681	COG3646										
7_FIG00471069_hypothetical_protein	7TM domains - Membrane carbohidrate transporter	120	+	cl22924	cd09321											
58_FIG00471069_hypothetical_protein	CRISPR/Cas Csx1 - Membrane signal transductor	306	+	cd09728	pfam16224											
16_Tetracycline_resistance_protein_TetO	Tetracycline resistance	1920	+	cd04168	cd01684	cd16258	cd03690	cd03711	COG0480							
43_hypothetical_protein	---	1245	-													
2_virulence_associated_protein_D_(vapD)	CRISPR/Cas2	288	+	COG3309	pfam09827											
1_hypothetical_protein	---	138	-													
3_hypothetical_protein	TIR_2 Bacterial toll-like receptor - Lipoprotein [membrane]	1227	+	pfam13676	TIGR04313											
4_hypothetical_protein	Cell-surface hemolisin	417	+	cl11371												
6_FIG00469644_hypothetical_protein	Acetyltransferase	705	+	pfam00583	cl24092											
54_FIG00470281_hypothetical_protein	Porin [membrane channel] - ABC transporter	444	+	TIGR04219	COG3637	TIGR03431										
14_FIG00638667_hypothetical_protein	---	855	-													
42_FIG00638667_hypothetical_protein	C39G peptidase [Bacteriocins resistance]	594	+	COG3271	cd02423	pfam03412	COG2274	TIGR01193								
73_hypothetical_protein	---	120	-													
50_Phage_Rha_protein	Phage regulatory protein	558	+	pfam09669	TIGR02681	cd02774	COG3646									
55_FIG00470802_hypothetical_protein	---	1137	+	TIGR04313	cd02774	pfam02414	COG1479	PHA03016	PRK07143	pfam13175	TIGR04499	COG3711	TIGR00912			
15_DNA_topoisomerase_II	Viral endonuclease	180	+	PHA02546												
12_FIG00471537_hypothetical_protein	AAA_23 ATPase - EpsG domain [Biofilm/EPS]	558	+	pfam14897	cd12823	pfam13476	PHA02590	COG4897	PRK03918							
11_FIG00469626_hypothetical_protein	MATE-like protein [Multidrug toxin extrusion]	513	+	cd13125	pfam07666	MTH00095	TIGR04370	COG2244	pfam00335							
17_hypothetical_protein	---	183	-													
78_hypothetical_protein	---	114	-													
9_FIG00469571_hypothetical_protein	Unknown - Chromosome structure [Eukariotic]	552	+	pfam05483												
77_FIG00472625_hypothetical_protein	LabA regulatory domain family	729	+	cd06167	pfam01936	COG1432										
76_hypothetical_protein	CRISPR/Cas6	129	+	COG1583												
75_FIG00471024_hypothetical_protein	GTPase [Regulatory - Eukariotic organisms]	414	+	cd05392												
69_hypothetical_protein	---	423	-													
68_TolA_protein	TolA colicins uptake - DNA mobilisation relaxase	663	+	pfam03432	pfam00841	TIGR02926	TIGR02794	pfam05672	PTZ00121	COG3064						
67_repA_replication_protein	---	1140	-													
60_hypothetical_protein	Nichel chaperone - DNA interaction	189	+	pfam01155	COG0375	COG1241	PRK09263									
79_pK2_Serinethreonine_protein_kinase_PrkC2C	Protein kinase	1173	+	cd00180	TIGR03724	PRK14879	COG3642	pfam02414	pfam00069	smart00220	COG0515	PTZ00024	TIGR04523	pfam14093		
47_IncQ_plasmid_conjugative_transfer_protein_TraG	Unkown function	279	+	PRK15396	pfam13978											
63_hypothetical_protein	Helix-Turn-Helix DNA interaction protein	945	+	cd00093	smart00530	pfam01381	COG1476	cd00093	pfam12844	COG1813						
66_hypothetical_protein	RNA polymerase subunit	642	+	pfam03874	TIGR04523	PTZ00440	pfam07218									
61_hypothetical_protein	AAA_18 ATPase - Protein kinase	552	+	pfam13238	cd00227	COG1102	pfam13671	PRK12338	TIGR02173	PRK14961						
59_aminoglycoside_3_phosphotransferase_aph3_HypA	Aminoglycoside 3'-phosphotransferase	918	+	cd05120	COG3173	pfam01636	COG2334									
65_peptidase_C15_partial	Peptidase C15	216	+	PRK13197	COG2039	cd00501	pfam01470	TIGR00504								

4.3 Discussion

In this chapter, a conserved domain search was used to identify the putative functions in the genes identified in the pTet-like plasmid pan genome.

For the majority of the proteins annotated as “hypothetical” by the automated pipeline RAST (Overbeek *et al.,* 2014), several conserved domains have been identified with a high degree of confidence (e-value lower than 10e-3).

The genes predicted to be involved in the conjugational DNA transfer and with the Type Four Secretion System, have all been confirmed to be present by the presence of the expected conserved protein domains in the translated sequence.

DNA replication-associated proteins (RepA, ParA/B) have also been identified, even if the CDS identified as “67_repA_replication_protein” does not show the presence of any conserved domains associated with the function. The sequence of the gene “67_repA_replication_protein” is homologous to the *repA* gene of the plasmid pTet, but no conserved domain is identified in the translated sequence. This might be a mis-annotation in the original pTet NC008790.1 plasmid (locus tag CJJ81176_pTet0001) annotated for the first time in 2004 (Batchelor *et al*, 2004).

Several CRISPR/Cas system associated hypothetical proteins have been identified in this study: in particular, two CRISPR/Cas2 domains associated with the virulence-associated proteins VapD and VapD2 and domains associated with Cas9, Cas8, Cas6, Csx1 in other hypothetical proteins.

CRISPR/Cas systems have been described extensively in literature recently and they are associated with bacterial resistance to phages (Boyaval *et al.,* 2007; reviewed by Marrafini in 2015). They are described as bacterial adaptive immunity: CRISPR loci are formed by arrays of short repeated sequences with intervening short sequences of viral or plasmid origins. The small RNAs produced by the

transcription of the spacers target the complementary viral sequence, and they are used as a guide by the Cas endonucleases as guide for the direct cleavage of the viral genome. The function of CRISPR/Cas Systems have also been associated to virulence in pathogenic bacteria (reviewed by Louwen *et al.*, 2014): in particular, it was observed that the deletion of *cas9* in *C. jejuni* affected the ability of the bacterium to translocate across intestinal epithelial cells (Louwen *et al.,* 2013).

A HicA/HicB system is part of the pTet-like plasmid pan genome. This system is a toxin/antitoxin system used for plasmid stabilisation in bacterial cells by Post Segregational Killing (PSK). The function of this system is reviewed by Van Melderen and De Bast in 2009. The system is organised into an operon: the toxin is expressed constitutively and is more stable than the antitoxin, which degrades the mRNA encoding for the toxin. If the bacterial cell does not inherit the plasmid after duplication, it loses the antitoxin and is killed the stable toxin protein still present in it. The HicAB system has been shown to be involved with other functions in other bacterial species, such as chemotaxis and biofilm formation in *E. coli*, to promote virulence in *Salmonella* and to have a role in persister formation in *Burkholderia pseudomallei* (De la Cruz *et al.,* 2013, Butt *et al.,* 2014, Kim *et al.,* 2009). The TA-system described here in particular was found to share between 70% and 80% protein sequence identity with TA-systems described in the chromosome of several strains of *H. pylori.* Recently the expression of one TA-system of *H. pylori* (*hp0968–hp0967*, a TA-system of the same family of the HicAB system described here in *C. jejuni*) was described to be regulated by environmental signals relevant in the context of human infection and to be induced in biofilm formation (Cárdenas-Mondragón *et al.,* 2016).

The presence of this TA-system may explain why it was not possible to cure a pTet-like plasmid when it is found without other plasmids in a *C. jejuni* strain (Smitd-Ott *et al.,* 2004), but its involvement in environmental adaptation or virulence cannot be excluded.

Several conserved domains associated with antibiotic resistance and membrane transport were identified in the pTet-like plasmid pan-genome. These include an ABC transporter in 54_FIG00470281_hypothetical_protein, an ABC transporter/glycopeptide antibiotic resistance in 20_FIG00471065_hypothetical_protein, a MATE-like protein involved in multidrug and toxin extrusion in 11_FIG00469626_hypothetical_protein, the tetracycline resistance gene *tet(O)*, a resistance to aminoglycoside antibiotics (*aph3)* a C39G peptidase involved in resistance to bacteriocins in 42_FIG00638667_hypothetical_protein. ABC (ATP-binding cassette) transporters rely on the hydrolysis of an ATP molecule for the uptake or the extrusion of molecules (reviewed by Wilkens in 2015) and are used by bacteria for uptake of nutrients or extrusion of toxic molecules. In *C. jejuni*, a class of amino acid ABC-transporters have been recently linked to abiotic stress resistance (peroxide, heat, and osmotic shock) and virulence in mouse and macrophage cell infection models (Lin *et al.,* 2009). As far as the antibiotic resistance genes are concerned, *Campylobacter* species are known to exhibit resistance to a wide variety of antibiotics (Luangtongkum *et al*, 2010), and it is not surprising to find several antibiotic-resistance genes and MATE-like transporters in this analysis.

A number of proteins carry conserved domains identified in other bacterial species, and some of them seem involved with pathogenicity. Examples are the conserved domains highlighted in the hypothetical proteins

26_FIG00470038_hypothetical_protein (a factor promoting the initiation of blood coagulation in *S. aureus*, Friedrich *et al., 2003*), 33_cag_pathogenicity_island_protein (a membrane protein expressed by the Cag12 pathogenicity island of *H. pylori*, Kim et al., 2006), and 4_hypothetical_protein (a cell-surface protein promoting the lysis of blood cells and associated with pathogenicity of the insect pathogen *Xenorhabdus nematophila*, Cowles et al., 2005). The presence of genes derived from other bacterial species has been already observed in this plasmid family, as described in the previous chapter for the gene *tet(O)* and in the mosaic structure of a multiple-drug resistance plasmid (Nirdnoy *et al*, 2005). Several studies highlighted the genome plasticity of *C. jejuni*: Gibreel *et al*. in 2004 were able to transfer a plasmid from *C. jejuni* to *E. coli*, and *C. jejuni* is naturally competent and can efficiently uptake DNA from the environment (De Boer 2002, Wang 1990, Vegge 2012). These mechanisms can be used by the bacterium in the process of adaptation.

The regulative function often appears in the conserved domains identified in the hypothetical proteins encoded by the pTet-like plasmid pan-genome. A large number of conserved domains are associated with ATPases, Protein Kinases, Acetyltransferases, and Membrane receptors for signal transduction. It is not clear in which regulative process these proteins are involved, but their presence highlights the possibility of the association of the plasmid family pTet-like in a large number of regulative functions.

A number of membrane transporter-associated conserved domains are also identified. For instance, a carbohydrate transporter domain is present in 70_FIG00469626_hypothetical_protein and 7_FIG00471069_hypothetical_protein; ABC transporter domains are present in the hypothetical proteins

20_FIG00471065_hypothetical_protein and 54_FIG00470281_hypothetical_protein. The C4-carbohydrate transporter domain identified in the hypothetical protein 7_FIG00471069_hypothetical_protein may be of particular interest in the *Campylobacter* species: it is reported that most *Campylobacter* strains cannot utilise sugars and several works have shown how they rely on the use of amino acids and C4 carbohydrates for survival and successful colonisation (Guccione *et al.,* 2008; Stahl *et al.*, 2011).

Because of the presence of these transporter proteins, one of the functions of the plasmid could be the uptake of molecules that are present just in particular environments, giving the bacterium a higher fitness in different ecological niches.

Finally, the comparison of the cluster analysis performed in the previous chapter (3.3.1) with the putative functions of the hypothetical proteins highlights different hypothetical functions associated with different gene frequencies in the set of *C. jejuni* strains analysed.

In *Campylobacter*, several studies showed genes for particular functions associated with different ecological niches, such as a cluster of six genes within the O-linked flagellin glycosylation locus associated with strains isolated from livestock and a series of strains isolated from a sandy beach environment lacking a putative serine protease (Champion *et al.,* 2005), or a gene cluster implicated in the production of B5 vitamin in strains isolated from cattle (Sheppard *et al.,* 2013). In these cases, the association was found with chromosomal genes.

The distribution of gene functions associated with frequencies at which those genes are found in a *Campylobacter* strain explains the composition of the plasmid pan-genome of the pTet-like plasmids. The “core” genes, found at a higher frequency, are associated with plasmid maintenance and transfer; the “accessory”,

found at a lower frequency, genes are associated with metabolism, regulation and resistance: it is possible that they give a higher fitness to the organism in determinate conditions and are added to the plasmid or maintained just in particular niches.

In the next chapters, I explain the functions of some of the genes identified in the pTet-like plasmids via a mutagenesis analysis and phenotype evaluation.

Chapter 5 - Mutagenesis of pTet-like plasmid pan-genome accessory and core genes

After assessing a possible function for the 79 genes in the pTet plasmid pan-genome, I next carried out a functional study of a set of these genes via targeted mutagenesis.

In the last 30 years, several protocols for targeted mutagenesis in this *C. jejuni* have been described and successfully applied. These protocols are based on the introduction of a plasmid vector carrying an antibiotic resistance marker in *C. jejuni* via electroporation or conjugation. These plasmid vectors were constructed to be unable to replicate in *Campylobacter* (suicide-vector) and to include an antimicrobial resitance cassette, surrounded by two regions homologous to *Campylobacter* genome—specifically upstream and downstream of the gene targeted for deletion. Once the plasmid is transformed in *Camptlobacter*, and under selective pressure, the homologous regions would facilitate the recombination event that will result in the stable integration of the antimicrobial resistance cassette in the genome of the bacterium, substituting the gene targeted for deletion with the resistance cassette. The use of these protocols was successfully described in several recent publications (Reuter *et al.,* 2013; Vegge *et al.,* 2009).

The main aims of this chapter are to develop a reliable method for the application of one these mutagenesis protocols and to mutate a subset of genes from the pTet-like plasmid pan-genome.

Moreover, I will describe the genomic set up of the only *C. jejuni* strain in which I was able to stably introduce foreign DNA.

5.1 Assess the possibility to introduce mutations in *C. jejuni* strains

I performed two studies in order to assess the possibility of producing mutants in different *C. jejuni* strains.

First, the copy number of the pTet-like plasmid in different strains was measured. This is important: if the plasmid is present at a copy number higher than 1 per cell, only one copy would be mutated. In this analysis, I evaluated the copy numbers of pCj2 and pCj3 and some of the plasmid genes present in the strain Cj1. Sequencing reads of Cj1, Cj2, and Cj3 were aligned back to the corresponding draft genome assemblies using Bowtie2 software. The “Mpileup” module of SamTools was then used to calculate the coverage per site of the alignment (Li *et al.,* 2009). An ad hoc perl script extracted the average coverage value for each contig, and the coverage of plasmid genes (*tetO*, *virD4*, *traQ* and *traR*) was compared to the coverage of chromosomal genes (extract_pileup_coverage_gene.pl). The results are reported in Figure 5.1 and Table 5.1. The average coverage value of the plasmid genes—compared to the rest of the genome—suggests a copy number of 1.

Once the low copy number of the plasmid pTet-like genes was established in the strains Cj1, Cj2, and Cj3, assuming the low copy number is conserved in the other strains analysed, the possibility of transforming *C. jejuni* strains with exogenous DNA was assessed. In order to do so, 19 strains of *C. jejuni* were transformed with pGEM::cj1411 via electroporation. Plasmid pGEM::cj1411 was designed in the University of Exeter’s Molecular Pathogenesis Laboratory (Olivia Champion, personal communication), to integrate fully into gene *cj1411* of *C. jejuni* strain 11168 (AL111168.1). This gene is broadly conserved in *C. jejuni* and encodes a cytochrome P450 protein (Alvarez *et al.,* 2013). Transformant strains could be identified thanks to the kanamycin cassette marker, in case of a successful

transformation and chromosomic integration of the resistance marker. The strains used in this analysis are reported in Table 5.2, together with the percentage of similarity of the gene cj1411, which was always higher than 98%. Only strain Cj1 gave positive results, showing 10^2 transformant cells/ml able to grow on kanamycin when transformed with 1μg of vector DNA (Materials and methods, Section 2.6). I decided to use the strain Cj1 to produce mutants of plasmid genes.

Figure 5.1 - Depths of genomic sequencing coverage for strains Cj1, Cj2, and Cj3. Genomic shotgun sequence reads were aligned against the respective genome assembly using BWA-mem and the depth of coverage was calculated for each site. Blue bars are the average depths across the entire genome sequence; red bars are the average depths for four plasmid genes (*tetO*, *virD4*, *traQ* and *traR*). The differences between whole genome and plasmid genes are not significant in any of the 3 strains (Single tailed, t-test, $p<0.01$)

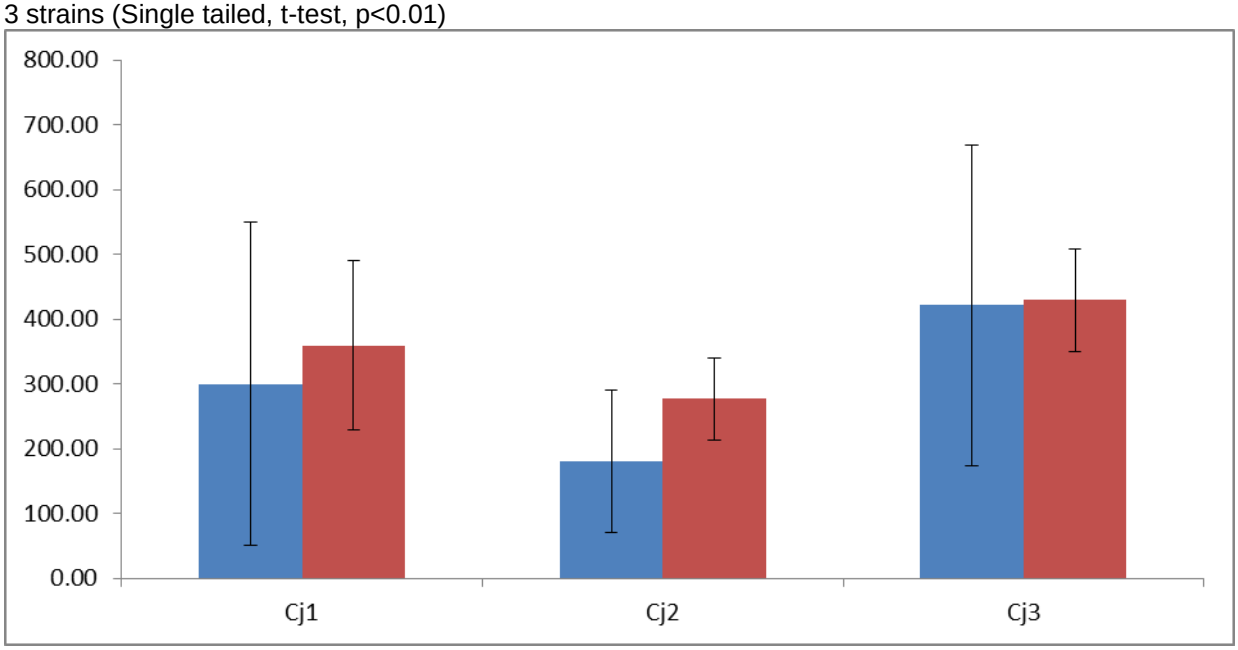


Table 5.1 - Data associated to Figure 5.1 Average coverage per site of the alignment of Cj1, Cj2, and Cj3 sequencing reads versus their own assemblies.

	Genome average	SD	Plasmid genes	SD
Cj1	299.70	249.28	359.39	131.47
Cj2	180.57	110.08	276.99	63.2
Cj3	421.15	246.87	429.44	79.03

Table 5.2 - Strains used to assess for transformability with exogenous DNA. Columns show (from left to right): strain name (accession number in bracket, where available), nucleotide sequence identities versus the *cj1411* gene, % of identity with the *cj1411* gene, +/- = positive or negative result of the transformation.

Strain	Identities	% identity	Transformation result
Cj1	1361/1362	99%	+
Cj2	1355/1362	99%	-
Cj3	1351/1362	99%	-
11818	1354/1362	99%	-
Hi40620300	1362/1362	100%	-
Pet_93_327	1331/1362	97%	-
Pig_PS623	1336/1362	98%	-
PoultryC5_T2_8	1361/1362	99%	-
Poultry_MB12	1333/1362	97%	-
Poultry_MB9	1361/1362	99%	-
81-176 (NZ_AASL01000001.1)	1362/1362	100%	-
Cj5	1349/1362	99%	-
Cj32799	1355/1362	99%	-
cow201	1355/1362	99%	-
Pig_PS549	1336/1362	98%	-
Poultry_D2	1355/1362	99%	-
Sheep_S120	1338/1362	98%	-
Sheep_S216	1362/1362	100%	-
11168 (AL111168.1)	1362/1362	100%	-

5.2 Description of the strain Cj1 genome

C. jejuni Cj1 is a strain originally isolated in Thailand from a hospitalised patient with bloody diarrhoea in 2009. That strain was positive when tested using the PCRs for plasmid marker genes and the tetracycline resistance gene reported in Chapter 3. However, I was unable to isolate a plasmid from strain Cj1 (see Figure 3.5). This strain exhibits the majority of the genes annotated in the plasmid pTet NC00079.1 (Table 3.5), as well as the majority of the genes (48) annotated in the pTet-like plasmid-pan genome. When compared to the global plasmid pan-genome of *C. jejuni*, it shows the presence of 99 genes out of 168. It possible that the plasmid is integrated into the chromosome or that the plasmid is larger than 100 kbp and impossible to extract with any of the techniques I used in Chapter 3.

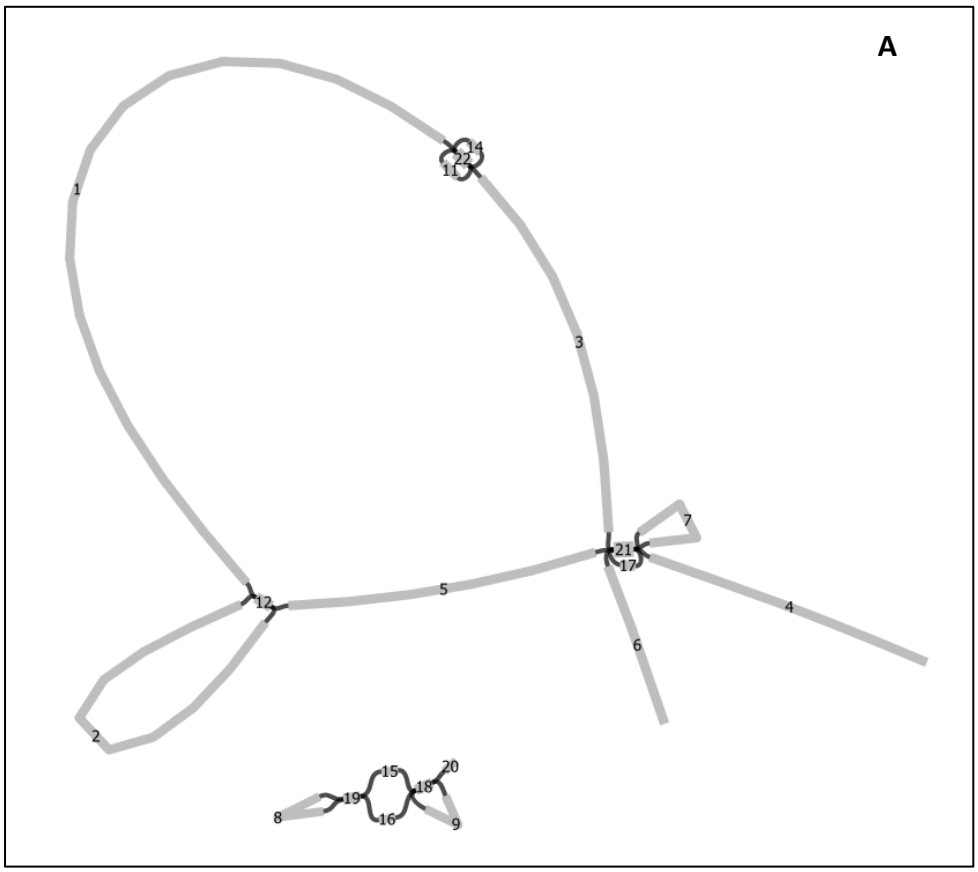
Strain Cj1 was sequenced using Illumina HiSeq (100bp paired-end reads, short insert) and Illumina MiSeq (300bp paired-end reads, short insert). An assembly was performed from this sequencing data using SPAdes (Bankevich *et al.,* 2012; Materials and Methods, Section 2.10.4). The final assembly statistics are reported in Table 3.3. The analysis of the graph file performed with Bandage (Wick *et al.,* 2015) shows two distinct contig sets (Figure 5.2 A), whose contiguity can be inferred via the paired end information. The longer contig set is 1.6 Mb and represents the chromosome of strain Cj1. The shorter contig set is 115 kbp total and includes the 99 genes from the plasmid pan genome (Figure 5.2 B). This is consistent with reports that some strains of *C. jejuni* possess megaplasmids (pCJDM202 and pCJDM67 L, Table 3.14). Analysing the coverage of the contigs included in the shorter set, it is possible that 2 of them (contig 18 and 19) are repeated in the mega plasmid of Cj1, assuming it is present, as they show a read coverage almost double when compared to the rest of the contigs included in this set (Figure 5.2 C).

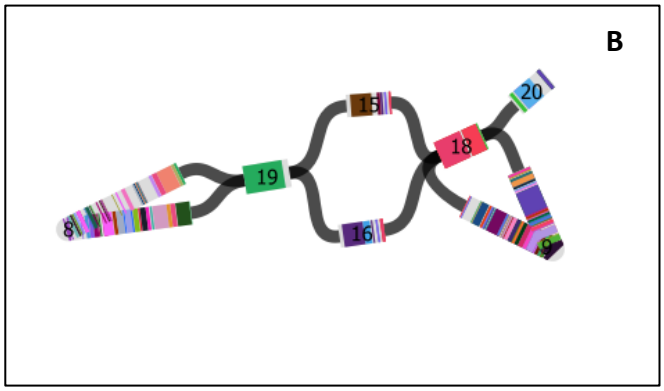
A Blastn search against the bacterial sequence database of contigs 8, 15, 16, 18, 19 and 20 against the NCBI database showed similarity between 95 and 100% with plasmids pCJDM202, pCJDM67 L, the plasmid from strain RM3194 and plasmid pcjDM (accession numbers CP014743.1, CP014745.1, CP014345.1, CP013117.1). The same contigs from the Cj1 genome assembly show a high percent of similarity (between 95% and 100%) with the genome of strains 00-1597 and RM1221, which are not reported to possess plasmids.

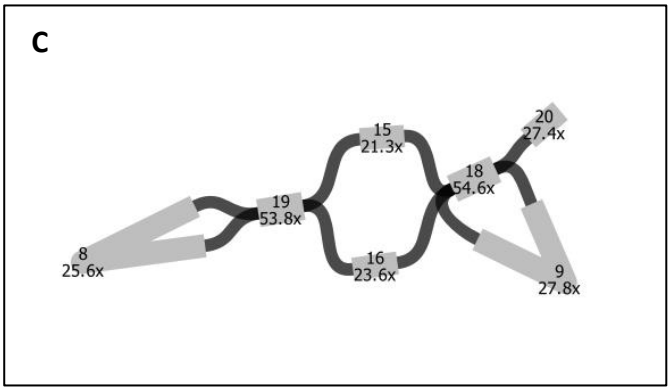
Table 5.3 shows the results of the Blastn search against the bacterial sequence database for the contigs 8, 9, 15, 16, 18, 19, and 20.

Contig 9 of strain Cj1 shows the presence of 45 genes listed in the pTet-like plasmid pan genome (Figure 5.2 D), and its sequence was used to design the mutants. An image representing the contig 9 is reported in Figure 5.3. The genes present in contig 9 of the genome sequence of Cj1 are described in Table 5.3.

Figure 5.2 - Genome assembly of the strain Cj1 (A) Graphical representation of the assembly of the strain Cj1 using Bandage software. Each grey segment of the figure represents a contig of the genome assembly (grey lines), linked to each other using the “graph” information provided by the software SPAdes. Numbers shown on the grey segments are referred to the number of the contig in the genome assembly. Two contig sets are identifiable; each contig in the different sets is linked thanks to the paired end information. (B) Zoom on the shorter contig set (total length 115 kbp). Colours are referred to the BLASTn hits of the *C. jejuni* plasmid pan-genome genes versus the strain Cj1. (C) Zoom on the shorter contig set. Average coverage of each contig, as reported by the automated assembler SPAdes, is shown on each contig. (D) Zoom on the shorter contig set. Colours are referred to the BLASTn hits of the pTet-like plasmid pan-genome genes versus the strain Cj1.







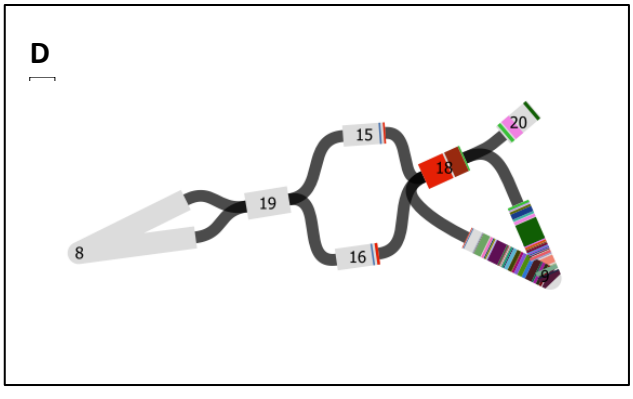


Table 5.3 – Results of the first four BLASTn hits of contigs 8, 9, 15, 16, 18, 19 and 20 of *C. jejuni* strain Cj1 genome assembly versus the bacterial genome database. For each of the seven contigs each column shows (from left to right): the description of the the sequence hit, the maximum BLASTn score, the total score, the % of coverage of the query, the e-value, the % of identity, the accession number of the hit sequence. Contigs 8, 15, 16, 18, 19, and 20 show hits both on the chromosome and on plasmids of *C. jejuni* strains.

Query	Description	Max score	Tot score	Query cov.	E-value	% identity	Acc. no.
NODE_8+_length_61577_cov_25.6027	*Campylobacter jejuni strain WP2202 plasmid pCJDM202, complete sequence*	56061	1.20E+05	100%	0	99%	CP014743.1
	Campylobacter jejuni strain OD267 plasmid pCJDM67 L, complete sequence	48592	1.18E+05	98%	0	99%	CP014745.1
	Campylobacter jejuni strain RM3194 plasmid, complete sequence	48228	1.17E+05	98%	0	99%	CP014345.1
	Campylobacter jejuni subsp. jejuni strain 00-1597, complete genome	31388	71009	60%	0	98%	CP010306.1
NODE_9+_length_43551_cov_27.7867	*Campylobacter coli strain OR12 plasmid pOR12TET, complete sequence*	36326	42594	54%	0	98%	CP013735.1
	Campylobacter coli strain CO2-160 plasmid pccdm3, complete sequence	36287	68965	88%	0	98%	CP013033.1
	Campylobacter jejuni subsp. jejuni 81-176 plasmid pTet, partial sequence	33323	69191	90%	0	99%	CP000549.1
	Campylobacter jejuni strain 81-176 plasmid pTet, complete sequence	33318	69156	90%	0	99%	AY714214.1
NODE_15+_length_3370_cov_21.3309	*Campylobacter jejuni strain WP2202 plasmid pCJDM202, complete sequence*	5526	8114	100%	0	99%	CP014743.1
	Campylobacter jejuni strain OD267 plasmid pCJDM67 L, complete sequence	5520	8114	100%	0	99%	CP014745.1
	Campylobacter jejuni strain RM3194 plasmid, complete sequence	3646	9254	100%	0	98%	CP014345.1
	Campylobacter jejuni RM1221, complete genome	3256	7949	100%	0	94%	CP000025.1
NODE_16+_length_2770_cov_23.6115	*Campylobacter jejuni strain OD267 plasmid pCJDM67 L, complete sequence*	4935	7292	100%	0	99%	CP014745.1
	Campylobacter jejuni strain WP2202 plasmid pCJDM202, complete sequence	4924	7292	100%	0	99%	CP014743.1
	Campylobacter jejuni strain RM3194 plasmid, complete sequence	4905	7693	100%	0	99%	CP014345.1
	Campylobacter jejuni RM1221, complete genome	3843	6909	100%	0	99%	CP000025.1
NODE_18+_length_1994_cov_54.6422	*Campylobacter jejuni strain WP2202 plasmid pCJDM202, complete sequence*	3659	6513	100%	0	99%	CP014743.1
	Campylobacter jejuni strain OD267 plasmid pCJDM67 L, complete sequence	3603	6242	100%	0	99%	CP014745.1
	Campylobacter jejuni strain T1-21 plasmid pcjDM, complete sequence	3090	3090	100%	0	95%	CP013117.1
	Campylobacter coli CVM N29710 plasmid pN29710-1, complete sequence	2191	2191	100%	0	87%	CP004067.1
NODE_19+_length_1237_cov_53.7802	*Campylobacter jejuni strain OD267 plasmid pCJDM67 L, complete sequence*	2285	4570	100%	0	100%	CP014745.1
	Campylobacter jejuni strain WP2202 plasmid pCJDM202, complete sequence	2285	4570	100%	0	100%	CP014743.1
	Campylobacter jejuni strain RM3194 plasmid, complete sequence	2257	4515	100%	0	99%	CP014345.1
	Campylobacter jejuni RM1221, complete genome	2252	4504	100%	0	99%	CP000025.1
NODE_20+_length_1130_cov_27.3659	*Campylobacter jejuni strain WP2202 plasmid pCJDM202, complete sequence*	2069	2468	100%	0	99%	CP014743.1
	Campylobacter jejuni strain RM3194 plasmid, complete sequence	2058	2129	100%	0	99%	CP014345.1
	Campylobacter jejuni strain OD267 plasmid pCJDM67 L, complete sequence	2052	2452	100%	0	99%	CP014745.1
	Campylobacter jejuni RM1221, complete genome	2041	2041	100%	0	99%	CP000025.1

Figure 5.3 - Graphical representation of the genes annotated on the Contig 9 (NODE_9+_length_43551_cov_27.7867) of the Cj1 draft genome sequence. Yellow arrows represent the genes; the gene names (as reported in Chapter 3.3) are linked to each arrow.

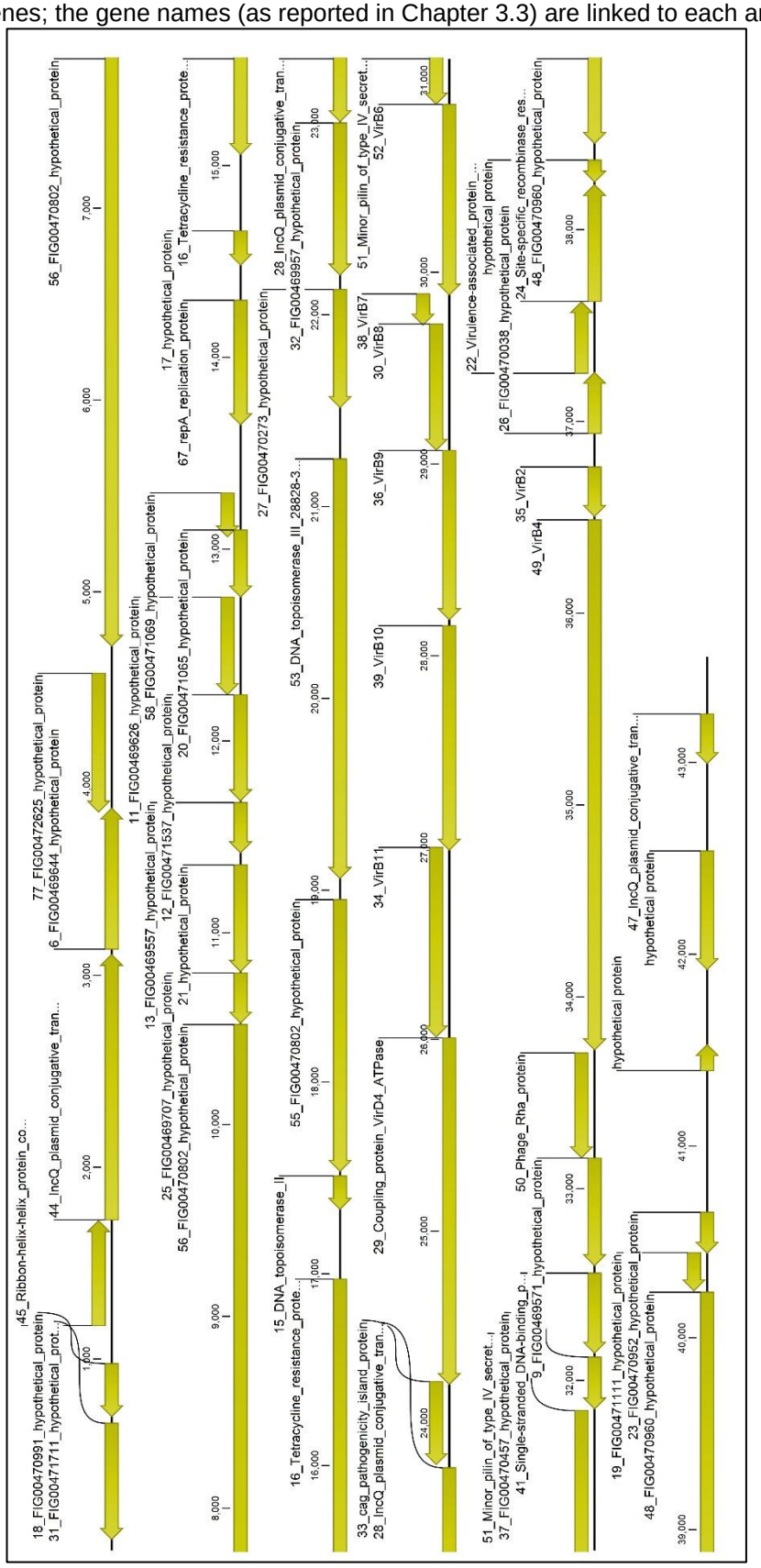


Table 5.4 - Genes annotated on the Contig 9 (NODE_9+_length_43551_cov_27.7867) of the draft genome sequence of strain Cj1. Columns show (from left to right): gene name assigned by RAST automated annotation, gene name as assigned in the pTet-like plasmid pan genome analysis (chapter 3, section 3), relative position of the first gene nucleotide, relative position of the end nucleotide, transcription direction, names of the mutant strains (for genes selected for mutagenesis, described in Chapter 5.3), cluster number of the genes selected for mutagenesis (according to the pTet-like plasmid pan genome analysis described in Chapter 3.3.2).

Gene name (RAST assembly)	Gene name (Homologous to pTet-like plasmsd pan-genome)	Start	End	Direction	Mutant name	Cluster (Figure 3.24)
fig|354242.102.peg.1781	31_FIG00471711_hypothetical_protein_NTPase	57	668	-		
fig|354242.102.peg.1782	18_FIG00470991_hypothetical_protein	697	978	-	*Dcri*	*1*
fig|354242.102.peg.1783	45_Ribbon-helix-helix_protein_copG_family_domain_protein	1172	1723	+		
fig|354242.102.peg.1784	44_IncQ_plasmid_conjugative_transfer_TraR_(VirD2)	1723	3111	+		
fig|354242.102.peg.1785	6_FIG00469644_hypothetical_protein	3134	3874	+	*Datr*	*3*
fig|354242.102.peg.1786	77_FIG00472625_hypothetical_protein	3849	4577	-	*DlabA*	*3*
fig|354242.102.peg.1787	56_FIG00470802_hypothetical_protein	4714	10524	-		
fig|354242.102.peg.1788	25_FIG00469707_hypothetical_protein	10527	10793	-	*Dhyp25*	*2*
fig|354242.102.peg.1789	21_hypothetical_protein	10796	11356	-	*DparA*	*2*
fig|354242.102.peg.1790	13_FIG00469557_hypothetical_protein	11415	11681	-		
fig|354242.102.peg.1791	12_FIG00471537_hypothetical_protein	11686	12243	-	*DepsG*	*3*
fig|354242.102.peg.1792	11_FIG00469626_hypothetical_protein	12240	12752	-	*Dmate*	*3*
fig|354242.102.peg.1793	20_FIG00471065_hypothetical_protein	12749	13102	-	*Dabc*	*1*
fig|354242.102.peg.1794	58_FIG00471069_hypothetical_protein	13062	13295	-		
fig|354242.102.peg.1795	67_repA_replication_protein	13647	14300	-		
fig|354242.102.peg.1796	17_hypothetical_protein	14479	14661	-		
fig|354242.102.peg.1797	16_Tetracycline_resistance_protein_TetO	15056	16975	-	*DtetO*	*3*
fig|354242.102.peg.1798	15_DNA_topoisomerase_II	17334	17513	-		
fig|354242.102.peg.1799	55_FIG00470802_hypothetical_protein	17532	18953	-		
fig|354242.102.peg.1800	53_DNA_topoisomerase_III_28828-31020_00-0949_pTet	19059	21251	-		
fig|354242.102.peg.1801	27_FIG00470273_hypothetical_protein	21514	22134	-	*Dhyp27*	*1*
fig|354242.102.peg.1802	32_FIG00469957_hypothetical_protein	22205	23002	-	*DparB*	*1*
fig|354242.102.peg.1803	28_IncQ_plasmid_conjugative_transfer_protein_TraQ	23005	23769	-		
fig|354242.102.peg.1804	33_cag_pathogenicity_island_protein	23781	24218	-	*Dcag*	*2*
fig|354242.102.peg.1805	29_Coupling_protein_VirD4_ATPase	24199	26010	-		
fig|354242.102.peg.1806	34_VirB11	26011	27003	-		
fig|354242.102.peg.1807	39_VirB10	26984	28159	-		
fig|354242.102.peg.1808	36_VirB9	28186	29073	-		
fig|354242.102.peg.1809	30_VirB8	29070	29732	-		
fig|354242.102.peg.1810	38_VirB7	29722	29889	-		
fig|354242.102.peg.1811	52_VirB6	29879	30877	-		
fig|354242.102.peg.1812	51_Minor_pilin_of_type_IV_secretion_complex_VirB5	30874	31845	-		
fig|354242.102.peg.1813	37_FIG00470457_hypothetical_protein	31849	32124	-		
fig|354242.102.peg.1814	41_Single-stranded_DNA-binding_protein	32140	32562	-		
fig|354242.102.peg.1815	9_FIG00469571_hypothetical_protein	32596	33162	-	*Dunk9*	*3*
fig|354242.102.peg.1816	50_Phage_Rha_protein	33159	33710	-		
fig|354242.102.peg.1817	49_VirB4	33721	36489	-		
fig|354242.102.peg.1818	35_VirB2	36502	36765	-	*DvirB2*	*1*
fig|354242.102.peg.1819	26_FIG00470038_hypothetical_protein	36935	37258	+	*Dcoag*	*1*
fig|354242.102.peg.1820	22_Virulence-associated_protein_2_(vapD)	37249	37626	+	*DvapD*	*1*
fig|354242.102.peg.1821	24_Site-specific_recombinase_resolvase	37623	38237	+		
fig|354242.102.peg.1822	hypotetical protein	38246	38365	-		
fig|354242.102.peg.1823	48_FIG00470960_hypothetical_protein	38447	40240	-		
fig|354242.102.peg.1824	19_FIG00471111_hypothetical_protein	40243	40446	-	*DhicA*	*2*
fig|354242.102.peg.1825	23_FIG00470952_hypothetical_protein	40439	40657	-		
fig|354242.102.peg.1826	hypotetical protein	41391	41531	+		
fig|354242.102.peg.1827	hypotetical protein	41918	42541	-		
fig|354242.102.peg.1828	47_IncQ_plasmid_conjugative_transfer_protein_TraG	42992	43255	-		

5.3 Genes selected for mutagenesis

A set of 17 genes were selected from the pTet-plasmid pan genome for mutagenesis. These genes were part of either the “core” or the “accessory” pTet plasmid pan-genome. They were selected on the basis of the automated annotation of the genome fo Cj1 and of the conserved domains identified, described in Chapter 4. I decided to design mutants on genes present in the genome of the strain *Cj1*, whose function was not fully described or on genes whose function could be linked to phenotypes previously undescribed in *C. jejuni*. The following targets were selected for mutagenesis. The names given to the mutants reflect the conserved domains identified in the hypothetical protein if a conserved domain was found.

Datr - 6_FIG00469644_hypothetical_protein (see S4.4) - A partial acetyl transferase domain is present in the hypothetical protein encoded by this gene (Neuwald *et al.,* 1997, Cort *et al.,* 2008).

DlabA - 77_FIG00472625_hypothetical_protein (see S4.64) - This hypothetical protein shows a LabA-like conserved domain, associated with a variety of diverse regulatory and metabolic functions (Taniguchi *et al.,* 2007; Taniguchi *et al.,* 2010; Tang *et al.,* 2008; Parsons *et al.,* 2002).

DepsG - 12_FIG00471537_hypothetical_protein (see S4.9) - Several conserved domains are here identified: a domain for an ATPase associated with diverse cellular activity (Iyer *et al.,* 2004), a CsbA domain involved in stress response in *B. subtilis* (Petersohn *et al.,* 2001), and an EpsG domain involved in production and maintenance of biofilm structure (Branda *et al.,* 2004).

Dabc - 20_FIG00471065_hypothetical_protein (see S4.15) - A conserved domain for an ABC (ATP-binding cassette) membrane transporter is described in this

hypothetical protein (Reizer *et al.,* 1992), together with a VanZ domain for glycopetide antibiotics resistance (Arthur *et al.,* 1995).

DparB - 32_FIG00469957_hypothetical_protein (see S4.26) - Conserved domains involved with nuclease activity are identified in this hypothetical protein (ParB, AAA-domain, Easter and Gober, 2002). A domain for the cellular sensing of L-fucose is also described (Felder *et al.,* 1999). Together with the ParA domain identified in the protein 21_hypothetical_protein, a putative function of this protein is the control and regulation of DNA replication.

Dcag - 33_cag_pathogenicity_island_protein (see S4.27) - A conserved domain (Cag12) belonging to the cag pathogenicity island of *Helicobacter pylori* is described in this hypothetical protein. The function of Cag12 is unknown (Kim *et al.,* 2006).

Dunk9 - 9_FIG00469571_hypothetical_protein (see S4.6) - The only conserved feature identified in this hypothetical protein is associated with chromosomal structure in meiotic phase in eukaryotic cells (Meuwissen *et al.,* 1992). Its function is not yet fully described but its structure shows motifs present in DNA-binding proteins and several potential targets for protein kinases. There is no function that can be associated with bacterial cells in this hypothetical protein.

Dcoag - 26_FIG00470038_hypothetical_protein (see S4.20) - A conserved domain from a blood coagulase of *S. aureus* is described in this hypothetical protein (Friedrich *et al.,* 2003).

DvapD - 22_Virulence-associated_protein_2_(vapD) (see S4.17) - The VapD domain identified here is part of the CRISPR/Cas superfamily, involved in the protection of the bacterium against foreign DNA (Boyaval *et al.,* 2007; Nam *et al.,* 2012).

DhicA - 19_FIG00471111_hypothetical_protein (see S4.14) - The HicA_toxin domain identifies this hypothetical protein as the cognate toxin of the HicA/B toxin-antitoxin system (see 4.3.18, 23_FIG00470952_hypothetical_protein; Yamaguchi and Inouye, 2011; Makarova *et al.,* 2006).

Dcri - 18_FIG00470991_hypothetical_protein (see S4.13) - This hypothetical protein shows several conserved domains, including a a Cas8a1 domain associated to CRISPR/Cas proteins (Boyaval *et al.,* 2007) and a PRK14149 domain, part of the GrpE superfamily which includes heat shock and bacterial chaperones proteins (Harrison, 2003).

Dhyp25 - 25_FIG00469707_hypothetical_protein - the conserved domains search did not highlight the presence an any conserved domain in thys hypothetical protein.

DparA - 21_hypothetical_protein (see S4.16) - This hypothetical protein shows the presence of a ParA domain, which is part of the ParA-ParB system and promotes the regulation of DNA replication (Easter and Gober, 2002). FlgG, a member of the ParA-ATPase superfamily, has been shown to be regulate the production of polar flagella in *C. jejuni* (Balaban and Hendrixson, 2011).

Dmate - 11_FIG00469626_hypothetical_protein (see S4.8) - The conserved domain identified in this protein is associared with the MATE protein family (“Multidrug And Toxic compound Extrusion”; Putman *et al.,* 2000; Hvorup *et al.,* 2003). The function of this protein family is exporting metabolites across the cell membrane and conferring multi drug resistance to bacteria.

Dhyp27 - 27_FIG00470273_hypothetical_protein (see S4.22) - With the exception of a single domain of unknown function conserved in *Mycoplasma* species

(Sasaki *et al.,* 2002), this protein does not show the presence of any conserved domain.

DvirB2 – 35_VirB2 (see S4.29) - According to the automated annotation and the conserved domains search, this protein encodes for the main component of the type four secretion system pilus (VirB2; Kalkum *et al.,* 2002; Schulein and Dehio, 2002).

A further mutant has been designed in the *tet(O)* gene in order to have a positive control for a mutation with a known phenotype.

The targets selected for mutagenesis are also highlighted in Table 5.4, indicating the associated mutant name. The characteristics of the genes selected for deletion are summarised in Table 5.5.

Table 5.5 – Summary of the characteristics of the genes selected for deletion. Columns show (from left to right): mutant name, gene name (accordin to the pTet-like plasmid pan-genome described in Chapter 3.3), predicted gen function, gene length, and protein length.

	Name (Chapter 3.3)	Predicted function from associated domains	Gene (bp)
Datr	6_FIG00469644_hypothetical_protein	Acetyltransferase	705
Dunk9	9_FIG00469571_hypothetical_protein	Unknown - Chromosome structure [Eukariotic]	552
DepsG	12_FIG00471537_hypothetical_protein	AAA_23 ATPase - EpsG domain [Biofilm/EPS]	558
Dabc	20_FIG00471065_hypothetical_protein	ABC transporter - Glycopeptide antibiotics resistance	399
Dcoag	26_FIG00470038_hypothetical_protein	Blood coagulase (S. aureus)	324
Dcag	33_cag_pathogenicity_island_protein	Membrane protein (Helicobacter pylori)	438
DparB	32_FIG00469957_hypothetical_protein	ParB-like domain - Ligand-binding of fucose receptor - AAA15 ATPase	798
DlabA	77_FIG00472625_hypothetical_protein	LabA regulatory domain family	729
DhicA	19_FIG00471111_hypothetical_protein	HicA toxin	204
DvapD	22_Virulence-associated_protein_2_(vapD)	CRISPR/Cas2	408
DtetO	16_Tetracycline_resistance_protein_TetO	Tetracycline resistance	1920
Dcri	18_FIG00470991_hypothetical_protein	CRISPR/Cas8 - Viral protein domain - Varous cellular functions	282
Dhyp25	25_FIG00469707_hypothetical_protein	---	267
DparA	21_hypothetical_protein	ParA ATPase - replication protein	561
Dmate	11_FIG00469626_hypothetical_protein	MATE-like protein [Multidrug toxin extrusion]	513
Dhyp27	27_FIG00470273_hypothetical_protein	Alpha helices rich domain	621
DvirB2	35_VirB2	TrbC/VirB2	264

5.4 Mutagenesis of *Campylobacter jejuni*

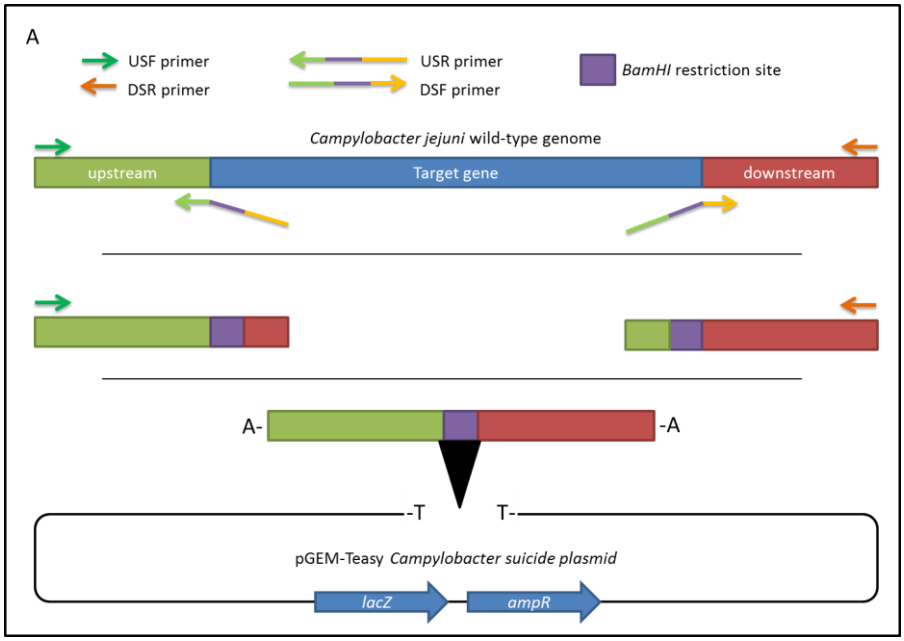
The protocol for the mutagenesis of *C. jejuni* was a variation of the allelic exchange protocol used previously for *Campylobacter* and other bacteria (Van Vliet *et al.,* 1998, Karlyshev and Wren 2005, Tunio *et al.,* 2010).

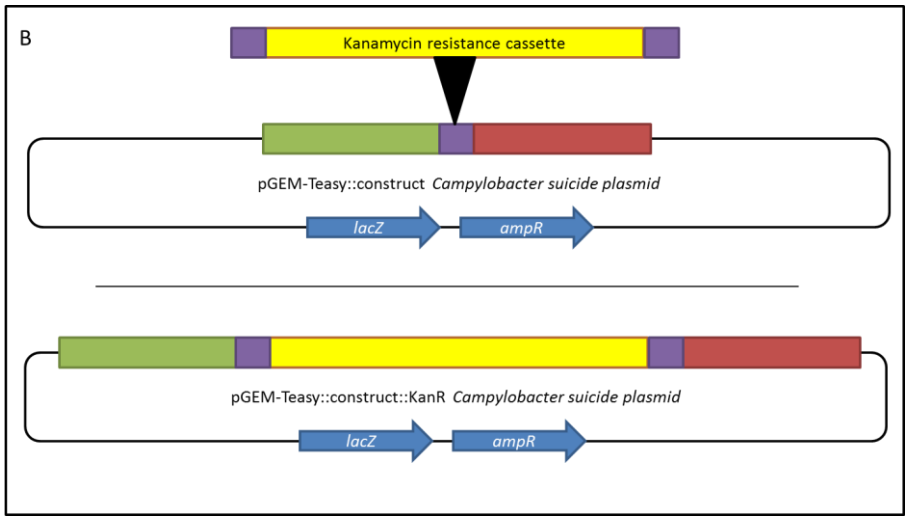
The protocol is summarised in Figure 5.4.1. PCR was used to amplify 500 to 700 base pair regions upstream and downstream the target gene. The PCR primers were designed so that amplification would introduce a 20 to 25 bp complementary overhang at the 3' end of the upstream region and the 5' end of the downstream region. This overhang included a *bamHI* restriction site (Figure 5.4 A).

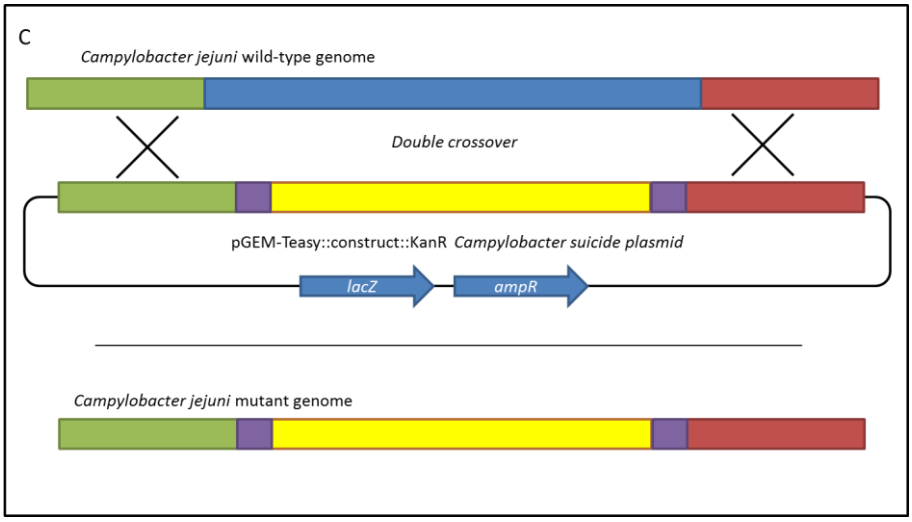
An overlapping extension PCR was performed using the upstream forward primer (USF) and the downstream reverse primer (DSR) and the two purified PCR products from the previous step as template DNA (Figure 5.4 A). The construct, composed of 500 to 700 bp flanking the 5' and the 3' extremity of the gene of interest, intervaled by a *bamHI* site, was cloned in pGEM-T easy (Promega), via TA cloning (Figure 5.4.1 A, Materials and Methods, Section 2.4). The pGEM::construct vector was cloned into chemically competent *E. coli* TOP10 and positive clones were selected (Materials and Methods, Section 2.5). The pGem::construct vector was then digested with *bamHI* enzyme. A plasmid vector pJMK30 containing a kanamycin resistance cassette (kan^R_cas, van Vliet *et al.,* 1998, kindly provided by Dr Mark Reuters) was digested with *bamHI,* resulting in the digestion of the cassette from the plasmid (Karlyshev and Wren, 2005). The kan^R_cas was subsequently cloned into the digested pGem::construct vector, resulting in pGem::construct::kan^R_cas vector (Figure 5.4 B). This vector was then introduced in *Campylobacter* via electroporation (Van Vliet *et al.,* 1998).

pGEM-T Easy is not able to replicate in *Campylobacter:* the homologous regions on the vector promotes the double recombination event on the *Campylobacter* genome. The result is a substitution of the gene of interest with the kanamycin cassette (Figure 5.4 C). *C. jejuni* transformants were screened on MH kan50 plates and patched replicated on CBA with Skirrow supplement (Materials and methods 2.6).

Figure 5.4 - *C. jejuni* mutagenesis protocol. (A) Primer pairs USF-USR and DSF-DRS are used to amplify 500-700 base pairs upstream and downstream of the gene of interest and to include a *BamHI* in the PCR product. Overlapping extension is used to join the upstream and the downstream segment, which is then cloned in a pGEM-Teasy vector (pGEM::construct). (B) A kanamycin resistance cassette is cloned in the pGEM::construct vector using the BamHI restriction site (pGEM::construct:: kan^R_cas). (C) The pGEM::construct:: kan^R_cas is transformed in *C. jejuni*. The homology regions promote the double recombination event, and the substitution of the target gene with the kan^R cassette.


A
USF primer
DSR primer
USR primer
DSF primer
BamHI restriction site
Campylobacter jejuni wild-type genome
upstream
Target gene
downstream
A-
-A
-T
T-
pGEM-Teasy Campylobacter suicide plasmid
lacZ
ampR


B
Kanamycin resistance cassette
pGEM-Teasy::construct Campylobacter suicide plasmid
lacZ
ampR
pGEM-Teasy::construct::KanR Campylobacter suicide plasmid
lacZ
ampR


C
Campylobacter jejuni wild-type genome
Double crossover
pGEM-Teasy::construct::KanR Campylobacter suicide plasmid
lacZ
ampR
Campylobacter jejuni mutant genome

5.5 Constructs design and mutagenesis

Primer pairs were used to amplify the upstream and the downstream section of each target gene according to Table 5.6. A gradient of temperatures was tested to find the optimal annealing temperature for each primer pair. Figure 5.5 A and B show the result of a temperature gradient of annealing temperature for the primer pairs Gac_USF/Gac_USR and Gac_DSF/Gac_DSR. The amplicons were then used in an overlapping extension PCR (Ho *et al.,* 1988), inserting a *BamHI* restriction site between the two fragments. Figure 5.5 C shows the result of such reaction for the construct necessary for the mutant *Dcag*.

The construct was cloned into pGEM-T Easy plasmid (Promega) obtaining the plasmids listed in Table 5.7 and transformed into chemically competent *E. coli* TOP10 cells. The presence of the construct in the pGEM-T Easy plasmid was assessed via amplification with SP6 and T7 primers (annealing outside the integration point of pGEM-T Easy) and sequencing. The recombinant plasmids pGEM carrying the construct of interest and plasmid pJMK30 carrying the kanamycin (kan^R) resistance cassette were digested with the *BamHI* restriction enzyme, obtaining a linearised pGEM::construct plasmid and a 1440 bp *kan* cassette. The result of the *BamHI* digestion for the plasmid vector pJMK30 is shown in Figure 5.6.

The kan cassette and the linearised pGEM::construct plasmid were purified from the agarose gel and ligated to obtain a pGEM::construct::kan_cassette plasmid; Table 5.7 lists the plasmids obtained. To check that the orientation of the kan^R cassette was the same of the gene to be deleted, the plasmid pGEM::construct:: kan^R _cassette was used as PCR template with the primer pairs kan_out_F/T7 and kan_out_R/SP6 (Figure 5.7, Table 5.8). The PCR with primers kan_out_F/T7 and kan_out_R/SP6 would only yield a product if the kan^R cassette were integrated with

the same orientation to the fragments of the gene to be deleted. The result of the PCR reaction with the primers kan_out_R/SP6 for the vector pGEM::cag::kan_R is shown in Figure 5.8. Each plasmid pGEM::construct::kan_cassette was also sanger-sequenced using the SP6 and T7 primers and maintained in *E. coli* TOP10 cells.

Finally, each plasmid was transformed into electro-competent *C. jejuni* strain Cj1 cells, promoting the double recombination and the substitution of the target gene with the kanamycin resistance cassette and therefore meaning recombinants should grow on MH plates supplemented with 50 µg/ml of kanamycin.

Six transformant strains did not grow on the screening plate. These were the mutant strains *Dcri, Dhyp25, DparA, Dmate, Dhyp27*, and *DvirB2.* This can be due to the lack of recombination between the construct and the plasmid vector (the substitution of the target gene with the kanamycin cassette did not happen), to fact that the mutation disrupted an essential gene, or due to the fact that the pTet-like plasmid carried that may by the strain Cj1 was lost after the mutation of a gene that stopped it from replicating or transmitting to the daughter cells.

The mutants that were able to grow on kanamycin were screened for successful mutation: five to ten colonies of transformant *C. jejuni* were tested for the substitution of the gene of interest with the kanamycin resistance cassette via PCR with the primers XXX_screenEXT_F and XXX_screenEXT_R (Table 5.9). The annealing site of these primers was 50-100 bp upstream of the USF primer and 50-100 bp downstream of the DSR primer, respectively. The resulting PCR products are shown in Figures 5.9 to 5.16, the predicted length of each product and the primer sequences are shown in Table 5.9.

The sequence of each PCR product was assessed via Sanger-sequencing in order to confirm the successful construction of the mutant.

Table 5.6 - Primers designed for the production of each mutant. Columns show (from left to right): mutant name as reported in Section 5.3, name of each primer, sequence of each primer, length of the amplified product, length of the overlapping PCR product.

Mutant Name	Primer Name	Primer Sequence	PCR product Length	Overlapping PCR product Length
DtetO	Dtet_USF	GCAAAATATAATGAATTTGCA	651	1240
	Dtet_USR	TATATGACTTTTGCAAGCTGggatccGTGATTTTCCTCCTATCAAC		
	Dtet_DSF	GTTGATAGGAGGAAAATCACggatccCAGCTTGCAAAAGTCATATA	635	
	Dtet_DSR	TTCATCAGCCGGATAAAGGT		
Datr	Datr_USF	GGCTAAAGAATTGGCTAAG	610	1118
	Datr_USR	AAATTATATAGGAGTAAGAAATAggatccTAAAAACAGACGAAAGGATG		
	Datr_DSF	CATCCTTTCGTCTGTTTTTAggatccTATTTCTTACTCCTATATAATTT	557	
	Datr_DSR	TAGAAAGGGTTGTAAGAG		
DlabA	Dlab_USF	ATAAGCGGAGAAAGAACAG	623	1194
	Dlab_USR	GTTTAATAAAAATAAAAGCAAAAAggatccTGTTGCGTAGTCCTAATTTATTA		
	Dlab_DSF	TAATAAATTAGGACTACGCAACAggatccTTTTTGCTTTTATTTTTATTAAAC	624	
	Dlab_DSR	GAAATAAGAGCGTTTGTG		
DepsG	Deps_USF	TAATGCCTATATTTGTCGCG	659	1238
	Deps_USR	TATCCATCTCTAGCATTTTAggatccTTATCCTCCTTTAATATTTC		
	Deps_DSF	GAAATATTAAAGGAGGATAAggatccTAAAATGCTAGAGATGGATA	625	
	Deps_DSR	ATGTTCTAAGTAAAGAATTG		
Dabc	Dabc_USF	ATGTGAAAGACGGTATAAAG	548	1132
	Dabc_USR	TGCTACTATTGTAATTTTTTggatccTAATCAATCTTTTTTCTTTTTC		
	Dabc_DSF	GAAAAAGAAAAAAGATTGATTAggatccAAAAAATTACAATAGTAGCA	632	
	Dabc_DSR	ATTTACTACTCTCATTTATC		
DparB	Dpar_USF	AGTTAAACCACCTGAATG	595	1217
	Dpar_USR	TTTGTATATAATTGTTCTGTggatccTGTTATTTCTCGTTTATCC		
	Dpar_DSF	GGATAAACGAGAAATAACAggatccACAGAACAATTATATACAAA	667	
	Dpar_DSR	TTGAAATTACAGCTATAAAC		
Dcag	Dcag_USF	TCATAACCTCTTTCCCAATC	594	1082
	Dcag_USR	AATTCAAATTATAGGAGAAAACAggatccAGGAGAGAAAAATGAAAAAAATAT		
	Dcag_DSF	ATATTTTTTTCATTTTTCTCTCCTggatccTGTTTTCTCCTATAATTTGAATT	541	
	Dcag_DSR	TTTATGGCTGGTTATTGG		
Dunk9	Dunk_USF	CAAGAAGTGCAATTTGGCAC	551	1130
	Dunk_USR	CTTTTTGTTGGTGTAATATTTTTTggatccTGCAACTTCTTTGGATCTTT		
	Dunk_DSF	AAAGATCCAAAGAAGTTGCAggatccAAAAAATATTACACCAACAAAAAG	629	
	Dunk_DSR	ATTTGATCTGAATTAAGATC		
Dcoag	Dcoa_USF	AAATTACTAAGTTCCATTTG	704	1373
	Dcoa_USR	GATTTTTGATGAAAGGATAAAAAggatccATCGCAAAGCAATTAATTTT		
	Dcoa_DSF	AAAATTAATTGCTTTGCGATggatccTTTTTATCCTTTCATCAAAAATC	718	
	Dcoa_DSR	TCTTAGACGCAGGATTAG		
DvapD	Dvap_USF	CCATCATCTTTGTATAATAAC	626	1130
	Dvap_USR	AAGCTGAAGGAGAAGGAATggatccATATAGCATACATAAGAGTATC		
	Dvap_DSF	GATACTCTTATGTATGCTATATggatccATTCCTTCTCCTTCAGCTT	551	
	Dvap_DSR	ACATAAATGCAATTGTGAGTAG		
DhicA	Dhic_USF	TAACTTATGGGTGATTGTATGC	505	1025
	Dhic_USR	AATAGTATTGCGTTTAGGCATTTggatccTCAGGCATTTGCTAAAACCATTG		
	Dhic_DSF	CAATGGTTTTAGCAAATGCCTGAggatccAAATGCCTAAACGCAATACTATT	550	
	Dhic_DSR	TACAGAGCTACCCATAAGAC		
Dcri	Dcri_USF	TTTCCACTTCTAAGAATC	610	1107
	Dcri_USR	CTCCTATGATAAATTTTTGTTggatccAACATCTCCTTTATAATATT		
	Dcri_DSF	AATATTATAAAGGAGATGTTggatccAACAAAAATTTATCATAGGAG	544	
	Dcri_DSR	GTTGCTAATTTCTCTATCTG		
Dhyp25	Dhyp25_USF	TAACCATAAAGAAGATAGAG	582	1154
	Dhyp25_USR	GCAAGGATATTGTAACATTTggatccTATTATTTATCCTTTCTAATAAAA		
	Dhyp25_DSF	TTTTATTAGAAAGGATAAATAATAggatccAAATGTTACAATATCCTTGC	622	
	Dhyp25_DSR	GCTTAGTTCGCTAACATC		
DparA	Dpaa_USF	AGCTTCAGGAATACTATG	594	1086
	Dpaa_USR	AAATGGTTTTTTTTCCATTAggatccAAAAAAACCTTTTTTATATTTTTG		
	Dpaa_DSF	CAAAAATATAAAAAAGGTTTTTTTggatccTAATGGAAAAAAAACCATTT	587	
	Dpaa_DSR	TTTAAAGCTTCTGCTATTG		
Dmate	Dmate_USF	GTTGAGTTTCAGCCGAATG	643	1218
	Dmate_USR	TTTGACAAATTTACTACTCTggatccCTTTGATTTACACTCCATGA		
	Dmate_DSF	TCATGGAGTGTAAATCAAAGggatccAGAGTAGTAAATTTGTCAAA	621	
	Dmate_DSR	TATCGCAAATGCCATAAC		
Dhyp27	Dhyp27_USF	AATCCTCCTAATGATGAAG	558	1100
	Dhyp27_USR	CCTTTATGAGGTGCATATTTggatccTTTTTCTCCTTTTTAAATTTTTTA		
	Dhyp27_DSF	TAAAAAATTTAAAAAGGAGAAAAAggatccAAATATGCACCTCATAAAGG	592	
	Dhyp27_DSR	GTAATCTTGTCGCTATGG		
DvirB2	Dvir_USF	ATATTGTCTATGCTCGAAAC	662	1227
	Dvir_USR	CTCCATAGCTAGTCCTATTggatccCCATATCCTTTCGTTAAAAAATTG		
	Dvir_DSF	CAATTTTTTAACGAAAGGATATGGggatccAATAGGACTAGCTATGGAG	614	
	Dvir_DSR	AGGTTAGTTTATCTTCGATAG		

Figure 5.5 - Production of the construct for the mutant *Dcag* (A) Result of the amplification with the primers Dcag_USF/Dcag_USR. Lane 1: 1kb plus ladder; lane 2-9: amplification result on Cj1 strain using a gradient of annealing temperature between 65℃ (lane 2), and 50 ℃ (lane 9); lane 10, NTC. (B) Result of the amplification with the primers Dcag_DSF/Dcag_DSR. Lane 1: 1kb plus ladder; lane 2-9: amplification result on Cj1 strain using a gradient of annealing temperature between 65℃ (lane 2), and 50 ℃ (lane 9); lane 10, NTC. Annealing temperature of 60 ℃ is shown in lane 4. (C) result of overlapping extension PCR used for the *Dcag* mutant. Lane 1-3, overlapping extension PCR (Dcag_USF/ Dcag_DSR primers, using the purified product from 3.5.1 A and 3.5.1 B as template); lane 4, 1kb plus ladder; lane 5, positive control (using the DNA extraction from strain Cj1 as templete); lane 6, NTC. Each gel was 1% agarose in TAE. Run 30-45 minutes at 120 V.

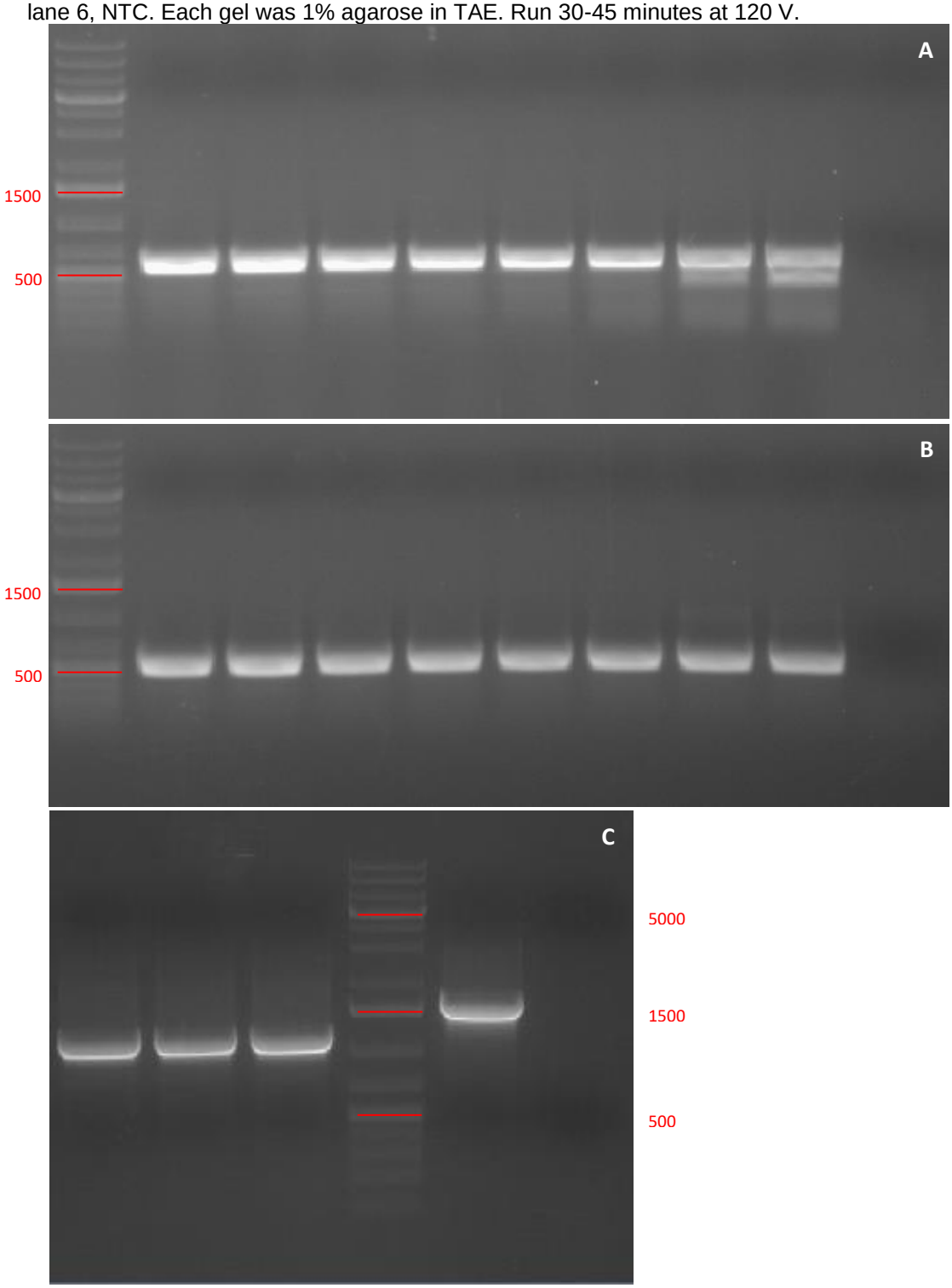


Table 5.7 - Plasmid vectors used for the production of the mutant strains. Columns show (from left to right): plasmid name, length of the plasmid, brief description of the plasmid vector, markers present on each plasmid, reference (where available).

Plasmid Name	Length	Description	Markers	
pGEM-T easy vector	3015	Vector system used for TA-cloning of the overlapping extension products	ampR, lacZ	Kobs, 1997
pJMK30	4185	Campylobacter coli Kanamycin resistance cassette in pUC19	kanR, ampR	Van Vliet et al., 1998
pGEM::tet	4255	pGEM-T easy plasmid vector carrying the 1240 pb overlapping pcr product for the *DtetO* mutant	ampR	This study
pGEM::tet::kan_R	5754	pGEM::tet carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *DtetO* construct	ampR, kanR	"
pGEM::atr	4133	pGEM-T easy plasmid vector carrying the 1118 pb overlapping pcr product for the *Datr* mutant	ampR	"
pGEM::atr::kan_R	5632	pGEM::atr carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Datr* construct	ampR, kanR	"
pGEM::lab	4209	pGEM-T easy plasmid vector carrying the 1194 pb overlapping pcr product for the *DlabA* mutant	ampR	"
pGEM::lab::kan_R	5708	pGEM::lab carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *DlabA* construct	ampR, kanR	"
pGEM::eps	4253	pGEM-T easy plasmid vector carrying the 1238 pb overlapping pcr product for the *DepsG* mutant	ampR	"
pGEM::eps::kan_R	5752	pGEM::eps carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *DepsG* construct	ampR, kanR	"
pGEM::abc	4147	pGEM-T easy plasmid vector carrying the 1132 pb overlapping pcr product for the *Dabc* mutant	ampR	"
pGEM::abc::kan_R	5646	pGEM::abc carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Dabc* construct	ampR, kanR	"
pGEM::par	4232	pGEM-T easy plasmid vector carrying the 1217 pb overlapping pcr product for the *DparB* mutant	ampR	"
pGEM::par::kan_R	5731	pGEM::par carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *DparB* construct	ampR, kanR	"
pGEM::cag	4097	pGEM-T easy plasmid vector carrying the 1082 pb overlapping pcr product for the *Dcag* mutant	ampR	"
pGEM::cag::kan_R	5596	pGEM::cag carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Dcag* construct	ampR, kanR	"
pGEM::unk	4145	pGEM-T easy plasmid vector carrying the 1130 pb overlapping pcr product for the *Dunk9* mutant	ampR	"
pGEM::unk::kan_R	5644	pGEM::unk carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Dunk9* construct	ampR, kanR	"
pGEM::coa	4388	pGEM-T easy plasmid vector carrying the 1373 pb overlapping pcr product for the *Dcoag* mutant	ampR	"
pGEM::coa::kan_R	5887	pGEM::coa carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Dcoag* construct	ampR, kanR	"
pGEM::vap	4175	pGEM-T easy plasmid vector carrying the 1190 pb overlapping pcr product for the *DvapD* mutant	ampR	"
pGEM::vap::kan_R	5644	pGEM::vap carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *DvapD* construct	ampR, kanR	"
pGEM::hic	4040	pGEM-T easy plasmid vector carrying the pb overlapping pcr product for the *DhicA* mutant	ampR	"
pGEM::hic::kan_R	5539	pGEM::hic carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *DhicA* construct	ampR, kanR	"
pGEM::cri	4122	pGEM-T easy plasmid vector carrying the 1107 pb overlapping pcr product for the *Dcri* mutant	ampR	"
pGEM::cri::kan_R	5621	pGEM::cri carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Dcri* construct	ampR, kanR	"
pGEM::hyp25	4169	pGEM-T easy plasmid vector carrying the 1154 pb overlapping pcr product for the *Dhyp25* mutant	ampR	"
pGEM::hyp25::kan_R	5668	pGEM::tet carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Dhyp25* construct	ampR, kanR	"
pGEM::paa	4101	pGEM-T easy plasmid vector carrying the 1086 pb overlapping pcr product for the *DparA* mutant	ampR	"
pGEM::paa::kan_R	5600	pGEM::tet carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *DparA* construct	ampR, kanR	"
pGEM::mate	4233	pGEM-T easy plasmid vector carrying the 1218 pb overlapping pcr product for the *Dmate* mutant	ampR	"
pGEM::mate::kan_R	5732	pGEM::tet carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Dmate* construct	ampR, kanR	"
pGEM::hyp27	4115	pGEM-T easy plasmid vector carrying the 1373 pb overlapping pcr product for the *Dhyp27* mutant	ampR	"
pGEM::hyp27::kan_R	5614	pGEM::tet carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *Dhyp27* construct	ampR, kanR	"
pGEM::vir	4242	pGEM-T easy plasmid vector carrying the 1227 pb overlapping pcr product for the *DvirB2* mutant	ampR	"
pGEM::vir::kan_R	5741	pGEM::tet carrying the 1499 pb kanamycin cassette between the upstream and the downstream fragment of the *DvirB2* construct	ampR, kanR	"

Figure 5.6 - Plasmid pJMK30 digested with BamHI. Lane 1 and 15, 1kb plus ladder (Thermo scientific); lane 2-14, digestion of pJMK30 with *BamHI* restriction enzyme. 1% agarose in TAE. Run for 1 hour at 120 V.

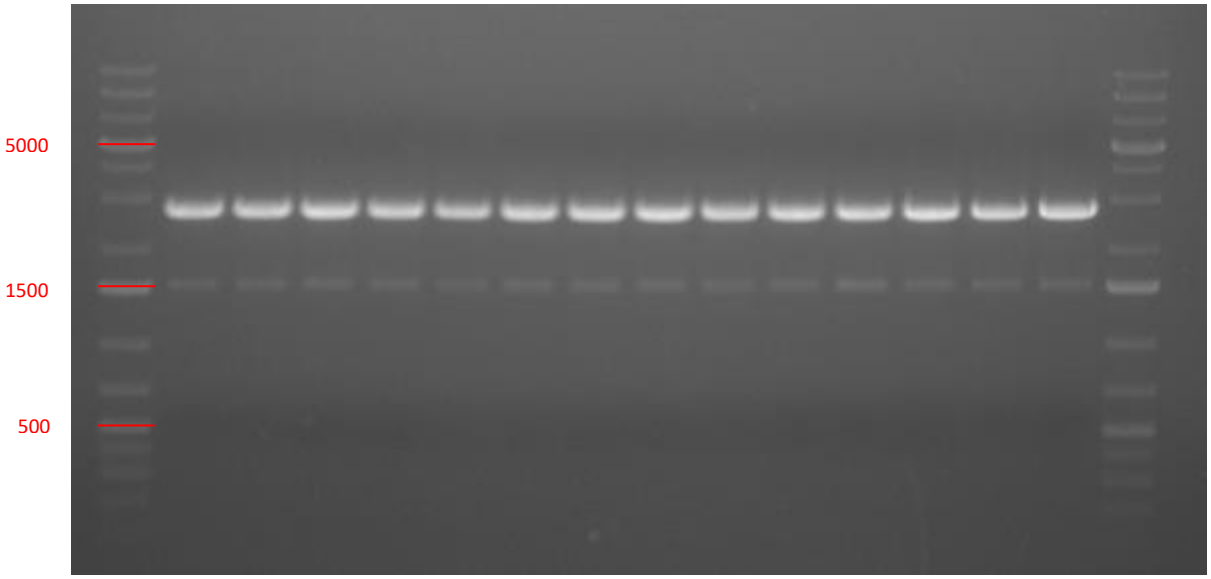


Figure 5.7 - Graphical representation of a pGEM::construct::kanR *Campylobacter* suicide vector. Positions of primers SP6, T7, Kan_out_F and Kan_out_R are shown.

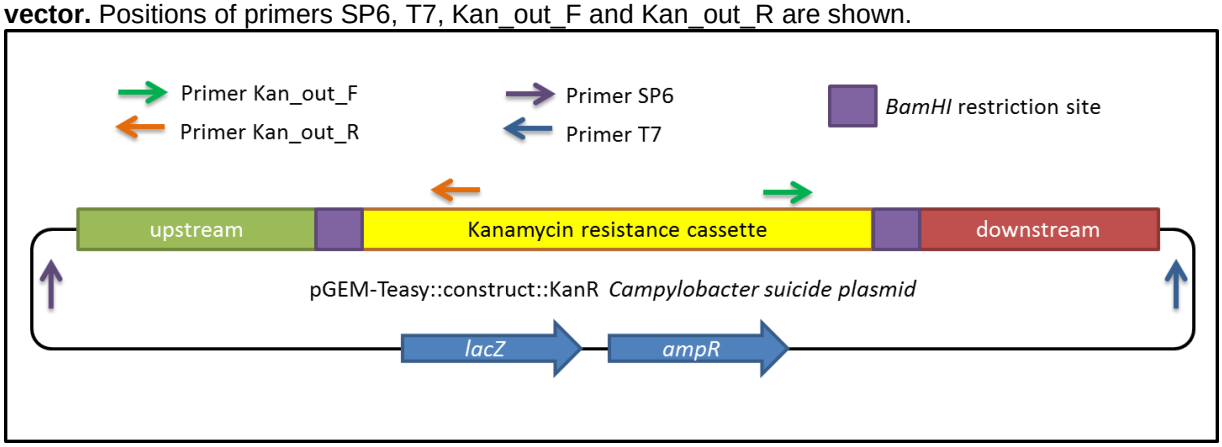


Table 5.8 - PCR primers used to check the integration and the orientation of the kanamycin cassette in the plasmid vectors described in Table 5.7.

Primer name	Primer sequence
kan_out_F	CATCCTCTTCGTCTTGGTAGC
kan_out_R	TTGCCTTCTGCGTCCGGTCG
SP6	ATTTAGGTGACACTATAG
T7	TAATACGACTCACTATAGGG

Figure 5.8 - Result of the screening of six *E. coli* clones transformed with the ligation between pGEM::cag::kan_R and the kanamycin resistance cassette (both digested with *BamHI*). Lane 1, 1kb plus ladder (Thermo scientific); lane 2-7, screening with kan_out_R/SP6 primers of 6 different *E. coli* clones; lane 8, NTC. 1% agarose in TAE. Run for 1 hour at 100 V.

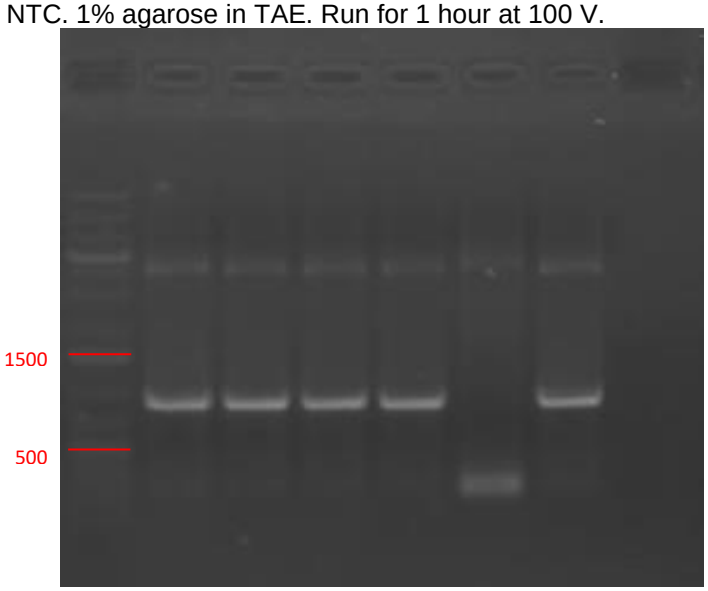


Figure 5.9 - PCR testing the successful production of *DhicA* mutant. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb ladder (Thermo scientific); Lane 2, *DhicA* mutant confirmation PCR; Lane 3, positive control (Cj1 WT); lane 4, NTC. 1% agarose in TAE. Run for 1 hour at 100 V.

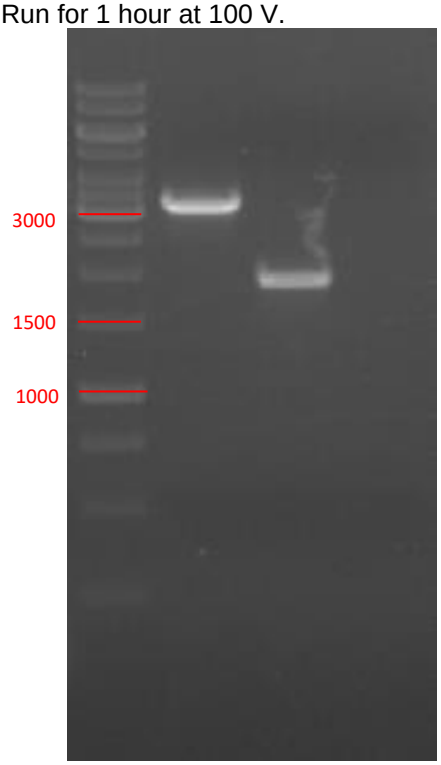


Figure 5.10 - PCR testing the successful production of *DlabA* and *Datr* mutant. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb plus ladder (Thermo scientific); lane 2, positive control (Cj1 WT) with *DlabA* mutant confirmation primers; lane 3, *DlabA* mutant confirmation PCR; lane 5, positive control (Cj1 WT) with *Datr* mutant confirmation primers; lane 6, *Datr* mutant confirmation PCR; lane 4 and 7, NTC. 1% agarose in TAE. Run for 45 minutes at 100 V.

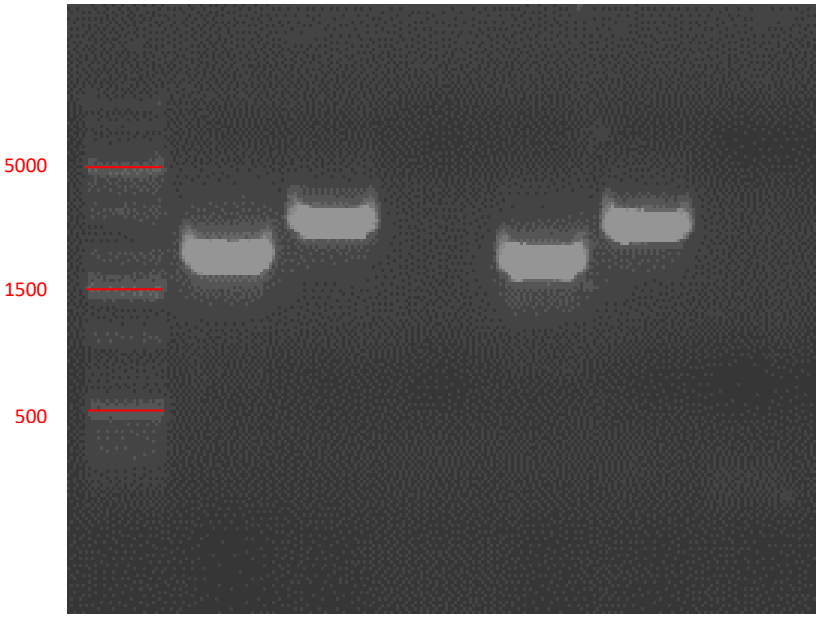


Figure 5.11- PCR testing the successful production of *DvapD* mutant. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb plus ladder (Thermo scientific); Lane 2, positive control (Cj1 WT); Lane 3, *DvapD* mutant confirmation PCR; lane 4, NTC. 1% agarose in TAE. Run for 45 minutes at 100 V.

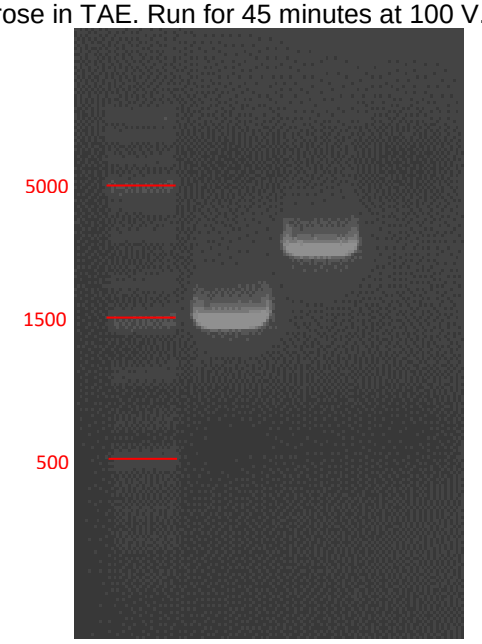


Figure 5.12 - PCR testing the successful production of *Dcoag* mutant. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb plus ladder (Thermo scientific); Lane 2, positive control (Cj1 WT); Lane 3, *Dcoag* mutant confirmation PCR; lane 4, NTC. 1% agarose in TAE. Run for 45 minutes at 100 V.

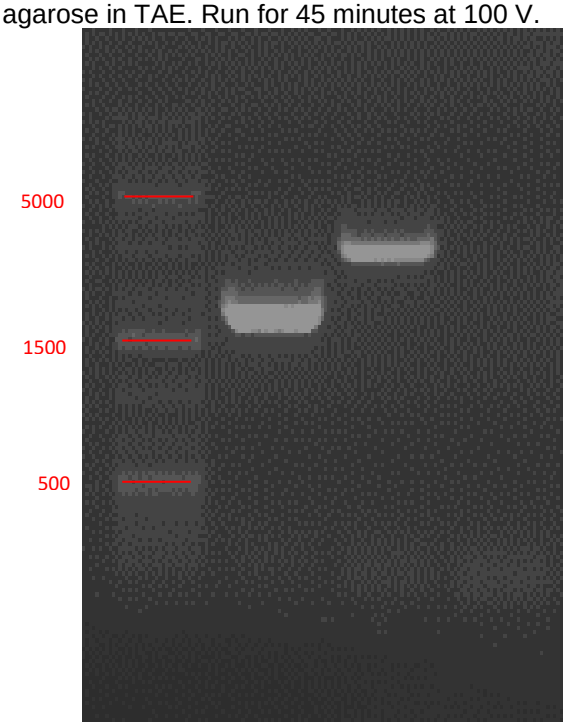


Figure 5.13 - PCR testing the successful production of *DparB* mutant. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb plus ladder (Thermo scientific); Lane 2, positive control (Cj1 WT); Lane 3, *DparB* mutant confirmation PCR; lane 4, NTC. 1% agarose in TAE. Run for 45 minutes at 100 V.

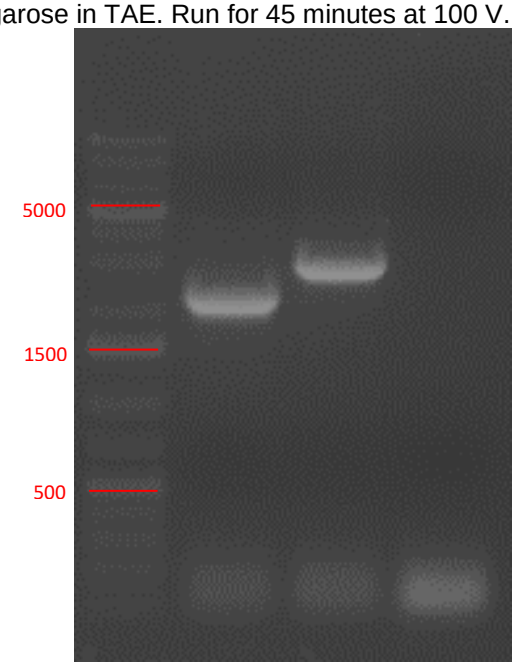


Figure 5.14 - PCR testing the successful production of *Dabc* mutant. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb plus ladder (Thermo scientific); lane 2, NTC; Lane 3, *Dabc* mutant confirmation PCR; Lane 4, positive control (Cj1 WT). 1% agarose in TAE. Run for 1 hour at 100 V.

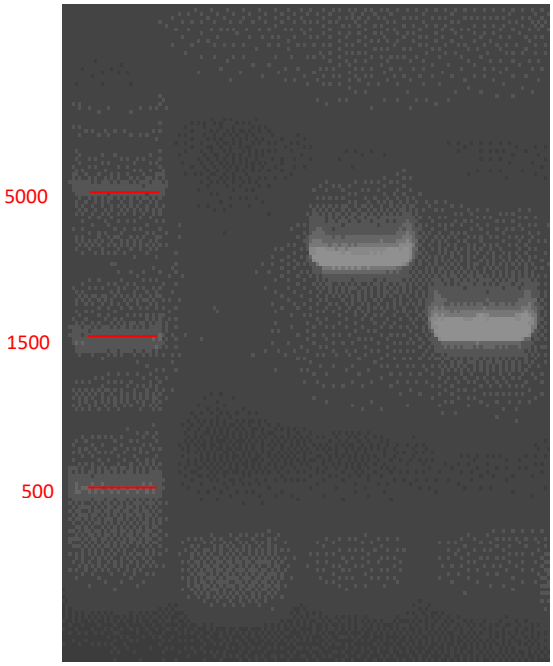


Figure 5.15 - PCR testing the successful production of *DepsG* and *Dunk9* mutant. (A) Confirmation of *DepsG* and *Dunk9* mutant production. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb plus ladder (Thermo scientific); lane 2, *DepsG* mutant confirmation PCR; lane 3, positive control (Cj1 WT) with *DepsG* mutant confirmation primers; lane 5, *Dunk9* mutant confirmation PCR; lane 6, positive control (Cj1 WT) with *Dunk9* mutant confirmation primers; lane 4 and 7, NTC. 1% agarose in TAE. Run for 45 minutes at 100 V. (B) Confirmation of *DepsG* mutant production, repeated at 55℃. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb ladder (Thermo scientific); Lane 2, *DepsG* mutant confirmation PCR; Lane 3, positive control (Cj1 WT); lane 4, NTC. 1% agarose in TAE. Run for 1 hour at 100 V.

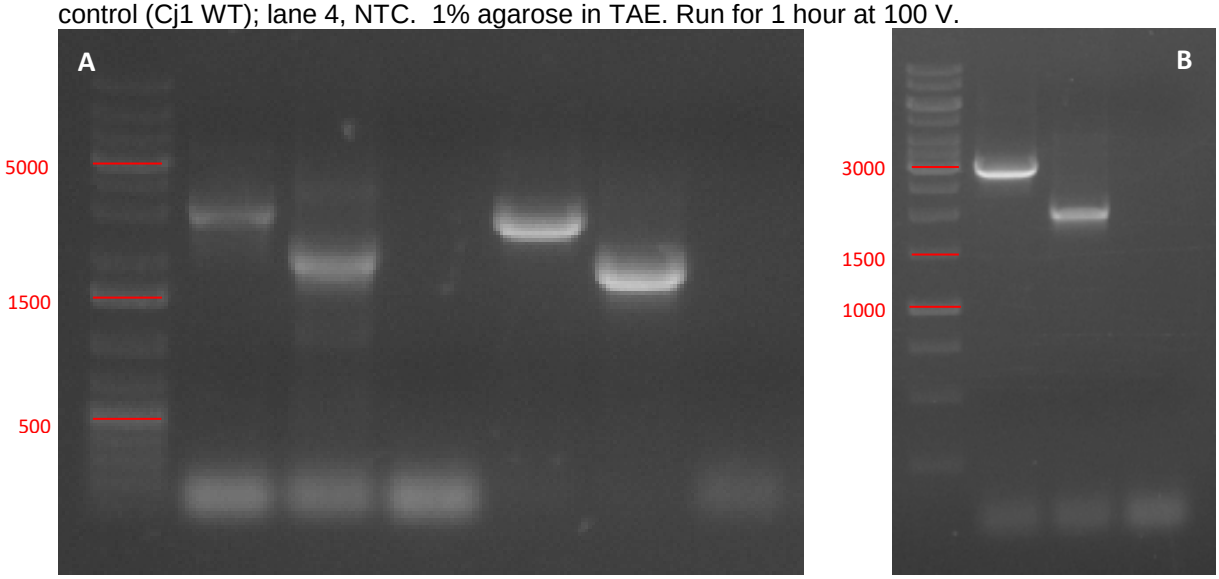


Figure 5.16 - PCR testing the successful production of *DtetO* and *Dcag* mutant. Amplification performed with primers annealing outside of the DNA fragments used for the construct production. Lane 1, 1kb plus ladder (Thermo scientific); lane 3, positive control (Cj1 WT) with *DtetO* mutant confirmation primers; lane 4, *DtetO* mutant confirmation PCR; lane 6, positive control (Cj1 WT) with *Dcag* mutant confirmation primers; lane 7, *Dcag* mutant confirmation PCR; lane 2 and 5, NTC. 1% agarose in TAE. Run for 45 minutes at 100 V.

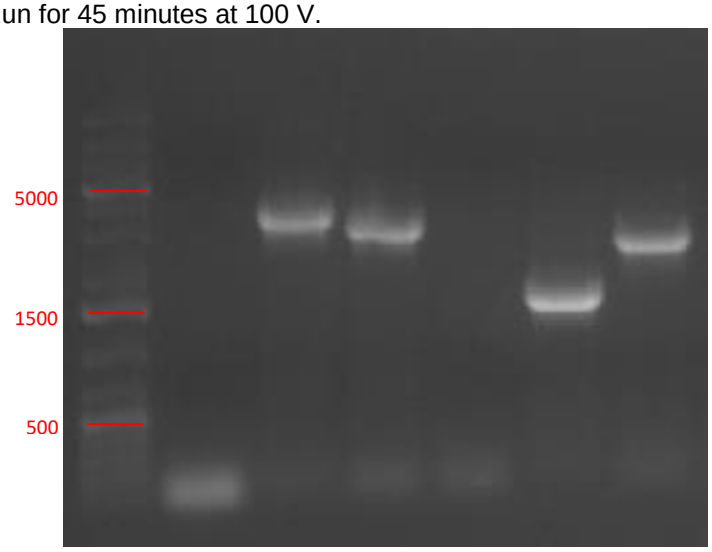


Table 5.9 - Expected lengths and primers sequences for the mutant confirmation PCRs. Columns show (from left to right): name of the mutant strain, length of the target gene, expected length of the PCR product in the Cj1 WT strain, expected length of the PCR in the mutant strain.

Mutant Name	Primer Name	Primer sequence	Gene Length	WT - screen PCR Length	Mutant - screen PCR Length
DtetO	Dtet_screen_F	AATTGCGTTTATACAAGAAG	1920	3454	3032
	Dtet_screen_R	CTTCTTCCCGGTTTCGATAC			
Datr	Datr_screen_F	TTCAAGCAATCCAAATATGG	741	2027	2785
	Datr_screen_R	GTTATTATGTGAGTTATAAAACC			
DlabA	DlabA_screen_F	AGCGATTTAAACAAAAGCAG	729	2083	2853
	DlabA_screen_R	TGAATTTTGAAATAGAGATAAG			
DepsG	DepsG_screen_F	TAGCTATTTGTGCTGTTTGG	558	2034	2975
	DepsG_screen_R	AAGAAGCTCCACTACTCATAG			
Dabc	Dabc_screen_F	TCTTATGGCTTGAAGTGTGG	354	1646	2791
	Dabc_screen_R	AACACGATTTATTCTCAAAAAC			
DparB	DparB_screen_F	CCCATTTTGTGTTTTTGGTG	798	2165	2866
	DparB_screen_R	GTTTTTGTGCTTTCATTTCAGC			
Dcag	Dcag_screen_F	CTTGTTCTCTTTGCGATAC	438	1573	2634
	Dcag_screen_R	TTTATGGCTGGTTATTGG			
Dunk9	Dunk9_screen_F	TGATGAAAAAGTATTAGAGC	567	1900	2832
	Dunk9_screen_R	TTCTTGTTTTTTGCAAAGCTC			
Dcoag	Dcoag_screen_F	CTAGCTCTAGCATTTTCTAG	324	1847	3022
	Dcoag_screen_R	TTTGTCTATAAGAATTTGTGCT			
DvapD	DvapD_screen_F	GGAGCGTTTTATTTTATTAC	378	1562	2674
	DvapD_screen_R	AAATAGGTGCAACTTCTCTAAG			
DhicA	DhicA_screen_F	ATAGAATTTACAAGCTATAAG	204	1925	3102
	DhicA_screen_R	TAAGTGGCTTTATTGATCTTG			

5.6 Discussion

In this chapter, I described how 11 deletion mutants have been obtained in *C. jejuni* strain Cj1.

I targeted plasmid genes, and consequently the first step was to assess the copy number of the plasmid in the selected strains. It is reported that large plasmids are present at low copy number in several strain of *C. jejuni* (Marasini, *et al.,* 2014). In strains Cj1, Cj2, and Cj3 I showed that the copy number was close to one, so compatible with the production of a series of mutants.

Even if the bacterium is reported to be naturally transformable (Vegge *et al*., 2012), conferring a high level of genome plasticity, it is reported that the ability of producing mutants in *C. jejuni* is highly strain-specific, and it is linked to the ability of each strain to accept foreign DNA (Van Vliet *et al.,* 1998). The technique described in this chapter to produce deletion mutants is a variation of the protocol described by Van Vliet *et al.,* 1998, and Karlyshev and Wren, 2005. We used pGEM-T Easy as the backbone vector for cloning of the construct produced via overlapping PCR and subsequent cloning of the kanamycin cassette (derived from pJMK30) in substitution of the gene to be removed in the genome of *C. jejuni*. The vector is then transformed in *C. jejuni* via electroporation. I was not able to induce the allelic exchange of a target gene with a kanamycin resistance cassette in strains such 11168 or 81-176, in which such protocol has been already used for the production of deletion mutants. The reason for this might be the vector used (pGEM-T Easy) or in the *E. coli* strain used for maintaining and replicate the vector (TOP-10).

I applied the targeted mutagenesis protocol to the several strains of *C. jejuni* available: the only strain that was able to accept the introduction of exogenous DNA was the strain Cj1. This strain has a particular genomic set-up; it shows the presence

of all the PCR markers tested in Chapter 3, together with the tetracycline resistance and a total of 48 genes of the pTet-like plasmid pan genome, but it is the only strain in which a full circular plasmid sequence of a pTet-like plasmid could not be determined by analysing the assembly graph with Bandage (Wick *et al.,* 2015), nor could the plasmid be visualised on an agarose gel following a plasmid extraction (as shown in Chapter 3).

The presence of a chromosomally-encoded tetracycline resistance has been observed in this work and previously (Albert *et al*, 2009; Pratt and Korolik, 2005), and although it has not been observed before, we cannot discard the possible integration of a full plasmid pTet-like in the chromosome of the strain Cj1. This strain could also carry a plasmid that cannot be extracted or properly assembled, similar to the megaplasmids isolated from strain RM3194 (NZ_CP014345.1) or pCJDM202 (NZ_CP014743.1).

Strain Cj1 showed the presence of a number of the genes of the pTet-like plasmid pan genome comparable to those strains showing a pTet-like plasmid, such as Cj2 or Cj3. It is unclear whether the strain possesses a larger plasmid or the plasmid pTet-like is integrated in the chromosome. According to the results shown in Section 5.2, the set of a small contigs set of the strain Cj1 shows a high percent of similarity (>90%) with the plasmids pCJDM202, pCJDM67 L, pcjDM, and the plasmid of strain RM3194 (accession NZ_CP014345.1) that are between 80 kbp and 120 kbp long, and the genome of the strain RM1221 that shows plasmid sequences integrated in the chromosome. The graph analysis of the genome assembly produced with the libraries derived by two different sequencing technologies show a structure that cannot be resolved but whose total length resembles the one of the plasmids pCJDM202 and pCJDM67 L. However, when trying to isolate the plasmid

from Cj1, I was never able to visualise it on an agarose gel. The kit used to isolate the plasmid from Cj1 is indicated for plasmid up to 50 kbp, so we cannot discount the possibility that the plasmid size might be greater.

The 17 genes selected for deletion were chosen amongst the genes present in the strain Cj1 and in the pTet-like plasmid pan genome. For the majority of the mutants, I did select genes whose functions were unknown and focussed on hypothetical proteins whose function in *C. jejuni* was not described previously. Several mutants I attempted to produce were in genes encoding for proteins showing conserved domains that could exhibit a bacterial phenotype different from the wild type when deleted, and potentially important for environmental or clinical adaptation, such as the ABC transporter of 20_FIG00471065_hypothetical_protein, the MATE-like domain of 11_FIG00469626_hypothetical_protein or the LabA-like domain in 77_FIG00472625_hypothetical_protein. Other genes I attempted to delete encoded for hypothetical proteins with domains of unknown functions, such as 9_FIG00469571_hypothetical_protein, or showing no conserved domains, such as 27_FIG00470273_hypothetical_protein.

Six genes targeted for mutation were not successfully deleted. These were the mutant strains *Dcri, Dhyp25, DparA, Dmate, Dhyp27*, and *DvirB2.* For each of these mutants, the DNA construct and the vector for mutation were produced, but once transformed into the wild-type strain of *C. jejuni*, the recombinant strain failed to grow on the selective plate.

This can be due to several reasons: the construct may have failed to integrate in the genome and the substitution of the target gene with the kanamycin cassette did not happen; the construct may have integrated in the genome, and the deletion of the target gene happened, but the mutation may have disrupted an essential

gene, and the resulting strain of *C. jejuni* may not be able to grow; or the pTet-like plasmid that may be carried by the strain Cj1 was lost after the mutation of a gene that stopped it from replicating or transmitting to the daughter cells. Regarding the latter case, if the HicA/B toxin-antitoxin system function is post-segregational killing, and if a plasmid is present in the strain Cj1 and the gene *parA* function is regulating its replication, we may speculate that the mutation of *parA* was successful, but the mutant strain lost the plasmid pTet-like, and this event could not be identified because the bacterium is killed as soon as the plasmid is lost either by the HicA/B system or by the presence of Kanamycin in the selective plate. In order to investigate, the first step should be completing the genome of the strain Cj1, in order to understand unequivocally whether it carries a plasmid pTet-like or only a set of genes integrated in the chromosome. If this is the case, then it would be necessary to understand what is the mechanism regulating its replication, including discovering the origin of replication of the plasmid and understanding which proteins regulate its replication. These may include the ParA/B system but also the RepA protein described in the contig of the genome assembly of the strain Cj1 used to design the mutants described in this chapter.

Further investigation is also required to understand whether one or more of the genes targeted for deletion are essential for the growth of *C. jejuni*, in particular the mutants *Dhyp25* and *Dhyp*27, which show no conserved domains, or whether the disruption of one of those genes may have facilitated the curing the plasmid pTet-like.

In the next chapter, I describe the design and the application of the phenotypic tests to the 11 mutants produced.

The deletion of the tetracycline resistance gene *tet(O)* will be used for testing the validity of the gene deletion protocol (as the *DtetO* mutant should be susceptible to tetracycline and have a known and identifiable phenotype) and will be used to assess the ability of the protocol described in section 2.8 to complement the phenotype of the mutant strains.

Chapter 6 - Phenotypic analysis of the *C. jejuni* mutations in pTet-like genes

The previous chapters described the plasmid pan-genome of *C. jejuni,* focusing on the pTet-like plasmid pan-genome. I described 79 genes present in all of the sequenced plasmids of *C. jejuni*, dividing them into 4 clusters defined according to the frequency of each gene in the sequenced strains of *C. jejuni* to date. The genes present in two of these clusters are present in *C. jejuni* with a higher frequency than the others, and I defined the genes belonging to these as “core” genes of the plasmid pan genome. I defined the genes in the clusters associated with lower frequency “accessory” genes of the pTet-like plasmid pan-genome. I described the conserved domains present in the proteins encoded by those 79 genes, and I constructed 11 mutants in core and accessory genes.

In this chapter, I will describe the application of a series of phenotypic assays to these mutants, and I will compare their phenotype with the wild-type *C. jejuni* strain Cj1.

For the mutants that showed a phenotype different from the wild type, I complemented the mutation (Jervis *et al.*, 2015), integrating the wild type gene into the chromosome of the mutant and under the control of a constitutive promoter.

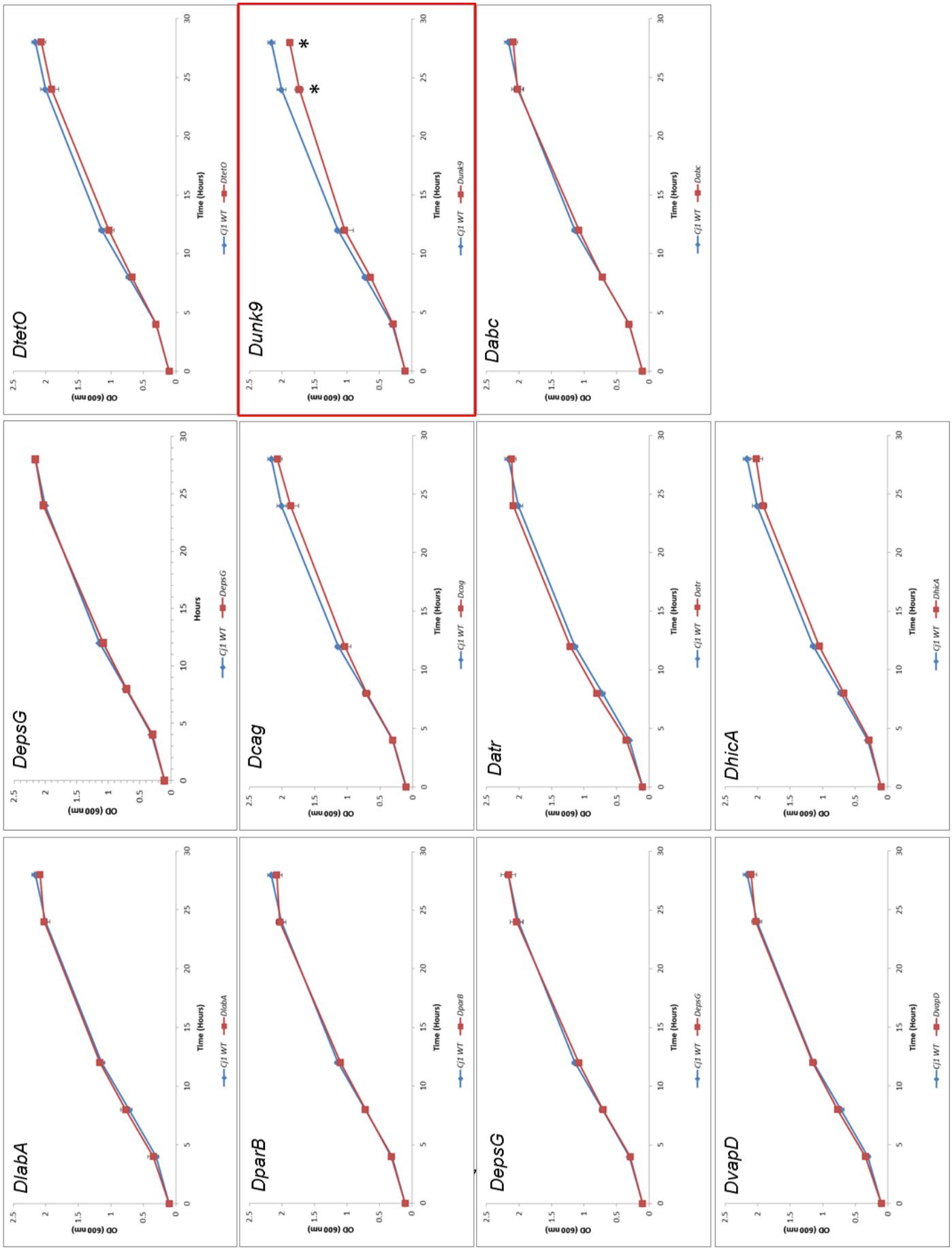
This chapter identifies previously undescribed functions for genes associated with the pTet-like plasmid.

6.1 Growth rate in MH

The growth rates of the mutants in MH broth were first measured. A 25 ml liquid culture was set up for each strain, with a starting OD (600nm) value of 0.1. The culture was incubated for 28 hours in microaerophilic conditions, and the OD_{600} value was read at intervals (Figure 6.1). The full protocol is described in Chapter 2.7.1.

With this test a difference between the WT and the mutant strain *Dunk9 was* identified, in particular the strain *Dunk9* grew significantly ($p<0.01$) more slowly than the strain Cj1.

Figure 6.1 - Growth rate of each mutant compared to the WT strain Cj1 in 25ml of MH broth measured for 28 hours. Each point is the average of three biological replicates; error bars show the SD. The growth rate of each mutant is shown in red the growth rate of the wild type is shown in blue. Each graph is annotated with the name of each mutant. The graph associated with the mutant *Dunk9* is highlighted in red. * = statistically significant difference. p<0.01 T-test, single tailed, paired.



6.2 Motility

To assess the motility of the different mutant strains of *C. jejuni,* I used a variation of the method described by Reuter *et al.,* in 2013. The full protocol is described in Chapter 2.7.5.

Briefly, 5 μl of a *C. jejuni* suspension was spotted on plates containing MHA with 0.4% agar. The MHA plates contained either a full dose or a half dose of nutrient. Each plate was photographed and the area covered by the bacterial movement was measured with ImageJ software (https://imagej.nih.gov/ij/index.html). Figures 6.2 and 6.3 show examples of the plates after 48 hours on MH+0.4% agar plates and after 72 hours on MH (half nutrient dose) +0.4% agar plates. Figures 6.4 to 6.8 shows the percentage of spreading of the colonies compared to the wildtype.

There was a statistically significant difference ($p<0.01$) between the mutant *Dunk9* and the wild type Cj1 strain in the conditions of MH with full nutrients, and between the mutant *DhicA* and the wild type strain Cj1, which moves at a slower rate than the wild type both in condition of full nutrients and half nutrients.

Figure 6.2 - Example of six motility plates (MH, 0.4% agar). On each plate strain Cj1 WT and one or two mutant strains were spotted. Plates were incubated for 48 hours in microaerophilic conditions at 37℃.

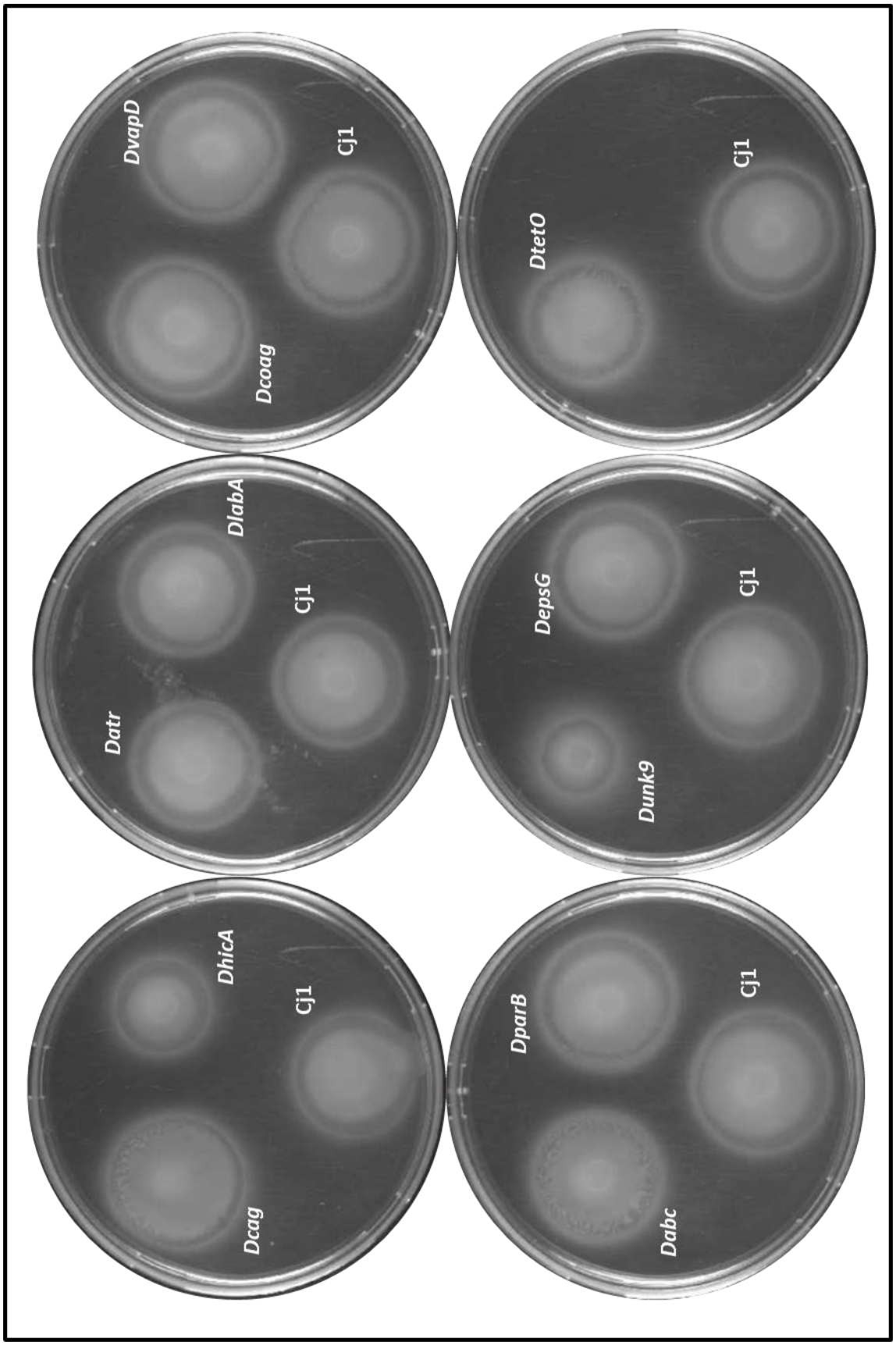


Figure 6.3 - Example of six motility plates (MH half nutrients concentration, 0.4% agar). On each plate strain Cj1 WT and one or two mutant strains were spotted. Plates were incubated for 72 hours in microaerophilic conditions at 37℃.

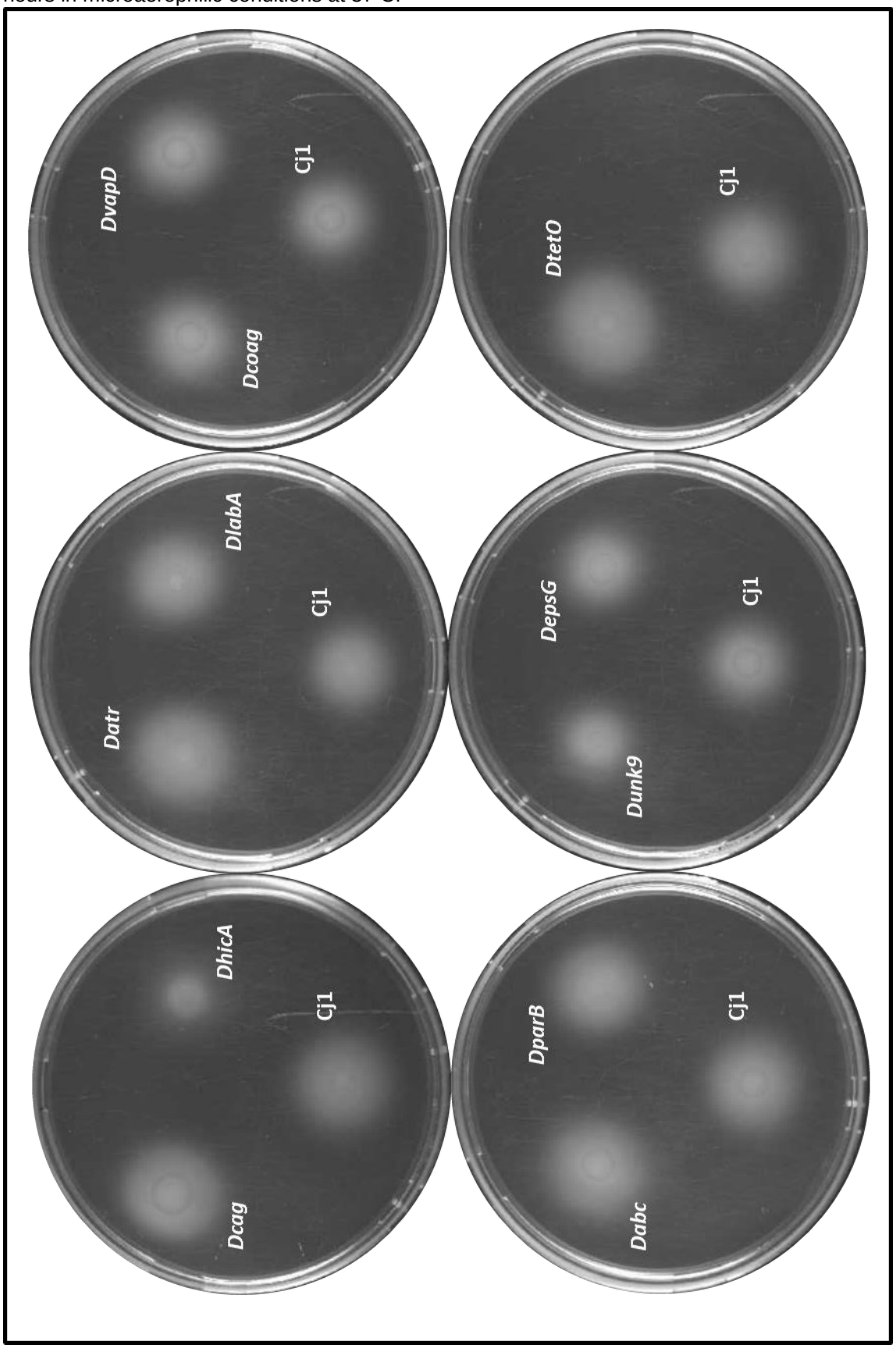


Figure 6.4 - Colony diameter after 24 hours of incubation in microaerophilic conditions at 37℃ of MH plates 0.4% agar. Each measurement is the average of at least three biological replicates. The error bars represent the SD. * = Significantly different from Cj1 WT. p<0.01 One-way ANOVA, Tukey's post-hoc.

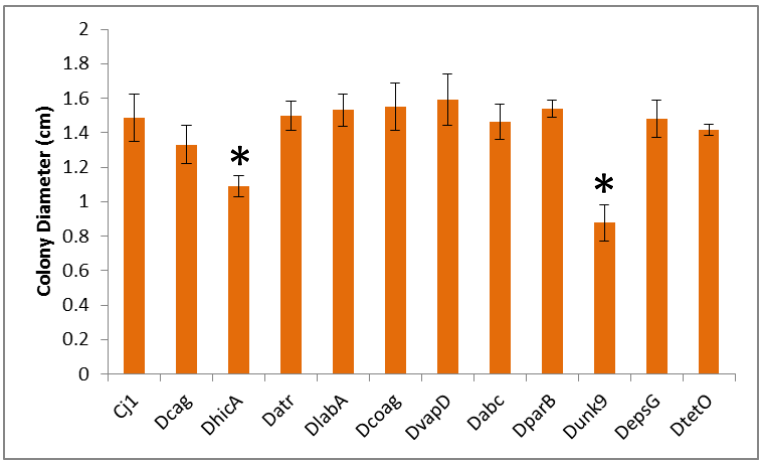


Figure 6.5 - Colony diameter after 48 hours of incubation in microaerophilic conditions at 37℃ of MH plates 0.4% agar. Each measurement is the average of at least three biological replicates. The error bars represent the SD. * = Significantly different from Cj1 WT. p<0.01 One-way ANOVA, Tukey's post-hoc.

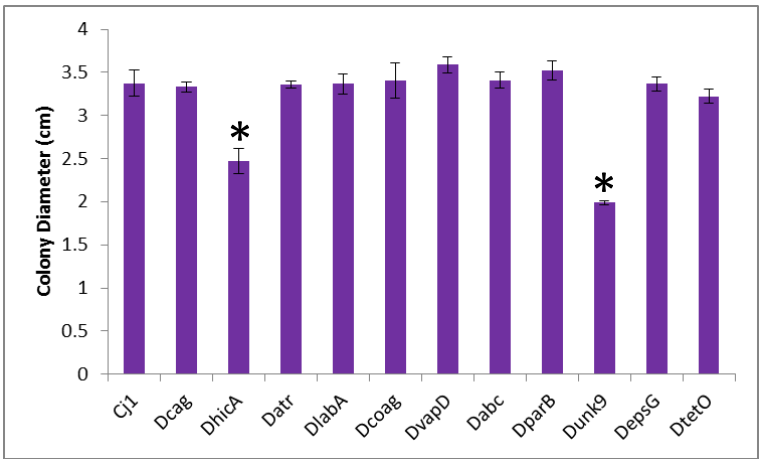


Figure 6.6 - Colony diameter after 48 hours of incubation in microaerophilic conditions at 37℃ of MH plates, with half the concentration of nutrients, 0.4% agar. Each measurement is the average of at least three biological replicates. The error bars represent the SD. * = Significantly different from Cj1 WT. p<0.01 One-way ANOVA, Tukey's post-hoc.

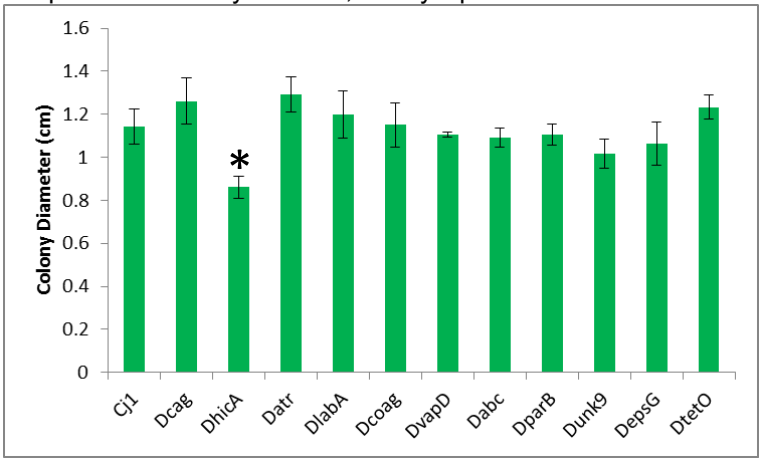


Figure 6.7 - Colony diameter after 72 hours of incubation in microaerophilic conditions at 37℃ of MH plates, with half the concentration of nutrients, 0.4% agar. Each measurement is the average of at least three biological replicates. The error bars represent the SD. * = Significantly different from Cj1 WT. p<0.01 One-way ANOVA, Tukey's post-hoc.

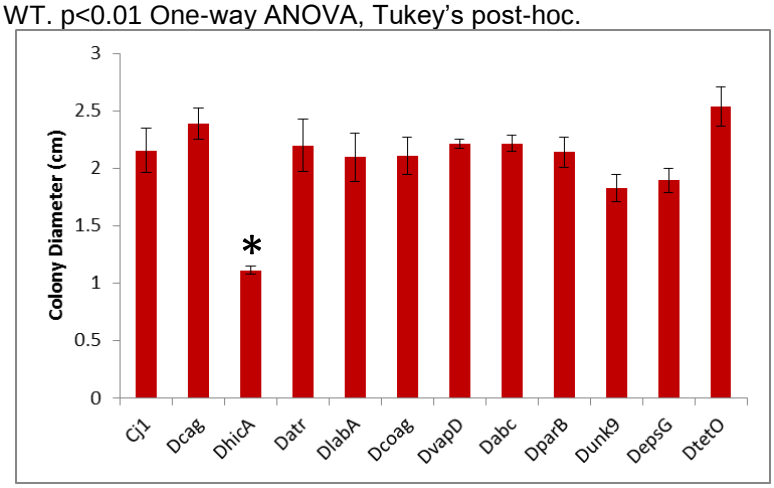
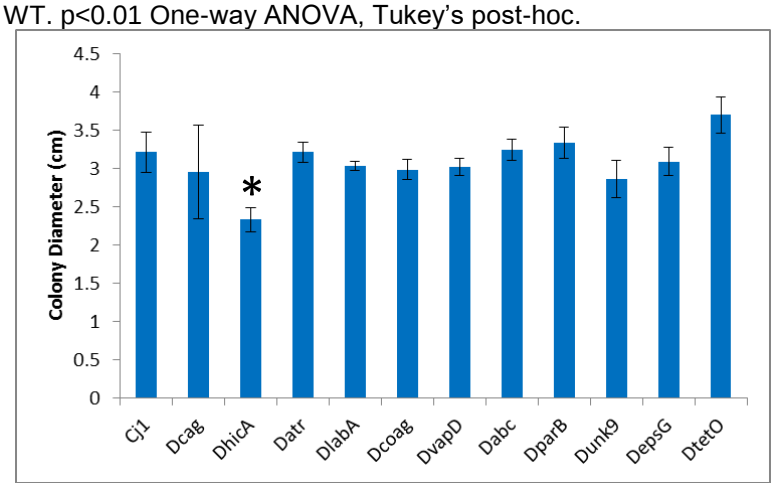


Figure 6.8 - Colony diameter after 96 hours of incubation in microaerophilic conditions at 37℃ of MH plates, with half the concentration of nutrients, 0.4% agar. Each measurement is the average of at least three biological replicates. The error bars represent the SD. * = Significantly different from Cj1 WT. p<0.01 One-way ANOVA, Tukey's post-hoc.



6.3 Biofilm production

6.3.1 Method design

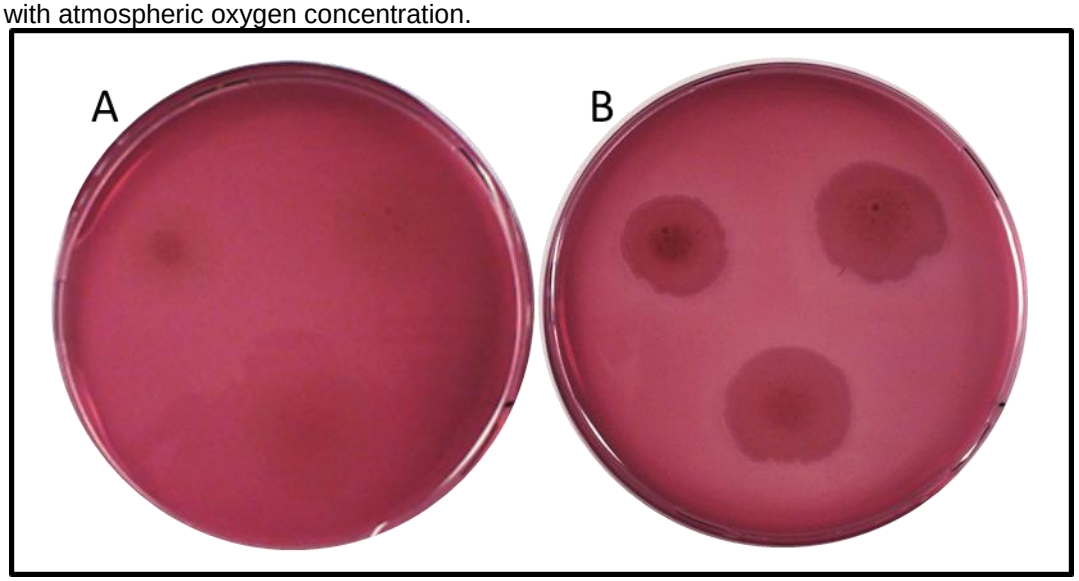
The test described by Oh *et al.* in 2014 was first used to measure biofilm production. The test is based on the incubation of a bacterial suspension in MH in a 96-well plate for 24 hours. After this time, the plate is rinsed with water and stained with a crystal violet solution, and de-stained with ethanol. Biofilm is quantified by reading the absorbance at 595 nm. This test was characterised by a high degree of variability, and I could not repeat the results between experiments (data not shown).

Therefore, I developed a test on agar plates that provides more repeatable results. This test is based on the use of Congo-red, which stains b-(1-*4)-linked D-glucopyranosyl units, as described in Theather and Wood in 1982, and Coomassie blue. Although the main component identified in *C. jejuni* biofilm is reported to be α-dextran (Jowila *et al.*, 2015), the Congo-red molecule have been successfully used to stain the biofilm of *C. jejuni* (Reuter *et al*, 2010): this study showed how the results obtained with Congo red staining were confirming the observations made with Crystal violet and concluded that *C. jejuni* biofilms bind Congo red.

After an optimisation of the duration of the test, temperatures, and concentration of agarose in the plates, the test involved a growing step, in which the plates were spotted with of 5 μl of bacterial suspension (1.0 OD_{600}) and incubated for 48 hours in microaerophilic conditions, and a biofilm development step, during which the plates were incubated in normal aerobic conditions (i.e. atmospheric) for 48 hours at 37 °C. As reported by Reuter *et al*. in 2010, *Campylobacter* shows an increased ability to produce biofilm when incubated in aerobic conditions, and this incubation was necessary to identify the biofilm production in this assay. During the growing phase (Figure 6.9 A), bacterial colonies do not develop an identifiable

colour. During the biofilm development phase, the colonies develop the colour shown in Figure 6.9 B in case of ability of producing biofilm. The final protocol adopted for this test is described in Chapter 2.6.7.

Figure 6.9 - Biofilm formation in MH plate (0.8% agar) supplemented with Congo Red and Coomassie Brilliant Blue (MHA-Congo red). The plate was spotted three times with 5 µl of Cj1 WT bacterial suspension (PBS, OD_{600}=1.0). (A) MHA-Congo red plate after 48 hours incubation at 37℃ in microaerophilic conditions. (B) MHA-Congo red plate from (A) after further 48 hours of incubation at 37℃ with atmospheric oxygen concentration.



6.3.2 Biofilm production of the mutant strains and persister cells formation

The biofilm production and the ability for the bacterium to produce persister cells are linked in several bacterial species (Wang and Wood, 2011, Wood *et al.,* 2013); I therefore decided to measure the ability to develop persister cells in the strains showing impairment in biofilm formation. Although the association between biofilm and persister cells was never shown in *C. jejuni,* previous unpublished work by Champion *et al.* had shown persisters cells features in the strain 11168 of *C. jejuni*. That data shows how the exposure of strain 11168 to up to 100 MIC of Penicillin and Ciprofloxacin for up to 24 hours does not kill the entire bacterial population and that the surviving cells remain susceptible to the drug when re-cultured.

The protocol applied here to identify the presister cells formation in *C. jejuni* was developed in the University of Exeter's Molecular Pathogenesis Laboratory (Olivia Champion, personal communication).

6.3.2.1 Biofilm production

10 of the 11 mutant strains showed the same phenotype of the wild-type strain Cj1 of *C. jejuni*, with colonies that diffuse in the MH plate 0.8% agar stained in pink/red after 2 days of incubation in microaerophilic conditions at 37 °C, followed by 2 days of incubation in normal atmospheric condition at 37 °C.

The *DhicA* mutant showed a different phenotype, with bacterial colonies that were not stained after the same incubation periods. A picture of the plates is shown in Figure 6.10 and a Table 6.1 summarises the results.

6.3.2.2 Persister cells formation in the mutant *DhicA*

The mutant *DhicA* was assayed for the ability to form persister cells after 24 hours of incubation in presence of concentrations of penicillin 10X, 50X, and 100X higher than the MIC (256 µg/ml) and concentrations of ciprofloxacin 10X and 20X higher than the MIC (32 µg/ml).

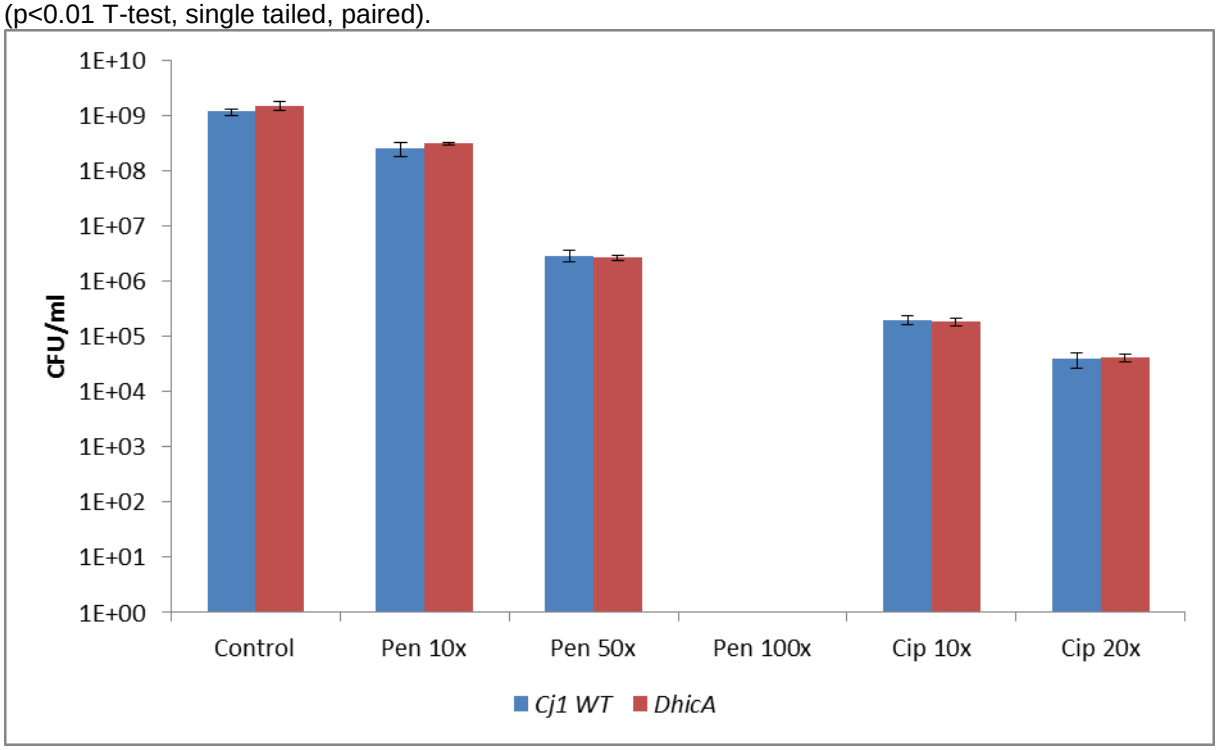
The results of this assay are shown in Figure 6.11: in each condition tested, the number of CFUs measured in the mutant strain is not statistically different from the number of CFUs measured in the WT Cj1 strain ($p<0.01$).

Table 6.1 - Biofilm production in the 11 mutant strains and in the Cj1 WT strain, evaluated with the MH-Congo red plate method. +/- = strain is able or unable to produce biofilm, as identified by the colony staining in at least three biological replicates.

Strains	Congo Red staining
Cj1 WT	+
DhicA	-
DlabA	+
DvapD	+
Dcoag	+
DparB	+
Dabc	+
Dcag	+
Dunk	+
Datr	+
DtetO	+
DepsG	+

Figure 6.10 - Example of six MH-Congo red plates for the identification of *C. jejuni* strains able to produce biofilm. On each plate strain Cj1 WT and one or two mutant strains were spotted. Plates were incubated for 48 hours in microaerophilic conditions at 37℃, followed by 48 hours with normal atmospheric concentration at 37℃.

Figure 6.11 - Persister cells formation after 24 hours incubation in strain Cj1 WT and strain *DhicA*. Control = MH broth; Pen 10x = MH + penicillin 2.56 mg/ml; Pen 50x = MH + penicillin 12.8 mg/ml; Pen 100x = MH + penicillin 25.6 mg/ml; Cip 10x = MH + Ciprofloxacin 320 µg/ml; Cip 20x = MH + Ciprofloxacin 640 µg/ml. Each measurement is the average of three biological replicates. The error bars represent the SD. The mean of each pair (Cj1 WT vs. *DhicA*) is not significantly different ($p<0.01$ T-test, single tailed, paired).

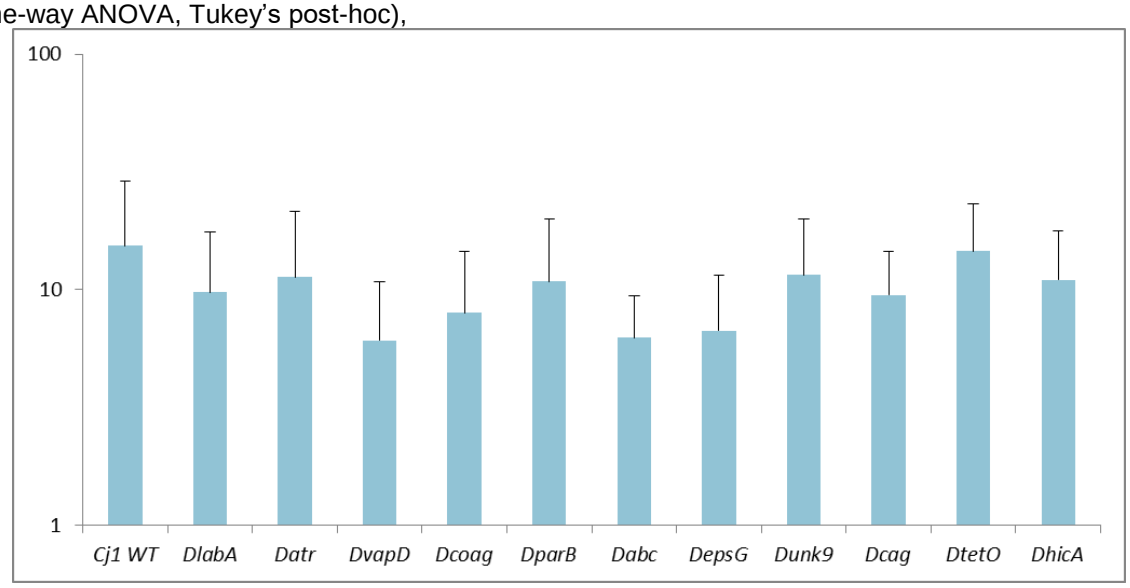


6.4 Survival in water

The ability to survive in tap water was tested as described in Chapter 2.7.8. Briefly, bacteria were grown on CBA+ plate, re-suspended and washed twice in PBS buffer. Bacterial were finally re-suspended in 10 ml of filter sterilised tap water, adjusting the bacterial concentration to 0.5 OD_{600}. At the beginning of the incubation and after 8 days of incubation at 4 °C, 100 µL of bacterial suspension was serially diluted and bacteria enumerated on agar.

The results are shown in Figure 6.12. None of the mutants showed a significant difference ($p<0.01$) in survival compared with the wild type.

Figure 6.12 - Percent of bacterial survival after 8 days of incubation in water of each mutant strain and the strain Cj1 WT. Each measurement is the average of three biological replicates. The error bars represent the SD. Each mean is not significantly different from the strain Cj1 WT ($p<0.01$, One-way ANOVA, Tukey's post-hoc),



6.5 Antibiotic resistance

Antibiotic resistance in the mutant strains was evaluated via MIC quantification on microtiter plates, except for the tetracycline resistance in *DtetO*, for which an e-strip test (Bio-Merieux) was used. The full protocols are described in Chapter 2.7.2. As expected, the mutant *DtetO* showed susceptibility to tetracycline, with an MIC >0.048 μg/ml, compared to 48-32 μg/ml in strain Cj1.

The MIC values for the antibiotics erythromycin, ciprofloxacin, penicillin, and ceftazidime are reported in Table 6.2. None of the mutants showed a significant difference ($p<0.01$) in MIC compared with the wild type.

Table 6.2 - Minimum inhibitory concentrations (MICs) for growth in Erythromycin, Penicillin, Ciprofloxacin, and Ceftazidime in each mutant strain and in the strain Cj1 WT. The values were measured with the plate dilution method in three biological replicates; values are shown in µg/ml.

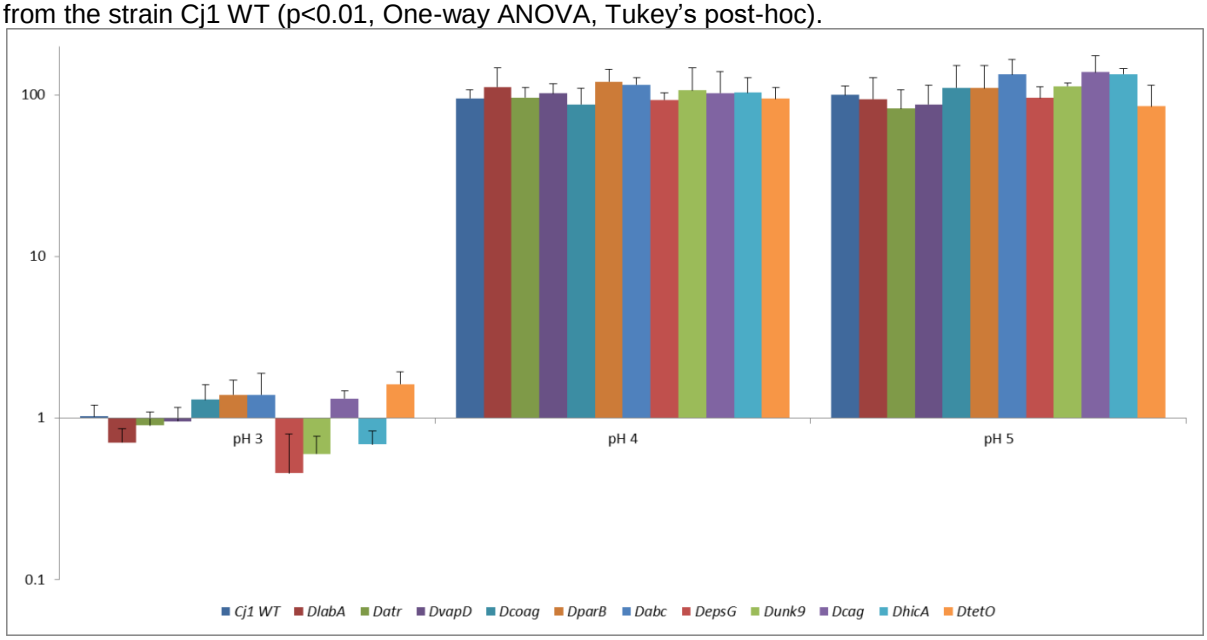
Strain	Erythromycin MIC	Penicillin MIC	Cyprofloxacin MIC	Ceftazidine MIC
Cj1 WT	>0.5	256-128	32-16	32-16
Dabc	>0.5	256-128	32-16	32-16
DepsG	>0.5	256-128	32-16	32-16
Dcag	>0.5	256-128	32-16	32-16
DhicA	>0.5	256-128	32-16	32-16
DtetO	>0.5	256-128	32-16	32-16
DlabA	>0.5	256-128	32-16	32-16
Datr	>0.5	256-128	32-16	32-16
DvapD	>0.5	256-128	32-16	32-16
Dcoag	>0.5	256-128	32-16	32-16
DparB	>0.5	256-128	32-16	32-16
DhicA	>0.5	256-128	32-16	32-16

6.6 Acid tolerance

Acid tolerance was evaluated using a modification of the method described by Reid *et al.,* in 2008 (1). Briefly, MH medium was adjusted to pH 3, 4 and 5 using HCl. Bacteria grown on CBA plates overnight were used to produce a 0.5 OD_{600} bacterial suspension in MH. One ml of culture was centrifuged and re-suspended in MH adjusted to different pH values. The tubes were incubated under microaerophilic conditions for 20 minutes and surviving bacteria enumerated on agar. The full protocol is described in Chapter 2.7.3.

The results are showed in Figure 6.13, expressed as percentage of bacteria surviving the treatment. None of the mutants showed a significant difference ($p<0.01$) in acid tolerance compared with the wild type.

Figure 6.13 - Percent of bacterial survival after 20 minutes of incubation in MH buffered at pH 3, pH 4 or pH 5 of each mutant strain and the strain Cj1 WT. Each measurement is the average of three biological replicates. The error bars represent the SD. Each mean is not significantly different from the strain Cj1 WT ($p<0.01$, One-way ANOVA, Tukey's post-hoc).



6.7 Tolerance to synthetic gastric fluid

The tolerance to synthetic gastric fluid buffered to pH 4.5 (SGF for the simulation of stomach environment, Beumer *et al.,* 1992; Reid *et al.,* in 2008 (2)) was evaluated using a similar method to the one described in previous section but using synthetic gastric fluid in place of MH medium. The ability to tolerate incubation in SGF has been evaluated for 6 hours, preparing 10 ml 0.5 OD_{600} bacterial suspension and enumerating the number of surviving bacteria after 20 minutes, 3 hours, and 6 hours. The full protocol is described in Chapter 2.7.4.

The results are reported in Figure 6.14 and Figure 6.15. None of the mutants showed a significant difference ($p<0.01$) in synthetic gastric fluid tolerance compared with the wild type. After 6 hours of incubation in SGF no bacteria grew after plating samples.

Figure 6.14 - Percent of bacterial survival after 20 minutes of incubation in synthetic gastric fluid (SGF) buffered at pH 4.5 of each mutant strain and the strain Cj1 WT. Each measurement is the average of three biological replicates. The error bars represent the SD. Each mean is not significantly different from the strain Cj1 WT ($p<0.01$, One-way ANOVA, Tukey's post-hoc).

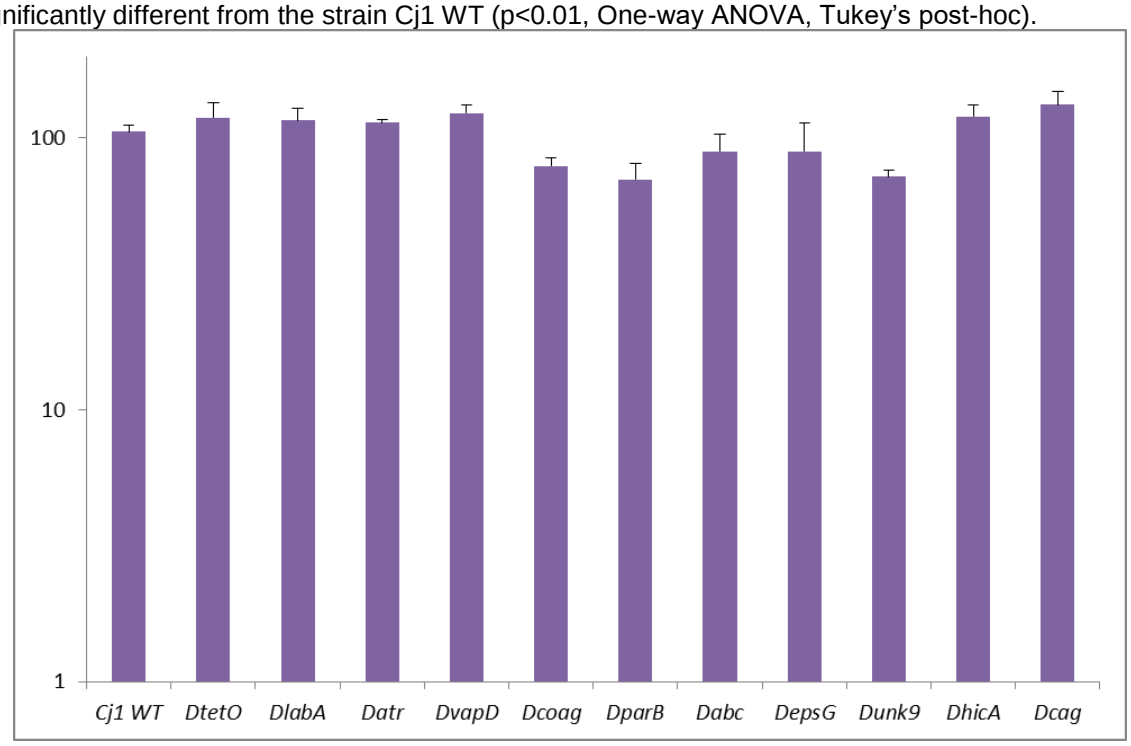
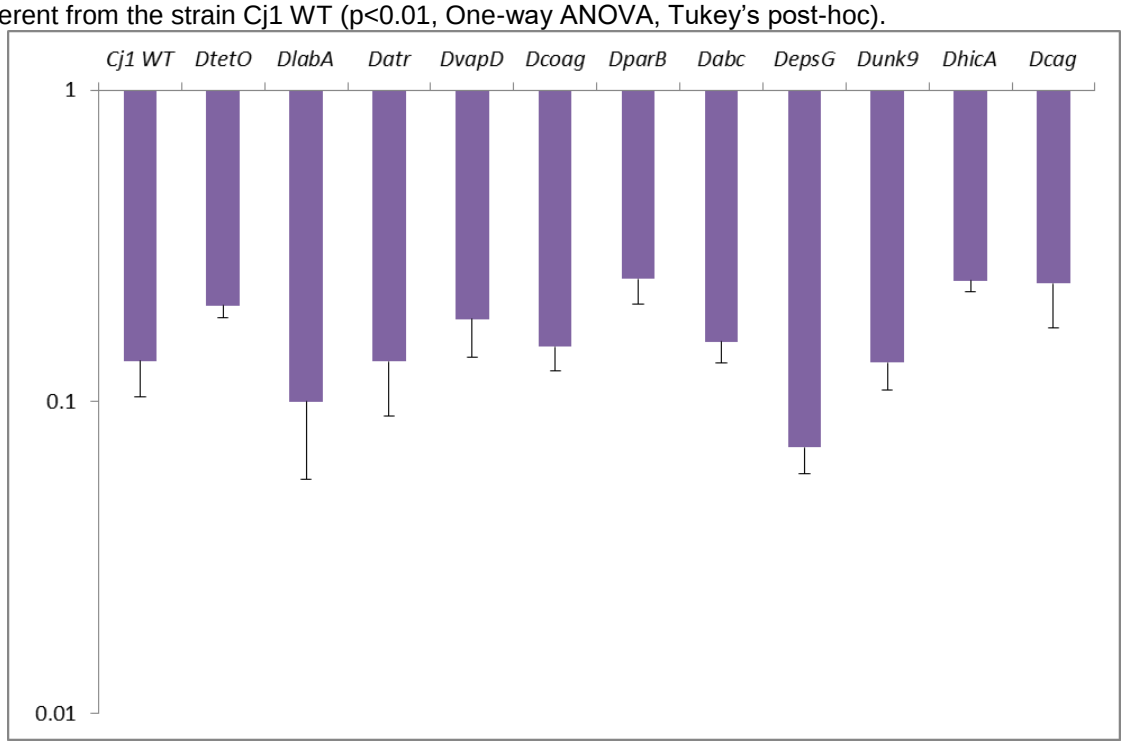


Figure 6.15 - Percent of bacterial survival after 3 hours of incubation in synthetic gastric fluid (SGF) buffered at pH 4.5 of each mutant strain and the strain Cj1 WT. Each measurement is the average of three biological replicates. The error bars represent the SD. Each mean is not significantly different from the strain Cj1 WT ($p<0.01$, One-way ANOVA, Tukey's post-hoc).



6.8 *Galleria mellonella* larvae infection

6.8.1 Method design

In order to evaluate the ability of killing the model organism *Galleria mellonella* larvae, bacteria were injected into the first right foreleg of the larva.

As with other studies (e.g. Elmi *et al.,* 2012), I was not able to obtain the reproducible killing of larvae incubated under aerobic conditions. I optimised the technique, incubating the larvae post-injection in different concentrations of oxygen, which promoted the growth of *Campylobacter*, and injecting different bacterial concentrations.

As described in Figure 6.16, when the larvae were injected with 10^4 to 10^7 bacterial cells and incubated for 48 hours at 37 °C in normal atmospheric conditions, the maximum killing registered was 20% of larvae. If the larvae were injected with the same doses of bacteria and then incubated for 48 hours at 37 °C in microaerophilic conditions, 90-100% of larvae were killed after challenge with 10^6 or 10^7 bacterial cells (Figure 6.17). There was no killing of control larvae injected with PBS or in not-injected.

Therefore, the virulence of the mutants compared to the wild type was evaluated injecting 10^7 bacterial cells and incubating the larvae for 48 hours in microaerophilic conditions. The full protocol is described in Chapter 2.7.9.

Figure 6.16 - Percent survival of *G. mellonella* larvae injected with 10^7, 10^6, 10^5, 10^4 bacterial cells (*C. jejuni* strain Cj1 WT), PBS or not injected (Non-stabbed control NSC). Larvae were incubated in normal atmospheric conditions at 37℃ for 48 hours post inoculation. Each measurement is the result of the injection of 10 larvae.

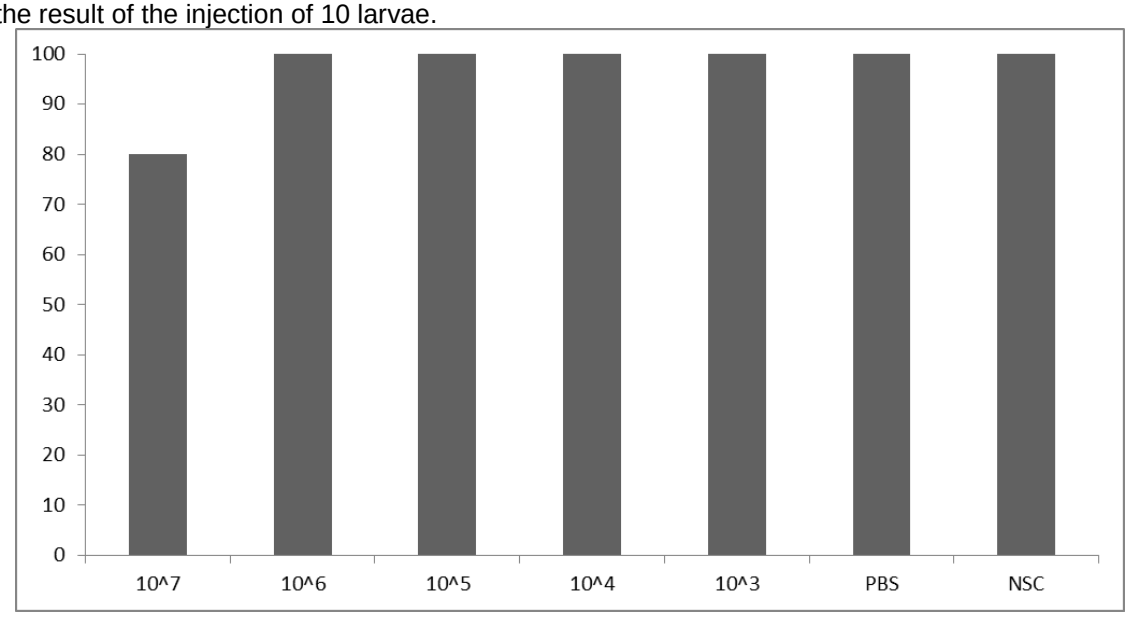
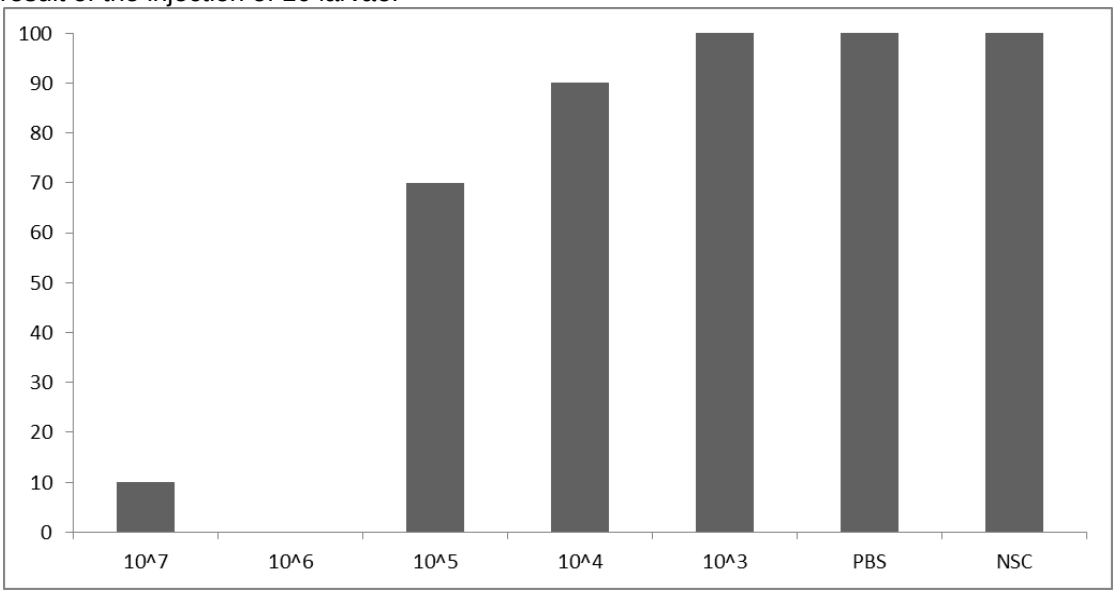


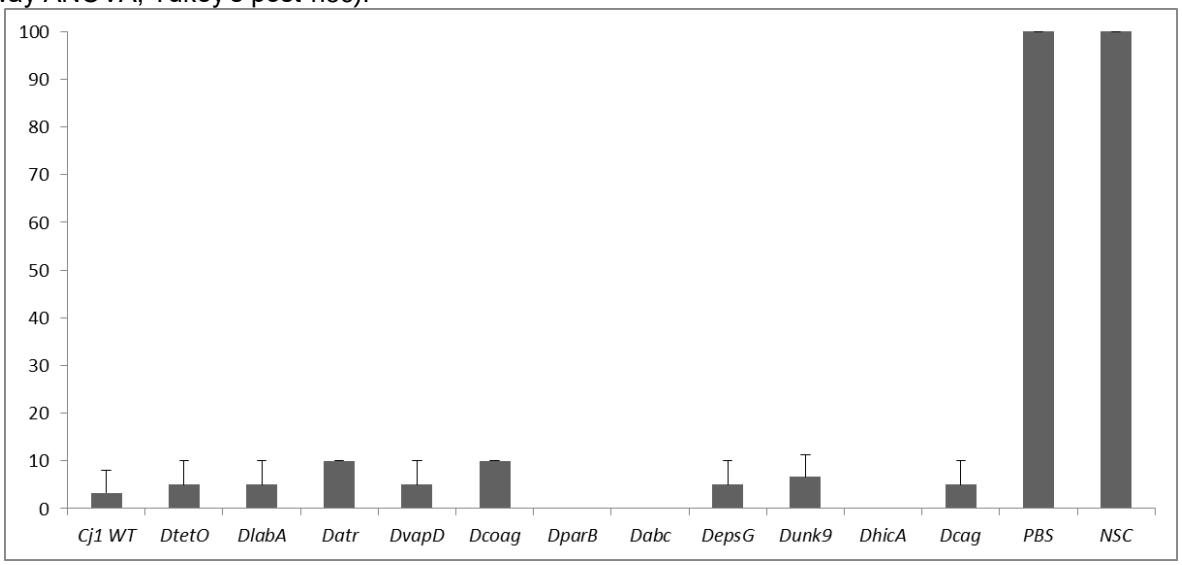
Figure 6.17 - Percent survival of *G. mellonella* larvae injected with 10^7, 10^6, 10^5, 10^4 bacterial cells (*C. jejuni* strain Cj1 WT), PBS or not injected (Non-stabbed control NSC). Larvae were incubated in microaerophilic conditions at 37℃ for 48 hours post inoculation. Each measurement is the result of the injection of 10 larvae.



6.8.2 Infection of larvae

The results of the injection of at least two groups of 10 larvae per mutant are shown in Figure 6.18. The controls (PBS injection and non-injected control) are shown in the same graphs, showing 100% survival in every repetition of the experiment. None of the mutants showed a significant difference ($p<0.01$) in killing of *G. mellonella* larvae compared with the wild type.

Figure 6.8 - Percent survival of *G. mellonella* larvae injected with 10^7 bacterial cells of each mutant strain, strain Cj1 WT, PBS or not injected (Non-stabbed control NSC). Larvae were incubated in microaerophilic conditions at 37℃ for 48 hours post inoculation. Each measurement is the average of three independent experiments; in each experiment 10 larvae were injected. The error bars represent the SD. Each mean is not significantly different from the strain Cj1 WT ($p<0.01$, One-way ANOVA, Tukey's post-hoc).



6.9 Complementation of the mutant strains

Three mutants exhibited a different phenotype to the wild type in some of the experiments described above. The mutant *Dunk9* grew significantly slower than the wild type in MH in microaerophilic conditions and reached a lower final OD_{600} (Figure 6.1); the same mutant *Dunk9* was impaired in motility (Figures 6.2.3 and 6.2.4). The *DhicA* mutant showed an impaired ability to produce a biofilm (Figure 6.10 and Table 6.1) and motility in particular nutrient and agarose concentrations (Figures 6.4 to 6.8). Finally, the *DtetO* mutant showed increased susceptibility to tetracycline, compared to the wild type. Although this phenotype of the *DtetO* mutant is not surprising given that *tet(O)* is known to be required for tetracycline resistance (Mazi *et al.,* 2008), it was useful to confirm the success of the mutagenesis protocol and will be used to confirm the success in the complementation of the phenotypes.

The strategy used here to complement the mutant strains was described by Jervis *et al*. in 2015 and detailed in Chapter 2.8.

Briefly, the three mutants were complemented by amplifying the gene of interest (*hicA, tetO*, or *unk9*) with the primers in Table 6.3. Each primer pair introduced a restriction site for the enzymes *NcoI* and *NheI* in the 5’ and the 3’ end of the PCR product respectively, which were then used to clone the gene in the complementation vector. The vectors built for the complementation of the mutants are reported in Table 6.4.

The correct insertion of each wild-type gene was checked by PCR, using primers annealing outside the integration site. The PCR product confirming the integration of the gene of interest is shown in Figure 6.19. Sanger sequencing confirmed the correct integration in each of the complemented mutants.

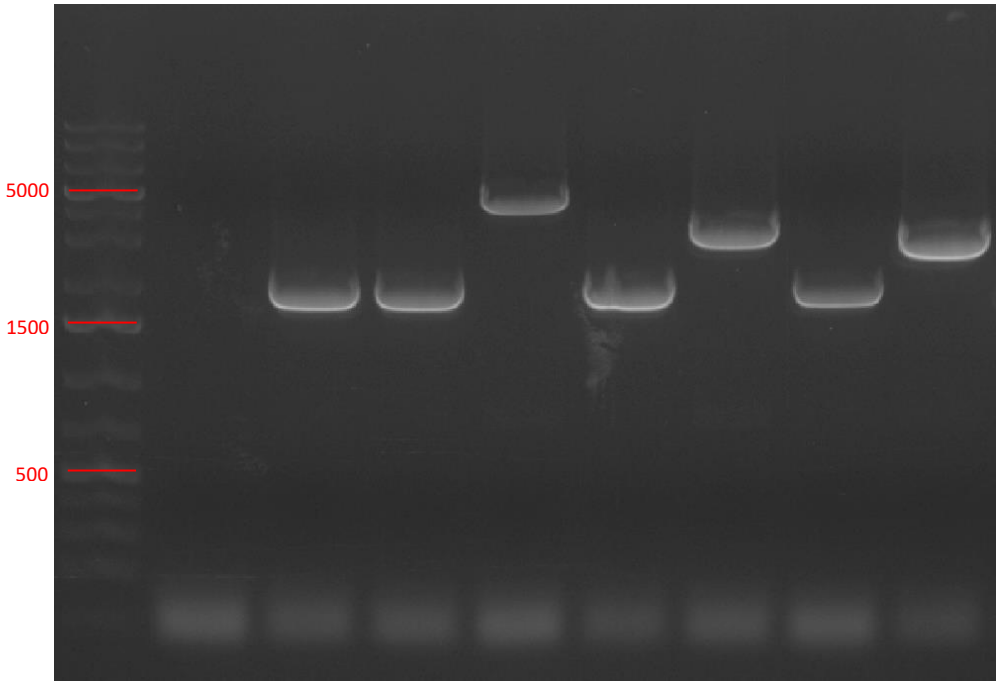
Table 6.3 - PCR primers designed for the complementation of mutants *DtetO*, *DhicA* and *Dunk9*. Columns show (from left to right): name of the complemented mutant, name of each primer, sequence of each primer, length of the amplified product.

Mutant Name	Primer Name	Primer Sequence	PCR product Length
Dunk9+unk9	unk9_complF	CATGCCATGGATGAGTGATAAAGAATTAGAA	576
	unk9_complR	CTAGCTAGCTCATCTTTTAAACTCCTTATT	
DhicA+hicA	hicA_complF	CATGCCATGGATGCCTGAATTACCAAGATTG	204
	hicA_complR	CTAGCTAGCTTAATCCTTTGCTACTTCTATA	
DtetO+tetO	tetO_complF	CATGCCATGGATGAAAATAATTAACTTAGGC	1920
	tetO_complR	CTAGCTAGCTTAAGCTAACTTGTGGAACAT	

Table 6.4 - Vectors designed for the complementation of mutants *DtetO*, *DhicA* and *Dunk9.* Columns shows (from left to right): plasmid name, length of the plasmid, brief description of the plasmid vector, markers present on each plasmid, reference (where available).

Plasmid Name	Length	Description	Markers	
pCJC1	5775	Vector designed for recombination-mediated delivery of genes onto the C. jejuni chromosome.	cat, ampR	Jervis et al., 2015
pCJC1::*tetO*	7695	pCJC1 for the complementation of the mutant *DtetO*	cat, ampR	This study
pCJC1::*unk9*	6342	pCJC1 for the complementation of the mutant *Dunk9*	cat, ampR	"
pCJC1::*hicA*	5979	pCJC1 for the complementation of the mutant *DhicA*	cat, ampR	"

Figure 6.19 - PCR for testing the successful complementation of the mutant strains *DtetO*, *Dunk9* and *DhicA*. Lane 1, 1kb plus ladder (Thermo scientific); lane 2, NTC; lane 3, Cj1 WT; lane 4, *DtetO*; lane 5, *DtetO+tetO* complemented strain; lane 6, *Dunk9*; lane 7, *Dunk9+unk9* complemented strain; lane 8, *DhicA*; lane 9, *DhicA+hicA* complemented strain.



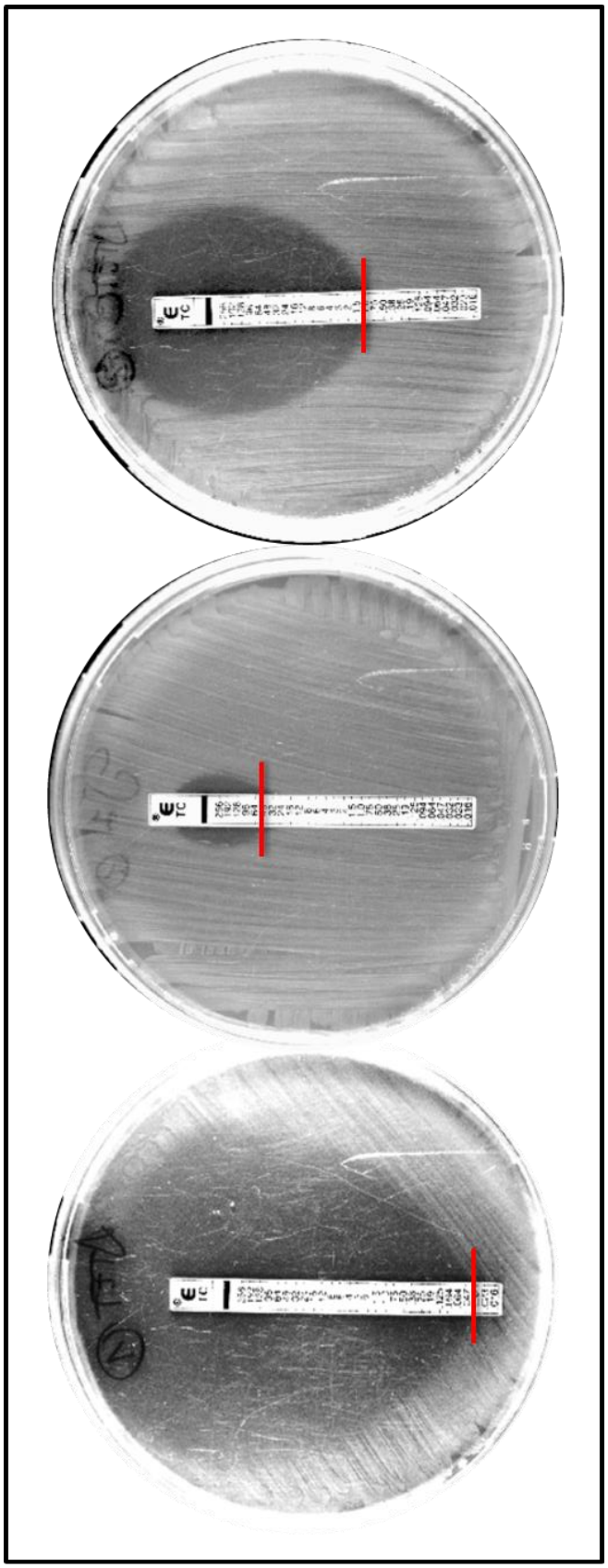
6.10 Phenotype of the complemented mutants

6.10.1 Complementation of the *tetO* mutant

The resistance to tetracycline was partially restored in the complemented *tetO* mutant. Figure 6.20 shows the comparison between the strains Cj1, *DtetO,* and *DtetO+tetO*.

The MIC of the complemented mutant *DtetO+tetO* was 1.0-1.5 µg/ml. In contrast the MIC was 32-48 µg/ml for the wild type strain and >0.048 µg/ml for the *DtetO* mutant.

Figure 6.20 - Measurement of MIC of tetracycline for strains *DtetO*, Cj1 WT and *DtetO+tetO* (from bottom to top) using Estrips (Biomerieux).



6.10.2 Complementation of the *hicA* mutant

It was not possible to restore the wild type phenotype to the *DhicA* mutant by complementation. The phenotype of the mutant and the complemented mutant for motility and biofilm assays were not statistically different ($p<0.01$), as shown in Figure 6.21 and 6.22.

A second type of complementation was next attempted by cloning the full *hicAB* operon into the pCJC1 plasmid. This experiment was performed in order to test whether both genes of the *hicAB* operon need to be under the control of the same promoter to complement the mutation. The *DhicA+hicAB* showed the same phenotype as the *DhicA* mutant, as indicated in Figures 6.23 and 6.24.

Finally, in order to ascertain that during the deletion of the gene *hicA* no other mutations were introduced, the *DhicA* mutant was genome sequenced (MiSeq, PE library 300+300, Illumina Inc.). The alignment between the genome sequence of the wild type strain Cj1 and *DhicA* is shown in Figure 6.25. No other mutation was identified in the genome of the mutant strain.

Figure 6.21 - Motility of strains Cj1, *DhicA*, and *DhicA+hicA* in MH 0.4% agar (half nutrients). (A) A plate used to assess the mobility. (B) Average colony diameter and SD (three biological replicates, * = significantly different from Cj1 WT. p<0.01 One-way ANOVA, Tukey's post-hoc).

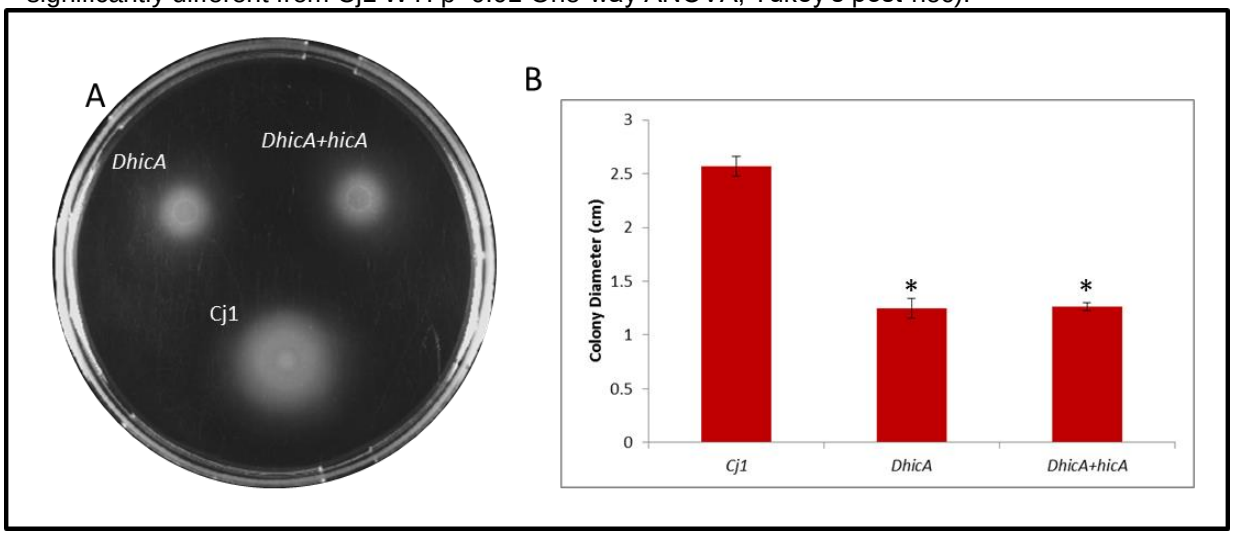
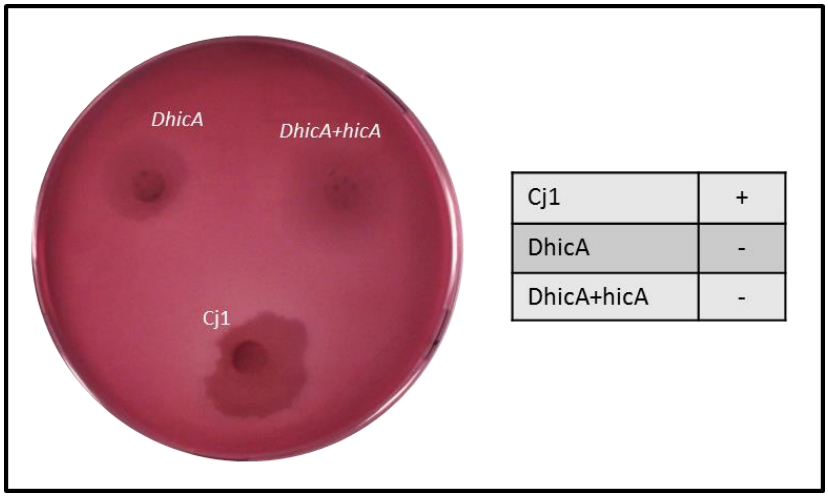


Figure 6.22 - Biofilm formation of strains Cj1, *DhicA* and *DhicA+hicA* in MH-Congo red 0.8% agar.



Cj1	+
DhicA	-
DhicA+hicA	-

Figure 6.23 - Motility of strains Cj1, *DhicA*, and *DhicA+hicAB* in MH 0.4% agar (half nutrients). (A) A plate used to assess the mobility. (B) Average colony diameter and SD (three biological replicates, * = significantly different from Cj1 WT. p<0.01 One-way ANOVA, Tukey's post-hoc).

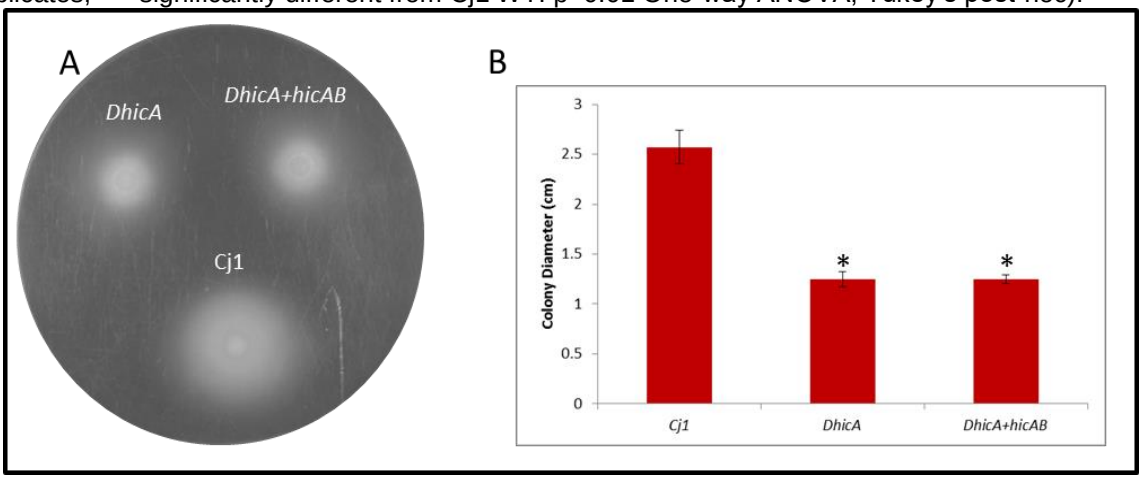
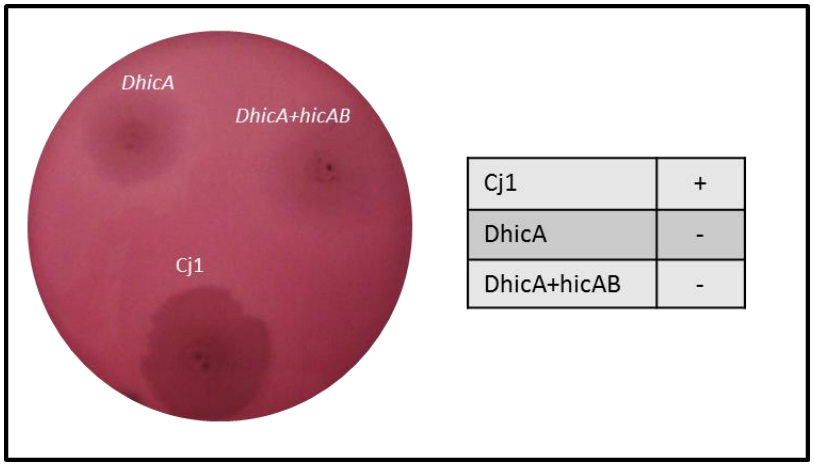
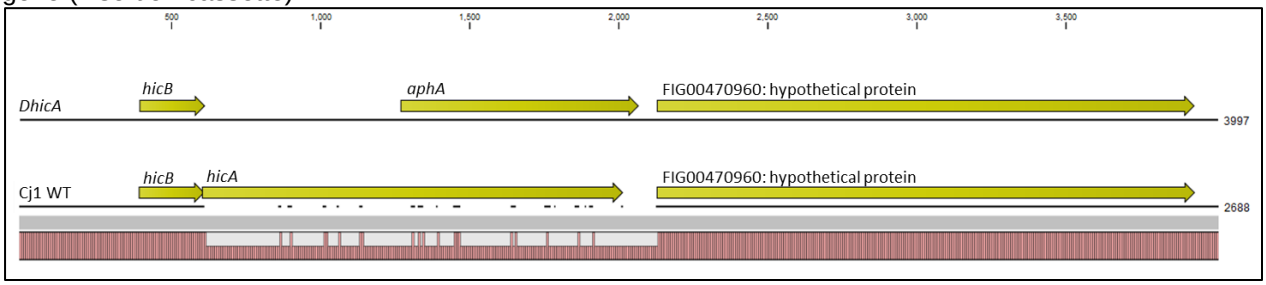


Figure 6.24 - Biofilm formation of strains Cj1, *DhicA* and *DhicA+hicAB* in MH-Congo red 0.8% agar.



Cj1	+
DhicA	-
DhicA+hicAB	-

Figure 6.25 - Sequence alignment of the region affected by the mutation in the strain Cj1 WT and *DhicA*. Sequence alignment was obtained with MUSCLE software (*Edgar,* 2004) and visualised with CLC Genome Wiewer (Qiagen Inc.). Full pink bars at the bottom of the alignment indicate a perfect match. Only the gene *DhicA* is affected by the mutation. *aphA*: chloramphenicol resistance gene (insertion cassette).



6.10.3 Complementation of the *unk9* mutant

The phenotype of the WT strain Cj1 was completely restored in the mutant *Dunk9* when complemented with the *unk9* gene. Figures 6.26 and 6.27 show the growth curves of the complemented mutant *Dunk9+unk9* compared to the *Dunk9* mutant and the Cj1 strain, and the motility phenotype in the three strains (mutant, complemented mutant and wild-type) in MH plates with full concentration of nutrients and 0.4% of agarose.

Figure 6.26 - Motility of strains Cj1, *Dunk9*, and *Dunk9+unk9* in MH 0.4% agar. (A) An example plate used to assess the mobility. (B) Average colony diameter and SD (three biological replicates, * = Significantly different from Cj1 WT. p<0.01 One-way ANOVA, Tukey's post-hoc).

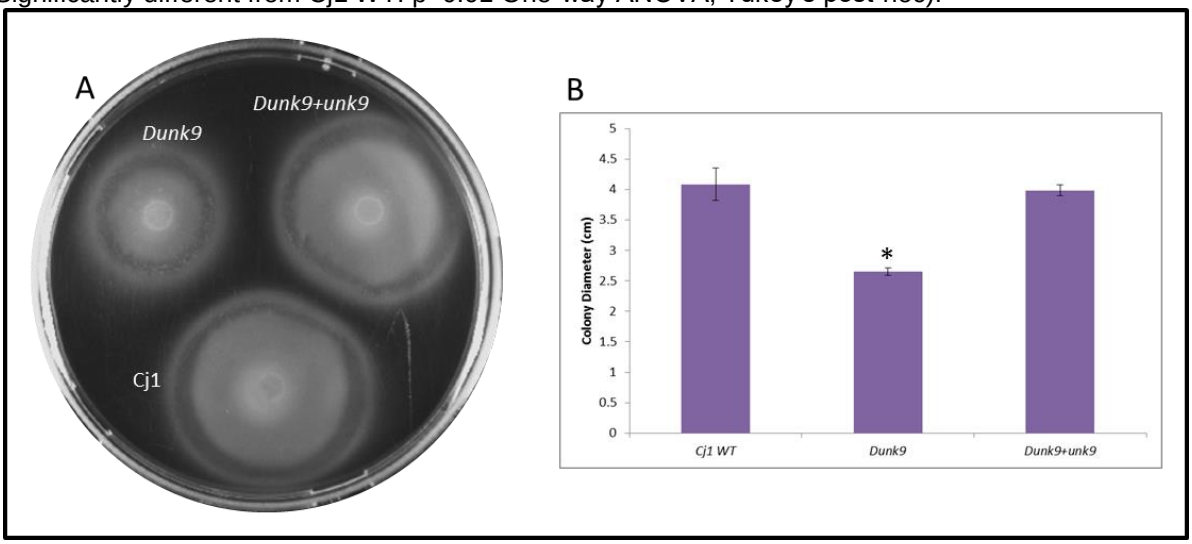
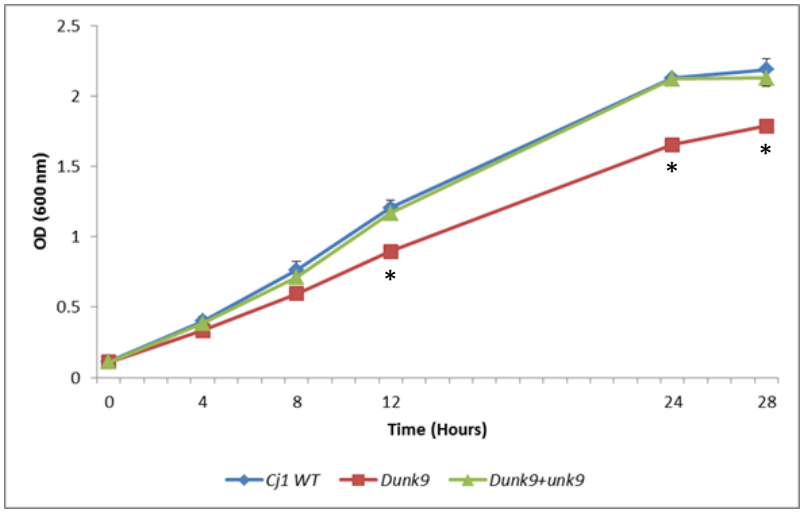


Figure 6.27 - Growth rate of the mutant *Dunk9*, the complemented mutant *Dunk9* and in the WT strain Cj1 in 25ml of MH broth measured for 28 hours. Each point is the average of three biological replicate, error bars show the SD. * = Statistically significant difference. p<0.01 T-test, single tailed, paired



6.11 Discussion

In this chapter, I described the phenotypic effects of the 11 mutations I introduced into the strain Cj1 of *C. jejuni* on selected phenotypes related to environmental survival and infection. In particular, I assessed the ability to grow in a controlled environment, the ability to producing biofilm, the motility of the bacterium in 0.4% agar plates with different concentration of nutrients, the survival at low pH, in synthetic gastric fluid and in tap water at low temperature, resistance to different antibiotics, and the ability to kill *G. mellonella.*

Two of these assays needed to be optimised in this study: the measurement of biofilm production and killing of *G. mellonella.*

I adapted a method that was not applied before to *C. jejuni* to assess the biofilm production on agar plates. Use of Congo Red for bacterial enumeration thanks to its interaction with beta-polysaccharides was described for the first time in 1981 (Teather and Wood, 1981). The method described in this chapter was previously applied to other bacterial species to judge the colony morphology and biofilm production, including *Salmonella* (Montiero *et al.*, 2011) and *Neisseria* (Kyungcheol *et al.*, 2007). Other methods for biofilm production evaluation in *Campylobacter* described in the literature are based on the staining with crystal violet of cells adhesive to borosilicate or plastic substrates (Oh *et al.*, 2014; Oh *et al.,* 2016; Naito *et al.,* 2010). These methods, although quantitative, are, in my experience, laborious and give results that are difficult to replicate. Moreover, the application of different gas concentrations to a static liquid bacterial suspension requires the tube to be prepared in a defined environment (i.e. gas cabinet), with broth that was equilibrated to the different oxygen concentrations and needs to be incubated open, which may lead to the contamination of the assay. Here, I describe a method that is qualitative,

hence it just highlights a positive or a negative phenotype, but it is simple and gives repeatable results.

Congo red was used previously to quantify the biofilm production of *C. jejuni* and the results obtained with this molecule were comparable to the one obtained with a less specific stain, such as Crystal violet (Reuter *et al*, 2010). Although only alpha-linked sugars have been identified as components of *C. jejuni* biofilm (Jowiya *et al.*, 2015), given the phenotype identified in this thesis and the previous results shown by Reuter *et al.* in 2010, we observed that Congo red binds with a component of the biofilm of *C. jejuni.* This could be explained with a previously undescribed beta-linked polysaccharide secreted by *C. jejuni*, or with an intereaction of the Congo red with the known exopolysaccharides secreted by *C. jejuni*, and requires further investigation.

In order to apply it to the microaerophilic species *C. jejuni*, the *G. mellonella* larvae killing assay needed to be optimised. *G. mellonella* larvae have been used as a model organism to evaluate the attenuation in virulence of several pathogenic bacteria including *Legionella pneumophila*, *Helicobacter pylori* and *Escherichia coli* (Giannouli *et al.,* 2014; Harding *et al.,* 2012; Leuko and Raivio 2012). For *C. jejuni*, the evaluation of virulence in *G. mellonella* larvae was described by Champion *et al*. in 2010 and Senior *et al*. in 2011. In both these papers, the larvae were challenged injecting 10^4 to 10^6 bacterial cells, causing close to 100% killing after two days at 37 °C. I could replicate this percentage of killing only when incubating the injected larvae under microaerophilic conditions. This result may be explained with a different growth and propagation conditions of the *G. mellonella* larvae: different pet shop suppliers grows the larvae under different conditions, feeding them different substrates or even treating them with antibiotics prophylactically to avoid infections

(Olivia Champion, personal communication). As shown by the PBS-injected and the non-injected groups, the larvae can survive for two days under microaerophilic conditions. It is not clear if the killing by *C. jejuni* is due to the extra stress imposed on the animal or it is due to a longer survival of the bacterium in the larvae, allowing colonisation. I was not able to recover any bacterial cells after incubation in atmospheric conditions, while I was able to count 10^3 to 10^4 bacterial cells in the haemocoel of the insect after incubation in microaerophilic conditions (data not shown). This might indicate that *C. jejuni* is not able to colonise the larvae in presence of atmospheric concentrations of oxygen, although I cannot exclude the presence of bacteria colonising the internal organs of the larvae (e.g. the larva's digestive system). The oxygen concentrations used in this study allowed estimating the attenuation of *Campylobacter* virulence, having a baseline of 90-100% killing in the wild type.

The conditions in which the genes of the plasmid pTet are expressed are poorly understood. The most valuable work in this area is the complete transcriptome of *C. jejuni* strain 81-176 (Taveirne *et al.,* 2013), showing the expression profiles amongst the whole genome of this strain of *C. jejuni* during stationary and mid-log phase and during an *in vivo* colonisation of chicken. That study highlighted how the genes on the plasmid pTet are only expressed during the stationary and mid-log phase of the bacterial growth and not during the colonisation of chicken. Even though the bacterium is considered to be non-pathogenic in poultry, it is fair to assume that in order to colonise the animal, *C. jejuni* has to resist stresses that are similar to those that it encounters when colonising other hosts (e.g. acid tolerance). The mutants analysed in this study, did not exhibit any difference in phenotypes associated with

acid tolerance, SGF tolerance or virulence in *Galleria mellonella* larvae, which accords with the observations described in the transcriptome analysis.

Another assay that did not show any difference between wild type and any of the mutant strains tested was survival in tap water. *C. jejuni* has been shown to survive up to 64 days in drinking water when incubated at 4 °C, while at 25 °C the number of viable cells drops to zero in the first day of incubation (Cools *et al.,* 2003; Trigui *et al.,* 2015). Although *Campylobacter* may be able to survive for such a long time at 4 °C, a difference between strains characterised by a different resistance to this particular stress can be identified in the first week (Trigui *et al.,* 2015), hence the time of incubation of 8 days was selected.

As far as the assays highlighting a phenotype in some of the mutant strains are concerned, the *DtetO* mutant showed a high level of susceptibility to tetracycline, the *Dunk9* mutant showed an impaired growth in liquid culture and impaired motility, and the *DhicA* mutant showed impaired motility and an a lack of biofilm production in the condition tested in this analysis.

The tetracycline susceptibility in *DtetO* is unsurprising, even though it is only partially restored by the complementation. Although this test only confirms that the tetracycline resistance phenotype is due to the presence of the *tetO* gene, it is useful to confirm that the technique used to produce the mutants works in these conditions. The reason why the complementation of *tetO* is only partial might reside in the different promoter that controls the expression of the gene in the complemented mutant. It may be useful to test the complementation of *DtetO* by *tetO* controlled by different promoters (for instance pCJC2, pCJC3, and pCJC4m described in Jervis *et al.,* 2015).

The *Dunk9* mutant showed two interesting phenotypes. They were both fully complemented when the *unk9* gene was integrated back in the chromosome of the mutant strain, a sign that these phenotypes are due solely to the absence of this gene. Unfortunately, the predicted gene product did not possess any conserved domains that may be associated to the phenotypes identified. The only conserved domain identified in the hypothetical protein is part of a superfamily identified in a protein involved in the meiotic phase in eukaryotic organism (Meuwissen *et al*, 2004, domain SCP-1, as reported in Section 4.3.6). The *unk9* gene expressing the hypothetical protein 9_FIG00469571_hypothetical_protein is much shorter than the one described in the original paper: the protein SCP1 is 946 amino acids long, while *unk9* encodes a protein of just 194 amino acids, and the fact that the only domain identified here is a partial multi-domain lowers the confidence of function association between these two proteins. However, recently an association between DNA supercoiling and motility in *C. jejuni* was described (Shortt *et al.,* 2016). The protein SCP1 interacts with chromatin and DNA, and it could be involved in processes that might influence DNA-supercoiling and tertiary structure. Certainly, a much deeper study of the gene *unk9* and its protein product is required to ascertain its cellular function. Motility is an extremely important phenotype for *C. jejuni*, involved in the colonisation of the host and in the environmental survival (Young *et al*, 2007), and being involved with this phenotype makes the *unk9* gene a candidate for future investigation. A follow up of this work should fully characterise the Unk9 protein (9_FIG00469571_hypothetical_protein). The protein should be isolated and its structure determined in order to gain a better understanding of its function, and the conditions in which the protein is produced by the bacterium should be identified.

Finally, the *DhicA* mutant showed a phenotype different from the wild-type but it could not be complemented. The gene *hicA* is part of a toxin-antitoxin system encoded by *C. jejuni.* These systems are formed by a toxin deleterious to the bacterial cell and an antitoxin acting like an antidote. Today, as many as six classes of toxin-antitoxin systems are known, characterised by different mechanisms of toxin inactivation (Page and Peti, 2016). Plasmid-encoded TA-systems were described initially to be involved in the maintenance and stabilisation of the plasmid by post-segregational killing (PSK), as described in Van Melderen and De Bast in 2009, but since then several chromosomal encoded TA-systems were discovered, which are not involved in PSK but work to ensure the survival of the bacterial population in response to various stresses (Page and Peti, 2016). For instance, in *Salmonella*, a toxin-antitoxin system *sehAB* was associated with virulence in mice (De La Cruz, *et al.,* 2013); in *Acidithiobacillus ferrooxidans* a series of TA-system were associated with the maintenance of integrated genetic elements (Bustamante *et al.,* 2014); in *Burkolderia pseudomallei* the toxin HicA was shown to have a role in persister formation (Butt *et al.,* 2014); in *Escherichia coli* multiple TA-systems were shown to influence biofilm formation and fimbriae (Kim *et al.,* 2009). In particular, the study of Kim *et al.* has found that 5 TA systems are involved in the upregulation of biofilm formation in the early phases of bacterial growth (8 H) through upregulation of fimbriae production (by repressing the expression of a single gene, *yjgK*, which encodes for an uncharachterised protein and has no homologues in *C. jejuni*) and in repression of biofilm production in the late phases of bacterial growth (24 H). Although an association between *hicA* and production of biofilm and motility is described *C. jejuni* for the first time, the phenotypes identified in this study are consistent with at least two of the previous works in other bacterial organisms. I

recorded impairment in motility and in biofilm production in the mutant *DhicA*. I also evaluated the possibility of forming persister cells in the mutant strain *DhicA,* because TA systems and ability of forming biofilm are linked to the ability of producing persister cells in several bacterial species (Xang and Wood, 2011). Persister cells are a small fraction of bacterial cells that survive to an antibiotic treatment but are not genetically resistant to the antibiotic molecule (Lewis *et al.,* 2007). I could not identify a difference between the WT strain Cj1 and the mutant *DhicA* in the number of persister cells formed after 24 hours. The impossibility of complementing the mutant that I observed applying the methods here described undermines the full association between the presence of the *hicA* gene and the phenotypes observed; however, the absence of other mutations in the genome of the strain *DhicA* was ascertained via whole genome sequencing. Several hypotheses could be formulated about the impossibility observed here of complementing the phenotype of this mutant strain. For instance, according to the mechanism described in 1986 for post-segregational killing of plasmid free cells (Gerdes *et al*, 1986), for the TA-system to work the quantity of toxin and antitoxin needs to be finely regulated: for the system to work, the quantity of RNAse translated (HicB antitoxin, in this case), has to be enough to quickly degrade all the toxin (HicA, in this instance). If the two genes are transcribed under the control of different promoters, as in this case with the complementation method with pCJC1, this could disrupt the correct synergy of the system. Moreover, it is reported that the TA-systems are transcribed in an operon, hence a single polycistronic mRNA (Yamaguchi *et al.,* 2011). Separating the toxin and the antitoxin gene might have an effect on the correct function of the system. Further investigation is required on this novel putative TA-system of *C. jejuni*: it is necessary to ascertain the involvement of this system in the

phenotypes identified; I suggest producing a mutant lacking for the entire system, and measuring its phenotype in the conditions identified in this work, then complementing the mutant with *hicBA* operon. I also suggest evaluating the mRNA levels of the toxin and the antitoxin in the complemented mutant via qPCR.

In conclusion, two genes have been associated with novel phenotypes in this mutagenesis analysis: this directly links the plasmids pTet-like to novel functions other than the maintenance of tetracycline resistance in a *Campylobacter* population. These functions are directly responsible for a greater fitness of the strains carrying those genes (the importance of chemotaxis and biofilm production for *Campylobacter* in the environment and in the pathogenesis was reviewed by Bolton in 2015 and by Young *et al*. in 2007) and may explain at least in part the great diffusion of the plasmid pTet, as expected by De Friis in 2007.

Concluding remarks

The importance of the pTet-like family of plasmids in *C. jejuni* has been neglected for the last 10 years. Since its description in the strain 81-176 of *C. jejuni* in 2000 (Bacon *et al.,* 2000), research has focused on the importance of the plasmid pVir, possibly because the plasmid pTet's function looked clear: the plasmid was transferred via the conjugative type four secretion system that it encodes, and it was maintained for tetracycline resistance (Batchelor *et al.,* 2004).

In this work, I studied the importance of this plasmid family through sequence analysis and mutagenesis.

Initially, I determined the sequence of 19 new pTet-like plasmids in *C. jejuni*. These plasmids were between ~37 kbp and ~48 kbp long and showed a high level of similarity with the plasmid pTet but also showed some remarkable differences: for example, three of them lacked the tetracycline resistance gene, and 10 of them lacked for the replication protein described in the pTet plasmid carried by the strain 81-176 (NC008790.1). I also assessed the presence of the plasmid pTet in a database of 4005 strains of *C. jejuni*, sequenced before February 2016. This study showed that the majority of the pTet genes are carried by around 15.4% percent of the strains, and of these, 25.3% lack the tetracycline resistance gene. Also, different clusters of genes of the plasmid pTet were characterised by different frequencies in the *C. jejuni* strain database. I approached this variability between the plasmids pTet-like producing a plasmid pan genome for the plasmids pTet-like that I sequenced, together with the ones that are already present in the NCBI database. The result of this analysis shows that the pTet-like plasmids are composed of 79 genes. The majority of them are core genes, which largely encode for maintenance and conjugation. The remainders are accessory genes, encoding for disposable

phenotypes, such as antibiotic resistance, *coli*cin production, or metabolism of particular compounds.

One strain of *C. jejuni* tested in this work (strain *Cj1*) was able to accept foreign DNA using a mutagenesis method developed for this bacterium, using a suicide plasmid constructed on pGEM-T easy as DNA vector. This strain has a particular genomic set-up: it shows the presence of the majority of the pTet-like plasmid pan-genome genes in a structure that appear to be either a much bigger plasmid, like the one present in the strains RM3194 (CP014745.1), or integrated in the chromosome. The sequencing experiments applied to this strain were not able definitively to ascertain the chromosomal or plasmid location of the pTet-like genes; however, 11 mutants were produced on this strain in order to study the importance of the pTet-encoded functions. The production of a further 7 mutants was attempted, but the mutant strains were not able to grow on selective plates for screening after transformation. Future work is required to understand whether this is due to a technical problem in the deletion protocol (and the kanamycin cassette used failed to substitute the target gene) or whether one or more of these seven genes are essential for the growth of *C. jejuni*. The genes targeted for mutation successfully encoded largely for hypothetical proteins, which contained conserved domains associated with several important functions: an acetyl transferase domain (*Datr*), a LabA-like domain associated with a variety of regulatory functions (*DlabA*), an ATPase associated with several cellular activity and a CsbA domain for stress response (*DepsG*), an ABC transporter (*Dabc*), a domain for nuclease activity and a domain for sensing extracellular L-fucose (*DparB*), a domain associated with a cag pathogenicity island in *H. pylori* (*Dcag*), a blood coagulase described in *S. aureus* (*Dcoag*), a conserved domain involved in chromatin structure in eukaryotic cells

(*Dunk9*), and a virulence-associated protein (*DvapD*). Furthermore, a mutant on the toxin protein *hicA* of the toxin-antitoxin system *hicAB* identified for the first time on the plasmids pTet-like was constructed, together with a mutant on the gene *tet(O)* responsible for tetracycline resistance.

Nine different phenotypic tests were applied to the 11 mutants produced on the strain Cj1. These tests were designed to simulate an array of conditions associated with pathogenesis and environmental survival and included a growth curve in Mueller-Hinton broth, resistance to different antibiotics, tolerance to low pH, survival in Synthetic Gastric Fluid, motility in soft agar, production of biofilm and formation of persister cells, survival in water and virulence in the model organism *G. mellonella* larvae. Two of these tests required an optimisation that led to the design of a novel protocol for the test of biofilm production in *C. jejuni* when grown in solid medium and a modification of the incubation conditions for testing the virulence in *G. mellonella* larvae.

Three of the mutants showed a different phenotype when compared to the wild-type strain in some of the conditions tested: the mutant *DtetO* exhibited a high susceptibility to the antibiotic tetracycline, the mutant *DhicA* exhibited an impaired ability of swimming in soft agar in particular concentrations of nutrients and an impaired ability to produce biofilm, the mutant *Dunk9* showed a reduced growth rate in MHB and an impaired ability of swimming in soft agar in particular concentrations of nutrients. If the phenotype of the mutant *DtetO* is unsurprising, as the gene *tet(O)* is reported to be responsible of the tetracycline resistance phenotype, the phenotypes of the other two mutants require a deeper study.

The hypothetical protein affected by the mutation in the mutant *Dunk9* carries a conserved domain involved in the interaction with chromatin and DNA in eukaryotic

organisms. In *Campylobacter,* an association between DNA supercoiling and motility was described recently (Shortt *et al.,* 2016), and it is certainly interesting that the mutant *Dunk9* shows a phenotype impaired in motility and growth.

The mutant *DhicA* exhibit an interesting phenotype on motility and biofilm production: these phenotypes were shown to be associated with a toxin-antitoxin system in several bacterial species, such as *E. coli* (Kim *et al.,* 2009), but never before in *C. jejuni.*

Complementation of the mutants was performed in order to ascertain the association of the genes deleted with the phenotype observed: the phenotype of mutant *Dunk9* was fully restored to the wild-type, the phenotype of the mutant *DtetO* was partially restored, and finally it was impossible to restore the wild-type phenotype in the mutant *DhicA.*

The work carried out in this thesis can be further extended both in the bioinformatics section and in the functional analysis of the pTet-associated genes.

In the production of the pTet-like plasmid pan-genome, I used a total of 27 pTet-like plasmids, present on the NCBI database, or sequenced in-house and whose sequence was ascertained using the paired-end information of the Illumina sequencing. I have demonstrated how 15.4% percent of the *C. jejuni* strains sequenced to date show at least half of the genes associated with the plasmid pTet. A step that can be taken in the direction of extending the knowledge on the variability of the plasmids pTet-like could be producing a pTet-like plasmid pan-genome using a larger number of plasmids pTet-like. These could be assembled from the strains that have been already sequenced, with the help of Bandage (Wick *et al.,* 2015) or a newer version of SPAdes software that promises to be plasmid-aware (PlasmidSPAdes, Bankevich *et al.,* 2012). A characteristic of pan-genomes is the

completeness: adding more genomes in the analysis corresponds to having more genes in the pan-genome result. When this growth in the number of genes after adding more genomes in the analysis reaches a plateau, the pan-genome is complete (Lefébure *et al.,* 2010; Méric *et al.,* 2014). It would be interesting to investigate whether the same were true for the pTet-like plasmid pan-genome: the 79 genes described in this study describe the variability of the 27 pTet-like plasmids I used in the pan-genome analysis, but further work is required to ascertain if this is the total variability of the plasmid or if new genes can be discovered by adding other pTet-like plasmids to the analysis.

Secondly, other mutants could be produced in order to study the phenotype of other genes identified in the pTet-like plasmid pan-genome. An array of hypothetical proteins has been identified in several different plasmids pTet-like, and many of them show interesting conserved domains that could not be studied in this work. A few examples of hypothetical proteins that would require further study are the one encoded by “7_FIG00471069_hypothetical_protein”, carrying a conserved domain for membrane carbohydrate transport or the one encoded by “42_FIG00638667_hypothetical_protein,” which shows a conserved domain encoding for a peptidase involved with bacteriocins resistance. Furthermore, several hypothetical proteins identified with automated annotation show no putative function and no conserved domains (Table 4.1.1), leaving the production of a mutant and a thorough phenotypic study the only way to identify their importance.

The mutants produced in this work were tested with an array of phenotypic tests designed to simulate conditions important for bacterial environmental survival and host colonisation, but they by no means cover all the possible characteristics of *C. jejuni*. The majority of the mutants did not show a phenotype different from the

wild-type, but different phenotypic assays could highlight differences that I was not able to identify. In particular, future work could point in the direction of high throughput metabolic analysis using Biolog system (Biolog INC., http://www.biolog.com/) and host invasion, evaluating the ability to invade the well-established models for *C. jejuni* caco-2 or INT407 cell lines.

The important phenotype shown by the mutant *Dunk9* has been fully complemented with the system developed by Jervis *et al.* in 2015. In the future, it is necessary to characterise fully the function of the deleted gene and the protein produced. As far as the deleted gene is concerned, it would be interesting to ascertain possible interactions with other genes and expression patterns in different conditions via transcriptomics study. It is also necessary to identify the deleted protein, studying its cellular localisation or secretion, structure, and interaction with other proteins.

The phenotype of the mutant *DhicA* was not complemented; however, the absence of mutations different from the expected one was ascertained via whole genome sequencing of the mutant strain. The impossibility of restoring the wild-type phenotype in this mutant strain could be due to the operon structure of the toxin-antitoxin system, and future analyses of this mutant could involve the study of the expression levels of the toxin and the antitoxin in the complemented strain and production of a mutant lacking of the entire system, if possible. I also suggest trying to complement the mutant with different pCJC plasmids, containing the gene to complement under the control of different promoters (Jervis *et al.,* 2015).

Overall, in this project, I demonstrated that plasmids pTet-like have a role in the lifestyle of *C. jejuni* other than the tetracycline resistance. I described a variability in the plasmids pTet-like and the absence of the tetracycline resistance characteristic

from a large set of them. I summarised the variability of the plasmids pTet-like describing the pTet-like plasmid pan-genome and a set of core and accessory genes for the plasmid pTet. I produced several mutants on pTet-like genes and tested them for phenotypes that simulated the environmental survival and host invasion, revealing that some of the genes encoded by the plasmids pTet-like are useful in previously undescribed conditions, such as the production of biofilm and motility.

This work could shed new light on the lifestyle of *C. jejuni*: I showed that a common characteristic like the presence of a plasmid pTet-like is not to be ignored. A better understanding of the genetic and the lifestyle of *C. jejuni*, achieved with studies of this kind, can lead to the design of better control measures for the bacterium, aiming to eliminate it while it colonises the environment and typical ecological niches, preventing the attack to the human host.

References

Aarestrup, F. M., & Engberg, J. (2001). Antimicrobial resistance of thermophilic *Campylobacter*. *Veterinary Research*, *32*(3/4), 311-321.

Albert, M. J., Udo, E., Jose, B. T., Haridas, S., & Rotimi, V. O. (2009). Tetracycline resistance is frequent among *Campylobacter jejuni* isolates from Kuwait. *Microbial Drug Resistance (Larchmont, N.Y.)*, *15*(2), 115-20.

Alfredson, D. a., & Korolik, V. (2003). Sequence analysis of a cryptic plasmid pCJ419 from *Campylobacter jejuni* and construction of an *Escherichia coli-Campylobacter* shuttle vector. *Plasmid*, *50*(2), 152-160.

Alfredson, D. A., & Korolik, V. (2007). Antibiotic resistance and resistance mechanisms in *Campylobacter jejuni* and *Campylobacter coli*. *FEMS Microbiology Letters*, *277*(2).

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, *12*(1), 402.

Anderson, L. B., Hertzel, A. V, & Das, A. (1996). Agrobacterium tumefaciens VirB7 and VirB9 form a disulfide-linked protein complex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(17), 8889-94.

Aravind, L., Leipe, D. D., & Koonin, E. V. (1998). Toprim - A conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Research*, *26*(18), 4205-4213.

Arcus, V. (2002). OB-fold domains: A snapshot of the evolution of sequence, structure and function. *Current Opinion in Structural Biology*. 12(6), 794-801.

Arthur, M., Depardieu, F., Molinas, C., Reynolds, P., & Courvalin, P. (1995). The vanZ gene of Tn1546 from enterococcus faecium BM4147 confers resistance to teicoplanin. *Gene*, *154*(1), 87-92.

Asakura, M., Samosornsuk, W., Hinenoya, A., Misawa, N., Nishimura, K., Matsuhisa, A., & Yamasaki, S. (2008). Development of a cytolethal distending toxin (cdt) gene-based species-specific multiplex PCR assay for the detection and identification of *Campylobacter jejuni*, *Campylobacter coli* and *Campylobacter* fetus. *Fems Immunology and Medical Microbiology*, *52*(2), 260-266.

Atack, J. M., & Kelly, D. J. (2009). Oxidative stress in *Campylobacter jejuni* : responses, resistance and regulation. *Future Microbiology*, *4*(6), 677-690.

Atmakuri, K., Cascales, E., & Christie, P. J. (2004). Energetic components VirD4, VirB11 and VirB4 mediate early DNA transfer reactions required for bacterial type IV secretion. *Molecular Microbiology*, *54*(5), 1199-1211.

Avrain, L., Vernozy-Rozand, C., & Kempf, I. (2004). Evidence for natural horizontal transfer of tetO gene between *Campylobacter jejuni* strains in chickens. *Journal of Applied Microbiology*, *97*(1), 134-140.

Babst, M., Katzmann, D. J., Estepa-Sabal, E. J., Meerloo, T., & Emr, S. D. (2002). Escrt-III: An Endosome-Associated Heterooligomeric Protein Complex Required for MVB Sorting. *Developmental Cell*, *3*(2), 271-282.

Bachtiar, B. M., Coloe, P. J., & Fry, B. N. (2007). Knockout mutagenesis of the kpsE gene of *Campylobacter jejuni* 81116 and its involvement in bacterium-host interactions. *FEMS Immunology and Medical Microbiology*, *49*(1), 149-154.

Backert, S., Ziska, E., Brinkmann, V., Zimny-Arndt, U., Fauconnier, A., Jungblut, P. R., … Meyer, T. F. (2000). Translocation of the *Helicobacter pylori* CagA protein in gastric epithelial cells by a type IV secretion apparatus. *Cellular Microbiology*, *2*(2), 155-164.

Bacon, D. J., Alm, R. A., Burr, D. H., Hu, L., Kopecko, D. J., Ewing, C. P., … Guerry, P. (2000). Involvement of a Plasmid in Virulence of *Campylobacter jejuni* 81-176. *Infection and Immunity*, *68*(8), 4384-4390.

Bacon, D. J., Alm, R. A., Hu, L., Hickey, T. E., Ewing, C. P., Batchelor, R. A., … Guerry, P. (2002). DNA Sequence and Mutational Analyses of the pVir Plasmid of *Campylobacter jejuni* 81-176. *Infection and Immunity*, *70*(11), 6242-6250.

Bacon, D. J., Szymanski, C. M., Burr, D. H., 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*(3), 769-777.

Balaban, M., & Hendrixson, D. R. (2011). Polar flagellar biosynthesis and a regulator of flagellar number influence spatial parameters of cell division in *Campylobacter jejuni*. PLoS Pathogens, 7(12), e1002420.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455-477.

Barrett, a. J., & Rawlings, N. D. (2001). Evolutionary lines of cysteine peptidases. *Biological Chemistry*. 382(5), 727-733.

Barrett, A. J., Tolle, D. P., & Rawlings, N. D. (2003). Managing peptidases in the genomic era. *Biological Chemistry*, *384*(6), 873-882.

Batchelor, R. a, Pearson, B. M., Friis, L. M., Guerry, P., & Wells, J. M. (2004). Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species. *Microbiology (Reading, England)*, *150*(Pt 10), 3507-17.

Bäumler, A. J., Tsolis, R. M., Bowe, F. A., Kusters, J. G., Hoffmann, S., & Heffron, F. (1996). The pef fimbrial operon of *Salmonella typhimurium* mediates adhesion to murine small intestine and is necessary for fluid accumulation in the infant mouse. *Infection and Immunity*, *64*(1), 61-68.

Bawden, a L., Glassberg, K. J., Diggans, J., Shaw, R., Farmerie, W., & Moyer, R. W. (2000). Complete genomic sequence of the *Amsacta moorei* entomopoxvirus: analysis and comparison with other poxviruses. *Virology*, *274*(1), 120-39.

Bayliss, R., Harris, R., Coutte, L., Monier, A., Fronzes, R., Christie, P. J., … Waksman, G. (2007). NMR structure of a complex between the VirB9/VirB7 interaction domains of the pKM101 type IV secretion system. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1673-1678.

Bergquist, P. L. (1987) Incompatibility. In: Hardy KG (ed.) *Plasmids: A Practical Approach*, pp. 37-78. Oxford: IRL Press.

Beumer, R. R., de Vries, J., & Rombouts, F. M. (1992). *Campylobacter jejuni* non-culturable coccoid cells. *International Journal of Food Microbiology*, *15*(1-2), 153-163.

Bolton, D. J. (2015). *Campylobacter* virulence and survival factors. *Food Microbiology*, *48*, 99-108.

Boucher, C., Martinel, A., Barberis, P., Alloing, G., & Zischek, C. (1986). Virulence genes are carried by a megaplasmid of the plant pathogen *Pseudomonas solanacearum*. *MGG Molecular & General Genetics*, *205*(2), 270-275.

Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*(March), 1709-1712.

Branda, S. S., González-Pastor, J. E., Dervyn, E., Ehrlich, S. D., Losick, R., & Kolter, R. (2004). Genes involved in formation of structured multicellular communities by *Bacillus subtilis*. *Journal of Bacteriology*, *186*(12), 3970-3979.

Buelow, D. R., Christensen, J. E., Neal-McKinney, J. M., & Konkel, M. E. (2011). *Campylobacter jejuni* survival within human epithelial cells is enhanced by the secreted protein CiaI. *Molecular Microbiology*, *80*(5), 1296-1312.

Burian, J., Guller, L., Macor, M., & Kay, W. W. (1997). Small cryptic plasmids of multiplasmid, clinical *Escherichia coli*. Plasmid, 37(1), 2–14.

Bustamante, P., Tello, M., & Orellana, O. (2014). Toxin-antitoxin systems in the mobile genome of *Acidithiobacillus ferrooxidans*. *PLoS ONE*, *9*(11), e112226.

Buswell, C. M., Herlihy, Y. M., Lawrence, L. M., McGuiggan, J. T. M., Marsh, P. D., Keevil, C. W., & Leach, S.A. (1998). Extended survival and persistence of *Campylobacter* spp. in water and aquatic biofilms and their detection by immunofluorescent-antibody and -rRNA staining. *Applied and Environmental Microbiology*. 64:733-741.

Butt, A., Higman, V. A., Williams, C., Crump, M. P., Hemsley, C. M., Harmer, N., … Christopher, W. (2014). The HicA toxin from *Burkholderia pseudomallei* has a role in persister cell formation. *Biochemical Journal*, *459*(19464), 333-44.

Butzler, J.-P., & Oosterom, J. (1991). *Campylobacter*: pathogenicity and significance in foods. *International Journal of Food Microbiology*, *12*(1), 1-8.

Byrne, C. M., Clyne, M., & Bourke, B. (2007). *Campylobacter jejuni* adhere to and invade chicken intestinal epithelial cells in vitro. *Microbiology*, *153*(2), 561-569.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421.

Cárdenas-Mondragón, M. G., Ares, M. A., Panunzi, L. G., Pacheco, S., Camorlinga-Ponce, M., Girón, J. A., … De la Cruz, M. A. (2016). Transcriptional profiling of type II toxin-antitoxin genes of *Helicobacter pylori* under different environmental conditions: Identification of hp0967-hp0968 system. Frontiers in Microbiology, 7(NOV), 1872.

Carrillo, C. D., Taboada, E., Nash, J. H. E., Lanthier, P., Kelly, J., Lau, P. C., … Szymanski, C. M. (2004). Genome-wide expression analyses of *Campylobacter jejuni* NCTC11168 reveals coordinate regulation of motility and virulence by flhA. *The Journal of Biological Chemistry*, *279*(19), 20327-38.

Castanié-Cornet, M. P., Bruel, N., & Genevaux, P. (2014). Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. *Biochimica et Biophysica Acta - Molecular Cell Research*.

Champion, O. L., Gaunt, M. W., Gundogdu, O., Elmi, A., Witney, A. A., Hinds, J., … Wren, B. W. (2005). Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(44), 16043-8.

Champion, O. L., Karlyshev, A. V, Senior, N. J., Woodward, M., La Ragione, R., Howard, S. L., … Titball, R. W. (2010). Insect infection model for *Campylobacter jejuni* reveals that O-methyl phosphoramidate has insecticidal activity. *The Journal of Infectious Diseases*, *201*(5), 776-782.

Champoux, J. J. (2001). DNA TOPOISOMERASES: Structure, Function, and Mechanism. *Annual Review of Biochemistry*, *70*(1), 369-413.

Chang, N. and Taylor, D.E. (1990) Use of pulsed-field agarose gel electrophoresis to size genomes of *Campylobacter* species and to construct a SalI map of *Campylobacter jejuni* UA580. *Journal of Bacteriology* 172, 5211-5217

Chatzipanagiotou, S., Papavasileiou, E., Lakumenta, A., Makri, A., Nicolaou, C., Chantzis, K., … Legakis, N. I. (2002). Antimicrobial susceptibility patterns of

Campylobacter jejuni strains isolated from hospitalized children in Athens, Greece. *The Journal of Antimicrobial Chemotherapy*, *49*(5), 803-805.

Chaveerach, P., ter Huurne, A. A. H. M., Lipman, L. J. A., & van Knapen, F. (2003). Survival and resuscitation of ten strains of *Campylobacter jejuni* and *Campylobacter coli* under acid conditions. *Applied and Environmental Microbiology*, *69*(1), 711-714.

Chen, C. M., Ye, Q. Z., Zhu, Z., Wanner, B. L., & Walsh, C. T. (1990). Molecular biology of carbon-phosphorus bond cleavage: Cloning and sequencing of the phn (psiD) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. *Journal of Biological Chemistry*, *265*(8), 4461-4471.

Christie, P. J. (2004). Type IV secretion: The Agrobacterium VirB/D4 and related conjugation systems. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1694*(1-3 SPEC.ISS.), 219-234.

Christie, P. J., & Vogel, J. P. (2000). Bacterial type IV secretion: Conjugation systems adapted to deliver effector molecules to host cells. *Trends in Microbiology*. 8(8), 354-360.

Clark, E., Manulis, S., Ophir, Y., Barash, I., & Gafni, Y. (1993). Cloning and characterization of iaaM and iaaH from *Erwinia herbicola* pathovar gypsophilae. *Phytopathology*, *83*, 234-240.

Cole, S. P., Harwood, J., Lee, R., She, R., & Guiney, D. G. (2004). Characterization of monospecies biofilm formation by *Helicobacter pylori*. *Journal of Bacteriology*, *186*(10), 3124-32.

Cools, I., Uyttendaele, M., Caro, C., D'Haese, E., Nelis, H. J., & Debevere, J. (2003). Survival of *Campylobacter jejuni* strains of different origin in drinking water. *Journal of Applied Microbiology*, *94*(5), 886-892.

Cort, J. R., Ramelot, T. A., Murray, D., Acton, T. B., Ma, L. C., Xiao, R., … Kennedy, M. A. (2008). Structure of an acetyl-CoA binding protein from *Staphylococcus aureus* representing a novel subfamily of GCN5-related N-acetyltransferase-like proteins. *Journal of Structural and Functional Genomics*, *9*(1-4), 7-20.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., & Lappin-Scott, H.M. (1995). Microbial biofilms. *Annual Reviews of Microbiology*. 49:711-745

Cowles, K. N., & Goodrich-Blair, H. (2005). Expression and activity of a *Xenorhabdus nematophila* haemolysin required for full virulence towards *Manduca sexta* insects. *Cellular Microbiology*, *7*(2), 209-219.

Crespo, M. D., Altermann, E., Olson, J., Miller, W. G., Chandrashekhar, K., & Kathariou, S. (2016). Novel plasmid conferring kanamycin and tetracycline resistance in the turkey-derived *Campylobacter jejuni* strain 11601MD. *Plasmid*, *86*, 32-37.

Cummins, P. M., & O'Connor, B. (1998). Pyroglutamyl peptidase: An overview of the three known enzymatic forms. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, *1429*(1), 1-17.

Rollins, D., & Colwell, R. (1986). Viable but Nonculturable Stage of *Campylobacter jejuni* and Its Role in Survival in the Natural Aquatic Environment, Applied and Environmental Microbiology. 52 (3), 531-538.

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394-403.

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*(4), 205-211.

De Boer, P., Wagenaar, J. A., Achterberg, R. P., Van Putten, J. P. M., Schouls, L. M., & Duim, B. (2002). Generation of *Campylobacter jejuni* genetic diversity in vivo. *Molecular Microbiology*, *44*(2), 351-359.

De Bois, M. H., Schoemaker, M. C., van der Werf, S. D., & Puylaert, J. B. (1989). Pancreatitis associated with *Campylobacter jejuni* infection: diagnosis by ultrasonography. *BMJ (Clinical Research Ed.)*, *298*(6679), 1004.

De la Cruz, M. A., Zhao, W., Farenc, C., Gimenez, G., Raoult, D., Cambillau, C., … Méresse, S. (2013). A Toxin-Antitoxin Module of *Salmonella* Promotes Virulence in Mice. *PLoS Pathogens*, *9*(12), 1-13.

Del Solar, G., & Espinosa, M. (2002). Plasmid copy number control: an ever-growing story. *Molecular Microbiology*, *37*(3), 492-500.

Delauney, A. J., Cheon, C. I., Snyder, P. J., & Verma, D. P. S. (1990). A nodule-specific sequence encoding a methionine-rich polypeptide, nodulin-21. *Plant Molecular Biology*, *14*(3), 449-451.

Derouiche, R., Gavioli, M., Bénédetti, H., Prilipov, a, Lazdunski, C., & Lloubès, R. (1996). TolA central domain interacts with *Escherichia coli* porins. *The EMBO Journal*, *15*(23), 6408-6415.

Dingle, K. E., Colles, F. M., Wareing, D. R. A., Ure, R., Fox, A. J., Bolton, F. E., … Maiden, M. C. J. (2001). Multilocus sequence typing system for *Campylobacter jejuni*. *Journal of Clinical Microbiology*, 39(1), 14–23.

Easter, J., & Gober, J. W. (2002). ParB-stimulated nucleotide exchange regulates a switch in functionally distinct ParA activities. *Molecular Cell*, *10*(2), 427-434.

Ebersbach, G., Sherratt, D. J., & Gerdes, K. (2005). Partition-associated incompatibility caused by random assortment of pure plasmid clusters. *Molecular Microbiology*, *56*(6), 1430-1440.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797.

EFSA. (2008). Report of the task force on zoonoses data collection on the analysis of the baseline survey on the prevalence of Salmonella in slaughter pigs, in the EU, 2006-2007, part A: Salmonella prevalence estimates. *EFSA J.* 135, 1-111.

Elmi, A., Watson, E., Sandu, P., Gundogdu, O., Mills, D. C., Inglis, N. F., … Dorrell, N. (2012). *Campylobacter jejuni* outer membrane vesicles play an important role in bacterial interactions with human intestinal epithelial cells. *Infection and Immunity*, *80*(12), 4089-4098.

Eucker, T. P., & Konkel, M. E. (2012). The cooperative action of bacterial fibronectin-binding proteins and secreted proteins promote maximal *Campylobacter jejuni* invasion of host cells by stimulating membrane ruffling. *Cellular Microbiology*, *14*(2), 226-238.

Felder, C. B., Graul, R. C., Lee, A. Y., Merkle, H.-P., & Sadee, W. (1999). The Venus flytrap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors. *AAPS pharmSci*, *1*(2), 7-26.

Fidelma Boyd, E., Hill, C. W., Rich, S. M., & Hard, D. L. (1996). Mosaic structure of plasmids from natural populations of *Escherichia coli*. Genetics, 143(3), 1091–1100.

Field, L.H., Headley, V.L., Payne, S.M., Berry, L.J. (1986) Influence of iron on growth, morphology, outer membrane protein composition, and synthesis of siderophores in *Campylobacter jejuni*. Infection and Immunity 54 , 126-132.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279-D285.

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*(6), 2399-407.

Fondi, M., Bacci, G., Brilli, M., Papaleo, M. C., Mengoni, A., Vaneechoutte, M., … Fani, R. (2010). Exploring the evolutionary dynamics of plasmids: the *Acinetobacter* pan-plasmidome. *Bmc Evolutionary Biology*, *10*(1).

Friedrich, R., Panizzi, P., Fuentes-Prior, P., Richter, K., Verhamme, I., Anderson, P. J., … Bock, P. E. (2003). Staphylocoagulase is a prototype for the mechanism of cofactor-induced zymogen activation. *Nature*, *425*(2001), 535-539.

Friis, L. M., Pin, C., Taylor, D. E., Pearson, B. M., & Wells, J. M. (2007). A role for the tet(O) plasmid in maintaining *Campylobacter* plasticity. *Plasmid*, *57*(1), 18-28.

Galinski, M. R., Xu, M., & Barnwell, J. W. (2000). *Plasmodium vivax* reticulocyte binding protein-2 (PvRBP-2) shares structural features with PvRBP-1 and the *Plasmodium yoelii* 235 kDa rhoptry protein family. *Molecular and Biochemical Parasitology* (Vol. 108).

Gallay, A., Prouzet-Mauléon, V., Kempf, I., Lehours, P., Labadi, L., Camou, C., … Mégraud, F. (2007). *Campylobacter* Antimicrobial Drug Resistance among Humans, Broiler Chickens, and Pigs, France. *Emerging Infectious Diseases*, *13*(2), 259-266.

Garbers, D. L., Chrisman, T. D., Wiegn, P., Katafuchi, T., Albanesi, J. P., Bielinski, V., … Burnett, J. C. (2006). Membrane guanylyl cyclase receptors: an update. *Trends in Endocrinology and Metabolism*. 17(6):251-258.

Garénaux, A., Guillou, S., Ermel, G., Wren, B., Federighi, M., & Ritz, M. (2008). Role of the Cj1371 periplasmic protein and the Cj0355c two-component regulator in the *Campylobacter jejuni* NCTC 11168 response to oxidative stress caused by paraquat. Research in Microbiology, 159(9–10), 718–726.

Garénaux, A., Jugiau, F., Rama, F., de Jonge, R., Denis, M., Federighi, M., & Ritz, M. (2008). Survival of *Campylobacter jejuni* Strains from Different Origins Under Oxidative Stress Conditions: Effect of Temperature. *Current Microbiology*, *56*(4), 293-297.

Gaujoux, R., & Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, *11*(1), 367.

Gaynor, E. C., Wells, D. H., MacKichan, J. K., & Falkow, S. (2005). The *Campylobacter jejuni* stringent response controls specific stress survival and virulence-associated phenotypes. *Molecular Microbiology*, *56*(1), 8-27.

Giannouli, M., Palatucci, A. T., Rubino, V., Ruggiero, G., Romano, M., Triassi, M., … Zarrilli, R. (2014). Use of larvae of the wax moth *Galleria mellonella* as an in vivo model to study the virulence of Helicobacter pylori. *BMC Microbiology*, *14*(1), 228.

Gibreel, A., Sköld, O., & Taylor, D. E. (2004). Characterization of plasmid-mediated aphA-3 kanamycin resistance in *Campylobacter jejuni*. *Microbial Drug Resistance (Larchmont, N.Y.)*, *10*(2), 98-105.

Gilbert, M., Brisson, J. R., Karwaski, M. F., Michniewicz, J., Cunningham, A. M., Wu, Y., … Wakarchuk, W. W. (2000). Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384. Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-MHz 1H and 13C NMR analysis. *Journal of Biological Chemistry*, *275*(6), 3896-3906.

Gogarten, J. P., & Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology*, *3*(9), 679-687.

Golden, N. J., & Acheson, D. W. K. (2002). Identification of motility and autoagglutination *Campylobacter jejuni* mutants by random transposon mutagenesis. *Infection and Immunity*, *70*(4), 1761-1771.

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, *27*(2), 221-224.

Guccione, E., Del Rocio Leon-Kempis, M., Pearson, B. M., Hitchin, E., Mulholland, F., Van Diemen, P. M., … Kelly, D. J. (2008). Amino acid-dependent growth of *Campylobacter jejuni*: Key roles for aspartase (AspA) under microaerobic and oxygen-limited conditions and identification of AspB (Cj0762), essential for growth on glutamate. Molecular Microbiology, 69(1), 77–93.

Guerry, P. (2007). *Campylobacter* flagella: not just for motility. *Trends in Microbiology*, *15*(10), 456-461.

Guerry, P., Ewing, C. P., Schirm, M., Lorenzo, M., Kelly, J., Pattarini, D., … Logan, S. (2006). Changes in flagellin glycosylation affect *Campylobacter* autoagglutination and virulence. *Molecular Microbiology*, *60*(2), 299-311.

Gunther, N. W., Reichenberger, E. R., & Bono, J. L. (2016). Complete Genome Sequence of UV-Resistant *Campylobacter jejuni* RM3194, Including an 81.08-Kilobase Plasmid. *Genome Announcements*, *4*(2), e00305-16.

Gunther, N. W., Reichenberger, E. R., & Bono, J. L. (2016). Complete Genome Sequence of UV-Resistant *Campylobacter jejuni* RM3194, Including an 81.08-Kilobase Plasmid. *Genome Announcements*, *4*(2), e00305-16.

Guo, B., Wang, Y., Shi, F., Barton, Y.-W., Plummer, P., Reynolds, D. L., … Zhang, Q. (2008). CmeR functions as a pleiotropic regulator and is required for optimal colonization of *Campylobacter jejuni* in vivo. *Journal of Bacteriology*, *190*(6), 1879-1890.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072-1075.

Hackett, J., Wyk, P., Reeves, P., & Mathan, V. (1987). Mediation of serum resistance in *Salmonella typhimurium* by an 11-kilodalton polypeptide encoded by the cryptic plasmid. *The Journal of Infectious Diseases*, *155*(3), 540-549.

Haft, D. H., Selengut, J., Mongodin, E. F., & Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes. *PLoS Computational Biology*, *1*(6), 0474-0483.

Hallet, B., & Sherratt, D. J. (1997). Transposition and site-specific recombination: Adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiology Reviews*. 21(2), 157-178.

Han, F., Pu, S., Hou, A., & Ge, B. (2009). Characterization of Clinical and Environmental Types of *Vibrio vulnificus* Isolates from Louisiana Oysters. *Foodborne Pathogens and Disease*, *6*(10), 1251-1258.

Handford, J. I., Ize, B., Buchanan, G., Butland, G. P., Greenblatt, J., Emili, A., & Palmer, T. (2009). Conserved network of proteins essential for bacterial viability. *Journal of Bacteriology*, *191*(15), 4732-4749.

Hapfelmeier, S., Domke, N., Zambryski, P. C., & Baron, C. (2000). VirB6 is required for stabilization of VirB5 and VirB3 and formation of VirB7 homodimers in Agrobacterium tumefaciens. *Journal of Bacteriology*, *182*(16), 4505-4511.

Harding, C. R., Schroeder, G. N., Reynolds, S., Kosta, A., Collins, J. W., Mousnier, A., & Frankel, G. (2012). *Legionella pneumophila* pathogenesis in the *Galleria mellonella* infection model. *Infection and Immunity*, *80*(8), 2780-2790.

Hardy, K. C. (1986). Bacterial plasmids. *Kluwer Academic Publishers Group*.

Harrisingh, M. C., & Lloyd, A. C. (2004). Ras/Raf/ERK signalling and NF1: Implications for neurofibroma formation. *Cell Cycle*, *3*(10), 1255-1258.

Harrison, C. (2003). GrpE, a nucleotide exchange factor for DnaK. Cell Stress & Chaperones, 8(3), 218–224.

Harrison, J. W., Dung, T. T. N., Siddiqui, F., Korbrisate, S., Bukhari, H., Vu Tra, M. P., … Champion, O. L. (2014). Identification of possible virulence marker from *Campylobacter jejuni* isolates. *Emerging Infectious Diseases*, *20*(6), 1026-1029.

Hazeleger, W. C., Wouters, J. A., Rombouts, F. M., & Abee, T. (1998). Physiological activity of *Campylobacter jejuni* far below the minimal growth temperature. *Applied and Environmental Microbiology*, *64*(10), 3917-3922.

He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W. H., & Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res*, *44*(W1), W236-W241.

Hendrixson DR. 2006. A phase-variable mechanism controlling the *Campylobacter jejuni* FlgR response regulator influences commensalism. Mol. Microbiol. 61:1646-59

Hendrixson DR. 2008. Restoration of flagellar biosynthesis by varied mutational events in *Campylobacter jejuni*. Mol. Microbiol. 70:519-36

Hendrixson, D. R. (2006). A phase-variable mechanism controlling the *Campylobacter jejuni* FlgR response regulator influences commensalism. *Molecular Microbiology*, *61*(6), 1646-1659.

Hendrixson, D. R., & DiRita, V. J. (2003). Transcription of sigma54-dependent but not sigma28-dependent flagellar genes in *Campylobacter jejuni* is associated

with formation of the flagellar secretory apparatus. *Molecular Microbiology*, *50*(2), 687-702.

Hendrixson, D. R., & DiRita, V. J. (2004). Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract. *Molecular Microbiology*, *52*(2), 471-484.

Henthorn, K. S., & Friedman, D. I. (1995). Identification of related genes in phages 80 and P22 whose products are inhibitory for phage growth in *Escherichia coli* IHF mutants. *Journal of Bacteriology*, *177*(11), 3185-3190.

Hermans, D., Van Deun, K., Martel, A., Van Immerseel, F., Messens, W., Heyndrickx, M., … Pasmans, F. (2011). Colonization factors of *Campylobacter jejuni* in the chicken gut. *Veterinary Research*, *42*(1), 82.

Hiett, K. L., Rothrock, M. J., & Seal, B. S. (2013). Characterization of the *Campylobacter jejuni* cryptic plasmid pTIW94 recovered from wild birds in the southeastern United States. *Plasmid*, *70*(2), 268-71.

Hofreuter, D., Odenbreit, S., & Haas, R. (2001). Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Molecular Microbiology*, *41*(2), 379-391.

Humphrey, T., Mason, M., & Martin, K. (1995). The isolation of *Campylobacter jejuni* from contaminated surfaces and its survival in diluents. *International Journal of Food Microbiology*, *26*(3), 295-303.

Hvorup, R. N., Winnen, B., Chang, A. B., Jiang, Y., Zhou, X. F., & Saier, M. H. (2003). The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *European Journal of Biochemistry*. *270*(5), 799-813.

Igimi, S., Okada, Y., Ishiwa, A., Yamasaki, M., Morisaki, N., Kubo, Y., … Yamamoto, S. (2008). Antimicrobial resistance of *Campylobacter*: prevalence and trends in Japan. *Food Additives & Contaminants. Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment*, *25*(9), 1080-3.

Iraola, G., Pérez, R., Naya, H., Paolicchi, F., Pastor, E., Valenzuela, S., … Morsella, C. (2014). Genomic evidence for the emergence and evolution of pathogenicity and niche preferences in the genus *Campylobacter*. *Genome Biology and Evolution*, *6*(9), 2392-405.

Iyer, L. M., Koonin, E. V, & Aravind, L. (2002). Extensive domain shuffling in transcription regulators of DNA viruses and implications for the origin of fungal APSES transcription factors. *Genome Biology*, *3*(3), RESEARCH0012.

Iyer, L. M., Leipe, D. D., Koonin, E. V., & Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *Journal of Structural Biology*, *146*(1-2), 11-31.

Izard, T. (2001). Structural basis for chloramphenicol tolerance in *Streptomyces venezuelae* by chloramphenicol phosphotransferase activity. *Protein Science*, *10*(8), 1508-1513.

J Sacks, S Lieb *et al.,* (1986), Epidemic Campylobacteriosis Associated with a Community Water Supply, American Journal of Public Health. 76 (4) 424-428.

J.L., S., & P.M., F. (2010). Fluoroquinolone resistance in *Campylobacter*. *Journal of Food Protection*, *73*(6), 1141-1152.

Jackson, R. W., Vinatzer, B., Arnold, D. L., Dorus, S., & Murillo, J. (2011). The influence of the accessory genome on bacterial pathogen evolution. *Mobile Genetic Elements*, *1*(1), 55-65.

Jagannathan, A., Constantinidou, C., & Penn, C. W. (2001). Roles of rpoN, fliA, and flgR in expression of flagella in *Campylobacter jejuni*. *Journal of Bacteriology*, *183*(9), 2937-42.

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*(1), 230.

Jesse, T. W., Pittenger-Alley, L. G., & Englen, M. D. (2006). Sequence analysis of two cryptic plasmids from an agricultural isolate of *Campylobacter coli*. *Plasmid*, *55*(1), 64-9.

Jin, S., Joe, A., Lynett, J., Hani, E. K., Sherman, P., & Chan, V. L. (2004). JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. *Molecular Microbiology*, *39*(5), 1225-1236.

Jin, S., Joe, A., Lynett, J., Hani, E.K., Sherman, P., Chan, V.L., (2001). JlpA, a novel surface- exposed lipoprotein specificto *Campylobacter jejuni*, mediates adherence to host epithelial cells. Mol. Microbiol. 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 90alpha and triggers signalling pathways leading to the activation of NF-kappaB and p38 MAP kinase in epithelial cells. *Cellular Microbiology*, *5*(3), 165-174.

Jolley, K. A., Maiden, M. C., Pettersson, E., Lundeberg, J., Ahmadian, A., Roumagnac, P., … Mavromatis, K. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, *11*(1), 595.

Jones, F. S., Orcutt, M., & Little, R. B. (1931). Vibrios (*Vibrio jejuni*, n. sp.) associated with intestinal disorders of cows and calves. The Journal of Experimental Medicine, 53(6), 853–864.

Jones, M. A., Marston, K. L., Woodall, C. A., Maskell, D. J., Linton, D., Karlyshev, A. V, … Barrow, P. A. (2004). Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract. *Infection and Immunity*, *72*(7), 3769-76.

Joshua, G. W. P., Guthrie-Irons, C., Karlyshev, A. V., & Wren, B. W. (2006). Biofilm formation in *Campylobacter jejuni*. *Microbiology*, *152*(2), 387-396.

Kado, C. I. (2000). The role of the T-pilus in horizontal gene transfer and tumorigenesis. *Current Opinion in Microbiology*. 3(6), 643-648.

Kalkum, M., Eisenbrandt, R., Lurz, R., & Lanka, E. (2002). Tying rings for sex. *Trends in Microbiology*. 10(8), 382-387.

Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C., & Manning, G. (2007). Structural and functional diversity of the microbial kinome. *PLoS Biology*, *5*(3), 0467-0478.

Karlyshev, A. V, & Wren, B. W. (2005). Development and application of an insertional system for gene delivery and expression in *Campylobacter jejuni*. *Applied and Environmental Microbiology*, *71*(7), 4004-4013.

Keck, J. L. (2000). Structure of the RNA Polymerase Domain of *E. coli* Primase. *Science*, *287*(5462), 2482-2486.

Kim, C., & Mobashery, S. (2005). Phosphoryl transfer by aminoglycoside 3-phosphotransferases and manifestation of antibiotic resistance. *Bioorganic Chemistry*. 33(3), 149-158.

Kim, S. J., Jun, D. Y., Yang, C. H., & Kim, Y. H. (2006). Expression of *Helicobacter pylori* cag12 gene in *Lactococcus lactis* MG1363 and its oral administration to induce systemic anti-Cag12 immune response in mice. *Applied Microbiology and Biotechnology*, *72*(3), 462-470.

Kim, Y., Wang, X., Ma, Q., Zhang, X. S., & Wood, T. K. (2009). Toxin-antitoxin systems in *Escherichia coli* influence biofilm formation through YjgK (TabA) and fimbriae. *Journal of Bacteriology*, *191*(4), 1258-1267.

Klančnik, A., Guzej, B., Jamnik, P., Vučković, D., Abram, M., & Možina, S. S. (2009). Stress response and pathogenic potential of *Campylobacter jejuni* cells exposed to starvation. *Research in Microbiology*, *160*(5), 345-352.

Konkel, M. E., Kim, B. J., Klena, J. D., Young, C. R., & Ziprin, R. (1998). Characterization of the thermal stress response of *Campylobacter jejuni*. *Infection and Immunity*, *66*(8), 3666-72.

Konkel, M. E., Kim, B. J., Rivera-Amill, V., & Garvis, S. G. (1999). Bacterial secreted proteins are required for the internaliztion of *Campylobacter jejuni* into cultured mammalian cells. *Molecular Microbiology*, *32*(4), 691-701.

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*(11), 3296-3303.

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*(1), 68-76.

Korlath, J. A., Osterholm, M. T., Judy, L. A., Forfang, J. C., & Robinson, R. A. (1985). A point-source outbreak of Campylobacteriosis associated with consumption of raw milk. *Journal of Infectious Diseases*, *152*(3), 592-596.

Krause-Gruszczynska, M., van Alphen, L. B., Oyarzabal, O. A., Alter, T., Hänel, I., Schliephake, A., … Backert, S. (2007). Expression patterns and role of the CadF protein in *Campylobacter jejuni* and *Campylobacter coli*. *FEMS Microbiology Letters*, *274*(1), 9-16.

Kumar, R. B., & Das, A. (2001). Functional analysis of the Agrobacterium tumefaciens T-DNA transport pore protein VirB8. *Journal of Bacteriology*, *183*(12), 3636-3641.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357-359.

Lau, S.K. (2014), Molecular interrogation of *Campylobacter* infection. PhD Thesis. University of Exeter.

Laursen, B. S., Mortensen, K. K., Sperling-Petersen, H. U., & Hoffman, D. W. (2003). A conserved structural motif at the N terminus of bacterial translation initiation factor IF2. *Journal of Biological Chemistry*, *278*(18), 16320-16328.

Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. (2011). Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*, *12*(1), 124.

Lederber, J., & Tatum, E. L. (1953). Sex in bacteria; genetic studies, 1945-1952. *Science*, *118*(3059), 169-175.

Lefébure, T., Bitar, P. D. P., Suzuki, H., & Stanhope, M. J. (2010). Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. *Genome Biology and Evolution*, *2*(1), 646-655.

Lepp, D., Roxas, B., Parreira, V. R., Marri, P. R., Rosey, E. L., Gong, J., … Prescott, J. F. (2010). Identification of Novel Pathogenicity Loci in *Clostridium perfringens* Strains That Cause Avian Necrotic Enteritis. *PLoS ONE*, *5*(5), e10795.

Lertpiriyapong, K., Gamazon, E. R., Feng, Y., Park, D. S., Pang, J., Botka, G., … Fox, J. G. (2012). *Campylobacter jejuni* type VI secretion system: Roles in

adaptation to deoxycholic acid, host cell adherence, invasion, and in vivo colonization. *PLoS ONE*, *7*(8), e42842.

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

Lesnick, M. L., Reiner, N. E., Fierer, J., & Guiney, D. G. (2001). The Salmonella spvB virulence gene encodes an enzyme that ADP-ribosylates actin and destabilizes the cytoskeleton of eukaryotic cells. *Molecular Microbiology*, *39*(6), 1464-1470.

Lessl, M., Balzer, D., Pansegrau, W., & Lanka, E. (1992). Sequence similarities between the RP4 Tra2 and the Ti VirB region strongly support the conjugation model for T-DNA transfer. *Journal of Biological Chemistry*, *267*(28), 20471-20480.

Letunic, I., Doerks, T., & Bork, P. (2015). SMART: Recent updates, new developments and status in 2015. *Nucleic Acids Research*, *43*(D1), D257-D260.

Leuko, S., & Raivio, T. L. (2012). Mutations that impact the enteropathogenic *Escherichia coli* Cpx envelope stress response attenuate virulence in galleria mellonella. *Infection and Immunity*, *80*(9), 3077-3085.

Levengood, S. K., & Webster, R. E. (1989). Nucleotide sequences of the tolA and *tolB* genes and localization of their products, components of a multistep translocation system in *Escherichia coli*. *Journal of Bacteriology*, *171*(12), 6600-6609.

Lewis, K. (2007). Persister cells, dormancy and infectious disease. *Nature Reviews Microbiology*, *5*(1), 48-56.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Li, J., Adams, V., Bannam, T. L., Miyamoto, K., Garcia, J. P., Uzal, F. A., … McClane, B. A. (2013). Toxin plasmids of *Clostridium perfringens*. *Microbiology and Molecular Biology Reviews: MMBR*, *77*(2), 208-33.

Li, L., Chen, O. S., Ward, D. M., & Kaplan, J. (2001). CCC1 Is a Transporter That Mediates Vacuolar Iron Storage in Yeast. *Journal of Biological Chemistry*, *276*(31), 29515-29519.

Lin, A. E., Krastel, K., Hobb, R. I., Thompson, S. A., Cvitkovitch, D. G., & Gaynor, E. C. (2009). Atypical roles for *Campylobacter jejuni* amino acid ATP binding cassette transporter components PaqP and PaqQ in bacterial stress tolerance and pathogen-host cell dynamics. *Infection and Immunity*, *77*(11), 4912-4924.

Lin, J., Michel, L. O., & Zhang, Q. (2002). CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*. *Antimicrobial Agents and Chemotherapy*, *46*(7), 2124-31.

Lipps, G. (2008). Plasmids: current research and future trends. Horizon Scientific Press.

Lopper, M., Holton, J. M., & Keck, J. L. (2004). Crystal structure of PriB, a component of the *Escherichia coli* replication restart primosome. *Structure*, *12*(11), 1967-1975.

Louwen, R. P. L., Van Belkum, a., Wagenaar, J. a., Doorduyn, Y., Achterberg, R., & Endtz, H. P. (2006). Lack of association between the presence of the pVir plasmid and bloody diarrhea in *Campylobacter jejuni* enteritis. *Journal of Clinical Microbiology*, *44*(5), 1867-1868.

Louwen, R., Horst-Kreft, D., De Boer, A. G., Van Der Graaf, L., De Knegt, G., Hamersma, M., … Van Belkum, A. (2013). A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barré syndrome. *European Journal of Clinical Microbiology and Infectious Diseases*, *32*(2), 207-226.

Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P., & van der Oost, J. (2014). The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiology and Molecular Biology Reviews: MMBR*, *78*(1), 74-88.

Luangtongkum, T., Jeon, B., Han, J., Plummer, P., Logue, C. M., & Zhang, Q. (2010). Antibiotic resistance in *Campylobacter*: emergence, transmission and persistence. *Future Microbiology*, *4*(2), 189-200.

Luangtongkum, T., Morishita, T. Y., Ison, A. J., Huang, S., McDermott, P. F., & Zhang, Q. (2006). Effect of Conventional and Organic Production Practices on the Prevalence and Antimicrobial Resistance of *Campylobacter* spp. in Poultry. *Applied and Environmental Microbiology*, *72*(5), 3600-3607.

Luo, N., & Zhang, Q. (2001). Molecular characterization of a cryptic plasmid from *Campylobacter jejuni*. *Plasmid*, *45*(2), 127-133.

Luscombe, N. M., Austin, S. E., Berman, H. M., & Thornton, J. M. (2000). An overview of the structures of protein-DNA complexes. *Genome Biology*, *1*(1), REVIEWS001.

Makarova, K. S., Grishin, N. V, & Koonin, E. V. (2006). The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria. *Bioinformatics*, *22*(21), 2581-2584.

Makarova, K. S., Grishin, N. V, Shabalina, S. a, Wolf, Y. I., & Koonin, E. V. (2006 (2)). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, *1*(1), 7.

Manival, X., Yang, Y., Strub, M. P., Kochoyan, M., Steinmetz, M., & Aymerich, S. (1997). From genetic to structural characterization of a new class of RNA-binding domain within the SacY/BgIG family of antiterminator proteins. *EMBO Journal*, *16*(16), 5019-5029.

Marasini, D., & Fakhr, M. K. (2014). Exploring PFGE for Detecting Large Plasmids in *Campylobacter jejuni* and *Campylobacter coli* Isolated from Various Retail Meats. *Pathogens (Basel, Switzerland)*, *3*(4), 833-844.

Marasini, D., & Fakhr, M. K. (2016). Whole-Genome Sequencing of a *Campylobacter jejuni* Strain Isolated from Retail Chicken Meat Reveals the Presence of a Megaplasmid with Mu-Like Prophage and Multidrug Resistance Genes. *Genome Announcements*, *4*(3), e00460-16.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., … Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*(D1), D222-D226.

Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature*, *526*(7571), 55-61.

Marsin, S., Mathieu, A., Kortulewski, T., Guérois, R., & Radicella, J. P. (2008). Unveiling novel RecO distant orthologues involved in homologous recombination. *PLoS Genetics*, *4*(8), e1000146.

Martins, A., Spengler, G., Molnár, J., & Amaral, L. (2001). Bacterial Antibiotic Resistance. In *eLS*. CHAP, John Wiley & Sons, Ltd.

Matthews, J. M., & Visvader, J. E. (2003). LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Reports*, *4*(12), 1132-1137.

Mazi, W., Senok, A., Al-Mahmeed, A., Arzese, A., Bindayna, K., & Botta, G. (2008). Trends in Antibiotic Sensitivity Pattern and Molecular Detection of tet(O)-Mediated Tetracycline Resistance in *Campylobacter jejuni* Isolates from Human and Poultry Sources. *Jpn. J. Infect. Dis*, *61*, 82-84.

McDermott, P. F., Bodeis, S. M., English, L. L., White, D. G., Walker, R. D., Zhao, S., … Wagner, D. D. (2002). Ciprofloxacin resistance in *Campylobacter jejuni* evolves rapidly in chickens treated with fluoroquinolones. *The Journal of Infectious Diseases*, *185*(6), 837-40.

McFarland, B. A., & Neill, S. D. (1992). Profiles of toxin production by thermophilic *Campylobacter* of animal origin. *Veterinary Microbiology*, *30*(2), 257-266.

Méric, G., Yahara, K., Mageiros, L., Pascoe, B., Maiden, M. C. J., Jolley, K. A., & Sheppard, S. K. (2014). A reference pan-genome approach to comparative bacterial genomics: Identification of novel epidemiological markers in pathogenic *Campylobacter*. *PLoS ONE*, *9*(3), e92798.

Meuwissen, R. L., Offenberg, H. H., Dietrich, A. J., Riesewijk, A., van Iersel, M., & Heyting, C. (1992). A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes. *Embo J*, *11*(13), 5091-5100.

Meyer, R. (2000). Identification of the *mob* genes of plasmid pSC101 and characterization of a hybrid pSC101-R1162 system for conjugal mobilization. *Journal of Bacteriology*, *182*(17), 4875-4881.

Mickelson, C., & Wiberg, J. S. (1981). Membrane-associated DNase activity controlled by genes 46 and 47 of bacteriophage T4D and elevated DNase activity associated with the T4 das mutation. *Journal of Virology*, *40*(1), 65-77.

Mihaljevic, R. R., Sikic, M., Klancnik, A., Brumini, G., Mozina, S. S., & Abram, M. (2007). Environmental stress factors affecting survival and virulence of *Campylobacter jejuni*. *Microbial Pathogenesis* (Vol. 43).

Miles, A. A., Misra, S. S., & Irwin, J. O. (1938). The estimation of the bactericidal power of the blood. The Journal of Hygiene, 38(6), 732–749.

Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, *95*(6), 315-27.

Miller, W. G., Heath, S., & Mandrell, R. E. (2007). Cryptic plasmids isolated from *Campylobacter* strains represent multiple, novel incompatibility groups. *Plasmid*, *57*(2), 108-17.

Misawa, N., & Blaser, M. J. (2000). Detection and characterization of autoagglutination activity by *Campylobacter jejuni*. *Infection and Immunity*, *68*(11), 6168-75.

Miyamoto, K., Fisher, D. J., Li, J., Sayeed, S., Akimoto, S., & McClane, B. A. (2006). Complete sequencing and diversity analysis of the enterotoxin-encoding plasmids in *Clostridium perfringens* type A non-food-borne human gastrointestinal disease isolates. *Journal of Bacteriology*, *188*(4), 1585-1598.

Monteiro, C., Fang, X., Ahmad, I., Gomelsky, M., & Römling, U. (2011). Regulation of biofilm components in *Salmonella enterica* serovar typhimurium by lytic transglycosylases involved in cell wall turnover. *Journal of Bacteriology*, *193*(23), 6443-6451.

Monteville, M. R. (2003). Maximal adherence and invasion of INT 407 cells by *Campylobacter jejuni* requires the CadF outer-membrane protein and microfilament reorganization. *Microbiology*, *149*(1), 153-165.

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, 153e165.

Moore, J. E. (2001). Bacterial dormancy in *Campylobacter*: abstract theory or cause for concern? *International Journal of Food Science and Technology*, *36*(6), 593-600.

Moore, J. E., Barton, M. D., Blair, I. S., Corcoran, D., Dooley, J. S. G., Fanning, S., … Tolba, O. (2006). The epidemiology of antibiotic resistance in *Campylobacter*. *Microbes and Infection*, *8*(7), 1955-1966.

Moreira, C. G., Palmer, K., Whiteley, M., Sircili, M. P., Trabulsi, L. R., Castro, A. F. P., & Sperandio, V. (2006). Bundle-Forming Pili and EspA Are Involved in Biofilm Formation by Enteropathogenic *Escherichia coli*. *Journal of Bacteriology*, *188*(11), 3952-3961.

Morooka, T., Umeda, A., & Amako, K. (1985). Motility as an Intestinal Colonization Factor for *Campylobacter jejuni*. *Microbiology*, *131*(8), 1973-1980.

Muzzin, O., Campbell, E. a., Xia, L., Severinova, E., Darst, S. a., & Severinov, K. (1998). Disruption of *Escherichia coli* HepA, an RNA polymerase-associated protein, causes UV sensitivity. *Journal of Biological Chemistry*, *273*(24), 15157-15161.

Nachamkin, I., Allos, B. M., & Ho, T. (1998). *Campylobacter* species and Guillain-Barré syndrome. *Clin Microbiol Rev*. 11, 555-567.

Nachamkin, I., Liu, J., Li, M., Ung, H., Moran, A. P., Prendergast, M. M., & Sheikh, K. (2002). *Campylobacter jejuni* from patients with Guillain-Barre syndrome preferentially expresses a GD1a-like epitope. Infection and Immunity, 70(9), 5299–5303.

Nachamkin, I., Yang, X. H., & Stern, N. J. (1993). Role of *Campylobacter jejuni* flagella as colonization factors for three-day-old chicks: analysis with flagellar mutants. *Applied and Environmental Microbiology*, *59*(5), 1269-1273.

Naito, M., Frirdich, E., Fields, J. A., Pryjma, M., Li, J., Cameron, A., … Gaynor, E. C. (2010). Effects of sequential *Campylobacter jejuni* 81-176 lipooligosaccharide core truncations on biofilm formation, stress survival, and pathogenesis. *Journal of Bacteriology*, *192*(8), 2182-2192.

Nakjang, S., Williams, T. A., Heinz, E., Watson, A. K., Foster, P. G., Sendra, K. M., … Martin Embley, T. (2013). Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. *Genome Biology and Evolution*, *5*(12), 2285-2303.

Nam, K. H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M. P., & Ke, A. (2012). Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *Journal of Biological Chemistry*, *287*(43), 35943-35952.

Nelson, J. M., Chiller, T. M., Powers, J. H., & Angulo, F. J. (2007). Fluoroquinolone-Resistant *Campylobacter* Species and the Withdrawal of Fluoroquinolones from

Use in Poultry: A Public Health Success Story. *Clinical Infectious Diseases*, *44*(7), 977-980.

Neuwald, A. F., & Landsman, D. (1997). GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends in Biochemical Sciences*. 22(5), 154-5.

Nguyen, C. C., & Saier, M. H. (1995). Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. *FEBS Letters*, *377*(2), 98-102.

Niedenzu, T., Roleke, D., Bains, G., Scherzinger, E., & Saenger, W. (2001). Crystal structure of the hexameric replicative helicase RepA of plasmid RSF1010. *Journal of Molecular Biology*, *306*(3), 479-487.

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.*, *67*(4), 593-656.

Nirdnoy, W., Mason, C. J., & Guerry, P. (2005). Mosaic structure of a multiple-drug-resistant, conjugative plasmid from *Campylobacter jejuni*. *Antimicrobial Agents and Chemotherapy*, *49*(6), 2454-9.

Nizan, R., Barash, I., Valinsky, L., Lichter, A., & Manulis, S. (1997). The Presence of *hrp* Genes on the Pathogenicity-Associated Plasmid of the Tumorigenic Bacterium *Erwinia herbicola* pv. *gypsophilae*. *Molecular Plant-Microbe Interactions*, *10*(5), 677-682.

Norström, M., Johnsen, G., Hofshagen, M., Tharaldsen, H., & Kruse, H. (2007). Antimicrobial resistance in *Campylobacter jejuni* from broilers and broiler house environments in Norway. *Journal of Food Protection*, *70*(3), 736-8.

Nuijten PJM, Bleumink-Pluym NMC, Gaastra W & Zeijst BAM van der (1989) Flagellin gene expression in *Campylobacter jejuni* is regulated at the transcriptional level. Infect. Immun. 57: 1084-1088

Nyengaard, N. R., Mortensen, K. K., Lassen, S. F., Hershey, J. W., & Sperling-Petersen, H. U. (1991). Tandem translation of *E. coli* initiation factor IF2 beta: Purification and characterization in vitro of two active forms. *Biochemical and Biophysical Research Communications*, *181*(3), 1572-1579.

Oggioni, M. R., Dowson, C. G., Smith, J. M., Provvedi, R., & Pozzi, G. (1996). The tetracycline resistance gene *tet(M)* exhibits mosaic structure. *Plasmid*, *35*(3), 156-163.

Oh, E., & Jeon, B. (2014). Role of alkyl hydroperoxide reductase (AhpC) in the biofilm formation of *Campylobacter jejuni*. *PLoS ONE*, *9*(1).

Oh, E., Kim, J.C., & Jeon, B. (2016). Stimulation of biofilm formation by oxidative stress in *Campylobacter jejuni* under aerobic conditions. *Virulence*, *7*(7), 846-851.

Oosterom, J., De Wilde, G. J. A., De Boer, E., De Blaauw, L. H., & Karman, H. (1983). Survival of *Campylobacter jejuni* during Poultry Processing and Pig Slaughtering. *Journal of Food Protection*, *46*(8), 702-706.

Oudot, M. P., Kloareg, B., & Loiseaux-de Goër, S. (1999). The mitochondrial *Pylaiella littoralis* nad11 gene contains only the N-terminal FeS-binding domain. *Gene*, *235*(1-2), 131-137.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., … Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1), D206-14.

Ozer, E. A., Allen, J. P., Hauser, A. R., Silby, M., Winstanley, C., Godfrey, S., … Edgar, R. (2014). Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics*, *15*(1), 737.

Page, R., & Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat Chem Biol*, *12*(4), 208-214.

Palyada, K., Sun, Y.-Q., Flint, A., Butcher, J., Naikare, H., & Stintzi, A. (2009). Characterization of the oxidative stress stimulon and PerR regulon of *Campylobacter jejuni*. *BMC Genomics*, *10*, 481.

Palyada, K., Threadgill, D., & Stintzi, A. (2004). Iron Acquisition and Regulation in *Campylobacter jejuni*. *Journal of Bacteriology*, *186*(14), 4714-4729.

Park, S. F. (2005). *Campylobacter jejuni* stress responses during survival in the food chain and colonization. *Campylobacter* molecular and cellular biology. Horizon Bioscience, Wymondham, 311-330.

Park, S., Tian, G., Roelofs, J., & Finley, D. (2010). Assembly manual for the proteasome regulatory particle: the first draft. *Biochemical Society Transactions*, *38*(Pt 1), 6-13.

Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., … Barrell, B. G. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*, *403*(6770), 665-8.

Parsons, J. F., Lim, K., Tempczyk, A., Krajewski, W., Eisenstein, E., & Herzberg, O. (2002). From structure to function: YrbI from *Haemophilus influenzae* (HI1679) is a phosphatase. *Proteins: Structure, Function and Genetics*, *46*(4), 393-404.

Pavlovskis, O. R., Rollins, D. M., Haberberger, R. L., Green, A. E., Habash, L., Strocko, S., & Walker, R. I. (1991). Significance of flagella in colonization resistance of rabbits immunized with *Campylobacter* spp. *Infection and Immunity*, *59*(7), 2259-64.

Perez, J., Castaneda-Garcia, a, Jenke-Kodama, H., Muller, R., & Munoz-Dorado, J. (2008). Eukaryotic-like protein kinases in the prokaryotes and the myxobacterial kinome. *Proc Natl Acad Sci U S A*, *105*(41), 15950-15955.

Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J. D., Völker, U., & Heckler, M. (2001). Global analysis of the general stress response of *Bacillus subtilis*. *Journal of Bacteriology*, *183*(19), 5617-5631.

Pickett, C. L., & Whitehouse, C. A. (1999). The cytolethal distending toxin family. *Trends in Microbiology*, *7*(7), 292-297.

Pickett, C.L., Auffenberg, T., Pesci, E.C., Sheen, V.L., Jusuf, S.S. (1992). Iron acquisition and hemolysin production by *Campylobacter jejuni*. *Infection and Immunity,* 60, 3872-3877.

Pijlman, G. P., Pruijssers, A. J. P., & Vlak, J. M. (2003). Identification of pif-2, a third conserved baculovirus gene required for per os infection of insects. *Journal of General Virology*, *84*(8), 2041-2049.

Poly, F., & Guerry, P. (2008). Pathogenesis of *Campylobacter*. *Current Opinion in Gastroenterology*, *24*(1), 27-31.

Pratt, A., & Korolik, V. (2005). Tetracycline resistance of Australian *Campylobacter jejuni* and *Campylobacter coli* isolates. *Journal of Antimicrobial Chemotherapy*, *55*(4), 452-460.

Price, L. B., Lackey, L. G., Vailes, R., & Silbergeld, E. (2007). The persistence of fluoroquinolone-resistant *Campylobacter* in poultry production. *Environmental Health Perspectives*, *115*(7), 1035-9.

Purdy, D., Cawthraw, S., Dickinson, J. H., Newell, D. G., & Park, S. F. (1999). Generation of a superoxide dismutase (SOD)-deficient mutant of *Campylobacter coli*: Evidence for the significance of SOD in *Campylobacter* survival and colonization. *Applied and Environmental Microbiology*, *65*(6), 2540-2546.

Putman, M., van Veen, H. W., & Konings, W. N. (2000). Molecular properties of bacterial multidrug transporters. *Microbiology and Molecular Biology Reviews : MMBR*, *64*(4), 672-93.

Quick, M., Yano, H., Goldberg, N. R., Duan, L., Beuming, T., Shi, L., … Javitch, J. A. (2006). State-dependent conformations of the translocation pathway in the tyrosine transporter Tyt1, a novel neurotransmitter:sodium symporter from Fusobacterium nucleatum. *Journal of Biological Chemistry*, *281*(36), 26444-26454.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841-842.

Ratledge, C., & Dover, L. G. (2000). Iron Metabolism in Pathogenic Bacteria. *Annual Review of Microbiology*, *54*(1), 881-941.

Reeser, R. J., Medler, R. T., Billington, S. J., Jost, B. H., & Joens, L. A. (2007). Characterization of *Campylobacter jejuni* biofilms under defined growth conditions. *Applied and Environmental Microbiology*, *73*(6), 1908-1913.

Reid, A. N., Pandey, R., Palyada, K., Naikare, H., & Stintzi, A. (2008) (2). Identification of *Campylobacter jejuni* genes involved in the response to acidic pH and stomach transit. *Applied and Environmental Microbiology*, *74*(5), 1583-1597.

Reid, A. N., Pandey, R., Palyada, K., Whitworth, L., Doukhanine, E., & Stintzi, A. (2008) (1). Identification of *Campylobacter jejuni* genes contributing to acid adaptation by transcriptional profiling and genome-wide mutagenesis. *Applied and Environmental Microbiology*, *74*(5), 1598-1612.

Reizer, J., Reizer, A., & Saier, M. H. (1992). A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. *Protein Science : A Publication of the Protein Society*, *1*(10), 1326-1332.

Ren, J., Sainsbury, S., Nettleship, J. E., Saunders, N. J., & Owens, R. J. (2010). The crystal structure of NGO0477 from Neisseria gonorrhoeae reveals a novel protein fold incorporating a helix-turn-helix motif. *Proteins*, *78*(7), 1798-1802.

Reuter, M., & van Vliet, A. H. M. (2013). Signal balancing by the CetABC and CetZ chemoreceptors controls energy taxis in *Campylobacter jejuni*. *PloS One*, *8*(1), e54390.

Reuter, M., Mallett, A., Pearson, B. M., & van Vliet, A. H. M. (2010). Biofilm formation by *Campylobacter jejuni* is increased under aerobic conditions. *Applied and Environmental Microbiology*, *76*(7), 2122-8.

Ridley, A. M., Toszeghy, M. J., Cawthraw, S. A., Wassenaar, T. M., & Newell, D. G. (2008). Genetic instability is associated with changes in the colonization potential of *Campylobacter jejuni* in the avian intestine. *Journal of Applied Microbiology*, *105*(1), 95-104.

Roberts, M. C. (2005). Update on acquired tetracycline resistance genes. *FEMS Microbiology Letters*. 245(2), 195-203.

Robertson, H. M., & Thomas, J. H. (2006). The putative chemoreceptor families of C. elegans. In *WormBook: the online review of C. elegans biology* (pp. 1-12).

Roca, J. (1995). The mechanisms of DNA topoisomerases. Trends in Biochemical Sciences, 20(4), 156–160.

Ruiz-Palacios, G. M. (2007). The health burden of *Campylobacter* infection and the impact of antimicrobial resistance: playing chicken. *Clinical Infectious Diseases:*

An Official Publication of the Infectious Diseases Society of America, *44*(5), 701-3.

Rychlik, I., Gregorova, D., & Hradecka, H. (2006). Distribution and function of plasmids in *Salmonella enterica*. *Veterinary Microbiology*, *112*(1), 1-10.

Sadhale, P. P., & Woychik, N. a. (1994). C25, an essential RNA polymerase III subunit related to the RNA polymerase II subunit RPB7. *Molecular and Cellular Biology*, *14*(9), 6164-6170.

Saikrishnan, K., Jeyakanthan, J., Venkatesh, J., Acharya, N., Sekar, K., Varshney, U., & Vijayan, M. (2003). Structure of *Mycobacterium tuberculosis* single-stranded DNA-binding protein. Variability in quaternary structure and its implications. *Journal of Molecular Biology*, *331*(2), 385-393.

Sasaki, Y., Ishikawa, J., Yamashita, A., Oshima, K., Kenri, T., Furuya, K., … Hattori, M. (2002). The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. *Nucleic Acids Research*. Oxford University Press.

Schmidt, R., Ahmetagic, A., Philip, D. S., & Pemberton, J. M. (2001). Catabolic Plasmids. *Encyclopaedia of life sciences*, DOI: 10.1002/9780470015902.a0000471.pub3.

Schmidt-Ott, R., Pohl, S., Burghard, S., Weig, M., & Gross, U. (2005). Identification and characterization of a major subgroup of conjugative *Campylobacter jejuni* plasmids. *The Journal of Infection*, *50*(1), 12-21.

Schulein, R., & Dehio, C. (2002). The VirB/VirD4 type IV secretion system of *Bartonella* is essential for establishing intraerythrocytic infection. *Molecular Microbiology*, *46*(4), 1053-1067.

Schulze, F., Hänel, I., & Borrmann, E. (1998). Formation of cytotoxins by enteric *Campylobacter* in humans and animals. *Zentralblatt Für Bakteriologie*, *288*(2), 225-236.

Senior, N. J., Bagnall, M. C., Champion, O. L., Reynolds, S. E., la Ragione, R. M., Woodward, M. J., … Titball, R. W. (2011). *Galleria mellonella* as an infection model for *Campylobacter jejuni* virulence. *Journal of Medical Microbiology*, *60*(5), 661-669.

Sheppard, S. K., Colles, F. M., McCarthy, N. D., Strachan, N. J. C., Ogden, I. D., Forbes, K. J., … Maiden, M. C. J. (2011). Niche segregation and genetic structure of *Campylobacter jejuni* populations from wild and agricultural host species. *Molecular Ecology*, *20*(16), 3484-3490.

Sheppard, S. K., Didelot, X., Meric, G., Torralbo, A., Jolley, K. A., Kelly, D. J., … Falush, D. (2013). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *Proceedings of the*

National Academy of Sciences of the United States of America, *110*(29), 11923-11927.

Shortt, C., Scanlan, E., Hilliard, A., Cotroneo, C. E., Bourke, B., & Cróinín, Ó. (2016). DNA Supercoiling Regulates the Motility of *Campylobacter jejuni* and Is Altered by Growth in the Presence of Chicken Mucus. *mBio*, *7*(5), e01227-16.

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.

Smith, M. C. M., & Thorpe, H. M. (2002). Diversity in the serine recombinases. *Molecular Microbiology*. 44(2), 299-307.

Smole Možina S., Kurincic, C. M., Kramar, A., Uršicc, S., and Katalinic, V. (2009). Prevalence and resistance against different antimicrobial compounds of *Campylobacter* spp. in/from retail poultry. *Meat Technology*. 50, 112-120.

Solomon, E. B., Hoover, D.G. (1999). *Campylobacter jejuni*: a bacterial paradox. *Journal of Food Safety*. 19, 121-136.

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(17), 7194–7199.

Storz, G., & Imlayt, J. A. (1999). Oxidative stress. *Current Opinion in Microbiology*, *2*(2), 188-194.

Sundin, G. W. (2007). Genomic Insights into the Contribution of Phytopathogenic Bacterial Plasmids to the Evolutionary History of Their Hosts. *Annual Review of Phytopathology*, *45*(1), 129-151.

Szymanski, C. M., King, M., Haardt, M., & Armstrong, G. D. (1995). *Campylobacter jejuni* motility and invasion of Caco-2 cells. *Infection and Immunity*, *63*(11), 4295-300.

Takkinen J., Ammon A., Robstad O., Breuer T., The *Campylobacter* Working Group (2003) European survey on *Campylobacter* surveillance and diagnosis 2001. *European communicable disease bulletin* 8, 207-213.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731-2739.

Tanaka, H., Chinami, M., Mizushima, T., Ogasahara, K., Ota, M., Tsukihara, T., & Yutani, K. (2001). X-ray crystalline structures of pyrrolidone carboxyl peptidase from a hyperthermophile, *Pyrococcus furiosus*, and its cys-free mutant. *Journal of Biochemistry*, *130*(1), 107-118.

Tang, H., Wang, S., Ma, L., Meng, X., Deng, Z., Zhang, D., … Xu, P. (2008). A novel gene, encoding 6-hydroxy-3-succinoylpyridine hydroxylase, involved in nicotine degradation by *Pseudomonas putida* strain S16. *Applied and Environmental Microbiology*, *74*(5), 1567-1574.

Tangney, M., & Fitzgerald, G. F. (2002). AbiA, a lactococcal abortive infection mechanism functioning in *Streptococcus thermophilus*. *Applied and Environmental Microbiology*, *68*(12), 6388-6391.

Taniguchi, Y., Katayama, M., Ito, R., Takai, N., Kondo, T., & Oyama, T. (2007). labA: A novel gene required for negative feedback regulation of the cyanobacterial circadian clock protein KaiC. *Genes and Development*, *21*(1), 60-70.

Taniguchi, Y., Takai, N., Katayama, M., Kondo, T., & Oyama, T. (2010). Three major output pathways from the KaiABC-based oscillator cooperate to generate robust circadian *kaiBC* expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(7), 3263-3268.

Tatusov, R. L., Koonin, E. V, Lipman, D. J., Fleischmann, R. D., Fraser, C. M., Himmelreich, R., … Koonin, E. V. (1997). A genomic perspective on protein families. *Science*, *278*(5338), 631-637.

Tauxe, R. V. (2002). Emerging foodborne pathogens. *International Journal of Food Microbiology*, *78*(1-2), 31-41.

Taveirne, M. E., Theriot, C. M., Livny, J., & DiRita, V. J. (2013). The complete *Campylobacter jejuni* transcriptome during colonization of a natural host determined by RNAseq. *PloS One*, *8*(8), e73586.

Teather, R. M., & Wood, P. J. (1982). Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Applied and Environmental Microbiology*, *43*(4), 777-780.

Teramoto, H., Shirai, T., Inui, M., & Yukawa, H. (2008). Identification of a gene encoding a transporter essential for utilization of C4 dicarboxylates in *Corynebacterium glutamicum*. *Applied and Environmental Microbiology*, *74*(17), 5290-5296.

Thomas, C. M. (2014). Plasmid Incompatibility. In E. Bell (Ed.), Molecular Life Sciences: An Encyclopedic Reference (pp. 1-3). Springer New York.

Thomas, C. M., & Summers, D. (2008). Bacterial Plasmids. In *eLS*. CHAP, John Wiley & Sons, Ltd.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178-192.

Todone, F., Brick, P., Werner, F., Weinzierl, R. O. J., & Onesti, S. (2001). Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Molecular Cell*, *8*(5), 1137-1143.

Top E.M., Moënne-Loccoz Y., Pembroke T. and Thomas C.M. (2000) Phenotypic traits conferred by plasmids. In: Thomas C.M. (ed.) *The Horizontal Gene Pool: Bacterial Plasmids and Gene Spread*, pp. 249-285. *Amsterdam: Harwood Academic Press*

Tracz, D. M., Keelan, M., Ahmed-Bentley, J., Gibreel, A., Kowalewska-Grochowska, K., & Taylor, D. E. (2005). pVir and bloody diarrhea in *Campylobacter jejuni* enteritis. *Emerging Infectious Diseases*, *11*(6), 838-843.

Trigui, H., Thibodeau, A., Fravalo, P., Letellier, A., & P. Faucher, S. (2015). Survival in water of *Campylobacter jejuni* strains isolated from the slaughterhouse. *SpringerPlus*, *4*(1), 799.

Tunio, S. A., Oldfield, N. J., Ala'Aldeen, D. A., Wooldridge, K. G., Turner, D. P., Caugant, D., … Ketley, J. (2010). The role of glyceraldehyde 3-phosphate dehydrogenase (GapA-1) in Neisseria meningitidis adherence to human cells. *BMC Microbiology*, *10*(1), 280.

Türck, M., & Bierbaum, G. (2012). Purification and activity testing of the full-length YycFGHI proteins of *Staphylococcus aureus*. *PLoS ONE*, *7*(1), e30403.

Van der Woude, M. W., & Bäumler, A. J. (2004). Phase and antigenic variation in bacteria. *Clinical Microbiology Reviews*, *17*(3), 581-611.

Van Melderen, L., & Saavedra De Bast, M. (2009). Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genetics*, *5*(3), e1000437.

Van Vliet, A. H. M., Wood, A. C., Wooldridge, K., Ketley, J. M., & Henderson, J. (1998). Genetic Manipulation of Enteric *Campylobacter* Species. *Methods in Microbiology*, 27(C), 407–419.

Van Vliet, A. H., & Ketley, J. M. (2001). Pathogenesis of enteric *Campylobacter* infection. *Journal of Applied Microbiology, 90*, 45S-56S.

Vegge, C. S., Brøndsted, 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*(16), 5308-5314.

Vegge, C. S., Brøndsted, L., Ligowska-Marzeta, M., & Ingmer, H. (2012). Natural Transformation of *Campylobacter jejuni* Occurs Beyond Limits of Growth. *PLoS ONE*, *7*(9), e45467.

Vegge, C. S., Brøndsted, L., Ligowska-Marzeta, M., & Ingmer, H. (2012). Natural Transformation of *Campylobacter jejuni* Occurs Beyond Limits of Growth. *PLoS ONE*, *7*(9), e45467.

Vestal, D. J., & Jeyaratnam, J. A. (2011). The guanylate-binding proteins: emerging insights into the biochemical properties and functions of this family of large interferon-induced guanosine triphosphatase. *Journal of Interferon & Cytokine Research*, *31*(1), 89-97.

Vivian, A., Murillo, J., & Jackson, R. W. (2001). The roles of plasmids in phytopathogenic bacteria: Mobile arsenals? *Microbiology*. 147(4), 763-780

Wallden, K., Rivera-Calzada, A., & Waksman, G. (2010). Type IV secretion systems: Versatility and diversity in function. *Cellular Microbiology*. 12(9), 1203-1212

Wallis, T. S., Paulin, S. M., Plested, J. S., Watson, P. R., & Jones, P. W. (1995). The *Salmonella dublin* virulence plasmid mediates systemic but not enteric phases of Salmonellosis in cattle. *Infection and Immunity*, *63*(7), 2755-2761.

Wang, G., Clark, C. G., Taylor, T. M., Pucknell, C., Barton, C., Price, L., … Rodgers, F. G. (2002). Colony multiplex PCR assay for identification and differentiation of *Campylobacter jejuni*, *C. coli*, *C. lari*, *C. upsaliensis*, and C. fetus subsp. fetus. Journal of Clinical Microbiology, 40(12), 4744–4747.

Wang, X., & Wood, T. K. (2011). Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Applied and Environmental Microbiology*, *77*(16), 5577-5583.

Wang, Y., & Taylor, D. E. (1990). Natural transformation in *Campylobacter* species. *Journal of Bacteriology*, *172*(2), 949-955.

Ward, J. E., Dale, E. M., Christie, P. J., Nester, E. W., & Binns, A. N. (1990). Complementation analysis of Agrobacterium tumefaciens Ti plasmid virB genes by use of a vir promoter expression vector: virB9, virB10, and virB11 are essential virulence genes. *Journal of Bacteriology*, *172*(9), 5187-5199.

Wassenaar, T. M., Bleumink-Pluym, N. M., & van der Zeijst, B. A. (1991). Inactivation of *Campylobacter jejuni* flagellin genes by homologous recombination demonstrates that flaA but not flaB is required for invasion. *The EMBO Journal*, *10*(8), 2055-2061.

Wassenaar, T.M. (1997) Toxin production by *Campylobacter*. Clinical Microbiology Reviews 10, 466-476.

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, *31*(20), 3350-3352.

Wilkens, S. (2004). Structure and mechanism of ABC transporters. *Current Opinion in Structural Biology*. 14(4), 426-431.

Winter, E., & Ponting, C. P. (2002). TRAM, LAG1 and CLN8: Members of a novel family of lipid-sensing domains? *Trends in Biochemical Sciences*, 27(8), 381-383.

Wintjens, R., & Rooman, M. (1996). Structural Classification of HTH DNA-binding Domains and Protein--DNA Interaction Modes. *Journal of Molecular Biology*, *262*(2), 294-313.

Wood, T. K., Knabel, S. J., & Kwan, B. W. (2013). Bacterial persister cell formation and dormancy. *Applied and Environmental Microbiology*. *79*(23), 7116-7121.

Woodall, C. A., Jones, M. A., Barrow, P. A., Hinds, J., Marsden, G. L., Kelly, D. J., … Maskell, D. J. (2005). *Campylobacter jejuni* gene expression in the chick cecum: evidence for adaptation to a low-oxygen environment. *Infection and Immunity*, *73*(8), 5278-5285.

Woolbridge, K.G. & Ketley, J.M. (1997) *Campylobacter*-host cell interactions. *Trends in Microbiology. 5*, 96-102.

Wösten, M. M. S. M., Wagenaar, J. A., & van Putten, J. P. M. (2004). The FlgS/FlgR two-component signal transduction system regulates the *fla* regulon in *Campylobacter jejuni*. *The Journal of Biological Chemistry*, *279*(16), 16214-16222.

Wu, B., Gong, J., Liu, L., Li, T., Wei, T., & Bai, Z. (2012). Evolution of prokaryotic homologues of the eukaryotic SEFIR protein domain. *Gene*, *492*(1), 160-166.

Wu, K. H., & Tai, P. C. (2004). Cys32 and His105 Ate the Critical Residues for the Calcium-dependent Cysteine Proteolytic Activity of CvaB, an ATP-binding Cassette Transporter. *Journal of Biological Chemistry*, *279*(2), 901-909.

Xia, W., Li, H., Sze, K. H., & Sun, H. (2009). Structure of a nickel chaperone, HypA, from *Helicobacter pylori* reveals two distinct metal binding sites. *Journal of the American Chemical Society*, *131*(29), 10031-10040.

Xu, F., Zeng, X., Haigh, R. D., Ketley, J. M., & Lin, J. (2010). Identification and characterization of a new ferric enterobactin receptor, CfrB, in *Campylobacter*. *Journal of Bacteriology*, *192*(17), 4425-4435.

Xuan, Y. H., Hu, Y. B., Chen, L.-Q., Sosso, D., Ducat, D. C., Hou, B.-H., & Frommer, W. B. (2013). Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), E3685-3694.

Yamada, T., Lee, P., & Kosuge, T. (1986). Insertion sequence elements of *Pseudomonas savastanoi*: Nucleotide sequence and homology with *Agrobacterium tumefaciens* transfer DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(21), 8263-8267.

Yamaguchi, Y., & Inouye, M. (2011). Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nature Reviews. Microbiology*, *9*(11), 779-790.

Yamaguchi, Y., Park, J.-H., & Inouye, M. (2011). Toxin-Antitoxin Systems in Bacteria and Archaea. *Annual Review of Genetics*, *45*(1), 61-79.

Yeo, H.-J., Yuan, Q., Beck, M. R., Baron, C., & Waksman, G. (2003). Structural and functional characterization of the VirB5 protein from the type IV secretion system encoded by the conjugative plasmid pKM101. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15947-15952.

Yi, K., Rasmussen, A. W., Gudlavalleti, S. K., Stephens, D. S., & Stojiljkovic, I. (2004). Biofilm formation by Neisseria meningitidis. *Infection and Immunity*, *72*(10), 6132-6138.

Young, K. T., Davis, L. M., & Dirita, V. J. (2007). *Campylobacter jejuni*: molecular biology and pathogenesis. *Nature Reviews. Microbiology*, *5*(9), 665-679.

Yun, J., Jeon, B., Barton, Y.-W., Plummer, P., Zhang, Q., & Ryu, S. (2008). Role of the DksA-like protein in the pathogenesis and diverse metabolic activity of *Campylobacter jejuni*. *Journal of Bacteriology*, *190*(13), 4512-4520.

Zhang, M., He, L., Li, Q., Sun, H., Gu, Y., You, Y., … Zhang, J. (2010). Genomic characterization of the Guillain-Barre syndrome-associated *Campylobacter jejuni* ICDCCJ07001 Isolate. *PloS One*, *5*(11), e15060.

Zhang, M., Xiao, D., Zhao, F., Gu, Y., Meng, F., & He, L. (2009). Comparative Proteomic Analysis of *Campylobacter jejuni* Cultured at 37 ° C and 42 ° C. *Japanese Journal of Infectious Diseases*, *62*(5), 356-361.

Zhang-Akiyama, Q. M., Morinaga, H., Kikuchi, M., Yonekura, S. I., Sugiyama, H., Yamamoto, K., & Yonei, S. (2009). KsgA, a 16S rRNA adenine methyltransferase, has a novel DNA glycosylase/ AP lyase activity to prevent mutations in *Escherichia coli*. *Nucleic Acids Research*, *37*(7), 2116-2125.

Zhao, D.-L., Yu, Z.-C., Li, P.-Y., Wu, Z.-Y., Chen, X.-L., Shi, M., … Zhang, Y.-Z. (2011). Characterization of a cryptic plasmid pSM429 and its application for heterologous expression in psychrophilic *Pseudoalteromonas*. *Microbial Cell Factories*, 10, 30.

Zhao, S., Tyson, G. H., Chen, Y., Li, C., Mukherjee, S., Young, S., … McDermott, P. F. (2016). Whole-genome sequencing analysis accurately predicts antimicrobial resistance phenotypes in *Campylobacter* spp. *Applied and Environmental Microbiology*, *82*(2), 459-466.

Ziprin, R. L., Young, C. R., Byrd, J. A., Stanker, L. H., Hume, M. E., Gray, S. A., … Konkel, M. E. (2014). Role of *Campylobacter jejuni* potential virulence genes in cecal colonization. *Avian Diseases*, *45*(3), 549-557.

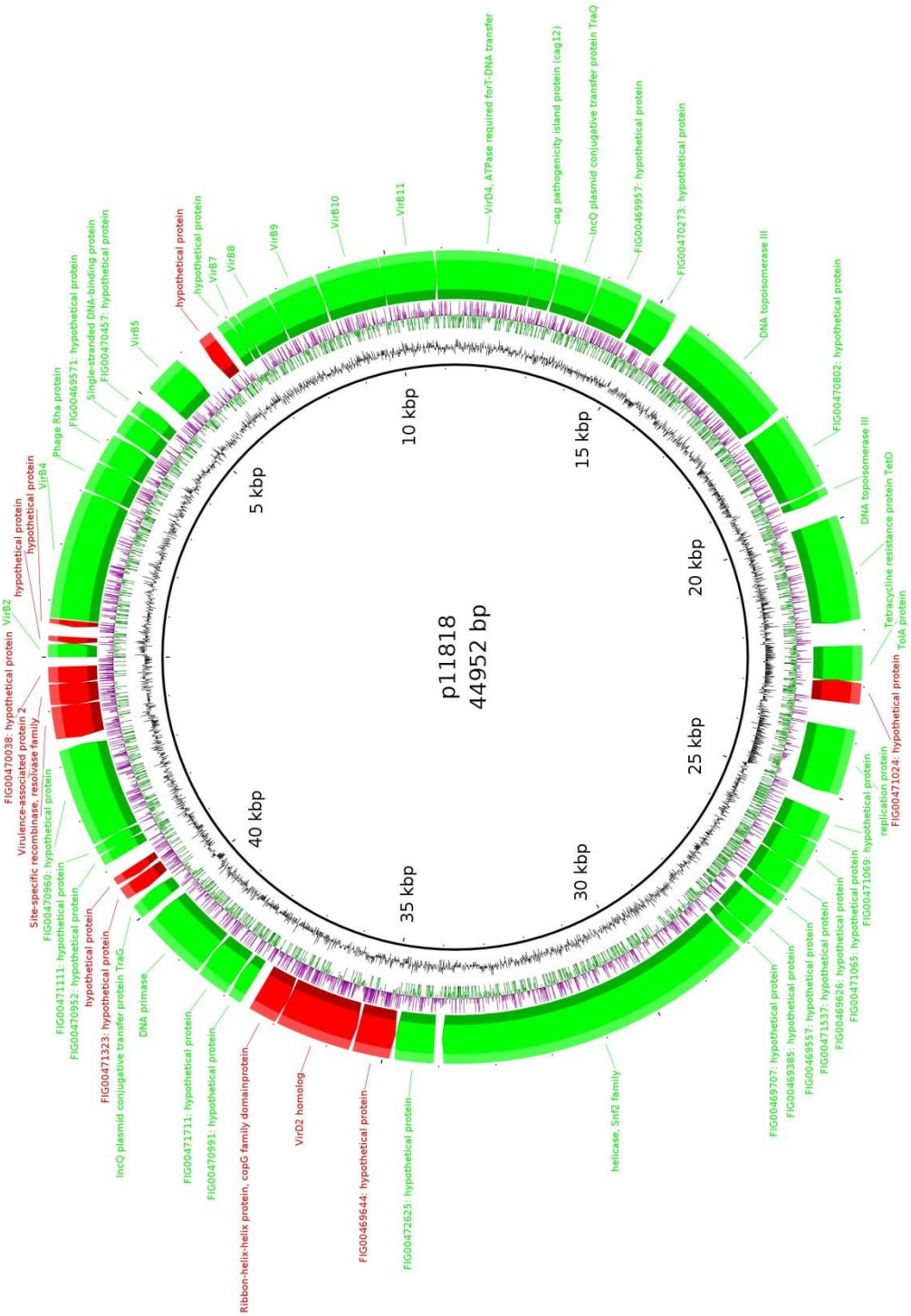
Zückert, W. R., & Meyer, J. (1996). Circular and linear plasmids of Lyme disease spirochetes have extensive homology: Characterization of a repeated DNA element. *Journal of Bacteriology*, *178*(8), 2287-2298.

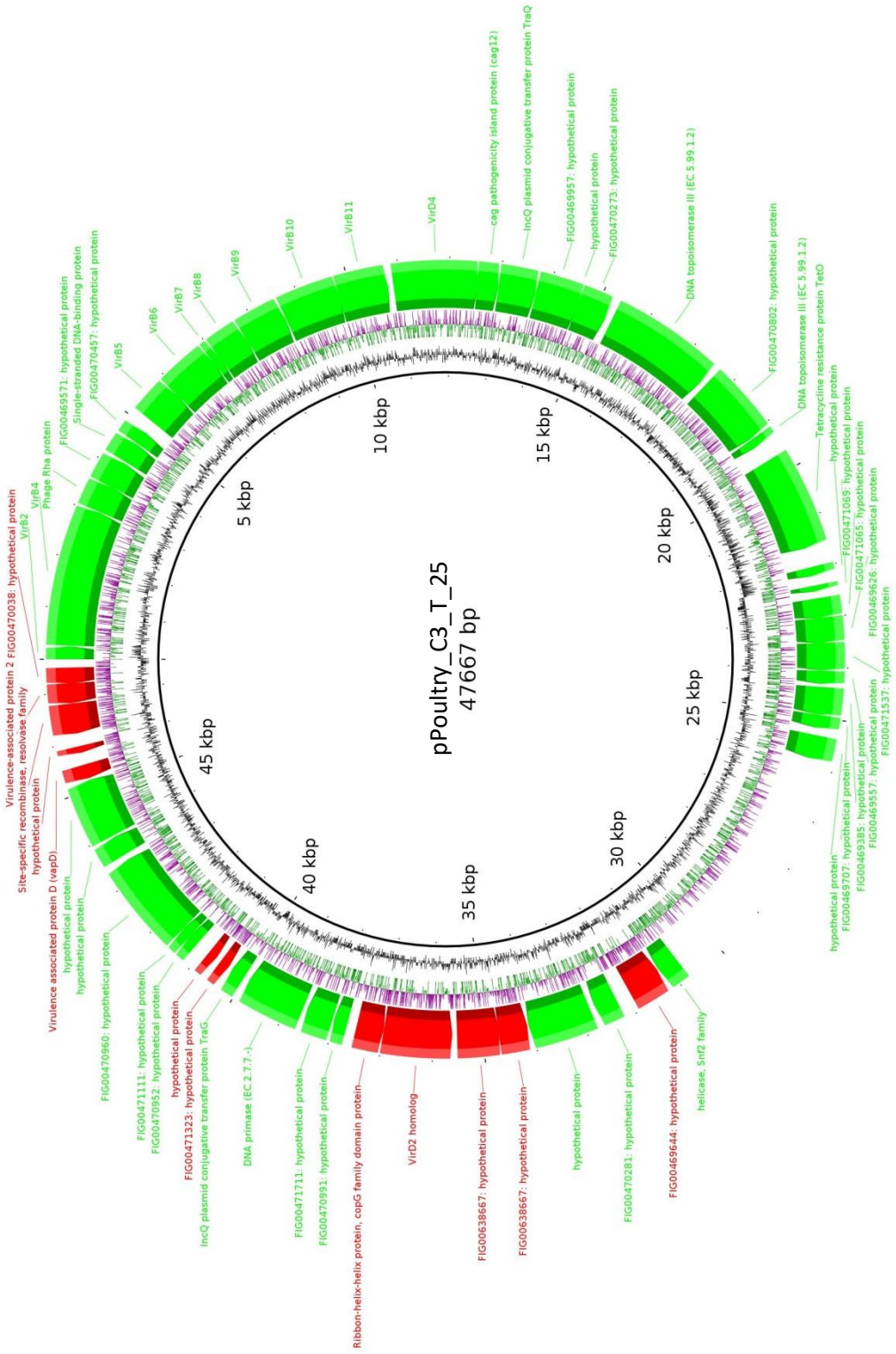
Supplementary material

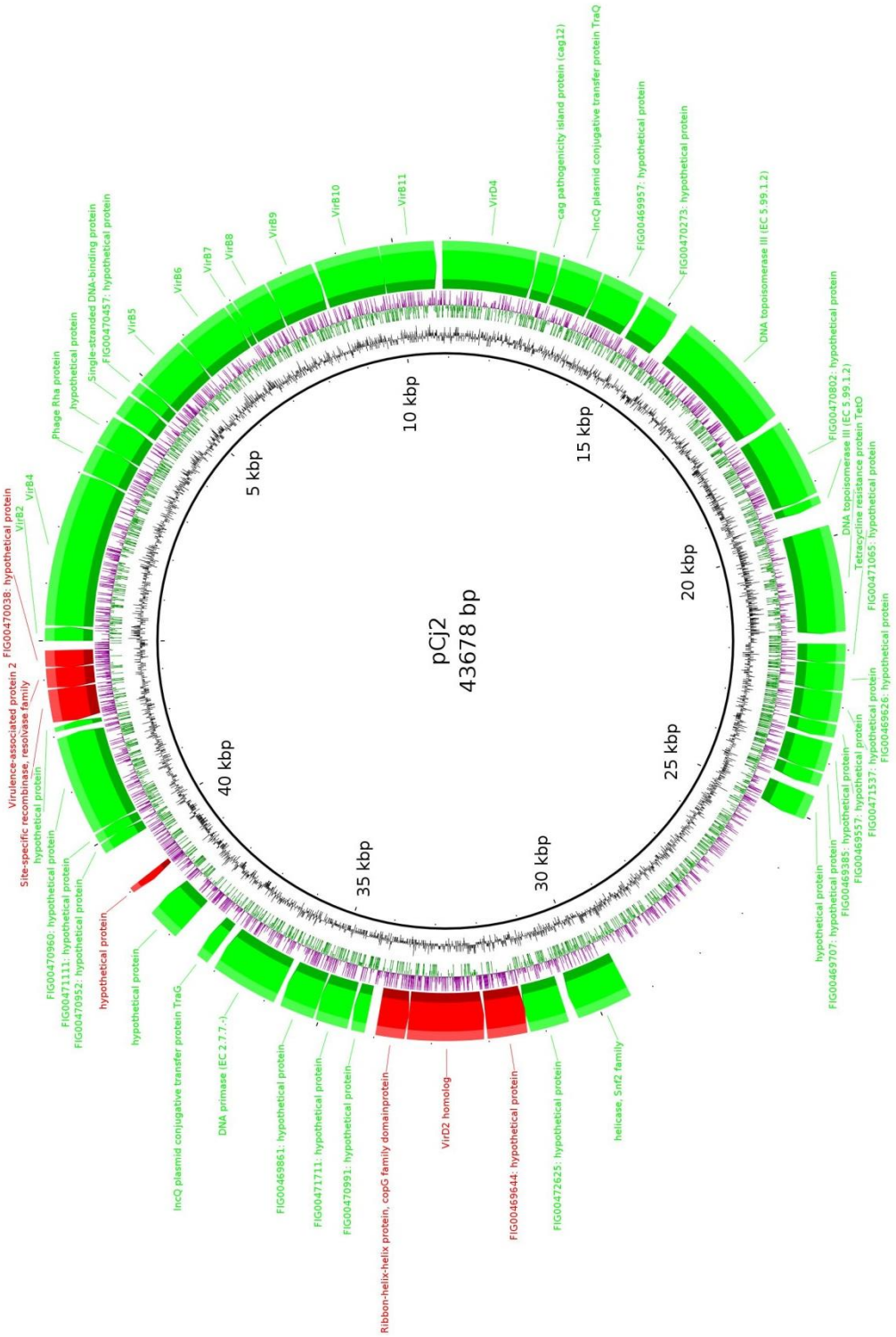
S1 - Strains from Molecular Microbiology Laboratory (Geoffrey Pope building, laboratory 401, EX44QD) used in this study.

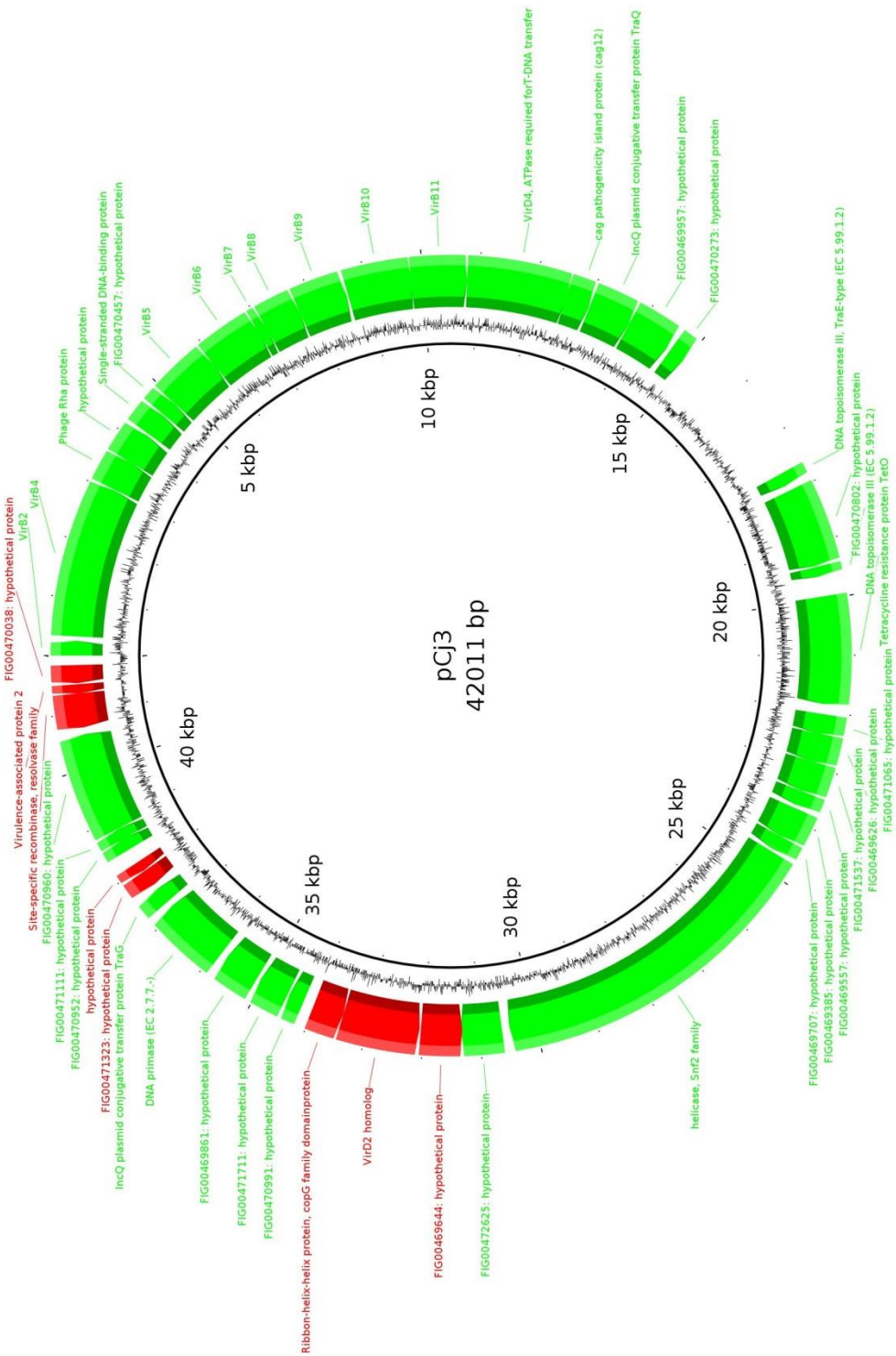
Strain	Origin	Growth	Tet R	virD4 mark	tet(O) ma	hel mark	hicA/B mark	Species
Calf_3	Cow	Y	N	N	X	X	X	X
99/118	Cow	N	X	X	X	X	X	X
99/194	Cow	N	X	X	X	X	X	X
99/201	Cow	Y	N	N	X	X	X	X
99/202	Cow	Y	N	N	X	X	X	X
C85-4-99-5	Cow	Y	N	N	X	X	X	X
C500-1-99-2	Cow	Y	N	N	X	X	X	X
C559-3-99-2	Cow	Y	N	N	X	X	X	X
KS_Cattle_8	Cow	Y	N	N	X	X	X	X
BB1267	Environmental	Y	N	N	X	X	X	X
Beach_28766	Environmental	Y	N	N	X	X	X	X
EX1182	Environmental	Y	N	N	X	X	X	X
47	Goose	Y	N	N	X	X	X	X
95	Goose	Y	N	N	X	X	X	X
222	Goose	Y	Y	Y	Y	Y	N	*jejuni*
32799	Human	Y	N	N	X	X	X	X
33084	Human	Y	N	N	X	X	X	X
K1	Human	Y	N	N	X	X	X	X
K3	Human	Y	N	N	X	X	X	X
K4	Human	Y	N	N	X	X	X	X
K5	Human	Y	Y	Y	Y	Y	N	*coli*
K6	Human	N	X	X	X	X	X	X
K7	Human	Y	Y	N	N	N	N	*coli*
Jan-43	Human	Y	N	N	X	X	X	X
99/97	Human	Y	N	N	X	X	X	X
99/188	Human	Y	N	N	X	X	X	X
99/189	Human	Y	N	N	X	X	X	X
99/197	Human	Y	N	N	X	X	X	X
99/212	Human	Y	N	N	X	X	X	X
99/216	Human	Y	N	N	X	X	X	X
11818	Human	Y	Y	N	Y	Y	Y	*jejuni*
11919	Human	Y	N	N	X	X	X	X
11974	Human	Y	N	N	X	X	X	X
12241	Human	Y	N	N	X	X	X	X
12450	Human	N	X	X	X	X	X	X
13305	Human	Y	N	N	X	X	X	X
18836	Human	Y	N	N	X	X	X	X
30280	Human	Y	N	N	X	X	X	X
30328	Human	Y	N	N	X	X	X	X
31467	Human	Y	N	N	X	X	X	X
31481	Human	Y	N	N	X	X	X	X
31485	Human	Y	N	N	X	X	X	X
32787	Human	Y	N	N	X	X	X	X
33106	Human	Y	N	N	X	X	X	X
34007	Human	Y	N	N	X	X	X	X
44119	Human	Y	N	N	X	X	X	X
47693	Human	Y	N	N	X	X	X	X
K2	Human	Y	Y	Y	Y	Y	N	*jejuni*
90843	Human	Y	N	N	X	X	X	X
Hi40500471	Human	Y	N	N	X	X	X	X
Hi40620300	Human	Y	Y	Y	Y	Y	Y	*jejuni*
Hi40980306	Human	Y	N	N	X	X	X	X
Hi80547	Human	Y	N	N	X	X	X	X
Hi80554	Human	N	X	X	X	X	X	X
Hi80586	Human	Y	N	N	X	X	X	X
Hi81006	Human	Y	Y	N	Y	N	N	*jejuni*
Hi81214	Human	Y	Y	N	Y	N	N	*jejuni*
Hi81266	Human	Y	N	N	X	X	X	X
K8	Human	Y	N	N	X	X	X	X
KSSAPSM6	Human	Y	N	N	X	X	X	X
S2471	Human	N	X	X	X	X	X	X
KSSHPSM4	Human	Y	N	N	X	X	X	X
93/372	Unknown	Y	N	Y	N	Y	Y	*jejuni*
PS304	Pig	Y	Y	N	Y	Y	Y	*coli*
PS549.1	Pig	Y	N	N	X	X	X	X
PS623	Pig	Y	Y	Y	Y	Y	Y	*jejuni*
PS762	Pig	Y	Y	Y	Y	Y	Y	*jejuni*
PS838	Pig	Y	N	N	X	X	X	X
PS843	Pig	Y	N	N	X	X	X	X
PS849	Pig	Y	N	N	X	X	X	X
PS852	Pig	Y	N	N	X	X	X	X
PS857	Pig	Y	N	N	X	X	X	X
Chicken_91_B1	Poultry	Y	N	N	X	X	X	X
A1.CF.12	Poultry	Y	Y	Y	Y	N	Y	*jejuni*
A6.T2.15	Poultry	Y	Y	Y	Y	Y	Y	*jejuni*
A8/35/15A	Poultry	Y	Y	Y	Y	Y	Y	*jejuni*
C1/C/2	Poultry	Y	Y	Y	Y	Y	Y	*jejuni*
C120/2	Poultry	Y	N	N	X	X	X	X
C132/1	Poultry	Y	X	X	X	X	X	X
C3/T/25col3	Poultry	Y	Y	N	Y	Y	Y	*jejuni*
C5/T2/8	Poultry	Y	Y	Y	Y	Y	Y	*jejuni*
D2/27/3	Poultry	Y	N	N	X	X	X	X
D2/T/8	Poultry	Y	N	N	X	X	X	X
D2/T/95	Poultry	Y	N	N	X	X	X	X
D5-20-9A	Poultry	Y	N	N	X	X	X	X
EX1286	Poultry	Y	N	N	X	X	X	X
MB1	Poultry	Y	N	N	X	X	X	X
MB2	Poultry	Y	N	N	X	X	X	X
MB3	Poultry	Y	N	Y	N	Y	Y	*jejuni*
MB4	Poultry	Y	N	N	X	X	X	X
MB5	Poultry	Y	N	N	X	X	X	X
MB6	Poultry	Y	N	N	X	X	X	X
MB7	Poultry	Y	N	N	X	X	X	X
MB8	Poultry	Y	Y	N	Y	N	N	*jejuni*
MB9	Poultry	Y	Y	Y	Y	Y	Y	*jejuni*
MB10	Poultry	Y	N	N	X	X	X	X
MB12	Poultry	Y	N	Y	N	N	Y	*jejuni*
MB13	Poultry	Y	N	N	X	X	X	X
MB14	Poultry	Y	N	N	X	X	X	X
MB15	Poultry	Y	N	N	X	X	X	X
MB16	Poultry	Y	N	N	X	X	X	X
MB17	Poultry	Y	Y	N	Y	N	N	*coli*
MB18	Poultry	Y	Y	Y	Y	Y	Y	*jejuni*
94/229	Poultry	Y	N	N	X	X	X	X
S39-2-99-3	Sheep	Y	N	N	X	X	X	X
S87-4-99-3	Sheep	Y	N	N	X	X	X	X
S120-4-99-4	Sheep	Y	N	N	X	X	X	X
S216-5-99-1	Sheep	Y	N	N	X	X	X	X
S372-5-99-4	Sheep	Y	N	N	X	X	X	X
S379-8-99-1	Sheep	Y	N	N	X	X	X	X
S435-3-99	Sheep	Y	N	N	X	X	X	X
S499-1-99-5	Sheep	Y	N	N	X	X	X	X
S585-3-99	Sheep	Y	N	N	X	X	X	X
30x2_tag60	Unknown	Y	N	N	X	X	X	X
KJSHPFELL_2	Unknown	Y	N	N	X	X	X	X
KJSHPM8	Unknown	N	X	292	X	X	X	X
Cj1	Human	Y	Y	Y	Y	Y	Y	*jejuni*
Cj2	Human	Y	Y	Y	Y	Y	Y	*jejuni*
Cj3	Human	Y	Y	Y	Y	Y	Y	*jejuni*
Cj4	Human	N	X	X	X	X	X	X
Cj5	Human	Y	N	N	X	X	X	*X*

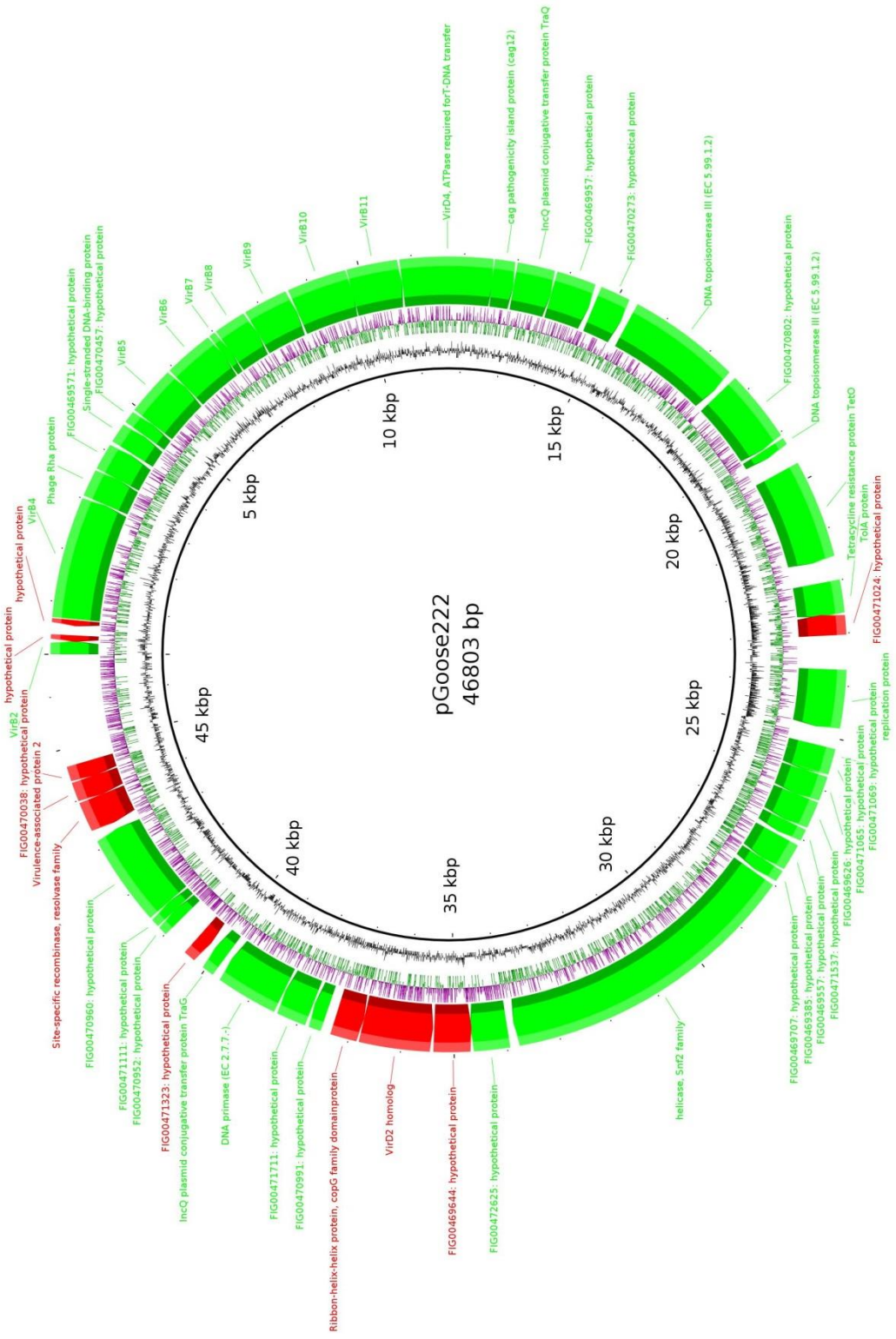
S2 - Plasmids maps sequenced in this study. Plasmids maps were obtained with BRIG software. Green and Red arrows represent the genes annotated on the plasmid, as reported by RAST automated annotation. Inner circles represent C+G % and GC skew (C+G % only for pCj3 and pK2).

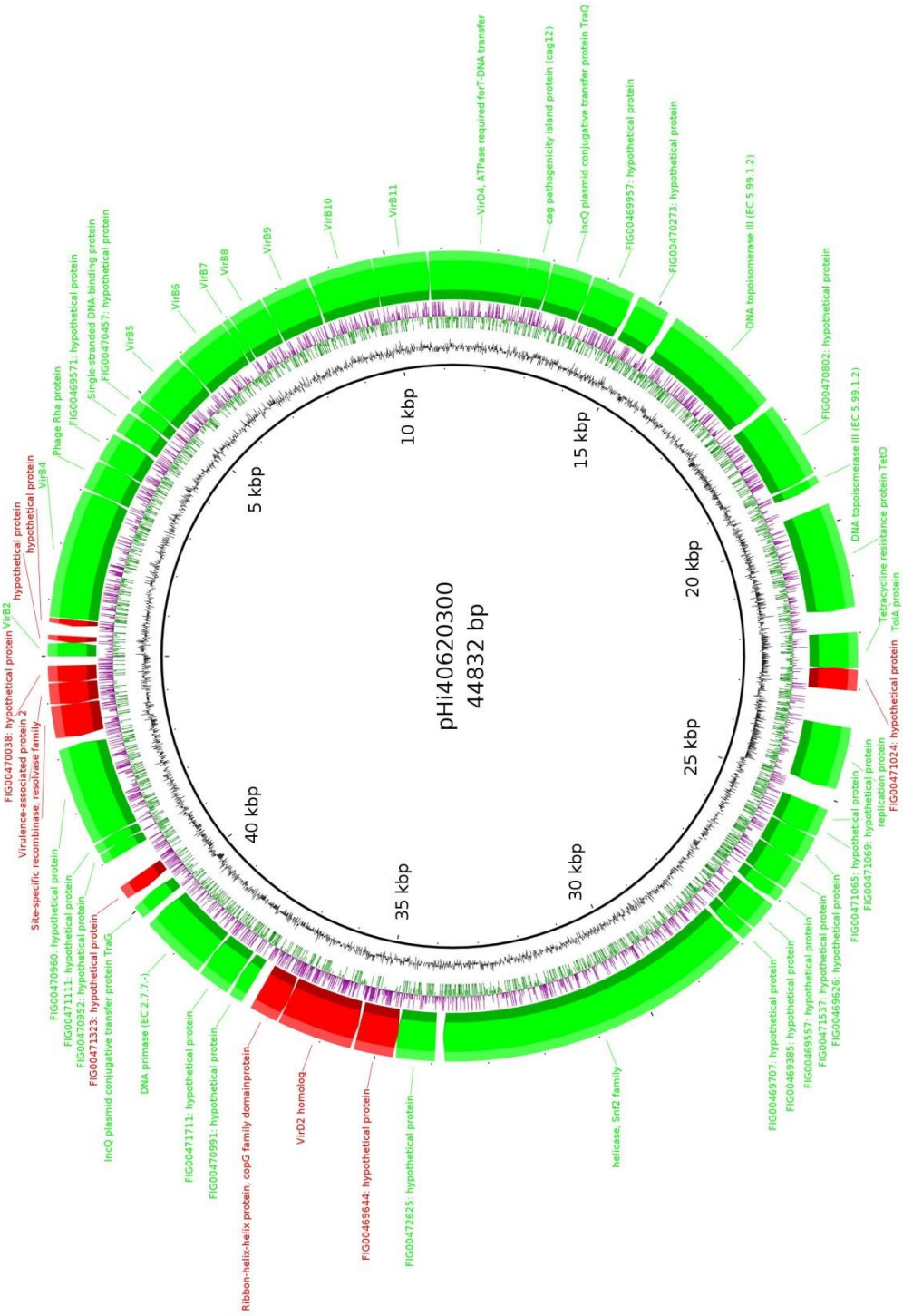

p11818
44952 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
VirD4, ATPase required for T-DNA transfer
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469557: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
FIG00470802: hypothetical protein
DNA topoisomerase III
Tetracycline resistance protein TetO
TolA protein
FIG00471024: hypothetical protein
replication protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00472625: hypothetical protein
FIG00469644: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domain protein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein
VirB2
hypothetical protein
hypothetical protein
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
hypothetical protein
hypothetical protein
VirB7
VirB8
VirB9
VirB10
VirB11

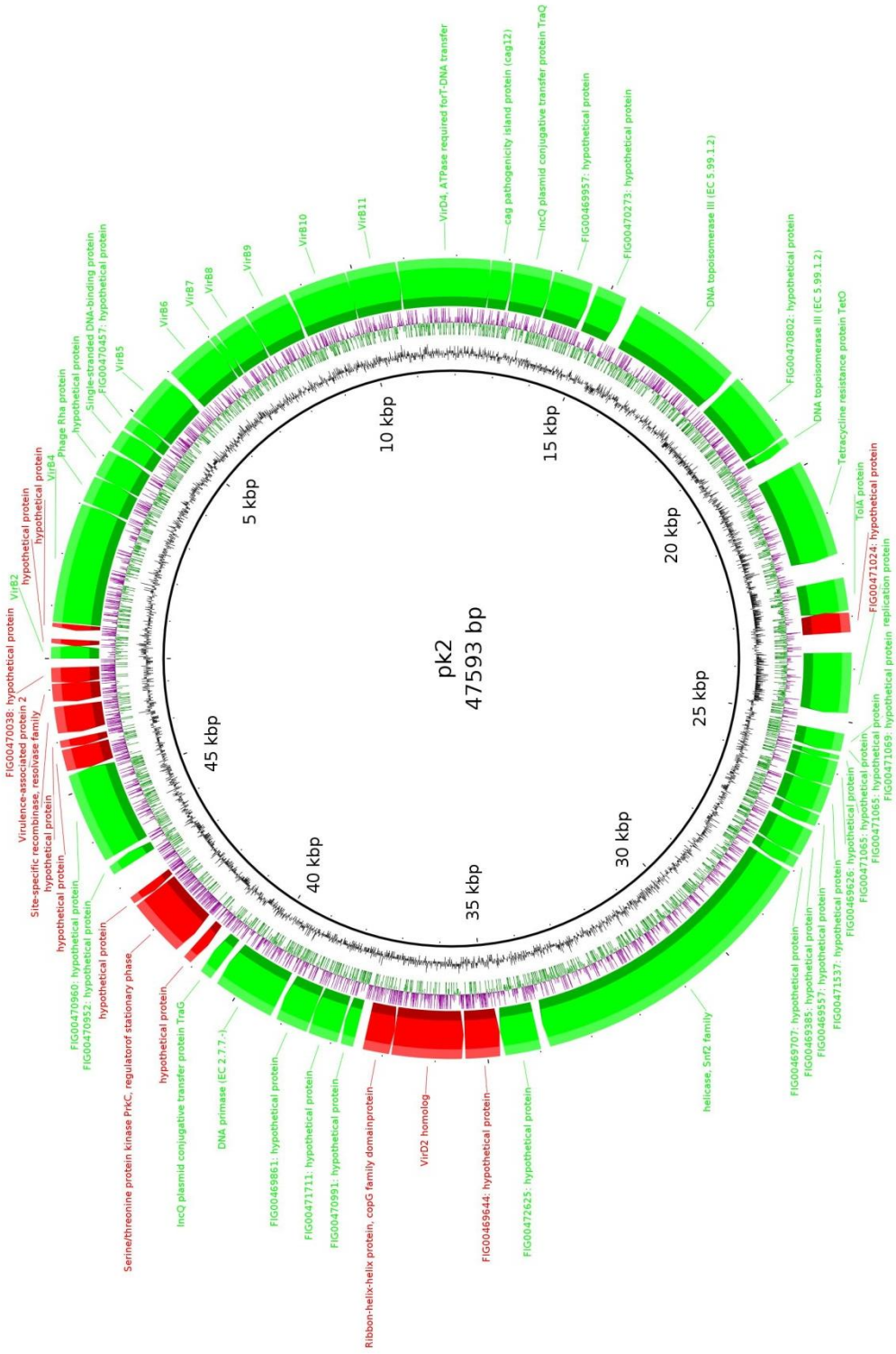

pPoultry_C3_T_25
47667 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
VirD4
VirB11
VirB10
VirB9
VirB8
VirB7
VirB6
VirB5
VirB4
VirB2
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
FIG00470802: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
Tetracycline resistance protein TetO
hypothetical protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
hypothetical protein
helicase, Snf2 family
FIG00469644: hypothetical protein
FIG00470281: hypothetical protein
hypothetical protein
FIG00638667: hypothetical protein
FIG00638667: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domain protein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
hypothetical protein
hypothetical protein
Virulence associated protein D (vapD)
hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2 FIG00470038: hypothetical protein
VirB2
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5

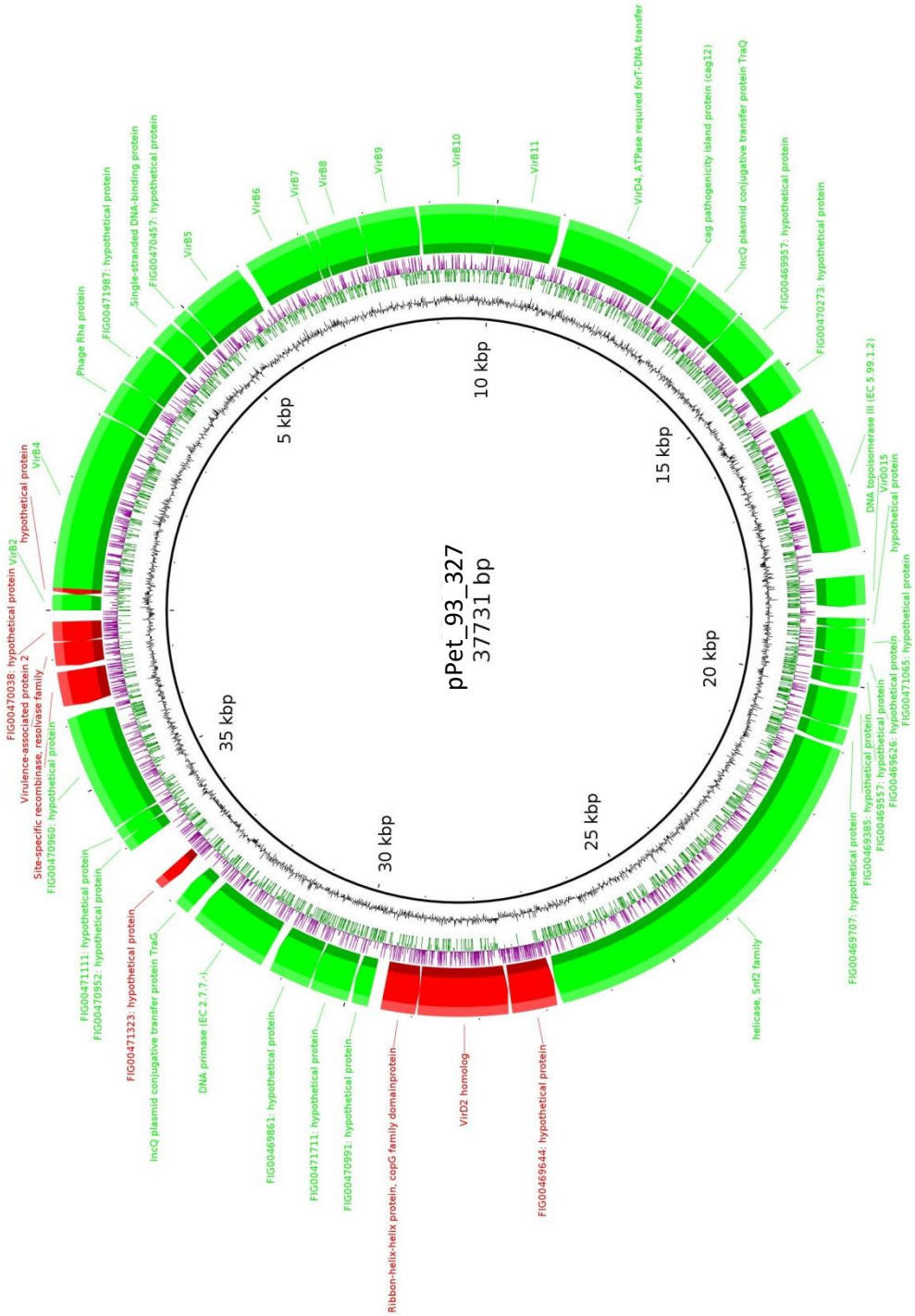

pCj2
43678 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
VirB11
VirD4
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
FIG00470802: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
Tetracycline resistance protein TetO
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
hypothetical protein
FIG00469707: hypothetical protein
FIG00469385: hypothetical protein
FIG00469557: hypothetical protein
FIG00471537: hypothetical protein
helicase, Snf2 family
FIG00472625: hypothetical protein
FIG00469644: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
FIG00469861: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein
VirB2
VirB4
Phage Rha protein
hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10

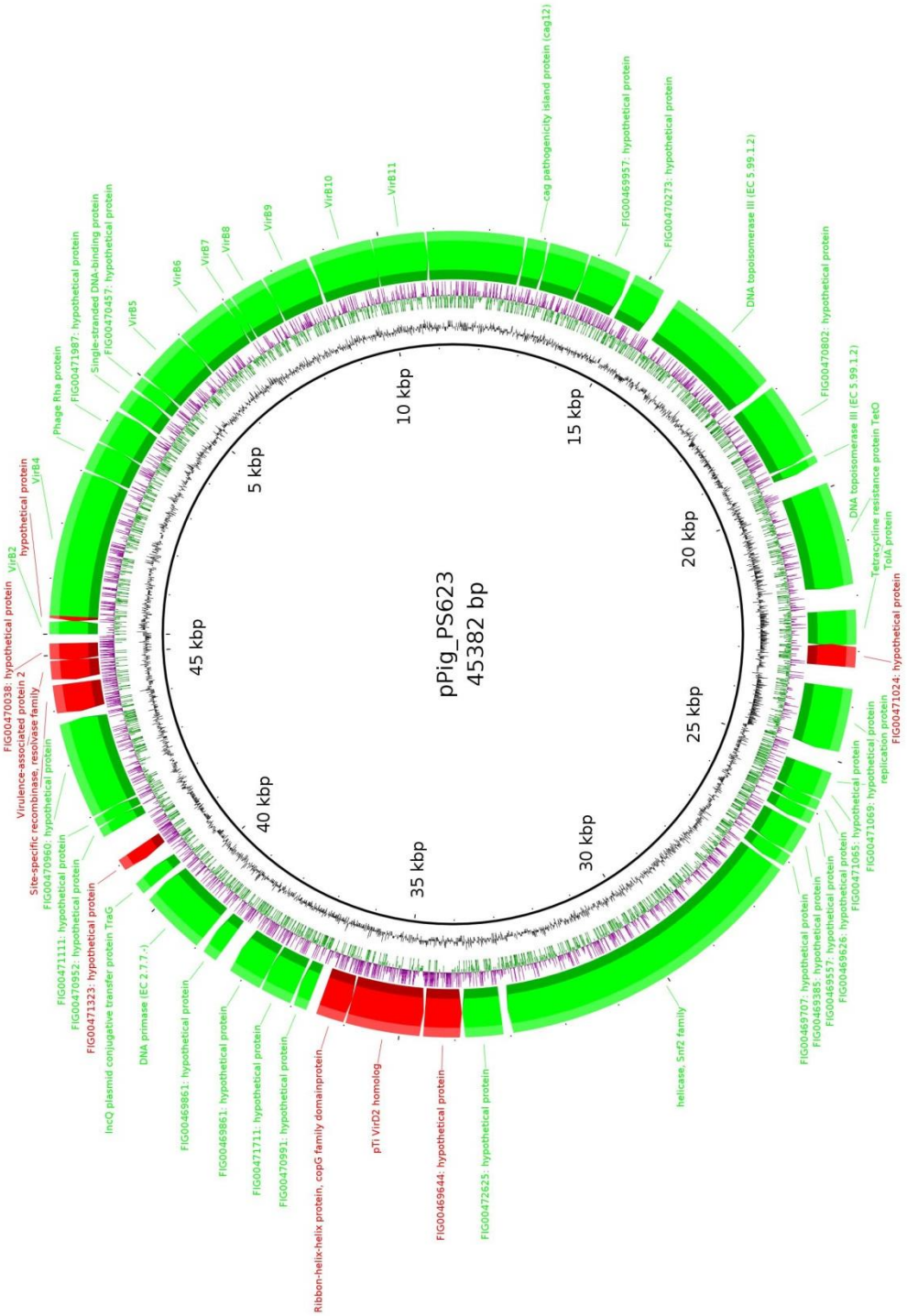

pCj3
42011 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
VirD4, ATPase required for T-DNA transfer
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
VirB11
VirB10
VirB9
VirB8
VirB7
VirB6
VirB5
FIG00470457: hypothetical protein
Single-stranded DNA-binding protein
hypothetical protein
Phage Rha protein
VirB4
VirB2
FIG00470038: hypothetical protein
Virulence-associated protein 2
Site-specific recombinase, resolvase family
FIG00470960: hypothetical protein
FIG00471111: hypothetical protein
FIG00470952: hypothetical protein
hypothetical protein
FIG00471323: hypothetical protein
IncQ plasmid conjugative transfer protein TraG
DNA primase (EC 2.7.7.-)
FIG00469861: hypothetical protein
FIG00471711: hypothetical protein
FIG00470991: hypothetical protein
Ribbon-helix-helix protein, copG family domainprotein
VirD2 homolog
FIG00469644: hypothetical protein
FIG00472625: hypothetical protein
helicase, Snf2 family
FIG00469707: hypothetical protein
FIG00469385: hypothetical protein
FIG00469557: hypothetical protein
FIG00471537: hypothetical protein
FIG00469626: hypothetical protein
FIG00471065: hypothetical protein
Tetracycline resistance protein TetO
DNA topoisomerase III (EC 5.99.1.2)
FIG00470802: hypothetical protein
DNA topoisomerase III, TraE-type (EC 5.99.1.2)

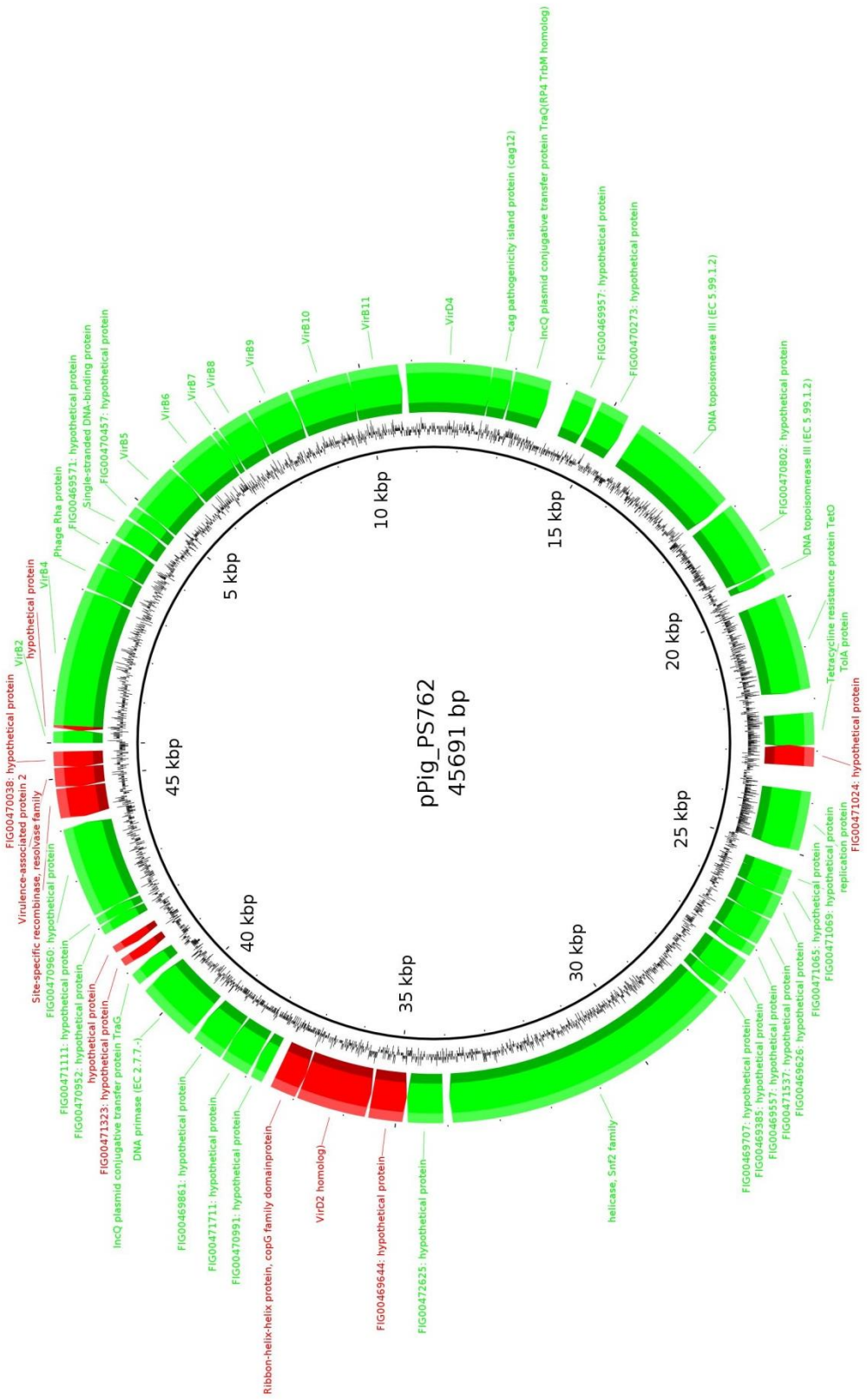

pGoose222
46803 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
VirD4, ATPase required for T-DNA transfer
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
FIG00470802: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
Tetracycline resistance protein TetO
TolA protein
FIG00471024: hypothetical protein
replication protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00472625: hypothetical protein
FIG00469644: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
FIG00470038: hypothetical protein
VirB2
hypothetical protein
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11

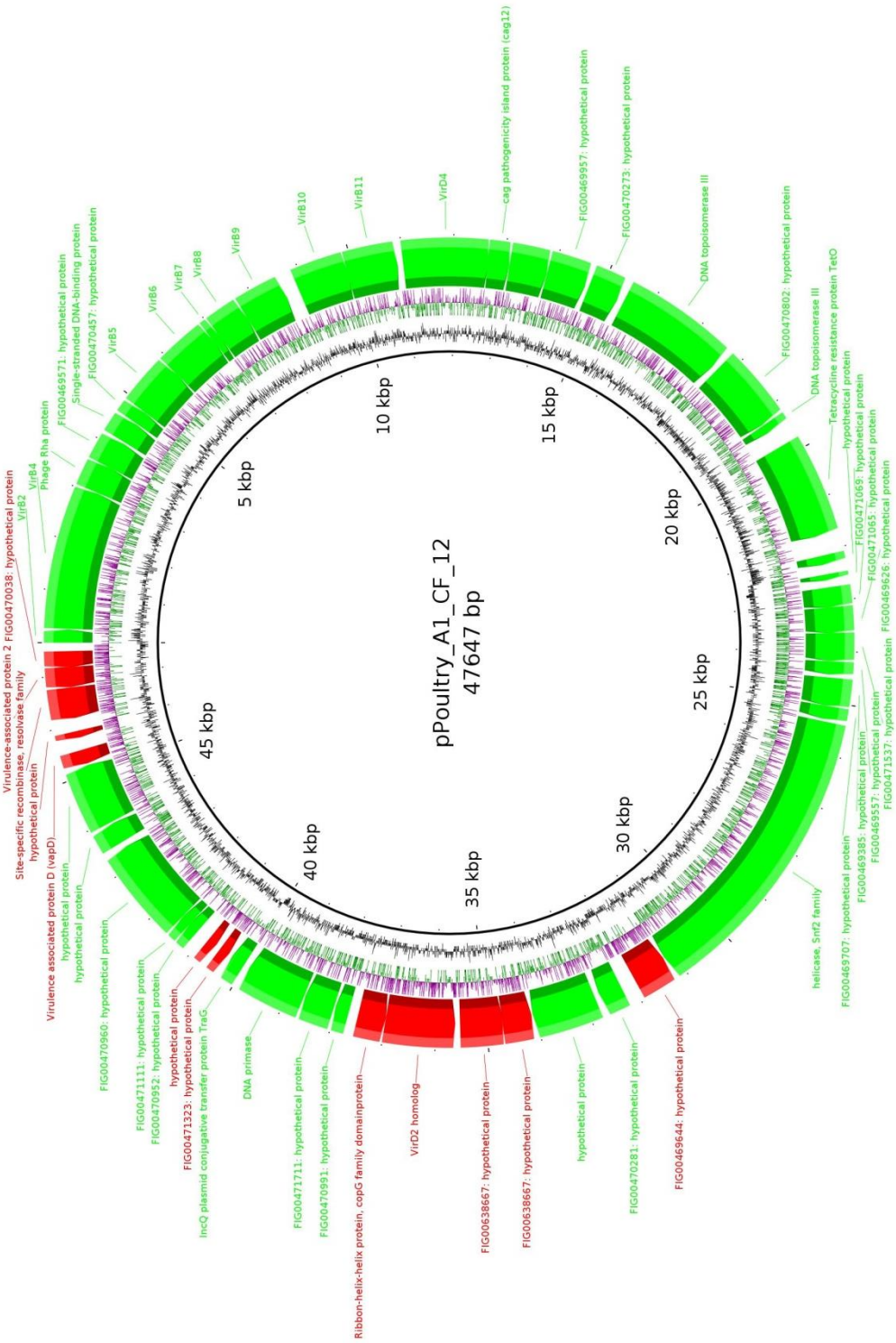

pHi40620300
44832 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
VirD4, ATPase required for T-DNA transfer
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
FIG00470802: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
Tetracycline resistance protein TetO
TolA protein
FIG00471024: hypothetical protein
replication protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00472625: hypothetical protein
FIG00469644: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein VirB2
hypothetical protein
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11

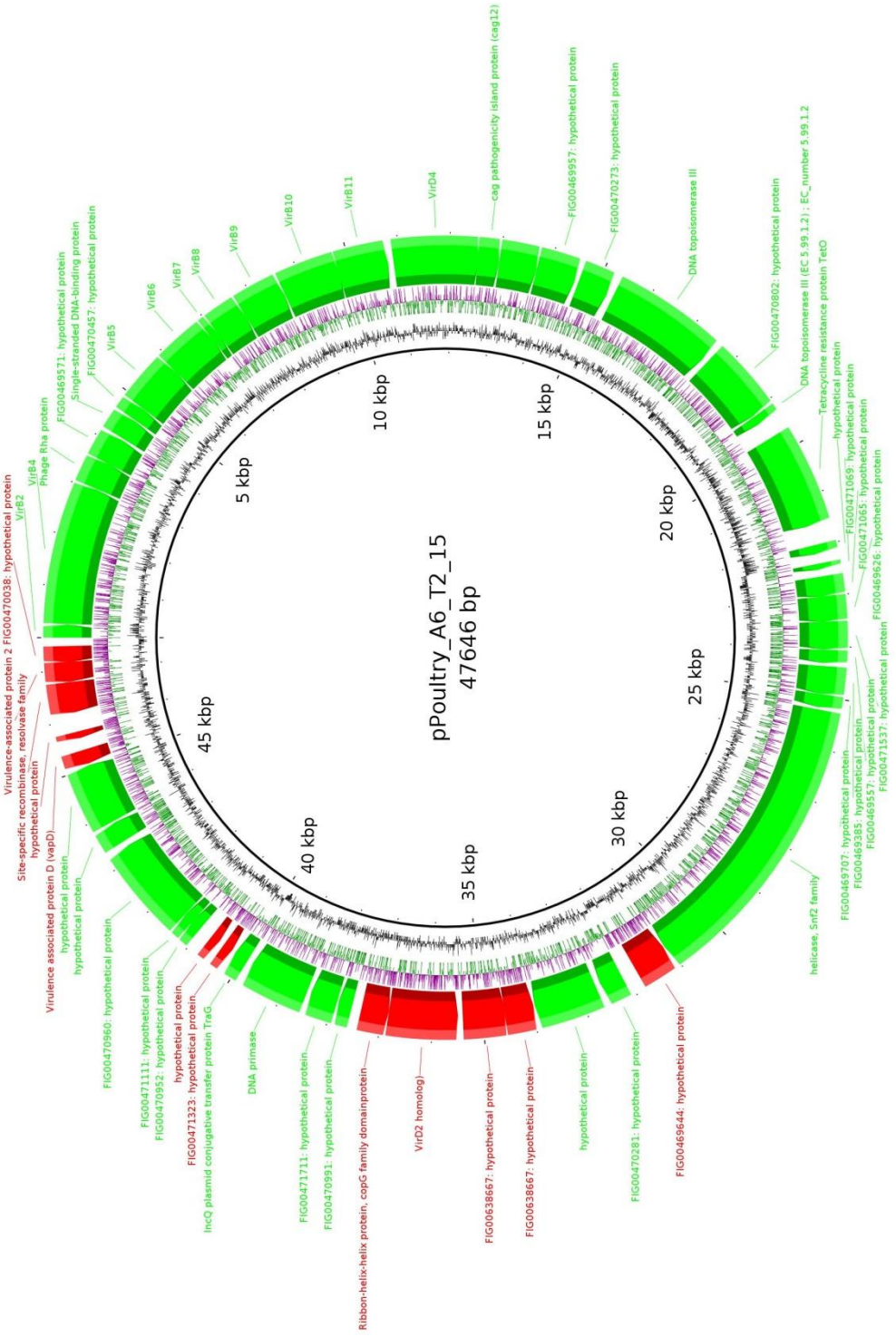

pk2
47593 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
VirD4, ATPase required forT-DNA transfer
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
FIG00470802: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
Tetracycline resistance protein TetO
TolA protein
FIG00471024: hypothetical protein
replication protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00472625: hypothetical protein
FIG00469644: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
FIG00469861: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
hypothetical protein
Serine/threonine protein kinase PrkC, regulatorof stationary phase
hypothetical protein
FIG00470552: hypothetical protein
FIG00470960: hypothetical protein
hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein
VirB2
hypothetical protein
VirB4
Phage Rha protein
hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11

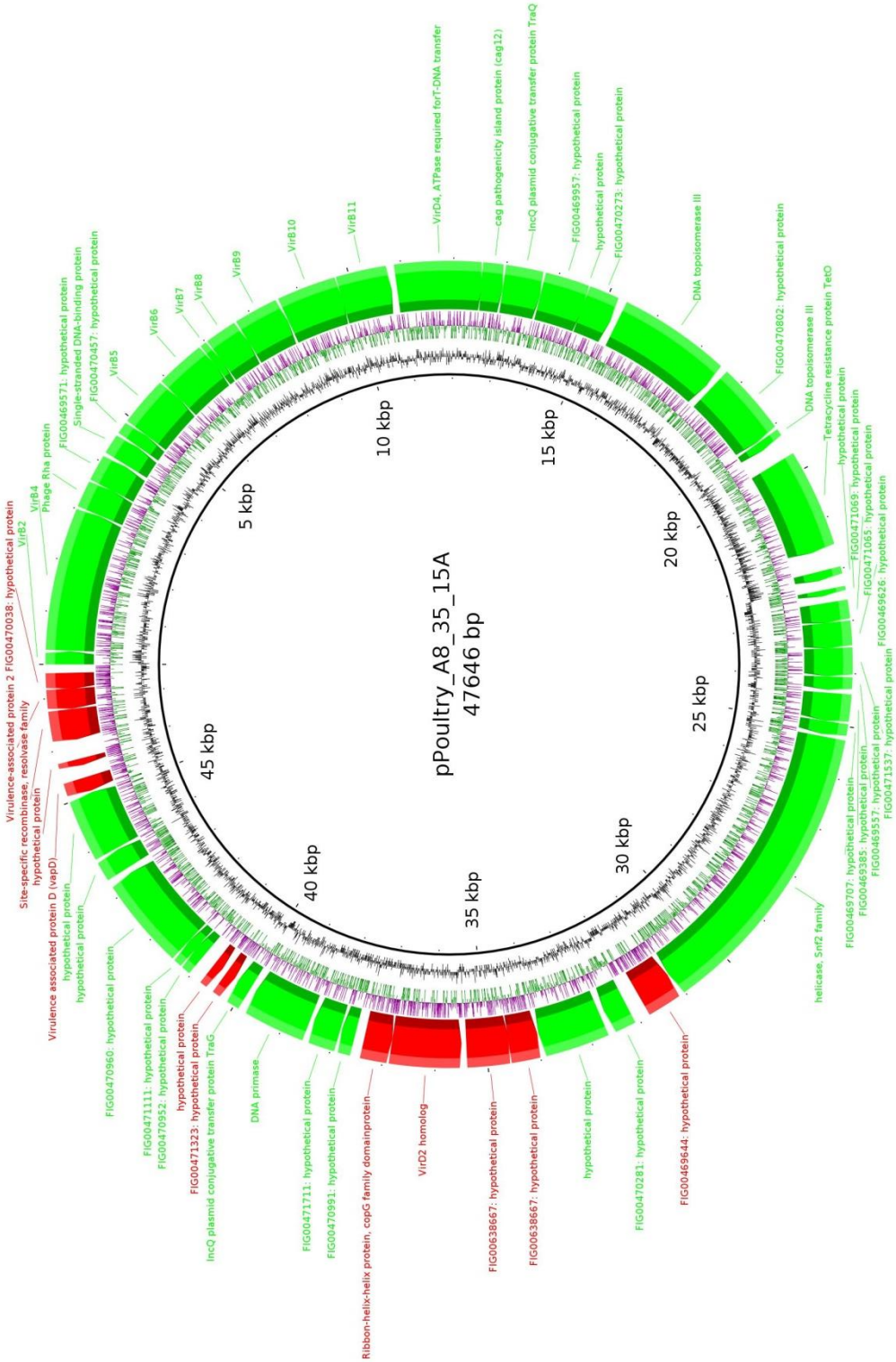

pPet_93_327
37731 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11
VirD4, ATPase required forT-DNA transfer
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
VirD015
hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00469644: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
FIG00469861: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein
VirB2
hypothetical protein
VirB4
Phage Rha protein
FIG00471987: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5

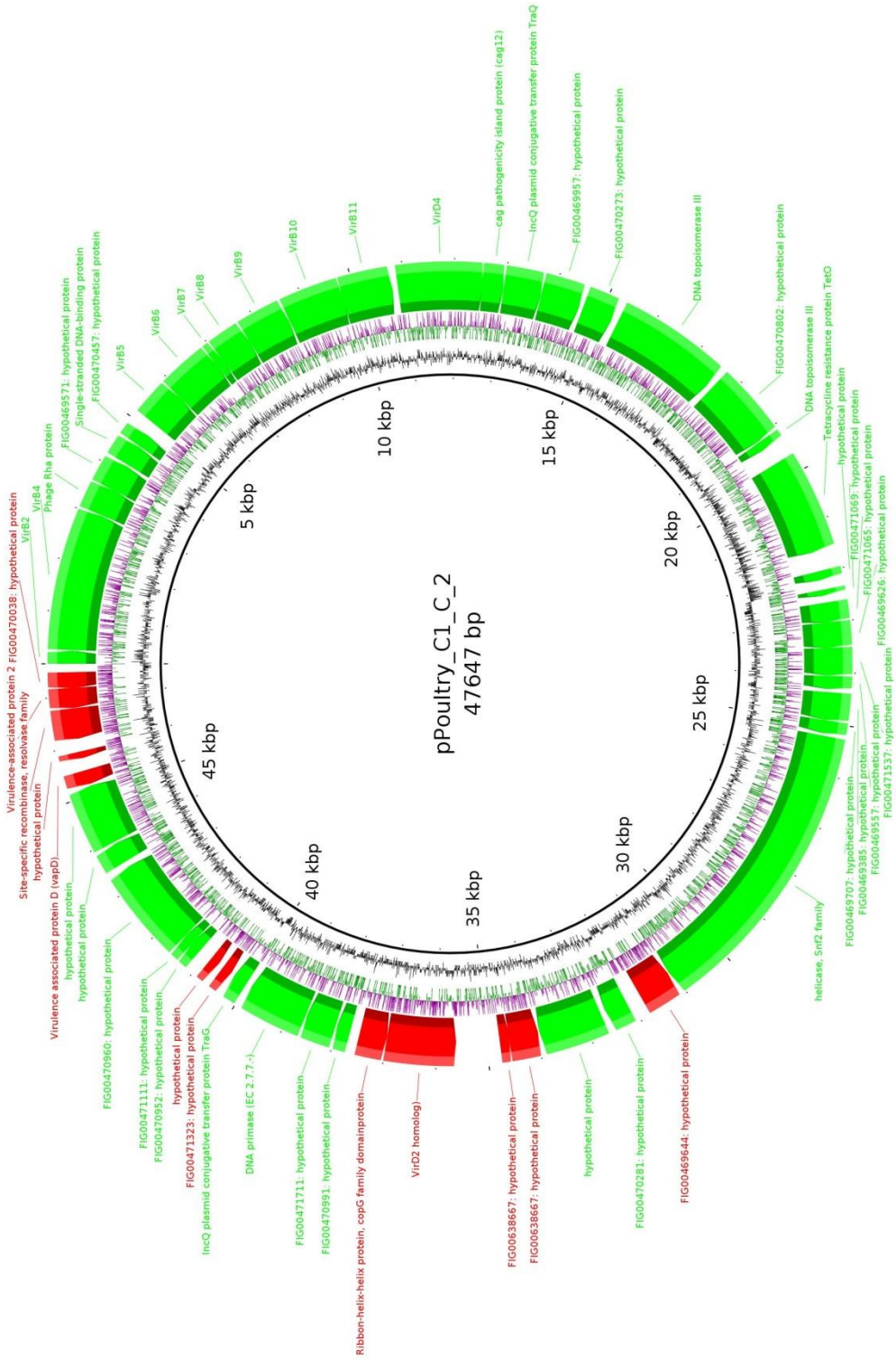

pPig_PS623
45382 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
cag pathogenicity island protein (cag12)
FIG00469557: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
FIG00470802: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
Tetracycline resistance protein TetO
TolA protein
FIG00471024: hypothetical protein
replication protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00472625: hypothetical protein
FIG00469644: hypothetical protein
pTi VirD2 homolog
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
FIG00469861: hypothetical protein
FIG00469861: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein
VirB2
hypothetical protein
VirB4
Phage Rha protein
FIG00471987: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11

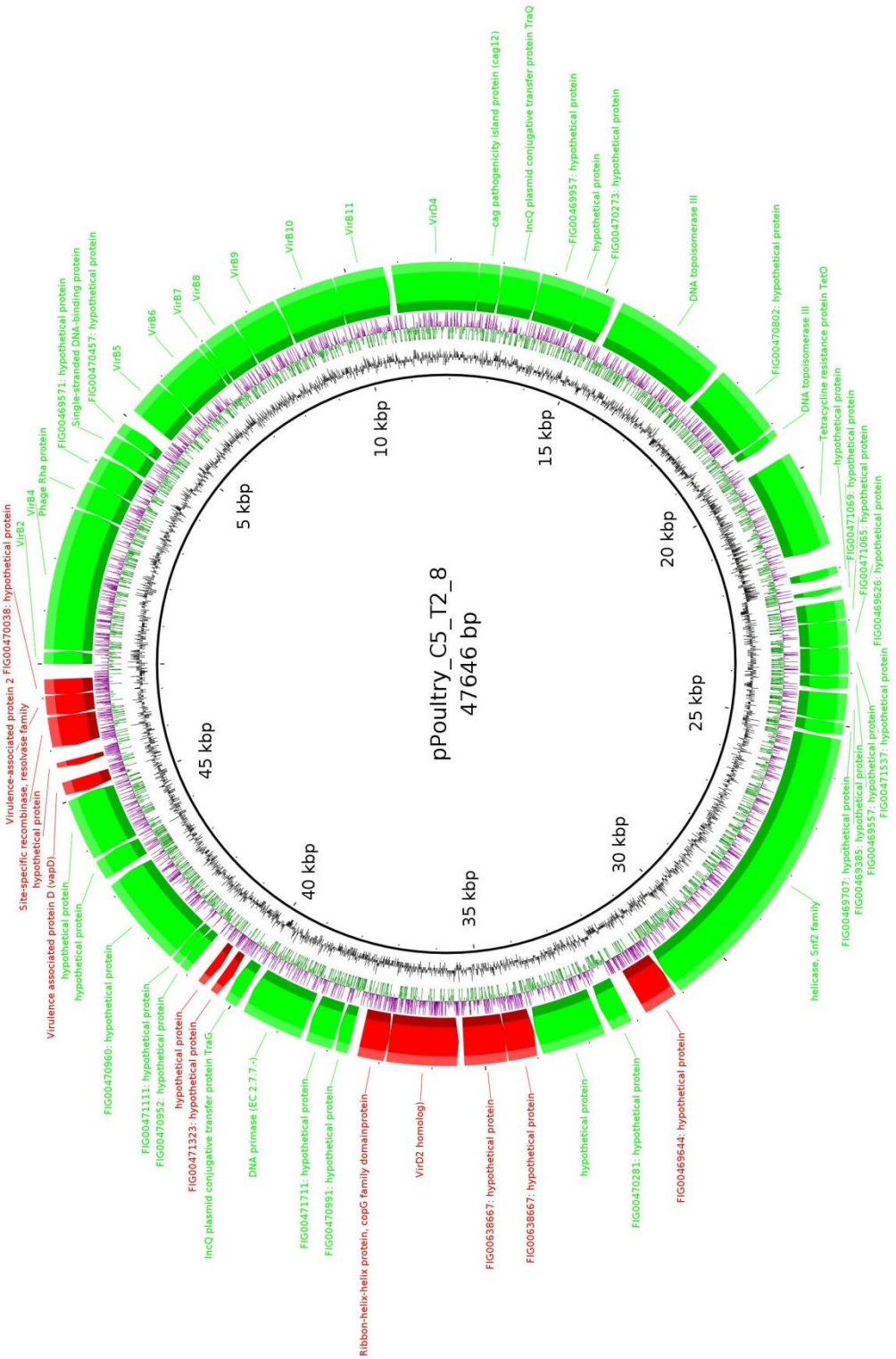

pPig_PS762
45691 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ(RP4 TrbM homolog)
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
FIG00470802: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2)
Tetracycline resistance protein TetO
TolA protein
FIG004071024: hypothetical protein
FIG00471069: hypothetical protein
replication protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00472625: hypothetical protein
FIG00469644: hypothetical protein
VirD2 homolog)
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
FIG00469861: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein
VirB2
hypothetical protein
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11
VirD4

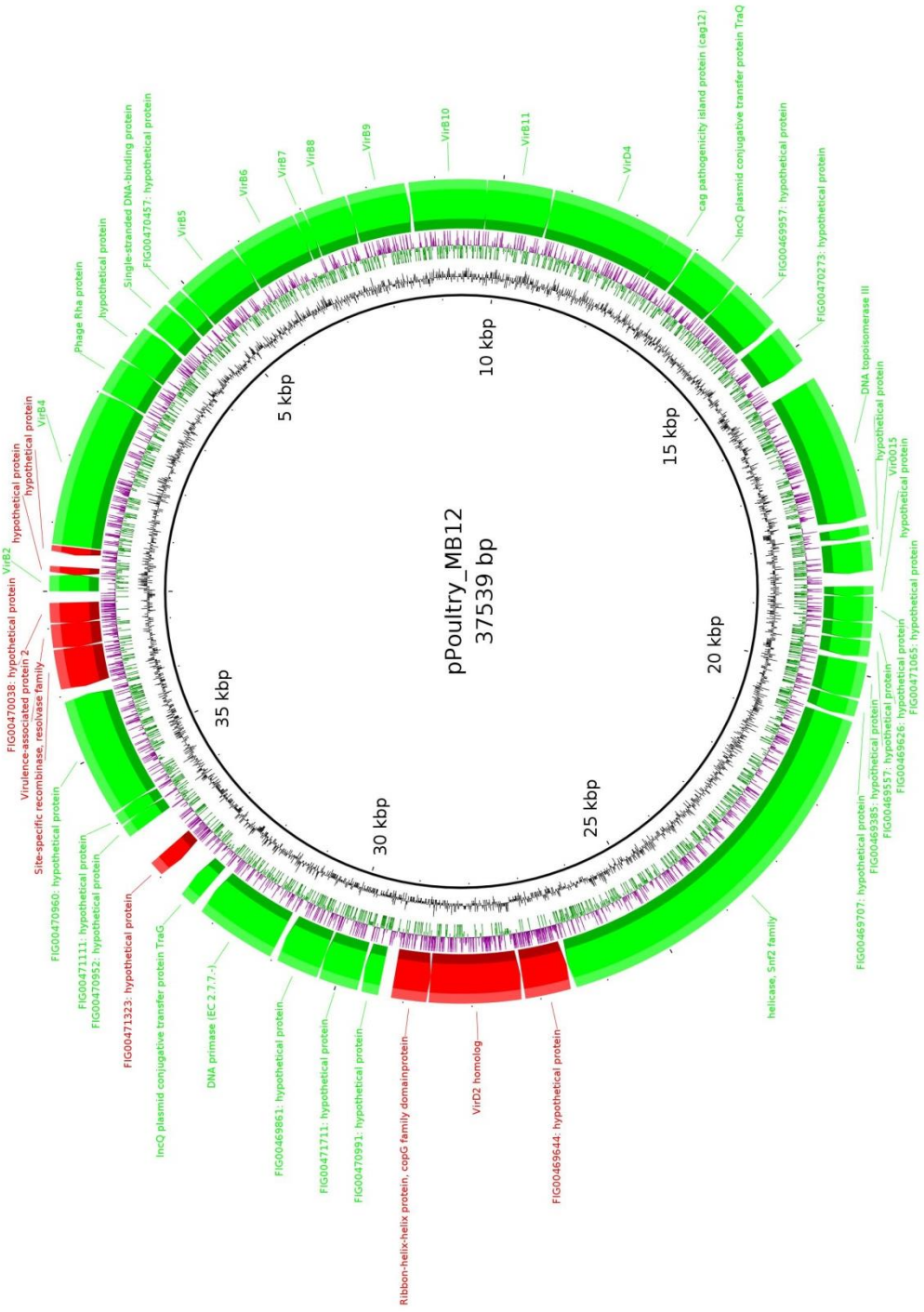

pPoultry_A1_CF_12
47647 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
VirB10
VirB11
VirD4
cag pathogenicity island protein (cag12)
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
FIG00470802: hypothetical protein
DNA topoisomerase III
Tetracycline resistance protein TetO
hypothetical protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00469644: hypothetical protein
FIG00470281: hypothetical protein
hypothetical protein
FIG00638667: hypothetical protein
FIG00638667: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domain protein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
hypothetical protein
hypothetical protein
Virulence associated protein D (vapD)
hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2 FIG00470038: hypothetical protein
VirB2
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9

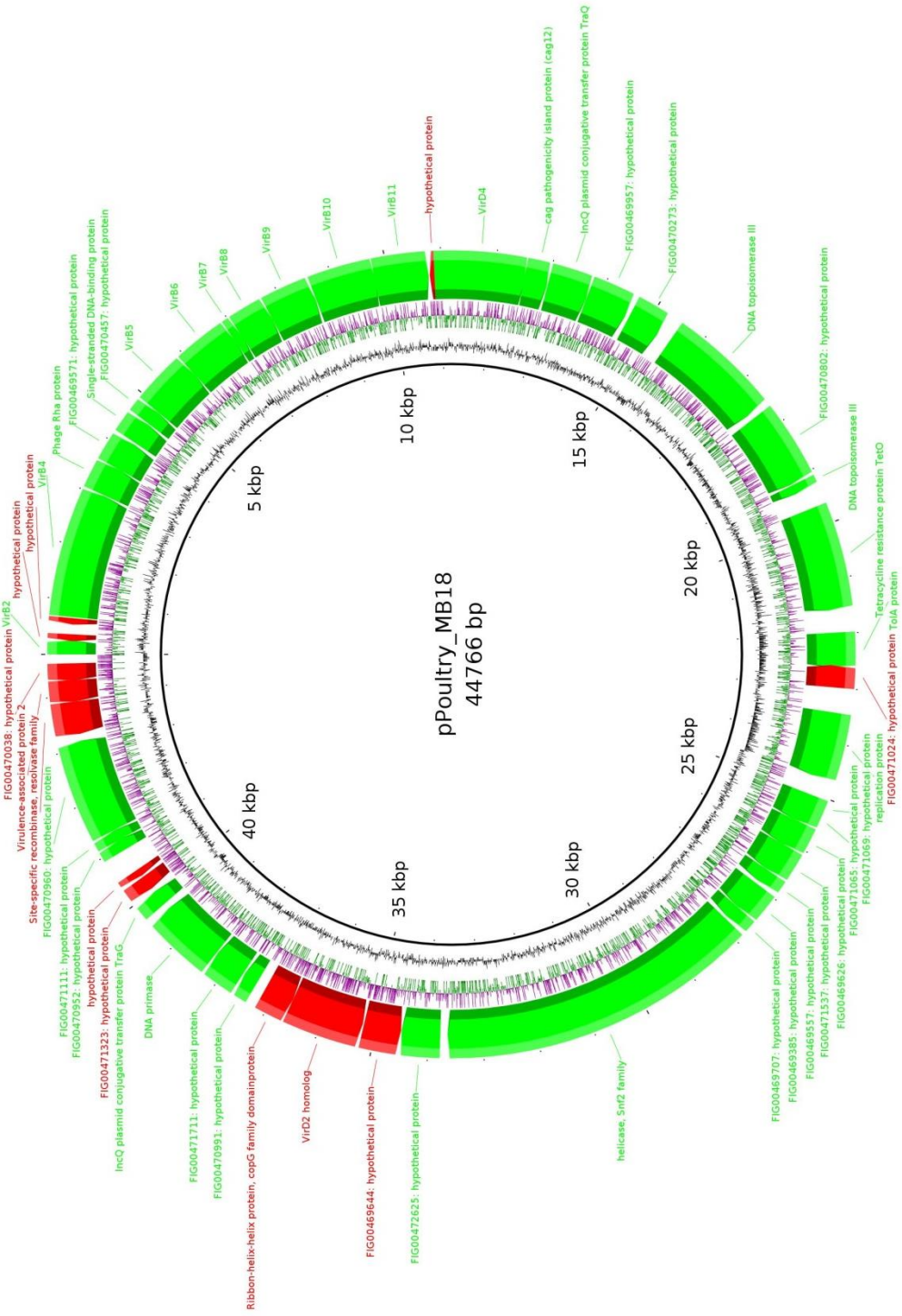

pPoultry_A6_T2_15
47646 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
VirD4
cag pathogenicity island protein (cag12)
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
FIG00470802: hypothetical protein
DNA topoisomerase III (EC 5.99.1.2) ; EC_number 5.99.1.2
Tetracycline resistance protein TetO
hypothetical protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
helicase, Snf2 family
FIG00469707: hypothetical protein
FIG00469385: hypothetical protein
FIG00469557: hypothetical protein
FIG00469644: hypothetical protein
FIG00470281: hypothetical protein
hypothetical protein
FIG00638667: hypothetical protein
FIG00638667: hypothetical protein
VirD2 homolog)
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase
IncQ plasmid conjugative transfer protein TraG
hypothetical protein
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
hypothetical protein
hypothetical protein
Virulence associated protein D (vapD)
hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2 FIG00470038: hypothetical protein
VirB2
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11

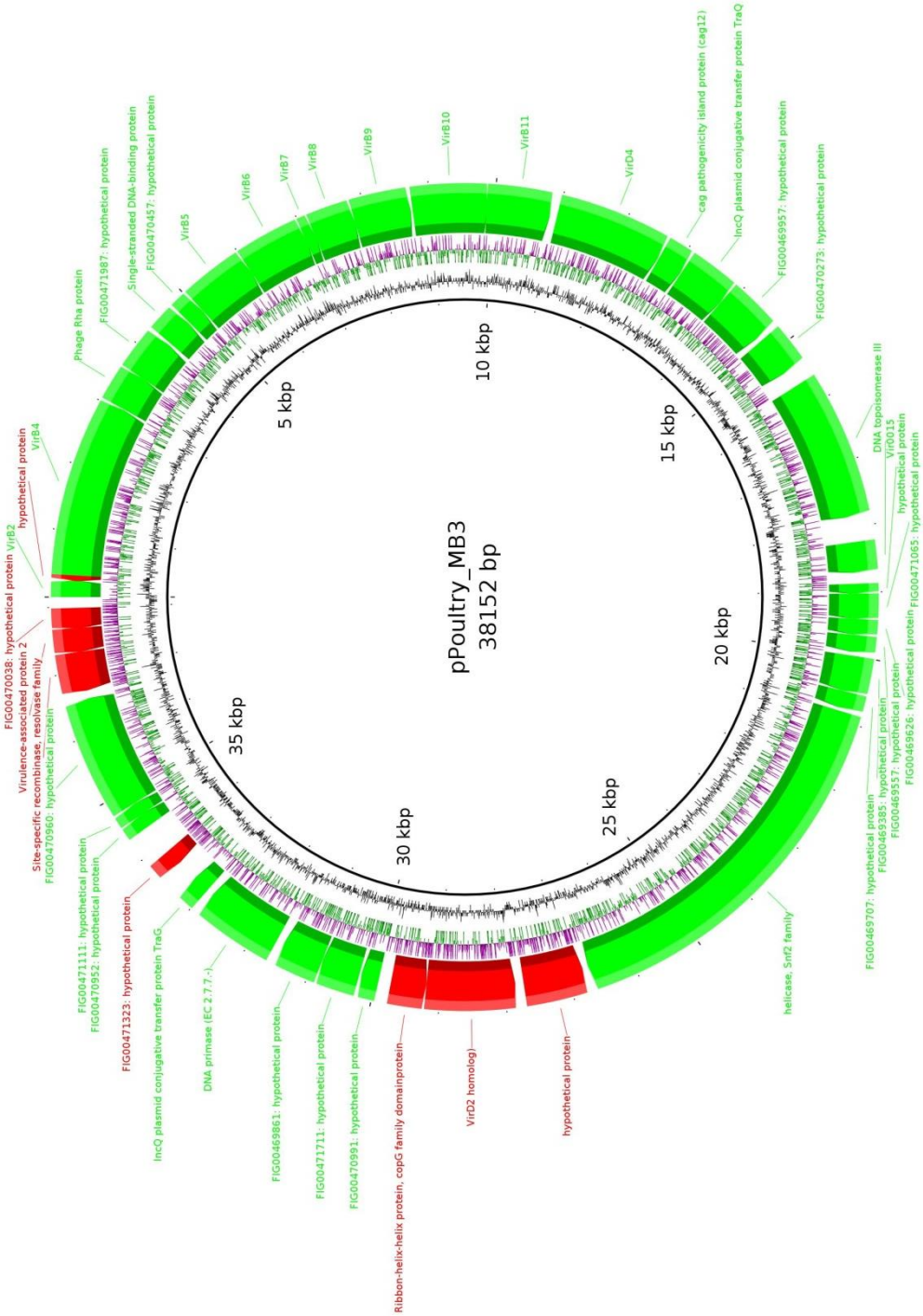

pPoultry_A8_35_15A
47646 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
VirD4, ATPase required for T-DNA transfer
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469557: hypothetical protein
hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
FIG00470802: hypothetical protein
DNA topoisomerase III
Tetracycline resistance protein TetD
hypothetical protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471337: hypothetical protein
helicase, Snf2 family
FIG00469707: hypothetical protein
FIG00469385: hypothetical protein
FIG00469557: hypothetical protein
FIG00469644: hypothetical protein
FIG00470281: hypothetical protein
hypothetical protein
FIG00638667: hypothetical protein
FIG00638667: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domain protein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
hypothetical protein
Virulence associated protein D (vapD)
hypothetical protein
Site-specific recombinase, resolvase family
hypothetical protein
Virulence-associated protein 2 FIG00470038: hypothetical protein
VirB2
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11

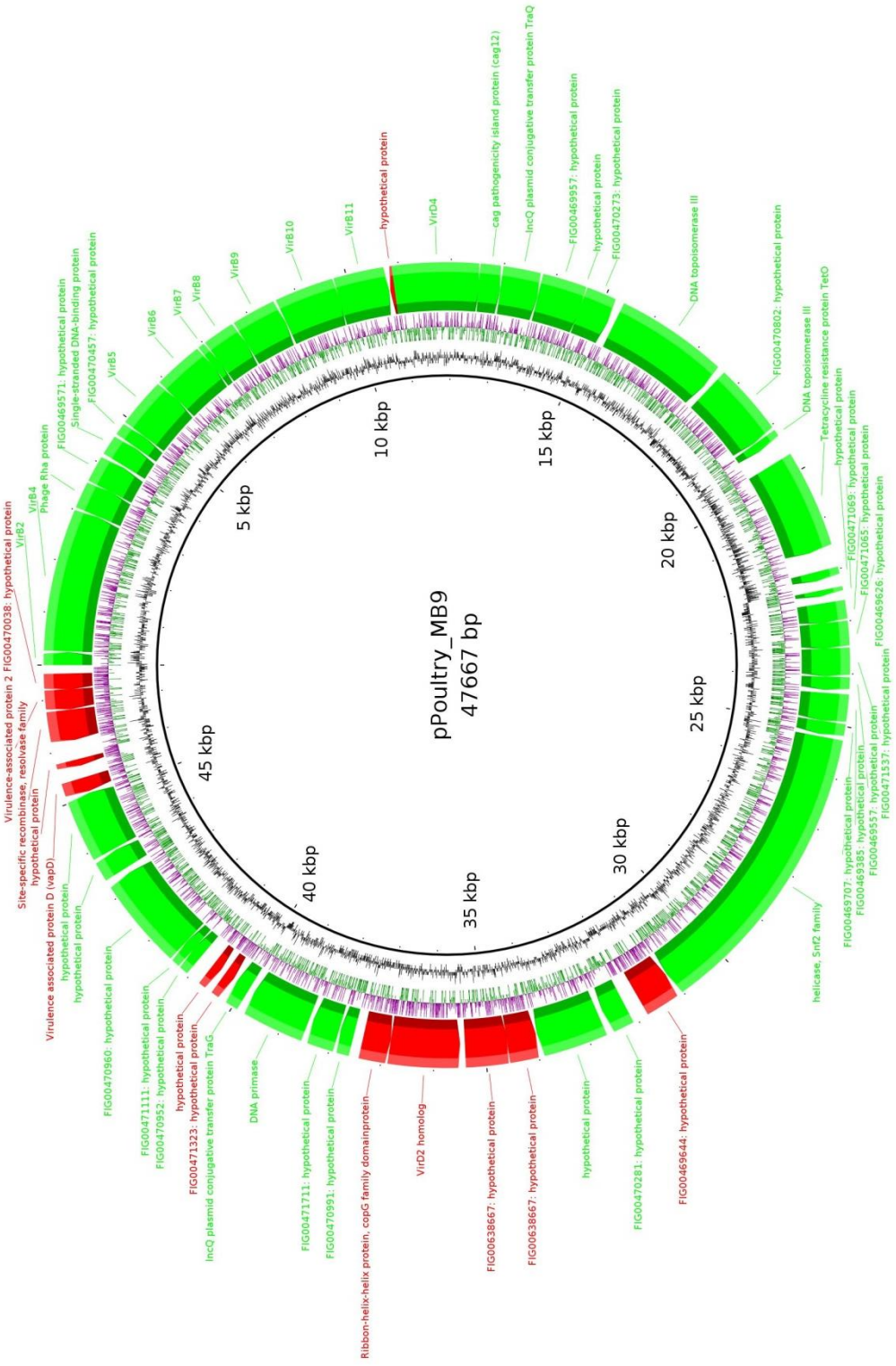

pPoultry_C1_C_2
47647 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
FIG00470802: hypothetical protein
DNA topoisomerase III
Tetracycline resistance protein TetO
hypothetical protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
helicase, Snf2 family
FIG00469707: hypothetical protein
FIG00469383: hypothetical protein
FIG00469557: hypothetical protein
FIG00469644: hypothetical protein
FIG00470281: hypothetical protein
hypothetical protein
FIG00638667: hypothetical protein
FIG00638667: hypothetical protein
VirD2 homolog)
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
hypothetical protein
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
hypothetical protein
hypothetical protein
Virulence associated protein D (vapD)
hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2 FIG00470038: hypothetical protein
VirB2
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11
VirD4


pPoultry_C5_T2_8
47646 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
VirB11
VirD4
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
FIG00470802: hypothetical protein
DNA topoisomerase III
Tetracycline resistance protein TetO
hypothetical protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00469644: hypothetical protein
FIG00470281: hypothetical protein
hypothetical protein
FIG00638667: hypothetical protein
FIG00638667: hypothetical protein
VirD2 homolog)
Ribbon-helix-helix protein, copG family domainprotein
FIG00471711: hypothetical protein
FIG00470991: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
hypothetical protein
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
hypothetical protein
hypothetical protein
Virulence associated protein D (vapD)
hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2 FIG00470038: hypothetical protein
VirB2
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10


pPoultry_MB12
37539 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
VirB2
hypothetical protein
hypothetical protein
VirB4
Phage Rha protein
hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11
VirD4
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
hypothetical protein
VirD015
hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00469644: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domain protein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
FIG00469861: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein


pPoultry_MB18
44766 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
hypothetical protein
VirD4
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469557: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
FIG00470802: hypothetical protein
DNA topoisomerase III
Tetracycline resistance protein TetO
FIG00471024: hypothetical protein TolA protein
replication protein
FIG00471069: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00471537: hypothetical protein
FIG00469557: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
FIG00472625: hypothetical protein
FIG00469644: hypothetical protein
VirD2 homolog
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
DNA primase
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein
hypothetical protein
VirB2
VirB4
Phage Rha protein
FIG00469571: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11


pPoultry_MB3
38152 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11
VirD4
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
VirD15
hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
FIG00469537: hypothetical protein
FIG00469385: hypothetical protein
FIG00469707: hypothetical protein
helicase, Snf2 family
hypothetical protein
VirD2 homolog)
Ribbon-helix-helix protein, copG family domainprotein
FIG00470991: hypothetical protein
FIG00471711: hypothetical protein
FIG00469861: hypothetical protein
DNA primase (EC 2.7.7.-)
IncQ plasmid conjugative transfer protein TraG
FIG00471323: hypothetical protein
FIG00470952: hypothetical protein
FIG00471111: hypothetical protein
FIG00470960: hypothetical protein
Site-specific recombinase, resolvase family
Virulence-associated protein 2
FIG00470038: hypothetical protein
VirB2
hypothetical protein
VirB4
Phage Rha protein
FIG00471987: hypothetical protein
Single-stranded DNA-binding protein
FIG00470457: hypothetical protein
VirB5


pPoultry_MB9
47667 bp
5 kbp
10 kbp
15 kbp
20 kbp
25 kbp
30 kbp
35 kbp
40 kbp
45 kbp
hypothetical protein
VirB11
VirB10
VirB9
VirB8
VirB7
VirB6
VirB5
FIG00470457: hypothetical protein
Single-stranded DNA-binding protein
FIG00469571: hypothetical protein
Phage Rha protein
VirB4
VirB2
Virulence-associated protein 2 FIG00470038: hypothetical protein
Site-specific recombinase, resolvase family
hypothetical protein
Virulence associated protein D (vapD)
hypothetical protein
hypothetical protein
FIG00470960: hypothetical protein
FIG00471111: hypothetical protein
FIG00470952: hypothetical protein
hypothetical protein
FIG00471323: hypothetical protein
IncQ plasmid conjugative transfer protein TraG
DNA primase
FIG00471711: hypothetical protein
FIG00470991: hypothetical protein
Ribbon-helix-helix protein, copG family domain protein
VirD2 homolog
FIG00638667: hypothetical protein
FIG00638667: hypothetical protein
hypothetical protein
FIG00470281: hypothetical protein
FIG00469644: hypothetical protein
VirD4
cag pathogenicity island protein (cag12)
IncQ plasmid conjugative transfer protein TraQ
FIG00469957: hypothetical protein
hypothetical protein
FIG00470273: hypothetical protein
DNA topoisomerase III
FIG00470802: hypothetical protein
DNA topoisomerase III
Tetracycline resistance protein TetO
hypothetical protein
FIG00471068: hypothetical protein
FIG00471065: hypothetical protein
FIG00469626: hypothetical protein
helicase, Snf2 family
FIG00469707: hypothetical protein
FIG00469385: hypothetical protein
FIG00469557: hypothetical protein
FIG00471537: hypothetical protein

S3 - Annotated genes in the 19 newly sequenced plasmids and in the plasmid pTet (NC_008790.1)**.** Plasmids have been linearised at the start of the virB2 gene. Each plasmid annotation shows the start nucleotide, the end nucleotide, the gene orientation (+/- = sense or antisense), gene name (as reported by RAST automated annotation pipeline).

pTet			
1	264	+	"VirB2"
249	371	-	"hypothetical protein"
337	3045	+	"VirB4"
3056	3622	+	"Phage Rha protein"
3619	4284	+	"FIG00471987: hypothetical protein"
4318	4743	+	"Single-stranded DNA-binding protein"
4759	5034	+	"FIG00470457: hypothetical protein"
5038	6009	+	"VirB5"
6006	7004	+	"VirB6"
6979	7161	+	"VirB7"
7151	7813	+	"VirB8"
7810	8697	+	"VirB9"
8703	9899	+	"VirB10"
9880	10872	+	"VirB11"
10873	12684	+	"VirD4, ATPase required forT-DNA transfer"
12698	13102	+	"cag pathogenicity island protein (cag12)"
13114	13878	+	"IncQ plasmid conjugative transfer protein TraQ"
13881	14678	+	"FIG00469957: hypothetical protein"
14749	15369	+	"FIG00470273: hypothetical protein"
15633	17825	+	"DNA topoisomerase III (EC 5.99.1.2)"
17931	19352	+	"FIG00470802: hypothetical protein"
19371	19550	+	"DNA topoisomerase III (EC 5.99.1.2)"
19909	21828	+	"Tetracycline resistance protein TetO"
22066	22179	+	"hypothetical protein"
22251	22919	+	"TolA protein"
22878	23261	-	"FIG00471024: hypothetical protein"
23214	23636	+	"hypothetical protein"
23725	24798	+	"replication protein"
25093	25455	+	"FIG00471069: hypothetical protein"
25597	25995	+	"FIG00471065: hypothetical protein"
25992	26504	+	"FIG00469626: hypothetical protein"
26579	27058	+	"FIG00471537: hypothetical protein"
27063	27329	+	"FIG00469557: hypothetical protein"
27388	27948	+	"FIG00469385: hypothetical protein"
27951	28217	+	"FIG00469707: hypothetical protein"
28234	34062	+	"helicase, Snf2 family"
34084	34788	-	"FIG00469644: hypothetical protein"
34814	36202	-	"VirD2 homolog"
36202	36753	-	"Ribbon-helix-helix protein, copG family domainprotein"
36954	37235	+	"FIG00470991: hypothetical protein"
37264	37875	+	"FIG00471711: hypothetical protein"
37879	38532	+	"FIG00469861: hypothetical protein"
38651	39877	+	"DNA primase (EC 2.7.7.-)"
39943	40248	+	"IncQ plasmid conjugative transfer protein TraG"
40470	40727	-	"FIG00471323: hypothetical protein"
40761	40946	-	"hypothetical protein"
41219	41437	+	"FIG00470952: hypothetical protein"
41430	41633	+	"FIG00471111: hypothetical protein"
41637	43430	+	"FIG00470960: hypothetical protein"
43553	44167	-	"Site-specific recombinase, resolvase family"
44164	44541	-	"Virulence-associated protein 2"
44532	44855	-	"FIG00470038: hypothetical protein"

pPoultryC3_T_25			
1	264	+	"VirB2"
277	3045	+	"VirB4"
3056	3613	+	"Phage Rha protein"
3610	4161	+	"FIG00469571: hypothetical protein"
4195	4620	+	"Single-stranded DNA-binding protein"
4636	4911	+	"FIG00470457: hypothetical protein"
5170	5904	+	"VirB5"
5901	6896	+	"VirB6"
6886	7050	+	"VirB7"
7040	7699	+	"VirB8"
7696	8586	+	"VirB9"
8596	9774	+	"VirB10"
9755	10762	+	"VirB11"
10871	12574	+	"VirD4"
12555	12986	+	"cag pathogenicity island protein (cag12)"
12989	13747	+	"IncQ plasmid conjugative transfer protein TraQ"
13750	14547	+	"FIG00469957: hypothetical protein"
14501	14644	+	"hypothetical protein"
14622	15242	+	"FIG00470273: hypothetical protein"
15371	17563	+	"DNA topoisomerase III (EC 5.99.1.2)"
17669	19090	+	"FIG00470802: hypothetical protein"
19109	19288	+	"DNA topoisomerase III (EC 5.99.1.2)"
19647	21566	+	"Tetracycline resistance protein TetO"
21961	22143	+	"hypothetical protein"
22339	22458	+	"FIG00471069: hypothetical protein"
22600	22998	+	"FIG00471065: hypothetical protein"
22995	23507	+	"FIG00469626: hypothetical protein"
23504	24061	+	"FIG00471537: hypothetical protein"
24066	24332	+	"FIG00469557: hypothetical protein"
24391	24951	+	"FIG00469385: hypothetical protein"
24954	25220	+	"FIG00469707: hypothetical protein"
25367	25846	+	"hypothetical protein"
30779	31192	+	"helicase, Snf2 family"
31214	31918	-	"FIG00469644: hypothetical protein"
32225	32677	+	"FIG00470281: hypothetical protein"
32781	34061	+	"hypothetical protein"
34085	34678	-	"FIG00638667: hypothetical protein"
34663	35517	-	"FIG00638667: hypothetical protein"
35613	37001	-	"VirD2 homolog"
37001	37552	-	"Ribbon-helix-helix protein, copG family domain protein"
37746	38027	+	"FIG00470991: hypothetical protein"
38055	38606	+	"FIG00471711: hypothetical protein"
38686	39912	+	"DNA primase (EC 2.7.7.-)"
40030	40323	+	"IncQ plasmid conjugative transfer protein TraG"
40402	40602	-	"FIG00471323: hypothetical protein"
40711	40896	-	"hypothetical protein"
41190	41408	+	"FIG00470952: hypothetical protein"
41401	41604	+	"FIG00471111: hypothetical protein"
41608	43401	+	"FIG00470960: hypothetical protein"
43516	43932	+	"hypothetical protein"
43966	45192	+	"hypothetical protein"
45226	45513	-	"Virulence associated protein D (vapD)"
45762	45899	-	"hypothetical protein"
46161	46775	-	"Site-specific recombinase, resolvase family"
46781	47185	-	"Virulence-associated protein 2"
47176	47499	-	"FIG00470038: hypothetical protein"

pCj2			
1	264	+	"VirB2"
277	3045	+	"VirB4"
3056	3607	+	"Phage Rha protein"
3604	4170	+	"hypothetical protein"
4204	4629	+	"Single-stranded DNA-binding protein"
4645	4920	+	"FIG00470457: hypothetical protein"
4924	5895	+	"VirB5"
5892	6890	+	"VirB6"
6880	7047	+	"VirB7"
7037	7699	+	"VirB8"
7696	8583	+	"VirB9"
8610	9785	+	"VirB10"
9766	10758	+	"VirB11"
10867	12570	+	"VirD4"
12584	12988	+	"cag pathogenicity island protein (cag12)"
13000	13764	+	"IncQ plasmid conjugative transfer protein TraQ"
13767	14564	+	"FIG00469957: hypothetical protein"
14635	15255	+	"FIG00470273: hypothetical protein"
15518	17710	+	"DNA topoisomerase III (EC 5.99.1.2)"
17816	19237	+	"FIG00470802: hypothetical protein"
19256	19435	+	"DNA topoisomerase III (EC 5.99.1.2)"
19794	21713	+	"Tetracycline resistance protein TetO"
21892	22245	+	"FIG00471065: hypothetical protein"
22242	22754	+	"FIG00469626: hypothetical protein"
22751	23308	+	"FIG00471537: hypothetical protein"
23313	23579	+	"FIG00469557: hypothetical protein"
23638	24198	+	"FIG00469385: hypothetical protein"
24201	24467	+	"FIG00469707: hypothetical protein"
24614	25093	+	"hypothetical protein"
29405	30439	+	"helicase, Snf2 family"
30576	31304	+	"FIG00472625: hypothetical protein"
31279	32019	-	"FIG00469644: hypothetical protein"
32042	33430	-	"VirD2 homolog"
33430	33981	-	"Ribbon-helix-helix protein, copG family domainprotein"
34175	34456	+	"FIG00470991: hypothetical protein"
34485	35096	+	"FIG00471711: hypothetical protein"
35100	35753	+	"FIG00469861: hypothetical protein"
35841	37070	+	"DNA primase (EC 2.7.7.-)"
37188	37451	+	"IncQ plasmid conjugative transfer protein TraG"
37902	38525	+	"hypothetical protein"
38912	39052	-	"hypothetical protein"
39788	40006	+	"FIG00470952: hypothetical protein"
39999	40202	+	"FIG00471111: hypothetical protein"
40205	41998	+	"FIG00470960: hypothetical protein"
42080	42199	+	"hypothetical protein"
42208	42822	-	"Site-specific recombinase, resolvase family"
42819	43196	-	"Virulence-associated protein 2"
43187	43510	-	"FIG00470038: hypothetical protein"

pCj3			
1	264	+	"VirB2"
343	3051	+	"VirB4"
3062	3628	+	"Phage Rha protein"
3625	4191	+	"hypothetical protein"
4225	4647	+	"Single-stranded DNA-binding protein"
4663	4938	+	"FIG00470457: hypothetical protein"
4942	5931	+	"VirB5"
5928	6923	+	"VirB6"
6913	7077	+	"VirB7"
7067	7729	+	"VirB8"
7726	8613	+	"VirB9"
8640	9815	+	"VirB10"
9796	10788	+	"VirB11"
10789	12600	+	"VirD4, ATPase required forT-DNA transfer"
12581	13018	+	"cag pathogenicity island protein (cag12)"
13030	13794	+	"IncQ plasmid conjugative transfer protein TraQ"
13797	14594	+	"FIG00469957: hypothetical protein"
14665	14958	+	"FIG00470273: hypothetical protein"
17638	17868	+	"DNA topoisomerase III, TraE-type (EC 5.99.1.2)"
17974	19395	+	"FIG00470802: hypothetical protein"
19414	19593	+	"DNA topoisomerase III (EC 5.99.1.2)"
19952	21871	+	"Tetracycline resistance protein TetO"
22050	22403	+	"FIG00471065: hypothetical protein"
22400	22912	+	"FIG00469626: hypothetical protein"
22909	23466	+	"FIG00471537: hypothetical protein"
23471	23737	+	"FIG00469557: hypothetical protein"
23796	24356	+	"FIG00469385: hypothetical protein"
24359	24625	+	"FIG00469707: hypothetical protein"
24672	30470	+	"helicase, Snf2 family"
30607	31335	+	"FIG00472625: hypothetical protein"
31310	32050	-	"FIG00469644: hypothetical protein"
32073	33461	-	"VirD2 homolog"
33461	34012	-	"Ribbon-helix-helix protein, copG family domainprotein"
34206	34487	+	"FIG00470991: hypothetical protein"
34515	35066	+	"FIG00471711: hypothetical protein"
35129	35782	+	"FIG00469861: hypothetical protein"
35901	37127	+	"DNA primase (EC 2.7.7.-)"
37234	37527	+	"IncQ plasmid conjugative transfer protein TraG"
37623	37919	-	"FIG00471323: hypothetical protein"
37932	38117	-	"hypothetical protein"
38390	38608	+	"FIG00470952: hypothetical protein"
38601	38804	+	"FIG00471111: hypothetical protein"
38808	40601	+	"FIG00470960: hypothetical protein"
40722	41336	-	"Site-specific recombinase, resolvase family"
41333	41500	-	"Virulence-associated protein 2"
41519	41842	-	"FIG00470038: hypothetical protein"

pGoose222			
1	264	+	"VirB2"
249	377	-	"hypothetical protein"
561	680	-	"hypothetical protein"
700	3051	+	"VirB4"
3062	3619	+	"Phage Rha protein"
3616	4182	+	"hypothetical protein"
4216	4641	+	"Single-stranded DNA-binding protein"
4657	4932	+	"FIG00470457: hypothetical protein"
4936	5907	+	"VirB5"
5904	6902	+	"VirB6"
6892	7059	+	"VirB7"
7049	7711	+	"VirB8"
7708	8595	+	"VirB9"
8622	9797	+	"VirB10"
9778	10770	+	"VirB11"
10879	12582	+	"VirD4, ATPase required forT-DNA transfer"
12563	13000	+	"cag pathogenicity island protein (cag12)"
13012	13776	+	"IncQ plasmid conjugative transfer protein TraQ"
13779	14576	+	"FIG00469957: hypothetical protein"
15657	17849	+	"DNA topoisomerase III (EC 5.99.1.2)"
17955	19376	+	"FIG00470802: hypothetical protein"
19395	19574	+	"DNA topoisomerase III (EC 5.99.1.2)"
19933	21852	+	"Tetracycline resistance protein TetO"
22247	22429	+	"hypothetical protein"
22281	22919	+	"TolA protein"
22878	23291	-	"FIG00471024: hypothetical protein"
23754	24977	+	"replication protein"
25329	25562	+	"FIG00471069: hypothetical protein"
25522	25875	+	"FIG00471065: hypothetical protein"
25872	26384	+	"FIG00469626: hypothetical protein"
26381	26938	+	"FIG00471537: hypothetical protein"
26943	27209	+	"FIG00469557: hypothetical protein"
27268	27828	+	"FIG00469385: hypothetical protein"
27831	28097	+	"FIG00469707: hypothetical protein"
28144	33942	+	"helicase, Snf2 family"
33964	34668	-	"FIG00469644: hypothetical protein"
34695	36083	-	"VirD2 homolog"
36083	36634	-	"Ribbon-helix-helix protein, copG family domainprotein"
36828	37109	+	"FIG00470991: hypothetical protein"
37138	37749	+	"FIG00471711: hypothetical protein"
37753	38412	+	"FIG00469861: hypothetical protein"
38494	39726	+	"DNA primase (EC 2.7.7.-)"
39841	40104	+	"IncQ plasmid conjugative transfer protein TraG"
40264	40473	-	"hypothetical protein"
40623	41795	-	"Serine/threonine protein kinase PrkC, regulatorof stationary phase"
41807	41989	-	"hypothetical protein"
42523	42741	+	"FIG00470952: hypothetical protein"
42832	44625	+	"FIG00470960: hypothetical protein"
44759	45055	-	"hypothetical protein"
45333	45881	-	"Site-specific recombinase, resolvase family"
45944	46321	-	"Virulence-associated protein 2"
46312	46635	-	"FIG00470038: hypothetical protein"

pHi40620300			
1	264	+	"VirB2"
249	377	-	"hypothetical protein"
561	680	-	"hypothetical protein"
700	3051	+	"VirB4"
3062	3628	+	"Phage Rha protein"
3625	4176	+	"FIG00469571: hypothetical protein"
4210	4632	+	"Single-stranded DNA-binding protein"
4648	4923	+	"FIG00470457: hypothetical protein"
4927	5913	+	"VirB5"
5917	6909	+	"VirB6"
6899	7063	+	"VirB7"
7053	7712	+	"VirB8"
7709	8599	+	"VirB9"
8609	9787	+	"VirB10"
9768	10775	+	"VirB11"
10776	12587	+	"VirD4, ATPase required forT-DNA transfer"
12568	12999	+	"cag pathogenicity island protein (cag12)"
13002	13760	+	"IncQ plasmid conjugative transfer protein TraQ"
13763	14560	+	"FIG00469957: hypothetical protein"
14636	15256	+	"FIG00470273: hypothetical protein"
15385	17577	+	"DNA topoisomerase III (EC 5.99.1.2)"
17683	19104	+	"FIG00470802: hypothetical protein"
19123	19302	+	"DNA topoisomerase III (EC 5.99.1.2)"
19661	21580	+	"Tetracycline resistance protein TetO"
21992	22660	+	"TolA protein"
22619	23032	-	"FIG00471024: hypothetical protein"
23696	24844	+	"replication protein"
25196	25429	+	"FIG00471069: hypothetical protein"
25389	25742	+	"FIG00471065: hypothetical protein"
25739	26251	+	"FIG00469626: hypothetical protein"
26248	26805	+	"FIG00471537: hypothetical protein"
26810	27076	+	"FIG00469557: hypothetical protein"
27135	27695	+	"FIG00469385: hypothetical protein"
27698	27964	+	"FIG00469707: hypothetical protein"
28011	33809	+	"helicase, Snf2 family"
33946	34674	+	"FIG00472625: hypothetical protein"
34649	35389	-	"FIG00469644: hypothetical protein"
35412	36800	-	"VirD2 homolog"
36800	37351	-	"Ribbon-helix-helix protein, copG family domainprotein"
37545	37826	+	"FIG00470991: hypothetical protein"
37854	38465	+	"FIG00471711: hypothetical protein"
38486	39712	+	"DNA primase (EC 2.7.7.-)"
39830	40108	+	"IncQ plasmid conjugative transfer protein TraG"
40224	40499	-	"FIG00471323: hypothetical protein"
40991	41209	+	"FIG00470952: hypothetical protein"
41202	41405	+	"FIG00471111: hypothetical protein"
41409	43202	+	"FIG00470960: hypothetical protein"
43332	43946	-	"Site-specific recombinase, resolvase family"
43943	44350	-	"Virulence-associated protein 2"
44341	44664	-	"FIG00470038: hypothetical protein"

pPoultry_A1_CF_12			
1	264	+	"VirB2"
277	3045	+	"VirB4"
3056	3613	+	"Phage Rha protein"
3610	4161	+	"FIG00469571: hypothetical protein"
4195	4620	+	"Single-stranded DNA-binding protein"
4636	4911	+	"FIG00470457: hypothetical protein"
4915	5904	+	"VirB5"
5901	6896	+	"VirB6"
6886	7050	+	"VirB7"
7040	7699	+	"VirB8"
7696	8586	+	"VirB9"
8861	9901	+	"VirB10"
9882	10889	+	"VirB11"
10998	12701	+	"VirD4, ATPase required forT-DNA transfer"
12682	13113	+	"cag pathogenicity island protein (cag12)"
13116	13874	+	"IncQ plasmid conjugative transfer protein TraQ(RP4 TrbM homolog)"
13877	14674	+	"FIG00469957: hypothetical protein"
14749	15369	+	"FIG00470273: hypothetical protein"
15498	17690	+	"DNA topoisomerase III (EC 5.99.1.2)"
17796	19217	+	"FIG00470802: hypothetical protein"
19236	19415	+	"DNA topoisomerase III (EC 5.99.1.2)"
19774	21693	+	"Tetracycline resistance protein TetO"
22088	22270	+	"hypothetical protein"
22466	22585	+	"FIG00471069: hypothetical protein"
22727	23125	+	"FIG00471065: hypothetical protein"
23122	23634	+	"FIG00469626: hypothetical protein"
23631	24188	+	"FIG00471537: hypothetical protein"
24193	24459	+	"FIG00469557: hypothetical protein"
24518	25078	+	"FIG00469385: hypothetical protein"
25081	25347	+	"FIG00469707: hypothetical protein"
25394	31192	+	"helicase, Snf2 family"
31214	31918	-	"FIG00469644: hypothetical protein"
32225	32677	+	"FIG00470281: hypothetical protein"
32781	34061	+	"hypothetical protein"
34085	34678	-	"FIG00638667: hypothetical protein"
34663	35517	-	"FIG00638667: hypothetical protein"
35613	37001	-	"VirD2 homolog"
37001	37552	-	"Ribbon-helix-helix protein, copG family domainprotein"
37746	38027	+	"FIG00470991: hypothetical protein"
38055	38666	+	"FIG00471711: hypothetical protein"
38687	39913	+	"DNA primase (EC 2.7.7.-)"
40031	40324	+	"IncQ plasmid conjugative transfer protein TraG"
40403	40603	-	"FIG00471323: hypothetical protein"
40712	40897	-	"hypothetical protein"
41170	41388	+	"FIG00470952: hypothetical protein"
41381	41584	+	"FIG00471111: hypothetical protein"
41588	43381	+	"FIG00470960: hypothetical protein"
43496	43912	+	"hypothetical protein"
43946	45172	+	"hypothetical protein"
45206	45493	-	"Virulence associated protein D (vapD)"
45742	45879	-	"hypothetical protein"
46141	46755	-	"Site-specific recombinase, resolvase family"
46761	47165	-	"Virulence-associated protein 2"
47156	47479	-	"FIG00470038: hypothetical protein"

pPoultry_A6_T2_15			
1	264	+	"VirB2"
277	3045	+	"VirB4"
3056	3613	+	"Phage Rha protein"
3610	4161	+	"FIG00469571: hypothetical protein"
4195	4620	+	"Single-stranded DNA-binding protein"
4636	4911	+	"FIG00470457: hypothetical protein"
4915	5904	+	"VirB5"
5901	6896	+	"VirB6"
6886	7050	+	"VirB7"
7040	7699	+	"VirB8"
7696	8586	+	"VirB9"
8596	9774	+	"VirB10"
9755	10762	+	"VirB11"
10871	12574	+	"VirD4, ATPase required forT-DNA transfer"
12555	12986	+	"cag pathogenicity island protein (cag12)"
12989	13747	+	"IncQ plasmid conjugative transfer protein TraQ"
13750	14547	+	"FIG00469957: hypothetical protein"
14622	15242	+	"FIG00470273: hypothetical protein"
15371	17563	+	"DNA topoisomerase III"
17669	19090	+	"FIG00470802: hypothetical protein"
19109	19288	+	"DNA topoisomerase III"
19647	21566	+	"Tetracycline resistance protein TetO"
21961	22143	+	"hypothetical protein"
22339	22458	+	"FIG00471069: hypothetical protein"
22600	22998	+	"FIG00471065: hypothetical protein"
22995	23507	+	"FIG00469626: hypothetical protein"
23504	24061	+	"FIG00471537: hypothetical protein"
24066	24332	+	"FIG00469557: hypothetical protein"
24391	24951	+	"FIG00469385: hypothetical protein"
24954	25220	+	"FIG00469707: hypothetical protein"
25267	31065	+	"helicase, Snf2 family"
31087	31791	-	"FIG00469644: hypothetical protein"
32098	32550	+	"FIG00470281: hypothetical protein"
32654	33934	+	"hypothetical protein"
33958	34551	-	"FIG00638667: hypothetical protein"
34536	35390	-	"FIG00638667: hypothetical protein"
35486	36874	-	"VirD2 homolog)"
36874	37425	-	"Ribbon-helix-helix protein, copG family domainprotein"
37619	37900	+	"FIG00470991: hypothetical protein"
37928	38479	+	"FIG00471711: hypothetical protein"
38559	39785	+	"DNA primase"
39903	40196	+	"IncQ plasmid conjugative transfer protein TraG"
40275	40475	-	"FIG00471323: hypothetical protein"
40584	40769	-	"hypothetical protein"
41042	41260	+	"FIG00470952: hypothetical protein"
41253	41456	+	"FIG00471111: hypothetical protein"
41460	43253	+	"FIG00470960: hypothetical protein"
43368	43784	+	"hypothetical protein"
43818	45044	+	"hypothetical protein"
45078	45365	-	"Virulence associated protein D (vapD)"
45614	45751	-	"hypothetical protein"
46140	46754	-	"Site-specific recombinase, resolvase family"
46760	47164	-	"Virulence-associated protein 2"
47155	47478	-	"FIG00470038: hypothetical protein"

pPoultry_A8_35_15A			
1	264	+	"VirB2"
277	3045	+	"VirB4"
3056	3613	+	"Phage Rha protein"
3610	4161	+	"FIG00469571: hypothetical protein"
4195	4620	+	"Single-stranded DNA-binding protein"
4636	4911	+	"FIG00470457: hypothetical protein"
4915	5904	+	"VirB5"
5901	6896	+	"VirB6"
6886	7050	+	"VirB7"
7040	7699	+	"VirB8"
7696	8586	+	"VirB9"
8596	9774	+	"VirB10"
9755	10762	+	"VirB11"
10871	12574	+	"VirD4, ATPase required forT-DNA transfer"
12555	12986	+	"cag pathogenicity island protein (cag12)"
12989	13747	+	"IncQ plasmid conjugative transfer protein TraQ"
13750	14547	+	"FIG00469957: hypothetical protein"
14501	14644	+	"hypothetical protein"
14622	15242	+	"FIG00470273: hypothetical protein"
15371	17563	+	"DNA topoisomerase III"
17669	19090	+	"FIG00470802: hypothetical protein"
19109	19288	+	"DNA topoisomerase III"
19647	21566	+	"Tetracycline resistance protein TetO"
21961	22143	+	"hypothetical protein"
22339	22458	+	"FIG00471069: hypothetical protein"
22600	22998	+	"FIG00471065: hypothetical protein"
22995	23507	+	"FIG00469626: hypothetical protein"
23504	24061	+	"FIG00471537: hypothetical protein"
24066	24332	+	"FIG00469557: hypothetical protein"
24391	24951	+	"FIG00469385: hypothetical protein"
24954	25220	+	"FIG00469707: hypothetical protein"
25267	31065	+	"helicase, Snf2 family"
31087	31791	-	"FIG00469644: hypothetical protein"
32098	32550	+	"FIG00470281: hypothetical protein"
32654	33934	+	"hypothetical protein"
33958	34551	-	"FIG00638667: hypothetical protein"
34536	35390	-	"FIG00638667: hypothetical protein"
35486	36874	-	"VirD2 homolog)"
36874	37425	-	"Ribbon-helix-helix protein, copG family domainprotein"
37619	37900	+	"FIG00470991: hypothetical protein"
37928	38479	+	"FIG00471711: hypothetical protein"
38559	39785	+	"DNA primase"
39903	40196	+	"IncQ plasmid conjugative transfer protein TraG"
40275	40475	-	"FIG00471323: hypothetical protein"
40584	40769	-	"hypothetical protein"
41042	41260	+	"FIG00470952: hypothetical protein"
41253	41456	+	"FIG00471111: hypothetical protein"
41460	43253	+	"FIG00470960: hypothetical protein"
43368	43784	+	"hypothetical protein"
43818	45044	+	"hypothetical protein"
45078	45365	-	"Virulence associated protein D (vapD)"
45614	45751	-	"hypothetical protein"
46140	46754	-	"Site-specific recombinase, resolvase family"
46760	47164	-	"Virulence-associated protein 2"
47155	47478	-	"FIG00470038: hypothetical protein"

pPoultry_C1_C_2			
1	264	+	"VirB2"
277	3045	+	"VirB4"
3056	3613	+	"Phage Rha protein"
3610	4161	+	"FIG00469571: hypothetical protein"
4195	4620	+	"Single-stranded DNA-binding protein"
4636	4911	+	"FIG00470457: hypothetical protein"
5170	5904	+	"VirB5"
5901	6896	+	"VirB6"
6886	7050	+	"VirB7"
7040	7699	+	"VirB8"
7696	8586	+	"VirB9"
8596	9774	+	"VirB10"
9755	10762	+	"VirB11"
10871	12574	+	"VirD4, ATPase required forT-DNA transfer"
12555	12986	+	"cag pathogenicity island protein (cag12)"
12989	13747	+	"IncQ plasmid conjugative transfer protein TraQ"
13750	14547	+	"FIG00469957: hypothetical protein"
14622	15242	+	"FIG00470273: hypothetical protein"
15371	17563	+	"DNA topoisomerase III"
17669	19090	+	"FIG00470802: hypothetical protein"
19109	19288	+	"DNA topoisomerase III"
19647	21566	+	"Tetracycline resistance protein TetO"
21961	22143	+	"hypothetical protein"
22339	22458	+	"FIG00471069: hypothetical protein"
22600	22998	+	"FIG00471065: hypothetical protein"
22995	23507	+	"FIG00469626: hypothetical protein"
23504	24061	+	"FIG00471537: hypothetical protein"
24066	24332	+	"FIG00469557: hypothetical protein"
24391	24951	+	"FIG00469385: hypothetical protein"
24954	25220	+	"FIG00469707: hypothetical protein"
25267	31065	+	"helicase, Snf2 family"
31087	31791	-	"FIG00469644: hypothetical protein"
32098	32550	+	"FIG00470281: hypothetical protein"
32654	33934	+	"hypothetical protein"
33958	34551	-	"FIG00638667: hypothetical protein"
34536	34742	-	"FIG00638667: hypothetical protein"
35613	37001	-	"VirD2 homolog)"
37001	37552	-	"Ribbon-helix-helix protein, copG family domainprotein"
37746	38027	+	"FIG00470991: hypothetical protein"
38055	38666	+	"FIG00471711: hypothetical protein"
38687	39913	+	"DNA primase (EC 2.7.7.-)"
40031	40324	+	"IncQ plasmid conjugative transfer protein TraG"
40403	40603	-	"FIG00471323: hypothetical protein"
40712	40897	-	"hypothetical protein"
41170	41388	+	"FIG00470952: hypothetical protein"
41381	41584	+	"FIG00471111: hypothetical protein"
41588	43381	+	"FIG00470960: hypothetical protein"
43496	43912	+	"hypothetical protein"
43946	45172	+	"hypothetical protein"
45206	45493	-	"Virulence associated protein D (vapD)"
45742	45879	-	"hypothetical protein"
46141	46755	-	"Site-specific recombinase, resolvase family"
46761	47165	-	"Virulence-associated protein 2"
47156	47479	-	"FIG00470038: hypothetical protein"

pPoultry_C5_T2_8				pPoultry_MB3				pK2				pPet93_327				pPig_PS623			
1	264	+	"VirB2"	1	264	+	"VirB2"	1	264	+	"VirB2"	1	264	+	"VirB2"	1	264	+	"VirB2"
277	3045	+	"VirB4"	249	371	-	"hypothetical protein"	249	377	-	"hypothetical protein"	249	371	-	"hypothetical protein"	249	371	-	"hypothetical protein"
3056	3613	+	"Phage Rha protein"	337	3045	+	"VirB4"	561	680	-	"hypothetical protein"	337	3045	+	"VirB4"	337	3045	+	"VirB4"
3610	4161	+	"FIG00469571: hypothetical protein"	3056	3622	+	"Phage Rha protein"	700	3051	+	"VirB4"	3056	3622	+	"Phage Rha protein"	3056	3622	+	"Phage Rha protein"
4195	4620	+	"Single-stranded DNA-binding protein"	3619	4284	+	"FIG00471987: hypothetical protein"	3062	3619	+	"Phage Rha protein"	3619	4284	+	"FIG00471987: hypothetical protein"	3619	4284	+	"FIG00471987: hypothetical protein"
4636	4911	+	"FIG00470457: hypothetical protein"	4318	4743	+	"Single-stranded DNA-binding protein"	3616	4182	+	"hypothetical protein"	4318	4743	+	"Single-stranded DNA-binding protein"	4318	4743	+	"Single-stranded DNA-binding protein"
5170	5904	+	"VirB5"	4759	5034	+	"FIG00470457: hypothetical protein"	4216	4638	+	"Single-stranded DNA-binding protein"	4759	5034	+	"FIG00470457: hypothetical protein"	4759	5034	+	"FIG00470457: hypothetical protein"
5901	6896	+	"VirB6"	5038	6009	+	"VirB5"	4654	4929	+	"FIG00470457: hypothetical protein"	5038	6009	+	"VirB5"	5038	6009	+	"VirB5"
6886	7050	+	"VirB7"	6006	7004	+	"VirB6"	4933	5919	+	"VirB5"	6120	7118	+	"VirB6"	6006	7004	+	"VirB6"
7040	7699	+	"VirB8"	6994	7161	+	"VirB7"	6019	7011	+	"VirB6"	7108	7275	+	"VirB7"	6994	7161	+	"VirB7"
7696	8586	+	"VirB9"	7151	7813	+	"VirB8"	7001	7168	+	"VirB7"	7265	7927	+	"VirB8"	7151	7813	+	"VirB8"
8596	9774	+	"VirB10"	7810	8697	+	"VirB9"	7158	7820	+	"VirB8"	7924	8811	+	"VirB9"	7810	8697	+	"VirB9"
9755	10762	+	"VirB11"	8724	9899	+	"VirB10"	7817	8704	+	"VirB9"	8838	10013	+	"VirB10"	8724	9899	+	"VirB10"
10871	12574	+	"VirD4, ATPase required forT-DNA transfer"	9880	10872	+	"VirB11"	8731	9903	+	"VirB10"	9994	10986	+	"VirB11"	9880	10872	+	"VirB11"
12555	12986	+	"cag pathogenicity island protein (cag12)"	10981	12684	+	"VirD4"	9884	10876	+	"VirB11"	11095	12798	+	"VirD4, ATPase required forT-DNA transfer"	10873	12684	+	"Coupling protein VirD4, ATPase required forT-DNA transfer"
12989	13747	+	"IncQ plasmid conjugative transfer protein TraQ"	12698	13102	+	"cag pathogenicity island protein (cag12)"	10877	12688	+	"VirD4, ATPase required forT-DNA transfer"	12812	13216	+	"cag pathogenicity island protein (cag12)"	12698	13102	+	"cag pathogenicity island protein (cag12)"
13750	14547	+	"FIG00469957: hypothetical protein"	13114	13878	+	"IncQ plasmid conjugative transfer protein TraQ"	12669	13106	+	"cag pathogenicity island protein (cag12)"	13228	13992	+	"IncQ plasmid conjugative transfer protein TraQ"	13114	13878	+	"IncQ plasmid conjugative transfer protein TraQ(RP4 TrbM homolog)"
14501	14644	+	"hypothetical protein"	13881	14678	+	"FIG00469957: hypothetical protein"	13118	13882	+	"IncQ plasmid conjugative transfer protein TraQ"	13995	14792	+	"FIG00469957: hypothetical protein"	13881	14678	+	"FIG00469957: hypothetical protein"
14622	15242	+	"FIG00470273: hypothetical protein"	14749	15369	+	"FIG00470273: hypothetical protein"	13885	14682	+	"FIG00469957: hypothetical protein"	14863	15483	+	"FIG00470273: hypothetical protein"	14749	15369	+	"FIG00470273: hypothetical protein"
15371	17563	+	"DNA topoisomerase III"	15633	17828	+	"DNA topoisomerase III"	14753	15373	+	"FIG00470273: hypothetical protein"	15747	17942	+	"DNA topoisomerase III (EC 5.99.1.2)"	15633	17825	+	"DNA topoisomerase III (EC 5.99.1.2)"
17669	19090	+	"FIG00470802: hypothetical protein"	18226	18699	+	"Vir0015"	15636	17828	+	"DNA topoisomerase III (EC 5.99.1.2)"	18340	18813	+	"Vir0015"	17931	19352	+	"FIG00470802: hypothetical protein"
19109	19288	+	"DNA topoisomerase III"	18871	19026	+	"hypothetical protein"	17934	19355	+	"FIG00470802: hypothetical protein"	18985	19140	+	"hypothetical protein"	19371	19550	+	"DNA topoisomerase III (EC 5.99.1.2)"
19647	21566	+	"Tetracycline resistance protein TetO"	19023	19418	+	"FIG00471065: hypothetical protein"	19374	19553	+	"DNA topoisomerase III (EC 5.99.1.2)"	19137	19532	+	"FIG00471065: hypothetical protein"	19909	21828	+	"Tetracycline resistance protein TetO"
21961	22143	+	"hypothetical protein"	19415	19660	+	"FIG00469626: hypothetical protein"	19912	21831	+	"Tetracycline resistance protein TetO"	19529	19774	+	"FIG00469626: hypothetical protein"	22251	22919	+	"TolA protein"
22339	22458	+	"FIG00471069: hypothetical protein"	19665	19931	+	"FIG00469557: hypothetical protein"	22254	22922	+	"TolA protein"	19779	20045	+	"FIG00469557: hypothetical protein"	22878	23291	-	"FIG00471024: hypothetical protein"
22600	22998	+	"FIG00471065: hypothetical protein"	19989	20549	+	"FIG00469385: hypothetical protein"	22881	23294	-	"FIG00471024: hypothetical protein"	20103	20663	+	"FIG00469385: hypothetical protein"	23681	24877	+	"replication protein"
22995	23507	+	"FIG00469626: hypothetical protein"	20552	20818	+	"FIG00469707: hypothetical protein"	23684	24880	+	"replication protein"	20666	20932	+	"FIG00469707: hypothetical protein"	25229	25462	+	"FIG00471069: hypothetical protein"
23504	24061	+	"FIG00471537: hypothetical protein"	20865	26660	+	"helicase, Snf2 family"	25232	25465	+	"FIG00471069: hypothetical protein"	20979	26777	+	"helicase, Snf2 family"	25422	25775	+	"FIG00471065: hypothetical protein"
24066	24332	+	"FIG00469557: hypothetical protein"	26752	27696	-	"hypothetical protein"	25425	25616	+	"FIG00471065: hypothetical protein"	26795	27505	-	"FIG00469644: hypothetical protein"	25772	25990	+	"FIG00469626: hypothetical protein"
24391	24951	+	"FIG00469385: hypothetical protein"	27826	29214	-	"VirD2 homolog)"	25653	25778	+	"FIG00471065: hypothetical protein"	27528	28916	-	"VirD2 homolog"	25995	26261	+	"FIG00469557: hypothetical protein"
24954	25220	+	"FIG00469707: hypothetical protein"	29214	29765	-	"Ribbon-helix-helix protein, copG family domainprotein"	25775	26287	+	"FIG00469626: hypothetical protein"	28916	29467	-	"Ribbon-helix-helix protein, copG family domainprotein"	26320	26880	+	"FIG00469385: hypothetical protein"
25267	31065	+	"helicase, Snf2 family"	29959	30240	+	"FIG00470991: hypothetical protein"	26284	26841	+	"FIG00471537: hypothetical protein"	29661	29942	+	"FIG00470991: hypothetical protein"	26883	27149	+	"FIG00469707: hypothetical protein"
31087	31791	-	"FIG00469644: hypothetical protein"	30269	30880	+	"FIG00471711: hypothetical protein"	26846	27112	+	"FIG00469557: hypothetical protein"	29971	30582	+	"FIG00471711: hypothetical protein"	27196	32991	+	"helicase, Snf2 family"
32098	32550	+	"FIG00470281: hypothetical protein"	30884	31537	+	"FIG00469861: hypothetical protein"	27171	27731	+	"FIG00469385: hypothetical protein"	30586	31239	+	"FIG00469861: hypothetical protein"	33128	33856	+	"FIG00472625: hypothetical protein"
32654	33934	+	"hypothetical protein"	31656	32882	+	"DNA primase (EC 2.7.7.-)"	27734	28000	+	"FIG00469707: hypothetical protein"	31358	32584	+	"DNA primase (EC 2.7.7.-)"	33866	34570	-	"FIG00469644: hypothetical protein"
33958	34551	-	"FIG00638667: hypothetical protein"	32990	33262	+	"IncQ plasmid conjugative transfer protein TraG"	28047	33845	+	"helicase, Snf2 family"	32692	32955	+	"IncQ plasmid conjugative transfer protein TraG"	34593	35981	-	"pTi VirD2 homolog"
34536	35390	-	"FIG00638667: hypothetical protein"	33598	33855	-	"FIG00471323: hypothetical protein"	33982	34710	+	"FIG00472625: hypothetical protein"	33177	33359	-	"FIG00471323: hypothetical protein"	35981	36532	-	"Ribbon-helix-helix protein, copG family domainprotein"
35486	36874	-	"VirD2 homolog)"	34347	34565	+	"FIG00470952: hypothetical protein"	34720	35424	-	"FIG00469644: hypothetical protein"	33926	34144	+	"FIG00470952: hypothetical protein"	36726	37007	+	"FIG00470991: hypothetical protein"
36874	37425	-	"Ribbon-helix-helix protein, copG family domainprotein"	34558	34761	+	"FIG00471111: hypothetical protein"	35447	36835	-	"VirD2 homolog"	34137	34340	+	"FIG00471111: hypothetical protein"	37036	37647	+	"FIG00471711: hypothetical protein"
37619	37900	+	"FIG00470991: hypothetical protein"	34765	36558	+	"FIG00470960: hypothetical protein"	36835	37386	-	"Ribbon-helix-helix protein, copG family domainprotein"	34344	36137	+	"FIG00470960: hypothetical protein"	37651	38304	+	"FIG00469861: hypothetical protein"
37928	38479	+	"FIG00471711: hypothetical protein"	36681	37295	-	"Site-specific recombinase, resolvase family"	37580	37861	+	"FIG00470991: hypothetical protein"	36260	36808	-	"Site-specific recombinase, resolvase family"	38579	38890	+	"FIG00469861: hypothetical protein"
38559	39785	+	"DNA primase (EC 2.7.7.-)"	37292	37669	-	"Virulence-associated protein 2"	37890	38501	+	"FIG00471711: hypothetical protein"	36871	37248	-	"Virulence-associated protein 2"	39009	40235	+	"DNA primase (EC 2.7.7.-)"
39903	40196	+	"IncQ plasmid conjugative transfer protein TraG"	37660	37983	-	"FIG00470038: hypothetical protein"	38505	39164	+	"FIG00469861: hypothetical protein"	37239	37562	-	"FIG00470038: hypothetical protein"	40343	40606	+	"IncQ plasmid conjugative transfer protein TraG"
40275	40475	-	"FIG00471323: hypothetical protein"					39247	40476	+	"DNA primase (EC 2.7.7.-)"					40828	41085	-	"FIG00471323: hypothetical protein"
40584	40769	-	"hypothetical protein"					40584	40862	+	"IncQ plasmid conjugative transfer protein TraG"					41577	41795	+	"FIG00470952: hypothetical protein"
41042	41260	+	"FIG00470952: hypothetical protein"					41023	41232	-	"hypothetical protein"					41788	41991	+	"FIG00471111: hypothetical protein"
41253	41456	+	"FIG00471111: hypothetical protein"					41382	42554	-	"Serine/threonine protein kinase PrkC, regulatorof stationary phase"					41995	43788	+	"FIG00470960: hypothetical protein"
41460	43253	+	"FIG00470960: hypothetical protein"					42566	42745	-	"hypothetical protein"					43911	44459	-	"Site-specific recombinase, resolvase family"
43368	43784	+	"hypothetical protein"					43282	43500	+	"FIG00470952: hypothetical protein"					44522	44899	-	"Virulence-associated protein 2"
43818	45044	+	"hypothetical protein"					43591	45384	+	"FIG00470960: hypothetical protein"					44890	45213	-	"FIG00470038: hypothetical protein"
45078	45365	-	"Virulence associated protein D (vapD)"					45388	45840	-	"hypothetical protein"								
45614	45751	-	"hypothetical protein"					45844	46020	-	"hypothetical protein"								
46013	46627	-	"Site-specific recombinase, resolvase family"					46122	46670	-	"Site-specific recombinase, resolvase family"								
46633	47037	-	"Virulence-associated protein 2"					46733	47110	-	"Virulence-associated protein 2"								
47028	47351	-	"FIG00470038: hypothetical protein"					47101	47424	-	"FIG00470038: hypothetical protein"								

pPig_PS762			
1	264	+	"VirB2"
249	377	-	"hypothetical protein"
343	3051	+	"VirB4"
3062	3628	+	"Phage Rha protein"
3625	4176	+	"FIG00469571: hypothetical protein"
4210	4632	+	"Single-stranded DNA-binding protein"
4648	4923	+	"FIG00470457: hypothetical protein"
4927	5913	+	"VirB5"
5917	6909	+	"VirB6"
6899	7063	+	"VirB7"
7053	7712	+	"VirB8"
7709	8599	+	"VirB9"
8609	9787	+	"VirB10"
9768	10775	+	"VirB11"
10884	12587	+	"VirD4"
12568	12999	+	"cag pathogenicity island protein (cag12)"
13002	13760	+	"IncQ plasmid conjugative transfer protein TraQ(RP4 TrbM homolog)"
14205	14687	+	"FIG00469957: hypothetical protein"
14758	15378	+	"FIG00470273: hypothetical protein"
15641	17833	+	"DNA topoisomerase III (EC 5.99.1.2)"
17939	19360	+	"FIG00470802: hypothetical protein"
19379	19558	+	"DNA topoisomerase III (EC 5.99.1.2)"
19917	21836	+	"Tetracycline resistance protein TetO"
22259	22927	+	"TolA protein"
22886	23299	-	"FIG00471024: hypothetical protein"
23788	24984	+	"replication protein"
25336	25569	+	"FIG00471069: hypothetical protein"
25529	25882	+	"FIG00471065: hypothetical protein"
25879	26391	+	"FIG00469626: hypothetical protein"
26388	26945	+	"FIG00471537: hypothetical protein"
26950	27216	+	"FIG00469557: hypothetical protein"
27275	27835	+	"FIG00469385: hypothetical protein"
27838	28104	+	"FIG00469707: hypothetical protein"
28151	33949	+	"helicase, Snf2 family"
34086	34814	+	"FIG00472625: hypothetical protein"
34824	35528	-	"FIG00469644: hypothetical protein"
35551	36939	-	"VirD2 homolog)"
36939	37490	-	"Ribbon-helix-helix protein, copG family domainprotein"
37684	37965	+	"FIG00470991: hypothetical protein"
37994	38605	+	"FIG00471711: hypothetical protein"
38609	39268	+	"FIG00469861: hypothetical protein"
39350	40576	+	"DNA primase (EC 2.7.7.-)"
40695	40973	+	"IncQ plasmid conjugative transfer protein TraG"
41089	41289	-	"FIG00471323: hypothetical protein"
41398	41577	-	"hypothetical protein"
41850	42068	+	"FIG00470952: hypothetical protein"
42061	42264	+	"FIG00471111: hypothetical protein"
42268	44061	+	"FIG00470960: hypothetical protein"
44191	44805	-	"Site-specific recombinase, resolvase family"
44802	45209	-	"Virulence-associated protein 2"
45200	45523	-	"FIG00470038: hypothetical protein"

pPoultry_MB9			
1	264	+	"VirB2"
277	3045	+	"VirB4"
3056	3613	+	"Phage Rha protein"
3610	4161	+	"FIG00469571: hypothetical protein"
4195	4620	+	"Single-stranded DNA-binding protein"
4636	4911	+	"FIG00470457: hypothetical protein"
4915	5904	+	"VirB5"
5901	6896	+	"VirB6"
6886	7050	+	"VirB7"
7040	7699	+	"VirB8"
7696	8586	+	"VirB9"
8596	9774	+	"VirB10"
9755	10762	+	"VirB11"
10783	10896	-	"hypothetical protein"
10871	12574	+	"VirD4, ATPase required forT-DNA transfer"
12555	12986	+	"cag pathogenicity island protein (cag12)"
12989	13747	+	"IncQ plasmid conjugative transfer protein TraQ"
13750	14547	+	"FIG00469957: hypothetical protein"
14501	14644	+	"hypothetical protein"
14622	15242	+	"FIG00470273: hypothetical protein"
15371	17563	+	"DNA topoisomerase III"
17669	19090	+	"FIG00470802: hypothetical protein"
19109	19288	+	"DNA topoisomerase III"
19647	21566	+	"Tetracycline resistance protein TetO"
21961	22143	+	"hypothetical protein"
22339	22458	+	"FIG00471069: hypothetical protein"
22600	22998	+	"FIG00471065: hypothetical protein"
22995	23507	+	"FIG00469626: hypothetical protein"
23504	24061	+	"FIG00471537: hypothetical protein"
24066	24332	+	"FIG00469557: hypothetical protein"
24391	24951	+	"FIG00469385: hypothetical protein"
24954	25220	+	"FIG00469707: hypothetical protein"
25267	31065	+	"helicase, Snf2 family"
31087	31791	-	"FIG00469644: hypothetical protein"
32098	32550	+	"FIG00470281: hypothetical protein"
32654	33934	+	"hypothetical protein"
33958	34551	-	"FIG00638667: hypothetical protein"
34536	35390	-	"FIG00638667: hypothetical protein"
35486	36874	-	"VirD2 homolog"
36874	37425	-	"Ribbon-helix-helix protein, copG family domainprotein"
37619	37900	+	"FIG00470991: hypothetical protein"
37928	38479	+	"FIG00471711: hypothetical protein"
38559	39785	+	"DNA primase (EC 2.7.7.-)"
39903	40196	+	"IncQ plasmid conjugative transfer protein TraG"
40275	40475	-	"FIG00471323: hypothetical protein"
40584	40769	-	"hypothetical protein"
41063	41281	+	"FIG00470952: hypothetical protein"
41274	41477	+	"FIG00471111: hypothetical protein"
41481	43274	+	"FIG00470960: hypothetical protein"
43389	43805	+	"hypothetical protein"
43839	45065	+	"hypothetical protein"
45099	45386	-	"Virulence associated protein D (vapD)"
45635	45772	-	"hypothetical protein"
46161	46775	-	"Site-specific recombinase, resolvase family"
46781	47185	-	"Virulence-associated protein 2"
47176	47499	-	"FIG00470038: hypothetical protein"

pPoultry_MB12			
1	264	+	"VirB2"
249	371	-	"hypothetical protein"
555	674	-	"hypothetical protein"
694	3045	+	"VirB4"
3056	3613	+	"Phage Rha protein"
3610	4176	+	"hypothetical protein"
4210	4635	+	"Single-stranded DNA-binding protein"
4651	4926	+	"FIG00470457: hypothetical protein"
4930	5901	+	"VirB5"
5898	6896	+	"VirB6"
6886	7053	+	"VirB7"
7043	7705	+	"VirB8"
7702	8589	+	"VirB9"
8616	9791	+	"VirB10"
9772	10764	+	"VirB11"
10765	12576	+	"VirD4"
12557	12982	+	"cag pathogenicity island protein (cag12)"
12994	13758	+	"IncQ plasmid conjugative transfer protein TraQ"
13761	14558	+	"FIG00469957: hypothetical protein"
14633	15253	+	"FIG00470273: hypothetical protein"
15517	17709	+	"DNA topoisomerase III"
17803	17976	+	"hypothetical protein"
18027	18500	+	"Vir0015"
18693	18848	+	"hypothetical protein"
18845	19240	+	"FIG00471065: hypothetical protein"
19237	19488	+	"FIG00469626: hypothetical protein"
19493	19759	+	"FIG00469557: hypothetical protein"
19819	20376	+	"FIG00469385: hypothetical protein"
20383	20649	+	"FIG00469707: hypothetical protein"
20696	26494	+	"helicase, Snf2 family"
26512	27219	-	"FIG00469644: hypothetical protein"
27242	28630	-	"VirD2 homolog"
28630	29181	-	"Ribbon-helix-helix protein, copG family domainprotein"
29375	29656	+	"FIG00470991: hypothetical protein"
29685	30296	+	"FIG00471711: hypothetical protein"
30300	30959	+	"FIG00469861: hypothetical protein"
31042	32268	+	"DNA primase (EC 2.7.7.-)"
32386	32649	+	"IncQ plasmid conjugative transfer protein TraG"
32985	33242	-	"FIG00471323: hypothetical protein"
33734	33952	+	"FIG00470952: hypothetical protein"
33945	34148	+	"FIG00471111: hypothetical protein"
34152	35945	+	"FIG00470960: hypothetical protein"
36068	36682	-	"Site-specific recombinase, resolvase family"
36679	37056	-	"Virulence-associated protein 2"
37047	37370	-	"FIG00470038: hypothetical protein"

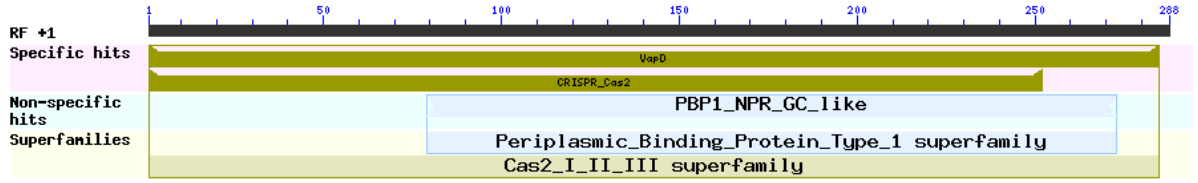
pPoultry_MB18			
1	264	+	"VirB2"
249	377	-	"hypothetical protein"
561	680	-	"hypothetical protein"
700	3051	+	"VirB4"
3062	3619	+	"Phage Rha protein"
3616	4167	+	"FIG00469571: hypothetical protein"
4201	4623	+	"Single-stranded DNA-binding protein"
4639	4914	+	"FIG00470457: hypothetical protein"
4918	5907	+	"VirB5"
5904	6899	+	"VirB6"
6889	7053	+	"VirB7"
7043	7702	+	"VirB8"
7699	8589	+	"VirB9"
8599	9777	+	"VirB10"
9758	10765	+	"VirB11"
10786	10899	-	"hypothetical protein"
10874	12577	+	"VirD4, ATPase required forT-DNA transfer"
12558	12995	+	"cag pathogenicity island protein (cag12)"
13007	13771	+	"IncQ plasmid conjugative transfer protein TraQ"
13774	14571	+	"FIG00469957: hypothetical protein"
14642	15262	+	"FIG00470273: hypothetical protein"
15525	17555	+	"DNA topoisomerase III"
17661	19082	+	"FIG00470802: hypothetical protein"
19101	19280	+	"DNA topoisomerase III"
19639	21558	+	"Tetracycline resistance protein TetO"
21981	22625	+	"TolA protein"
22584	22997	-	"FIG00471024: hypothetical protein"
23460	24683	+	"replication protein"
25035	25268	+	"FIG00471069: hypothetical protein"
25228	25581	+	"FIG00471065: hypothetical protein"
25578	26090	+	"FIG00469626: hypothetical protein"
26087	26644	+	"FIG00471537: hypothetical protein"
26649	26915	+	"FIG00469557: hypothetical protein"
26974	27534	+	"FIG00469385: hypothetical protein"
27537	27803	+	"FIG00469707: hypothetical protein"
27850	33648	+	"helicase, Snf2 family"
33785	34513	+	"FIG00472625: hypothetical protein"
34523	35227	-	"FIG00469644: hypothetical protein"
35250	36638	-	"VirD2 homolog)"
36638	37189	-	"Ribbon-helix-helix protein, copG family domainprotein"
37383	37664	+	"FIG00470991: hypothetical protein"
37692	38303	+	"FIG00471711: hypothetical protein"
38324	39550	+	"DNA primase"
39668	39961	+	"IncQ plasmid conjugative transfer protein TraG"
40040	40312	-	"FIG00471323: hypothetical protein"
40346	40483	-	"hypothetical protein"
40925	41143	+	"FIG00470952: hypothetical protein"
41136	41339	+	"FIG00471111: hypothetical protein"
41343	43136	+	"FIG00470960: hypothetical protein"
43266	43880	-	"Site-specific recombinase, resolvase family"
43877	44284	-	"Virulence-associated protein 2"
44275	44598	-	"FIG00470038: hypothetical protein"

p11818			
1	264	+	"VirB2"
249	377	-	"hypothetical protein"
561	680	-	"hypothetical protein"
700	3051	+	"VirB4"
3062	3619	+	"Phage Rha protein"
3616	4167	+	"FIG00469571: hypothetical protein"
4201	4623	+	"Single-stranded DNA-binding protein"
4639	4914	+	"FIG00470457: hypothetical protein"
5173	5907	+	"VirB5"
6326	6622	-	"hypothetical protein"
6772	7026	+	"hypothetical protein"
7016	7180	+	"VirB7"
7170	7829	+	"VirB8"
7826	8716	+	"VirB9"
8726	9904	+	"VirB10"
9885	10892	+	"VirB11"
10893	12704	+	"VirD4, ATPase required forT-DNA transfer"
12685	13122	+	"cag pathogenicity island protein (cag12)"
13134	13898	+	"IncQ plasmid conjugative transfer protein TraQ"
13901	14698	+	"FIG00469957: hypothetical protein"
14769	15389	+	"FIG00470273: hypothetical protein"
15652	17844	+	"DNA topoisomerase III"
17950	19371	+	"FIG00470802: hypothetical protein"
19390	19569	+	"DNA topoisomerase III"
19928	21847	+	"Tetracycline resistance protein TetO"
22270	22938	+	"TolA protein"
22897	23310	-	"FIG00471024: hypothetical protein"
23773	24996	+	"replication protein"
25348	25581	+	"FIG00471069: hypothetical protein"
25541	25894	+	"FIG00471065: hypothetical protein"
25891	26403	+	"FIG00469626: hypothetical protein"
26400	26957	+	"FIG00471537: hypothetical protein"
26962	27228	+	"FIG00469557: hypothetical protein"
27287	27847	+	"FIG00469385: hypothetical protein"
27850	28116	+	"FIG00469707: hypothetical protein"
28163	33961	+	"helicase, Snf2 family"
34098	34826	+	"FIG00472625: hypothetical protein"
34836	35540	-	"FIG00469644: hypothetical protein"
35563	36951	-	"VirD2 homolog"
36951	37502	-	"Ribbon-helix-helix protein, copG family domainprotein"
37696	37977	+	"FIG00470991: hypothetical protein"
38005	38616	+	"FIG00471711: hypothetical protein"
38637	39863	+	"DNA primase"
39981	40274	+	"IncQ plasmid conjugative transfer protein TraG"
40353	40625	-	"FIG00471323: hypothetical protein"
40659	40838	-	"hypothetical protein"
41111	41329	+	"FIG00470952: hypothetical protein"
41322	41525	+	"FIG00471111: hypothetical protein"
41529	43322	+	"FIG00470960: hypothetical protein"
43452	44066	-	"Site-specific recombinase, resolvase family"
44063	44470	-	"Virulence-associated protein 2"
44461	44784	-	"FIG00470038: hypothetical protein"

S4 - Putative proteins of the pTet-like plasmid pan-genome

This section will provide a summary of the conserved domains identified in the pTet-like plasmid pan-genome. The protein names reported in this section refer to the pTet-like plasmid pan-genome (Section 3.3.1). For each hypothetical protein, the detail of each conserved domain identified is described: this includes the results retrieved from NCBI's CD-Search search engine for each hypothetical protein, together with the name, the accession number and the e-value of each conserved domain and a brief description of the putative function of each conserved domain identified.

S4.1 2_virulence_associated_protein_D_(vapD)



Virulence-associated protein VapD (function unknown) CRISPR associated protein Cas2 Ligand-binding domain - PBP1_NPR_GC_like		
COG3309	1-285	7.63e-42
pfam09827	1-252	8.48e-07
cd06352	79-273	1.32e-03

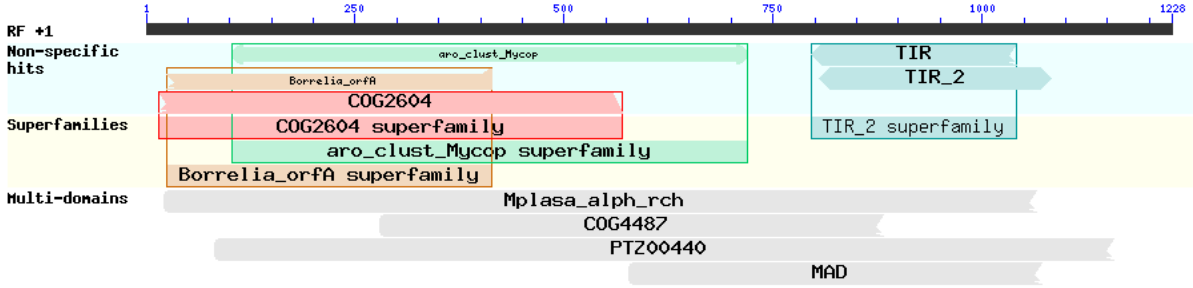
This protein shows a VapD (virulence associated protein) domain. The function of this domain is unknown, but it belongs to the "Cas2_I_II_III Superfamily", a CRISPR/Cas system associated domain (cl11442). A CRISPR_Cas2 domain is also identified in the same region of the VapD domain.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and associated Cas proteins comprise a system for heritable host defense by prokaryotic

cells against phage and other foreign DNA, this domain is present in most CRISP/Cas systems (Barrangou *et al.,* 2007).

Further, a non-specific hit against a PBP1_NPR_GC_like domain is reported in this hypothetical protein (cd06352). This domain is a ligand-binding domain of membrane guanylyl-cyclase receptors, and it is found in several mammalian tissues (Garbers *et al.,* 2006).

S4.2 3_hypothetical_protein

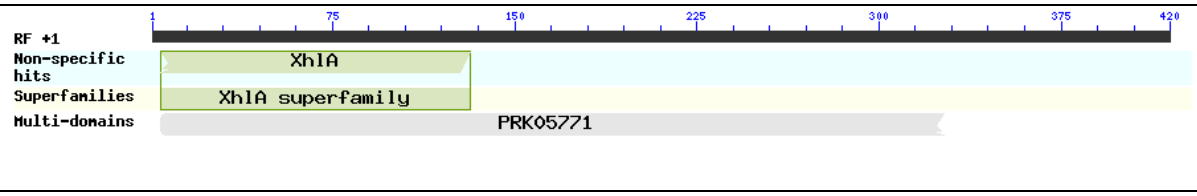


TIR domain; family of bacterial Toll-like receptors aro_clust_Mycop; aromatic cluster surface protein Borrelia_orfA COG2604; Uncharacterized conserved protein [Function unknown]		
pfam13676	805-1083	1.99e-06
TIGR04313	103-720	1.95e-05
pfam02414	25-414	2.72e-04
COG2604	16-570	5.83e-03

The conserved domains shown in this protein suggest a surface expression, possibly involved in signal transduction. A domain of superfamily TIR_2 is described as a bacterial toll-like receptor (Wu *et al.,* 2012) involved with signal transduction. The aro_clust_Mycop, is a domain restricted to *Mollicutes*, far from gram negative bacteria such as *C. jejuni*, and is associated with a lipoprotein suggesting membrane expression. COG2604 is a conserved domain in bacterial proteins with unknown function.

Borrelia ORF-A is a conserved domain often found in the putative proteins from the pTet-like pan-genome and it is derived from plasmid-borne DNA repeats in *Borrelia* species (Zückert *et al.,* 1996). Its putative function is unknown.

S4.3 4_hypothetical_protein

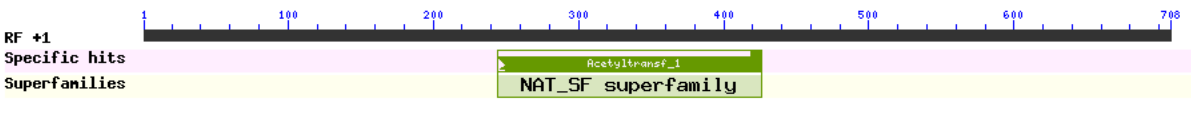


Haemolysin XhlA PRK05771 - V-type ATP synthase subunit I		
pfam10779 PRK05771	4-132 3-327	8.67e-03 8.00e-04

A partial domain related to XhlA is identified in this protein. This is described as a cell-surface haemolysin in an insect pathogen, also able to lyse mammalian horse cell (Cowles *et al.,* 2005).

A multi domain PRK05771 is also identified in this putative protein, associated with the subunit 1 of a V-type ATP synthase.

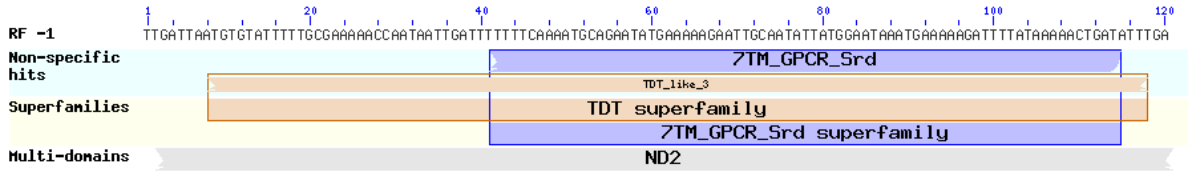
S4.4 6_FIG00469644_hypothetical_protein



Acetyltransf_1		
pfam00583	244-426	2.21e-03

In this hypothetical protein a partial acetyl transferase domain is identified. It belongs to a NAT_SF superfamily (cl17182), a large superfamily of enzymes that catalyses the transfer of an acyl group with several associated functions, from histone acylation to antibiotic resistance in bacteria (Neuwald *et al.,* 1997, Cort *et al.,* 2008).

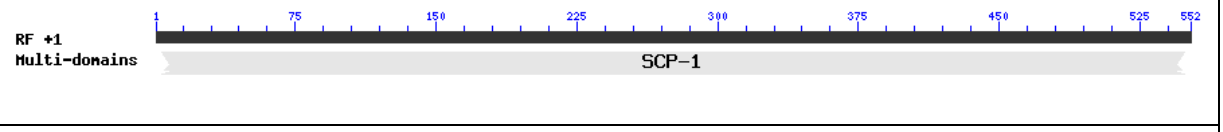
S4.5 7_FIG00471069_hypothetical_protein



Serpentine type 7TM GPCR chemoreceptor Srd Tellurite-resistance/Dicarboxylate Transporter (TDT) family		
pfam10317	40-114	4.39e-05
cd09321	7-117	9.73e-04

In this small hypothetical protein two partial non-specific hits are identified: the first is far from the bacterial kingdom, associated with a 7 transmembrane domain chemoreceptor in *Caenorhabditis elegans* (Robertson *et al.,* 2006); the second is associated with a partial Tellurite-resistance/Dicarboxylate Transporter (TDT). This protein family is conserved in all kingdoms of life. In bacteria a member of this family has been identified as an uptake channel for C4 carbohydrates (Teramoto *et al.,* 2008). This may be particularly relevant in *Campylobacter*, as most *Campylobacter* strains cannot utilise sugars and they rely on the use of amino acids and C4 carbohydrates for survival and successful colonisation (Guccione *et al.,* 2008; Stahl *et al.*, 2011).

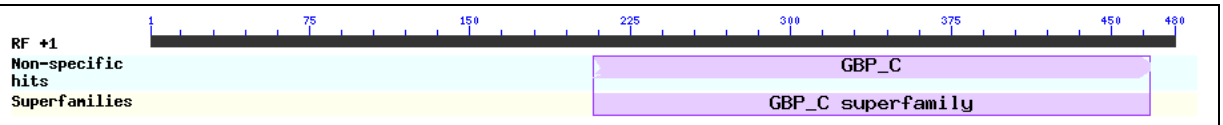
S4.6 9_FIG00469571_hypothetical_protein



Synaptonemal complex protein 1 (SCP-1)		
pfam05483	4-549	1.77e-03

In this hypothetical protein, the only hit is related to a superfamily that is involved with a chromosomal structure that is formed during the meiotic phase in eukaryotic cells (Meuwissen *et al.,* 1992).

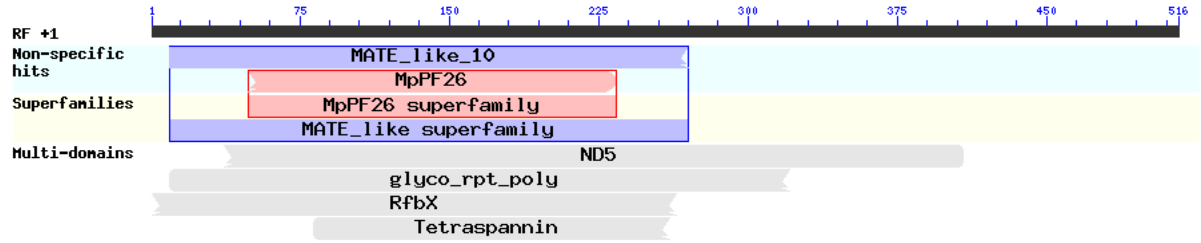
S4.7 10_hypothetical_protein



Guanylate-binding protein, C-terminal domain		
cd16269	208-468	9.22e-03

The conserved domain GBP_C is the only domain found in this hypothetical protein. It is the C-terminal domain of a guanylate-binding protein, a regulative protein synthesised in eukaryotic cells after activation with interferons (Vestal *et al.,* 2011).

S4.8 11_FIG00469626_hypothetical_protein



Subfamily multidrug and toxic compound extrusion (MATE) proteins; MATE_like_10

M. penetrans paralogue family 26; MpPF26

NADH dehydrogenase subunit 5 - ND5

oligosaccharide repeat unit polymerase - glyco_rpt_poly

Membrane protein - RfbX

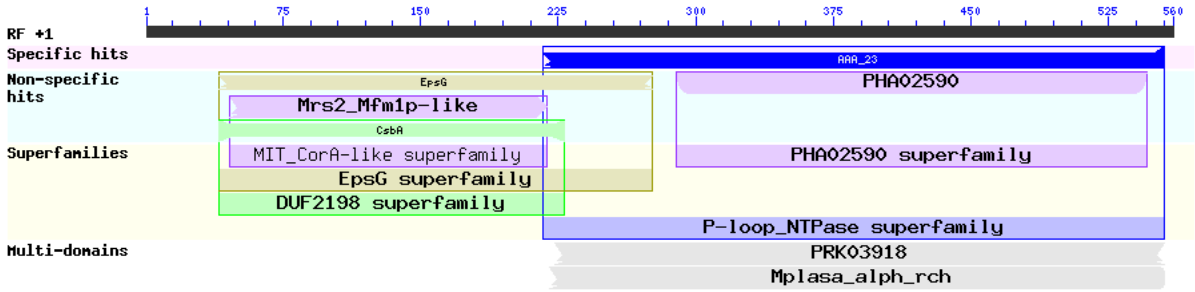
Tetraspanin family

cd13125	10-270	8.86e-05
pfam07666	49-234	1.42e-03
MTH00095	37-408	5.71e-04
TIGR04370	10-321	1.18e-03
COG2244	1-264	2.83e-03
pfam00335	82-261	3.90e-03

This hypothetical protein shows several partial domains the majority of which seem related to a membrane transporter or a membrane protein. One of the conserved domains identified is a MATE-like domain (Putman *et al.,* 2000; Hvorup *et al.,* 2003). It encodes for a family of proteins associated with the function of multidrug and toxic compound extrusion (MATE), which is involved in exporting metabolites across the cell membrane and conferring multi drug resistance to bacteria.

Further, it is possible to identify the domain MpPF26, a paralogue of a protein family in *Mycoplasma penetrans* (Sasaki *et al.,* 2002). The domains RfbX and glyco_RPT_poly are involved with the synthesis and export of cell walls component. The domain ND5 associated with a NADH dehydrogenase subunit and an undescribed Tetraspannin family domain associated with eukaryotic membrane proteins.

S4.9 12_FIG00471537_hypothetical_protein



EpsG family

S. cerevisiae inner mitochondrial membrane Mg2+ transporters Mfm1p and Mrs2p-like

AAA domain - AAA_23

PHA02590

General stress protein CsbA

PRK03918

helix-rich Mycoplasma protein - Mplasa_alph_rch

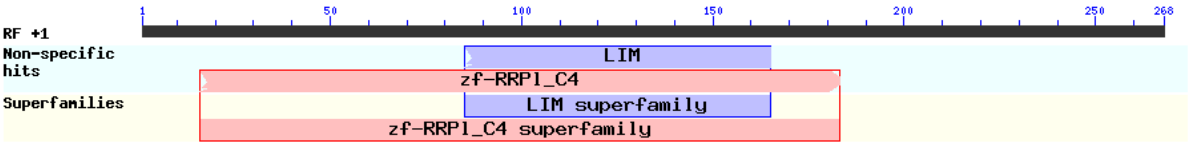
pfam14897	40-276	9.78e-05
cd12823	46-219	1.66e-04
pfam13476	217-555	7.78e-04
PHA02590	289-546	1.71e-03
COG4897	40-228	4.51e-03
PRK03918	223-555	3.33e-04

In the hypothetical protein 12_FIG00471537_hypothetical_protein, one specific hit is identified as AAA protein (ATPase associated with diverse cellular activities; Iyer *et al.,* 2004). Other non-specific hits include an EpsG domain, which appears often in the plasmidic proteins of *C. jejuni*: it is related to the EpsG protein of *Bacillus subtilis,* which appears to be involved in the production of exopolysaccharides

required for biofilm maintenance (Branda *et al.,* 2004). Moreover, a general stress protein CsbA domain is identified, related to a protein of unknown function in *Bacillus subtilis* involved in stress response (Petersohn *et al.,* 2001).

Finally, this hypothetical protein includes domains associated with viral proteins (PHA02590), mitochondrial cation transport (cd12823), chromosome segregation (PRK03918), and a membrane protein of unknown function occurring in members of *Mycoplasma* species.

S4.10 13_FIG00469557_hypothetical_protein

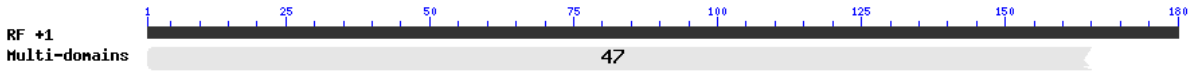


Small protein-protein interaction domain - LIM Putative ribonucleoprotein zinc-finger pf C4 type - zf-RRPl_C4		
cd08368	85-165	2.97e-03
pfam17026	16-183	6.73e-03

The LIM domain found in this protein is identified in several proteins with various functions in eukaryotic cells, such as DNA expression regulation and cell development (Matthews *et al.,* 2003). The domain contains two zinc finger motifs.

Furthermore, a domain for a putative ribonucleoprotein is identified. This family (zinc finger of C4 type) is largely represented in Microsporidia and carries residues (in a pattern of CxxC) that suggest DNA binding (Nakjang *et al.,* 2013).

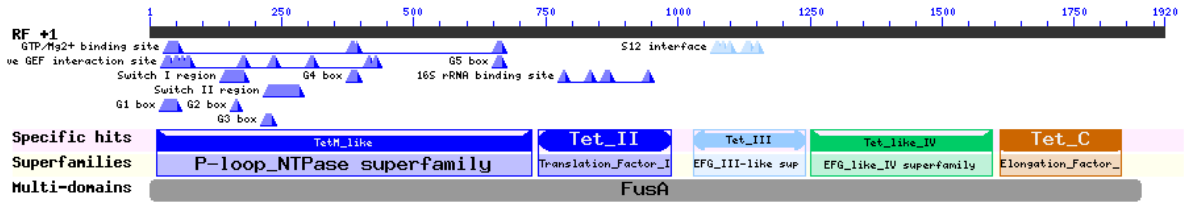
S4.11 15_DNA_topoisomerase_II



endonuclease subunit - 47		
PHA02546	1-165	1.77e-03

This putative protein, annotated as a DNA topoisomerase II carries a viral endonuclease domain (Mickelson and Wiberg, 1981).

S4.12 16_Tetracycline_resistance_protein_TetO



Tet(M)-like family

EF-G_domain IV_RPP - Tet_like_IV

Domain III of Tetracycline resistance protein Tet - Tet_III

Domain II of ribosomal protection proteins Tet(M) and Tet(O) - Tet_II

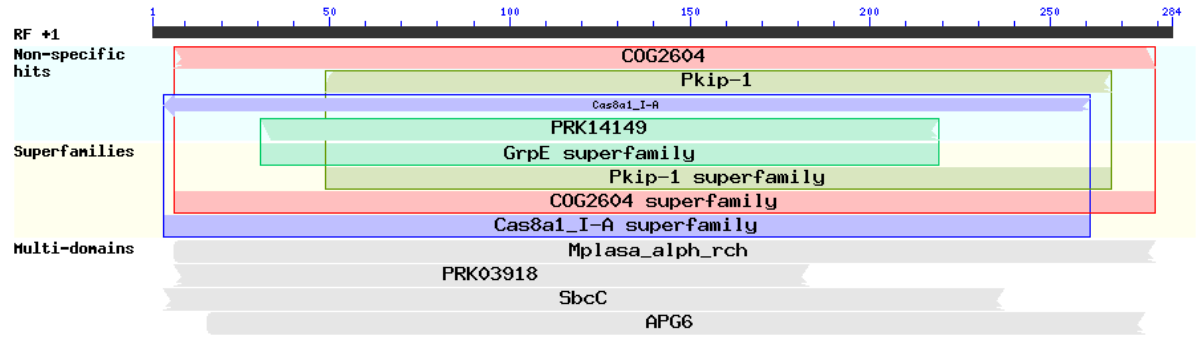
Tet_C: C-terminus of ribosomal protection proteins Tet(M) and Tet(O)

Translation elongation factor EF-G - FusA

cd04168	13-723	3.32e-149
cd01684	1252-1596	5.89e-65
cd16258	1030-1242	1.34e-35
cd03690	736-987	1.12e-34
cd03711	1609-1839	2.22e-32
COG0480	1-1875	3.25e-137

This putative protein shows all the typical domains of a tetracycline resistance protein Tet(O) or Tet(M) (Roberts, 2005; Oggioni *et al.,* 1996). This protein acts protecting the ribosome and it is a homologue to the elongation factors EF-Tu and EF-G.

S4.13 18_FIG00470991_hypothetical_protein



Uncharacterized conserved protein - COG2604

Pkip-1 protein

CRISPR/Cas system-associated protein - Cas8a1

heat shock protein GrpE - PRK14149

chromosome segregation protein - PRK03918

DNA repair exonuclease SbcCD

Autophagy protein Apg6

COG2604	7-279	1.78e-04
pfam06878	49-267	5.78e-04
cd09730	4-261	7.02e-04
PRK14149	31-219	8.81e-04
PRK03918	7-183	1.39e-04
COG0419	4-237	1.50e-04
pfam04111	16-276	5.76e-04

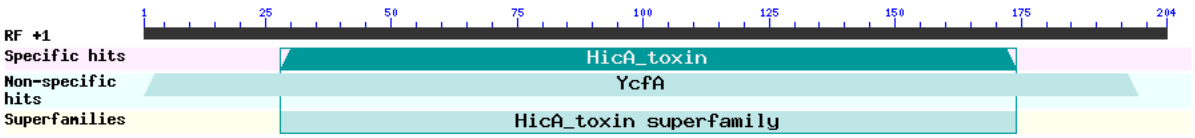
Several conserved domains appear in this protein.

Pkip-1 is related to viral proteins of unknown function. PRK14149 is a part of the GrpE superfamily, including heat-shock proteins and an essential factor for bacterial chaperone proteins. Other non-specific domains are mainly involved in

eukaryotic cellular processes (PRK03918 and Apg6) (Pijlman *et al.,* 2003; Castanié-Cornet *et al.,* 2014).

Cas8a1 domain is associated with CRISPR/Cas proteins (Boyaval *et al.,* 2007).

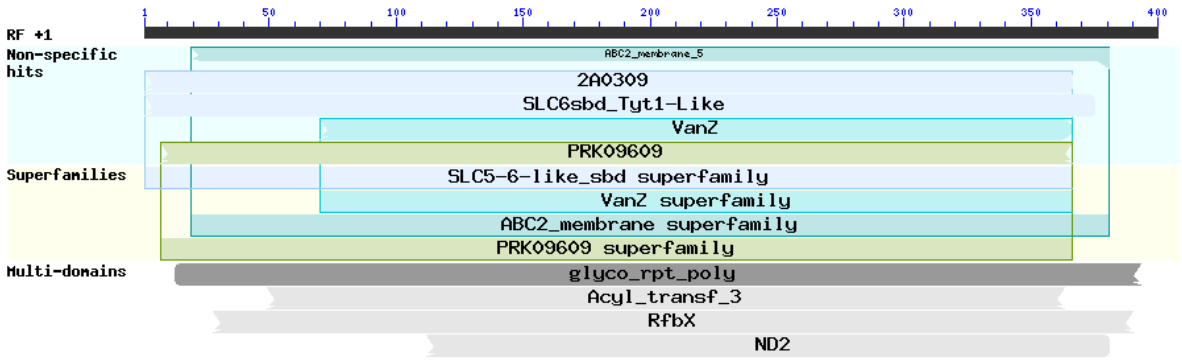
S4.14 19_FIG00471111_hypothetical_protein



Predicted RNA binding protein YcfA HicA toxin of bacterial toxin-antitoxin		
COG1724 pfam07927	1-198 28-174	8.40e-19 1.00e-12

This protein shows the conserved domain for YfcA/HicA toxin. The HicA toxin acts as an mRNA interferase and is neutralised by the cognate antitoxin HicB, usually found in tandem in the sequenced plasmids and described later on in this chapter (23_FIG00470952_hypothetical_protein). The HicAB system is often used in the bacterial cell for plasmid stabilisation (Makarova *et al.,* 2006; Yamaguchi and Inouye, 2011).

S4.15 20_FIG00471065_hypothetical_protein



ABC-2 family transporter protein - ABC2_membrane_5

spore germination protein (amino acid permease) - 2A0309

solute carrier 6 subfamily - SLC6sbd_Tyt1-Like

glycopeptide antibiotics resistance protein - VanZ

hypothetical protein - PRK09609

oligosaccharide repeat unit polymerase - glyco_rpt_poly

acyltransferase family - Acyl_transf_3

membrane protein, O-antigen and teichoic acid export - RfbX

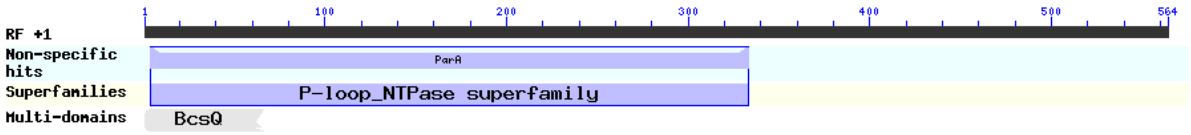
NADH dehydrogenase subunit 2 - ND2

pfam13346	19-381	2.75e-09
TIGR00912	1-366	4.17e-06
cd10336	1-375	1.50e-05
COG4767	70-366	1.60e-04
PRK09609	7-366	3.38e-04
TIGR04370	13-393	1.21e-04
pfam01757	49-363	1.17e-03
COG2244	28-390	1.69e-03
MTH00091	112-381	2.77e-03

This hypothetical protein shows several domains linked with membrane protein and transport. In particular, the domain with a lower p-value associated with this sequence is an ABC transporter domain, a bacterial system associated with export of drugs and carbohydrates (Reizer *et al.,* 1992). Another domain hit associated with molecules transport is the solute carrier from *Fusobacterium nucleatum* Tyt1-like. Bacterial variants of this protein family are involved in transport of amino acids (Quick *et al.,* 2006).

Furthermore, a VanZ domain is identified and this is involved in the resistance to glycopeptide antibiotics such as vancomicin and teicoplanin (Arthur *et al.,* 1995).

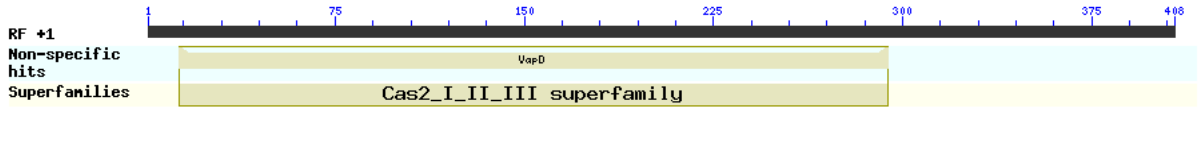
S4.16 21_hypothetical_protein



ParA Cellulose biosynthesis protein BcsQ [Cell motility]		
cd02042 COG1192	4-333 1-66	3.01e-04 3.29e-03

The conserved domain ParA is recognisable in this hypothetical protein. It is part of a ParA-ParB system involved in bacterial chromosome segregation. ParA, in particular, appears to have an ATPase activity that promotes the regulation of the DNA replication process. It shares similarity with a conserved family of ATPases involved in the same processes that includes RepA (Easter and Gober, 2002).

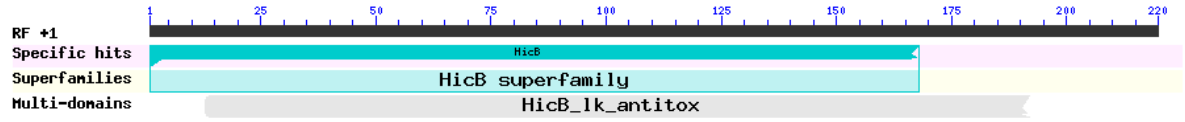
S4.17 22_Virulence-associated_protein_2_(vapD)



Virulence-associated protein VapD CRISPR/Cas system-associated protein Cas2 superfamily		
COG3309 cl11442	13-294 13-294	7.41e-12 7.41e-12

The conserved domain found here is consistent with the automated annotation. VapD, is a virulence-associated protein of unknown function, part of a superfamily of CRISPR/Cas system-associated protein involved in the protection of the bacterium against phages and other foreign DNA (Boyaval *et al.,* 2007; Nam *et al.,* 2012).

S4.18 23_FIG00470952_hypothetical_protein

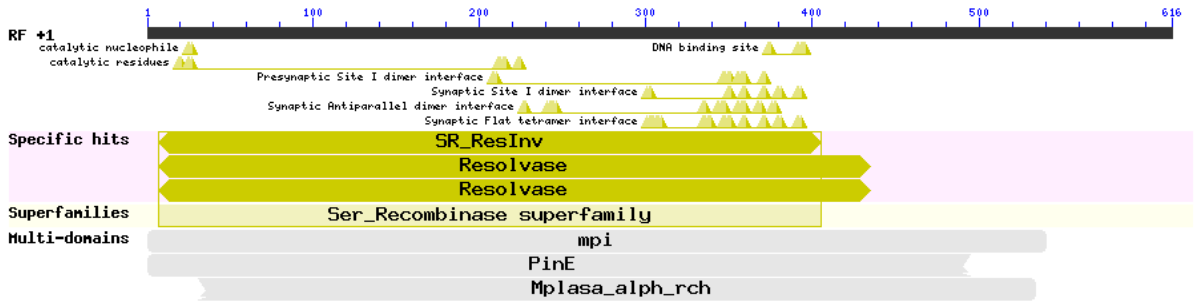


Predicted nuclease of the RNAse H fold, HicB family HicB_like antitoxin of bacterial toxin-antitoxin system - HicB_lk_antitox		
COG1598 pfam15919	1-168 13-192	2.34e-16 3.78e-09

This putative protein harbours an HicB domain, antitoxin of the HicAB toxin-antitoxin system. The conserved domain associated with the cognate toxin is

encoded by the “19_FIG00471111_hypothetical_protein” (Yamaguchi and Inouye, 2011; Makarova *et al.,* 2006).

S4.19 24_Site-specific_recombinase_resolvase



Serine Recombinase (SR) family - SR_ResInv

Resolvase, N terminal domain

Resolvase, N terminal domain

multiple promoter invertase - Mpi

Site-specific DNA recombinase - PinE

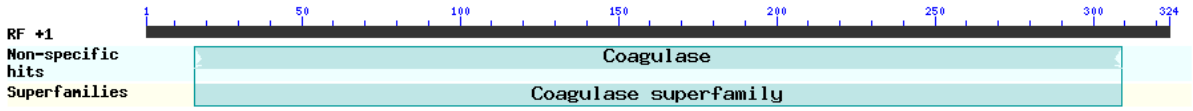
helix-rich Mycoplasma protein - Mplasa_alph_rch

cd03768	7-405	7.34e-41
pfam00239	7-435	6.13e-38
smart00857	7-435	4.63e-37
PRK13413	1-540	2.65e-47
COG1961	1-495	2.87e-28
TIGR04523	31-534	1.47e-04

As predicted by the automated annotation, this hypothetical protein contains conserved domain associated with a resolvase/inverase of the serine recombinase family. This family of proteins catalyses the site-specific recombination of DNA

molecules (Hallet and Sherratt, 1997; Smith and Thorpe, 2002). Several catalytic features associated with this protein family are also identified in the sequence.

S4.20 26_FIG00470038_hypothetical_protein



Staphylococcus aureus coagulase		
pfam08764	16-309	9.78e-03

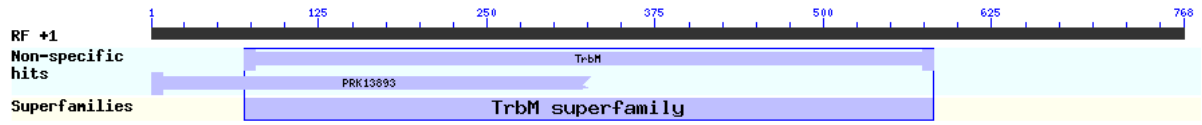
The coagulase conserved domain shown here derives from a coagulase protein described in *Staphylococcus aureus*, which initiates blood clotting (Friedrich *et al.,* 2003).

S4.21 27_FIG00470273_hypothetical_protein

helix-rich Mycoplasma protein		
TIGR04523	328-606	1.51e-04

A partial domain of unknown function is present in this hypothetical protein. It is associated with *Mycoplasma* species, strong amino acid periodicity and it is almost completely formed by alpha helices (Sasaki *et al.,* 2002).

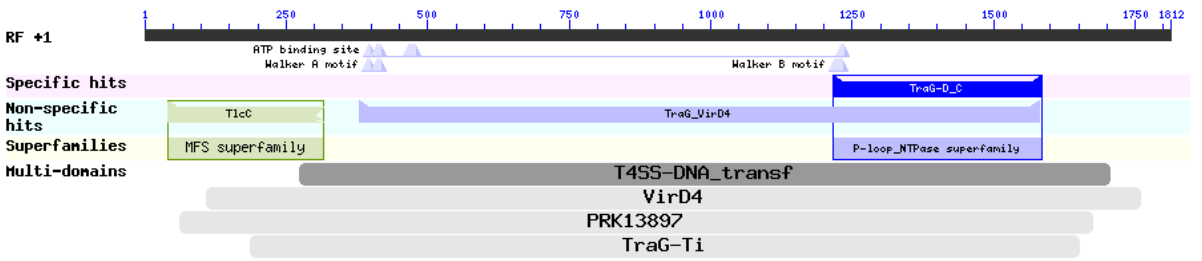
S4.22 28_IncQ_plasmid_conjugative_transfer_protein_TraQ



TrbM conjugal transfer protein TrbM - PRK13893		
pfam07424 PRK13893	70-582 1-327	1.60e-35 4.67e-27

The protein family described here is consistent with the automated annotation and represents a TrbM protein (homologous with TraQ), involved with DNA conjugal transfer (Lessl *et al.,* 1992).

S4.23 29_Coupling_protein_VirD4_ATPase

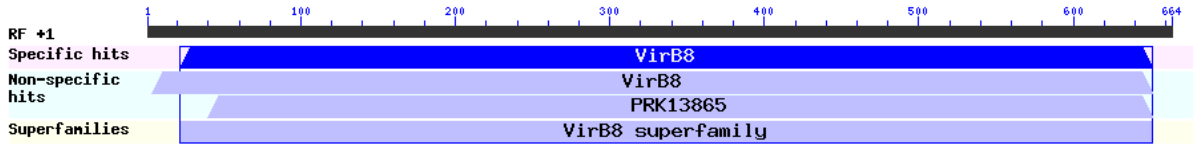


TraG_VirD4 bacterial conjugation protein TraM recognition site of TraD and TraG - TraG-D_C ATP/ADP translocase - TlcC Type IV secretory system Conjugative DNA transfer - T4SS-DNA_transf Type IV secretory pathway, VirD4 type IV secretion system component VirD4 - PRK13897 Ti-type conjugative transfer system protein TraG		
cd01126	379-1581	5.96e-82

pfam12696	1216-1584	5.53e-28
COG3202	40-315	9.86e-03
pfam02534	274-1704	4.98e-104
COG3505	109-1758	6.78e-107
PRK13897	61-1674	2.63e-76
TIGR02767	187-1650	7.79e-15

This hypothetical protein, annotated as VirD4 ATPase, shows conserved domain consistent with the function of ATPase providing the energy to the Type IV Secretion System (Hofreuter *et al.,* 2001; Kado, 2006).

S4.24 30_VirB8



VirB8 protein

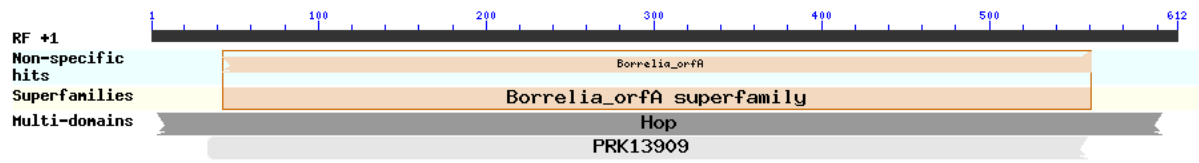
Type IV secretory pathway, component VirB8

type IV secretion system protein VirB8 - PRK13865

pfam04335	22-651	5.25e-60
COG3736	4-651	1.89e-57
PRK13865	40-651	8.52e-07

This hypothetical protein, annotated as VirB8, shows a complete conserved domain associated with that Type IV Secretion System protein (Kumar and Das, 2001; Backert *et al.,* 2000).

S4.25 31_FIG00471711_hypothetical_protein_NTPase

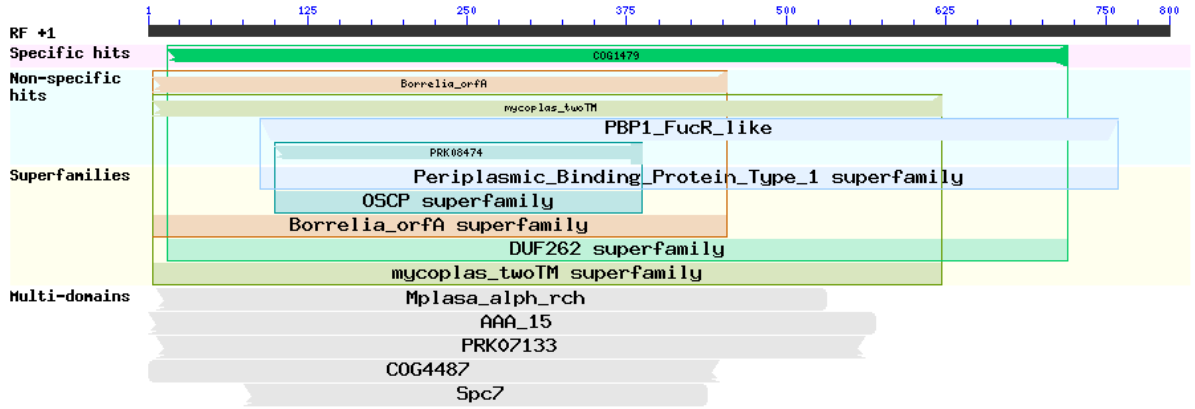


Borrelia_orfA Intein/homing endonuclease - Hop putative recombination protein RecB - PRK13909		
pfam02414 COG1372 PRK13909	43-561 4-603 34-558	1.77e-03 1.25e-03 1.29e-03

This hypothetical protein shows a conserved domain associated with an ORF of unknown function identified in plasmids of *Borrelia bugdoferi* (Zückert and Meyer, 1996).

Two further domains are involved with DNA mobilisation and recombination (Marsin *et al.,* 2008).

S4.26 32_FIG00469957_hypothetical_protein



Uncharacterized conserved protein - COG1479

Borrelia_orfA

two transmembrane protein - mycoplas_twoTM

Ligand-binding domain of a transcription repressor - PBP1_FucR_like

F0F1 ATP synthase subunit delta - PRK08474

helix-rich Mycoplasma protein - Mplasa_alph_rch

AAA ATPase domain

DNA polymerase III subunits gamma and tau - PRK07133

Uncharacterized protein - COG4487

Spc7 kinetochore protein

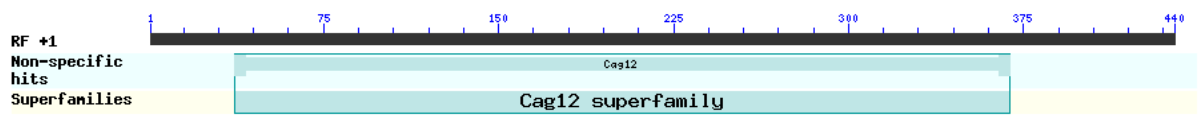
COG1479	16-720	1.39e-05
pfam02414	4-453	2.90e-05
TIGR04527	4-621	5.46e-05
cd06276	88-759	8.66e-05
PRK08474	100-387	5.53e-04
TIGR04523	7-531	1.64e-08
pfam13175	1-570	8.69e-06

PRK07133	7-561	3.74e-05
COG4487	1-447	3.11e-03
smart00787	76-438	6.49e-03

Several domains of different functions are identified in this protein. A partial uncharacterised protein domain (COG1479) is present which contains a ParB-like nuclease domain, making it a candidate to complete the ParA-ParB nuclease system described with the 21_hypothetical_protein (Easter and Gober, 2002). Other Domains are consistent with the nuclease activity, such as the AAA-domain pfam13175.

The domain PBP1_FucR_like (cd06276) is also found in this hypothetical protein: it is described as a ligand binding domain sensor for L-fucose which function as a transcription repressor. The protein FucR, from which this domain family is derived, is composed of a helix-turn-helix DNA-binding domain at the N-termius, and a C-terminal sugar binding domain, homologous to other sugar binding domain found in other membrane transport systems, such as ABC transporters. (Felder *et al.,* 1999; Nguyen and Saier. 1995).

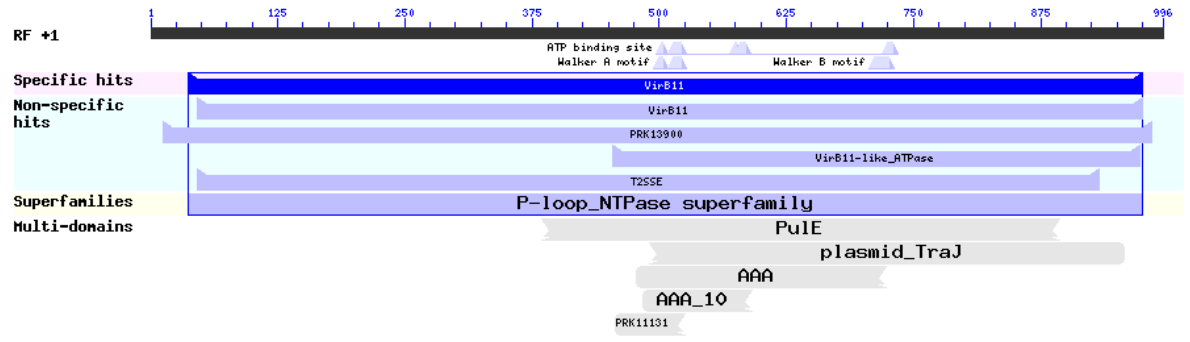
S4.27 33_cag_pathogenicity_island_protein



Cag pathogenicity island protein Cag12		
pfam13117	37-369	1.05e-22

This putative protein shows a Cag12 conserved domain, which is consistent with the predicted function of the automated annotation. The protein Cag12 is part of the Cag pathogenicity island of *Helicobacter pylori*. The Cag-PAI encodes for the Type IV Secretion System of *H. pylori*, used by the bacterium to deliver cytotoxins to the host cells. Although the function of Cag12 is not clear, it is a component located on the outer membrane of the bacterium and it has been proved that it could be used to develop a protective immunity against the pathogen (Kim *et al.,* 2006).

S4.28 34_VirB11



P-type DNA transfer ATPase VirB11

Type IV secretory pathway ATPase VirB11

Type IV secretion system ATPase VirB11 - PRK13900

Type IV secretory pathway component VirB11

Type II/IV secretion system protein - T2SSE

Type II secretory pathway ATPase GspE/PulE or T4P pilus assembly pathway ATPase PilB - PulE

plasmid transfer ATPase TraJ

ATPases associated with a variety of cellular activities - AAA

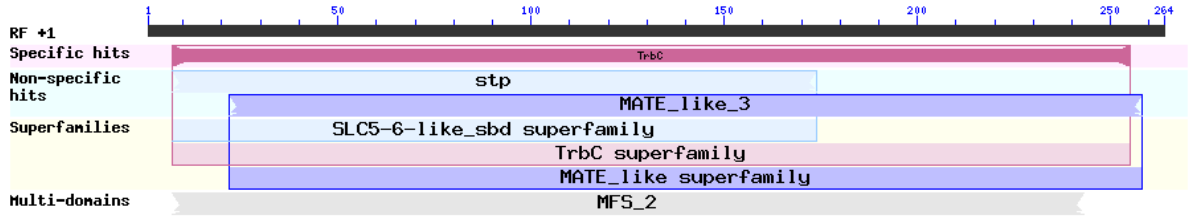
AAA-like domain - AAA_10

ATP-dependent RNA helicase HrpA - PRK11131

TIGR02788	46-975	3.32e-132
COG0630	37-975	4.49e-74
PRK13900	13-984	3.35e-67
cd01130	454-972	5.63e-67
pfam00437	46-933	5.90e-58
COG2804	385-894	2.11e-12
TIGR02525	490-957	4.43e-11
smart00382	478-723	7.72e-06
pfam12846	484-591	2.79e-04
PRK11131	457-525	1.92e-03

As predicted by the automated annotation, this putative gene encodes for conserved domains strictly related to the VirB11 ATPase of the Type Four Secretion System (Atmakuri *et al.,* 2004).

S4.29 35_VirB2



TrbC/VIRB2 family

serine transporter - stp

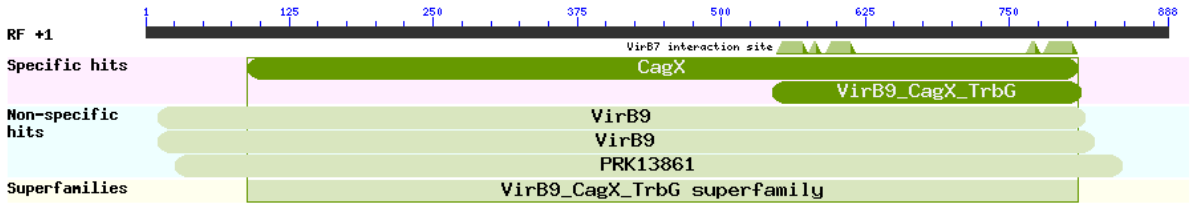
Subfamily of the multidrug and toxic compound extrusion (MATE) proteins - MATE_like_3

MFS/sugar transport protein - MSF2

pfam04956	7-255	1.64e-19
TIGR00814	7-174	2.21e-03
cd13148	22-258	7.04e-03
pfam13347	7-243	2.16e-05

Consistent with the automated annotation, the most reliable hit according to the p-Value is the TrbC/VirB2 family domain. It is the main structural component of the Type IV secretion system (Kalkum *et al.,* 2002; Schulein and Dehio, 2002).

S4.30 36_VirB9



P-type conjugative transfer protein VirB9

Conjugal transfer protein - CagX

Type IV secretory pathway, VirB9 component

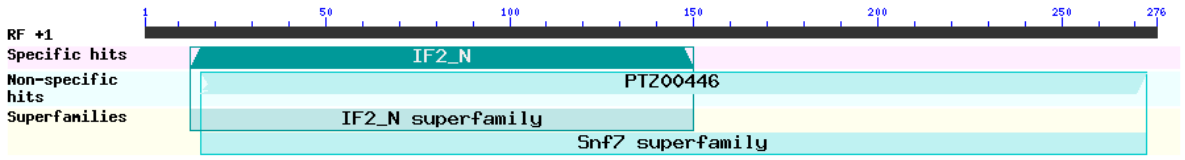
VirB9/CagX/TrbG, a component of the type IV secretion system - VirB9

type IV secretion system protein VirB9 - PRK13861

TIGR02781	10-816	4.75e-79
pfam03524	10-825	1.73e-34
cd06911	544-813	2.70e-28
PRK13861	25-849	3.69e-15

The conserved domains identified in this protein are consistent with the component VirB9 of the Type IV secretion system. A VirB7 interaction site is also identified: the two proteins form a complex when the Type IV secretion system is assembled (Anderson *et al.,* 1996; Bayliss *et al.,* 2007).

S4.31 37_FIG00470457_hypothetical_protein

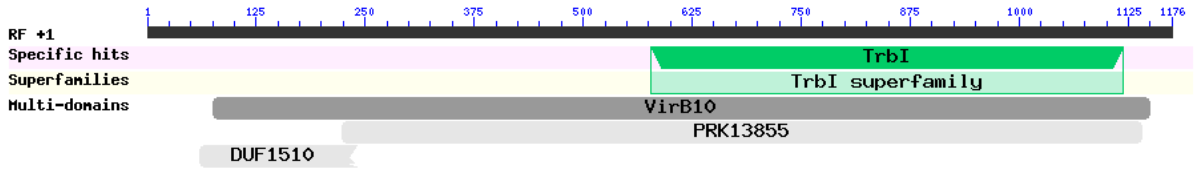


Translation initiation factor IF-2, N-terminal region vacuolar sorting protein SNF7-like - PTZ00446		
pfam04760	13-150	1.86e-05
PTZ00446	16-273	4.96e-03

The N-termini of the IF2 translation initiation factor of *E. coli* is identified in this hypothetical protein. The IF2 factor ensures the binding of the MET-tRNA to the 30S ribosomal subunit (Nyengaard *et al.,* 1991; Laursen *et al.,* 2003).

A non-specific hit is also identified with a superfamily associated with vacuolar sorting and transport of membrane protein in eukaryotic cells (Babst *et al.,* 2002).

S4.32 39_VirB10



Bacterial conjugation TrbI-like protein

Type IV secretory pathway, VirB10

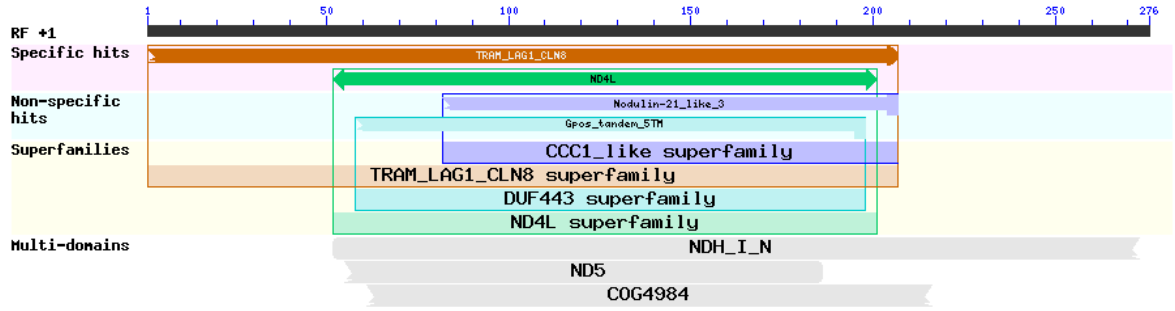
type IV secretion system protein VirB10 - PRK13855

Protein of unknown function - DUF1510

pfam03743	577-1119	5.96e-55
COG2948	76-1149	7.49e-54
PRK13855	223-1140	1.69e-39
pfam07423	61-240	3.05e-03

The conserved domains identified in this protein are associated with the VirB10 component of the Type IV secretion system, with the exception of a small domain of unknown function (pfam07423) (Ward *et al.,* 1990).

S4.33 40_FIG00471323_hypothetical_protein



Nodulin-21 and CCC1-related protein family

TLC domain - TRAM_LAG1_CLN8

NADH dehydrogenase subunit 4L - NDL4

tandem five-transmembrane protein - Gpos_tandem_5TM

proton-translocating NADH-quinone oxidoreductase - NDH_I_N

NADH dehydrogenase subunit 5 - ND5

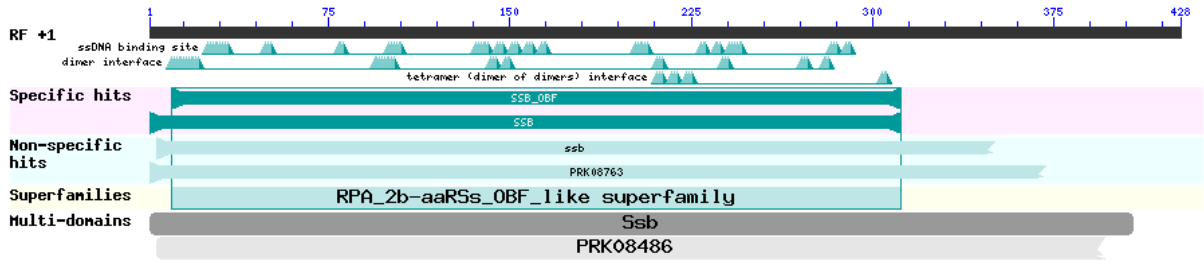
Uncharacterized membrane protein - COG4984

cd02434	82-207	2.39e-04
pfam03798	1-207	5.64e-04
MTH00093	52-201	6.01e-04
TIGR01218	58-198	1.35e-03
TIGR01770	52-273	9.93e-04
MTH00095	55-186	2.04e-03
COG4984	61-216	8.19e-03

This hypothetical protein includes conserved domains associated with several putative functions. A nodulin related protein associated to nitrogen fixation in plants and metals transporter in yeast (Delauney et al, 1990; Li *et al.,* 2001); a domain associated with lipid sensing in yeast and mammalian cells (Winter and Ponting,

2002); a domain of unknown function present in proteins with five transmembrane domains; and domains associated with a NADH dehydrogenase function. A putative function is not possible to predict, but the majority of the conserved domains suggest that this is likely to be a membrane protein.

S4.34 41_Single-stranded_DNA-binding_protein



SSB_OBF: A subfamily of OB folds similar to the OB fold of ssDNA-binding protein (SSB)

Single-strand binding protein family - SSB

single stranded DNA-binding protein - ssb

single-stranded DNA-binding protein - PRK08763

Single-stranded DNA-binding protein - Ssb

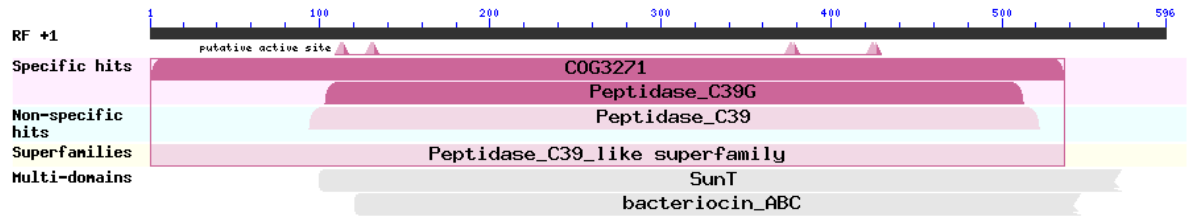
single-stranded DNA-binding protein - PRK08486

cd04496	10-312	4.37e-33
pfam00436	1-312	4.60e-29
TIGR00621	4-351	5.91e-24
PRK08763	1-372	7.54e-17
PRK08486	4-396	1.56e-16

The conserved domains identified in this protein suggest the possible function of single-strand DNA binding protein. This protein family are known to bind ssDNA in order to aid several processes including transcription, replication and recombination.

In *E. coli* this protein is known to form stable homodimer, while in other bacterial species (such as *Mycobacterium tuberculosis*) it forms tetramer. The sequence here described shows features for dimer and tetramer interaction, highlighted in the figure (Lopper *et al.,* 2004; Saikrishnan *et al.,* 2003; Arcus, 2002).

S4.35 42_FIG00638667_hypothetical_protein



Predicted double-glycine peptidase - COG3271

Peptidase_C39G

Peptidase C39 family

ABC-type bacteriocin/lantibiotic exporters - SunT

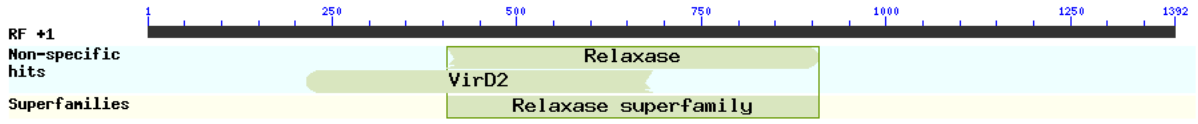
ABC-type bacteriocin transporter - bacteriocin_ABC

COG3271	1-537	2.73e-39
cd02423	103-513	6.64e-38
pfam03412	94-522	2.89e-21
COG2274	100-570	8.50e-12
TIGR01193	121-546	7.79e-06

This hypothetical protein shows several conserved domains associated with the C39 peptidase, proteins necessary to process various bacteriocins. These are antibiotic proteins produced by bacteria to inhibit the growth of other bacterial species. This family of peptidase is usually the N-terminal of an ABC-transporter: the

processing of the bacteriocin involves the cleavage of a Glycine-Glycine bond, followed by the export of the molecule across the cytoplasmic membrane. The putative active site is also identified in the protein sequence (Wu and Tai, 2004; Barrett and Rawlings, 2001).

S4.36 44_IncQ_plasmid_conjugative_transfer_TraR_(VirD2)

<table>
<tr><td colspan="3"> 
</td></tr>
<tr><td colspan="3">Relaxase/Mobilization nuclease domain

Type IV secretory pathway - VirD2</td></tr>
<tr><td>pfam03432

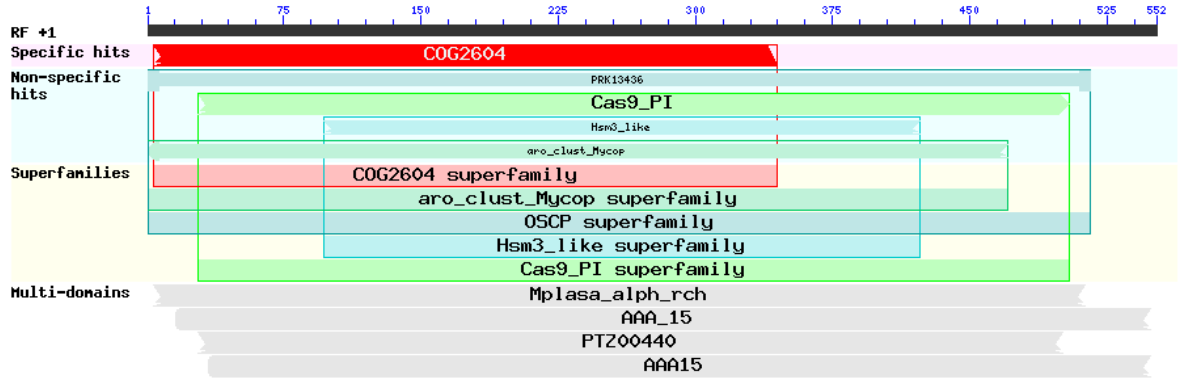
COG3843</td><td>406-909

214-684</td><td>5.24e-08

2.61e-04</td></tr>
</table>

The domains identified in this putative protein are consistent with the VirD2/TraR protein. This protein is involved with the mobilisation of DNA during conjugational transfer (Meyer, 2000).

S4.37 45_Ribbon-helix-helix_protein_copG_family_domain_protein



F0F1 ATP synthase subunit delta - PRK13436

PAM-interacting domain of CRISPR-associated endonuclease Cas9 - Cas9_PI

Uncharacterized conserved protein - COG2604

Proteasome chaperone - Hsm3_like

aromatic cluster surface protein - aro_clust_Mycop

helix-rich Mycoplasma protein - Mplasa_alph_rch

AAA ATPase domain - AAA_15

reticulocyte binding protein 2-like - PTZ00440

ATPase/GTPase - AAA15

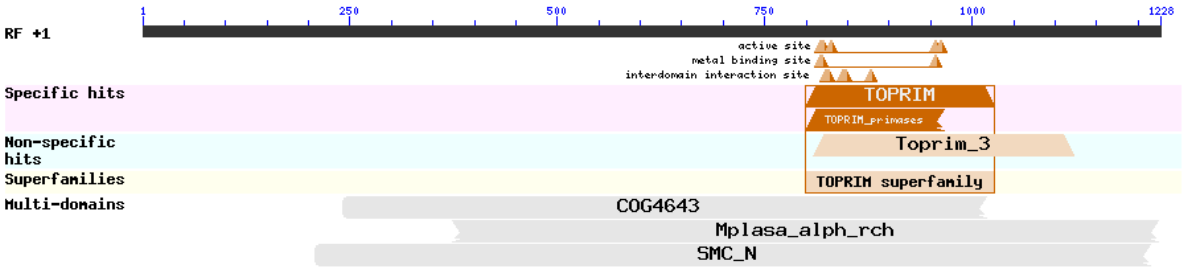
PRK13436	1-516	1.77e-06
pfam16595	28-504	5.81e-06
COG2604	4-345	1.52e-04
cd12794	97-423	3.04e-04
TIGR04313	1-471	5.07e-04
TIGR04523	4-513	9.72e-07
pfam13175	16-549	1.73e-06
PTZ00440	28-501	6.90e-05
COG1106	34-549	1.48e-04

This hypothetical protein presents several domains of different nature. One of these is a complete conserved domain associated with the proteasome assembly in yeast showing an ATPase function (Park *et al.,* 2010).

A CRISPR/Cas9 associated protein domain is also identifiable (Boyaval *et al.,* 2007).

Other domains are of unknown function, an ATP synthetase and several ATPase putative domains.

S4.38 46_DNA_primase_(EC_2.7.7.-)



topoisomerases, DnaG-type primases - TOPRIM

Toprim domain - Toprim_3

Topoisomerase-primase - TOPRIM_primases

Uncharacterized domain associated with phage/plasmid primase COG4643

helix-rich Mycoplasma protein - Mplasa_alph_rch

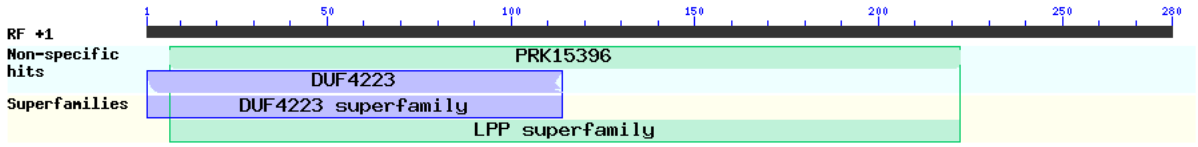
RecF/RecN/SMC N terminal domain - SMC_N

smart00493	799-1026	2.79e-06
pfam13362	808-1122	6.89e-06
cd01029	799-966	7.45e-06
COG4643	241-1017	6.85e-28

TIGR04523 pfam02463	373-1224 208-1215	4.80e-04 6.07e-03

This protein carries a Topoisomerase/Primase domain. Bacterial primases synthesise RNA primer to initiate DNA replication. The reaction is Mg2+ dependent. The active site together with the metal binding site are identified in the conserved domain (Keck *et al.,* 2000; Aravind *et al.,* 1998).

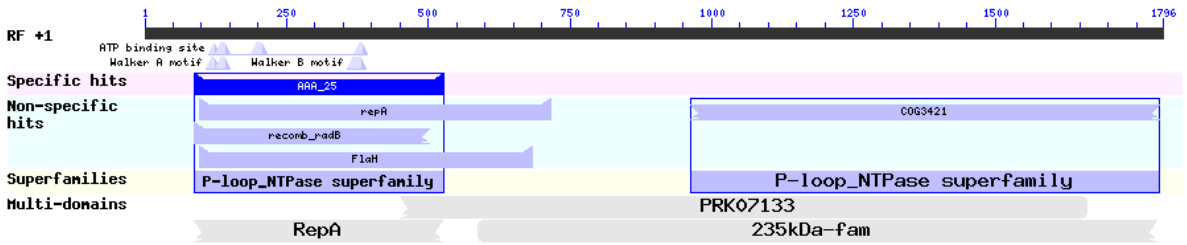
S4.39 47_IncQ_plasmid_conjugative_transfer_protein_TraG



Murein lipoprotein - PRK15396 Protein of unknown function - DUF4223		
PRK15396 pfam13978	7-222 1-114	1.50e-04 4.78e-03

In this hypothetical protein, annotated as a plasmid conjugative transfer protein, it is possible to identify just two conserved domains of unknown function.

S4.40 48_FIG00470960_hypothetical_protein



AAA domain, DNA repair proteins - AAA_25

Hexameric Replicative Helicase - RepA

DNA repair and recombination protein - recomb_RadB

Archaellum biogenesis protein - FlaH

Uncharacterized protein - COG3421

DNA polymerase III subunits gamma and tau - PRK07133

RecA-family ATPase

reticulocyte binding/rhoptry protein - 235kDa-fam

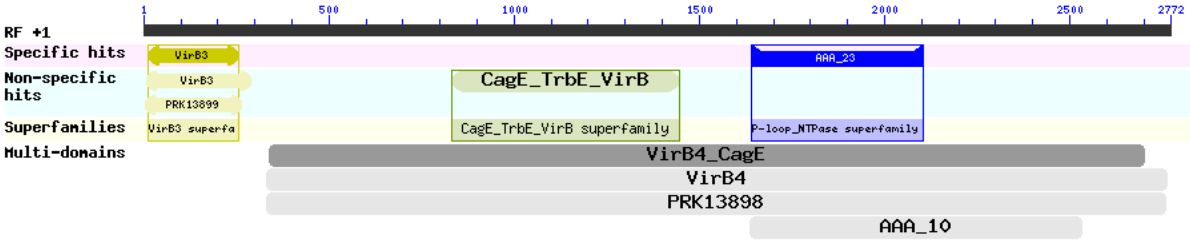
pfam13481	88-528	1.68e-14
cd01125	97-717	4.06e-13
TIGR02237	88-501	4.06e-05
COG2874	97-684	1.02e-03
COG3421	964-1791	3.28e-03
PRK07133	451-1662	1.20e-06
COG3598	88-525	1.37e-05
TIGR01612	589-1785	5.21e-03

This hypothetical protein carries domains for DNA repair and replication. In particular, one of the highest p-value is identified for a hexameric replicative protein RepA domain including its catalytic and polymerisation sites. The RepA protein is an

ATP-dependent helicase encoded by plasmids in gram-negative bacteria, which initiates the replication of DNA (Niedenzu *et al.,* 2001).

The other conserved domains are consistent with the ATPase function of the protein RepA, with the exception of 235kDa-fam (TIGR01612), which is usually found in plasmodium species.

S4.41 49_VirB4



CagE, TrbE, VirB family, component of type IV transporter - CagE_TrbE_VirB

Type IV secretory pathway, VirB3-like protein - VirB3

Type IV secretory pathway - VirB3

type IV secretion system protein VirB3 - PRK13899

AAA domain - AAA_23

type IV secretion/conjugal transfer ATPase, VirB4 - VirB4_CagE

Type IV secretory pathway - VirB4

type IV secretion system ATPase VirB4 - PRK13898

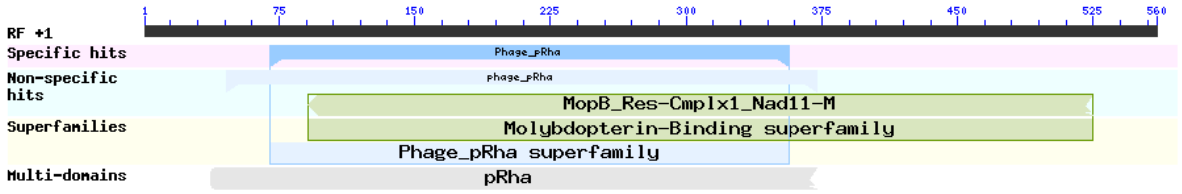
AAA-like domain associated with conjugative transfer protein - AAA_10

pfam03135	832-1446	4.45e-51
pfam05101	10-255	5.23e-24
COG3702	4-294	1.86e-15
PRK13899	1-267	5.55e-10
pfam13476	1642-2103	5.73e-04

TIGR00929	340-2700	0e+00
COG3451	331-2763	0e+00
PRK13898	331-2760	5.71e-134
pfam12846	1636-2532	4.05e-36

This hypothetical protein annotated as VirB4 shows conserved domains for both the protein VirB3 and the protein VirB4 of bacterial Type IV secretion system (Christie and Vogel, 2000). These domains might appear together in this hypothetical protein because of an error in the automated annotation and CDS prediction pipeline.

S4.42 50_Phage_Rha_protein



Phage regulatory protein Rha - Phage_pRha

phage regulatory protein - phage_pRha

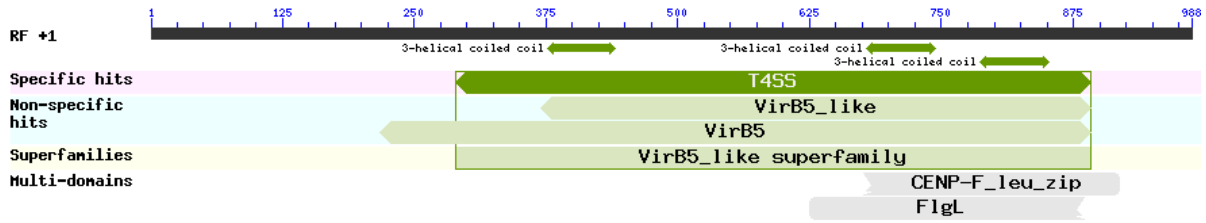
NADH-quinone oxidoreductase/respiratory complex I - MopB_Res-Cmplx1_Nad11-M

Phage regulatory protein Rha - pRha

pfam09669	70-357	2.24e-40
TIGR02681	46-372	3.62e-29
cd02774	91-525	5.16e-04
COG3646	37-372	1.89e-26

The conserved domains included in this hypothetical protein encode for a phage regulatory protein, with the exception of the mitochondrial NADH oxidoreductase, which has the highest e value. This protein, found in temperate phages and bacterial prophages, interferes with the infection of the bacterial cell (Iyer *et al.,* 2002).

S4.43 51_Minor_pilin_of_type_IV_secretion_complex_VirB5



Type IV secretion system proteins - T4SS

VirB5 protein family - VirB5_like

P-type DNA transfer protein - VirB5

Leucine-rich repeats of kinetochore protein Cenp-F/LEK1 - CENP-F_leu_zip

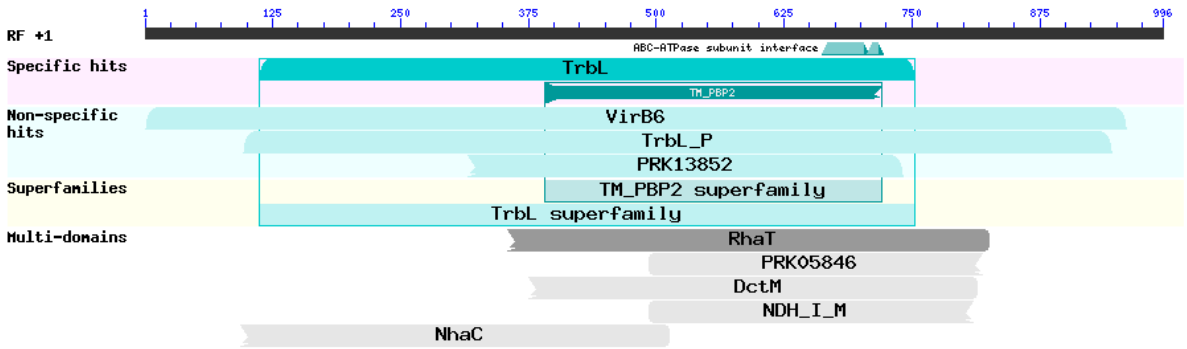
Flagellin and related hook-associated protein - FlgL

pfam07996	289-891	5.48e-30
cd14262	370-891	9.30e-10
TIGR02791	217-891	7.73e-09
pfam10473	676-918	1.81e-03
COG1344	625-885	2.37e-03

This hypothetical protein is associated with several VirB5-like conserved domains. VirB5 is required for the assembly of the Type IV secretion system extracellular pilus. It was shown to be involved in protein-protein interaction with

VirB2, and coiled-coiled motifs that could explain this interaction are identified in the protein sequence (Yeo *et al.,* 2003; Christie, 2004).

S4.44 52_VirB6



TrbL/VirB6 plasmid conjugal transfer protein

Type IV secretory pathway - VirB6

Transmembrane subunit (TM) found in Periplasmic Binding Protein - TM_PBP2

P-type conjugative transfer protein - TrbL_P

type IV secretion system protein VirB6 - PRK13852

Permease of the drug/metabolite transporter (DMT) superfamily - RhaT

NADH:ubiquinone oxidoreductase subunit M - PRK05846

Tripartite ATP-independent periplasmic transporter - DctM

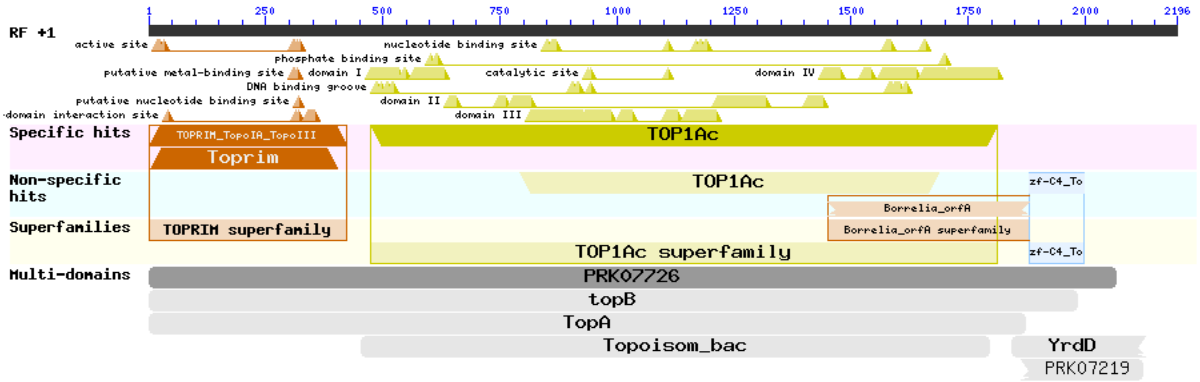
Na+/H+ antiporter - NhaC

pfam04610	112-753	1.14e-49
COG3704	1-960	1.63e-12
cd06261	391-720	3.72e-06
TIGR02783	97-945	4.13e-06
PRK13852	316-741	7.71e-06
COG0697	355-825	3.08e-05

PRK05846	493-819	6.14e-05
pfam06808	376-813	7.11e-05
COG1757	94-513	8.00e-03

The conserved domains in this protein sequence are associated with the TrbL/VirB6 protein of the Type IV secretion system (Hapfelmeier *et al.,* 2000).

S4.45 53_DNA_topoisomerase_III_28828-31020_00-0949_pTet



DNA Topoisomerase, subtype IA - TOP1Ac

Bacterial DNA topoisomerase I DNA-binding domain - TOP1Ac

Topoisomerase-primase domain - TOPRIM_TopoIA_TopoIII

Toprim domain

Topoisomerase DNA binding C4 zinc finger - zf-C4_Topoisom

Borrelia ORF-A

DNA topoisomerase III - PRK07727

DNA topoisomerase III - TopB

DNA topoisomerase IA - TopA

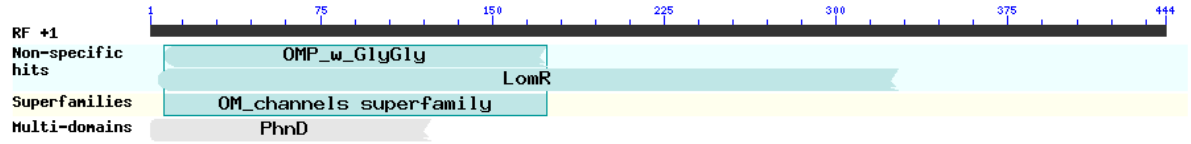
DNA topoisomerase - Topoisom_bac

ssDNA-binding Zn-finger and Zn-ribbon domains of topoisomerase 1 - YrdD

DNA topoisomerase I - PRK07219		
cd00186	475-1812	3.76e-86
smart00437	793-1686	8.16e-54
cd03362	1-423	3.93e-46
pfam01751	4-402	6.58e-07
pfam01396	1882-1998	2.55e-06
pfam02414	1450-1881	1.82e-04
PRK07727	1-2067	0e+00
TIGR01056	1-1983	2.64e-173
COG0550	1-1872	6.76e-151
pfam01131	454-1794	1.99e-111
COG0551	1843-2130	1.28e-10
PRK07219	1864-2124	3.48e-05

The conserved domains identified in this protein sequence are consistent with the automated annotation and they are related to a bacterial topoisomerase. Topoisomerases break the single or double stranded DNA and re-join the phosphodiester-backbone: these enzymes tune the supercoil state of DNA, facilitating interactions with other proteins (Champoux, 2001).

S4.46 54_FIG00470281_hypothetical_protein

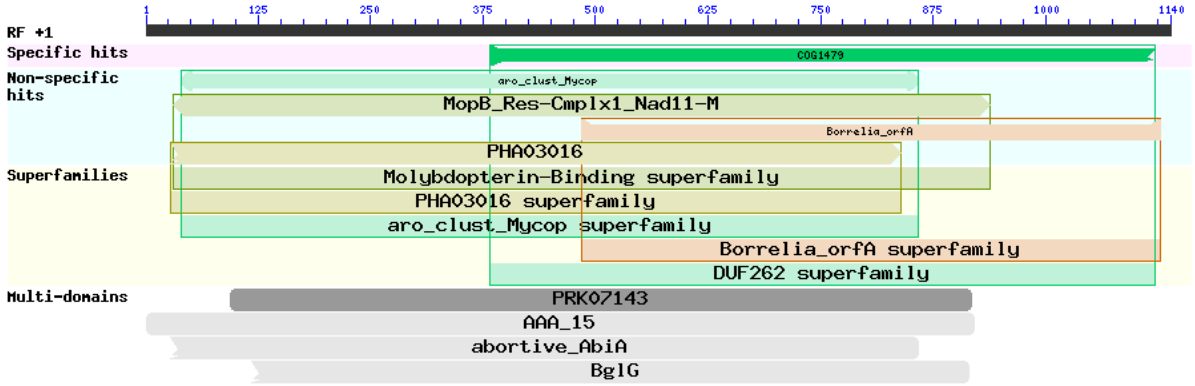


Outer membrane protein - OMP_w_GlyGly Opacity protein and related surface antigens - LomR Alkylphosphonate ABC transporter - PhnD		
TIGR04219	7-174	1.79e-03
COG3637	4-327	6.55e-03
TIGR03431	1-123	1.85e-03

This hypothetical protein shows hits with conserved domain associated with membrane transport. Both OMP_w_GlyGly and LomR are part of a superfamily of porins (cl21487), which are non-specific channels for transport of various molecules, for instance hydrophilic molecules or various sugars (Nikaido, 2003).

A further partial conserved domain is involved in the uptake of alkylphosphonate in *E. coli* (Chen *et al.,* 1990).

S4.47 55_FIG00470802_hypothetical_protein



aromatic cluster surface protein - aro_clust_Mycop

Mitochondrial-encoded NADH-quinone oxidoreductase - MopB_Res-Cmplx1_Nad11-M

Borrelia ORF-A

Uncharacterized conserved protein - COG1479

hypothetical protein - PHA03016

hypothetical protein - PRK07143

AAA ATPase domain - AAA_15

Abortive phage infection protein - abortive_AbiA

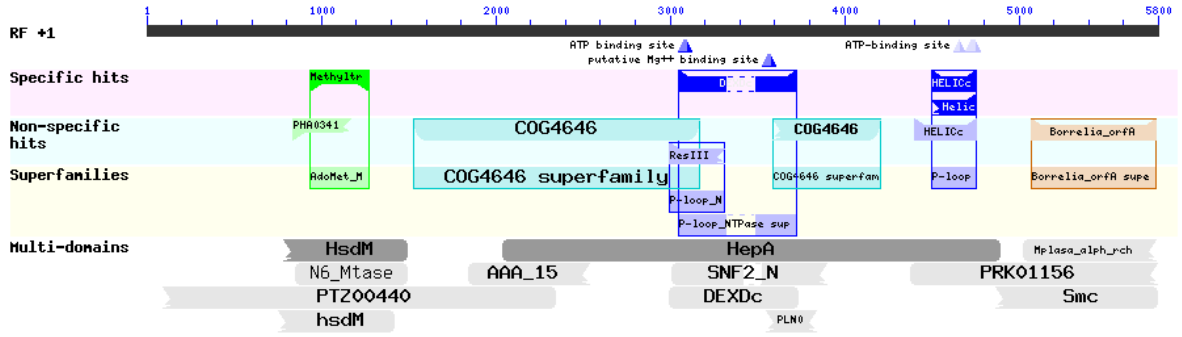
Transcriptional antiterminator - BglG

TIGR04313	40-858	1.35e-13
cd02774	31-939	1.21e-10
pfam02414	484-1128	1.29e-09
COG1479	382-1122	2.26e-09
PHA03016	28-840	1.74e-08
PRK07143	94-918	1.09e-06
pfam13175	1-921	4.12e-11
TIGR04499	28-858	8.62e-09
COG3711	118-915	1.84e-07

Several conserved domains identified in this hypothetical protein are uncharacterised or of unknown function: these include TIGR04313, pfam02414, COG1479, PHA03016 (Bawden *et al.,* 2000).

Other conserved domains are associated with a eukaryotic respiratory complex (Mitochondrial-encoded NADH-quinone oxidoreductase, Oudot *et al.,* 1999), a protein for phage defence in *Lactococcus lactis* (Abortive phage infection protein, Tangney and Fitzgerald, 2002) and a transcriptional anti-terminator protein (Manival *et al.,* 1997).

S4.48 56_FIG00470802_hypothetical_protein



Adenine-specific DNA methylase - COG4646

Adenine-specific DNA methylase - COG4646

Methyltransferase domain - Methyltransf_26

helicase superfamily c-terminal domain - HELICc

Type III restriction enzyme - ResIII

Helicase conserved C-terminal domain - Helicase_C

DEAD-like helicases superfamily - DEXDc

Helicase superfamily c-terminal domain, DEXDc-, DEAD-, and DEAH-box proteins - HELICc

Borrelia ORF-A - Borrelia_orfA

putative methyltransferase - PHA03412

Superfamily II DNA or RNA helicase, SNF2 family - HepA

Type I restriction-modification system, DNA methylase subunit - HsdM

N-6 DNA Methylase; Restriction-modification (R-M) systems - N6_Mtase

SNF2 family N-terminal domain - SNF2_N

DEAD-like helicases superfamily - DEXDc

helix-rich Mycoplasma protein - Mplasa_alph_rch

Glucosyl transferase GtrII - B3:C24 PTZ00440

type I restriction system adenine methylase (hsdM) - HsdM

chromosome segregation protein - PRK01156

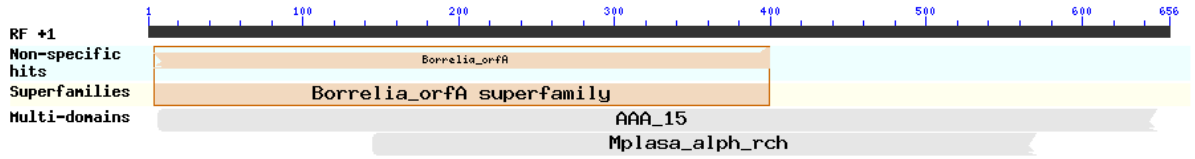
Probable chromatin-re-modeling complex ATPase chain - PLN03142 Chromosome segregation ATPase - Smc AAA ATPase domain - AAA_15 EpsG family - EpsG Glucosyl transferase GtrII - Glucos_trans_IIB2B1:C24B24B3:C24		
COG4646	1528-3168	1.15e-74
COG4646	3592-4209	5.07e-44
pfam13659	937-1269	1.01e-17
smart00490	4498-4752	1.13e-07
pfam04851	2995-3312	1.32e-06
pfam00271	4498-4752	2.97e-06
cd00046	3049-3723	3.69e-04
cd00079	4402-4758	1.03e-03
pfam02414	5068-5784	1.90e-03
PHA03412	838-1167	8.87e-03
COG0553	2041-4887	8.16e-13
COG0286	784-1494	5.77e-15
pfam02384	853-1494	4.53e-13
pfam00176	3013-3894	9.04e-12
smart00487	2998-3732	4.04e-11
TIGR04523	5026-5784	3.27e-08
PTZ00440	91-2343	3.51e-06
TIGR00497	757-1419	8.63e-06
PRK01156	4378-5790	1.89e-05
PLN03142	3553-3834	2.97e-04

COG1196	4870-5796	8.73e-04
pfam13175	1846-2538	1.31e-03
pfam14897	5017-5799	6.00e-07
pfam14264	94-828	2.40e-05

This is the largest predicted hypothetical protein. In total, 24 conserved domains are predicted. The lowest E-values are identified in three domains associated with an adenine-specific methyl transferase domain, involved in DNA protection and repair (Zhang-Akiyama *et al.,* 2009).

Furthermore, a group of Helicase-associated conserved domains is characterised by an E-value between 1.0e-13 and 1.0e-15. This class of proteins is associated with the unwinding of supercoiled DNA in order to regulate transcription and replication (Muzzin *et al.,* 1998).

S4.49 57_FIG00469861_hypothetical_protein



Borrelia ORF-A - Borrelia_orfA

AAA ATPase domain - AAA_15

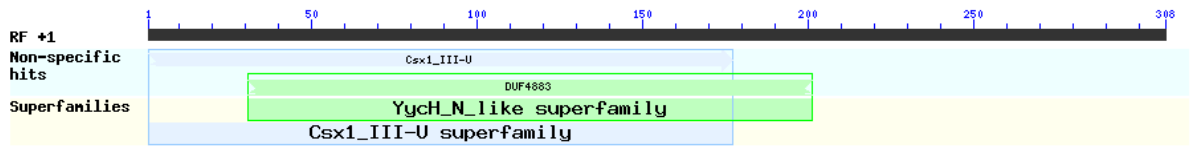
helix-rich Mycoplasma protein - Mplasa_alph_rch

pfam02414	4-399	1.96E-03
pfam13175	7-648	1.69E-04
TIGR04523	145-570	9.48E-04

This hypothetical protein is associated with an uncharacterised domain found in plasmid DNA of the spirochete *Borrelia* (Zückert and Meyer, 1996).

Other multi-domains family are associated with an ATPase domain and an alpha-helix rich uncharacterised domain.

S4.50 58_FIG00471069_hypothetical_protein



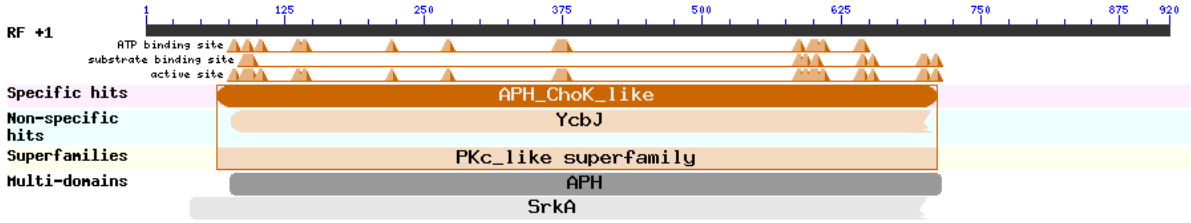
CRISPR/Cas system-associated protein Csx1 - Csx1_III-U

DOmain of unknown function - DUF4883

cd09728	1-177	2.64E-03
pfam16224	31-201	7.03E-03

Two domains of different functions are identified in this hypothetical protein: a CRISPR/Cas system domain, and a domain of unknown function belonging to the superfamily YycH_N_like. This superfamily includes the membrane signal transductor of a two component system YycG-YycC in several bacteria, including *S. aureus* (Makarova *et al.,* 2006 (2); Türck and Bierbaum, 2012).

S4.51 59_aminoglycoside_3_phosphotransferase_aph3_HypA



Aminoglycoside 3'-phosphotransferase and Choline Kinase family - APH_ChoK_like

Predicted kinase, aminoglycoside phosphotransferase (APT) family - YcbJ

Phosphotransferase enzyme family - APH

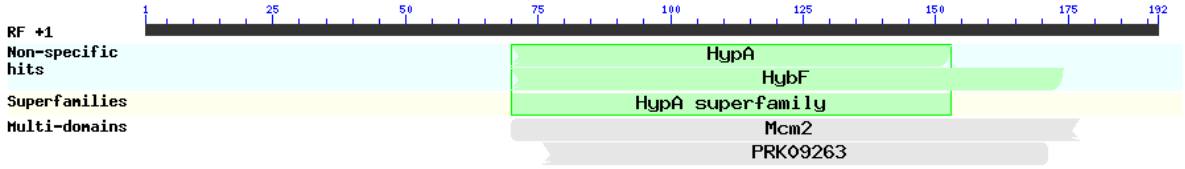
Ser/Thr protein kinase RdoA involved in Cpx stress response - SrkA

cd05120	64-711	1.15E-25
COG3173	76-705	4.95E-10
pfam01636	76-714	1.29E-22
COG2334	40-702	1.49E-09

This putative protein is associated with an aminoglycoside phosphotransferase function. The enzyme catalyses the transfer of a phosphate group to a small molecule such as an antibiotic belonging to the aminoglycoside or macrolide class in order to inactivate it. This protein may also catalyse the transfer of a phosphate group to other small molecules, such as choline, ethanolamine, and homoserine,

which function as precursors for the synthesis of several biological compounds, like phospholipids or amino acids (Kim and Mobashery, 2005).

S4.52 60_hypothetical_protein

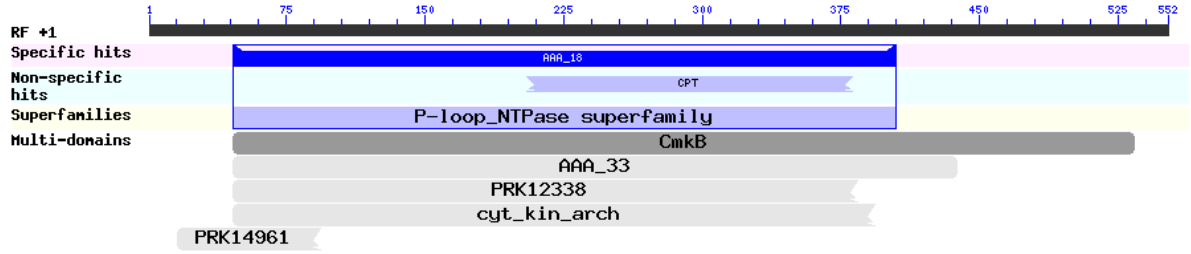


Hydrogenase/urease nickel incorporation, metallochaperone - HypA Hydrogenase maturation metallochaperone HypA/HybF - HybF DNA replicative helicase MCM subunit Mcm2 - Mcm2 anaerobic ribonucleoside triphosphate reductase - PRK09263		
pfam01155	70-153	3.44E-04
COG0375	70-174	3.86E-03
COG1241	70-177	6.19E-04
PRK09263	76-171	1.86E-03

A domain for the incorporation of nichel in Nichel-containing enzyme is identified in this protein. This type of protein was identified and described in *H. pylori* (Xia *et al.,* 2009).

Two further multi-domains are identified, involved with DNA interaction (helicase) and with NTP reduction.

S4.53 61_hypothetical_protein



AAA domain - AAA_18

Chloramphenicol (Cm) phosphotransferase (CPT) - CPT

Cytidylate kinase [Nucleotide transport and metabolism] - CmkB

AAA domain - AAA_33

hypothetical protein; Provisional - PRK12338

cytidylate kinase - cyt_kin_arch

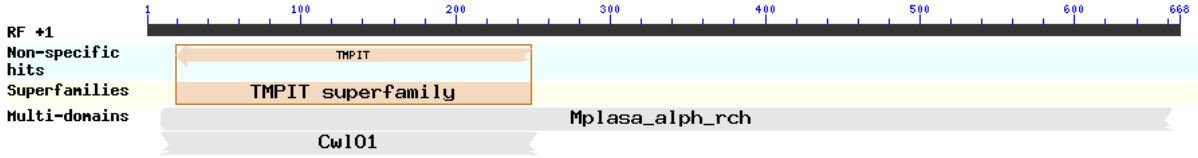
DNA polymerase III subunits gamma and tau - PRK14961

pfam13238	46-405	3.35E-11
cd00227	205-381	4.85E-03
COG1102	46-534	6.09E-04
pfam13671	46-438	4.75E-06
PRK12338	46-384	1.41E-03
TIGR02173	46-393	3.10E-03
PRK14961	16-93	6.64E-03

The lowest e-value score for this hypothetical protein is associated with an ATPase domain associated with a variety of cellular functions, from DNA replication to signal transduction and the regulation of gene expression (AAA_18) (Iyer *et al.*, 2004).

A partial conserved domain (CPT, cd00227) of the same superfamily (P-loop_NTPase) is identified and is associated with resistance to chloramphenicol. (Izard, 2001).

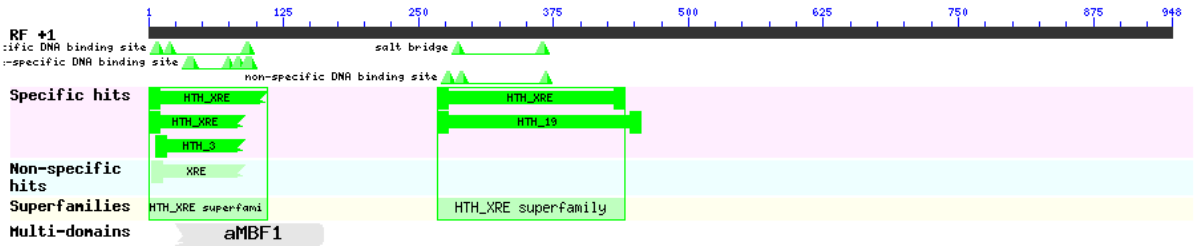
S4.54 62_hypothetical_protein



TMPIT-like transmembrane protein helix-rich Mycoplasma protein - Mplasa_alph_rch Uncharacterized N-terminal domain of peptidoglycan hydrolase CwlO1		
pfam07851	19-249	5.76E-04
TIGR04523	10-663	1.96E-04
COG3883	10-252	3.52E-03

The conserved domains associated with this hypothetical protein are associated with a membrane localisation but unknown function, and no literature in support.

S4.55 63_hypothetical_protein



Helix-turn-helix XRE-family like proteins - HTH_XRE

Helix-turn-helix XRE-family like proteins - HTH_XRE

Helix-turn-helix - HTH_3

DNA-binding transcriptional regulator - XRE

Helix-turn-helix XRE-family like proteins - HTH_XRE

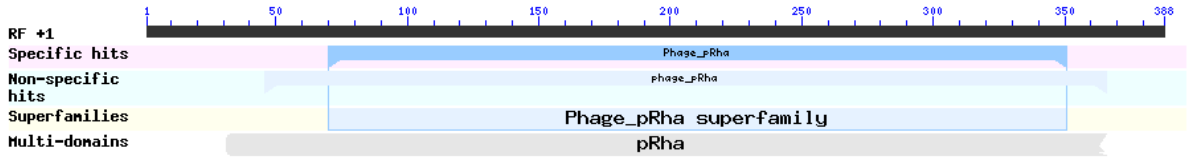
Helix-turn-helix domain, DNA-binding helix-turn-helix domain - HTH_19

Archaeal ribosome-binding protein aMBF1

cd00093	1-111	6.56E-04
smart00530	1-90	8.42E-04
pfam01381	7-90	1.63E-03
COG1476	4-90	2.35E-03
cd00093	268-441	4.72E-03
pfam12844	268-456	7.53E-03
COG1813	25-162	3.02E-03

In this hypothetical protei,n several DNA-protein interactions conserved domains are identified. The HTH_XRE domain is described as a DNA binding protein regulating gene expression in response to xenobiotic compounds (Wintjens and Rooman, 1996; Luscombe *et al.,* 2000; Ren *et al.,* 2010).

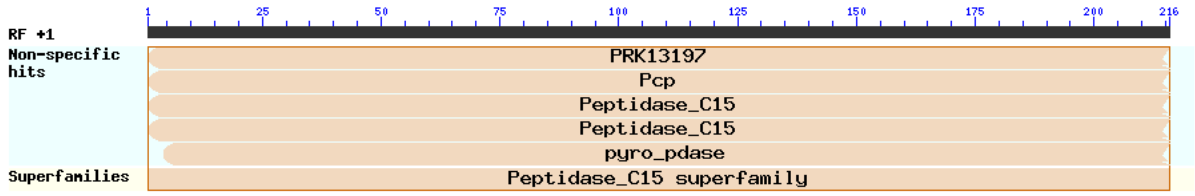
S4.56 64_antirepressor_partial



Phage regulatory protein Rha - Phage_pRha phage regulatory protein, rha family [Mobile element, Prophage functions] - phage_pRha Phage regulatory protein Rha - pRha		
pfam09669	70-351	1.32E-34
TIGR02681	46-366	4.23E-24
COG3646	31-366	2.68E-25

This hypothetical protein is associated with phage regulatory proteins. The protein pRha is a phage regulative protein. The presence of this gene interferes with infection of bacterial strains without integration host factor (IHF) (Henthorn and Friedman, 1995).

S4.57 65_peptidase_C15_partial



pyrrolidone-carboxylate peptidase - PRK13197

Pyrrolidone-carboxylate peptidase - Pcp

Pyroglutamyl peptidase (PGP) type I, pyrrolidone carboxyl peptidase (pcp) type I - Peptidase_C15

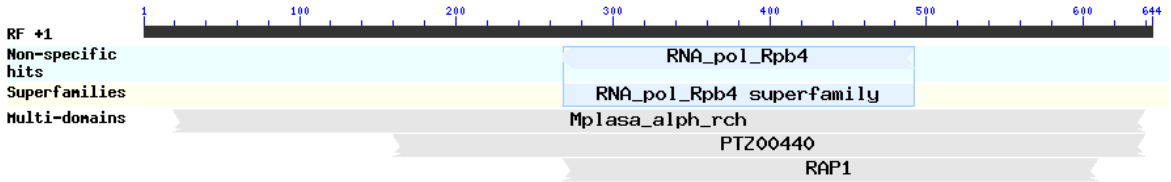
Pyroglutamyl peptidase - Peptidase_C15

pyroglutamyl-peptidase I - pyro_pdase

PRK13197	1-216	6.35E-38
COG2039	1-216	1.17E-30
cd00501	1-216	1.22E-30
pfam01470	1-216	3.46E-30
TIGR00504	4-216	8.48E-21

The conserved domains identified in this protein are consistent with the automated annotation: a peptidase associated with the cleavage of C-terminal pyroglutamate. The role of this protein in the various species is not fully understood (Tanaka *et al.,* 2001; Barrett *et al.,* 2003; Cummins and O'Connor, 1998).

S4.58 66_hypothetical_protein



RNA polymerase Rpb4 - RNA_pol_Rpb4

helix-rich Mycoplasma protein - Mplasa_alph_rch

reticulocyte binding protein 2-like protein - PTZ00440

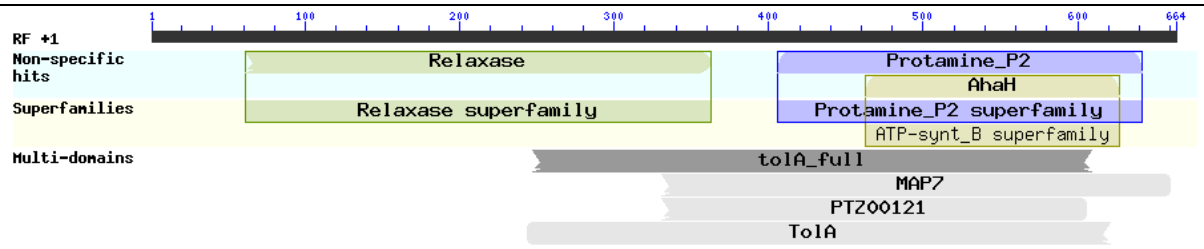
Rhoptry-associated protein 1 (RAP-1) - RAP1

pfam03874	268-492	3.62E-05
TIGR04523	19-639	1.34E-05
PTZ00440	160-639	8.95E-04
pfam07218	268-609	1.02E-03

The partial conserved domain (RNA_pol_Rpb4) is identified in this hypothetical protein. It is associated with an essential subunit of the RNA polymerase III.

Other partial multi-domains families are identified, including a Mycoplasma associated-helix rich domain, and two domains associated with immune response in mycoplasma (Todone *et al.,* 2001; Sadhale and Woychik, 1994; Galinski *et al.,* 2000).

S4.59 68_TolA_protein



Relaxase/Mobilization nuclease domain - Relaxase

Sperm histone P2; substitute for histones in the chromatin of sperm - Protamine_P2

ATP synthase archaeal, H subunit - AhaH

TolA protein - tolA_full

MAP7 (E-MAP-115) family; microtubule-stabilizing protein - MAP7

MAEBL; Provisional - PTZ00121

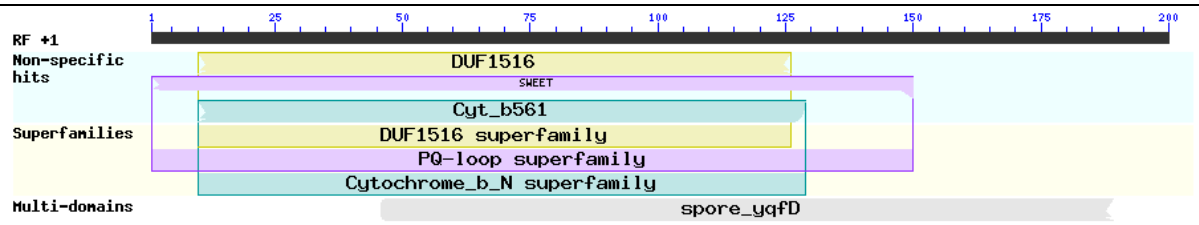
Membrane protein involved in *coli*cin uptake [Cell wall/membrane/envelope biogenesis] - TolA

pfam03432	61-363	4.10E-07
pfam00841	406-642	2.56E-04
TIGR02926	463-627	3.21E-03
TIGR02794	247-609	3.81E-06
pfam05672	331-660	7.92E-06
PTZ00121	331-606	3.42E-05
COG3064	244-621	8.51E-05

This protein is annotated as a TolA protein, part of the Tol-Pal complex. It is required for maintaining outer membrane integrity and also involved in uptake of *coli*cins and filamentous DNA, and implicated in pathogenesis. The TolA and tolA_full Multi-domains are identified in this protein, alongside other domains of different functions including mobilisation of DNA (pfam03432) and an ATP synthase

associated with archaeal organisms (TIGR02926) (Levengood and Webster, 1989; Derouiche *et al.,* 1996; Meyer, 2000).

S4.60 70_FIG00469626_hypothetical_protein



Protein of unknown function - DUF1516

Sugar transporter, SemiSWEET family [Carbohydrate transport and metabolism] - SWEET

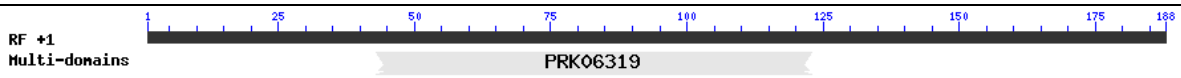
Eukaryotic cytochrome b(561) - Cyt_b561

sporulation protein YqfD - spore_yqfD

pfam07457	10-126	2.55E-04
COG4095	1-150	4.81E-04
cd08554	10-129	5.60E-04
TIGR02876	46-189	8.06E-03

In this hypothetical protein, three non-specific hits are identified: one of them is associated with a sucrose membrane transporter (SWEET), part of a PQ-loop superfamily (Xuan *et al.,* 2013).

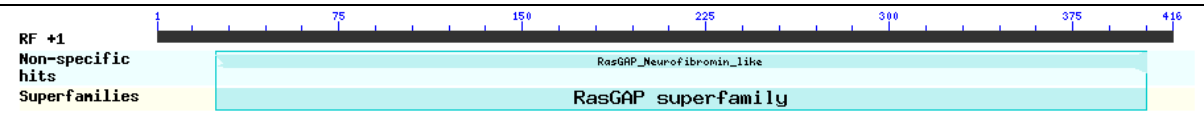
S4.61 72_DNA_topoisomerase_III_TraE_(EC_5.99.1.2)



DNA topoisomerase I/SWI domain fusion protein - PRK06319		
PRK06319	43-123	2.32E-04

This protein is annotated by the automatic pipeline as DNA topoisomerase III, and the conserved domain identified in it confirms this.

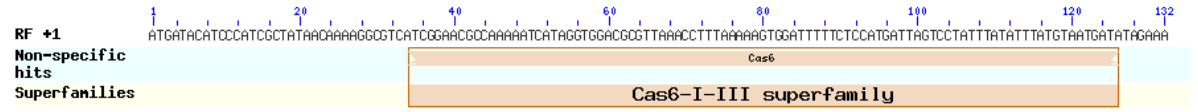
S4.62 75_FIG00471024_hypothetical_protein



Ras-GTPase Activating Domain of proteins similar to neurofibromin - RasGAP_Neurofibromin_like		
cd05392	25-405	6.91E-05

The partial conserved domain identified in this protein is associated with eukaryotic organisms. Ras proteins have been described in yeast and human cells to activate the cAMP/protein kinase A (PKA) pathway, controlling metabolism, stress resistance, growth, and meiosis (Harrisingh and Lloyd, 2004).

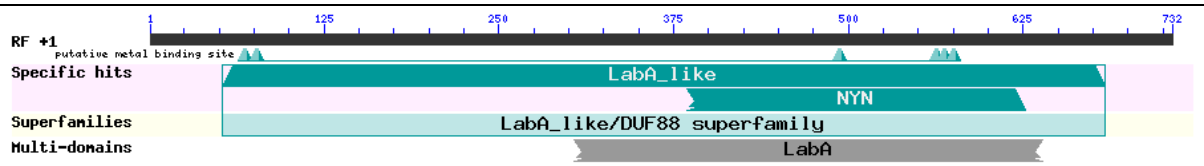
S4.63 76_hypothetical_protein



CRISPR/Cas system endoribonuclease Cas6		
COG1583	34-126	8.62E-03

This small hypothetical protein shows the presence of a CRISPR/Cas system conserved domain belonging to the Cas6 superfamily. As described in other hypothetical proteins, CRISPR/Cas systems are associated with defence from phage infection in bacteria (Makarova *et al.,* 2006 (2), Haft *et al.,* 2005).

S4.64 77_FIG00472625_hypothetical_protein

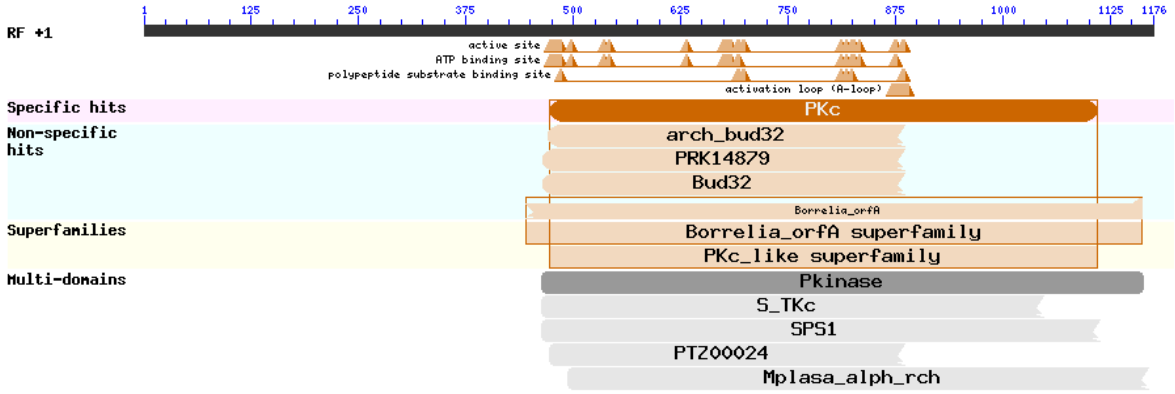


LabA_like proteins NYN domain Uncharacterized conserved; LabA/DUF88 family [Function unknown] - LabA		
cd06167	52-684	6.40E-20
pfam01936	385-627	2.09E-08
COG1432	304-639	2.72E-10

The LabA domain found in the hypothetical protein encoded by this gene has no defined function but it is well conserved in all kingdoms of life. In cyanobacteria, it is required to regulate the circadian clock system via negative feedback of an

autokinase (KaiC). In *Pseudomonas putida* a protein falling in the “LabA family” is involved in the bacterial degradation of nicotine. The hypothetical protein also exhibits a putative metal binding site, structurally similar to proteins like YrbI phosphatase from *Haemophilus influenzae* (Taniguchi *et al.,* 2007; Taniguchi *et al.,* 2010; Tang *et al.,* 2008; Parsons *et al.,* 2002).

S4.65 79_pK2_Serinethreonine_protein_kinase_PrkC2C_regulator_of _stationary_phase



Catalytic domain of Protein Kinases - PKc

Kae1-associated kinase Bud32 - arch_bud32

serine/threonine protein kinase - PRK14879

tRNA A-37 threonylcarbamoyl transferase component Bud32 - Bud32

Borrelia ORF-A

Protein kinase domain - Pkinase

Serine/Threonine protein kinases, catalytic domain - S_TKc

Serine/threonine protein kinase [Signal transduction mechanisms] - SPS1

cyclin-dependent protein kinase - PTZ00024

helix-rich Mycoplasma protein - Mplasa_alph_rch

Domain of unknown function; integral membrane proteins - DUF4271

cd00180	472-1110	2.32E-46
TIGR03724	469-885	1.44E-10
PRK14879	463-885	1.15E-09
COG3642	463-885	5.99E-06
pfam02414	445-1161	4.91E-05
pfam00069	463-1164	5.85E-34
smart00220	463-1047	7.44E-36
COG0515	463-1113	2.43E-30
PTZ00024	472-885	3.91E-17
TIGR04523	493-1170	2.42E-06
pfam14093	382-990	1.67E-07

The conserved domains identified with a solid hit in this hypothetical protein are associated with a protein kinase function. Protein kinases in bacteria are often involved in signal transduction and response regulation (Kannan *et al.,* 2007; Handford *et al.,* 2009; Pérez *et al.,* 2008).

Acknowledgements

This work wouldn’t have happened without the help and support of a few very important people.

In no particular order (Daly, 2016):

My family. The support and love that my mum and my sister gave me in the last few years is impossible to describe; it is actually unbelievable. Thank you, I’ll love you forever. And my dad, without your support this would have been much harder. And my Turkish family, you just arrived, you are already so important. And it’s such a shame that nana won’t see this thesis, she would have read it all. I miss you.

My wife. I came for the science and I found the one love of my life, again, unbelievable. I am the luckiest man on earth; you make every day a better one. Together, we can do everything. We got each other and that’s a lot (Jovi, 1986). I love you.

My supervisors. I had to wait 26 years to find the best teachers, but it was definitely worth the wait. What Rick, David and Olivia thought me in science and in life is invaluable. It was a pleasure and an honour working with you.

My friends close. Callan, Tom, Nathan, Alex, Silvia, Ed, Juliette (also Jani, Melek, Benny, David and Pam, even if unfortunately, they are not so close anymore). If I never felt home sick it is because of you guys, you are the best. You made England, home.

My friends far. Luca, Giacomo, Giorgio, Giovanni, Roberto, Dario. I miss you every day. Why don’t you just move here, already? And Mavi (you are pretty close now!), Stefano, Elisa, Betta, Mirko, Cami, Elena. You all guys are my roots, and is good to have solid roots, that will always be there.

My workmates. The people I encountered during this journey, and made it the very best. Vanessa (thank you for the massive help with the congo red assay), Babs,

Michael, Harriet, Rebecca, Nicola, Phil, Selina, Sariqa, Tom, Jamie and everyone else. And Steve Porter for the supervision in my first year project.

My SWDTPmates. Alex, Mickey, Odette and Emma, I will never forget my first year, and that's also because of you. And Ron Yang. Best lecturer, best courses.

Dr Callan Davies, again. He is a Shakespeare laureate and went through the whole thing. I can't wait to read a poem on pTet-like plasmids. Thank you, mate.

Of course I want to acknowledge the University of Exeter for the support and the BBSRC for the generous funding of my research.

That's it, I think. If I forgot someone, please don't be mad.

Thank you all, very very much.