Mechanistic and functional analysis of Cj0031: a phase variable methyltransferase in

Campylobacter jejuni

Thesis submitted for the degree of Doctor of Philosophy
at the University of Leicester

By

Awais Anjum

Department of Genetics

University of Leicester

May 2013
Statement of Originality

This accompanying thesis submitted for the degree of PhD entitled: Mechanistic and functional analysis of Cj0031: a phase variable methyltransferase in Campylobacter C. jejuni is based on work carried out by the author in the Department of Genetics at the University of Leicester mainly during the period between January 2009 and December 2012. All of the work recorded in this thesis is original and consequence of authors’ research endeavour unless otherwise acknowledged in the text or by references. None of the work has been submitted for another degree in this or any other university.

Name AWAIS ANJUM

Date 15/05/2013
Abstract

*Campylobacter jejuni* constitutes the major cause of food-borne diarrheal disease in developed countries. The genome of this species comprises many surface genes having mononucleotide repeat tracts (PolyG/PolyC) which undergo reversible switching between ON and OFF phases, termed phase variation. Phase variation helps bacteria to colonize the host effectively as most phase variable genes are involved in expression of cell surface structures mediating interactions with the host. The mutation rates of phase variable genes are important determinants for genetic diversity and overall fitness of the bacterial population residing in various niches. The major aim of the project was to determine the phase variation rates for the simple sequence repeat tracts of varying lengths in *cj0031* and *capA* of *C. jejuni* strain 11168. The PV rates were determined by using chromosomally-located reporter construct for *cj0031* and by an immunoblotting assay for *capA*. The mutation rates of a G10 tract were 1.5 fold higher than a G9 repeat tract in *cj0031*. Similarly, a 6-fold increase in the PV rate was recorded for G12 in *capA* over a G10 repeat tract. The mutational spectra for G9 and G10 were predominantly insertions and were shifted to mainly deletions for G11 and G12 repeats.

Major shifts in the ON/OFF status of phase variable genes of *C. jejuni* strain 11168 were detected by performing a multiplex PCR on isolates following passage through chickens. Thirteen novel genotypes found in the output population indicated a high level of genetic diversity was generated by changes in repeat tract lengths of phase variable genes.

Bioinformatics analysis of *cj0031* revealed a homology to type IIG restriction modification systems. A Southern blot analysis demonstrated that *cj0031* possessed methyltransferase activity and led to the conclusion that 5’ CCCGA 3’/5’ CCCGAA 3’ were putative recognition sequences of Cj0031 methyltransferase. An investigation of functional abilities showed that Cj0031 enhanced the capability of adhesion, invasion and biofilm formation without having any affect on the motility of *C. jejuni*. A 5-fold restriction activity was exerted by Cj0031 on one phage type, showing that this enzyme also possessed restriction activity although this was marginal in comparison to restriction endonucleases in *E. coli*. It is postulated that Cj0031 mainly controls the phase variation of other genes in *C. jejuni* through methylation of target sequences located either in promoter region or intergenic regions near the promoters of such genes, rather than having active involvement in protection of hosts from phages. The high PV rates of *cj0031* might be compatible with its role as phase variation to facilitate the rapid adaptation of *C. jejuni* to the micro-environment of hosts.
Acknowledgement

I would like to offer my sincere thanks and gratitude to Almighty Lord, ALLAH who showed me enlightenment and guidance to complete this great work successfully. I would like to express my deep sense of appreciation my worthy supervisor, DR. Christopher D. Bayliss – a man of substance and ingenuity, for his guidance, inspiration and motivation and absorbing attitude through my research project. I am very much impressed of his all round co-operation, sympathetic behaviour, salutary suggestions and illuminating criticism during this research project.

I am unable to find appropriate words to pay my zealous thanks to members of thesis committee whose sympathetic attitude and beneficial cooperation enabled me to complete this project. Words are lacking to express my thanks to Pak Foods Ltd (UK) and HEC Pakistan and the Charles Wallace Pakistan Trust for funding my project. I feel immense pleasure to express my deepest gratitude and sincere thanks to all members of my lab and the Genetics department.

My expressions are still begging the words to pay my gratitude to my loving and caring great father (Late) and mother who helped me in each and every way during completion of this work. Their supreme guidance and support burned the light for me to let me pass through dark alleys of life safely; and they have always been a beacon house of guidance for me. I must not forget to extend my thanks to my sweet brothers, sisters and my loving wife who remembered me in their prayers and supported me in all respects along the awful avenues of my academic achievements.

AWAIS ANJUM
### Table of Contents

**STATEMENT OF ORIGINALITY**  
2

**ABSTRACT**  
2

**ACKNOWLEDGEMENT**  
3

**CHAPTER 1. INTRODUCTION**  
9

1.1: ADAPTIVE BEHAVIOUR OF PROKARYOTES:  
9
1.2: PHASE AND ANTIGENIC VARIATION  
12
1.3: PHASE VARIABLE PHENOTYPES IN PATHOGENS  
17
1.3.1: CAPSULE  
17
1.3.2: FIMBRIAE AND PILI  
18
1.3.3: FLAGELLA  
19
1.3.4: LPS AND LOS MODIFICATIONS  
20
1.3.5: DNA RESTRICTION-MODIFICATION SYSTEMS  
21
1.4: MOLECULAR MECHANISMS OF PHASE VARIATION  
26
1.4.1: SHORT SEQUENCE REPEATS AND SLIPPED-STRAND MISPARRING  
28
1.4.2: HOMOLOGOUS (GENERAL) RECOMBINATION  
32
1.4.3: SITE-SPECIFIC RECOMBINATION  
36
1.4.4: TRANSPOSITION:  
39
1.4.5: EPIGENETIC MECHANISM OF PHASE VARIATION:  
41
1.4.6: PHASEVARIN:  
45
1.5: DETERMINANTS OF PHASE VARIATION:  
48
1.5.1: PHASE VARIATION IN CAMPYLOBACTER JEJUNI  
52
1.5.1.1: THE GENUS CAMPYLOBACTER:  
52
1.5.2: GROWTH AND SURVIVAL CHARACTERISTICS:  
54
1.5.3: CAMPYLOBACTERIOSIS: OCCURRENCE AND ITS CLINICAL MANIFESTATIONS  
54
1.5.4: TRANSMISSION OF CAMPYLOBACTER TO HUMANS  
55
1.5.5: GENOME OF CAMPYLOBACTER JEJUNI:  
56
1.5.6: HOMOPOLYMERIC TRACTS IN C. JEJUNI-----ROLE IN PATHOGENICITY  
58
1.5.8: PHASE VARIABLE PHENOTYPES IN CAMPYLOBACTER JEJUNI  
61
1.6: DNA RESTRICTION-MODIFICATION SYSTEMS  
67
1.6.1: CLASSIFICATION OF RM SYSTEMS:  
68
1.7: BACKGROUND TO THE PROJECT:  
78
1.8: LIST OF OBJECTIVES  
78

**CHAPTER 2. MATERIALS AND METHODS**  
80

2.1: BACTERIAL STRAINS AND GROWTH CONDITIONS:  
80
2.2: LIST OF PRIMERS USED DURING THE STUDY:  
81
2.3: POLYMERASE CHAIN REACTION (PCR):  
82
2.4: DNA SEQUENCING:  
84
2.5: DNA MANIPULATIONS:  
85
2.5.1: LARGE SCALE CSCL PREPARATION OF CHROMOSOMAL DNA  
85
2.5.2: SMALL SCALE PREPARATION OF CHROMOSOMAL DNA:  
86
2.5.3: ETHANOL PRECIPITATION:  
87
2.5.4: RESTRICTION ENZYME DIGESTION:  
88
2.5.5: ESTIMATION OF DNA CONCENTRATION:  
88
2.6: TRANSFORMATION OF E. coli WITH PLASMID DNA: 88
2.6.1: CALCIUM CHLORIDE METHOD: 89
2.6.2: ELECTROPORATION: 89
2.6.3: NATURAL TRANSFORMATION: 90
2.7: COLONY IMMUNOBLOTTING: 91
2.8: PHASE VARIATION ASSAYS: 92
2.9: ANALYSIS OF PV GENOTYPES FOR CAMPYLOBACTER JEJUNI: 93
2.9.1: A-TAILING: 95
2.9.2: GENEScan ASSAY: 96
2.9.3: DNA SEQUENCING OF SAMPLES: 96
2.10: CHICKEN COLONIZATION EXPERIMENTS: 97
2.11: C. JEJUNI INTERACTIONS WITH CULTURED HUMAN EPITHELIAL CELL LINES: 98
2.11.1: MAINTENANCE OF TISSUE CULTURE CELLS: 98
2.11.2: ADHESION ASSAY: 99
2.11.3: INVASION ASSAY: 100
2.12: CRYSTAL VIOLET STAINING METHOD FOR MEASUREMENT OF BIOFILM FORMATION 101
2.13: MOTILITY ASSAY: 101

2.14: SOUTHERN BLOTTING ASSAY FOR DETERMINATION OF METHYLATION ACTIVITY OF 
Cj0031: 102
2.14.1: AMPLIFICATION OF PROBES: 102
2.14.3: DETERMINATION OF LABELLING EFFICIENCY: 103
2.14.4: DNA TRANSFER AND FIXATION: 104
2.14.5: HYBRIDIZATION REACTION: 105
2.14.6: STRINGENCY WASHES: 105
2.14.7: IMMUNOLOGICAL DETECTION: 106

2.15: PROCEDUREs FOR PERFORMING PHAGE ASSAY: 106
2.15.1: PLAQUE FORMING ABILITY OF PHAGES ON C. JEJUNI HOST STRAINS: 106
2.15.2: PSEUDOLYSOGENIC BACTERIOPHAGE PROPAGATION 107
2.15.3: BACTERIOPHAGE TITRATION 107
2.16: STATISTICAL ANALYSIS: 108

CHAPTER 3. THE IN VIVO AND IN VITRO EXAMINATION OF POLYMORPHISM 
IN THE PHASE VARIABLE GENES OF C. JEJUNI 109

3.1. INTRODUCTION: 109
3.2. RESULTS: 112
3.2.1. EXAMINATION OF THE CHANGES IN THE PHASE VARIABLE GENES OF C. JEJUNI DURING IN 
VIVO PASSAGE IN CHICKENS 112
3.2.2. AMPLIFICATION AND FRAGMENT ANALYSIS OF REPEAT TRACT LENGTHS OF SIX PHASE 
VARIABLE LOCI 113
3.2.3. DETECTION OF A HIGH FREQUENCY OF OFF-TO-ON SWITCHING IN Cj0031 118
3.2.4. SWITCHING FROM OFF TO ON DETECTED FOR Cj0045 DURING IN VIVO PASSAGE: 119
3.2.5. BIRD-TO-BIRD VARIATIONS DETECTED IN FREQUENCY OF ON/OFF VARIANTS OF Cj0628 
(CAPA): 120
3.2.6. SWITCHING FROM OFF TO ON WAS OBSERVED FOR Cj0685 (CIPAA) DURING IN VIVO 
PASSAGE: 121
3.2.7. HIGH FREQUENCY OF ON-TO-OFF SWITCHING OF Cj1139 (WLAN) DETECTED AFTER 
PASSAGE: 123
3.2.8. MINOR CHANGES IN REPEAT TRACT LENGTH OF GENE Cj1326: 124
3.2.9. ASSESSMENT OF THE VALIDITY OF GENEScan ANALYSIS OF REPEAT TRACT LENGTH 
THROUGH SEQUENCING: 125
3.2.10. SIGNIFICANT CHANGES IN ON/OFF GENOTYPES DETECTED DURING IN VIVO PASSAGE: 126
3.2.11. DIFFERENCES IN THE DISTRIBUTION OF THE COMBINED GENOTYPES FOR ALL SIX 
GENES AMONG BIRDS: 129
CHAPTER 4. INVESTIGATION OF METHYLATION ACTIVITY OF CJ0031 METHYLTRANSFERASE

4.1: INTRODUCTION: ________________________________________________________ 165
4.2: RESULTS: ______________________________________________________________ 168
   4.2.1: BIOINFORMATICS ANALYSIS OF CJ0031 R/M SYSTEM: ________________________ 168
   4.2.2: INVESTIGATION OF THE METHYLATION ACTIVITY OF CJ0031: _____________ 177
   4.3.2: DETERMINATION OF TRACT LENGTH IN CJ0031 IN THE SAMPLES: __________ 178
   4.2.3: IDENTIFICATION OF RECOGNITION SITE OF CJ0031 METHYLTRANSFERASE: ______ 193
   4.2.5: LOSS OF THE METHYLATION PHENOTYPE BY MUTATION OF THE CJ0031 METHYLTRANSFERASE: ________________________ 208
   4.2.6: POTENTIAL RECOGNITION SITE FOR THE CJ0031 METHYLTRANSFERASE: ______ 209
   4.2.7: DISCUSSION: _________________________________________________________ 214

CHAPTER 5. PHENOTYPIC ASSAYS FOR CJ0031 METHYLTRANSFERASE  ___________ 224

5.1: INTRODUCTION: ________________________________________________________ 224
5.1.2: ADHESION AND INVASION: _____________________________________________ 225
5.1.3: MOTILITY: __________________________________________________________________ 227
5.1.4: BIOFILM FORMATION PHENOTYPE: _______________________________________ 228
5.2: RESULTS: ______________________________________________________________ 231
   5.2.1: NATURAL TRANSFORMATION OF C. JEJUNI CHICKEN STRAIN WITH GENOMIC DNA OF C. JEJUNI LAB STRAIN: _______________ 231
   5.2.2: INVESTIGATION OF THE IMPACT OF THE CJ0031 METHYLTRANSFERASE ON THE ADHESION AND INVASION PHENOTYPES OF C. JEJUNI: ____________________________________________ 231
   5.2.3: INVESTIGATION OF THE IMPACT OF THE CJ0031 METHYLTRANSFERASE ON BIOFILM FORMATION PHENOTYPE BY C. JEJUNI: ________________________________________ 237
   5.2.4: INVESTIGATION OF THE IMPACT OF MUTATION OF THE CJ0031 METHYLTRANSFERASE ON MOTILITY OF C. JEJUNI: ____________________________________________ 239
5.3: DISCUSSION: ___________________________________________________________ 241
   5.3.1: EVIDENCE FOR A PHASEVARION FROM MOTILITY, ADHESION AND INVASION EXPERIMENTS ___________ 241
   5.3.2 EVIDENCE FOR A PHASEVARION FROM BIOFILM EXPERIMENTS ________________ 242
   5.3.3: EVIDENCE FOR A PHASEVARION FROM IN VIVO EXPERIMENTS _______________ 245
   5.3.4: PUTATIVE MECHANISM FOR A ROLE OF CJ0031 IN REGULATION OF OTHER GENES 246
Chapter 1. INTRODUCTION

1.1: Adaptive behaviour of prokaryotes:

Bacteria encounter immense pressures from their environments and hosts. The pressures faced by bacterial populations in their natural environment may include oxidative stress and starvation, heat and cold shocks, damages to DNA during exposure to UV radiation and antibiotics and predation by bacteriophages etc. Many disease causing bacteria are restrained by constantly evolving hosts showing varieties in their behavioural patterns and defense mechanisms such as immunological responses. The number of microflora residing in the gut of human beings is tenfold higher that of human cells (Berg, 1996). The competition between the microflora of the gut of a host and the invading bacterial population offers a severe challenge to survival and adaptability of pathogens along with genetic individuality and the variability within the host populations. These challenges from both environment and host elicit adaptive responses from the pathogens and commensals which help to maintain their fitness in heterogeneous environments (Moxon et al., 1994).

Bacterial populations have programmed responses for coping with threats to their survival and adaptation. This programmed response system recognizes the threats, produces the particular response to eliminate these stresses and ultimately allows the microorganisms to adapt to the environmental stress. Some of these responses are outcomes of mutagenesis within the stress related genes, which is often classified as stress-induced mutagenesis (Galhardo et al., 2007; Foster et al., 2007). Nevertheless, the programmed stress response system is inadequate for the adaptability and survival
of bacterial populations as a response to the large number and variety of these stresses produced by host and natural environments. Besides, some of these stresses are undetectable and unpredictable in nature. The evolution and maintenance of a stress response for each and every stress goes beyond the capability of bacteria, especially considering the energy and cost requirements for a bacterial population (Jayaraman, 2011).

The evolution of genetic strategies to cope with the environmental and host-enabled challenges may be a better option for a bacterial population in certain situations. These strategies include the employment of mutations and horizontal transfer of genetic material and are commonly used to introduce heterogeneity into the bacterial population. Spontaneous mutations are prevalent in bacterial populations occurring at a frequency of $10^{-9}$, indicating that this mutational process may be too slow to match with the speed of changes taking place in some internal and external environments of host and hence for survival in and adaptation to many constantly evolving environments. Should the mutation rate be increased, then it is possible to produce the mutations at multiple loci, favouring enhanced diversity at a population level and making it possible to cope with threats and challenges more efficiently. This is achieved by hypermutators which generate mutations at multiple loci concomitantly; and some of them combat efficiently with environmental stresses and accelerate the process of adaptation to them. Due to the high probability of survival and adaptation presented by hypermutators to the bacterial population, many pathogenic species contain high proportions of hypermutators in their clinical isolates such as *Neisseria menigitidis*, *Haemophilus influenzae*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Salmonella* etc (de Visser, 2000; Oliver, 2005; Labat *et al.*, 2005; Denamur and Matic, 2006).
However only a few out of several mutations generated by hypermutators turn out to be beneficial, whilst the rest of them are accumulated as neutral or deleterious mutations. That is why, hypermutability can have the status of a great asset for bacterial populations but is also reckoned as a great liability for them due to its intrinsic tendency of promoting the accumulation of deleterious mutations. It has been reported that rates of beneficial mutations in *E. coli* are in the range of $4 \times 10^{-9}$ mutations/cell/division (Kibota *et al.*, 1996; Boe *et al.*, 2000), in comparison rates of deleterious mutations are in the range of $4 \times 10^{-4}$ mutations/cell/division (Imhoff *et al.*, 2001). This indicates that hypermutators have a higher rate of deleterious mutations compared to that of beneficial mutations. Despite its drawbacks, hypermutability is a common phenomenon among several species of bacterial kingdom, indicating that advantages accrued through it can outweigh the liabilities caused by it in some situations (Jayaraman, 2011).

This discussion highlights how a mutational process causing genetic diversity without inheritance of associated risks would be ideal for bacteria. This would help bacteria to both adapt and to get rid of the burden of deleterious mutations. One way to achieve this ideal is the stress induced transient hypermutation process and the second is the localised hypermutation process occurring at only selected loci termed as contingency loci/genes (Wright, 2004; Roth *et al.*, 2006; Galhardo *et al.*, 2009). The hypermutation process at localised loci has the ability to shift between two phases, ON and OFF, without producing a risk of deleterious mutations. This process is called phase variation or antigenic variation.
1.2: Phase and Antigenic Variation

The term “phase variation” was originally coined to illustrate the switching patterns of antigenic flagella in *Salmonella* (Andrewes, 1922). The term “phase variation” is often used to describe the reversible mutation process shifting between the ON and OFF expression of contingency genes. Phase variation has three important signatures which are its reversibility, a high frequency of ON/OFF switching, and an association with genes encoding surface structures such as capsule, LPS, LOS, iron acquisition proteins, pili/fimbriae, flagella, excreted enzymes etc. However, an association of phase variation with non-surface coding genes has also been reported, for example, phase variation in restriction modification genes of several pathogenic bacteria (Srikhanta *et al*., 2005, Srikhanta *et al*., 2010, Srikhanta *et al*., 2012).

Antigenic variation denotes changes in the expression of functionally conserved and antigenically distinct moieties within a clonal population. Here it is important to differentiate antigenic variation from phase variation. Woude and Baumler (2004) described the three important criteria which must be utilized in order to declare any variation as antigenic variation. Firstly the variation must be implicated in the evasion of immune responses, secondly it must be multiphasic and thirdly it must result from the mechanism of gene conversion. The term phase variation is broadly applied to encompass the biphasic switching of particular antigen (phase variation) and multiple switching of expression of different forms of a particular antigen (Deitch *et al*., 2009; Bayliss, 2009).

The host immune system applies many pressures on antigenic characteristics of pathogens, normally based on the surface of pathogen, directly exposed to host
environment. Against the backdrop of this ever changing portfolio of immune system, the pathogens manoeuvre to escape these pressures by alterations in the antigenic moieties in the subsequent generations (Brunham et al., 1993). A host of such adaptive strategies, found in the pathogenic world and created by a range of molecular mechanisms, have caused the evolution of genetic variation at individual loci called “contingency loci” (Moxon et al., 1994). This term is used in order to differentiate contingency loci from “house-keeping” genes. These contingency loci are responsible for the antigenic and phase variation referred to above and described in detail below. The majority of contingency loci are characterized by the presence of simple sequence repeats (SSR). Thus these genes were called simple sequence contingency loci (SSCL) [Bayliss et al., 2001]. SSCL undergo phase variation due to mutations in their simple sequence repeats through slipped strand mispairing or some other mechanism, the details of which will be described in section 1.4.1. Phase variation at contingency loci mediates biphasic changes with a high frequency and an order of magnitude in the range of $10^{-4}$-$10^{-3}$ mutations/cell/generation in either direction. Hence, the frequency of phase variation in either direction is many folds higher than that of spontaneous mutational frequency. Phase variation helps with generating genetic diversity within bacterial clonal population of pathogens, commensals and symbionts. It facilitates the survival of microorganisms in fluctuating environments, evasion of immune responses in the host environment, niche adaptation, and enhancement of virulence characteristics (Wisniewski-Dyke and Vial, 2008). The key features of phase variation have been illustrated in figure 1.
Figure 1: The role of phase variation during adaptation of organisms to a hostile

A mixed population (ON and OFF phenotype) of bacteria switched from a completely OFF phenotype during growth in non-selective conditions. The ON variants from this original population are positively selected during propagation in the host due to their adherence and invasion properties. After invasion, the population multiplies in the host resulting in the production of ON and OFF phase variants. The host immune system is activated against the ON variants. So the selective forces drive a switch from ON to OFF variants. The OFF variants, under non-selective conditions, replicate to regenerate a mixed population. Thus natural selection always selects the fittest phenotype for survival in the host.

The events of phase variation are normally reversible: ON↔OFF, and stochastic giving rise to heterogeneous populations of microbes without any dependence on sensing
mechanisms or environmental signals. The PV events are termed stochastic because the timing of these events and cells of a bacterial population in which these events occur cannot be predicted (Moxon et al., 2006). However, these events may be the result of programmed coding in the genome of organism, which is considered as a highly organised entity in the organism. So PV events may occur without any reliance on the theory of probability what does this mean? can you explain it? (Robertson and Meyer, 1992).

Phase variation has evolved to confer short term and selective adaptive advantages to the founder population and its subsequent generations. This is an adaptive process by which the pathogenic bacteria undergo frequent and reversible phenotypic changes originating from genetic alterations at particular loci of their genomes. The principle target of the phase variation phenomenon, which looks like a multi-cellular adaptive strategy, is the production of genetic diversity, leading into the selection of the fittest variants and the subsequent expansion of the selected individuals (Hallet, 2001). This genetic diversity resulting from changes at contingency loci presents the host with constantly evolving population of invading microbes, whose elimination is difficult or impossible for the host.

Phase variation has the ability to produce a multitude of phase variants. Genes regulated by phase variation undergo switching from ON-to-OFF or OFF-to-ON independently. One gene can produce two possible phenotypes, ON or OFF, through phase variation. Similarly two phase variable genes can contribute 4 possible phenotypic combinations to the growing population. If there are “N” numbers of phase variable genes, the potential number of phenotypes arising from them as a result of phase variation can be $2^N$. Suppose there are as few as 7 phase variable genes in the genome of a pathogen, they can generate 128 different phenotypes. This indicates that PV has the potential to
make a major contribution to the genetic diversity of pathogenic microbes having contingency loci as a means to adapt to their changing environments.

Figure 2 Number of phase variable genes and number of phenotypes resulting from phase variation event.

The red and green bars represent the OFF and ON phases of phase variable genes, Double arrow symbols denote the reversible process of phase variation, black arrow before each gene shows the promoter region of each gene. Blocks A, B and C show the number of phenotypes resulting from phase variation events occurring at a single gene, two genes together and three genes together, respectively. The total number of
phenotypes calculated as \(2^N\). For one phase variable gene, \(2^1 = 2\), for two phase variable genes, \(2^2 = 4\), for 3 phase variable genes, \(2^3 = 8\).

1.3: Phase variable phenotypes in pathogens

Phase variation has great importance in the resistance of pathogens from the host’s immune system. As the surface exposed structures are in direct contact with all the attacks targeted to kill the bacteria by the host, many of the structures and moieties, made of proteins and carbohydrates, present on the bacterial surface have evolved the strategy of phase variation to control their expression and to facilitate escape from the continuous attacks from the host. Some examples of the phase variation of surface structures are discussed in the following sections.

1.3.1: Capsule

Capsule constitutes an important virulence determinant in many pathogenic bacteria. Capsule undergoes phase variation in both gram-positive and gram-negative bacterial species, including *Campylobacter jejuni* (Bacon et al, 2001), *Streptococcus pneumoniae* (Wait et al., 2003), and specific sero-groups of *N. meningitidis* (Hammerschmidt et al., 1996). Vogel and Frosch (1999) reported that phase variable expression of capsule in vitro resulted in resistance against the phagocytosis and killing of meningococcal cells through complement pathways. Srivastava et al. (2009) offered evidence that phase variable expression of capsular polysaccharides is required for *Vibrio vulnificus* infections in oysters. They reported that CPS phase variation mediated environmental survival and induction of virulence phenotypes. Phase variation in
capsular polysaccharides also interfered with invasion of host cells by *N. meningitidis* (Hammerschmidt *et al*., 1996; Nassif, 1999). A mononucleotide repeat tract is found in the *siaD* gene, and is responsible for switching this gene between ON and OFF phase. The *siaD* gene is involved in the biosynthesis of the capsule of serogroup B *N. meningitidis* strains (Hammerschmidt *et al*., 1996). Capsule was found absent in meningococcal carriage strains which formed a significant proportion of the population residing in the upper respiratory tract (Cartwright *et al*., 1995). Thus capsular phase variation may have significance for survival and adaptation of bacterial cells to host defenses such as opsonophagocytosis and complement.

### 1.3.2: Fimbriae and Pili

Fimbriae or pili are proteinaceous structures which mediate the adhesion of pathogenic bacteria to host tissues through specific receptors found on the surface of host cells (Loessner *et al*., 2002). The phase variation of fimbriae occurs in the promoter region of the fimbrial operon, resulting in a change at the transcriptional level, and leads into the ON/OFF expression of genes in the fimbrial operon (Loessner *et al*., 2002).

The fimbriae of *H. influenzae*, encoded by *hifA* and *hifB*, undergoes phase variation by changes in length of repetitive TA motifs clustering in the promoter region of these genes (Marieke van Ham *et al*., 1993). Changes in expression of Fimbriae of uropathogenic *E. coli* are mediated by epigenetic mechanism in the promoter region of the *pap* operon (Holden *et al*., 2007). In some species, Fimbriae undergo antigenic phase variation, though multiphasic phase variation. Thus in the genome of *Salmonella enterica* serotype *Typhimurium*, phase variable expression has been established for *pef, lpf, and fim* (Neyrolles, 1999, Norris, 1998).
1.3.3: Flagella

Flagella are considered important determinants of motility, adhesion and virulence in several pathogenic strains of bacteria (van der Woude and Baulmer, 2004). Flagella are recognized by Toll-like receptors of the innate immune system, indicating their role in virulence and pathogenesis (Liaudet et al., 2002). The presence of flagella as an antigenic property is employed to classify many bacterial isolates serologically. Though this indicates that the presence of flagella is assumed to be an invariant property, in fact this phenotype undergoes phase variation in some clonal populations. Antigenic variation of flagella was reported by Andrewes (1922) in *S. enterica* serotype *Typhimurium* (Censini *et al.*, 1996). The phase variation in *fljC* and *fljB* encoding flagellar subunits caused the emergence of two phenotypes H1 and H2, which is biphasic antigenic variation (O'Toole and Wolf-Watz, 1996). Jack *et al.* (2001) provided empirical evidence that the phase variable expression of the *fljB* and *fljC* has a selective advantage for adaptation of *S. enterica* serovar Typhimurium in the mouse model. *C. jejuni* and *Helicobacter pylori* are devoid of a fimbrial system to facilitate adhesion with host tissues. So the flagella are major adherence determinants which mediate the interaction of bacterial cells to the host tissues (Jeremy *et al.*, 2011). Flagellar mutants of *H. pylori* (Eaton, 1996), *C. jejuni* (Nachamkin, 1993) and *Campylobacter coli* (Pavlovskis, 1991) are less virulent than the wild type strains. The phase variants of *C. jejuni* having mutations in flagellar motility were unable to invade the chicken caecum properly during a chicken infection, indicating the requirement of flagellar motility for colonisation (Hendrixson and DiRita, 2004; Hendrixson, 2006; Wassenaar *et al.*, 1993; Wosten *et al.*, 2004). Nuijtan *et al.* (1995) showed that the *flaA* gene encoding the flagella subunit undergoes phase variation at the transcriptional level in *C. jejuni*. 
1.3.4: LPS and LOS Modifications

LPS plays an important role by affecting the serum sensitivity and adhesion properties of pathogenic microbes (Marjan et al., 2004). Lipopolysaccharides are regarded as the main constituent of the outer membrane of Gram-negative bacteria. The LPS core lacking the O-linked saccharide subunits is referred to as Lipooligosacchride LOS (Preston et al., 1996). LPS contains a lipid moiety acting as a powerful stimulant of the immune system in the host, so it is also termed as endotoxin (Beutler, 2002). Phase variation, antigenic mimicry and sialylation are the key signatures of LOS in many pathogens (Preston et al., 1996).

LPS structures in *H. pylori* are post modified with carbohydrate moieties by phase variable enzymes. These phase variable carbohydrate modifications have mimicry with structures of the Lewis group of antigens of human blood groups. The *fuctT* genes (*futA, futB, and futC*) undergo phase variation and are involved in the LPS modifications (Wang et al., 2000).

*H. influenzae* contains five LPS biosynthetic genes carrying simple sequence repeat tracts. These genes undergo phase variation, resulting in the production of 32 genotypically different antigenic variants of LPS. Nevertheless, given the complexity of LPS and its presentation of several epitopes to escape from the immune system, not all of the antigenic variants are equally immunogenic (Bayliss et al., 2001). The *licl* locus contains 4 genes essential for adding ChoP (Phosphorylcholine) to LPS. The phase variation in *liclA* present in that locus results in the presence or absence of ChoP on LPS, which has a dramatic influence on the evasion of bacterial cells from killing by antimicrobial peptides in the upper respiratory tract. The presence of Chop on LPS confers resistance against an antimicrobial peptide but it increases the sensitivity to C-
reactive protein mediated killing (Lysenko et al., 2000a). The isolates recovered from respiratory tract of mice after infection showed a higher proportion of lic1A in frame, indicating its importance for the adaptation of *H. influenzae* to the mucosal surfaces. Thus the phase variation in lic1A increases the colonization potential at one place and resistance to immune responses at another place of the host (Lysenko et al, 2000b).

Similarly, the simple sequence contingency loci encode enzymes required for post-modification of LPS in *N*. Four genes out of six genes (*igt*A, *igt*C, *igt*D and *igt*G) encoding LPS modifiers are phase variable, which can generate a repertoire of 12 different immunotypes by switching between ON and OFF phenotypes. The sialylated modification of these immunotypes not only prevents the killing mediated by phagocytosis and complement system but it also plays an important part in Opa/Opc mediated adhesion to and invasion of host tissues (van Putten and Robertson, 1995). This means that *N. meningitidis* strains can use various combinations of these immunotypes and their sialylated derivatives under particular environments in order to evade the immune responses. The non-sialylated immunotypes are normally characteristic of carriage strains while the sialylated immunotypes have the potential to be part of the repertoire of phenotypes important in invasive strains (Bayliss, 2001)

### 1.3.5: DNA Restriction-Modification Systems

DNA restriction modification systems are considered as very important tools in the recognition and subsequent destruction of unmodified foreign DNA. bacteria have evolved RM systems to safeguard themselves against the attack of phages (Marjan et al., 2004). The methylation subunits of RM systems and other orphan methylases (e.g. Dam) are thought to perform many functions apart from the defense of bacterial cells
from bacteriophages, which have been summarised by Marinus and Casadesus (2009) [figure 3]. You would be better making your own figure as this one is not very clear.

**Figure 3 Role of Adenine Methylation in Host Pathogenic Interaction**

The figure displays the overview of key roles of adenine specific (N6) MTase. The two main methylation systems possessed by bacteria are Dam-dependent methylation system and restriction modification system. Restriction modification system involved in host defence and epigenetic control of unlinked genes (details can be found in text). The Dam dependent methylation system control cellular processes like replication, mismatch repair, phage packaging, nucleotide structure, phase variation, conjugation. The proteins affected by Dam methylation are highlighted red and put in brackets along each process.
A phase variable R/M system was first discovered in the rodent pathogen, *Mycoplasma pulmonis*. The chromosome of *M. pulmonis* possesses two *hsd* loci, each encodes a separate type I RM system (*hsdM, hsdR, hsdS* units). These loci undergo recombinatorial rearrangement by an inversion event relative to their promoters, resulting in generation of ON/OFF phase variants (Dybvig *et al.*, 1998; Dybvig and Yu, 1994). Phase variants of *hsdS* from these loci have different sequence specificity, indicating their importance for methylation at different sequences under different circumstances (Dybvig, 1998).

It is evident from the complete genome and sequence analysis of many species that R/M systems are subjected to phase variation in a variety of species (De Bolle *et al.*, 2000, Sauer *et al.*, 2000). The summary of phase variable R/M loci in various pathogenic bacteria can be viewed (Table 1).

Table 1: Phase variable R/M loci present in genome of pathogenic bacteria
<table>
<thead>
<tr>
<th>R/M class</th>
<th>Organism</th>
<th>Locus</th>
<th>SSR location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td><em>Helicobacter pylori</em></td>
<td>hP0464</td>
<td>Within ORF</td>
<td>Guo and Mrazek, 2008</td>
</tr>
<tr>
<td></td>
<td><em>H. influenzae</em></td>
<td>hind1 (HsdM)</td>
<td>Within ORF</td>
<td>Zaleski <em>et al.</em>, 2005</td>
</tr>
<tr>
<td></td>
<td><em>Mycoplasma pulonalis</em></td>
<td>hsd1</td>
<td>Within ORF</td>
<td>Dybvig and Yu, 1994</td>
</tr>
<tr>
<td>Type II</td>
<td><em>H. pylori</em></td>
<td>hP1471</td>
<td>Within ORF</td>
<td>Salaun <em>et al.</em>, 2004</td>
</tr>
<tr>
<td></td>
<td><em>C. jejuni</em></td>
<td>cj0031</td>
<td>Within ORF</td>
<td>Parkhill <em>et al.</em>, 2000</td>
</tr>
<tr>
<td>Type III</td>
<td><em>H. pylori</em></td>
<td>hP1521 (res)</td>
<td>Promoter</td>
<td>Guo and Mrazek, 2008</td>
</tr>
<tr>
<td></td>
<td><em>H. influenzae</em></td>
<td>Mod</td>
<td>Within ORF</td>
<td>Hood <em>et al.</em>, 1996</td>
</tr>
<tr>
<td></td>
<td><em>Pasteurella hemolytica</em></td>
<td>Mod</td>
<td>Promoter</td>
<td>Ryan and Lo, 1999</td>
</tr>
</tbody>
</table>

The confirmation of phase variability of R/M system comes from the *mod* gene in *H. influenzae*. (De Bolle *et al.*, 2000). A *mod* gene contained 40 tetranucleotide (5'-AGTC) repeats within its open reading frame. They had shown that *mod* phase varied at high frequency ($10^{-3}$) by using an in-frame translational *Lac-Z* reporter construct. Bayliss *et
al. (2006) observed that the same phase variable mod type III RM system was responsible for high allelic diversity within natural population of strains of *H. influenzae*. The central portion of mod gene is subjected to variations, which is likely to encode TRD (target recognition domain). The variations in TRD of mod type III RM system provide a logical explanation for the differences in the recognition sequences of these enzymes. Moreover, the variations in the number of repeats were also found in this region, suggesting that diversifying selection is operating on this locus.

*Streptomyces coelicolor*, a soil bacterium, inhibits the growth of φC31 phage by switching its DNA methyltransferase (PglX) into an ON phase under the stressful conditions of bacteriophage attack (Struve and Krogfelt, 1999). Sumby and Smith (2003) further discovered that ON state of PglX also turns on the phase-variable phage growth limitation (Pgl) system of *S. coelicolor*, which offers resistance to propagation of phages on that strain. This suggests that phase variation in type III R/M system (PglX) interfere with pgl phenotype.

Similarly, phase variable type I RM system in *H. influenzae* also provides protection to the host against the bacteriophage attack. The *hsdM* gene encoding Hind1-a type I R/M system undergoes phase variation due to presence of pentanucleotide repeats within its ORF; and this is active against phage HPIc1. A 4 repeats of pentanucleotide in *hsdM* ORF led to the translation of active type I methyltransferase and were associated with phage resistance. However, 3/5 repeats caused production of an inactive type I methyltransferase, and were associated with phage sensitivity (Zaleski et al., 2005). The *H. influenzae* become susceptible to phages on changes in LPS structure or phase variation in Hind1 type I RM system. However, the exact role of type I RM system phase variation is yet to be revealed in *H. influenzae* (Zaleski et al., 2005).
1.3.6. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR):

Bacteria have acquired various methods to fight against the viruses and to avoid the stresses resulting from infection of phages and plasmids. These mechanisms involve the restriction modification systems, evolution of cell surface receptors or mutation in these receptors necessary for phage infection, abortive infections and CRISPR systems (Labrie et al., 2010; Petty et al., 2007). CRISPR system is a special mechanism which is heritable and passes the genetic memory of phage infections down to next generations, thus equipping the bacteria with “adaptive immune system”. CRISPR employs the RNAs (non-coding) in association with Cas proteins for defense against the phage infection. The resistance acquired via CRISPR includes the incorporation of a part of foreign genome invading the bacteria into repetitive DNA sequences called CRISPR arrays. The incorporated foreign DNA constitutes new spacer sequences within CRISPR arrays. The CRISPR arrays undergo transcription, resulting in the generation of small RNAs termed as crRNAs which are further processed into mature crRNAs with help of Cas proteins. Finally, these mature crRNAs complex with Cas proteins and target the invading the phages or plasmids by interfering with foreign DNA in sequence specific manner (Marraffini and Sontheimer, 2010; Terns and Terns, 2011, Richter et al., 2012) (Figure 3.1)
Figure 3.1. Overview of Mechanism of CRISPR/Crispr associated adaptive immunity (Source: Richter et al., 2012)

A. shows the CRISPR arrays with repetitive DNA sequences along with cas gene cluster, B. shows the incorporation of part of foreign DNA as new spacer into CRISPR arrays, and transcription and processing of crRNAs, C. displays the targeting action of crRNAs and Cas proteins complex on phage/plasmid DNA.

1.4: Molecular mechanisms of phase variation

Changes in DNA sequence can contribute to the phenotypic changes. They can alter the expression of a gene from ON to OFF or OFF to ON. These changes can range from a change of single nucleotide, in the case of SSM, to rearrangements of fragments of up to several kilobases in the case of recombination. Phenotypic variation resulting from DNA rearrangement, regardless of the molecular mechanisms responsible for them, has
been termed as a shufflon (Komano, 1999). The following sections describe the major mechanisms of phase variation.

1.4.1: **Short sequence repeats and slipped-strand mispairing**

The repeat units within SSCL are subject to frequent contraction and expansion by the deletion or addition of repeat units, resulting in phase variation in the expression of the associated gene (Marjan and Baulmer, 2004). The mechanism operating behind these phase variation events is likely to be slipped strand mispairing (SSM) [Bayliss, 2009]. Several bacterial species employ SSM mediated phase variation as a strategy to manipulate their population diversity within hostile environments (van der Broek *et al*., 2005). SSM involves the misalignment or mispairing between the daughter and the template DNA strands on the leading or lagging strand at reiterated tracts during the process of replication (Henderson *et al*., 1999). During replication, the transient separation between the nascent and template strands takes place (figure 4). During reannealing, the daughter strand can slip forward, resulting in the creation of a bulge on the template strand and subsequent contraction of repeat tract length by deletion of a repeat unit. If the backward slippage of the nascent strand occurs, it results in the formation a bulge on the nascent strand and subsequent expansion of a repeat tract length by addition of a repeat unit (Hallet, 2001; Moxon *et al*., 2006; Lerouge and Vanderleyden, 2002).
Figure 4 Slipped Strand Mispairing Mechanism

(1) The repeat tract G9 present is in the ORF of a phase variable gene; (2) the replication fork reaches the polyG tract and slippage occurs; (3) backward slippage causes a bulge to appear on the nascent strand (4) and forward slippage results in a bulge on the template strand; (5) the backward slippage causes an increase in length of polyG by one nucleotide; (6) however, the forward slippage causes the shortening of the repeat tract by one nucleotide.

The regions susceptible to slipped strand mispairing contain short, contiguous homogenous or heterogeneous repeat tracts of repetitive DNA sequence in the range of 1 to 7 bp. These reiterated sequences are designated as short sequence repeats (SSR) or microsatellites (van Belkum et al., 1998; Bayliss et al., 2006).

A change in the number of these repeat units can bring about an alteration of gene expression at the transcriptional level or translational level, depending on the location of
these short sequence repeats within a gene. The repeats can be found within the reading frame of contingency loci. The repeats found towards the 5’ end of the contingency gene can phase vary to produce multiple expression states of a gene while the repeats located towards the end of phase variable genes can undergo phase variation to produce altered function of the relevant gene (Bayliss et al., 2009). Thus the repeats located within an ORF of a gene can control gene expression at the translational level by alterations in their repeat units, causing a frameshift mutation. When the reiterated sequences are present in the coding region of a gene, then SSM can mediate changes in the translation of proteins by alerting the number of repeat units that are not a multiple of three, and thereby, disrupting the open reading frame. In this way, a truncated or non functional protein is synthesized. For example, the mod gene of H. influenzae undergoes phase variation in its expression by SSM mediated phase variation events within the tetranucleotide repeat units (5’-AGTC-3’) in its coding sequence (De Bolle et al., 2000). There are 40 tetranucleotide repeat units (5’-AGTC-3’) in the reading frame of the mod gene, which undergoes phase variation at a high rate. Mutation rates were determined for repeat tract lengths of 17-38 repeats and rates were found to increase linearly with increase of tract length, suggesting that repeat tract length is an important determinant of mutation rate for genes containing repeats.

However, the repeats may be located within a promoter or upstream of the promoter. If the repeats are found within the promoter region between the -10 and -35 sequences, a binding site of RNA polymerase, then the modulation of gene expression occurs at the transcriptional level. In contrast to translational control, phase variation of transcriptional control can lead to generation of varying levels of expression from low to high or otherwise. The phase variation events at the repeats associated with a promoter region are likely to change the binding capacity of RNA polymerase. For
instance, phase variation in *hifA* and *hifB* responsible for encoding fimbrial proteins and chaperons respectively in *H. influenzae* is mediated by alterations in dinucleotide (TA) repeat units (9, 10, 11) located between the -10 and -35 sequences. The optimal expression is produced when 10 repeat units (ON phase) are located between the -10 and -35 sequences, which results in the optimal spacing of 16bp between the -35 and -10 region of the *hifA* promoter, while 9 or 11 TA repeat units (OFF phases) cause the suboptimal spacing of 14bp or 18bp respectively between the -35 and -10 region in OFF variants of *hifA* (Van Ham *et al.*, 1993; Torres-Cruz and Woude, 2003; Woude and Baulmer, 2004). The *porA* gene in *N. meningitidis*, a vaccine candidate, contains a polyG repeat tract in the promoter region between the -35 and -10 sequences. An alteration in the repeats of *porA* caused by slipped strand mispairing affects promoter strength and polymerase binding to the promoter. G11, G10 and G9 tract lengths cause high, low and no expression, respectively (Van der Ende *et al.*, 1995; Van der Ende *et al.*, 2000). The *fetA* gene, a gonococcal ferric enterobactin receptor, undergoes phase variation between high expression and low expression due to alterations in the polyC repeat tract located between the -35 and -10 sequences (Carson *et al.*, 2000).

The repeats located upstream of the promoter region of a gene have also the potential to regulate gene expression at the transcriptional level by modulating the interaction between transcriptional regulatory proteins and RNA polymerase (Bayliss and Moxon, 2005). An example of such phase variation events has been found in the case of *nadA* expression. The tetranucleotide repeats (TAAA) found within the upstream region of the *nadA* promoter undergoes phase variation through slipped strand mispairing with a high frequency (4.4 x 10^{-4}) [Martin *et al.*, 2003], resulting in the modulation of promoter strength of *nadA*. The repeat units, 13, 10, or 8, induced the high expression level of *nadA* as opposed to 11 and 12 causing the medium expression. Nine repeat
units was responsible for induction of low expression of nadA (Martin et al., 2003). Thus phase variation in the repeats of nadA yields promoters with high, medium and low activities. Changes in the repeat motifs located upstream of the nadA promoter affect the binding of IHF protein to the promoter in vitro (Martin et al., 2005). However, Metruccio et al. (2009) showed that alterations in the repeat tracts change the affinity of IHF and a repressor, nadR, to the operators in the nadA promoter, leading to alterations in regulation of nadA expression. They also showed that a small molecule present in saliva interacts with NadR in the repression of nadA expression.

Another classical example is the modulation of expression of fimbrial genes in Bordetella by simple sequence repeats located upstream of the promoter. The main feature of the SSRs of these fim genes (fim2, fim3, fimX) is that the repeats are overlapping with the -35 sequences of the promoter. The overlapping region carries the binding site for the BvgA activator. So, the alteration in the polyC repeat tract of fim genes optimizes the interaction between the activator (bvga) and RNA polymerase (RNAP). The optimal length of C-stretch is 14 base pairs of which 10 bp overlaps with the -35 region [Chen et al., 2010].

The regulation of gene expression by the alteration in repeat motifs located downstream of the promoter, just before the transcriptional start site, has also been reported by Laity et al. (1993) in Moraxella catarrhalis. They have concluded that changes in the G-stretch located between the -10 and translational start site mediate the differential expression of the uspA gene, an adhesion protein in M. catarrhalis.

1.4.2: Homologous (general) recombination
Homologous or general recombination occurs at long stretches of homology (>50bps), and is accomplished by proteins involved in the DNA repair and maintenance machinery of the cell. This process requires the presence of two copies of a gene (i.e. two alleles) or two regions of DNA sharing considerable homology with each other. The recombination event happens between a transcriptionally active gene and one of its silent alleles by involving a unidirectional exchange of parts of homology, resulting in the formation of recombinant DNA. This is termed as gene conversion and it is associated with antigenic variation in many bacterial species (van Leeuwenhoek, 2008). If numerous copies of silent alleles of a gene exist in the genome, theoretically several recombinant molecules (antigens) can be produced, leading to generation of heterogeneity within bacterial population. The recombination mediated gene conversion can be distinguished from recA dependent recombination due to occurrence of the former with a higher frequency than is usually associated with the latter.

There is no common mechanism for gene conversion (Wisniewski-Dyé and Vial, 2008). However, it requires the machinery of homologous recombination (Marjan et al., 2004). Antigenic variation occurring in type IV pilin in Neisseria gonorrhoeae is a well documented example of recA-independent recombination. There are one or two functionally active loci of pilE gene in N. gonorrhoeae and one to six copies of transcriptionally inactive pilS alleles. These pilS loci are promoterless, containing the 5’ end of the gene and variable sequences. Each copy of a silent pilS has the potential to exchange its variable sequences with the pilE locus through gene conversion which utilises the proteins from the machinery of homologous recombination.

The gene conversion event is composed of two steps (Figure. 5). In the first step, the recombination of pilE and pilS happens with very short sequences of homology in the donor chromosome, yielding the intermediate hybrid recombinant molecule pilE-pilS
which is an extra-chromosomal structure (Howell-Adams and Seifert, 2000). This step is an example of recA-independent recombination. In the second step, the hybrid pilE-pilS molecule approaches to a recipient chromosome and transfers the pilS sequences to the pilE locus in the recipient chromosome. The second step is mediated by recA, recF and recX proteins which precipitate the recombination events at larger and shorter portions of homology of pilE locus (Kline et al., 2003; van der Woude and Balumer, 2004). Thus the homologous recombination mediated gene conversion causes the antigenic variation at pil locus by altering pilE loci whereas keeping pilS loci unaltered. The gene-conversion mediated antigenic variation in gonococcal pilin occurs at a high frequency during the human infection cycle, which is 0.13 recombination events per cell (Criss et al., 2005). These recombination events are triggered by the iron starvation conditions, suggesting that antigenic diversity is produced in gonococcal pili for survival and adaptation to the host environment (Serkin and Seifert, 2000).

The homologous mediated recombination event causing the antigenic variation can also lead to generation of ON/OFF states of pili (Figure 5). This happens when the pilS sequences containing premature stop codons are exchanged with pilE locus undirectionally, resulting in generation of a non-functional pilin gene or truncated protein product. However, the deletion of pilE may take place as a result of recombination events, generating the irreversible OFF phenotype for gonococcal pili (Manning et al., 1991; Segal et al., 1985).

Homologous recombination can also mediate DNA inversion or nested DNA rearrangements and gene duplication. One of the best examples of gene inversion is in the sap genes (8 genes) of Campylobacter fetus. All eight sap genes encoding eight antigenically distinct SLPs are transcribed under a single sap promoter. The homologous mediated gene inversion results in the positioning of sap promoter in such
a way that it can only transcribe one of the eight sap genes. The sap gene inversion event is decreased in a recA mutant, suggesting partial dependence of the inversion event on recA (Dworkin and Blaser, 1996; Dworkin et al., 1997).

Figure 5 Gene conversion mediated antigenic variation event at pilin in Neisseria gonorrhoeae

The open portions of rectangles represent the conserved sequences, while the coloured patterns depict the variable sequences within pilS and pilE loci. The DNA exchange occurs between a silent pilS and the pilE loci at a short region of homology, resulting in generation of a pilE-pilS intermediate hybrid molecule in the second step. In the third step, this hybrid molecule donates sequence to the pilE locus within a recipient chromosome. The reaction involves two cross-overs. One cross-over happens at a shorter region of homology while the other occurs at a longer one. The recombination event at both cross-overs is a RecA-dependent mediated recombination. Thus in the last step, a recombinant pilE locus is obtained.
The duplication cum excision events happening at capsular gene cap3A in *S. pneumonia* serotypes result in the disruption of the ORF, leading to generation of an irreversible OFF phenotype and subsequent loss of capsular synthesis. The duplication occurs from 11bps to 239bps, switching the gene from ON to OFF. However, the excision mediated reversion restores the ON phenotype. The duplications occurring for 10bp and 100bp are reverted at frequencies of $10^{-5}$ and $10^{-3}$ respectively. The mechanism behind this kind of gene duplication is not well characterized, however, the possibility of involvement of *recA* can not be ruled out (Waite *et al.*, 2001, Waite *et al.*, 2003).

### 1.4.3: Site-specific recombination

Another mechanism of phase variation involves a non-homologous recombination event involving participation of specific enzymes which recognize particular sequences normally encompassing 30 bps. These conservative site-specific recombination (CSSR) events carry the ability to produce a range of genetic rearrangements, some of which result in the creation of a repertoire of antigenic variations (Hallet and Eherrat, 1997, Johnson, 2002). Conservative site specific recombination leads into a range of genetic rearrangements through inversion, insertion or excision of a DNA region (Hallet and Sherratt, 1997).

CSSR mediates phase variation events if the promoter region of a phase variable gene contains the invertible element. The invertible elements can be recognized by the presence of flanking short inverted repeats which are utilized by recombinases for gene inversion (Cerdeno-Tarraga *et al.*, 2005). For gene inversion, recombinases belonging
to serine and tyrosine families recognize short inverted repeat sequences located upstream and downstream of the invertible element. The inversion of regulatory elements (e.g. the promoter) of phase variable genes acts as a switch triggering alterations in the expression of a gene. In one orientation, the gene is switched to an ON phase while it switches to OFF for the other orientation (Jayaraman, 2009; Jayaraman, 2011).

There is also evidence that invertible elements located at the 5’ end of untranslated transcripts leads to the phase variation. The invertible element at the 5’ end of cwpV transcript undergoes an inversion event, causing the generation of ON and OFF phases of translation of cwpV in Clostridium difficile (Emerson et al., 2009).

Another example of site-specific recombination can be provided from phase variation of the fimA gene responsible for transcription of type I fimbriae in E. coli. The fimA gene is separated from fimB and fimE genes located upstream of fimA-by a 300bp invertible DNA element containing the promoter for fimA (Figure. 6). The fimB and FimE genes encode site-specific recombinases which mediate the inversion of the invertible element.
Figure 6 DNA inversion mediated phase variation in fimbrial operon in *E. coli*

The orange rectangles represent *fimA*, *fimB*, and *fimE* genes of type I fimbrial operon. The open rectangles are for inverted repeats IRR and IRL at *fim*, and blue arrows show the relative positions of promoters. The invertible repeat sequences are in between open triangles. IRL and IRR are located within the binding sites of recombinases FimB and FimE. The details of regulatory proteins managing ON-to-OFF transition can be found in the text.

The inversion event changes the orientation of the promoter of *fimA*, leading to ON/OFF switching of *fimA*. FimB and FimE are recombinases with different specificities. FimB acts on both ON and OFF phases of *fimA* and mediates the phase variation from ON to OFF direction or from OFF to ON direction. However, FimE has specificity for only “ON” variants of *fimA*, therefore, resulting in the switching of *fimA* from only ON to OFF direction. The frequency of switching mediated by FimB and FimE are calculated as $10^{-3}$-$10^{-4}$ and $10^{-1}$ respectively. Thus the relative expression of these recombinases in various niches determines the net PV rates for type I fimbriae.
The expression of *fimB* and *fimE* are regulated by many cis-acting factors such as H-NS, DNA supercoiling and silencers and activators located upstream of the *fimB* promoter, and trans-acting factors such as N-acetylneuraminic acid (van der Woude and Baulmer, 2004; Bayliss, 2009; Jayaraman, 2011). Given the importance of *fim* genes for encoding the fimbriae required for adhesion to host tissues, the phase variation of *fimA* plays an important role in adaptation to the host environment.

Inversion of a DNA element causes phase variation of expression of the *fim* and *fot* operons in *E. coli* and *mrp* in *P. mirabilis*, encoding type 1, CS18, and MR/P Fimbriae, respectively (Honarvar et al., 2003). This shows that phase variation of fimbriae alters the bacterial binding to the host epithelial cells, thereby affecting the adaptation to the host.

### 1.4.4: Transposition:

Transposons are mobile genetic elements which move from one site to another site in the genome. Their movement is mediated by transposases from donor site and insertion into the target site. The excision of some transposons is specific and precise and results in the restoration of the original DNA sequence of the recipient DNA. Therefore, insertion and excision of transposons can mediate phase variation, only if these processes are precise and specific (van der Woude and Baulmer, 2004).

For instance, the IS492 element contains a 5-bp spacer sequences sharing homology with the target locus *eps*. The *eps* locus encodes the enzymes required for biosynthesis of extracellular polysaccharide (EPS), essential for biofilm formation in *Pseudoalteromonas atlantica*. The insertion of IS492 in the target locus occurs in a
specific way, and switches the \textit{eps} into an OFF phase. Circular structures formed by IS492 at the target locus create the strong promoter for expression of a putative transposase, MooV. The enhanced level of MooV is required for precise excision of IS492 element. Besides, 5 and 7 bp located within flanking sequences of \textit{epsG} are also required for precise excision, indicating that transposition of the IS492 element is mediated through site-specific recombination. The precise excision of IS492 switches \textit{eps} gene into an ON phase. Thus the phase variation of \textit{eps} is controlled by precise insertion and excision of IS492 from \textit{epsG} (Perkins-Balding \textit{et al.}, 1999; Higgins \textit{et al.}, 2009).

The insertion and subsequent precise excision of IS2561 transposon from the \textit{icaA} locus mediates phase variation of the \textit{ica} operon in a similar fashion as described for IS492 element. The \textit{ica} operon in \textit{Staphylococcus epidermis} is responsible for production of polysaccharides essential for adhesion and biofilm formation (Ziebuhr \textit{et al.}, 1999; O’Gara, 2007). Transposition mediated phase variation events in EPS loci confer adaptive benefits to \textit{P. atlantica} and \textit{S. epidermis} by switching the expression from an OFF to an ON phase, and allowing them to inhabit their niche through adhesion with solid surfaces.

Insertion and excision of IS elements mediated by transposition cause the reversible phase variation. Transposition mediated by the putative transposes MooV causes the phase variation in its expression. A requirement for the transposition of IS elements is the assembly of transposases which recognize a short sequence identity between the insertion and excision elements and the target sequence (Pericone \textit{et al.}, 2000). This process has limited application, and is restricted only to IS elements.
1.4.5: Epigenetic Mechanism of Phase Variation:

The foregoing mechanisms involve alteration in the DNA sequences for generation of biphasic and multiphase antigenic variations. However, epigenetic mechanism does not involve alteration of sequence for mediating ON and OFF phases of gene expression (Deitsch et al., 2009). As this mechanism generates inheritable but reversible changes for gene expression without involving the sequence modulation, it is called an epigenetic mechanism of phase variation (van der Woude and Baulmer, 2004). The site specific methylases methylate specific DNA sequences in the promoter of a gene, thereby, altering the interaction of regulatory proteins, transcription factors and DNA binding proteins with promoter sequences. The alteration of interactions between DNA and transcriptional regulatory proteins mediates the switching from ON to OFF or from OFF to ON states, or decreasing/increasing the level of expression of differentially methylated phase variable loci. Dam methylase in *E. coli* is best studied for its role in regulation of gene expression through differential methylation of 5’-GATC sequences at the promoters of phase variable genes (Bayliss, 2009).

In *E-coli*, the Dam mediated phase variation events were reported for the *pap* operon encoding pyelonephritis-associated pili and *ag43* encoding an outer membrane protein.

1.4.5.1: Pap phase variation:

The *pap* operon encodes PapB, PapI, PapA and a number of other proteins required for formation and regulation of the P-pili. The promoter of *papBA* contains distal and
proximal binding sites for Lrp (Figure 7). Each binding site carries the recognition signature, GATC, for Dam methylase. The ON/OFF phase variation is mediated through methylation of one of the binding sites of Lrp. If Lrp binds at the proximal binding site, the methylation occurs at the distal GATC site, leaving the proximal GATC unmethylated. The unmethylated proximal GATC site switches the papBA into an OFF phase. However, when the distal GATC site is bound by a complex of Lrp and PapI, the proximal GATC site becomes available for Dam methylation. The methylation at the proximal GATC site results in ON expression of papBA. Beside Dam methylase, the interaction of PapB, catabolite activator protein (CAT), RimJ and H-NS plays an important role in mediating the transitions of pap expression. The description of this mechanism has been extensively reviewed (van der Woude and Baulmer, 2004; Hernday et al., 2002, Bayliss et al., 2009). The epigenetic control of pap phase variation has been illustrated in the figure below:
Figure 7 Regulation of *pap* operon by epigenetic mechanism

(A) The regulatory region of the *pap* operon showing six Lrp binding sites and two GATCdist and GATCprox sites within them at positions 2 and 5 respectively for Dam-dependent methylation, blue arrows indicate the two divergent promoters of *papI* and *papBA*. (B) & (C) the hexagon for Dam, ovals for
Lrp, triangles for papI, diamonds for papB, red arrows for low affinity, green arrows for high affinity, red cross for OFF phase and promoter without red cross for ON phase are shown. The circle on the top strand indicates the Dam-dependent methylation. The mechanism of OFF-to-ON transition has been explained in the text.

1.4.5.2: Ag43 phase variation

Ag43 protein is encoded by the flu gene of E. coli. Ag43 protein is a multifunctional outer membrane protein affecting the phage adsorption, influencing the biofilm formation, and promoting cell aggregation (Wallencha et al., 2002; Jayaraman, 2011). The promoter region of the flu gene is characterized by three GATC sites located within the binding site of OxyR protein, an oxidative stress response protein. Both Dam and OxyR compete with each other in a concentration-dependent manner for binding to the promoter region of flu. In the case of binding of OxyR to the promoter, the GATC motifs are masked and become unavailable for binding of Dam and subsequent methylation on them. Besides, binding of OxyR to the flu promoter serves as an inhibitor for binding of RNA polymerase (Haagman and van der Woude, 2000; Waldron et al., 2002). Hence, the binding of OxyR to its operator results in switching the expression of ag43 into an OFF state. Otherwise, binding and methylation by Dam of its recognition sequence GATC blocks the binding of OxyR, therefore, causing the switching of expression of Ag43 protein into an “ON” state. In this way, phase variation of flu is mediated by an interplay of Dam and OxyR in a concentration-dependent fashion (van der Woude and Baulmer, 2004; van der Houde, 2008).
1.4.6: Phasevarion:

Generally, the phenomenon of phase variation is associated with genes showing potential candidacy for expression of the surface structures involved in virulence of pathogens and in commensal activities of commensal microbes. However, this is not a universal principle. Phase variation is also reported to be associated with those genes which are not obviously related to transcription of surface structures. The phase variation in restriction modification systems (RM) is one such example. The genes encoding RM systems in many pathogenic bacteria are reported to be either experimentally confirmed or putative phase variable candidates (Moxon et al., 2006). There are a number of functions associated with phase variation of RM systems:- protection of the host genome from restriction by cognate restriction enzymes, defense against bacteriophage attack, differential expression of multiple genes via epigenetic control, and prevention of entry of foreign DNA into host bacteria.

The phase variation in this class of genes was considered an obscure phenomenon, but recently a few studies have been conducted on them to show the relevance of their phenotypic switching to regulation of multiple and non linked genes in the genome of pathogens (De Bolle et al., 2000; Fox et al., 2007). Due to their role of global regulation of other genes, the phase variable RM systems are termed as a phasevarion (a phase variable regulon of genes) which is a novel mechanism to generate phase variation. The mod gene, a component of a type III RM system, is reported to be part of a repertoire of phase variable genes in H. influenzae strains. The tetranucleotide repeats (AGTC or AGCC) regulate the phase variation mediated switching in mod genes of H. influenzae (Fox et al., 2007, Srikhanta et al., 2005, De Bolle, 2000). It has been proposed by
Srikhanta et al (2005; 2010) that the mod as a phasevarion influences the expression by differentially methylating its cognate sequences in the promoter region of other genes. Thus the mod role as a phasevarion is reminiscent of dam regulation of genes through differential methylation.

Srikhanta et al. (2011) have revealed that the modA gene, part of a phase variable RM system, controls a phasevarion through differential methylation of the genome in H. pylori. They performed microarray analysis of modA ON and modA OFF variants of modA and the subsequent comparison between them revealed that 6 genes were up-regulated and nine genes were down-regulated. Some of the up-regulated genes were associated with virulence factors such as fla and hopG genes encoding flagella and outer membrane proteins respectively. Thus the comparison of modA ON and modA OFF variants indicated that phase variable RM systems play a vital role in generating antigenic phase variation.

Srikhanta et al (2009) also showed the phasevarion a common strategy employed by H. influenzae, N. meningitidis and N. gonorrhoeae to switch between “differentiated” cell types to increase their survival and adaptability in the host. The micro-array analysis of modA ON and modA OFF states showed the up-regulation and down-regulation of multiple genes via differential methylation of a phasevarion. They further revealed that almost 80 genes were regulated upward or downward by the phase variation of modA alleles in N. meningitidis MC58 and N. gonorrhoeae FA1090. Some of the differentially expressed phase variable genes in MC58 (N. meningitidis) were found to be potential vaccine candidates such as lactoferrin binding proteins A/B. In N. gonorrhoeae, the modA ON and OFF variants showed different levels of resistance to antimicrobial agents in the animal model of infection, and various degrees of biofilm formation,
suggesting the regulation of expression of stress related genes via differential methylation.

Dam also mediates the regulation of genes via differential methylation, but what differentiates the phasevarions from the Dam dependent regulation is that mod alleles are phase variable as opposed to Dam which is not phase variable. In addition, mod alleles are known for their diversity in pathogenic strains of *H. influenzae* and *N. meningitidis*. The 18 different alleles of mod gene have been detected in different strains of *H. influenzae* and *N. meningitidis*, having various sequence specifies for methylation at different loci in the genome. This suggests that the diversity of mod alleles, in variety and number, has a potential to generate a repertoire of different phenotypes of genes regulated by differential methylation (Srikhanta *et al.*, 2010; Jayaraman, 2011; Gawthorne *et al.*, 2012). However, this situation does not hold true for Dam which is phase invariant.

SSR mediated phase variation has also been documented for type I restriction modifications system in *H. influenzae*, which controls the restriction phenotype against phage infection (Glover and Piekarowicz, 1974; Zaleski *et al.*, 2005). However, phase variation in type I RM genes, unlike phase variable type III, does not control the global gene expression via methylation (Jayaraman, 2011).

In short, whatever the regulatory mechanisms may be behind the process of phase variation, it is a stochastic event. The understanding of mechanisms underlying the phase variation will help us to build some models as to how the phase variation process in different genes will enable the bacterium to adapt to hostile and unpredictable environments. So the efforts to determine the mechanisms of phase variation are part of
ongoing endeavours to enhance our understanding about the extent to which phase variation can play a role in the survival of bacteria in challenging environments.

1.5: Determinants of phase variation:

Simple sequence repeat tracts are found in many bacterial pathogens, commensal and symbionts. These simple sequence repeats mediate phase variation, which plays a vital role in survival and adaptability of bacterial species to constantly evolving environments. The frequency and rate at which these phase variation events occur, determines the extent of genetic diversity mediated by them within a bacterial population (Bayliss, 2009). Different cis- and trans-acting factors are involved in regulating the frequencies of phase variation within SSCL of bacterial pathogens.

The sequence of repeats and their lengths constitute the major cis-factors controlling the switching frequencies of several phase variable genes in bacteria (De Bollie et al., 2000), in yeast (Tran et al., 1997; Wierdle et al., 1997) and in humans (Rolfsmeier et al., 2001).

Richardson et al (2002) showed that a linear relationship exists between an increase in PV frequency of hmbR and polyG/C tract length located in its ORF. De Bolle et al (2002) determined the phase variation frequencies and rates for the repeats present in phase variable genes of H. influenzae Rd strain by utilizing a LacZ reporter construct. They concluded that phase variation rates were increased linearly for the range off 17-38 tetrancleotide repeats. The findings of their research suggested that the number of repeats located in the phase variable genes were an important determinant of phase variation rates in H. influenzae. Bayliss et al. (2012) showed that phase variation
frequencies of *cj1139*, *cj0031* and *capA* in *C. jejuni* increased linearly with an increase in tract length, showing the involvement of tract length in controlling the PV rates and PV frequencies.

A link was investigated between phase variability and repeat tract length in 51 putative phase variable genes of MC58 strain. This investigation revealed that poly G/C tracts containing 4/5 residues or more were subjected to phase variation. In the case of a polyA tract, an A10 was declared as the threshold value for phase variation in MC58 strain. Tetranucleotide repeats containing more than 4 copies were likely to undergo phase variation. This study demonstrated that the length of repeat tract affects the potential of genes to be phase variable; therefore, it was an important determinant for regulation of genes through phase variation (Martin *et al.*, 2003; Saunders *et al.*, 2000). Parkhill *et al* (2000) reported 27 phase variable genes containing polyG/C repeat tracts in *C. jejuni*. They concluded that repeat tracts containing less than 8 residues were unlikely to undergo changes in expression through phase variation in *C. jejuni NCTC11168*.

However, the repair systems are responsible for controlling the integrity and fidelity of genomes during replication/post replication processes. Thus the replicative/post replicative machinery of the cell, particularly the mismatch repair system, forms an important trans-factor affecting the frequency of alterations in simple sequence repeats of phase variable genes.

Tran *et al* (1997) investigated the impact of a mismatch repair system and DNA polymerase proofreading on the frameshift mutations in homopolymeric runs of yeast. They used various variants of repeats ranging from 4-14 bp. They concluded that DNA polymerase proof reading activity and the MMR pathway were efficient in correcting
the mutations for short runs (4-7 bp), however, they could not correct efficiently the mutations arising in the long runs of homopolymeric repeats (> or = 8 bp)

The impact of trans-acting factors in controlling phase variation rates have been investigated by Bayliss et al (2004) in *H. influenzae*. They investigated the pilin genes (*hifA* and *hifBCDE*), whose products mediate biosynthesis and assembly of pilin, containing 10 dinucleotide (5'-AT) repeats and determined the influence of mutations in the MMR pathway on phase variation rates of *hif* genes. Their results suggested that the defects in MMR pathway did not affect the dinucleotide repeat mediated phase variation of pilin in *H. influenzae*. However, the mutation in *mutS* influenced the phase variation rate of another locus carrying a similar dinucleotide repeat motif in the chromosome, indicating that the effect of trans-acting factors is not parallel for different genes (Bayliss et al., 2004).

Mutations in many DNA repair genes, such as *mutS*, *mutL*, *mutH*, *uvrD*, *polI*, *recA* and *mfd*, were constructed for determining their influence on dinucleotide and tetranucleotide repeats in *H. influenzae* strain Rd. The mutations in *polI* but not other repair genes affected the PV rates of tetranucleotide repeats. Conversely, mutations in MMR genes but not *polI*, *mfd* and *recA*, affected the dinucleotide repeat mediated phase variation rates. However, a deletional pattern was more prominent than an insertional one among the changes in tetranucleotide repeats due to end-joining occurring without DNA polymerization during *polI* deficient Okazaki fragment processing (Bayliss et al., 2002). The inactivation of the Klenow fragment, a part of DNA polymerase I, and RNAseHI also resulted in an increase of phase variation in *H. influenzae* (Bayliss et al., 2002; Bayliss et al., 2005). These studies suggested that lagging strand synthesis is a major determinant for controlling the switching rates of genes containing repeats of 4 or more nucleotides (Bayliss et al., 2009).
In *N. meningitidis*, a mutator phenotype was associated with high phase variation rates. The phase variation mediated phenotypic and genotypic diversity in mutator clones of *N. meningitidis* was caused by defects in mismatch repair genes (MMR). The defects in MMR genes resulted in >100 fold increase in phase variation rates as compared to wild type isolates (Richardson *et al.*, 2002). Another study conducted by Richardson and Stoljilkovic (2001) showed that an increase in mutational frequencies and phase variation rates of polyG/C tracts were observed in those clinical isolates of *N. meningitidis*, which were carrying defective versions of *mutS* and *mutL* (MMR genes). The switching frequencies of *hmbR* and *hpu*, encoding Hb receptors, were increased 500 and 250 fold in *mutS* and *mutL* mutants respectively. Oliver *et al* (2002) described similar results by showing that high switching frequencies were associated with alterations in the mutator phenotype mediated by mutations in MMR genes such as *mutS, mutL*, and *urvD* in *P. aeruginosa*.

The induction of MMR pathway gene, *xseB*, was also reported in *N. meningitidis* during contact with host cells. The XseB of *N. meningitidis* possesses a high degree of homology with Exonuclease VII from *E. coli*, which plays a vital role in methyl directed MMR pathway. The induction of XseB during interaction with host cells increased the phase variation rate by frameshifting (Morelle *et al.*, 2005). There is another genetic factor *dinB* which is reported to control the switching frequency in *N. meningitidis*. The overexpression of dinB increased the switching frequency of polyG tracts of *siaD* in *N. meningitidis*. Similar results were found for homopolymeric runs in *E. coli*, when the overexpression of dinB, encoding error-prone DNA polymerase IV, caused an increase in switching frequency of simple sequence repeats (Kim *et al.*, 1997). However, the inactivation of Dam did not affect the switching rates in *N.
meningitidis MC58 strain, which is consistent with the absence of \textit{mutH}, which links Dam methylation with the MMR pathway in \textit{E. coli} (Martin, 2004; Bayliss, 2009).

1.5: Phase variation in \textit{Campylobacter jejuni}

1.5.1: The genus \textit{Campylobacter}:

The first report on \textit{C. jejuni} is dated back to 1886, and was presented by Escherich who recognized the “non-cultural spiral shaped bacteria” (Vandamme \textit{et al.}, 2010; King and Adams, 2008). However, in 1906, \textit{C. jejuni} was discovered in the uterine mucus of sheep by two British veterinarians (Skirrow, 2006; Zilbauer \textit{et al.}, 2008). In 1913, two scientists isolated the microorganisms from cattle suffering from diarrhoea, and named them as \textit{Vibrio jejuni}, followed by isolation of different \textit{Vibrio} from pigs suffering from diarrhoea by Doyle who named them as \textit{Vibrio coli} (Vandamme, 2000; Vandamme \textit{et al.}, 2010). Based on different base composition, microaerophilic requirements for growth and non-fermentative metabolism, they were distinguished from \textit{Vibrio} species (On, 2001; Cowan and Steel, 1993; Annable \textit{et al.}, 1997). Debruyne \textit{et al} (2005) described 14 species of \textit{C. jejuni} belonging to the genus \textit{Campylobacter}. However, 20 species were described by Fernandez \textit{et al} (2008), including 16 major species with 4 subspecies for genus \textit{Campylobacter} (On, 20001; Foster \textit{et al.}, 2004).

The family Camplobacteraceae contains two genera: \textit{Campylobacter} and \textit{Arcobacter}. Both genera assume a commensal life-style in domestic animals (Vandamme, 2000). \textit{Campylobacter} is curved, rod shaped, Gram-negative, flagellate and spiral shaped bacterium. It is closely related to \textit{H. pylori}, a gastric pathogen, and shares many properties with it (Blaser, 1997). The flagellum is located at both ends of the cell,
resulting in the corkscrew like motion in the majority of *Campylobacter* species (Figure 8).

![Individual cells of *C. jejuni* with bipolar flagella](image)

**Figure 8 Individual cells of *C. jejuni* with bipolar flagella**

The *Campylobacter* strains were mostly found resistant to antibiotics such as cephalothin and fluoroquinolones employed to treat human and animal diseases (Koenraad *et al.*, 1995). The majority of strains of *Campylobacter* species have a propensity to form non-culturable but viable cells under hostile conditions (Portner *et al.*, 2007), which can assume the cultivable status on passage through the GIT tract of chickens (Cappelier, 1997).
1.5.2: Growth and Survival Characteristics:

*Campylobacter* species growing within a temperature range of 37-42°C are called thermophilic *Campylobacter*. Thermophilic *Campylobacter* can not grow below 30°C due to the absence of cold shock proteins necessary for survival at low temperature. However, the optimum temperature for growth of campylobacter was reported to be 41.5°C (Silva *et al*., 2011), therefore, they are not true thermophiles which normally grow at 55°C. They were named as thermotolerant by Levin (2007). *Campylobacter* can survive on the surface of food or any solid surface with relative humidity 60-62% for 4 hours at a temperature of 27°C. Continuous freezing and thawing reduces the survival time of *Campylobacter* (Stern and Kazmi, 1989), however, the freezing of contaminated food does not eliminate *Campylobacter* (Lee *et al*., 1998). The optimal pH for *Campylobacter* survival lies in the range of 6.5 to 7.5. Below pH 4.9 and above pH 9.0, the cells of *Campylobacter* are unlikely to survive (Hazeleger *et al*., 1995). The *Campylobacter* species are microaerophilic and grow best under low O₂ pressure. The conditions for optimum growth are O₂ = 5%; CO₂ = 10% and N₂ = 85% (Gareaux *et al*., 2008; Silva *et al*., 2011).

1.5.3: Campylobacteriosis: Occurrence and its Clinical Manifestations

*C. jejuni* is a major documented cause of gastroenteric infection. Its epidemiology is poorly understood although it is well distributed in the environment and throughout the food chain (Frost, 2001). The clinical spectrum of Campylobacteriosis ranges from manifestations of watery diarrhoea to severe bloody diarrhoea accompanied by abdominal pain and fever (Prendergas *et al*., 2004). Campylobacteriosis includes the
symptoms of Travellers diarrhoea (Scott, 1997). *C. jejuni* is a commensal organism of chickens (Young *et al.*, 2007). Campylobacteriosis is an important antecedent to Guillain-Barre Syndrome, a neuromuscular paralysis (Nachamkin *et al.*, 1998). *C. jejuni* is reported to cause 5% to 14% of diarrheal cases worldwide, which is suggestive of 400-500 million cases per year (Friedman *et al.*, 2000).

EFSA (2010) reported the clinical cases of Campylobacteriosis in Europe as: “there may be not less than 2 million and possibly as high as 20 million cases of clinical Campylobacteriosis per year in the EU 27 MS” In 2009, the reported cases of Campylobacteriosis in UK broke its previous records (ACMSF, 2010). Slaughterhouses, consumption of chickens and reservoir of chickens account for 5-80% this is a huge range shouldn't it be 80% cases of Campylobacteriosis in humans (EFSA, 2010).

### 1.5.4: Transmission of *Campylobacter* to Humans

*C. jejuni* has many reservoirs in the environment (Figure. 8). First, it colonizes the caecum and colon of chicken guts and multiplies, primarily, in the mucosal layer (Newell and Wagenaar, 2000). Its transmission to chicks in the flock occurs through the fecal oral route. It also enters the water supply and becomes attached to protozoan like amoebae. Humans are infected by poorly cooked contaminated chicken meat (Friedman *et al.*, 2001) or unpasteurized milk or through the consumption of contaminated drinking water. After entry into the human gut, *C. jejuni* invades, especially, the epithelial stratum of the intestine and causes inflammation and mild to severe diarrhoea (Young *et al.*, 2007). The Fig 9 illustrates the transmission routes.
1.5.5: Genome of *Campylobacter jejuni*:

The length of the genome of *C. jejuni* strain NCTC11168 is 1,641,481 base pairs (bp), which comprises of 1654 putative coding sequences. It suggests that 94.3% of the genome is specified for coding different proteins, declaring it the densest genome of bacteria ever sequenced (Lobry, 1996). The salient feature of *C. jejuni*’s genome is that
it is almost devoid of any IS sequence, transposons, prophages /retrons (Parkhill et al., 2000).

The C. jejuni NCTC11168 possesses a striking feature of having a circular chromosome with a 30.6% G+C content. This indicates that the C. jejuni genome is AT rich, reflecting the deviation from the GC rich genomes of other pathogens. There are no insertion sequences, very few repeats, less organization of genes into operons or clusters and existence of hypervariable sequences (Parkhill et al., 2000). Another intriguing feature of C. jejuni is that it lacks repetitive sequences. The genome of C. jejuni shows little organization into operons like that of its close relative H. pylori and other pathogenic bacteria. There are only few operons and gene clusters such as the ribosomal operon and gene clusters including LOS and EP biosynthesis.

Shotgun assembly of the genome of C. jejuni displayed the presence of homopolymeric tracts of G:C which were subjected to length variations among identical clones. Variations in these homopolymeric tracts are potentially subject to variations due to slipped strand mispairing during DNA replication. The hypervariable homopolymeric tracts were mainly clustered in the genome regions specialized for encoding flagellar modifications, LOS and EP biosynthesis (Parkhill et al., 2000). This can be viewed in the figure 9:
Figure 10 Genomic Organization of the *Campylobacter jejuni* NCTC11168

The clusters of genes responsible for LOS biosynthesis, EP biosynthesis and flagellar modification are marked.

1.5.6: Homopolymeric Tracts in *C. jejuni*----Role in Pathogenicity
The homopolymeric tracts within the contingency loci of *H. pylori* were found to be exhibiting phase variation in their sequences, which was confirmed by sequencing reactions. The phase variation mediated changes in the expression of phase variable genes of *H. pylori* result in adaptation to the microenvironment within the host (Alm *et al*., 1999). Homopolymeric tracts have been identified in 30 genes (27 coding genes plus 3 pseudogenes) of *C. jejuni* which is a close relative to *H. pylori*, indicating the significance of these repeat tracts in understanding the pathogenicity and biology of that pathogen (Parkhill *et al*., 2000). The majority of these repeat tracts are intragenic having G/C residues between 8 and 13. The length polymorphism in repeat tracts of *C. jejuni* results in translational frameshifts, leading to switching of expression between an ON or an OFF phase. A significant proportion of these repeat tracts in the NCTC1168 strain are clustered on the genome in three loci: LOS biosynthetic locus, capsular polysaccharide biosynthetic locus and flagellin modification locus. This is shown in the table below:

**Table 1 Potentially phase variable genes of NCTC 11168 strain of *Campylobacter jejuni***

<table>
<thead>
<tr>
<th>Polymeric tracts (No. of G/C)</th>
<th>Target genes</th>
<th>Putative functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>G (8-10)</td>
<td><em>Cj0031/0032</em></td>
<td>Putative type IIS restriction modification system</td>
</tr>
<tr>
<td>C (9-11)</td>
<td><em>Cj0045</em></td>
<td>Putative Iron binding protein (Hemerythrin-like)</td>
</tr>
<tr>
<td>(G9, 11)</td>
<td><em>Cj0170/0171</em></td>
<td>Conserved hypothetical</td>
</tr>
</tbody>
</table>
G (9-13)  
**Cj0046**  Pseudogene

T (4-5)  
**Cj0628/0629**  Lipoprotein

G (9-10)  
**Cj0685c**  Putative sugar transferase

G (9-10)  
**Cj676**  Pseudogene

T (7)  
**Cj1677/1678**  Unknown

G (10-11)  
Non-coding  Upstream of rRNA

G (8-9)  
**Cj0275**  Clp Protease ATP-binding protein

G(12-13)  
**Cj0565**  Noncoding, upstream pseudogene

G (9-10)  
**Cj0617/0618**  Unknown

**LOS biosynthetic locus**

G (8-9)  
**Cj1139**  Galactosyltransferase

C (8-9)  
**Cj1144/1145**  Unknown

**Flagellin modification locus**

G (9-10)  
**Cj1325/1326**  Unknown

G (9-10)  
**Cj1305**  Unknown, 617 family

G (8-9)  
**Cj1306**  Unknown, 617 family

G (9-10)  
**Cj1310**  Unknown, 617 family

G (10-11)  
**Cj1318**  Unknown, 1318 family

G (10-11)  
Upstream Cj1321 Transferase
There are many surface exposed moieties and phenotypes in *Campylobacter jejuni*, which undergo the phase variation. They are listed and discussed below.

### 1.5.8.1: Capsule

The capsule in both Gram-positive and Gram-negative bacteria undergoes phase variation. Capsular polysaccharides are considered to protect the bacteria from the host innate immune system. In *C. jejuni* capsule is involved in providing serum resistance, the ability to invade the human epithelial cells and to colonize the chicken gut, and
promotion of diarrheal disease in ferrets (Jones et al., 2004; Bacon et al., 2001). However, capsular polysaccharides are also implicated in provision of target for phage binding sites on the bacterial surface and the consequent opportunity for phage infection (Roberts, 1996; Comstock and Kasper, 2006). Coward et al (2006) identified that mutation in capsular polysaccharides genes resulted in increased phage resistance against a group of phages tested in their study.

In addition, CPS also serve as the basis for grouping the C. jejuni strains under the Penner serotyping scheme; and 47 different serotypes have been detected (Guerry et al., 2012). This reflected the variation in the genetic composition of capsular loci of various C. jejuni strains (Karlyshev et al., 2005). However, the strains with identical genetic composition of CPS genes may show variations in structures of CPS due to the phenomenon of phase variation mediated by intragenic repeat tracts located at genes encoding CPS (Linton et al., 2001; St Michael et al., 2002; Karlyshev et al., 2005; Chen et al., 2008).

There are 28 genes located in the CPS locus on the genome of C. jejuni NCTC 11168. These genes undergo phase variation due to polyG/C tracts in their ORFs, which are named as cjl420, cjl421, cjl422, cjl426, cjl429 and cjl437 (Parkhill et al., 2000). Only three of these genes have been characterized, while the other three have unknown functions. McNally et al (2007) have shown that MeOPN transferases are encoded by cjl421 and cjl422 and that these enzymes translocate sugar units to the GalNAc and the Hep residues, respectively. Sternberg and his colleagues demonstrated that cjl426 encoded a 6-O-Me transferase (Sternberg et al, submitted). Variations in the presence of residues of ethanolamine, phosphoramidate, methyl and aminoglycerol have been detected in the CPS (St Michael et al. 2002, Szymanski et al., 2003; McNally et al.,
These variations may have been brought about by the phase variation of the other three phase variable genes located in the CPS locus (Sorensen et al., 2012).

1.5.8.2: Flagella

Flagella are considered necessary factors for the pathogenesis and the colonization in the gastrointestinal tissues and infection of C. jejuni (Guery, 2007). C. jejuni’s flagella are composed of glycosylated filaments whose expression is regulated by FlgS (the sensor protein) and FlgR (the response proteins), a two component system (Datia et al., 2010).

Moreover, the flagella of C. jejuni consist of major and minor flagella proteins, FlaA and FlaB respectively (Gaury et al., 1991). Mutation of the flaA gene in C. jejuni caused the formation of a non-functional flagellar filament containing FlaB, and showed a serious reduction in motility. However, a mutation in flaB had no impact on motility (Gaury et al., 2007). Hence, the expression of flaA is essential for colonization, adherence and invasion of the host epithelial cells (Jain et al., 2008). Apart from the motility function, C. jejuni flagellar apparatus also performs the secretory function by secreting non-flagellar proteins and hence playing a role in virulence (Poly and Guerry, 2008).

The flagella in C. jejuni retain the inherent ability to undergo reversible phase variation (Diker et al., 2006). Karlyshev and his colleagues have demonstrated that maf1 and maf4 genes involved in flagellar biosynthesis had in-frame polyG tracts. The SSR mediated phase variation in them led to changes in the presence or absence of flagella in
*C. jejuni*. This reversible switching of flagellar proteins is critical for adaptation of *C. jejuni* to *in vivo* and *ex vivo* environments (Karlyshev *et al*., 2001).

Glycosylation of flagella is supposed to be involved in the evasion of the immune system. The phase variable *cj1295* is located at the flagellin glycosylation locus (Table 1). The phase variation mediated by changes in the polyG tract of *cj1295* leads to generation of structural changes in the flagellar glycoprotein structure. This helps *C. jejuni* to evade the immune system of the host (Hitchen *et al*., 2010). However, the majority of phase variable genes in the flagellin modification locus (Table 1) still need to be characterized. They may have significance for producing variety in the glycosylated flagellin, thereby, providing advantages for adaptation and survival in the evolving environment of the host.

### 1.5.8.3: LOS Structures

Lipooligosaccharide molecules in *C. jejuni* contain a lipid A moiety and oligosaccharide units (OS) forming inner and outer cores of LOS. In some strains, the outer core of OS mimics structures of GM1 and other gangliosides; and this strategy is used by *C. jejuni* to avoid the immune system of humans (Oldfield *et al*., 2002; Michael *et al*., 2002). The mimicry of GM1 ganglioside structures of the human host by *C. jejuni* may be the likely explanation for the cause of Guillain-Barré syndrome (GBS) and Miller Fisher syndrome (Aspinall *et al*., 1995; Moran *et al*., 1996).

Phase variation is observed with several of the LOS genes (Karlyshev *et al*., 2005). Parker *et al*. (2008) showed that LOS biosynthesis regions in 15 strains of *Campylobacter* are hotspots for antigenic variation mediated by gene inactivation,
variable presence/absence of genes and phase variation. Gilbert et al. (2004) identified a total of nine genes undergoing phase variation due to the presence of poly G/C repeat tracts in their ORFs, but the function of the phase variation phenomenon has only been studied in some of these genes. During genome sequencing, Parkhill et al. (2000) marked *cj1138* (*wlaN*) and *cj1144/cj1145* as genes containing homopolymeric repeat tracts (G/C). The *wlaN* gene (*cj1139*) in *C. jejuni* NCTC11168 strain encodes for a galactosyltransferase enzyme which phase varies to produce GM1 and GM2 like modifications of LOS, (Linton et al., 2000). The *wlaN* gene contains an intragenic homopolymeric tract which is subjected to phase variation through a slipped strand mispairing mechanism. They showed that G9 repeat tract introduces a frameshift mutation which causes the emergence of pre-mature stop-codon and subsequently results in the conversion of a GM1 structure into a GM2 structure. In contrast, G8 results in the translation of full length protein and an LOS mimicking a GM1 structure.

*cj1144/1145* contains C8, C9 and C10 repeat tracts. C8/C9 truncates the translation of *cj1145* while contrastingly C10 allows the translation into *cj1144*. The fused protein product resulting from *cj1144/1145* has an unknown function. These genes might play a role in generating further structural diversity within LOS structures. In *C. jejuni* strain 81176, the *cgtA* gene undergoes phase variation due to presence of G9, G10 and G12 tract lengths; and G12 allows the translation of full-length protein resulting in LOS variants mimicking GM2 gangliosides. In contrast, LOS variants arising from truncated products from G9 and G12 mimic GM3 structure (Guerry et al., 2002).

Thus phase-variation mediated differences in the LOS structure and their mimicry with gangliosides is thought to participate in the clinical features of GBS. However, all LOS variants are not involved in the generation of these features (Nachamkin et al., 1999),
indicating that there may be other factors implicated in the pathogenesis of *C. jejuni*, which are yet to be characterised.

### 1.5.8.4: Restriction-Modification System

The type I restriction modification system in *C. jejuni* is well characterised. In 73 strains of *C. jejuni*, the type I restriction–modification (*hsd*) systems were recognized and characterized based on their DNA and amino acids sequence (Miller *et al.*, 2005). *C. jejuni* NCTC11168 strain also possesses a putative type IIS restriction modification system (*cj0031/0032*), which is yet to be characterized. This putative type IIS RM system *cj0031/0032* is the only phase variable RM system in NCTC11168 strain. A repeat tract length ranging from G9 to G10 is present in the reading frame of *cj0031/0032*. The repeat tract length G9 mediates the in-frame fusion of *cj0031* and *cj0032*, resulting in production of a functional product. However, G10 separates their reading frames, leading to production of a non-functional or truncated methyltransferase enzyme (Fig. 11).
Figure 11 Phase variation mediated by repeat tract of cj0031/0032

G9 repeat tract represents the ON phase, G8 and G10 represent OFF phases. The red line indicates the active transcript while blue line with dumbbell shape shows the truncated or inactive protein.

1.6: DNA Restriction-Modification Systems

The restriction endonuclease recognizes a specific DNA sequence in order to introduce a double-strand break. The cognate modification enzyme recognizes the same target
sequence, methylates it and protects it from degradation by the endonuclease. When both enzymes (endonuclease and methyltransferase) act in a concerted manner, they constitute a restriction-modification system. The genes encoding the restriction endonuclease (ENase) and cognate methytransferase (MTase) are usually tightly linked in the genome (Kobayashi, 2004).

A “Non-self” or foreign DNA molecule exists in an unmethylated condition on both strands which are recognized and destroyed by the host endonucleases. Thus, R-M systems in bacteria serve as a cellular immune system acting on the invading DNA from phages or other sources, and destroy it (Pingoud and Jeltsch, 2000; Naito et al., 1995; Kobayashi, 2004).

MTases confer immunity to restriction mediated post-segregational killing through restriction upon the host DNA by methylating the endonuclease recognition sites. During replication of the bacterial DNA, MTases methylate a specific base within their recognition sequence, so they bestow the immunization upon the nascent strands of replicating DNA to an attack by endonuclease. The bacterial R-M systems, however, do not serve as a perfect immune system for bacteria, as sometimes, the bacteriophages or parasitic DNA molecules gain the host modification and escape from inactivation, (Bickle and Kruger, 1993; Kobayashi, 2001).

1.6.1: Classification of RM systems:

RM systems have been classified into 3 major types based upon their sequence specificities, co-factor requirements, mode of action, number and architecture of subunits and regulation of their expression (Wilson et al., 1991; Bujnicki, 2001). However, with the discovery of novel RM systems, further types and subtypes had been
proposed. The major types and subtypes of RM systems have been illustrated in this section.

1.6.1.1: Type I restriction-modification enzymes

Type I R-M systems consist of restriction subunits to restrict the DNA molecule, a specificity subunit (S1) to recognize the specific binding site, and MTase subunits (M2) for methylation of a specific DNA sequence in the host DNA (Dryden et al, 2001, Murry, 2000). One S1 and two M2 subunits combine to form the M2S1 complex, a methyltransferase component, which recognizes and modifies a specific base within the target recognition sequence. The M2S1 utilizes only hemimethylated DNA as a substrate for modification. M2S1 further joins with two restriction subunits R2 to make
the whole complex R2M2S1 acting as type I restriction modification system (figure 12).

Figure 12 Schematic representation of Type I restriction modification system
(Source: Bujnicki, 2001)
a) The M (HsdM) subunit comprising a single MTase module with N- and C-terminal extensions, b) the S (HsdS) subunit that exhibits circular pseudo-symmetry, comprising variable TRDs and conserved spacer domains, c) the R (HsdR) subunit comprising modules implicated in DNA cleavage, DNA translocation and binding to the M·S complex. d) proposed architecture of the M2R2S complex recognizing its bipartite target sequence using two TRDs, generating DNA loops and cleaving DNA at a distance. For the sake of clarity, only one $M_R S_2$ complex is shown, although dimerization is necessary for DNA translocation and cleavage to occur [16], and aspects of other possible interactions between the domains cleavage is ignored.

The endonuclease component of a type I RM system requires ATP, AdoMet (SAM) and Mg$^{2+}$ to perform the restriction reaction. However, the MTase component requires only AdoMet (SAM) to carry out the methylation reaction at the N$^6$ position of an adenine residue in the recognition sequence. Cleavage caused by the restriction subunit (R) occurs at a considerable distance from the recognition sequence. The recognition sequence of type I R-M is asymmetric and bipartite (Powell et al., 1993, Powell et al., 1998).

Restriction subunits only target the unmethylated DNA sequence, dimerizes rapidly and initiates translocation of DNA towards itself from both directions simultaneously to cause cleavage non-specifically at random target sites in DNA (Berge et al., 2000) while the methylation subunit (MTase) pinpoints the hemimethylated DNA of the host for modification. Thus, the fully methylated DNA is resistant to restriction mediated by the restriction ENase.
1.6.1.2: Type II restriction-modification enzymes

Type II R-M enzymes comprise a repertoire of over 3650 different R-M systems (Tock and Drydon, 2005). They are the simplest and the most prevalent among bacterial species (Roberts, 1990). These systems consist of separate subunits of ENase and MTase. The recognition sequences of Type II R-M systems are palindromic and 4-8 bp in length. A type II R/M cleaves both DNA strand within or adjacent to its recognition sequence, yielding defined products with both 5′-PO₄ and a 3′-OH termini (Pingoud and Jeltsch, 2000; Bujnicki, 2001). Some of these systems act as homo-dimers, and each monomer cuts DNA in a coordinated fashion to make a double-strand cut.

Type II MTases are usually found in a monomeric forms and perform modification through methylation of adenine or cytosine residues in the recognition sequence on each strand of the duplex. Methylation of a cytosine occurs at either the N⁴ or the C⁵ position, while that of an adenine occurs only at the N⁶ position (Sistla and Rao, 2004).

1.6.1.2.1: Type IIS restriction-modification enzymes:

Type IIS restriction modification systems shares some features with type III R-M systems. The ENase cleaves both strands of DNA at a fixed position, usually several base pairs (1-20bp) away from the asymmetric recognition sequence. Nevertheless, a type IIS RM system does not depend on the presence of ATP, SAM or an MTase subunit for the cleavage reaction. These biochemical properties of type IIS separates them from the type III RM system.
The type IIS have been classed separately from type II R/M systems due to their asymmetric recognition sequences. They assume the monomeric form with distinct domains for sequence recognition and DNA cleavage (Arber, 1965). Dimerisation, however, is an obligatory requirement for some of the type IIS RM systems such as FolkI (Bitinaite et al., 1998; Vanamee et al., 2001) [Figure 13]

**Figure 13** Type IIS restriction modification system, exemplified by *FolkI* (Source: Bujnicki, 2001)

- a) the MTase component comprises two type II-like MTase domains fused within a single polypeptide or two separate enzymes (a dotted line shows the presence of a possible linker sequence) or a single MTase able to methylate different sequences on both strands of the target, b) the type IIS ENase homodimer bound to two targets [43] and generating a ds break at a fixed distance in respect to one of the sites

The MTase, in some cases, exists in a transient dimeric form carrying out methylation on both strands of DNA (Brown and Jiricny, 1987, Bath et al., 2002). Most of the Type IIS enzymes normally communicate between two copies of the recognition sequences on a DNA substrate, and so cleave DNA substrates with two sites more rapidly than a one-site substrate (Bath et al., 2002).
Some type IIS systems like *EcoR571* are irregular in their structure and mode of action, containing a REase-MTase subunit and a separate MTase subunit. The former cleaves or modifies one strand of DNA, while the latter modifies both strands of DNA. Cleavage is stimulated by AdoMet (SAM) (Bickle and Kruger, 1993).

As type IIS RM system methylates asymmetric recognition sequences, they require the presence of two MTase subunits which may methylate both strands of a target substrate. Two MTase subunits may methylate different bases on the top and the bottom strand of DNA. For instance, *NgoBVIII* methylates adenine (GGTG\textsubscript{A}) on the top strand and cytosine (TCAC\textsubscript{C}) on the bottom strand.

In certain instances, such as *FolkI*, two MTases with distinct specificities are part of a single fusion protein, methylating adenine (GG\textsubscript{A}TG) on the top strand and adenine (CAT\textsubscript{T}CC) on the bottom strand.

### 1.6.1.3: Type III restriction-modification enzymes

Type III R/M systems shares some similarities with type I R/M systems (Drydon *et al.*, 2001). They were first placed along with type I systems as a group of ATP-dependent RM enzymes (Boyer, 1977). Later, it was discovered that, unlike type I RM systems, a type III complex consists of only two genes, a methyltransferase (Modification, mod) and an endonuclease (Restriction, res). The products from both genes form two enzymes—Mod and Res. Mod can act independently as a MTase on their cognate recognition sequence and methylates a specific base in the recognition sequence (Tock and Drydon, 2005). In contrast, the Res subunit cannot restrict its substrate
independently. Unusually, it forms a complex with the Mod subunit to be functional (Figure 14).

Figure 14: The schematic representation of architecture of EcoP1: type III restriction-modification system ((Source: Bujnicki, 2001)

1) The M (Mod) subunit comprising a MTase module with the TRD localized within an insert, b) the R (Res) subunit comprising modules implicated in DNA cleavage and DNA translocation, c) proposed architecture of the M-R complex comprising two enzymes bound to sites in a head to head orientation.

For the sake of clarity only one R and one M subunit in each complex interacts with the DNA and possible contacts between elements other than the ENase domains are ignored.
Mg$^{2+}$ and ATP are requirements for the normal restriction reaction catalysed by the res subunit, while the modification reaction carried out by the mod subunit depends on the presence of SAM. SAM is also required by the R subunit (res) for efficient cleavage [Bist et al., 2001].

Type III enzymes recognize 5-6 bp asymmetric recognition sequence for cleavage, and require two inversely oriented copies (Meisel et al., 1992). Cleavage of substrate DNA is mediated by a mod subunit on one strand, about 25 bps away from the recognition site Wilson, 1991). The intriguing feature of Type III RM system is that they do not cut all of the target sites, always leaving some proportion of target sites uncut in the genome. Methylation of Adenine on the N$^6$ position by mod on one strand of the substrate results in the formation of a hemi-methylated DNA molecule. This type of methylation is aimed at the protection of hemimethylated DNA from the action of the type III res subunit which only cuts unmethylated DNA substrates (Tock and Drydon, 2005).

Recently, it has been demonstrated that type III Enases act on the substrate DNA in the form of R2M2 complexes restricting double stranded DNA in a cooperative manner. The ENase of one complex cuts at its cleavage site on the top strand, while that of the second complex restricts the bottom strand (Janscak et al., 2001; Bujnicki et al., 2001).

EcoP11I and EcoP15I are well-known examples of type III R-M systems, and were discovered in Gram-negative bacteria (Bickle and Kruger, 1993).

1.6.1.4: Type IV restriction-modification enzymes
Type IV RM systems are composed of two complexes. One of the complexes acts as an MTase only and methylates its asymmetric target recognition sequence on both strands of DNA. However, the second complex of type IV RM possesses both ENase and MTase activities; and it only methylates one strand of the asymmetric recognition site. The architecture of type IV RM system has been shown in the figure 15:

![Figure 15 Simple Architecture of Type IV restriction-modification system](Source: Bujnicki, 2001)

The ENase component of a Type IV RM complex cuts the substrate at 16/14 bp away from the recognition site in a 3’ direction (Janulaitis et al., 1992). This biochemical property of Type IV RM systems is similar to that of Type IIS and Type III RM systems.

Type IV RM systems, like type III RM systems, restrict their target sites partially and cleavage is catalysed in the presence of SAM. This is true in the case of cleavage carried out by Eco57I. However, SAM is essential for the cleavage reaction conducted by BseMII. This class comprises of 227 putative enzymes. McrBC from E. coli K12 and Eco57I are the best characterized examples among them (Pieper et al., 1997).
1.7: Background to the project:

The majority of the phase variable genes of *C. jejuni* possess putative functions. One of the phase variable genes, *cj0031*, was reported as having a putative type IIS restriction modification function. These homologies were re-examined in Objective 2. Preliminary work indicated that this gene was subject to phase variation during colonization of chickens by *C. jejuni* strain 11168 (Bayliss, personal communication and Objective 1). This suggested that phase variation of *cj0031* might have some significance at some stage of infection for the bacterial population inside the host. So characterization of this gene was important to know its real function and to generate important hypotheses about its in vivo role. The phase variation rates and putative restriction functions were examined (Objective 2). It was also hypothesized that if *cj0031* encodes an RM system then it may help *C. jejuni* to protect itself against infection by bacteriophages (Objective 7) and/or have some role in increasing pathogenicity by controlling expression of some other genes in the genome. Thus, it was also hypothesized that this gene may be important for adhesion or invasion of *C. jejuni* into the host tissues (Objective 5 & 6). This project was therefore designed to characterize the functions of *cj0031* and its effect on different phenotypes during *in vitro* growth.

1.8: List of objectives
1. Determination and comparison of phase variation rates of both *cj0031* and *cj0628* (*capA*) genes.

2. Phylogenetic analysis of the *cj0031* gene


5. Assay for colonization, adhesion and invasion in eukaryotic cells by *cj0031* mutants and phase variants.


7. Assay for resistance by *cj0031* mutants and phase variants against phages.
Chapter 2. Materials and Methods

2.1: Bacterial strains and growth conditions:

Two *Campylobacter* strains were used in the present study: the motile but non-swarming strain NCTC11168 and a chicken adapted hypermotile version of the *C. jejuni* NCTC11168H. NCTC11168 strain was obtained from NCTC while the NCTC1168H was provided by Prof Julian Ketley. The *Campylobacter* strains were grown on Muller-Hinton agar (MHA) supplemented with vancomycin (10mg/ml) and trimethoprim (5mg/ml in 50% EthOH) in a VA500 Variable Atmosphere Incubator (Don Whitley, UK). The microaerophilic conditions maintained during the growth of the *C. jejuni* strains were listed as 4% oxygen, 10% carbon dioxide, and 86% nitrogen and at a temperature of 42°C. NCTC1168 and NCTC1168H were utilized for characterisation and phenotypic assays for *cj0031*. However, NCTC1168 was used by Christopher Bayliss (before I joined the project) for inoculation of chickens to check the phase variation in the repeat tracts of different (six) phase variable genes.

*E. coli* strains, DH5α and Rosetta (DE3), were used for cloning and expression purposes. They were grown at 37°C on Luria agar plates or in Luria broth containing the appropriate antibiotic. Ampicillin and Kanamycin were used at a 50 µg/ml concentration; however, Chloramphenicol was added to the growth medium at a 10 µl/ml concentration.
2.2: List of Primers used during the study:

The list of primers used during the present study can be found (Table 1)

### Table 1: The list of primers used during the study

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5’-3’)</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>lacZb-fam</td>
<td>[6FAM]tccccagtcagcaagttgt</td>
<td>Sequencing across polyG tract of <em>cj0031</em> in reporter construct</td>
</tr>
<tr>
<td>0031-for2108-pst</td>
<td>aactgcagaaattctcatcaggtag</td>
<td>Construction of knock-out mutant of <em>cj0031</em> and lacZ reporter construct</td>
</tr>
<tr>
<td>0031-rev2832-bam</td>
<td>cggtatccatcagtcgaattc</td>
<td></td>
</tr>
<tr>
<td>0032-rev-bam-kpn</td>
<td>cggtatccagctcagggcttggaagtg</td>
<td>Construction of knock-out mutant of <em>cj0031</em> and lacZ reporter construct</td>
</tr>
<tr>
<td>0032-for-bam</td>
<td>cggtatccagatagaaacgcattag</td>
<td></td>
</tr>
<tr>
<td>0032-rev-end</td>
<td>aagcacctgtctctaagtgag</td>
<td>Construction of knock-out mutant of <em>cj0031</em> and lacZ reporter construct</td>
</tr>
<tr>
<td>3544-for</td>
<td>tggcgtgaaggttgcgtggta</td>
<td>Construction of probes for Southern blotting to identify the methylation activity</td>
</tr>
<tr>
<td>4029-rev</td>
<td>gatccatcaacatccgcaac</td>
<td></td>
</tr>
<tr>
<td>425130-for</td>
<td>ttctctgctctgctctcg</td>
<td>Construction of probes for Southern blotting to identify the methylation activity</td>
</tr>
<tr>
<td>425852-rev</td>
<td>tacagcaggtatggcaagtg</td>
<td></td>
</tr>
<tr>
<td>1036450-for</td>
<td>ttaatctactactccagac</td>
<td>Construction of probes for Southern blotting to identify the methylation activity</td>
</tr>
<tr>
<td>1036809-rev</td>
<td>aagcaaaacactcttcattc</td>
<td></td>
</tr>
<tr>
<td><em>cj</em>0031-Nterm</td>
<td>caccatgtcatttcattttgtaaaag</td>
<td>Cloning of <em>cj0031</em> into expression vector</td>
</tr>
<tr>
<td><em>cj</em>0031-Cterm</td>
<td>gttcttccattttcccttcttg</td>
<td></td>
</tr>
<tr>
<td><em>cj</em>0031-comp-for</td>
<td>aaagctctcactgtcatttcttttctttcttcttcttcttg</td>
<td>Designing complementation of <em>cj0031</em> knock-out mutant</td>
</tr>
<tr>
<td><em>cj</em>0031-comp-rev</td>
<td>Tttccatggtctcattttcccttcttg</td>
<td></td>
</tr>
<tr>
<td>Probe IV-For</td>
<td>atatatcaaaaataaaacaattgagtg</td>
<td>Construction of probes for Southern blotting to identify the methylation activity</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>probe IV-Rev</td>
<td>caacaattgcaatcgccaaaag</td>
<td></td>
</tr>
<tr>
<td>PSR1-For</td>
<td>tgggaaagatcatactatcttttg</td>
<td></td>
</tr>
<tr>
<td>PSR1-Rev</td>
<td>gtttttaaatggtgtgccaaag</td>
<td></td>
</tr>
<tr>
<td>PSR2-F</td>
<td>tttataatcagcccaaatttatg</td>
<td></td>
</tr>
<tr>
<td>PSR2-Rev</td>
<td>ctttaaatcgtccatattgtaaac</td>
<td></td>
</tr>
<tr>
<td>PSR3-F</td>
<td>atattaaaaaccagattggaac</td>
<td></td>
</tr>
<tr>
<td>PSR3-Rev</td>
<td>ttaaacttttcttttggccgttaattc</td>
<td></td>
</tr>
<tr>
<td>PSR4-F</td>
<td>aagaaacactagccaaatattaag</td>
<td></td>
</tr>
<tr>
<td>PSR4-Rev</td>
<td>ataggtatggtttcgccaaatattg</td>
<td></td>
</tr>
<tr>
<td>PSR2-Hind-F</td>
<td>cctgataaattttaagtcttttg</td>
<td></td>
</tr>
<tr>
<td>PSR2-Hind-Rev</td>
<td>tttttagaacaatgaatgcaacc</td>
<td></td>
</tr>
<tr>
<td>PSR1-Hind-F</td>
<td>gttgcaaaagcaataaacctgaacctttaaatc</td>
<td></td>
</tr>
<tr>
<td>PSR1-Hind-R</td>
<td>cttgtaaaagacaagcttcataagaaattgc</td>
<td></td>
</tr>
<tr>
<td>PSR3-F4-Kpn</td>
<td>cgggtacctttatctcgtatcttagac</td>
<td></td>
</tr>
<tr>
<td>PSR3-R4-Bam</td>
<td>tggatccaagcaaacacattctctcact</td>
<td></td>
</tr>
</tbody>
</table>

### 2.3: Polymerase Chain Reaction (PCR):

All PCR reactions for the present work described in the later chapters were performed by preparing a 10 ul PCR reaction unless stated in the text (Table 3). The source and composition of the reagents for a typical PCR reaction were 10 x buffer (500 mM KCl, 100mM Tris-HCl (pH 8.3), 1mg/ml BSA); 100mM MgCl₂ , (Applied Biosystems),
reverse and forward primers (SIGMA-ALDRICH), 10mM dNTPs (Ultra Pure, Phamacia), Taq polymerase (5 U/ul; Applied Biosystems).

**Table 2: The composition of each amplification reaction**

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>Quantities (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>1</td>
</tr>
<tr>
<td>10 x PCR Buffer</td>
<td>1</td>
</tr>
<tr>
<td>25 mM MgCl₂</td>
<td>1.8</td>
</tr>
<tr>
<td>10 mM dNTPs</td>
<td>1</td>
</tr>
<tr>
<td>Primer mix (each primer at 2 µM)</td>
<td>1</td>
</tr>
<tr>
<td>TAQ DNA polymerase (5 units/µl)</td>
<td>0.1</td>
</tr>
<tr>
<td>dH₂O</td>
<td>4.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10</strong></td>
</tr>
</tbody>
</table>

PCR reactions were performed by using the Hybaid Omnigene Thermal Cycler. For PCR from a colony, the template DNA was prepared by suspending a single colony in 50 ul distilled water, followed by incubation at 100°C for five minutes. The mixture was centrifuged for three minutes and the supernatant was removed to be used as template DNA. The general PCR conditions for all amplification reactions involved initial
denaturation step (96\(^{\circ}\)C, 5 minutes), followed by 25-30 cycles of denaturation at 96\(^{\circ}\)C for 1 minute, annealing at 45-50\(^{\circ}\)C for 1 minute and elongation step at 72\(^{\circ}\)C for 1 minute (1 minute/1kb).

2.4: DNA Sequencing:

ABI PRISM™ technology (Perkin Elmer Applied Biosystems) was employed to carry out all sequencing reactions for the present work. The samples were handed in to PNACL (Protein and Nucleic Acid Chemistry Laboratory) which analysed them on an ABI PRISM™ 377 DNA sequencer.

The samples for DNA sequencing were prepared by following the protocol described in the ABI PRISM™ BIG DYEX Terminator Cycle Sequencing kit. The reaction contained 1 ul of DNA template, 1ul primer (from a 2 µM stock) and 8 ul of ABI PRISM™ BIG DYEX reaction mix and distilled water adding up to a 20ul final volume of reaction. The thermocycler conditions for sequencing reactions were 95\(^{\circ}\)C for 5 minutes (step1), 95\(^{\circ}\)C for 30 seconds (step 2), 50\(^{\circ}\)C for 30 seconds (step 3) and 60\(^{\circ}\)C for 4 minutes (step 4), and allowing the 25-30 cycles to be completed from step 2 to step 4.

The products thus obtained were purified by using a cartridge clean-up method. According to this method, 2ul 2.2% SDS was added to a 20 ul reaction, followed by incubation at 98\(^{\circ}\)C for five minutes. The cartridge was spun for 3 minutes at 850 x g (3100rpm) prior to addition of sample. The cartridge was transferred to a 1.5 ml eppendorf tube. The sample was added to the cartridge and spun for 3 minutes at 850 x g (3100rpm). The purified sample was stored at -20\(^{\circ}\)C prior to analysis.
2.5: DNA manipulations:

2.5.1: Large scale CsCl preparation of chromosomal DNA

This method was adapted from Sambrook et al (1989) with inclusion of an extra step of CsCl purification. This method is useful for the purification of the DNA for sequencing, restriction digestion and cloning. NCTC11168 strain was grown as a lawn on two MHA+VT plates. After an overnight growth, the cells were harvested from plates by adding and scraping cells into 2ml of MHB and pooled in a single tube. The cells in the tube were pelleted by centrifuging the culture at 10,000xg for 10 minutes. The cell pellet was resuspended in 10 ml of TES buffer (10mM Tris-HCL, pH, 7.6, 1 mM EDTA, 0.15 M NaCl). In order to perform cell lysis, 2ml of 20% SDS and 100µl of proteinase K (20 mg/ml) were added to the cell suspension, and subsequent incubation was performed at 37°C for 60 minutes. After cell lysis, the proteins were removed by the addition of an equal volume of chloroform/ethylalcohol in a 24:1 ratio. The mixing was carried out by gently inverting the tube for 5-10 times, and centrifuged at 10,000xg for 10-15 minutes. The top viscous layer was removed by using a plastic Pasteur pipette and transferred into a new tube.

Ethanol aided precipitation of DNA was performed by the addition of a 1/10th volume of 3M Na-acetate (pH 5.2) and 2 volumes of 95% ethanol. After precipitation of DNA, the glass Pasteur pipette was put in a flame to produce a hook on its head. The hooked Pasteur pipette was dipped in the solution and swirled slowly; the DNA spooled onto its hook. The hooked DNA was rinsed in 70% ethanol for a brief period of time. Excess ethanol was allowed to drain out of the DNA while briefly holding the Pasteur pipette.
The DNA was transferred into a tube containing 0.8 ml of dH2O or TE buffer and 0.8 ml of 3% N-lauryl sarcosine. The mixture was incubated at 4°C overnight, in order to redissolve the DNA. After dissolution of DNA, 1.76 g of CsCl: and 66µl of ethidium chloride (from a 10 mg/ml stock solution) were added and the CsCl was dissolved completely by incubating it briefly at 37°C and stirring the mixture with a magnetic stirrer. The mixture was transferred to a heat-sealable bell-top centrifuge tube (Beckman). The tubes were placed in the Beckman TL-100 benchtop ultracentrifuge; and centrifugation was performed for 4 hours (minimum) at 10,000 rpm at 20°C. This will cause the formation of a CsCl gradient. In order to remove the DNA, a wide-bore needle and syringe was used. DNA was transferred to an 1.5 ml eppendorf tube, ethidium bromide was removed by performing repeated extractions with H2O/CsCl saturated isobutanol. The DNA thus obtained was diluted with H2O and precipitated by the addition of 0.6 volumes of isopropanol. The DNA was pelleted by centrifugation for 10 minutes at 10,000xg. The washing of DNA with 70% ethanol was repeated twice to obtain highly pure DNA. The ethanol washed DNA was dried under vacuum and redissolved in 1000 µl of dH2O.

2.5.2: Small Scale Preparation of Chromosomal DNA:

Small scale preparation of chromosomal DNA was performed by employment of procedures described in the instruction manual of the DNeasy Blood and Tissue kit (Qiagen) for extraction of total DNA from bacterial cells. DNA obtained through this procedure was used for restriction digestion and Southern blotting assays. NTCTC11168 strain was grown on MHA+VT plates and incubated in the VAIN overnight. The cells were harvested after addition of MHB (2ml) onto the plate, and
transferred into an eppendorf tube. The cells were pelleted after centrifugation of the mixture at 10,000xg for 10 minutes. The pellet was resuspended in a solution containing 180 µl of buffer ATL and 20 µl of proteinase K. The mixture was vortexed and incubated at 56°C overnight to perform the complete cell lysis. The resulting mixture was vortexed for 15 seconds. 200 µl of Buffer AL was added to the mixture and mixed thoroughly by vortexing it, followed by addition of 200 µl of 96% ethanol and mixed again thoroughly by vortexing the mixture. The mixture was pipetted into the DNeasy Mini spin column placed in a 2 ml collection tube. The mixture placed in mini-spin column was centrifuged at 6000 x g for 60 seconds. Flow-through was discarded and 500 µl of Buffer AW1 was added into the column and centrifuged at 6000 x g. The flow-through was discarded. In the next step, 500 µl of buffer AW2 was added and centrifuged at 20,000 x g (14,000rpm). The flow-through and 2ml collection tube were discarded. The mini-prep spin column was placed in a new 1.5 ml microcentrifuge tube. The 200 µl of dH2O was added on to the spin column and centrifuged at 6000 x g for 60 seconds in order to elute ultra pure DNA.

2.5.3: Ethanol Precipitation

This technique was employed to obtain a highly concentrated DNA from samples having a lower concentration of DNA than the required one. This method was employed to obtain the high concentration of probes used for southern blotting. PCR samples (15-20 PCR tubes) of DNA fragments of interest were pooled into two or three tubes. A 1/10th volume of sodium acetate and 2.5 volumes of 96% ethanol were added to the samples. They were vortexed briefly and incubated for 20 minutes on ice. After that, these samples were centrifuged at 11000 x g for 5 minutes in the 1.5 ml microcentrifuge
tubes. The supernatant was discarded and the resulting pellet was dried in a vacuum or air dried. The appropriate volume of dH₂O (20 µl) was added to dissolve the pellet of DNA.

2.5.4: Restriction Enzyme Digestion:

All the restriction enzymes used in this study were purchased from New England Biolabs unless stated otherwise in the text. The concentrations of enzymes for digestion of genomic DNA was used in accordance with the manufacturer’s recommendations or optimized for achievement of a sufficient level of digestion. In general, restriction reactions contained the target DNA (0.5-1.0 µg), 1-2 ul of restriction enzyme (10-20 units), 2ul of the appropriate 10X buffer and dH_2O in a 20 ul digestion reaction; however, exceptions are stated in the text.

2.5.5: Estimation of DNA concentration:

DNA concentration was either estimated by running the sample on an agarose gel or by using the Nanodrop spectrophotometer (ND 1000). The calibration was done by loading the distilled water as diluents. 1ul of sample was loaded onto the pedestal and the concentration was measured in ng/ul.

2.6: Transformation of E. coli with plasmid DNA:

Two different kinds of transformation were employed based on the required level of efficiency of transformation. E. coli strain DH5 α was used for transforming with
plasmids. In order to transform the plasmids into *E. coli* strain DH5α, a simple calcium chloride method was utilised. However, for transformation of recombinant plasmids into NCTC11168 and NCTC11168H strains of *C. jejuni*, the electroporation method was favoured over the calcium chloride method.

### 2.6.1: Calcium Chloride Method:

*E. coli* DH5α strain was grown in LUB for overnight at 37°C. A 1ml of overnight culture was mixed with 100 ml of LUB with 10mM MgSO₄. The culture was incubated at 37°C for 1.5-2 hours until an OD600 of ~0.4-0.7 (preferably 0.4) was obtained. The culture was incubated on ice for 5 minutes followed by centrifugation for 15 minutes at 2000 x g at 4°C to form the cell pellet. The resulting cell pellet was resuspended in 20 ml of TFB1 (see the composition in section). The mixture was centrifuged again under the same conditions to pellet the cells. 4 ml of TFB2 (see the composition in section) was added to resuspend the cells, which was aliquoted into tubes having 100 µl of cell mixture and stored at -80°C for later use. 100 µl of competent cells were mixed with 10 µl of ligation mixture, and incubated on ice for 20 minutes. The mixture was heat-shocked for 90 seconds in a 37°C water bath. Following heatshock, the cells were incubated on ice for 5 minutes. The cells transformed with plasmid of interest were mixed with 900 µl of LUB and incubated for 60 minutes at 37°C in order to allow the expression of resistance determinants. Finally, cells were plated out on the LUA plates containing the required antibiotics.

### 2.6.2: Electroporation:

The *C. jejuni* strains (NCTC11168; NCTC11168H) were grown confluently on two MHA+VT plates. The cells were scrapped off from the plates and transferred to a tube and centrifuged at 4000rpm at 4°C for 20 minutes. The supernatant was removed and discarded, while the cells pellet was mixed with 1 ml of ice-cold wash buffer (272 mM sucrose and 15% (v/v) glycerol). The washing step was repeated twice. Finally, a cell pellet was obtained, which was resuspended in a 700 µl-1000 µl wash buffer before freezing 50 µl aliquots on dry ice. These aliquots were stored at -80°C. For electroporation, a 50 µl aliquot of electrocompetent cells was mixed with DNA/plasmid. The mixture was transferred to the pre-chilled electro-cuvettes (Biorad) and mixed properly by pipetting up and down. The cuvettes were placed in the Biorad Electroporation Pulse™ adjusted to 2.5kV (voltage), 200Ω (resistance) and 25 µF (capacitance). The cells were pulsed once with a time constant ~4.6. The electorporated cells were removed from cuvettes and mixed with 100 µl of MHB before plating them out on the non-selective MHA plates and subsequent incubation for a minimum of 5 hours or maximum overnight in the VAIN to allow the expression of antibiotic genes. The cells were harvested from the MHA plates with a 1ml MHB and spread on plates containing an appropriate antibiotic. The plates were incubated for 3-5 days in the VAIN.

### 2.6.3: Natural Transformation:

*C. jejuni* is a naturally competent microbe. That is why the natural transformation method was preferably used to transfer the genomic DNA from NCTC11168 to the NCTC11168H strain (hypermotile strain). For natural transformation, 11168 Δcj0031::kan was grown up confluently on MHA+kan plates. The cells were scrapped
off with a sterile spreader after addition of 2ml of MHB. The OD$_{600}$ was adjusted to 0.5 ($3 \times 10^9$ cells/ml) by diluting the cells with an appropriate volume of MHB. In order to determine the volume of culture which needed to be diluted to achieve OD$_{600} = 0.5$, the following formula was used:

$$\text{Volume of culture} = \frac{\text{Desired OD}_{600} = 0.5}{\text{Measured OD}_{600}} \times \text{dilution factor}$$

After adjustment of the OD of the culture, a 500 μl volume was transferred into a tube for incubation in the VAIN for 5 hours. A 3 μl sample of chromosomal DNA was added into the culture followed by overnight incubation in the VAIN. The culture was plated onto selective MHA plates and incubated in the VAIN for 3-5 days to obtain the transformants.

### 2.7: Colony immunoblotting:

This method was employed to determine the phase variation rate of *capA* in vitro (Weiser *et al.*, 1989). The *C. jejuni* strain NCTC 11168 was streaked onto the MHA plates and incubated for 3 three days in the VAIN. After 3 days, single colonies were picked by means of a loop and mixed with MHB to make serial dilutions in the order of $10^{-2}$, $10^{-3}$, $10^{-4}$, and $10^{-5}$. A 30 μl volume from each dilution was spread on the MHA plates. After 3 days, the number of colonies on each plate was counted. The plates with low dilutions were selected, which contained enough colonies to be counted easily. Usually the plates with colony counts in the range of 500-5000 were selected. The nitrocellulose filter paper was laid on the plate to cause the transfer of colonies onto it. Phosphate buffered saline (PBS) with 0.1% Tween 20 (PBST) and 10% dried milk...
(OXOID), a blocking buffer, was added and incubated for an hour. An additional 30 minutes incubation was performed after removal of excess colony material. After washing thrice in PBST, the filters were probed with a 1:500 dilution in blocking buffer of anti-CapA rabbit antiserum (primary antibody, a gift from Dr. Karl Wooldridge) for two hours under shaking conditions (Asghar et al., 2007). The washing step was repeated thrice with PBST in order to remove the unbound antibodies. A 1:2000 dilution in blocking buffer of a goat anti-rabbit antibody conjugated to alkaline phosphatase (secondary antibody, SIGMA-ALDRICH) was added to probe the filter and incubated for an hour. Filters were washed thrice again with PBST in order to remove the unbound secondary antibodies. The filter paper was developed with a commercial NBT/BCIP solution (PerkinElmer). The number of dark purple or light purple white phase variants was counted and the frequency was determined by dividing the number of phase variants by the total number of CFUs.

\[
\text{No. of Phase variants} \quad \frac{\text{Frequency}}{\text{No. non phase variants} + \text{No. Phase variants}}
\]

2.8: Phase Variation Assays:

The \emph{C. jejuni 0031-lacZ} reporter construct was grown on MHA plates. After three days of incubation in the VAIN, single colonies were picked and serial dilutions were made in the range of $10^{-2}$ to $10^{-5}$ in Mueller-Hinton broth (MHB) before plating them onto
MHA plates supplemented with a 1:1000 dilution of a 10mg/ml X-gal and IPTG solution (Melford). These plates were subjected to 3-4 days of incubation in the VAIN. The total number of colonies on each plate was counted and recorded. The total number of blue colonies in the case of OFF-to-ON phase variation was counted. Similarly, the total number of white colonies in the case of OFF-to-ON phase variation was enumerated. The number of white or blue phase variants on each plate was divided by the total number of colonies on the relevant plate in order to calculate the frequencies. From these frequencies phase variation rates were derived by employing the method described by (De Bolle et al., 2000).

2.9: Analysis of PV genotypes for C. jejuni:

Six phase variable genes from NCTC1168 were selected to determine the genotypes. The oligonucleotides were designed to span across the repeat tract region of these genes. Thus a total of 6 pairs of oligonucleotides were made, and one of each pair was labelled with a fluorescent dye (FAM, VIC, NAD). The names of these oligonucleotides and the resulting product sizes are presented (Table 3).
Table 3: The names of these oligonucleotides and the resulting product sizes

<table>
<thead>
<tr>
<th>Names of Oligos</th>
<th>Product Sizes (bps)</th>
<th>Colour</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{capA} (For)-FAM,</td>
<td>458</td>
<td>Blue</td>
</tr>
<tr>
<td>\textit{capA} (Rev)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{cj1139} (For)-Vic,</td>
<td>244</td>
<td>Green</td>
</tr>
<tr>
<td>\textit{cj1139} (Rev)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{cj0031} (For)-FAM,</td>
<td>220</td>
<td>Blue</td>
</tr>
<tr>
<td>\textit{cj0031} (Rev)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{cj1326} (For)-FAM,</td>
<td>165</td>
<td>Blue</td>
</tr>
<tr>
<td>\textit{cj1326} (Rev)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{cj0045} (For)-FAM,</td>
<td>280</td>
<td>Blue</td>
</tr>
<tr>
<td>\textit{cj0045} (Rev)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{cj0685} (For)-NED,</td>
<td>241</td>
<td>Black</td>
</tr>
<tr>
<td>\textit{cj0685} (Rev)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The boiled lysates of colonies were prepared and the repeat tracts were amplified by using two or three pairs of primers. The composition of each amplification reaction is shown in table 2. All steps of the amplification reaction were performed under the conditions (Table 4)
Table 4: Thermocycler conditions for performing an Amplification reaction

<table>
<thead>
<tr>
<th># steps</th>
<th>Steps</th>
<th>Temperature (°C)</th>
<th>Time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial Denaturation</td>
<td>94</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Denaturation</td>
<td>94</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Annealing</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Elongation</td>
<td>72</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Go to 2 REP 25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.9.1: A-Tailing:
The amplification reaction mixture was subjected to an A-tailing reaction. The reaction mix for A-tailing contained the following ingredients and conditions of incubation in thermocycler as given in Table 5.

Table 5: Ingredients for an A-tailing reaction for PCR product

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>Quantities (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 x PCR Buffer</td>
<td>0.4</td>
</tr>
<tr>
<td>25 mM Mgcl₂</td>
<td>0.4</td>
</tr>
<tr>
<td>TAQ DNA polymerase (5 units)</td>
<td>0.05</td>
</tr>
<tr>
<td>dH₂O</td>
<td>3.15</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
</tr>
</tbody>
</table>
The resulting reaction mixture was incubated for 45 minutes at 72°C. These A-tailed PCR products were analysed by using the GeneScan assay.

2.9.2: GeneScan Assay:

This assay was performed as described by (De Bolle et al., 2000). According to this method, the A-tailed PCR products were diluted in the ratio 1:10 or 1:5. A 9.25 µl of de-ionized formamide and 0.25 µl of GS500LIZ were mixed with 0.5 µl of diluted A-tailed PCR products. GS500LIZ was used as size standard from Applied Biosystems. These preparations were added in each well of 96-well GeneScan plates. The samples were electrophoresed on an ABI3700. The resultant files were analysed through Peak Scanner (Applied Sciences).

The peak Scanner showed multiple peaks for each sample. The highest peaks were selected and taken as representatives of sizes of samples (in base pairs) depending on the relative areas of highest peaks and of flanking peaks. Thus the sizes of the tract lengths were estimated.

2.9.3: DNA sequencing of Samples:

The samples were selected randomly in order to sequence them across the repeat tract of the relevant genes. The lengths of repeat tracts determined through the sequencing were correlated with the sizes of the tract lengths obtained through GeneScan analysis.
2.10: Chicken Colonization Experiments:

*Campylobacter* strain NCTC11168H was grown up on the sheep blood agar for 3 days under micro aerobic conditions in the VAIN. A cotton swab was used to take a sweep of *Campylobacter* cells from the plate prior to inoculating MHB with that sweep. Inoculated MHB was grown overnight at 37°C for preparation of the inoculum. The overnight MHB culture was used to inoculate the chickens through the oral route. Two weeks old chickens were provided by PD Hook, UK. They were fed with 0.1 ml of *Campylobacter*-free gut flora which was prepared as described by Jones *et al* (2004). Different dilutions of the inoculum were prepared and grown on *Campylobacter* selective blood free agar plates (Oxoid). After incubation of three days, 30-60 colonies were picked from these plates along with colony counts and immunoblots prepared from these plates. The remaining cells of the inoculum were concentrated by centrifugation and were utilized in preparation of boiled DNA lysates. After one day, some of the chickens were sacrificed and caecal samples were collected. Later on, the rest of chickens were culled at 14 days post-inoculation for collection of the caecal samples. Caecal materials collected from both events were resuspended in PBS. The mixture was vortexed properly to mix them thoroughly prior to plating different dilutions of caecal material onto *Campylobacter* selective blood free agar plates (Jones *et al*., 2004). A total of 30 colonies for each bird were picked from the relevant plates and subjected to a similar analysis as for the inoculum.
2.11: *C. jejuni* interactions with cultured human epithelial cell lines:

2.11.1: Maintenance of Tissue Culture Cells:

Caco2 cells were obtained from Dr Richard Haig, Department of Genetics, University of Leicester. Caco2 cells were grown in DMEM (with GlutaMAX™, 4500 mg/L D—glucose, Sodium Pyruvate: GibcoBRL) along with 20% fetal bovine serum (GibcoBRL). The cells were grown in 75 cm² flasks under standard growth conditions for tissue culture (37°C in a humidified 5% CO₂ incubator). The growth was continued until 70% confluency was achieved. After that, cells were washed twice with PBS gently so that the cell monolayer should not be peeled off, prior to adding 2 ml of trypsin. Cells were separated from the monolayer and divided into single cells by incubating them at 37°C and banging the side of the flask against the palm of the hand. This process may be repeated twice or thrice to dislodge all of the monolayer to single cells. After dislodgment of cells from the monolayer, a 10 ml aliquot of pre-warmed media (DMEM/FCS) was added immediately for resuspension of cells. A 3 ml aliquot of the suspension was added into 4 flasks, followed by addition of 30 ml of pre-warmed media (DMEM/FCS) to top up the culture. The whole preparation was incubated at 37°C under humidified 5% CO₂. The cells were kept and fed with 25 ml of DMEM/FCS (medium) per flask after every 2 days depending on the change of medium colour from red to yellow and until the growth of cells reached approximately 70% confluency (approximately two weeks). Once they had become confluent (2-3 days), they start differentiating (which may take 5-7 days) with appearance of domes and elaboration of microvilli characteristic of enterocytes. Only differentiated Caco2 cells can be used for
bacterial-cells interaction (Pinto et al., 19983). Differentiated cells were washed twice with PBS buffer gently prior to addition of 2 ml of trypsin and incubation at 37\(^{0}\)C for 5 minutes. The process of dislodging the cells from the monolayer was repeated as before. Once cells were dislodged from the monolayer, cells were suspended in 10 ml of DMEM/FCS medium (pre-warmed) and transferred to a 15 ml tube for spinning down at 700 x g for 5 minutes. The supernatant was discarded, followed by suspension of cells in freshly made FBS + 10% DMSO, which were used for adhesion and invasion assays. For storage of cells, they were stored first at -20\(^{0}\)C in a polystyrene box followed by storage at -80\(^{0}\)C for one hour prior to transferring of tubes to liquid N\(_2\) for long term storage (maximum 5 years).

2.11.2: Adhesion Assay:

*Campylobacter* strains 11168H-\(\Delta\)0031::kan, 11168H-0031-ON (G9), and 11168H-0031-OFF (G10) were grown on fresh MHA with antibiotics (if applicable) plates in the VAIN. The MHB (with no antibiotic) was inoculated with a sweep of cells from plates and incubated in the VAIN under shaking conditions overnight. The OD600 was adjusted to 0.5 (~2 x 10\(^9\) cells) by concentrating the cells by centrifugation and diluting them with MHB as required prior to adding 10 µl (1 x 10\(^7\)) into wells containing a Caco-2 monolayer at the bottom. The wells were provided with fresh pre-warmed DMEM + 1% FBS and incubated for 30 minutes at 37\(^{0}\)C before challenging them with *C. jejuni* 11168H strains. Cultures were incubated for 4 hours in the CO\(_2\) incubator. Media were removed with an aspirator and washed gently twice with 2 ml of PBS/well. After that, 1ml of mixture containing PBS and 1% Triton X100 were added to each well, followed by incubation for 10 minutes on the bench to allow the cell lysis of
Caco-2 cells. Lysed cells were carefully mixed by pipetting the liquid up and down, followed by serial dilutions (10^0, 10^1, 10^2) with PBS in the 96-well plate. Dilutions were made in triplicate. A 50 µl aliquot of each dilution was plated onto freshly made MHA plates divided into three sectors (1/3). Plates were incubated in the VAIN for 2-3 days. CFUs were counted and the number of adherents to Caco-2 cells was estimated.

2.11.3: Invasion Assay:

Caco-2 cells were grown in a 24-well tissue culture plate with 50% confluency and were allowed to grow until the confluency covers the bottom of the well. The media was replaced with fresh DMEM + 1% FBS, and was given a 30 minutes incubation at 37°C in the CO2 incubator. The strains under investigation were grown and processed as described above. The media were removed by aspirator and the Caco-2 monolayer was challenged with 2 x 10^7 cells of the strains used for this experiment, followed by incubation for 4 hours at 37°C. The monolayer of Caco-2 cells was washed gently twice with 2 ml of PBS. Fresh media containing 200 µl gentamycin was added to the wells and incubated for another 4 hours at 37°C in the CO2 incubator. The media was removed by aspirator and cells were washed gently thrice with PBS. For cell lysis, PBS containing Triton X-100 was added in each well and mixed well by pipetting the liquid in each well. The plates were incubated on the benchtop for 10 minutes. The number of CFUs was counted by plating serial dilutions of each strain (50 µl) onto fresh MHA plates. Assays were performed in duplicate.
2.12. Crystal Violet staining method for measurement of biofilm formation

The method is the slightly modified version of a methodology used by Reeser et al. (2007). The *C. jejuni* strains were grown in Mueller-Hinton broth (MHB; Difco) at 37°C overnight under shaking conditions. The optical density of bacterial cells at 600nm (OD$_{600}$) was adjusted to 0.25. The MHB was inoculated with an overnight culture and the optical density (OD$_{600}$) was adjusted to 0.025 (~ 2.5 x 10$^7$ CFU). 1ml of culture having OD (0.025) was added in to each well of 24-well polystyrene plates (Corning). The plates were incubated in VAIN (10% CO$_2$ and 37°C) for 48 hours and 72 hours. Following incubation, the media was discarded and washing of wells was performed twice with water. The wells were dried by incubating plates at 55°C for 30 minutes. 1m of 0.1 % Crystal violet was pipetted into each well, and incubated for 10 minutes at room temperature. The crystal violet solution was discarded from wells and washed twice with water to remove the unbound CV, following that wells were dried at 55°C for 15 minutes. The bound CV was decolorized with addition of 1 ml of 80%-ethanol-20% acetone solution. A 100 ul aliquot of this solution was taken out from each well and placed in a separate 96-well plate. In order to measure the amount of biofilm mass, the absorbance at 570nm (A$_{570}$) was determined by using a micro-plate reader (BMG-Labtech).

2.13. Motility Assay:

The motility assay was performed with *C. jejuni* strains in order to test their motility phenotype, by using 0.4% Muller-Hinton Agar (MHA; Difco). Bacterial cells were harvested from MHA plates after 36 hours incubation in VAIN and suspended into
MHB. The optical density at 600nm was adjusted to 0.45 (~ \(1 \times 10^9\) CFU/ml). Following that, 2μl of bacterial suspension was stabbed into motility agar, and plates were incubated in VAIN for 48 hours and 72 hours.

### 2.14: Southern Blotting Assay for determination of methylation activity of Cj0031:

#### 2.14.1: Amplification of Probes:

Five probes having lengths in the range of 450-900 bp were amplified by using the primers described in the list of primers. The PCR steps and conditions employed for the amplification of probes are given in the table:

**Table 6: The PCR steps and conditions employed for the amplification of probes**

<table>
<thead>
<tr>
<th># steps</th>
<th>Steps</th>
<th>Temperature (°C)</th>
<th>Time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial Denaturation</td>
<td>94</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Denaturation</td>
<td>94</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Annealing</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Elongation</td>
<td>72</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Go to 2 REP 30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The PCR products obtained were concentrated to 1 µg/20 µl by using the ethanol precipitation as described in section “Ethanol Precipitation”, which was required for efficient detection of target hybrid molecules.

2.14.2: Labeling of DNA:

A 16 µl aliquot of probe DNA was mixed with 4 µl of dH₂O to make the total reaction volume 20 µl. The reaction mixture was boiled for 10 minutes for denaturation and quickly placed on ice. Complete denaturation was necessary for efficient labelling. A 4 µl aliquot of DIG-High prime solution was mixed with denatured probe, and incubated at 37°C overnight. The labeling reaction was stopped by adding 2 µl of 0.2M EDTA (pH 8.0)

2.14.3: Determination of Labelling Efficiency:

All the incubations steps were carried out at room temperature unless stated otherwise.

In order to determine the labelling efficiency, labelled probes and control DNA were serially diluted as described in the kit. A 1 µl volume from each dilution was spotted on a nylon membrane (Osmonics Inc.) which was placed in the UV crosslinker set at 700 nm to fix the nucleic acid to the membrane, followed by transfer of the membrane into a container having 20 ml of maleic acid buffer (0.1 M Maleic acid, 0.15 M NaCl; pH (7.5) adjusted with NaOH). The membrane was incubated for 2 minutes under shaking
conditions. Maleic acid buffer was replaced with 10 ml of blocking solution (DIG labelling kit II; Roche Applied Science) which was used to block the non-specific binding sites on the membrane, and incubated for 30 minutes. Following that, 10 ml of an anti-DIG antibody solution was added and incubated for 30 minutes, followed by washing the membrane twice with washing buffer for 15 minutes each. The membrane was equilibrated for 2-5 minutes in 10 ml of detection buffer. The membrane was sandwiched between two pieces of cling film after application of CSPD-ready-to-use solution. The membrane was exposed to X-ray film for 1-2 hours to check the labelling efficiency by comparing the signals from labelled probe and control DNA.

2.14.4: DNA Transfer and Fixation:

Genomic DNA from 11168-0031-ON (G9), 11168-0031-OFF (G10), 11168-Δ0031::kan and 11168-complΔ0031::cat were digested with various methylation sensitive restriction enzymes by following the procedure described in section "Restriction digestion of Plasmids and genomic DNA". 15 µl of restriction digestion mixtures were run on a 1% agarose gel for 2 hours at 80V. An image of the gel was taken by placing it in a UV illuminator. A scale was also placed along a side of the gel so that the distance between fragments could be measured later on.

The fragments of genomic DNA were transferred from the gel to a positively charged membrane (Osmonic Inc.) by following the protocol described by Sambrook et al (1989). A 20 x SSC was used as transfer buffer. The incubation time for transferring the fragments efficiently from gel to nylon membrane was 16-20 hours.
The fragments of genomic DNA transferred to membrane were fixed by UV-crosslinking the wet membrane. After UV cross linking, the membrane was washed briefly in distilled H₂O.

2.14.5: Hybridization Reaction:

Optimum hybridization temperatures for each probe were determined using the following formula:

\[ T_m = 49.82 + 0.41 \times \%GC - \frac{600}{L} \]  
\[ T_{opt} = T_m - 20-25^0C = 41-42^0C \text{ for all the probes} \]

For hybridization reactions, 20 ml of DIG Easy Hyb (prewarmed at hybridization temperature) was added in a glass container containing the membrane for prehybridization, under agitating conditions.

DIG labelled DNA probes were denatured by boiling for 5 minutes and rapidly cooling by placing in ice prior to addition of denatured probes to pre-warmed DIG-Easy Hyb. Probe/hybridization mixture was transferred to a container carrying membrane and incubated at the optimal hybridization temperature overnight with gentle agitation.

2.14.6: Stringency Washes:

The membrane was washed twice for 5 minutes each with ample 2 x SSC, 0.1% SDS at room temperature under constant agitation. Following that, the membrane was washed twice for 15 minutes each at 61^0C under constant agitation conditions.
2.14.7: Immunological detection:

After hybridization and stringency washing, the membrane was washed briefly with washing buffer for 5 minutes prior to its incubation with 10 ml blocking solution for 30 minutes. Following that, the membrane was incubated with 20 ml of anti-DIG antibody solution for half an hour. The membrane was washed twice for 15 minutes each in an ample volume of washing buffer, followed by equilibration of the membrane in 20 ml of detection buffer for 5 minutes. CSPD ready-to-use solution was applied and spread across the membrane. After that the membrane was sandwiched between two pieces of cling film. Extra liquid was squeezed out and edges were sealed. The membrane was exposed to X-ray film and incubation was continued until the visible signal intensity was detected on X-ray film.

2.15. Procedures for performing phage assay:

2.15.1. Plaque forming ability of phages on C. jejuni host strains:

The plaque forming ability assay was adopted from Frost et al. (1999). The C. jejuni strains were subcultured on MHA and incubated at 42°C in VAIN under microaerobic conditions for 24 hours. Bacterial cells were scrapped off plates and suspended into 10ml of 10mM MgSO₄. A 500ul aliquot of cell suspension adjusted to cell density (~10⁹ CFU/ml) was added into pre-warmed NZCYM basal agar, and poured on to the NZCYM plates, followed by 20 min incubation of plates at room temperature for solidification of the overlay agar. The plates were kept for 10 minutes in VAIN, followed by spotting of 10 ul of diluted phage stock (10⁶ PFU/ml) on the lawns of
plates. The plates were incubated in VAIN for 24 hours. The degree of lysis of host strain by phages was recorded.

2.15.2. Pseudolysogenic Bacteriophage Propagation

The pseudolysogenic *C. jejuni* strains - PT14CP8PL, PT14CP30PL, HPC5CP8PL, HPC5CP30PL - were subcultured on MHA plates and incubated at 42 °C under microaerobic conditions in VAIN overnight. The *C. jejuni* cells were removed from the plates and suspended into 10ml of 10mM MgSO$_4$. A 500 µl aliquot of cell suspension was pipetted into 4ml of pre-warmed agar, and poured on to the lawn of NZCYM plates. After solidification, the plates were incubated in VAIN for 24 hours. The *C. jejuni* cells were scrapped and suspended into a 5ml of SM buffer and incubated at 4°C for overnight under shaking condition. The *C. jejuni* phages were separated from bacterial cells and purified by filtration using 0.2 µm membrane filter. Phage titration was carried out by following the method described in the following section.

2.15.3. Bacteriophage Titration

The titration of phages was performed by using the method of Frost et al (1999) with a slight modification. The propagating *C. jejuni* strains were cultured on MHA plates, and incubated in VAIN for overnight. The bacterial cells were harvested into 10ml of 10mM MgSO$_4$. The cell density was adjusted to $10^9$ CFU/ml. A 500 ul aliquot of cell suspension of the appropriate propagating strain was added into 4ml of molten overlay agar, mixed properly and poured on to level surface of NZCYM plates. After allowing the plates to solidify, the plates were incubated for 30 minutes at 37 °C for drying.
Serial dilutions of the phages ranging from $10^1$ to $10^6$ were prepared by taking 20 ul of undiluted phage into 180 ul of SM buffer, followed by spotting 10 µl of each serial dilution in triplicate on to the surface of NZCYM plates. Plates were incubated in VAIN for 24 hours. The number of plaque forming units per ml (PFU/ml) of undiluted phage suspension was calculated.

2.16. Statistical Analysis:

Statistical analysis for data obtained from adhesion/invasion, biofilm formation and motility assays was formed by using Graphpad Prism version 5. A paired t-test was used to determine the $p$ values and the level of significance between two pairs of a data set.
Chapter 3. The *In vivo* and *in vitro* examination of polymorphism in the phase variable genes of *C. jejuni*

3.1. Introduction:

The prevalence of simple sequence repeats (SSR) or homopolymeric tracts in the genomes of several pathogenic bacteria are known to be having played an important role in the adaptation of these bacteria to various niches. Many pathogens contain phase variable genes containing repeat tracts in their ORFs or promoter regions; and are known to control their expression. Examples of such pathogens include *N. menigitidis*, *Yersinia pestis*, *H. influenzae*, *Bordetella pertussis*, *H. pylori* and *C. jejuni*. The homopolymeric tracts, if present in an ORF, can mediate the reversible switching between ON and OFF phases by causing the frameshift mutations via slipped strand mispairing. This phenomenon is termed as phase variation. Phase variation is mediated by simple sequence repeats due to their high reversible switching and mutation rates. The simple sequence repeats may be composed of mononucleotide repeats, dinucleotide repeats or tetra nucleotide repeats. The phase variation rates and mutational spectra for many of the genes containing varying number of repeat type and lengths had been characterized in pathogens. However, the implication of phase variation associated with repeats containing genes in the pathogenesis of organisms remains an ongoing debate in the field of molecular biology.

Although frequencies of mutations have been determined for many phase variable loci, the mutational patterns associated with these mutational frequencies have not been thoroughly explored especially in *C. jejuni* whose genome contains 27 homopolymeric repeat tracts consisting of 8-12 runs of G residues (Parkhill, 2000). The phase variation
of these repeat tracts is reported to be high but there is dearth of data showing the accurate measurement of the phase variation rates and frequencies (Karlyshev et al., 2005; Linton et al., 2000; Wassenaar et al., 2002). The accurate measurement of phase variation rates and mutational patterns is important in relation to understanding their role in enhancing the genetic diversity within a bacterial population and for facilitating adaptation to environmental fluctuations, as shown by the modelling of phase variation in H. influenzae (De Bolle et al., 2000). We assume that the hypermutability of C. jejuni strains, arising from the polyG repeat tracts, accounts for the high PV rates. The characterization of these aspects of C. jejuni populations will enhance our understanding of the role and function of SSR-mediated phase variation as an adaptive strategy in the fluctuating environments of the hosts colonised by this species. Besides, the determination of mutational patterns and PV rates will also increase our insight into what are the predominant phase variation states in the host environment and what factors lead to their emergence and maintenance in the host environment.

Allelic variations of contingency loci in C. jejuni resulting from phase variation during in vivo passage have been described by Kim et al. (2012) and Wilson et al (2010). They have shown that alterations in some contingency loci cj0045, cj0685 and cj1139 were found in high abundance after passage through mice. Jerome et al (2011) came up with similar conclusions after their analysis of contingency genes in C. jejuni isolates subject to passage through mice. They found variations in the tract lengths of seven phase variable genes including cj0031, cj1325, and cj1295. There is also experimental evidence for PV of the flgR gene whose short polyA tract length was altered during selection for the motility required by C. jejuni to colonize chickens (Hendrixson, 2006). This indicates that the demands of the host environment may be the driving force behind changes in repeat tract length for the phase variable genes of C. jejuni. But there
is scarcity of evidence about the PV-driven changes in the polyG/C repeat tracts in *C. jejuni*. We hypothesize that variation in the polyG/C tracts of phase variable genes, driven by phase variation, are selected for during adaptation to the host.

In all the studies, researchers attempted to compare the frequency of genotypes of populations passaged through chickens with those passaged through mice. It is interesting to speculate that phase variable genes (*cj0045, cj0031, cj1325/26, cj0685, cj0628 (capA) and cj1139*) may undergo a major shift in their frequencies during in vivo passage through chickens, resulting in the prevalence of one phase or other in comparisons between the output and input populations (inoculum). This can lead to the discovery of the phenotypes/genotypes propagated at a high frequency within bacterial populations trying to adapt this host. Functions of these genes are poorly understood and most of them have putative functions of encoding surface antigens or modifiers of surface antigens except *cj0031* which is a putative type IIS restriction modification system. The phase variable *capA* gene is required for colonization of chickens by *C. jejuni* (Ashgar *et al.*, 2007). However, Coward *et al.* (2008) reported varying abilities of isogenic strains to colonize the chicken gut, which suggests that these strains may have had different expression states for the phase variable genes required for colonization.

In order to test the above hypotheses, chickens were inoculated with *C. jejuni* NCTC11168H strain. After that, the inoculum and caecal samples were analysed. The frequencies of variants within both samples were determined and compared in order to find out the predominant or modal allele for the genes under investigation. The colony immunoblotting and lacZ reporter construct tools were employed to measure the phase variation frequencies and PV rates for *capA* and *cj0031* in vitro.
3.2. Results:

3.2.1. Examination of the changes in the phase variable genes of *C. Jejuni* during *in vivo* Passage in chickens

When I started my PhD project, there was an ongoing project of examination of variations in repeat tract lengths of *C. jejuni*. Chicken infection studies and sample collection had already been conducted by Christopher Bayliss in Nottingham. I analysed the in vivo variation in repeat tract length of *cj0031, cj0045, capA, cj0685, cj1139, cj1325/26* genes in birds: - B7, B8, B9, B10, B11. Following that in vitro determination of in vitro phase variation frequencies, mutational spectra and mutation rates for *cj0031* were determined by me. For *capA*, I analysed 12 colonies for G10 and 15 colonies for G12 and 12 colonies for G11 repeat tracts to determine their switching frequencies and phase variation rates in the OFF-to-ON switching direction. Some of phase variation rates were previously determined by Fadil Bidmos, an MSc student of Christopher Bayliss. In order to present consistent and coherent results exhibiting the overall picture of *in vivo* variations of repeat tracts in phase variable genes and switching rates of *capA*, I included the data analysed for the inoculum and B6 by Rebecca Richards (a BSc student) and some of data pertaining to analysis of *capA* phase variation rates (8 colonies analysis for G12 and 2 colonies analysis for G11) from Fadil Bidmos [an MSc student].

Ten two-week old chickens were challenged with high bacterial dose (1×10⁸ CFU) of a hypermotile strain of *C. jejuni* (NCTC11168H). Boiled lysates of colonies prepared from dilutions of the inoculum were prepared and preserved at -80°C. After passage of one day, five birds were sacrificed and caecal samples were collected from their
caecum. When these samples were analysed for any discernible *C. jejuni* population, no growth was detected. This indicated that the NCTC11168H strain has a low level of initial colonization. The other five chickens were sacrificed after two weeks of inoculation. The caecal contents were recovered from these birds and examined for the presence of a *C. jejuni* population.

The persistence of high numbers of bacterial cells was detected, which ranged from $1.6 \times 10^7$ to $4 \times 10^8$ CFU/g caecal contents. This suggested a high degree of colonization and persistence of NCTC11168H strain in the alimentary canal of these chickens. It was hypothesized that high level of persistence and colonization of *C. jejuni* in the guts of these chicken resulted from phase variation mediated by the simple sequence repeats of some of the phase variable genes. The phase variable genes containing simple sequence repeats within them are known to play an important role in adaptation of *C. jejuni* to the host (Wassenaar *et al.*, 2002; Jerome *et al.*, 2011, Kim *et al.*, 2012).

Parkhill *et al.* (2000) first documented 27 genes containing polyG repeat tracts. However, only six genes having polyG homopolymeric tracts, *cj0031/cj0032, cj0685 (cipA), cj0628 (capA), cj1139, cj0045, cj1325/26*, were chosen for the present study, because there was scarcity of data showing the changes in repeat tract lengths of these genes during passage through chicken. Changes in repeat tract lengths and frequency of the alleles of contingency loci were determined by performing a GeneScan analysis which utilizes the principle of “sizing the PCR product”.

### 3.2.2. Amplification and fragment analysis of repeat tract lengths of six phase variable loci
Inoculum and caecal contents from all birds were subject to serial dilution and dilutions were grown on campylobacter-selective plates. Thirty colonies were picked from the appropriate dilution of the inoculum while 23-29 colonies were picked from dilutions of caecal samples of all birds. The DNA was isolated from samples of the inoculum and the caecal contents by applying a DNA isolation procedure.

Oligonucleotides, flanking the repeat tract regions of these six genes, had been designed (Bayliss, personal communication). The repeat tract lengths of six genes were readily amplified from thirty colonies derived from caecal samples of each bird and inoculum by using different sets of primers. Each primer set included at least one fluorescently labelled primer (usually the forward primer). Either 2-primer mixes or 3-primer mixes were used for economy and convenience. These primers amplified their cognate regions, resulting in the production of fluorescently labelled amplicons. The PCR products of these genes were run on the gel to check the size and purity of amplicons as shown in Figure 1.
Figure 1 Amplification of genomic DNA from bird samples by using primer mix (capA and cj0031)  PCR products obtained from amplification of genomic DNA from B11 and B9 bird samples by using a primer mix for capA and cj0031; the product size of 220bp obtained from cj0031 primers and 457bp obtained from capA primers. Lane 1: standard molecular marker, Lane 2: B11.1; Lane 3: B11.2, Lane 4: B11.3; Lane 5: B11.4, Lane 6: B11.5, Lane 7: B11.6; Lane 8: B9.1, Lane 9: B9.2, Lane 10: B9.3

Taq DNA polymerase adds a non-templated Adenine (A) residue on the 3 prime end of amplicons. However, this can create problems in determination of length changes especially in cases where the changes at mononucleotide tract are being detected. So in order to overcome this problem, an A-tailing reaction was performed to append A residues on both sides of PCR amplicons in order to produce a uniform product. A-tailed PCR products derived from 23-30 colonies of each bird and inoculum were run, as part of a service run by PNACL, on a denaturing acrylamide gel using an ABI3300 Autosequencer. Each amplicon showed the characteristic peak profile, which consists of

400bp
200bp
three peaks, two minor (left and right) and one major (middle). The major and minor peaks are shown in figure 2 for two different amplicons.

**Figure 2: Screenshot of a plot obtained from a GeneScan analysis of the amplicons for cj0031 and cj0045 from sample B9.14.** The orange peaks represent peaks arising from the fluorescently-labeled GS500LIZ size standard. The blue peaks represent the 220 bp product obtained for cj0031 using FAM-labelled PCR primers. The size obtained denotes a repeat tract of G9. The green peaks represent the 241bp product obtained for cj0685. The peak sizes obtained from GeneScan were confirmed by sequencing across the repeat region.

The sizes of major peaks and minor peaks were recorded and the area under each peak was also recorded to infer the ratio between major and minor peaks. The size of the largest peak was defined as a major peak and this corresponds to the major repeat type present in the sample, while the other two peaks were termed as minor peaks. The minor peaks may arise from replication slippage during the PCR reaction. Nevertheless, these minor peaks may be present as representatives of minor proportions of these repeat types in the sample. Ratios between major and minor peaks were used as an indicator for detection of the major and minor repeat types. Should the ratio between the major and minor peak of a sample be low (i.e. less than 5), it means that the sample carries a mixture of both kinds of variants. Low ratios of this type were observed in B10 samples for genes cj0045 and cj0686.
A sub-set of the PCR fragments were sequenced and this data was used to assign a repeat tract length against each peak size; for all of the six genes. These approximations have been shown in the Table 1.

Table 1: Presentation of ON repeat numbers for six genes, and their equivalent major peak sizes obtained from GeneScan analysis

<table>
<thead>
<tr>
<th>Genes</th>
<th>ON-Repeats</th>
<th>Peak Size</th>
<th>OFF-Repeats</th>
<th>Peak Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>capA</td>
<td>G11 (blue)</td>
<td>457.8</td>
<td>G12</td>
<td>458.8</td>
</tr>
<tr>
<td>cj1139</td>
<td>G8</td>
<td>243.5</td>
<td>G9</td>
<td>244.5</td>
</tr>
<tr>
<td>cj0031/cj0032</td>
<td>G9</td>
<td>220.4</td>
<td>G10</td>
<td>221.6</td>
</tr>
<tr>
<td>cj0045</td>
<td>G10</td>
<td>280.7</td>
<td>G11</td>
<td></td>
</tr>
<tr>
<td>cj0685</td>
<td>G9</td>
<td>241.5</td>
<td>G8</td>
<td>240.8</td>
</tr>
<tr>
<td>cj1325/26</td>
<td>G9</td>
<td></td>
<td>G10</td>
<td>165.2</td>
</tr>
</tbody>
</table>

The repeat tract lengths of all six genes in samples from birds B5-B11 (23-29 samples in each bird) were determined through peak sizes in GeneScan analysis. These repeat lengths were compared with each other within all of the birds (5) and with those derived from samples of the inoculum. The results of this data are described in the following sections.
3.2.3. Detection of a high frequency of OFF-to-ON switching in \textit{cj0031}

The \textit{cj0031} gene encodes a putative type IIS restriction modification system. A remarkable shift from G10 (in the inoculum) to G9 (in the samples from birds) was detected (Figure 3). A G9 repeat tract causes the fusion of ORFs \textit{cj0031} and \textit{cj0032} because it allows the translation to continue from \textit{cj0031} into \textit{cj0032}. Whereas a G8 or G10 repeat tract does not allow translation into \textit{cj0032}. The CDSs for both genes were merged and reannotated as \textit{cj0031}. A high frequency of G9, reaching 93\%, was recorded in B7 (30 samples), B8 (27 samples), B10 (19 samples) and B11 (30 samples). While in B6 and B8, the frequencies of ON variants were 70\% and 80\% respectively. Contrastingly, inoculum carried only 17\% of G9 variants.

![Figure 3: A graph showing the frequency (percentage) of variants of G9, G10 and G11 in the inoculum and output population of six birds: B6, B7, B8, B9, B10 and B11. The number of variants of each polyG tract length for \textit{cj0031} were determined from GeneScan](#)
analysis. The percentage frequency was determined as follows: number specific allele for a gene divided by total number of alleles present in samples for that gene. The frequencies were obtained by analysis of samples from inoculum and birds: Inoculum (30 samples), B6 (30 samples), B7 (30 samples), B8 (26 samples), B9 (30 samples), B10 (19 samples), and B11 (28 samples). Blue column: G9 repeat tract; Red column: G10 repeat tract; Yellow column: G11 repeat tract

3.2.4. Switching from OFF to ON detected for *cj0045* during in vivo passage:

The frequency of ON repeat tracts of *cj0045*, a putative iron binding protein, showed high frequency during in vivo passage (Figure 4). The inoculum was split between two genotypes: G9 (43%) and G10 (57%) but none of a G11 (ON) genotype. Following in vivo passage, the colonizing population became enriched with G10 (OFF) variants going as high 90% in B6 and 93% in B11, and 73% in B10. However, B8 and B9 showed variation as compared to other birds because the frequency of G10 variants went as low as 40% in both birds. A novel genotype G11 (ON) emerged during in vivo passage though it was not present in the inoculum. The results suggest that both alleles G10 and G11 of *cj0045* existed in varying proportions in the colonizing population of all of the birds. The results are shown in the figure below.
Figure 4: A graph showing the frequency (percentage) of variants of G9, G10, G11, and G12 in the inoculum and output populations of Birds: B6, B7, B8, B9, B10 and B11. The data was recorded after counting the number of alleles of polyG of *cj0045* determined from GeneScan analysis. The percentage frequency was determined by (No. specific allele for a gene divided by total number of alleles present in samples for that gene). The frequencies were obtained by analysis of samples from inoculum and birds: Inoculum (30 samples), B6 (28 samples), B7 (24 samples), B8 (27 samples), B9 (30 samples), B10 (26 samples), and B11 (29 samples); Blue column: G9 repeat tract; Red column: G10 repeat tract; Yellow column: G11 repeat tract.

3.2.5. Bird-to-bird variations detected in frequency of ON/OFF variants of *cj0628* (*capA*):  
The reiterated tract length of G12 in the capA gene switched (OFF) to G11 (ON) during in vivo passage in some of the birds. The inoculum carried G11 and G12 in the proportions of 38% and 35% respectively. The frequency of G11 (ON) variants was increased to 67%, 76% and 80% in birds B7, B8 and B10 respectively.
Figure 5: A graph showing the frequency (percentage) of variants of G9, G10, G11, G12 and G13 in ancestral inoculum and output population of Birds: B6, B7, B8, B9, B10 and B11. The data was recorded after counting the number of alleles of polyG of *cj0628 (capA)* determined from GeneScan analysis. The percentage frequency was determined by (No. specific allele for a gene divided by total number of alleles present in samples for that gene). The frequencies were obtained by analysis of samples from inoculum and birds: Inoculum (29 samples), B6 (28 samples), B7 (30 samples), B8 (25 samples), B9 (29 samples), B10 (15 samples), and B11 (28 samples). Blue column: G9 repeat tract; plum red column: G10 repeat tract; cream white column: G11 repeat tract; sky blue column: G12 repeat tract; purple column: G13 repeat tract.

3.2.6. Switching from OFF to ON was observed for *cj0685 (CipaA)* during in vivo passage:

CipA is an invasive protein which is required for colonization during mouse infection model (Lynett, 1999). After in vivo passage, a high level of switching from G8 to G9
was found indicating that insertion of single nucleotide in the polyG tract of \textit{cj0685} was prevalent (Figure 6). This gave rise to the enrichment of an ON (G9) phenotype. In the inoculum the frequency of G8 and G9 were 97\% and 3\% respectively. After in vivo passage, switching from a G8 to a G9 genotype resulted in the shift of frequencies in the output population such that the frequency of G9 reached a level of 100\%, in birds B7, B9 and B11 whereas in birds B6, B8 and B10 the frequencies of G9 were found to be 86\%, 93\% and 96\% respectively. These results provide evidence for the selection of G9 (ON) against the G8 (OFF) genotype, indicating that an ON phenotype might be required for enhanced colonization and adaptation to the host.

\textbf{Figure 6:} the graph shows the frequency (percentage) of variants of G8, G9 and G10 in ancestral inoculum and output population of Birds: B6, B7, B8, B9, B10 and B11. The data was recorded after counting the number of alleles of polyG of \textit{cj0685 (cipA)} determined from GeneScan analysis. The percentage frequency was determined by (No. specific allele for a gene divided by total number of alleles present in samples for that gene). The frequencies were obtained by analysis of samples from inoculum and birds: Inoculum (30 samples), B6 (28 samples), B7 (24
samples), B8 (27 samples), B9 (30 samples), B10 (26 samples), and B11 (29 samples). Blue column: G8 repeat tract; plum red column: G9 repeat tract; cream white column: G11 repeat tract

3.2.7. High Frequency of ON-to-OFF switching of *cj1139 (wlan)* detected after passage:

*Cj1139 (Wlan)*, a β 1,3 galactosyltransferase, showed reversion from ON to OFF (Figure 7). In the inoculum, the frequency of G8 (ON) tract was recorded as 83% and G9 (OFF) was found to be only 7%. However, after in vivo passage, the frequency of variants of G9 increased to 76%, 90%, 97% and 100% in B6, B11, B7, B9 and B8 respectively. This result suggested that selection against an ON phenotype had happened during in vivo passage:

![Figure 7: A graph shows the frequency (percentage) of variants of G7, G8, G9, and G10 in ancestral inoculum and output population of Birds: B6, B7, B8, B9, B10 and B11: The data was recorded after counting the number of alleles of polyG of *cj1139 (wlan)*](image)
determined from GeneScan analysis. The percentage frequency was determined by (No. specific allele for a gene divided by total number of alleles present in samples for that gene). The frequencies were obtained by analysis of samples from inoculum and birds: Inoculum (30 samples), B6 (28 samples), B7 (24 samples), B8 (26 samples), B9 (30 samples), B10 (27 samples), and B11 (29 samples)

3.2.8. Minor changes in repeat tract length of gene *cj1325/26*:

A 100% frequency of G10 variants was found for *cj1325/26* in bird B9 with similar levels of 89%, 94% and 85% in output samples from B8, B9 and B10, respectively (Figure 8). In contrast the inoculum contained G8 variants with a frequency of 26% and G10 with a frequency of 70%. The frequency of G8(ON) variants decreased to zero in B9 and B10 and to 3-4% in B6, B7, B8, and B11, suggesting the disappearance of G8(ON) genotype during passage through chicken. Nevertheless, a G10 repeat tract length was prevalent in both the inoculum and the output populations from all birds.

![Figure 8: A graph showing the frequency (percentage) of variants of G9, G10, G11, and G12 in ancestral inoculum and output population of Birds: B6, B7, B8, B9, B10, and B11.](image)
**B10 and B11.** The data was recorded after counting the number of alleles of polyG of *cj1325/26* determined from GeneScan analysis. The percentage frequency was determined as follows: number of specific allele for a gene divided by total number of alleles present in samples for that gene. The frequencies were obtained by analysis of samples from inoculum and birds: Inoculum (30 samples), B6 (28 samples), B7 (24 samples), B8 (26 samples), B9 (30 samples), B10 (27 samples), and B11 (29 samples)

### 3.2.9. Assessment of the validity of GeneScan analysis of repeat tract length through sequencing

Some samples having varying tract lengths from Bird7, Bird8, Bird9 and Bird10 and Bird 11 were picked and sequenced in order to determine the repeat tract length in the respective genes and to check whether the repeat tract lengths obtained through GeneScan analysis were correct or not. The results obtained through sequencing of all samples, except *cj0031*, were consistent with the GeneScan results (see Table 2), indicating that GeneScan results were not an artifact of the GeneScan analysis.

Some samples were analysed twice by GeneScan to confirm the conservation of peak patterns. A consistency was found in the peak patterns for both GeneScan analysis runs, which confirmed the reliability and reproducibility of data obtained through GeneScan analysis.

**Table 2 Comparison of the repeat tract lengths of *C. jejuni* genes as determined by GeneScan and sequence analyses**
### Table 3: Conversion of repeat tract lengths into ON/OFF genotypes for six genes

<table>
<thead>
<tr>
<th>Samples</th>
<th>Gene</th>
<th>Sequenced Repeats</th>
<th>GeneScanned Repeats (First run)</th>
<th>GeneScan Repeats (Second run)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>B9.7</td>
<td>cj0045</td>
<td>G10</td>
<td>280.8</td>
<td>280.9</td>
<td>Same</td>
</tr>
<tr>
<td>B9.15</td>
<td>cj0045</td>
<td>G11</td>
<td>281.6</td>
<td>281.8</td>
<td>Same</td>
</tr>
<tr>
<td>B9.20</td>
<td>cj0045</td>
<td>G10</td>
<td>280.7</td>
<td>280.9</td>
<td>Same</td>
</tr>
<tr>
<td>B11.13</td>
<td>cj0628</td>
<td>G10</td>
<td>242.6</td>
<td>242.8</td>
<td>Same</td>
</tr>
<tr>
<td>B11.14</td>
<td>cj0628</td>
<td>G9</td>
<td>241.9</td>
<td>241.7</td>
<td>Same</td>
</tr>
<tr>
<td>B11.27</td>
<td>cj0628</td>
<td>G10</td>
<td>242.9</td>
<td>242.9</td>
<td>Same</td>
</tr>
<tr>
<td>B11.2</td>
<td>cj0031</td>
<td>G10</td>
<td>221.5</td>
<td>221.7</td>
<td>Same</td>
</tr>
<tr>
<td>B11.7</td>
<td>cj0031</td>
<td>G9</td>
<td>220.3</td>
<td>220.3</td>
<td>Same</td>
</tr>
<tr>
<td>B11.13</td>
<td>cj0031</td>
<td>G9</td>
<td>220.5</td>
<td>220.6</td>
<td>Same</td>
</tr>
</tbody>
</table>

#### 3.2.10. Significant Changes in ON/OFF Genotypes detected during in Vivo Passage:

Following the analysis of changes in repeat tract lengths of all of the six genes — cj0031, cj0045, cj0685, cj1325/26, capA and cj1139; the ON and OFF genotypes for these genes were determined from genome of *C. jejuni* by using Artemis sequence visualization and annotation software (Table 3).
<table>
<thead>
<tr>
<th>Repeat lengths</th>
<th>ON/OFF genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>cj0031</td>
<td>cj0045</td>
</tr>
<tr>
<td>G8</td>
<td>-</td>
</tr>
<tr>
<td>G9</td>
<td>ON</td>
</tr>
<tr>
<td>G10</td>
<td>OFF</td>
</tr>
<tr>
<td>G11</td>
<td>OFF</td>
</tr>
<tr>
<td>G12</td>
<td>-</td>
</tr>
</tbody>
</table>

In order to obtain the relative frequencies, the number of ON or OFF genotypes derived from GeneScan data was divided by total number of ON and OFF genotypes present in all samples analysed for each bird for each gene. A Chi-Squared test was applied to each data set in order to determine whether significant changes had occurred between input and output samples for each bird and for each gene. The changes in the genotypes of all of the six genes were analysed. Significant fluctuations were observed in the genotypes of input populations (inoculum) and the output populations (caecal samples) during in vivo passage through chickens for genes *cj0031*, *cj1139* and *cj0685*.

For *cj0031*, a massive switch from OFF (G10) in the inoculum to ON (G9) in output samples was detected in all birds. Thus the inoculum carried 11% ON variants and 89% OFF variants whilst the frequency of ON (G9) variants in the outputs had increased to 93% for birds B7, B8, B10 and B11. This increase from OFF (G10) to ON (G9)
switching frequencies in output samples in comparison with input samples was found to be more than 8-fold.

The switching from OFF to ON variants of *cj0031* was significant with p values of <0.0001 obtained in a Chi-squared test for each bird as compared to the inoculum using Graphpad prism version 5. Similarly, a significant change from ON to OFF was detected for *cj1139* in all of the birds involved in this study. In the inoculum, the relative frequencies of ON and OFF variants of *cj1139* were 82% and 18% respectively.

Following infection of chickens, the ON genotype was replaced by the OFF phenotype with birds, B7, B8, B9 and B11, containing the OFF (G8) genotype at almost 100% whilst 77% samples from B6 displayed the OFF genotype. A p value of <0.0001 (obtained from Chi-squared test for these genes) was obtained for these values indicating a significant change from the inoculum genotype in the proportions of OFF variants for this gene. Similarly, *cj0685* switched from an OFF (G8) to an ON (G9) genotype, and demonstrated a significant change (p value <0.0001 obtained from Chi-squared test for that gene) when output samples were compared to the inoculum.

The analysis of genotypes demonstrated bird-to-bird variations in the G10/G11 variants for *cj0045* in birds B8 and B9 (p values of <0.001 were obtained for comparisons with data from B6 and B11). Bird-to-bird variations were also observed in the case of *capA*, and this was noticed in B7 and B8 (p values of <0.01 were obtained for B7 versus B11, B8 versus B9 and B8 versus B11 and 0.04 for B7 versus B9. Furthermore, the G11 for *cj0045* emerged as a novel genotype in B7, B8 and B9, where it constitutes a frequency of 33%, 57% and 52% of the total output populations from these birds.

So, three genes (*cj0031*, *cj0685*, and *cj1139*) manifested significant changes in the output samples with a switch from one phase (genotype) to another phase as compared
to the input samples of the inoculum. However, \emph{capA} and \emph{cj0045} exhibited significant bird-to-bird differences between the input and output samples whilst \emph{cj1326} exhibited no significant difference for changes in its repeat tract length.

### 3.2.11. Differences in the Distribution of the Combined Genotypes for All Six Genes among Birds:

Before going on to present the results pertaining to the distribution of genotypes, it is paramount to understand the nomenclature used for the genotypes and phenotypes of all of the six genes involved in this study. The nomenclature places “1” for every ON genotype and “0” for every OFF genotype of phase variable genes. Furthermore, the six genes were listed in each combined genotype in the following order: - \emph{cj1326-cj0031-cj1139-cj0685-cj0045-capA}. A total of 22 genotypes were detected in the output populations, whereas, nine genotypes were found in the inoculum (input population).

Only three of the nine inoculum genotypes - 0-0-0-0-0-1, 0-0-1-0-0-1 and 1-0-1-0-0-1 - were found in the output populations. Furthermore only the B6 output population was found to be carrying these three genotypes, while these genotypes were not detected in the output populations of B7, B8, B9 and B11.

Thirteen out of 22 genotypes were novel genotypes which had evolved within the colonizing population during the course of adaptation and colonization of the chicken gut. These results suggest that evolution of novel genotypes plays an important role in helping the bacterial populations of \emph{C. jejuni} to colonize the birds effectively. However, these genotypes did not all have an equal importance for the colonization as the frequencies of some genotypes in these birds were detected as negligibly low (i.e. <4\%) [Table 4].
Table 4: Distribution of major genotypes among the inoculum and birds

<table>
<thead>
<tr>
<th>Genotypes/Birds</th>
<th>Frequencies of Genotypes (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inoculum</td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-1</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
<tr>
<td>0-0-0-1-0-0</td>
<td></td>
</tr>
</tbody>
</table>

Variations between two related genotypes were noted while analysing the bird to bird variations in the frequencies of genotypes. The 0-1-0-1-0-0 variant was found at 50% in B6 but at 17%, 17%, 34% and 54% in B7, B8, B9 and B11 respectively. Parallel differences in the level of variant 0-1-0-1-0-1 were detected with levels of 12%, 38%, 26%, 7% and 19% in B6, B7, B8, B9 and B11 respectively. This is important as these genotypes represent the opposing ON/OFF phenotypes for capA. The variation in the level of variant 0-1-0-1-1-1 was also detected, which was found to be at 4%, 17%, 39%, 21% and 8% in birds B6, B7, B8, B9 and B11 respectively. This genotype (0-1-0-1-1-1)
represents capA ON/OFF and cj0045 ON/OFF phenotypes when this is viewed in combination with the other two genotypes (0-1-0-1-0-0, 0-1-0-1-0-1). Thus the differences in levels of these three genotypes, 0-1-0-1-0-0, 0-1-0-1-0-1, 0-1-0-1-1-1, were detected which represent the opposing phenotypes (ON/OFF) of capA and cj0045.

### 3.2.12. Distribution of ON phenotypes of genes among birds

During *in vivo* passage of *C. jejuni* through chickens, certain genotypes (ON/OFF) of phase variable genes may be subjected to selection for ON or OFF phases. This can enhance the colonization and adaptation potential of *C. jejuni* in the micro-environment of a new host. The ON phenotypes for *cj1326* and *cj1139* decreased during the in *vivo* passage in relation to the input populations. This decrease was detected in all of the birds under study. In contrast, a trend for an increase in ON phenotypes for *cj0031* and *cj0685* was found in B6, B7, B8, B9 and B11 in comparison with the input population. Significant bird-to-bird differences in the level of ON-variants were observed for *cj0045* and capA (p <0.001).
Figure 9: Changes in the proportions of ON genotypes for genes cj1326, cj0031/cj0032, cj1139, cj0685, cj0045, capA in the output populations from six birds inoculated with C. jejuni. Birds - B6, B7, B8, B9, B11 - were inoculated with C. jejuni strain 11168. The frequencies of ON genotypes were derived from the data obtained through GeneScan analysis of 23-30 samples from the inoculum and output population from B6-B11.

These results suggest that ON variants of cj0031 and cj0685 are subject to positive selection in the colonizing population of the C. jejuni, while the OFF phenotypes of cj1326 and cj1139 were selected against ON, resulting in an increase of the OFF variants of these genes in the bacterial population colonizing the chicken gut. In the case of cj0045 and capA, variations in the level of ON variants were found possibly indicating that some birds preferably promoted the ON phenotypes over the OFF phenotypes in relation to these genes.

3.2.13. Determination of phase variation rates in vitro:

Major changes in the frequencies of phase variants were observed for capA and cj0031 during in vivo passage of C. jejuni through chickens. The repeat tract length for cj0031 switched from G10 to G9. A low level of reversion from G10 to G11 but a high level of switching from G12 to G11 were observed for the capA gene. Based on these observations, we hypothesized that repeat tract length affected the phase variation frequency of contingency loci such as capA and cj0031 during in vivo growth. It was also hypothesized that longer repeats were more prone to deletion rather than insertion in comparison with shorter repeat tract lengths (G8-G10) which were subjected to insertion more often than deletion.

Moreover, it was important to determine the numbers of mutations occurring per cell division at the contingency loci i.e. the phase variation rates. The phase variation rates
could not be determined \textit{in vivo} directly due to lack of molecular tools. So, there was a need for the development for strategies to determine the phase variation rates in vitro for these genes. These rates could then be used to simulate phase variation \textit{in vivo} for a bacterial population propagating and colonizing a host.

The \textit{capA} phase variation rate was determined \textit{in vitro} through colony immunoblotting. The aim of this experiment was to estimate the influence of tract length on the switching frequencies and rates, and to determine the mutational patterns for various lengths of repeat of \textit{capA}.

In order to determine the mutational patterns and phase variation rates, NCTC 11168 strain was grown on MHA agar plates supplemented with vancomycin and trimethoprim, Some parental colonies were picked and the repeat tract length was determined to be G11 through sequencing across the repeat region. Various dilutions ranging from $10^{-1}$ to $10^{-5}$ were prepared and spread on MHA+VT plates. The total number of colonies on each dilution was counted. Confluent growth was found on the $10^{-1}$ and $10^{-2}$ dilution plates but countable colonies were found for dilutions $10^{-3}$, $10^{-4}$ and $10^{-5}$ dilutions. These dilutions were employed for colony immunoblotting experiments by following a standard protocol as described in Material and Methods chapter section (2.7) using anti-CapA antibodies raised in rabbit against the CapA proteins as a primary antibody to detect phase variants expressing the CapA surface protein. The pink dots resulting from reaction of \textit{capA} (ON) variants with a CapA-specific polyclonal sera were obtained during OFF-to-ON switching direction; and white dots arising from capA (OFF) variants were obtained during analysis of ON-to-OFF switching direction on nitrocellulose membrane (Figure 10)
Show of CapA Phase Variation (ON \rightarrow OFF) On Colony Blot

- CapA-ON (starter)
- Mixed Colony
- CapA-OFF Variant

Display of CapA ON Variation
Figure 10: A colony immunoblotting blot for the capA ON-to-OFF and OFF-to-ON switching directions.

Panel A: blot obtained from ON-to-OFF switching direction of capA by using G11 (ON) as starter colonies; shows the OFF variants, mixed variants and capA-OFF starter colonies; Panel B: blot showing the analysis of capA from OFF-to-ON switching direction by using G12 (OFF) as a starter colony, display the presence of capA (OFF) variants used as tarter colonies, mixed colonies representing mixture of capA ON/OFF variants, capA-ON variants.

3.2.13.1. Determination of Average Frequency of Phase Variants and Frequency Medians:

For ON-to-OFF switching with G11 colonies, 12 colonies were analysed, while 11 colonies and 15 colonies were analysed for the OFF (G10)-to-ON and OFF (G12)-to-ON switching directions respectively. Frequencies were measured for each colony and then used to determine the median frequency for the G11, G10 and G12 genotypes. The phase variations rates were estimated from these median frequencies by using the equation proposed by Drake (1991). The confidence intervals for G10, G11 and G12 were calculated according to Koskoska et al. (1998). The values of median frequencies, phase variation rates and confidence intervals for capA G10, G11, and G12 are presented in Table 4.

Table 5: Phase variation rates of Campylobacter jejuni CapA
## Relevant Genotype  Direction of Switching

<table>
<thead>
<tr>
<th>Gene</th>
<th>ON-to-OFF</th>
<th></th>
<th></th>
<th>OFF-to-ON</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Repeats</td>
<td>Colonies</td>
<td>Freq. (× 10⁻³)</td>
<td>Rates (× 10⁻⁴)b</td>
<td>Repeats</td>
<td>colonies</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>capA</em></td>
<td>G11</td>
<td>14*</td>
<td>18.44</td>
<td>16.41</td>
<td>G10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>{3.36}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[11.2-30.9]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G12</td>
<td>23**</td>
<td>32.59</td>
<td>29.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>{6.04}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 2 colonies analysis was taken from the analysis done by Fadil Bidmos

**8 colonies analysis was derived from OFF-to-ON switching analysis conducted by Fadil Bidmos

bPV rates estimated by using PV equation of Drake (1991), number in curly brackets indicate the fold increase over shortest tract, the numbers in square brackets indicate the 95% confidence interval calculated by using the method developed by Koskoska et al., (1998)

The results in the above table indicated that phase variation rates increased with increasing repeat tract length. A 2.3 fold increase in the phase variation rate for 11168-capA-G10 was detected relative to PV rate of 11168-cj1139-G10 LacZ reporter construct (2.15 x 10⁻⁴). The phase variation rate for 11168-capA-G10 (4.9 x 10⁻⁴) was higher than that of 11168-cj1139-G10 LacZ reporter construct (3.67 x 10⁻⁴ mutations/division) observed for LacZ reporter construct of cj1139 (Bayliss et al., 2012). This was also higher than PV rates observed for MC58-spr-C10 LacZ reporter
construct (1.2 x 10^{-5} mutations/division) in *N. meningitides* strain MC58 (Martin *et al.*, 2004).

A 3.36 fold increase in the phase variation rate for 11168-capA-G11 was observed relative to 11168-capA- G10. The 11168-capA-G12 variant exhibited a 6.04 fold increase in the phase variation rate relative to 11168-capA-G10. These differences in the mutation rates while going from 11168-capA-G10 to 11168-capA-G12 indicated that mutation rates varied as a function of repeat tract length (Figure 1).

![Figure 11: Relationship between repeat length and PV rates (capA)](image)

A graph shows that the phase variation rates vary as a function of repeat tract length. The phase variation rates for various tract lengths of *capA* are: G10 = 4.88 x 10^{-4} mutations per division, G11= 16.41 x 10^{-4} mutations per division, and G12 = 29.49 x 10^{-4} mutations per division. The phase variation rates were determined from median frequencies by following method of Drake (1991)

### 3.2.13.2. The mutation patterns of *C. jejuni* phase variable genes are the function of repeat tract length

For the analysis of repeat tract length changes for the ON-to-OFF switching experiment, white colonies representing the OFF genotype were picked from the respective plates of
growth and re-streaked on MHA+VT plates. Boiled lysates preps were prepared from these colonies and the starting colony. These lysates analysed through PCR and GeneScan to determine any change in the repeat tract length of the \textit{capA} gene. Most \textit{capA} variants showed the expected change in repeat tract length (see below), but in some of the samples tract length was found to be the same as the parental colonies. These samples were categorized as false positive and might have arisen due to picking the wrong colonies because the colonies on some plates were packed very close to each other. The lengths of homopolymeric tract of \textit{capA} determined through GeneScan analysis were confirmed by sequencing of a random set of samples prepared from OFF variants.

As the GeneScan is a very sensitive technique, which detects changes of even a single nucleotide. The analysis of variants through GeneScan provided data which showed both mononucleotide insertion (+1) and deletion (-1) at repeat tract of \textit{capA} gene. The data showed that mutational spectrum of G12 repeat tract of \textit{capA} showed only deletions (-1) while there was no insertional events (+1). This resulted in a contraction of repeat tract length. During ON-to-OFF switching, the repeat tract length G11 (ON phenotype) of \textit{capA} showed a mutational spectrum comprising of both insertions (+1) and deletions (-1). However, -1 deletional mutations (63%) were more significant than the +1 insertional mutations (38%). Only insertional events (+1) were detected in the mutational spectrum for G10 repeat tract length of \textit{capA} during OFF-to-ON switching, suggesting that ON variants were generated due to the gain of single repeat unit (+1) during the switching from OFF-to-ON. Of note, we had detected that frequency of +1 insertional events decreased as the tract length of \textit{capA} increases, and reverse was true for -1 deletional events at the repeat tract of \textit{capA}. The mutational spectra for G10, G11, G12 of \textit{capA} are shown in the figure 12.
Figure 12: mutational spectra for G10, G11, and G12. G10 undergoes +1 insertional mutations while G12 is subjected to -1 deletional mutations only. The repeat tract G11 mutates in both direction from G11 to G10, and from G11 to G12, showing -1 and +1 mutational patterns. However, the -1 deletions are prominent during G11 switching rather than +1 insertions

3.2.14. Construction of a *cj0031*-LacZ reporter construct

The phase variability of cj0031 was established in the in vivo experiments. This gene was selected for further studies which were aimed at checking the phase variation rate of *cj0031* in vitro. A *cj0031*-LacZ reporter construct was designed to estimate phase variation rates for *cj0031 in vitro.*

A primer set flanking the polyG repeat tract of *cj0031* was designed and used to clone fragment *L0031-up.* A second primer pair was designed to clone fragment *S0032-down,*
which includes a region downstream of the repeat tract and was required for recombining the vector into the *C. jejuni* genome. A *LacZ* gene lacking an initiation codon was derived and purified from *cj1139-lacZ* cat plasmid and was then fused with the repeat tract of *cj0031* in order to detect the changes in the expression of *lacZ* due to changes in the repeat tract of *cj0031*. A chloramphenicol cassette was obtained from pAV-cat plasmid and used for selection of recombinant clones. These fragments were employed to construct a *cj0031-LacZ* reporter construct containing a cat cassette.

The primers for *cj0031/0032* were designed to flank the region containing the repeat tract. The primer upstream of the repeat tract was carrying a PstI restriction site on the 5’-end whilst two sites, BamHI and KpnI, were on the 5’-end of the reverse primer. As the *lacZ* fragment obtained from digestion of the *cj1139-lacZ* vector (with BamHI and KpnI) contained BamHI and KpnI sites on its 5’-end and 3’-end, respectively (Figure 13).

**Figure 13: Fusion of LacZ-fragment near poly G region of *cj0031***

The figure was derived from pUC0031-LacZ-cat clone of *cj0031* reporter construct in the clone manager program; the solid green line shows the sequences of L0031-up fragment containing repeat tract and
S0032-down fragment; the dotted green line displays the sequence of LacZ; red line indicates the position of polyG in cj0031; position of KpnI and BamHI on 3-end and 5-end, respectively.

The fragment was amplified with pfu DNA polymerase and ligated into pGEMT-easy vector, which was transformed into *E. coli* strain (DHα5). The white colonies containing recombinant plasmids were picked and checked for the presence of an insert by using an *EcoRI* digest. The plasmids containing the inserts of the correct size were sequenced which verified that the right insert was present. A double digestion of the recombinant plasmid was performed with *PstI* and *BamHI* to recover the insert having a size of 935bps.

A fragment from 3’-end of 0032 gene was amplified with a forward primer carrying a *BamHI* site on its 5’-end and a reverse primer with no restriction site. This fragment was ligated into pGEMT easy vector and afterwards transformed into *E. coli* strain (DHα5). The white colonies were picked and a mini-preparation of the plasmid was made to get the high quality of plasmid. The plasmid obtained in this way was digested with enzymes *EcoRI* and *BamHI* to confirm insertion and to recover the insert. The resulting insert obtained from this digestion had a size of 609bp and carried a *BamHI* sticky end on its 5’-end and an *EcoRI* sticky end on its 3’-end. Moreover, sequencing of the plasmid by using forward and reverse primers of the insert was performed, which further confirmed the cloning of the correct insert.

A recombinant plasmid of pUC19 was constructed carrying the two inserts S0031 and M0031. The detailed description of the construction process for this plasmid will be illustrated in the next chapter under section “Construction of cj0031 Mutant”. S0031 was needed along with L0031 fragment to construct the 0031-LacZ reporter construct. Therefore, the pUC19-S0031-M0031 was used to proceed with the construction of the reporter plasmid. A digestion of pUC19-S0031-M0031 was performed with KpnI and
BamHI to release the M0031 fragment and a linear fragment pUC19-S0031 with KpnI and BamHI sticky ends. The clone of the L0031 fragment was digested with BamHI and KpnI to produce the compatible ends with pUC19-S0031. Following ligation and transformation, the resulting clones were digested with BamHI/EcoRI, which yielded the expected fragment sizes: 3400bp, 609bps, and 242bps (Figure 14). The clones were also digested with BamHI and PstI in a double digestion reaction in order to obtain the fragments having sizes of 935bps and 3300bps. A third diagnostic digest was performed with PstI and EcoRI, and expected fragments of sizes of 700bps, 851bps and 2700bps.

Figure 14: Digestion of six cj0031 lacZ reporter construct clones with BamHI, EcoRI and PstI

A 5μl of plasmid construct was digested with respective enzyme (s) in a total 20 ul reaction volume; 10μl of digestion mixture was loaded on to 1% agarose gel. Lane 1 and 14: standard molecular weight marker; Lanes 2, 5, 8, 11, 15, 18: BamHI/PstI double digest for six clones; Lanes 3, 6, 9, 12, 16, 19:
BamHI/EcoRI double digest for six clones; Lanes 4, 7, 10, 13, 17, 20: PstI/EcoRI double digest for six clones.

3.2.14.1. Digestion of pUC19-S0031-L0031 and Ligation with CAT cassette:

The CAT cassette with sticky ends of *BamHI* was obtained by digesting pAV35-CAT plasmid with *BamHI*. The pUC19-S0031-L0031 plasmid was digested with *BamHI* and KpnI. A full-length *LacZ* gene was recovered from pUC19-1139-LacZ by performing a digestion reaction with *KpnI* and *BamHI*. A three way ligation reaction was performed involving the pUC19-S0031-L0031 (cut with *BamHI* and *KpnI*) and CAT (cut with *BamHI*) and *LacZ* (cut with *BamHI*/KpnI). Following transformation, 20 colonies were picked and tested for the presence of the relevant fragments. Digestion reactions with various relevant restriction enzymes including single and double digestion reactions were carried out, which confirmed the successful production of pUC0031-LacZ-Cat construct. BamHI and KpnI double digest yielded the expected fragment sizes: - 3528, 4195, 834. The second diagnostic tested was performed with BamHI and EcoRI double digest, which generated the desired fragments having sizes: 3316bps, 3348bps, 814bps, 609bps and 450bps. Similarly third test to verify the *LacZ* reporter construct clones involving the double digestion of clones with BamHI and PstI also certified the clones by providing the fragments sizes: - 4463bps, 3260bps, 935bps (Figure 15).
Figure 15: Digestion of six cj0031 lacZ-cat reporter construct clones with BamHI, EcoRI and PstI

A 5ul of plasmid construct was digested with respective enzyme(s) in a total 20 ul reaction volume; 10ul of digestion mixture was loaded on to 1% agarose gel. Lane 1 and 14: standard molecular weight marker; Lanes 2, 5, 8, 11, 15, 18: BamHI/KpnI double digest for six clones; Lanes 3, 6, 9, 12, 16, 19: BamHI/EcoRI double digest for six clones; Lanes 4, 7, 10, 13, 17, 20: PstI/BamHI double digest for six clones.
A confirmatory PCR was performed on pUC19-0031-LacZ-cat construct by using the LacZB1-R and 0031-F primers. The PCR gave the expected fragments of 1100bps (Figure 16).

**Figure 16: PCR product using LacZB1 and cj0032-R primers**

A 10ul of PCR product using the LacZB1 and cj0032-R primers obtained; a 5ul of PCR product loaded on to 1% agarose gel; Lane 1: standard molecular weight marker; Lane 2, 3, 4, 7: PCR products from five pUC0031-LacZ-cat construct clones; Lane 6: blank sample

Ligation reaction may yield two products: pUC19-0031-LacZ-cat (with reverse orientation) and pUC19-0031-LacZ-cat (with Forward orientation). The pUC19-0031-LacZ-cat clone with forward orientation would cause the expression of cat cassette. HindIII enzyme in the clone was found be in position, whose restriction sites were subjected to changes if the orientation of cat fragments changes. In order to check the orientation of cat in pUC19-0031-LacZ-cat constructs, the five clones were digested.
with HindIII which was expected to produce different set of fragments for cat with forward orientation and for cat with reverse orientation (Figure 17).

A

B

<table>
<thead>
<tr>
<th>Digestion patterns of HindIII</th>
<th>Fragment Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC19-0031-LacZ-cat(R)</td>
<td>pUC19-0031-LacZ-cat (F)</td>
</tr>
<tr>
<td>4238bps, 2800bps, 1031bps, 276bps, 229bps</td>
<td>4560bps, 2800bps, 709bps, 276bps, 229bps</td>
</tr>
</tbody>
</table>
Figure 17: Digestion of clones of pUC0031-LacZ-cat construct

Panel A: A 10 from 20ul restriction digestion of pUC0031-LacZ-cat clones with HindIII was loaded on to 1% agarose gel; Lane 1: standard molecular marker; Lanes 2, 3, 4, 5, 6: HindIII digest of clone1, clone 2, clone 3 and clone 4 and clone 5, respectively; Panel B: the expected fragments obtained digestion of reporter construct clones with reverse and forward orientation of cat.

Two clones were found with LacZ in forwards orientation was being in frame with cj0031. These clones were used to perform the experiments to determine the expression of cj0031 under influence of changes in its repeat tract.

3.2.14.3. Recombination of cj0031-LacZ reporter construct into C. jejuni chromosome

The pUC19-0031-lacZ-cat reporter construct was recombined into genome of C. jejuni strain 11168. The chromosome of C. jejuni strain 11168 and insertion was confirmed by performing a PCR with primers lacZb1-R (which binds to the locus inside the lacZ gene) and 0031-F which binds to the cj0031 gene.

3.2.15. Mutational Patterns of ON-to-OFF and OFF-to-ON switching of cj0031:

3.2.15.1. A G10 repeat tract undergoes both insertion and deletion mutations:

NCTC11168 carrying the cj0031-LacZ reporter construct was grown on the MHA+VT plates supplemented with X-gal. A total of 15 colonies were picked in two separate experiments to be used as starter colonies. The length of the repeat tract was found to be
G9 (ON) upon sequencing of these colonies. This confirmed that LacZ (lacking initiation codon) is fused in frame to \textit{cj0031} and any changes in the tract length may push it out of frame with respect to the \textit{cj0031} gene, which would ultimately alter the expression of β-galactosidase. The \textit{C. jejuni} cells lacking production of β-galactosidase would not be able to utilize X-gal as a substrate, which would result in creation of white colonies. So the blue colonies would have an ON phenotype and a G9 repeat tract length while the white OFF phenotypic colonies might have a G10 or G8 repeat tract length. Thus the principle of screening white and blue colonies was used as the basis for detecting changes in the repeat tract length of \textit{cj0031}.

Dilutions (10\(^{-2}\) to 10\(^{-5}\)) of all of 15 colonies having a G9 repeat tract length were grown on the MHA+VT+X-gal plates. The ideal plates were those where the cells had such a sufficient gap between each other that they could be counted easily. The 10\(^{-3}\) (in certain cases), 10\(^{-4}\) and 10\(^{-5}\) (in most of the cases) met those criteria. These dilution plates were selected, and the total number of white colonies (OFF variants of \textit{cj0031}) was enumerated along with the total number of colonies (white + blue) on the plate.

**3.2.15.2. The G9 repeat tract of \textit{cj0031} undergoes only Insertion:**

The white variants/OFF variants of \textit{cj0031} were picked and analysed through sizing PCR products (GeneScan analysis). The length of the repeat tract was measured and found to be G10 in all of the samples analysed as opposed to the progeny colonies which had G9. This indicated that insertion of one nucleotide at the polyG locus of \textit{cj0031} caused the frameshift mutation resulting in the switching of expression from ON
to OFF. Furthermore, the sequencing of samples across the poly G repeat tract also confirmed the presence of a G10 tract length.

The phase variation frequency for OFF-to-ON switching of *cj0031* was measured by analyzing a total of 22 colonies in three separate experiments. The dilution of colonies carrying the G10 repeat tract length for *cj0031* was grown on the MHA+VT+X-gal. Blue variants were observed on the plates, indicating the presence of phase variants which reverted from OFF to ON and caused the expression of β-galactosidase from lacZ by placing it in frame with *cj0031*. The genetic constitution of these ON revertants and starter colonies (OFF variants) were confirmed by sequencing and GeneScan, which was G10 for OFF starter variants and G9 for ON revertants. A total of 58 colonies representing the ON phenotype for *cj0031* were analysed and only G9 repeat tracts were found in all of the ON revertants of *cj0031*, indicating that G10 only switches to G9 with deletion of a single nucleotide during OFF-to-ON switching direction (Table 10).

### 3.2.15.3. Estimation of switching frequencies and mutation rates for G9 and G10 repeats of *cj0031*:

The frequencies of phase variation were measured for each of 15 starter colonies for ON-to-OFF. The median was estimated from these frequency values, which was regarded as the overall phase variation frequency. A similar approach was adopted to calculate the frequency and median frequencies of total of 22 colonies analysed for OFF-to-ON switching direction. The median frequencies estimated from the average frequencies were estimated as $13.5 \times 10^{-3}$ and $17.8 \times 10^{-3}$ for ON-to-OFF and OFF-to-ON switching directions respectively.
### 3.2.15.4. Estimation of phase variation rates for G9 and G10:

The phase variations rates were estimated from these median frequencies by using the equation proposed by Drake (1991). The confidence intervals for were calculated according to Koskoska et al. (1998). The values of median frequencies, phase variation rates and confidence intervals for G9 and G10 repeat tracts are presented in table 11.

**Table 11: Phase variation rates of *C. jejuni* cj0031**

<table>
<thead>
<tr>
<th>Relevant Genotype</th>
<th>Direction of Switching</th>
<th>Repeats</th>
<th>colonies</th>
<th>Freq. (× 10⁻³)</th>
<th>Rates (× 10⁻⁴)</th>
<th>Repeats</th>
<th>colonies</th>
<th>Freq. (× 10⁻³)</th>
<th>Rates (× 10⁻⁴)</th>
</tr>
</thead>
</table>

bPV rates estimated by using PV equation of Drake (1991), number in curly brackets indicate the fold increase over shortest tract, the numbers in square brackets indicate the 95% confidence interval calculated by using the method developed by Koskoska et al., (1998)

### 3.2.15.5. Phase variation rates for *cj0031* showed variation as function of repeat tract length:

As is clear from table 11 the phase variation rates increased with increasing repeat tract length. The phase variation rate for G9 repeat tract phase registered 2.9 fold increase relative to that of the 11168-cj1139-G8 reporter construct. Similarly, the PV rate for 11168-cj0031-G9 was also higher than PV rates of 11168-cj0031-G9 \((2.15 \times 10^{-4})\) and 11168-cj1139-G9 \((2.15 \times 10^{-4})\) [Bayliss et al., 2012], suggesting that G9 repeat tract of cj0031 has higher tendency of mutations compared to the same length of polyG in cj1139. However, the increase in PV rate of G10 repeat tract of 11168-cj0031-OFF \((G10)\) was found to be 1.5 fold higher than 11168-cj0031-G9; and furthermore 8.3 fold higher than PV rate of the 11168-cj1139-G9 repeat tract determined by Bayliss et al. (2012), suggesting that the order of magnitude of mutation rates continued to increase with increase in length of repeat tract (Figure 18).

**Figure 18:** Relationship between PV rates of cj0031 and length of repeat tract

A graph shows that the phase variation rates vary as a function of repeat tract length. The phase variation rates for various tract lengths of cj0031 are: G9 = 12.30 \(x\) 10\(^{-4}\) mutations per division, G10= 17.82 \(x\) 10\(^{-4}\) mutations per division. The phase variation rates were determined from median frequencies by following method of Drake (1991)
3.16. Discussion:

*C. jejuni* has 27 contingency loci or phase variable genes carrying a polyG repeat tract one type of simple sequence repeat (SSR) [Parkhill *et al.*, 2001]. These SSR are prone to variations in their lengths during the replication process mainly due to slipped strand mispairing (Levinson and Gutman, 1987). These variations in the SSR lengths may cause changes in the expression of the phase variable genes which are usually located in chromosomal regions encoding surface structures such as Lipooligosacchride (LOS), capsular polysaccharide (CPS) and flagellar proteins (Parkhill *et al.*, 2000). The phase variation frequencies of most of these contingency loci in *C. jejuni* have not been studied extensively.

The phase variation of contingency loci in the genome of *C. jejuni*, a commensal of chickens but a pathogen of humans, is thought to play vital roles in adaptation and colonization in the host. The frequencies of variation in the homopolymeric tracts of these contingency loci may provide important evidence about the strategies employed by *C. jejuni* to adapt to the host. Furthermore, the mutation frequencies and mutation rates may have implications for long term persistence and colonization of bacterial population within the host. Hence, the study of variations in the frequencies of the simple sequence repeats may hold epidemiological significance for researchers studying the propagation and penetration of *C. jejuni* in chickens.

Since two week old chickens are susceptible to rapid colonization of the GI tract by invading *C. jejuni* strain 11168H, a two-week model of chicken inoculation (Jones *et al.*, 2004) was chosen in order to study the frequency of phase variation in vivo. The two-week old chickens were administered an inoculum dose of (1x 10$^8$) cfu. On the second day, some birds were sacrificed to check the status of colonization but no
bacterial populations of *C. jejuni* were found in the caecum, indicating that *C. jejuni* 11168H strain initially poorly colonised the chickens after the first day of inoculation.

### 3.16.1. Major differences in the proportions of ON and OFF variants

Major differences in the proportions of ON and OFF variants were detected for *cj0031, cj1326, capA, cj0685, cj1139c and cj1326* during in vivo passage through chickens as compared to input population (inoculum) as shown in figures 3-8. The high abundance of ON variants of these phase variable genes may have implications for the necessity of these genes during the colonization and infection processes. All of the six genes except *cj0031* are coding for surface exposed structures. So alterations in these simple sequence repeats may result in variations in the expression of these genes, thus causing the varying level of interactions with host surfaces and immune responses. However, a tendency for favouring an abundance of “ON” genotypes of phase variable genes in vivo was observed in our study.

The repeat tract G9, therefore, exhibits the ON phenotype/genotype for the preceding genes except *cj1139* whose ON genotype is G8. The modal repeat tract lengths having high abundance (frequency in percent) in vivo are G10 (OFF) for *cj0045*, G11 (ON) for *capA* and G10 (OFF) for *cj1326* among all of the birds under investigation. So during in vivo passage, selection for ON or OFF phenotypes may have taken place in vivo. High transcript abundance of ON genotypes for *cj0031, cj0685, capA* and *cj0045* in the output population in vivo was found in comparison with input population where these genotypes were present in low abundance. It is possible that the high prevalence of these ON or OFF genotypes is the outcome of adaptive strategies employed by *C. jejuni* to overcome the barriers to colonization or adaptation such that only those
phenotypes/genotypes were selected, which were best fit for survival in the host changing environment during the passage. However, some other genotypes were also present in small percentages along with the predominant genotype, which might have arisen due to the stochastic nature of slipped strand mispairing yielding sub-populations of other genotypes present in the background (Jerome et al., 2011).

3.16.2. *cj0031* with high frequency of ON variants may act as Phasevarion:

Another explanation of high transcript abundance of these genes after in vivo passage is that some genotypes might affect the expression of the other contingency loci and this mechanism is called a “phasevarion” as proposed for Type III RM systems in *N. meningitidis*, which regulates the expression of other genes (Srikhanta, 2011). A similar mechanism was reported for the contingency locus of a type III RM system in *H. influenza* (Srikhanta et al., 2005). As *cj0031* is a type IIS RM system, so its increased frequency after in vivo passage may be attributed to phasevarion activity of its “putative restriction-modification enzyme”. This RM system may influence the expression of other genes in the genome by turning them from ON or OFF through methylation of specific sites in the ORF or promoters of these genes. Besides, Jerome et al. (2011) also reported an increase in the frequency of a G9 repeat tract and suggested that an increase in the frequency of G9-ON genotype of *cj0031* might be required for adaption of *C. jejuni* to the mice. We postulate that *cj0031* expression in vivo may be required to coordinate the adaptation to the host by controlling the expression of other genes in the *C. jejuni* genome.
3.16.3. High Frequency of phase variation among genes encoding surface proteins enhances adaptation to host:

The proteins taking part in the biosynthesis of LOS and CPS of *C. jejuni* are important factors for the colonization of the GI tract of chickens (Muller *et al.*, 2006). The high frequency of G9 (ON) of *cj0685* and G11 (ON) of *capA* may be due to their potential roles as invasion proteins. The predominance of an OFF phenotype of *cj0045* in the output population of *C. jejuni* after in vivo passage was observed. However, in birds B8 and B9 their frequencies were comparable. Though an ON phenotype of *cj0045* was reported to play an important role for colonization during passage through mice (Kim *et al.*, 2012). Another study also described the *cgtb* system which works in complement with *cj0045* for LOS synthesis/modification (Muller *et al.*, 2006). They found that *C. jejuni* strains lacking the *cj0045* function but having the active function of *cgtb* could colonize the alimentary tract of avian birds successfully. So, the possible explanation of these differences within genetic diversity of *cj0045* may be due to the presence of some other complementary gene (e.g. *cgtb*) which may be expressed and allow the bacterial cells to colonize.

So, all six genes have shown a high frequency of phase variation in vivo. The frequencies and phase variation rates of these genes are yet to be explored in order to estimate the frequency and phase variation rates *in vivo* conditions. Thus, it would be interesting to measure the phase variation rates.

3.16.4. LacZ reporter constructs give higher PV rates:
The phase variation frequencies and phase variation rates were determined for *cj0031* and *capA* genes having a poly G repeat tract within their reading frames. The phase variation rates for *capA* were measured by using a colony immunoblotting assay. An antiserum specific for CapA was provided by Dr. Karl Wooldridge from the University of Nottingham. As antisera for *cj0031* were not available, a LacZ reporter was constructed having LacZ without an initiation codon and depending on the *cj0031*-polyG tract length (G9) for its expression. The *cj0031*-lacZ reporter construct was used to estimate the phase variation rates for the G9/G10 repeat tract lengths of *cj0031*. This technique was successfully employed previously by Bayliss et al. (2004) and Bayliss et al. (2005) for the determination of phase variation rates for dinucleotide and tetranucleotide repeats in *H. influenzae*. The *cj0031*-lacZ reporter construct was designed such that the polyG repeat tract of *cj0031* was located in close proximity to the upstream end of lacZ. The repeat tract length, G9, enables expression of β-galactosidase from lacZ. However, any change in the repeat tract of G9 will cause a frameshift mutation which will not allow the transcription of *cj0031* into the lacZ. This method was efficient in terms of reliability of estimation of phase variation rates due to direct enzymatic detection.

The phase variation rates of *cj0031* repeat tract were measured for G9 and G10 and found in the range of $12.30 \times 10^{-4}$ and $17.88 \times 10^{-4}$ mutations/division, respectively. Similarly the phase variation rates were estimated for G10, G11 and G12 of *capA* by using the colony immunoblotting, and they were $4.88 \times 10^{-4}$, $18.44 \times 10^{-4}$ and $29.49 \times 10^{-4}$ mutations/division, respectively. This shows that phase variation rates are high for these repeat tracts. Recently, Bayliss et al. (2012) measured the PV rates for *cj1139* by using a lacZ reporter construct. The ON-to-OFF PV rate for a G11 tract in *cj1139* was found to be $40.54 \times 10^{-4}$. The PV rate for G11 in the case of *cj1139* was higher than that
for ON-to-OFF PV rate for \( \text{capA-G11} \). A difference in the PV rates of OFF-to-ON switching rate for G10 of \( \text{cj0031} \) and that for G10 of \( \text{capA} \) was also noticed. These differences in the PV rates for \( \text{capA, cj1139} \) and \( \text{cj0031} \) for the same tract lengths may be attributed to the different contexts of the repeats of \( \text{capA} \), which may have a bearing on the expression of these genes. If the \( \text{capA} \) context is explored, it can be noticed that a poly T (T6) tract is found just before the start of poly G of \( \text{capA} \). This context of \( \text{capA} \) repeats is absent from the reporter constructs and it may have a stabilizing impact on the structure of the DNA (Dornberger et al., 1999). The slightly lower PV rates for repeat tracts of \( \text{capA} \) may, nevertheless, arise due to differences in the technique applied to measure it. PV rates for \( \text{capA} \) have been measured through colony immunoblotting while those for \( \text{cj0031} \) and \( \text{cj1139} \) were measured through the lacZ reporter constructs for these genes. The level of efficiency of both techniques for detection of phase variants could be different which might be responsible for these differences. In both cases, \( \text{cj0031}-\text{lacZ} \) and \( \text{cj1139}-\text{lacZ} \) reporter constructs, the PV rates are higher than those measured in colony blotting. We propose that measurement of PV rates through a lacZ reporter construct is more efficient than that measured through colony immunoblotting.

### 3.16.5. High Phase Variation Rates for \( \text{C. jejuni} \) Phase Variable Genes:

Phase variation rates were previously measured for the repeat tracts of \( \text{H. influenzae} \) and \( \text{N. meningitidis} \) by Bayliss et al. (2004), Bayliss et al. (2005), De Bolle et al. (2000), Martin et al. (2004) and Richardson and Stojiljkovic (2001). The switching rates measured for genes in our analysis are similar to the PV rates mediated by tetranucleotide repeats in \( \text{H. influenzae} \), which range from \( 1.4 \times 10^{-4} \) to \( 5.6 \times 10^{-4} \) mutations/division, for tracts of tetranucleotide repeat units having a length in the range
of 17-38 (De Bolle et al., 2000). The phase variation rates determined for poly G repeat tracts of *C. jejuni* genes in our study are, nevertheless, significantly higher than estimated for polyG tracts in *N. meningitidis*. But they are closely related to the PV rates determined for MMR mutants of this species. The PV frequencies reported for meningococci were in the range $1 \times 10^{-5}$ to $3 \times 10^{-5}$ in wild type strains for G10 and G12 tracts, which rose to the range of $2 \times 10^{-3}$ to $8.3 \times 10^{-3}$ for a *mutS* mutant for these tract lengths (Richardson and Stojiljkovic, 2001); whereas the PV frequencies for *capA* G11 and *cj0031*-G10 tracts were $1.8 \times 10^{2}$. This provides empirical evidence that phase variation frequencies are higher for the poly G tracts in *C. jejuni* than those detected for polyG tracts in *N. meningitidis*. The higher phase variation rates and frequencies in *C. jejuni* can be as a result of the absence of *mutS* and *mutL* genes in that species, which take part in the mismatch repair system (MMR). Though *C. jejuni* lacks the MMR pathway it contains multiple functional DNA repair systems such as *uvrB* and *recA* orthologues which may reduce the level of generation of genetic diversity. The spontaneous mutation frequencies estimated by Gaasbeek et al. (2009) lie in the range of $1 \times 10^{-8}$ for nalidixic acid for *C. jejuni* strain NCTC 11168, which oscillated between $1 \times 10^{-6}$ to $4 \times 10^{-9}$ as reported by Hanninen and Hannula (2007) for ciprofloxacin in other *C. jejuni* strains. This evidence suggests that *C. jejuni* does possess mismatch repair activity but is deficient for repairing deletions or insertion mutations. Even, in some cases of simple sequence contingency loci, MMR does not seem to be affecting the high phase variation rates of contingency loci. Bayliss et al. (2002) reported that phase variation rates of tetranaucleotide repeats of *Hi* genome (*H. influenzae*) were not affected by mutations in the MMR genes. However, MMR played a role in controlling the phase variation rate of dinucleotide repeats. Similarly, this is true for controlling the phase variation rates of mononucleotide repeats. So PV rates may vary independently
of the MMR system depending on the contents (tetranucleotide, dinucleotide or mononucleotides) of repeats of contingency loci, however, mutability resulting from the variations in the tract length of contingency loci does affect the phase variation rates of the contingency loci (De Bolle et al., 2000)

3.16.6. PV Rates vary as a function of repeat tract length:

Mutability in the simple sequence repeats is the function of length of SSR whether it comes to microsatellites of eukaryotes or the simple sequence repeats of prokaryotes. In the figures 13 and 20, it has been shown that PV rates increases as the length of repeat tract increases from G9 to G12, which indicates that repeat tract length of SSR is the major determinant of phase variation rates. The phase variation rate for G0 of cj0031 was estimated to be 8.3 fold higher than that of G9. The data pertaining to phase variation rates for capA also registered similar increases in the PV rates for the following repeat tracts: G10 (2.3 fold), G11 (4 fold) and G12 (13.7 fold). Bayliss et al. (2012) presented that a 10 fold increase in the PV rates was caused by the alteration of G8 repeat tract of cj1139 to G11 through site directed mutagenesis. Richardson et al. (2002) also reported a similar 10-fold increment to the PV rates of polyG tracts of phase variable haemoglobin receptors of Meningococci.

3.16.7. Mutational spectra caused by deficiency of repair system rather than selection:
During ON-to-OFF and OFF-to-ON switching directions of \( \text{capA} \) and \( \text{cj0031} \), various mutational spectra were observed. For repeat tracts G9 and G10 of \( \text{cj0031} \) and \( \text{capA} \) respectively, the bias towards +1 insertional events was observed for G9. Contrastingly, the bias towards -1 deletional events was prominent for G10 (\( \text{cj0031} \)), G11 (\( \text{capA} \)) and G12 (\( \text{capA} \)). Bayliss et al. (2012) also showed that G8 of \( \text{cj1139} \) were subjected to +1 insertions while G11 of \( \text{cj1139} \) was biased for -1 deletional events. This suggest that longer repeats are substrate for -1 deletional events while relatively shorter repeats undergo +1 insertional mutations. The studies conducted by Gawel et al. (2002) and Gragg et al. (2002) provided an empirical evidence by showing that deletional and insertional events within runs of G residues and T residues were outcomes of mutations resulting from deficiencies of MMR system and replicative slippage. Gragg et al. (2002) employed a G10 model in yeast whereas Gawel et al. (2002) used a G6 model in \( \text{E-coli} \). The analyses of mutational spectra from their data showed that MMR-proficient wild-type strains favoured the bias for +1 insertion rather than -1 deletion. However, MMR-deficient strains exhibited a shift in bias toward deletions in G10 model of Yeast used by Gragg et al. (2002). Corresponding loss of bias for insertion was observed for the G6 model in \( \text{E. coli} \) but only for the lagging strand. These shifts in the mutation spectra for MMR deficient yeast and \( \text{E. coli} \) strains from insertion to deletion were recorded for the repeat tracts having ≥8 nucleotides due to diminishing efficiency of polymerase selectivity, exonuclease proofreading and mismatch repair (Kroutil et al., 1996; Tran, 1997). In such circumstances, only the MMR system is responsible for correction of indels for longer repeat tracts ≥8. Species lacking efficient MMR system but carrying repeats ≥8 showed deletional patterns in their mutational spectra. That is why, the mutational spectra of G9 and G10 shifted more towards insertion and that of G11 and G12 were shifted towards deletions due to lack of MMR pathway in \( \text{C. jejuni} \).
Alternatively, there may be some other repair system active behind the shifts of these mutational spectra. The mutational spectra of \textit{cj0031} and \textit{capA} showed the prevalence of G9 and G10 repeat tracts in the phase variable genes of \textit{C. jejuni}, which suggested that selection for these polyG (G9/G10), were directed by the molecular drivers rather than by the selection for particular switching rates. Kim \textit{et al.} (2012) also showed that selection for G9 and G10 for phase variable genes of NCTCT 11168 strain imparts a selective advantage to bacterial populations over the G11 repeat tract length in vivo. If it is assumed that only selection for high phase variation would be operating, the prevalence of G11, G12 or higher should be expected. But the actual scenario of phase variable genes of \textit{C. jejuni} in vivo is different. The longer repeat tracts, however, exist within \textit{N. Meningococci} genomes where the selection for the longer repeat tracts might be the driving force operating behind the presence of longer repeat tracts just for generation of higher phase variation rate. As these strains contain an active and functional MMR system which might decrease the mutation rates of shorter repeat tracts, so the genome of \textit{Meningococci} tends to be having longer repeat tract lengths.

Some other \textit{Campylobacter} species, apart from \textit{C. jejuni}, possess larger number of repeats and longer than the repeat tracts of \textit{C. jejuni}. For instance, \textit{C. upsaliensis} RM 3195 harbours total of 89 repeat tracts carrying G7, however, 59 repeat tracts out of 89 should have G12 or longer repeat tracts (Fouts \textit{et al.}, 2005). These species might have an active MMR system, which is putting a selective evolutionary pressure for longer repeats. So our results suggest that divergence between the \textit{Campylobacter} genomes in terms of genetic constitution and sequence of repeat tracts may be driven by the varying abilities and efficiencies of replicative machinery or MMR system encoded by them.
3.16.8. Hypermutability of phase variable genes is the cause of genetic diversity:

A total of 22 genotypes were observed in the output populations, while there were only three genotypes out of 9 from input population were present in the genotypes of the output. So the massive difference was observed between the output and input populations in terms of ON and OFF phases of multiple genes. The quasispecies concept describes the population of *C. jejuni* as full of genotypes and phenotypes living in the host (Parkhill *et al*., 2000). This is what we detected during in vivo passage of NCTC11168H.

We suggest that differences in the genotypes of output and input are mediated by selection in the host. The Model presented by Bayliss *et al.* (2012) for the impact of mutational drift on genetic diversity could not replicate these differences, which reinforced the suggestion that these variations at the level of genotypes are not mediated by stochastic process of mutation, but in matter of fact, they are emerged under pressures of selection during in vivo passage. Three genotypes (*capA*-OFF, *capA*-ON, *cj0045*-ON) were prevalent in the output population along with some minor genotypes of multiple phase variable genes. These variations were also seen in the case of bird-to-bird variations for the distribution of these genotypes. These minor genotypes can result from analysis of only a few colonies. The bird to bird variations among these minor genotypes suggest that these variations may have been generated randomly at a low level in vivo, which would be available for selection depending on the requirement for colonization and motility as suggested by Jones *et al.* (2004). However, another possibility of variation in genotypes is that hosts may differ from each other in terms of behavior; daily excretion of caecal contents etc. may produce bottlenecks among some
birds but not in the rest of them. So the oscillation in the genotypes of \textit{cj0045} and \textit{capA} was observed during in vivo passage. Apart from them, the genetic diversity in the form of various genotypes (22 for six genes) may need to be maintained for survival of the selective pressures exerted by the host on the bacterial population. So it is proposed that genetic diversity for a particular combination of multiple genes is required for adaptation and colonization.

\textbf{3.2.16.9. Genetic diversity of ON/OFF status causes the impact on colonization of C. jejuni in the host:}

After observation of the data generated by our study, it seems obvious that genetic variation is present in all of these six genes which often results in changes to their ON or OFF status. We can not say with confidence that these genes would have been expressed during in-vivo passage. But the prevalent genotypes in the output population showed their bearing on the varying level of colonization and persistence between isolates of the same strain of \textit{Campylobacter}. The problems encountered by Coward \textit{et al.} (2008) during the screening of WITS-tagged isolates and their varying abilities in terms of colonization may be related to the phase variable structure of these populations and the differences in the level of expression status of these phase variable genes. These high frequencies of phase variation presented in this study may have implications not only for adaptation to different niches and a range of hosts, but they also have an important bearing on the survival of bacteriophages which infect \textit{Campylobacter} through their attachment with phase variable epitopes on the surface of the bacterial cells (Sorensen \textit{et al.}, 2011; Coward \textit{et al.}, 2006). Jerome \textit{et al.} (2011) reported that alterations in the frequency of phase variation in the avian host facilitated the
transmission and colonization of *C. jejuni* in mice (a non-native host). This finding further reinforced the suggestion presented herein that changes in the frequency and rates of mutation at repeats of contingency loci have evolved to enable rapid adaptation and colonization in the host.
Chapter 4. Investigation of methylation activity of Cj0031 methyltransferase

4.1: Introduction:

The methyltransferases are normally separated into two classes: orphan methyltransferases which function without being part of a restriction system in the bacterial host such as Dam methyltransferase in *E. coli*; and RM systems where the methyltransferase acts in concert with a restriction endonuclease to perform various biological functions in a host (Blumenthal and Cheng, 2002). The primary function of the latter class of methyltransferases is to methylate the host genome in order to prevent its digestion by the cognate restriction endonuclease component of the RM system. A restriction endonuclease of an RM system restricts invading foreign DNA entities, which can enter the bacterial cell during natural transformation, conjugation or infections by phages (Kobayashi, 2004).

RM systems are classified into type I, Type II, type III and type IIG which were discussed extensively in the chapter 1, section 1.6.1. Type I RM systems encode multi-subunit complexes in the form of M2R2S and normally their sequence specificities spans over 13-24bp. A Type II RM system recognizes a 4-8bp target sequence which may be palindromic or non-palindromic. Most of the type II RM systems discovered until now possess recognition sequences ranging from 4-6bp (Kobayashi, 2001). Type II RM systems cleave at the target sequence with a high precision level and sequence specificity. However, type IIG RM systems mediate non-specific and variable cleavage at their target sequence. Like, type I RM systems, a type III RM system consists of multi-subunits forming a Res2Mod2 complex which requires DNA translocation and ATP mediated hydrolysis in order to precipitate the cleavage.
The methyltransferase of type II RM in prokaryotes methylate the cytosine ($m^4$C; $m^5$C) or adenine ($m^6$N) in their recognition sequences, which are known to influence many biological processes such as DNA repair, gene silencing and particularly gene expression (Berger and Bird, 2005; Jelinic and Shaw, 2007; Iqbal et al., 2011; Marinus, 2010). The implication of type III RM in epigenetic regulation of gene expression in prokaryotes has been reported (Belland et al., 1996; de Varies et al., 2000; Srikhanta et al., 2007, Srikhanta et al., 2010). Moreover, the phase variation phenomenon associated with expression of the methyltransferase component of many type III RM systems has been ascribed to influencing the pathogenesis of pathogenic bacteria such as *N. meningitidis*, *H. pylori*, and *H. influenzae* Rd strain. The phase variable methyltransferases of type III RM systems constitute phasevarions affecting the switching of multiple genes in the genome of pathogens (Srikhanta et al., 2010, Srikhanta et al., 2012, Bayliss et al., 2006).

*cj0031* has a polyG repeat tract which fuses the two genes *cj0031* and *cj0032* together. Current annotation of these genes places both genes under a single promoter for ORFjc0031. The G9 length causes both genes to be in the same frame and allows the translation of a functional protein. However, switching of G9 to G10 or to G8 turns *cj0031* off and only truncated product is obtained. Hence, *cj0031* holds the status of phase variable gene expression. The *cj0031* gene product is a putative type IIG RM system and its biological functions might be related to restriction/modification, recombination, repair and recombination (http://www.xbase.ac.uk/genome/campylobacter-jejuni-subsp-jejuni-nctc-11168/search?s=cj0031).

The genome of *Campylobacter* NCTC11168 strain contains 2 putative methyltransferases, and two putative RM systems. *cj0031* is unique as it undergoes
phase variation, while the rest of the RM systems and methyltransferases in NCTC11168 are not phase variable. In *C. jejuni*, none of the methyltransferase and RM systems have been characterized so far in terms of their methylation or restriction activity.

However, the effect of *cj1461*, a putative methylate, has been determined by Kim *et al* (2008) on various phenotypes such as motility, adherence and invasion. This indicates that there is a paucity of empirical evidence about the biological activities of putative RM system in *C. jejuni*, especially, in the backdrop of current advances of understanding and elucidation of phase variable type II and type III RM systems in other pathogenic bacteria such as *H. pylori, N. meningitidis, H. influenzae* etc. *cj0031*, the only phase variable RM system, does not occur in *C. jejuni* without important implications for its pathogenesis and life-style.

In the previous chapter 3 section 3.2.3, it had been shown during in vivo passage of *C. jejuni* through chickens that a *cj0031*-ON phenotype was selected for in colonizing bacteria as against *cj0031*-OFF which was predominant in the inoculum. This may have reflected its importance during the colonization of chickens by *C. jejuni* in chicken. This empirical evidence also provided the motivation to extend the research work to investigate the biological activity of *cj0031*, to examine various phenotypes implicated in the pathogenesis of *C. jejuni*.

Thus, the aim of the present work is to characterize the putative type IIIS RM system encoded by *cj0031*. There are thousands of putative type II RM system which are yet to be characterized in terms of finding their methylation and restriction activities and their recognition sequences. Commonly, there are generally two strategies employed to characterise the RM systems. Firstly, an *in vitro* method used to find the recognition sites of type II RM systems is to cleave the genomic or plasmid DNA from a bacterial
strain expressing the putative RM system with various restriction endonucleases of known specificities. Then a comparison is performed of the observed restriction patterns with a computer generated set of fragments for digests of these substrates with the respective enzymes (Gann et al., 1987; Cowan et al., 1989).

Secondly, an in vivo strategy is frequently applied by molecular biologists to determine the recognition activity through the application of M13 phage vectors containing known DNA sequence. The principle behind this strategy is infection of a bacterial host by the phages the host and determination of the plating efficiency (EOP) for the phages. The modified phages have higher EOP than the unmodified ones by an order of magnitude of $10^{-1}$ to $10^{-3}$ (Ellrott et al., 2002). This strategy helps to determine recognition by restricting the phage genomic DNA supposed to be having putative recognition sites with purified methyltransferase, then mapping the produced fragments by running on gel leads to find a recognition site for putative methyltransferases.

As the goal of present work was to determine the methylation activity and recognition sequence, so the strategy of inhibition of restriction of modified target sequences by the relevant methylation sensitive restriction enzymes was adopted.

4.2: Results:

4.2.1: Bioinformatics analysis of Cj0031 R/M system:

4.2.1.1: Construction of a phylogenetic tree:

The phylogenetic tree was constructed by using the TreeFam version4 software on the UniprotKB website (http://pbil.univ-lyon1.fr/cgi-bin/acnuc-
ac2tree?query=Q0PC94&db=HOGENOM). The software needs the query sequence to perform an analysis. The UniprotKB accession number Q0PC94 for *cj0031* was provided to the software which was also linked with BLAST and HMMER to search for sequence matches in the protein data sets from bacterial species whose genome were fully sequenced. The software first created a link to the family to which query sequence belongs to. For instance, the restriction modification system was grouped in family B and methylation only systems were grouped in family A. The query sequence was placed in family B seed due to its highest-scoring HMMER match with that family. The seed was, then, expanded into full/clean family. The alignments of protein sequences in each full and/or clean family were performed using Muscle version 3.6 (Edgar, 2004). The alignments pertaining to the only conserved sequences were filtered, and subsequently employed by the software as input data in a neighbour-joining algorithm (NJA). The speciation and duplication nodes in the tree were created by utilizing the Duplication/Loss interface algorithm NJA which ultimately used the amino acid mismatch distances in order to build a phylogenetic tree. The small trees built at a species level were merged by using the PHYML release 2.4.5 which utilized a maximum likelihood method (ML) and the tree merging algorithm to put forward the final phylogenetic tree (Figure
Figure 1: Phylogenetic tree showing the relationship of *cj0031* with its orthologs in other species

The phylogenetic analysis revealed an association of *cj0031* with putative restriction modification systems of other strains of *C. jejuni* such as *C. jejuni* 81116RMS, *C. lari* RM2100, *C. jejuni* RM1221RMS, and *C. jejuni doylei* 269.97. The similarity between *cj0031* of *C. jejuni* 11168 with other strains was found to be in the range of 88-100% for the conserved regions (Figure 4). Orthologs of *cj0031* were found in many strains of *H. pylori*, the closest neighbour of *C. jejuni*, indicating the horizontal transfer of *cj0031* from other bacterial strains into NCTC11168 strain. A distant association of *cj0031* was also found with putative restriction modification systems of cyanobacterial strains (*Microcystis aeruginosa*) as indicated in the phylogenetic tree (Figure 4). This analysis showed the distribution of orthologs of *cj0031* over a wide range of bacterial strains and species.

4.2.1.2: Hypothetical Conserved Domain Model of Cj0031 RM system:

The conservation of motifs and domains was also reported among R/M systems (Blumenthal and Cheng, 2002). *cj0031* as a potential candidate for a type II R/M system showed homology with conserved domains and motifs of R/M systems in other bacterial species (figure 2A and figure 2B), indicating that *cj0031* has homology with conserved domain sequences of other type IIG restriction modification system such as *Eco57I* (T257_ECOLI), *HP1517* (*H. pylori*), BanIII (MTB3_BACAR). The conserved domain sequence from Eco57I restriction-methylase domain and Ado-MTase domain of Cj0031 RM system matching with those of type IIG RM systems in other bacterial strains is presented (Figure 2A and 2B)
<table>
<thead>
<tr>
<th>GI</th>
<th>Definition</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>158345116</td>
<td>Type II restriction, methylase unit</td>
<td>Salmonella</td>
</tr>
<tr>
<td>11496616</td>
<td>Type IIG restriction, methylase unit</td>
<td>E. coli</td>
</tr>
<tr>
<td>11496712</td>
<td>Type IIG restriction, methylase unit</td>
<td>Bacillus</td>
</tr>
<tr>
<td>11496797</td>
<td>Type IIG restriction, methylase unit</td>
<td>C. jejuni</td>
</tr>
<tr>
<td>15791430</td>
<td>Type II restriction, methylase unit</td>
<td>C. jejuni</td>
</tr>
<tr>
<td>15791431</td>
<td>Type II restriction Eco57I, methylase</td>
<td>H. pylori J99</td>
</tr>
<tr>
<td>1171052</td>
<td>Type IIG restriction, methylase unit</td>
<td>Deinococcus</td>
</tr>
<tr>
<td>16082324</td>
<td>Type II restriction, methylase unit</td>
<td>Clostridium</td>
</tr>
<tr>
<td>15804879</td>
<td>Type IIG restriction, methylase unit</td>
<td>H. pylori</td>
</tr>
</tbody>
</table>

B:
BpMI-Restriction endonucleases-MTase

<table>
<thead>
<tr>
<th>GI</th>
<th>Definition</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>81747716</td>
<td>Putative restriction endonuclease</td>
<td><em>Mycoplasma penetras</em></td>
</tr>
<tr>
<td>81584974</td>
<td>Restriction endonuclease</td>
<td><em>Enterococcus</em></td>
</tr>
<tr>
<td>75357976</td>
<td>Restriction endonuclease</td>
<td><em>Lactobacillus</em></td>
</tr>
<tr>
<td></td>
<td><strong>BpMI-Restriction endonucleases-MTase</strong></td>
<td></td>
</tr>
<tr>
<td>74499887</td>
<td>(IIG)</td>
<td><em>Archaea samples</em></td>
</tr>
<tr>
<td>52001476</td>
<td>Eco571-restriction endonuclease-MTase</td>
<td><em>E. coli</em></td>
</tr>
<tr>
<td>81625849</td>
<td>putative type II adenine specific MTase</td>
<td><em>H. pylori J99</em></td>
</tr>
<tr>
<td>148261408</td>
<td>Adenine-specific MTase</td>
<td><em>Acidiphilus JF-5</em></td>
</tr>
<tr>
<td>160886778</td>
<td>putative type II RM system</td>
<td><em>Bacteroid ovatis</em></td>
</tr>
<tr>
<td>120600856</td>
<td>Eco571-restriction endonuclease-MTase</td>
<td><em>Shewanella W3-18-1</em></td>
</tr>
</tbody>
</table>
**Figure 2: Homology of Cj0031 R/M system with Eco57I and methyltransferases in other bacterial species.** The homologies in specific domains between Cj0031 and counterparts from other bacterial strains are shown for the alignment of the methyltransferase (Panel A), Eco57I restriction-modification endonuclease (Panel B), and TaqI-like C-terminal specificity domains (Panel C). Panels-AK, BK and CK- show keys to understand the “GI” numbers against each alignment on the left sides of panels A, B and C, respectively. The alignment of domain motifs of Cj0031 with conserved motifs in other bacterial strains was obtained by feeding the protein sequence of Cj0031 to BLAST Conserved
Domains version 2.2.27. The red coloured amino acids represent the highly conserved residues between the various stains and query sequence (Cj0031). The blue coloured amino acids represent the less conserved but highly similar sequences.

Based on these conserved domains, the putative domain structure was drawn from the BLASTX searches. The protein sequence of Cj0031 was fed to the BLASTX (2.2.27) and alignments were made with other protein sequences having similar functions. The BLAST program then curated the conserved sequences from these alignments and a hypothetical domain model was constructed (Figure 3). The model showed that Cj0031 had signatures of Eco571 and AdoMet-MTase super family. A Taq-I binding domain was found at the C-terminus in the putative domain structure model. The N-terminal end of Cj0031 contained HsdR-N domain motifs characteristic of the endonuclease component of RM systems.
176

Figure 3: Hypothetical conserved domain model of Cj0031 R/M enzyme:

The model shows two domains-COG1002 and TaqI-C like domain. COG1002 is notation used to denote Ado-MTases superfamily, represents type II restriction enzyme, methylase subunits, represents multi-domains, AdoMet-MTase domain. The TaqI-C like specificity domain is towards the C-terminus and is a single domain. The yellow line above the protein query Seq. shows the positions of cj0031 and cj0032 plus the repeat tract (poly G) separating both reading frames. The length of cj0031 spans over 2.90 kb length while cj0032 covers 800bp length.

4.2.1.3. An in silico search for the recognition site for the Cj0031 methyltransferase:

The REBASE data base contains a collection of 315 type II RM systems and restriction enzymes (NEB: REBASE). A search for the 5’-CCCGAA-3’ or 5’-CCCGA-3’ sites was performed in the REBASE data base to find any restriction enzymes containing recognition sites identical to the Cj0031 methyltransferase but none of them was found.
to have sites identical to the putative site of Cj0031. The search was further extrapolated to find the specificity of any other characterized RM enzyme in the REBASE by using "REBASE Tool" option. Through this tool, the protein sequence was blasted in REBASE, two enzymes- CjeFV and HP99XIII- were found showing 81% and 38% identities with Cj0031. These enzymes are described as putative type IIG restriction-modification systems:- CjeFV is present in C. jejuni strain 81176, having specificity GGGCA, while HP99XIII belonged to H. pylori and had sequence specificity for GCCTA. These results suggest that the proposed recognition sequence for Cj0031 might be one of three options (5’-GGGCA-3’ , 5’-CCCGA-3’ or 5’-GCCTA-3’).

4.2.2: Investigation of the methylation activity of Cj0031

After having established the function of Cj0031 as a putative restriction modification enzyme by using the computer based analysis of the nucleotide and protein sequences of cj0031/cj0032, the next step was to provide direct empirical evidence for the methylation activity. cj0031 is a phase variable system which means it can exist either in ON or OFF states. Cj0031 methyltransferase in the Cj0031-ON variant was supposed to methylate the genomic DNA of NCTC11168 or vice versa for an OFF variant. Thus there was a need to obtain both ON and OFF variants of cj0031 in order to establish the methylation activity.

4.2.2.1: Selection of cj0031-OFF phenotype

The C. jejuni NCTC11168 strain was streaked on MHA agar plates. After 3-days growth, a single colony was picked and serial dilutions in the order of 10^{-2} to 10^{-5} were prepared. These dilutions were plated on MHA plates in order to obtain an isogenic population of C. jejuni. 110 colonies were randomly selected from 10^{-4} and 10^{-3}
dilutions, followed by streaking them on fresh MHA plates. After 3-days growth, some portion of each streaked colony was picked and utilized for making a boilate DNA preparation, while the rest was saved in glycerol at -80°C for further reference and processing.

4.3.2: Determination of tract length in *cj0031* in the Samples:

The boilate DNA preparations were subjected to amplification reactions in order to amplify the region across the repeat tract joining *cj0031* and *cj0032* by using primers: 0031-F (Fam) and 0031-R (Material and Methods). These PCR products were Genescanned (as described in the Materials and Methods) in order to determine the length of repeat tract in *cj0031*. It was found through GeneScan analysis that 75 samples showed the tract length G10, an OFF-phenotype, and 13 samples showed the G9 repeat tract representing an ON phenotype.

4.2.2.3: Sequencing of samples with *cj0031*-OFF /ON phenotype:

Three samples from the pools of G9 (ON) and G10(OFF) each were selected for sequencing reactions in order to verify the GeneScan results. The sequenced data of ON/OFF phenotypes of these samples was found to be consistent with the GeneScan data showing the validity of GeneScan results for the determination of repeat tract lengths of phase variable genes (Table. 1).
Table 1: Comparison of sequencing data with GeneScan Data

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Gene</th>
<th>Repeat Tract Length</th>
<th>Sequencing Result</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-OFF</td>
<td>cj0031</td>
<td>221.4</td>
<td>G10</td>
<td>same</td>
</tr>
<tr>
<td>8-OFF</td>
<td>cj0031</td>
<td>221.4</td>
<td>G10</td>
<td>same</td>
</tr>
<tr>
<td>17-OFF</td>
<td>cj0031</td>
<td>221.6</td>
<td>G10</td>
<td>same</td>
</tr>
<tr>
<td>57-ON</td>
<td>cj0031</td>
<td>220.3</td>
<td>G9</td>
<td>Same</td>
</tr>
<tr>
<td>59-ON</td>
<td>cj0031</td>
<td>220.6</td>
<td>G9</td>
<td>Same</td>
</tr>
<tr>
<td>60-ON</td>
<td>cj0031</td>
<td>220.5</td>
<td>G9</td>
<td>Same</td>
</tr>
</tbody>
</table>

4.2.2.4: Preparation of chromosomal DNA with cj0031-OFF & cj0031-ON phenotype:

Two variants, 3-OFF and 57-ON, from the glycerol stocks were streaked on MHA plates. They were further swabbed on MHA plates by using cotton swabs in order to obtain confluent growth of *C. jejuni*. The confluent growth of *Campylobacter* from these plates was utilized for preparation of ultra high quality chromosomal DNA by employing the CsCl DNA preparation method as described in Materials and Methods chapter 2 section 2.5.1.

The overview of whole process of selection of cj0031-ON and cj0031-OFF variants has been shown (Figure 4).
Figure 4: Selection of *cj0031*-ON and *cj0031*-OFF phenotypes:

*C. jejuni* strain 11168 was streaked on MHA plates, after three days of incubation in VAIN, 110 colonies were picked and streaked on MHA plates. A half of these colonies were used to prepare boilate DNA samples (blue tubes), while the rest was used to make glycerol stock of colonies (brown tubes). DNA extracted in blue tubes was subjected to GeneScan to determine repeat tract length, length was verified by sequencing analysis.
Based on repeat tract length data, *cj0031* ON/OFF variants were selected, the relevant glycerol stock was streaked on MHA plates and ultra high quality DNA was prepared.

### 4.2.2.5: Selection of methylation sensitive restriction enzymes:

As the putative Cj0031 methyltransferase is presumed to have adenine specific activity, most of the restriction enzymes tested to determine the methylation target site were sensitive to adenine methylation. By using the REBASE site, the adenine sensitive restriction enzymes were selected. There were hundreds of enzymes having sensitivity to adenine methylation on the REBASE site. The 30 methylation sensitive enzymes having recognition sites between 4 and 6bp were chosen to test whether their recognition sites overlap or match with that of Cj0031 methyltransferase. The hypothesis was that if the recognition site of these restriction enzymes overlaps or matches with that of Cj0031 methyltransferase, then restriction of that site by the particular restriction enzymes will be prohibited, otherwise the sites would be cut and two fragments can be visualized on a gel or can be detected by Southern blotting.

The digestion of genomic DNA by restriction enzymes does not leave a distinct pattern of fragments which can be visualised on an agarose gel, especially if this DNA contains hundreds of target sites of particular restriction sites. The digestion of genomic DNA by such an enzyme often leaves a smeary appearance on the gel. The decision of which target site was protected can not therefore be made on the basis of just visualisation of a gel picture. So there was a need of another accurate strategy which can be used to detect any change in the restriction pattern of a particular restriction enzyme in order to detect the protected or restricted target sites. A Southern blotting technique was employed to detect the fragmentation patterns of restriction enzymes. This technique made it
possible to detect any changes in the restriction pattern of target sites of particular enzymes located in the region recognized by a specific probe.

4.2.2.6: Design of Probes:

Initially three probes were designed by selecting fragments of genomic DNA from different loci of the *C. jejuni* strain 11168 genome. These fragments were selected by using Clone Manager and potential restriction sites of methylation sensitive enzymes were checked. Three probes were selected- probe I, probe II and probe III- and these probes contained 13, 10, and 10 restriction sites, respectively, for different methylation sensitive restriction enzymes (figure 5A, figure 5B, figure 5C). The lengths of probe I, probe II, and probe III were 700bp, 900bp and 485bp respectively. The frequency of restriction sites for a particular restriction enzyme can be observed and it should be noted that some sites occur more than once within a probe and between the probes. The frequency of occurrence of the restriction sites of a particular restriction enzyme at more than one place in the probes was preferably selected in order to enhance the chance of an overlap with the putative recognition site of the Cj0031 methyltransferase.
Figure 5A: Location of different methylation sensitive restriction enzymes sites on Probe I:
Figure 5B: Location of different methylation sensitive restriction enzymes sites on Probe II

Figure 5C: Probe III with display of various restriction sites for methylation sensitive enzymes
Primers were designed to amplify each of these probes. A PCR reaction was prepared to amplify the probes. The concentrated PCR product of a probe (~ 1ug) was required to perform labelling reactions. For this purpose, about 600 µl PCR reaction was prepared and concentrated by applying the ethanol precipitation method as mentioned in Material and Methods chapter. The labelling reaction was performed and labelling efficiencies of the probes were determined by following the instructions in the DIG-labelling kit (Roche) as detailed in the Material and Methods chapter. The amount of label incorporated per µg of probe were estimated to be 3-fold (probe I), 4-fold (probe II) and 7 fold (probe III) compared to the control labelled DNA provided along with DIG-labelling kit (Roche) (data not shown).

4.2.2.7: Optimization of Digestion and Gel separation Conditions:

Different DNA concentrations ranging from 1ug to 4ug were used in digestion reactions containing 2 units of restriction enzyme. The incubation time was varied from 2 hours, 4 hours and to overnight. The optimized digestion result was achieved with a 2 hour incubation. Separation was facilitated by running digestion mixtures on a 1% agarose gel in TAE at 80 V for 2.30 hours.

4.2.2.8: Detection of inhibition of digestion at a HindIII site

4.2.2.8.1: Probing with probe I:

Ultra-pure DNA preps of the cj0031-ON and cj0031-OFF phase variants were obtained (Material and Methods section 2.5.1). The chromosomal DNA was digested with PstI,
AlwI, HindIII, KpnI, HinfI, BglII and DraI, all of which are sensitive to adenine methylation in their target sequences. In Clone Manager, the fragments generated by the digestion of each restriction enzyme (unmethylated) and their respective distances migrated by them on an agarose gel were determined. Moreover, the fragments produced by each restriction enzyme in the case of inhibition of relevant restriction sites (due to potential methylation by Cj0031) were also calculated (Table 2).

**Table 2: No. of restriction enzymes sites in Probe I and length of fragments generated by them following digestion**

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>No. of sites</th>
<th>No. fragments produced following digestion</th>
<th>Length of Fragments (Bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MboI</td>
<td>3</td>
<td>4</td>
<td>283, 267, 58, 60</td>
</tr>
<tr>
<td>MboII</td>
<td>3</td>
<td>4</td>
<td>342, 58, 296, 97</td>
</tr>
<tr>
<td>AluI</td>
<td>7</td>
<td>8</td>
<td>255, 6, 16, 11, 27, 60, 102, 291</td>
</tr>
<tr>
<td>DraI</td>
<td>2</td>
<td>3</td>
<td>731, 33, 294</td>
</tr>
<tr>
<td>AlwI</td>
<td>1</td>
<td>2</td>
<td>3033, 1264</td>
</tr>
<tr>
<td>HindIII</td>
<td>1</td>
<td>2</td>
<td>417, 525</td>
</tr>
<tr>
<td>BglII</td>
<td>1</td>
<td>2</td>
<td>3773, 452</td>
</tr>
<tr>
<td>HinfI</td>
<td>1</td>
<td>2</td>
<td>1972, 866</td>
</tr>
<tr>
<td>PstI</td>
<td>1</td>
<td>2</td>
<td>2329, 3749</td>
</tr>
<tr>
<td>Sau3AI</td>
<td>3</td>
<td>4</td>
<td>283, 267, 58</td>
</tr>
<tr>
<td>RsaI</td>
<td>1</td>
<td>2</td>
<td>1764, 1208</td>
</tr>
<tr>
<td>KpnI</td>
<td>1</td>
<td>1</td>
<td>28792</td>
</tr>
<tr>
<td>EcoRV</td>
<td>1</td>
<td>1</td>
<td>3182</td>
</tr>
<tr>
<td>DpnI</td>
<td>2</td>
<td>3</td>
<td>283, 267, 58</td>
</tr>
</tbody>
</table>
A Southern blot analysis was performed and the resulting probe revealed the expected band patterns resulting from digestion of respective unmethylated restriction sites for both ON and OFF variants of *cj0031*, indicating that respective recognition sites for PstI, AlwI, HindIII, KpnI, HinfI, BglII and DraI did not overlap with the potential recognition site of Cj0031 (Figure 6).

**Figure 6: Southern blot probed with probe I**

Southern blot showing the digestion of restriction sites contained in probe I locus in the genome extracted from ON and OFF variants by various methylation sensitive restriction enzymes: positive (+) and negative (-) signs in the figure represents the ON and OFF variants of *cj0031*, respectively.

**4.2.2.8.2: Probing with Probe II:**

A Southern blot analysis was performed by probing two blotting membranes with probe II on two separate occasions. One set of digestion reactions was performed on genomic
DNA by using DraI, MboI, PsiI, MboII, BsmI, SacI, and BsgI. Computer based restriction digest was performed on the locus of probe II in *C. jejuni* strain 11168, and the fragments generated were recorded (Table 3).

**Table 3: No. of restriction enzymes sites in Probe II and length of fragments generated by them following digestion**

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>No. of sites</th>
<th>No. fragments produced following digestion</th>
<th>Length of Fragments (Bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MboI</td>
<td>1</td>
<td>2</td>
<td>537, 1488</td>
</tr>
<tr>
<td>MboII</td>
<td>3</td>
<td>4</td>
<td>156, 155, 17, 1499</td>
</tr>
<tr>
<td>DraI</td>
<td>5</td>
<td>6</td>
<td>163, 163, 7, 32, 135, 189</td>
</tr>
<tr>
<td>SacI</td>
<td>1</td>
<td>2</td>
<td>2762, 6430</td>
</tr>
<tr>
<td>PsiI</td>
<td>1</td>
<td>2</td>
<td>1126, 1322</td>
</tr>
<tr>
<td>BsmI</td>
<td>1</td>
<td>1</td>
<td>2775</td>
</tr>
<tr>
<td>BsgI</td>
<td>1</td>
<td>2</td>
<td>185, 7135</td>
</tr>
</tbody>
</table>

A difference in the restriction patterns was not observed between ON and OFF variants of *cj0031* (Figure 7), indicating that the target sites of respective restriction enzymes used did not overlap with the methylated recognition sequence of Cj0031 methyltransferase.
Figure 7: Southern blot probed with probe II

Southern blot probed by probe II showing the digestion of restriction sites contained in probe II locus in the genome extracted from ON and OFF variants by various methylation sensitive restriction enzymes: positive (+) and negative (-) signs in the figure represents the ON and OFF variants of \textit{cj0031}, respectively.

4.2.2.8.3: Probing with Probe III:

A third probe, Probe III, was also used to analyse the restriction digests of genomic DNA extracted from ON and OFF phase variants and digested with adenine methylation sensitive enzymes (MboI, MboII, BglII, AluI, DpnI, Sau3AI, Dral and AlwI, HindIII). In silico, the length of restriction fragments following digestion with respective restriction enzymes were determined and recorded (Table 4).
Table 4: No. of restriction enzymes sites in Probe III and length of fragments generated by them following digestion

<table>
<thead>
<tr>
<th>PROBE III (485bps)</th>
<th>Enzymes</th>
<th>No. of sites</th>
<th>No. fragments produced following digestion</th>
<th>Length of Fragments (Bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MboI</td>
<td>4</td>
<td>5</td>
<td>200, 311, 21, 31, 30</td>
</tr>
<tr>
<td></td>
<td>MboII</td>
<td>5</td>
<td>6</td>
<td>175, 172, 104, 21, 52, 51</td>
</tr>
<tr>
<td></td>
<td>AluI</td>
<td>2</td>
<td>3</td>
<td>376, 152, 411</td>
</tr>
<tr>
<td></td>
<td>DraI</td>
<td>2</td>
<td>3</td>
<td>408, 240, 266</td>
</tr>
<tr>
<td></td>
<td>AlwI</td>
<td>2</td>
<td>3</td>
<td>4493, 311, 411</td>
</tr>
<tr>
<td></td>
<td>HindIII</td>
<td>1</td>
<td>2</td>
<td>998, 425</td>
</tr>
<tr>
<td></td>
<td>BglII</td>
<td>1</td>
<td>2</td>
<td>5471, 354</td>
</tr>
<tr>
<td></td>
<td>HinfI</td>
<td>1</td>
<td>2</td>
<td>1508, 920</td>
</tr>
<tr>
<td></td>
<td>DpnI</td>
<td>4</td>
<td>5</td>
<td>200, 311, 21, 31, 30</td>
</tr>
<tr>
<td></td>
<td>Sau3AI</td>
<td>4</td>
<td>5</td>
<td>200, 311, 22, 31, 30</td>
</tr>
</tbody>
</table>

The digestion patterns generated by MboI, MboII, BglII, AluI, DpnI, Sau3AI, DraI and AlwI did not demonstrate any difference on comparison of bands on a Southern blot (Figure 9) with Table 6, showing that their restriction sites did not overlap with the respective methylated target site of Cj0031. However, A Southern blot analysis revealed a difference in band pattern generated by HindIII digestion of *cj0031*-ON and *cj0031*-OFF variants. The presence of a 1.425kp fragment in the case of *cj0031*-ON variants (Cj0031 methylated genomic DNA) and 1kb fragment in the case of WT-OFF variants.
(unmethylated genomic DNA) [Figure 8; Table 6], indicating that there was a loss of cleavage by HindIII within probe III.

**Figure 8:** Southern blot showing the digestion of restriction sites contained in probe III locus in the genome extracted from ON and OFF variants by various methylation sensitive restriction enzymes:

The top of the figure shows the names of various enzymes used to digest the genomic DNA derived from ON and OFF variants, which are MboI, MboII, BglII, AluI, DpnI, Sau3AI, DraI and AlwI. Positive (+) and negative (-) signs in the row below the figure represent the ON and OFF variants of *cj0031*. The sizes are shown by arrows for the restriction fragments following digestion with HindIII of ON and OFF variants.

The sizes of fragments produced by HindIII for protected or restricted HindIII sites in probe III were calculated (Figure 9; Table 4). In the case of inhibition of the HindIII site in probe III, a 1.425 kb fragment should be generated otherwise it should be cleaved.
into two fragments (425bp and 1kb). The loss of cleavage for \textit{cj0031}-ON variant on HindIII site suggested that terminal adenine residue was methylated by Cj0031 on this site (5’-AAGCTT-3’). A 425bp fragment for the OFF variant could not be detected on the Southern blot because it shared a very short sequence of homology (40bp) with the 3’ end of probe III. During the washing step, the hybrid might not be stable and hence binding is destabilised. However, the hybrid of length 980bp shared a long stretch of homology (450bp) with probe III, which was detected on the Southern blot.

\textbf{Figure 9: Diagram of the expected cleavage pattern detected by Probe III of HindIII/cj0031 methyltransferase site overlap in DNA from methylated (\textit{cj0031}-ON) and non-methylated (\textit{cj0031}-OFF) PV states:}

The position of the HindIII site in probe III has been shown at point B. The HindIII sites at positions A and C were located outside the probe. In the case of \textit{cj0031}-OFF variant, the HindIII/Cj0031MTase site overlap remained unmethylated resulting in the cleavage of the B site. Thus a HindIII digest gives rise to two fragments of sizes 1kb and 425bp. However, in the case of a \textit{cj0031}-ON variant, HindIII/Cj0031MTase sites overlap had been methylated by Cj0031 methyltransferase (Cj0031MTase) causing the inhibition of digestion, leading to generation of a single fragment of size 1425bp.
4.2.3: Identification of recognition site of Cj0031 methyltransferase:

It was hypothesized that the restriction site of HindIII corresponded to the target site of Cj0031 methyltransferase. However, another HindIII site was present in Probe I, whose digestion did not show any difference in restriction pattern between cj0031-ON and cj0031-OFF variants. Thus, the restriction site of HindIII found in probe I (figure 7) was not protected by the putative Cj0031 MTase mediated methylation; however, the site in probe III was protected. This clearly indicated that the HindIII target site did not correspond to the recognition site of the Cj0031 methyltransferase site but overlapped with it in this one location in Probe III.

The flanking sequences up to 6bp downstream and 6bp upstream of HindIII sites (5’-AAGCTT-3’) were scanned within probe I and probe III, and an obvious difference was revealed on one side of the HindIII site (Figure 9). The region on downstream 5’-end of Hind III in probe III contained AT rich sequence (AAAGA, AAAGAA), while the upstream 5’-end carried GC rich sequence (GGGCAA). Based on the 99% homology of Cj0031 with cjeFV whose recognition sequence was characterized to be CCCGA (Murray et al., 2012), the potential recognition sequence for Cj0031 was proposed to be GGGCA or GGGCA which was located on upstream 5’-end of Hind III in probe III.
Figure 10: Comparison of digestion patterns of HindIII sites located within probe I and probe III

The top panel shows a Southern blot of the genomic DNA from cj0031-ON and cj0031-OFF variant digested with Hind III digest and hybridised with either probe I or probe III. Digestion was performed by incubating a mixture of 1ug of genomic DNA of ON/OFF variants with 2 units of HindIII at 37°C for 2 hours. 15ul of digest loaded on 0.1% agarose. Lane 1 (both Southern blots) from left side: cj0031-ON, Lane 2: cj0031-OFF. The red arrows show the observed sizes of the bands. In the bottom panel, the flanking sequences of the Hind III site and positions of the HindIII and putative Cj0031 sites are shown for probe I and probe III.

This led to the hypothesis that either 5’-CGAA-3’, 5’-CGAAG-3’, 5’-CCGAA-3’, 5’-CCCGAA-3’ or 5’-CCGAAG-3’ was the target recognition site for Cj0031 methyltransferase. These recognition sequences were proposed for Cj0031 based on the fact that length of recognition sites for type II RM system can vary between 4-6bp.
There are two adenine residues in the recognition site of HindIII. In order to determine which adenine residue is methylated in the target site of HindIII (5’-AAGCTT-3’), the REBASE site was consulted with regard to the methylation sensitivity status of HindIII. It was found that inhibition of digestion could take place due to a methylated cytosine or methylation of the terminal adenine residue within the HindIII restriction site (Table 5), indicating that the Cj0031 MTase could either potentially methylate adenine or cytosine residues.

Table 5: Methylation sensitivity status of HindIII site

<table>
<thead>
<tr>
<th>HindIII</th>
<th>Site cut</th>
<th>Cleavage impaired</th>
<th>Sites not cut</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m6</td>
<td></td>
<td>m6</td>
</tr>
<tr>
<td></td>
<td>AAGCTT</td>
<td></td>
<td>AAGCTT</td>
</tr>
<tr>
<td></td>
<td>TCGAAA</td>
<td></td>
<td>TCGAAA</td>
</tr>
<tr>
<td></td>
<td>(5% cleaved)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Methylation sensitivity status of HindIII site

<table>
<thead>
<tr>
<th>HindIII</th>
<th>Site cut</th>
<th>Cleavage impaired</th>
<th>Sites not cut</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m6</td>
<td></td>
<td>m4</td>
</tr>
<tr>
<td></td>
<td>AAGCTT</td>
<td></td>
<td>AAGCTT</td>
</tr>
<tr>
<td></td>
<td>TCGAAA</td>
<td></td>
<td>TCGAAA</td>
</tr>
<tr>
<td></td>
<td>m6</td>
<td></td>
<td>m4</td>
</tr>
</tbody>
</table>
In order to examine whether methylation at adenine (m6A) or cytosine (m4C) was responsible for the inhibition of digestion, the AluI restriction enzyme was selected. The restriction site of AluI (5’-AGCT-3’) overlaps with that of HindIII (5’-AAGCTT-3’). The methylation status of AluI was examined in REBASE, which showed that digestion of an 5’-AGCT-3’ site by AluI is inhibited if the cytosine or terminal adenine residue (second adenine in the HindIII site) were methylated (Table 6).

Table 6: Methylation sensitivity status of HindIII site

<table>
<thead>
<tr>
<th>Site cut</th>
<th>Cleavage impaired</th>
<th>Sites not cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGCT</td>
<td>m6A</td>
<td>AGCT</td>
</tr>
<tr>
<td>TTCCG</td>
<td></td>
<td>TTCCG</td>
</tr>
<tr>
<td>AGCTT</td>
<td>m4A</td>
<td>AGCTT</td>
</tr>
<tr>
<td>TTCCGA</td>
<td></td>
<td>TTCCGA</td>
</tr>
</tbody>
</table>

It was found that the digestion patterns from the AluI (AGCT) restriction site overlapping with the HindIII site in Probe III did not display any difference for cj0031-
ON (methylated by Cj0031MTase) and cj0031-OFF (unmethylated by Cj0031MTase) (Figure 10) variants and had sizes indicative of no protection. Thus, only one band of size 425bp was detected on the Southern blot, which is the size expected for an unprotected AluI site. The other band, which should be 188bp, was not observed as it shares a very short region of homology of 50bp with Probe III and so it might have been washed away during the high stringency washing resulting in the visualisation of only the 425bp fragment on the Southern blot. This result shows that the HindIII restriction site was protected due to methylation of the terminal adenine residue.

**Figure 10: The restriction pattern of AluI site overlapping with HindIII site in Probe III:** Panel A shows the fragments detected by Probe III in the genomic DNA of Cj0031 variants following cleavage by AluI. A 425bp fragment is observed on both variants. Panels B and C display the positions of AluI sites and the restriction pattern for Probe III expected for the methylated and non-methylated states. The dotted line shows the length of the sequences lying beyond actual length of probe III. This shows that if the A or C residues within the AluI site in probe III were methylated, its digestion
would have been inhibited, generating a fragment of size 613bp whilst a lack of protection results in fragments of 425bp and 188.

4.2.3.1: Construction of *cj0031* Knockout Construct

In the section 4.2.2.8.3, it was shown that the HindIII recognition site in probe III demonstrated inhibition of digestion in a *cj0031*-ON variant but not an OFF variant. A mutant of *cj0031* and its complement was constructed in order to confirm the methylation activity of Cj0031MTase.

4.2.3.1.1 Amplification, Ligation and Transformation of M0031-Up fragment and S0032-down

In order to prepare the knock-out mutant, primers were designed to amplify 728bp fragment (M0031-up) from upstream of Cj0031. The fragment was amplified by using the primers: 0031-2108-F with a PstI restriction site on its 5’ end and 0031-R-2832-Bam. The second fragment (S0031-down) having a size of 621bp was amplified from the downstream region of *cj0032* by using the primer set: 0032-Bam-F and 0032-R (Material and Methods). The M0031-up and S0032-down fragments were A-tailed by incubating them at 72°C for 45 minutes with TAQ DNA polymerase, purified using the PCR clean-up kit, and separately ligated into pGEMT easy vector. Following transformation into *E. coli* strain (DHα5), the white colonies representing the transformants were selected, and plasmid minipreps (pGMET-M0031-up; pGMET-S0032-down) were performed by following the protocol provided with Qiagen Minipreps kit. In order to confirm the identity of the inserts (M0031-up and S0032-down), recombinant plasmids were sequenced by utilizing the primer set binding inside the pGEMT easy vector on both sides of the insert. The sequencing of recombinant
plasmids confirmed the cloning of respective insert into the cloning vector (data not shown).

4.2.3.1.2: Sub-cloning of M0031-up and S0032-down fragments into pUC18 vector

The pGMET-M0031-up recombinant plasmid was subjected to digestion with PstI and BamHI to release the M0031-up fragment. The recombinant plasmid pGMET-S0032-down was digested with BamHI and EcoRI to recover the S0031-down fragment. Similarly, pUC18 vector was digested with PstI and EcoRI in order to linearise it. All three fragments were purified and then a three-way ligation reaction was performed by combining the M0031-up, S0032-down and pUC18 fragments in a 3:3:1 ratio, followed by transformation into *E. coli* strain (DHα5). Five recombinant clones were selected and subjected to digestion with EcoRI+BamH and BamHI only in separate reactions. The digestion with EcoRI+BamHI released the S0032-down (621 bp) fragment. However, the BamHI digestion resulted in linearisation of the recombinant plasmid as was expected.

In order to insert a kanamycin (KAN) antibiotic cassette, pUC18-M0031-up-S0032-down recombinant plasmid was digested with BamHI to linearize it. The KAN cassette fragment (~1.4kb) was recovered from pJKM30 vector by digesting it with BamHI. The KAN cassette fragment and linearized pUC18-M0031-up-S0032-down recombinant plasmid were ligated and a subsequent transformation into *E. coli* strain (DHα5) was carried out. Five recombinants plasmids were digested with PstI+BamHI and BamHI+EcoRI. The digestion of recombinants with PstI+BamHI produced the expected fragments of sizes: 3.2kb, 724bp, and 1.5kb. Similarly expected fragments of sizes: 3.348kb, 1.5kb and 600bp were obtained following digestion of recombinants with BamHI+EcoRI (Figure 11).
Figure 11: Digestion of ∆0031::kan clones with PstI + BamHI and BamHI + EcoRI:

The PstI + BamHI and BamHI + EcoRI digests of five ∆cj0031::kan clones, 5ul of recombinant plasmids were digested with respective enzyme(s) in a total volume of 20ul, a 10ul sample of digestion mixture was loaded onto a 1% agarose gel, Lane 1: standard molecular weight marker, Lane 2, Lane 4, lane 6, lane 8 & lane 10: minipreps 1-5 digested with PstI and BamHI, Lane 3, lane 5, lane 7, lane 9 and lane 11: minipreps digested with BamHI and EcoRI.

HindIII generates a different restriction pattern for the KAN cassette with forward and reverse orientations in recombinant plasmids The HindIII restriction analysis revealed that four clones were found with KAN in a reverse orientation while one of them possessed the KAN cassette in a forward orientation. The digestion of four recombinant plasmids with HindIII yielded 229bp, 276bp, 273bp, 1.9bp, 2.8kb, indicating a reverse orientation of Kan cassette in them. However, one out of five plasmids displayed 229bp,
276bp, 483bp, 1.7kb and 2.8kb following the HindIII digestion, showing KAN in a forward orientation in this recombinant plasmid (Figure 12).

**Figure 12: Determination of Orientation of KAN cassette by digesting Δ0031::kan with HindIII:**

The HindIII digest of five Δcj0031::kan clones, 5ul of recombinant plasmid was digested with respective enzyme (s) in a total volume of 20ul; a 10ul sample of digestion mixture was loaded onto a 1% agarose gel, Lane 1: standard molecular weight marker. Lane 2, Lane 3, lane 5 & lane 6: - minipreps digested with HindIII.

The overview of construction of knock-out mutant has been shown in figure 13.
4.2.3.1.3: Recombination of Δ0031::kan construct into *C. jejuni* chromosome:

The Δ0031::kan construct was recombined into the chromosome of *C. jejuni* strain 11168 by electroporation with pUC0031-kan and then kanamycin resistant
transformants were isolated. The insertion was verified by performing PCR reactions using different sets of primers (Figure 14).

Figure 14: Confirmation of recombination of ∆0031::kan construct into C. jejuni 11168 chromosome by PCR reaction:

Panel A shows the primers used, their binding sites and expected sizes of PCR products. Panel B shows the sizes of PCR products following amplification reactions using DNA from one CjΔ0031::kan transformant. 5ul out of 10ul a PCR reaction was loaded onto a 1% agarose gel, Lane 1: standard molecular size marker, Lane 2: 2.8 kb product obtained by using primer set: cj0031-2108-F and 0032-End-R, Lane 3: 2.4 kb product obtained by using primer set: cj0031-N-term and 0031-2802-R, Lane 4: 5 kb product by using primer set: 0031-N-term and 0031-CTerm.

4.2.4: Construction of complementation system for Cj0031-knock-out mutant:
4.2.4.1: Acquisition of Vector:

The pCfdxA complementation vector was acquired from Gaskin et al. (2007). The vector contains two fragments of cj0046, a pseudogene, allowing recombination into the locus of cj0046 in the C. jejuni NCTC11168 strain genome after transformation.

4.2.4.2: Amplification of target gene:

Primers were designed for amplification of the entire cj0031 gene. The N-terminal primer contained a BsmBI (Esp3I) site at its 3’-terminal while the C-terminal primer contained an NcoI site at its 5’-terminal. The cj0031 gene was amplified from NCTC11168 genomic DNA by using the Phusion PCR reaction (Material and Methods).

4.2.4.3: Processing of Vector and Amplified cj0031 product for Ligation:

The pCfdxA vector was linearized by digesting it with BsmBI (Esp3I). The linearized vector was then dephosphorylated by following the instructions on the Dephosphorylation Kit (Roche). In order to digest the PCR product of cj0031 with Esp3I, the product was first cleaned up with a PCR clean-up Kit. As Esp3I does not fully digest DNA treated with ethanol, the cj0031 amplified product was subjected to ethanol precipitation in order to get rid of ethanol, and then digested with Esp3I. The Esp3I digested cj0031 PCR product was then subjected to further digestion with NcoI. The resulting digestion mixture was cleaned up by using a PCR clean-up kit. Both
processed pCfdxA vector and *cj*0031 PCR product (insert) were ligated together and then transformed into DH5α competent cells. A total of 30 chloramphenicol resistant colonies were picked. Half of each colony was streaked back on the LA+CAT plates while the rest was utilized in making the boilate DNA preparations. The colonies were checked by PCR with an internal primer set (2108-F; 2809-R) to verify the presence of the correct insert. Two colonies (5&6) were found to be positive. Both colonies were further tested by performing PCR reactions using different combination of primers (Figure 15A). PCR products of the expected sizes were obtained (Figure 15B). Three primer sets: CAT-INV-F, 0031-2108-R; 0046-INV-F, 0032-R; and CAT-INV-F, 0031-2832-R were used to amplify the *cj*0031 gene recombined into the vector, which generated the expected sizes of 2.4kb, 1.1kb and 2.4kb, respectively (Figure 17A). These results suggested that the *cj*0031 gene was cloned successfully into the vector.
Figure 15: Digestion and PCR products of pCfdx-cj0031 construct:

Panel A shows the primers used, their binding sites and expected sizes of PCR products. Panel B shows the PCR products obtained from pCfdxa-cj0031 and recombinant plasmid DNA digested with EcoRI, loaded onto a 1% agarose gel. Lane 1: standard molecular weight marker, Lane 2: EcoRI digest, Lane 2: PCR using primer set: 0046-INVF and 0032-F, Lane 3: PCR product using primer set: CAT-INVF and 0031-2108-R, Lane 4: PCR using primer set: CAT-INVF and 0031-2832-R.

The insert in pCfdxA must be oriented such that the inserted gene is downstream of the CfxA promoter in order to be expressed. In order to confirm the orientation in pCfdxA-cj0031, digestion of clones 5 & 6 with EcoRI was performed. cj0031 gene cloned in the
correct orientation would generate fragments of sizes: 1.2kb, 1.8kb, 2.2kb and 2.4kb whilst those for the incorrect orientation would be 1.2kb, 1.4kb, 1.8kb, and 3.2kb (Figure 16). The digestion profile of EcoRI showed that insert was in the correct orientation with respect to the promoter in both clones (Figure 15B). Moreover, a PCR with the CAT-INVf and 0031-2832-R primers (Figure 15B) also provided evidence about the orientation of *cj0031* relative to the CAT promoter.

The sequencing of pCfdxA-*cj0031* was also orchestrated in order to verify the sequence of *cj0031* recombined into the vector. The sequencing of the polyG tract region was performed by using the primer set (0031-F and 0031-R) flanking that region in *cj0031*. It was found that clones contained a G9 repeat tract, indicating that *cj0031* was in frame in all of the recombinant clones.

**4.2.4.1: Recombination of the *cj0031* complement construct into the *C. jejuni* chromosome:**

The recombinant plasmid was transformed by electroporation into *C. jejuni* 11168Δ0031::kan strain in order to integrate it into the chromosome. A total of 15 chloramphenicol-resistant colonies were picked and insertion of the complementation sequences was verified by PCR using the strategy for verification (Figure 16A, 16B). Primers- *cj0046*-INVf and *cj0032*-F- were used to amplify the C-terminus of *cj0031* and part of *cj0046*. This PCR gave the expected size of fragment (1.1kb). Moreover, a 3.2 kb fragment of the expected size was generated when CAT-INVf and *cj0031*-2832-R were used to amplify the N-terminus of *cj0031*. This indicated that recombination of
the complementation construct into *C. jejuni* strain 11168 genome occurred at the right locus. This construct is referred to as NCTC11168Δ0031c.

### 4.2.5: Loss of the Methylation Phenotype by Mutation of the Cj0031 Methyltransferase:

Chromosomal DNA was extracted from *cj0031*-ON, *cj0031*-ON, NCTC11168Δ0031:kan and NCTC11168Δ0031c, and digested with HindIII under optimum conditions. A Southern blotting analysis was carried out to transfer the restricted fragments from an agarose gel to a Nylon membrane. After DNA fixation, the membrane was probed with DIG-labelled PCR product of probe III. The HindIII site in this probe remained protected in the *cj0031*-ON variant, and this phenotype was partially restored by complementation with full-length *cj0031* as shown by the presence of 1.425kb band in lane 2 of the Southern blot (Fig. 16). However, it was discovered that this target site remained unprotected in *cj0031*-OFF and in the knock-out mutant Δ0031:kan as shown by the presence of 1kb band (Figure 16). This indicated that Cj0031 methyltransferase possessed methylation activity and that the HindIII target site overlapped with its recognition sequence.
Figure 16: Analysis of the HindIII digestion patterns probed with probe III for the knock-out and complementation mutants.

The Southern blot was developed by transferring the DNA fragments from a 0.1% agarose gel to positively charged Nylon blotting membrane, followed by incubation with DIG-labelled probe III and washing steps. Bands on a southern blot were obtained by digesting genomic DNA with Hind III, all digestions performed by incubating a mixture of 1ug of genomic DNA from all strains with 2 units of HindIII at 37°C for 2 hours. A 15ul of digestion mixture loaded on 0.1% agarose, followed by Southern blot analysis. Lane 1: cj0031 mutant, Lane 2: complemented cj0031 mutant, Lane 3: cj0031-OFF variant (clone 2), Lane 4: cj0031-ON variant, Lane 5: cj0031-OFF (clone 1)

4.2.6: Potential Recognition Site for the Cj0031 methyltransferase:

The HindIII site overlapping with the putative Cj0031 methyltransferase site in probe III was 5’-CCCGAAGCTT-3’ . The genome of NCTC11168 was searched for some more sites with the same sequence but unfortunately, the specific HindIII site displaying the inhibition (5’-CCCGAAGCTT-3’) was unique in the genome of NCTC11168 strain. Based on the knowledge of recognition sequence length for type II RM systems (Introduction Chapter, Section 1.6.1.2), it was hypothesized that the length of the
Cj0031 methyltransferase site should be 4-6bp. The overlap of the Cj0031 recognition sequences must be 5’ to the HindIII site, based on the 99% homology of Cj0031 with cjeFV whose recognition sequence was characterized to be CCCGA (Murray et al., 2012), the potential recognition sequence for Cj0031 was proposed to be GGGCA or GGGGCA which was located on upstream 5’-end of Hind III in probe III.

![Figure 17: Nucleotide sequence flanking Hind III site in Probe I and Probe III:](image)

Green coloured sequence is HindIII site, red coloured small “m” represents the methylation state of adenine residue, if it is methylated by Cj0031 methyltransferase, should be protected from the HindIII digestion. HindIII cuts its recognition site if adenine remains unmethylated.

Thus a set of putative recognition sequences were constructed based on flanking sequences of the HindIII site in probe III, which are 5’-CCCGAA-3’, 5’-CCGAGA-3’, 5’-CCCGA-3’ and 5’-CGAGA-3’. The frequencies of these sites were analysed in the
C. jejuni strain 11168 genome and recorded (Table 7). The proposed sites of Cj0031 were GC rich. For comparison, the frequency of 6-bp AT rich sites of HindIII were quantified through the genome. It was found that the AT rich sites of HindIII occur at a 4-8 fold higher rate than the proposed GC sites of Cj0031.

Table 7: The frequencies of proposed recognition sites of Cj0031 methyltransferase

<table>
<thead>
<tr>
<th>Putative Recognition Sites</th>
<th>Frequency of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-CCCGAA-3’</td>
<td>87</td>
</tr>
<tr>
<td>5’-CCGAAG-3’</td>
<td>121</td>
</tr>
<tr>
<td>5’-CCCJA-3’</td>
<td>157</td>
</tr>
<tr>
<td>5’-CGAA-3’</td>
<td>187</td>
</tr>
<tr>
<td>5’-CGAA-3’</td>
<td>227</td>
</tr>
<tr>
<td>5’-AAGCTT-3’ (HindIII)</td>
<td>821</td>
</tr>
</tbody>
</table>

In order to test the above mentioned potential recognition sites of Cj0031-methytransferase, C. jejuni 11168 genome was searched for the other Hind III sites overlapping with CCG sequences to its upstream 5’-end. 121 such sites in the genome were discovered. One of these sites was selected for further analysis. A PSR-HindIII-CCG probe was designed containing a HindIII/CCGAA or HindIII/CGAA overlaps. A PSR-HindIII-CCG probe was amplified by using primers: PSR-HindIII-CCG-F and PSR-HindIII-CCG-R (Materials and Methods). The resulting PCR product was ethanol precipitated and DIG labelled as described in the Materials and Methods section (2.5.3,
The genomic DNA extracted from *cj0031*-ON, *cj0031*-OFF, ∆*cj0031*::kan and complement ∆*0031*::cat were digested with HindIII, separated, Southern blotted and probed with a DIG labelled PSR-HindIII-CCG probe. It was assumed that if the putative recognition sites overlapping with the HindIII restriction site were CCGAA or CGAAG, then restriction by HindIII would be inhibited at a HindIII site in PSR-HindIII-CCG probe for *cj0031*-ON and complement ∆0031::cat resulting in the presence of three fragments 96bps, 930bp and 230bp. Otherwise, a pattern of fragments after restriction by HindIII would appear on the Southern blot in the form of four fragments: 96bps, 170bps, 760bp and 30bps. The Southern blot analysis revealed the presence of 4 fragments (96bps, 170bps, 760bp and 30bps) (Figure 3A, Figure 3B), ruling out the possibility of 5’-CCGAA-3’ and 5’-CGAAG-3’ as a recognition site for Cj0031-methytransferase. This strengthened the assumption that a Cj0031MTase recognition site should be CCCGAA. The proposed recognition site 5’CCCGA is similar to the 5’GGGCA site of CjeFV found through an in silico search in REBASE (Section 4.2.1.3).
Figure 18: Digestion pattern following HindIII digestion and probing by PSR4-HindIII-CCG probe:

**Panel A:** the number of fragments produced by HindIII following digestion of the HindIII site overlapping with 5'CCGAA, these fragments were predicted by feeding the locus of PSR4-HindIII-CCG probe from C. jejuni strain 11168 into Clone Manager programme where hypothetical restriction by HindIII was performed on it. **Panel B:** shows bands on a southern blot were obtained by digesting genomic DNA with Hind III, all digestions performed by incubating a mixture of 1ug of genomic DNA from all strains with 2 units of HindIII at 370C for 2 hours. A 15ul of digestion mixture loaded on 0.1% agarose, followed by Southern blot analysis. Lane 1: cj0031-ON, Lane 2: cj0031-OFF, Lane 3: cj0031 mutant, Lane 4: complemented mutant of cj0031. The arrows shows the fragment pattern following Hind III.
4.2.7: Discussion:

A phase variable *cj0031* - a putative type IIG methyltransferase- was characterized in this study. There are two MTases and two restriction modification system in the REBASE database for *C. jejuni* 11168 strain. This is the only RM system containing repeat tract and is subjected to ON/OFF phase variation events. The main findings of the study describe that *cj0031* shares a considerable homology with other putative restriction modifications systems. First hand empirical evidence has been provided about the methylation activity of Cj0031 and an attempt to propose its recognition site was performed. These findings will be elaborated in the subsequent sections.

4.2.7.2: Cj0031 is a putative RM system- an in-silico analysis:

DNA MTases are categorised into different groups based on the position and base of methylation in the genome and nature of the catalytic reactions. These methyltransferases methylate either adenine (m6A) or cytosine (m4C or m5C). The in-silico analysis revealed that Cj0031 shared some homology with a group of previously uncharacterized adenine MTases. An obvious feature of Cj0031 is that it possesses the adenine methylation signature motif (GNPPY) which serves as the catalytic motif IV for adenine specific methyltransferases (Posfai *et al.*, 1989)
Figure 19: Position of Adenine MTase signature motif in Cj0031: An arrow shows the adenine MTase signature motif positioned at 813-817aa in the protein sequence of cj0031. The protein sequence segment shown (here) was derived from UniprotKB database.

This conserved motif IV is considered as a characteristic of the adenine subfamily (Furuta and Kobayashi, 2011) in contrast to Pro-Cys type IV motif characteristic of cytosine subfamily (Henderson et al., 2010)

Moreover, Cj0031MTase shared a considerable homology (24%) with the N-terminus of T257_ECOLI type IIG restriction enzyme Eco57I whose methylation and endonuclease activities were previously shown empirically (Janulaitis et al; 1992). In particular, there was high conservation of the putative functional domains of Eco57I (Figure 2) , suggesting that Cj0031MTase might be in possession of both methylation and endonuclease activity. It also had 40.5% homology with a putative RM system HP1571 of H. pylori. The C-terminus of Cj0031 is similar to the C-terminus of MTB3_BACAR modification methylase BanII (24%) and HP1571 (47.4%),(Figure 2A and 2B) indicating the possibility of the presence of a methylation domain towards the C-terminus of Cj0031MTase (www.xbase.ac.uk). These findings suggest that both methylation and endonuclease domains are potentially located at the N-terminus of
Cj0031MTase (900aa) and the C-terminus may be encoding a second methylation domain (300aa).

The phylogenetic analysis of cj0031 showed that alleles are present in other strains of *Campylobacter*, indicating that cj0031 gene may be adapted by various strains to suit their own biological functions. In addition, orthologs of cj0031 were also present in *H. pylori*, a close relative to *C. jejuni*, as well as *Bacteriodes, Lactobacillus, Streptococcus* and *Microcystis*. This finding highlighted the fact that RM modules were disseminated to various phyla through horizontal gene distribution (Dziewit *et al.*, 2011). The homologs of cj0031 in other strains of *C. jejuni* were 90-99% similar in sequence to the cj0031 in NCTC11168 strain, showing the conservation of function of cj0031 in other strains of *Campylobacter*. As one goes down the phylogenetic tree, it can be found that the homology index is lowered, which indicates that only conserved motifs are retained in species along the phylogenetic tree. Most of these orthologs of cj0031 were uncharacterised RM systems, and hence there was a need to characterise cj0031 to show whether it is functionally active or not.

**4.2.7.3: Structural similarity of cj0031 RM to type IIG RM system**

A hypothetical domain model was developed by using the BLAST tool, which showed that Cj0031 contained two separate domains for the methyltransferase and endonuclease functions with the latter domain being fused to the methyltransferase domain, an unusual feature of RM systems. Intriguingly, this fusion of endonuclease and methyltransferase domains is the key aspect of Type IIG restriction modification systems differentiating them from the type IIS restriction-modification (Kobayashi, 2001). This can be viewed in a hypothetical model of Cj0031 (figure 5) in which the
restriction modification methylase Eco57I domain (808-919aa) possesses the similarity with Eco57IR and N6-MTase motif (GNPPY) signature, demonstrating the fusion of endonuclease and methylation activity in the same domain of the protein. The restriction modification Eco57I domain is located in the methyltransferase subunit of Eco571RM system which is referred to as a typical example of a type IIG RM system. Hence, it is suggested that the Cj0031 RM system appears more like a type IIG RM system rather than a type IIS RM system. The function and purpose of these two domains in Cj0031 as a putative type IIG RM system may be to protectively methylate the recognition sequence in a cooperative manner as suggested by Strieps and Yasbin (2002) for type IIG systems. Towards the C-terminal end of Cj0031, a TaqI-like C-terminal specificity domain (1029-1182 aa) is located containing a Proline-Lysine motif necessary for binding with the recognition sequence of a methyltransferase (Figure 5). Proline-Lysine motifs represent the conserved motif III(CMIII) for all N6-MTases (Timinskas et al., 1995). The N-terminus of Cj0031 contains another conserved motif (F-G-G) named CMI (based on scheme of Timinskas et al., 1995) used by methyltransferases for binding with AdoMet substrate and acts cooperatively with the N6-MTase motif to transfer adenosyl-methionine from substrate to the target base in the recognition sequence (Timinskas et al., 1995; Klimasauskas et al., 1989). The N6-MTase motif (GNPPY) was numbered as CMIII according to the classification scheme proposed by Timinskas et al. (1995) and as CMIV by Klimasauskas et al. (1989) (Figure 20).
Figure 20: Domains and Conserved Motifs of Cj0031. The numbers on the left side of lines represent the EBI reference numbers CMI, CMII/CMIV, CMIII are the conserved motifs in the domain designated on left side of lines. The figure was derived by feeding the protein sequence of Cj0031 in “EBI Domain and Motif Search Database”, motifs and domains were curated by searching the homology with other methyltransferases present in the database.

4.2.7.4: Cj0031 methyltransferase possesses methylation activity:

In order to investigate the functional status of methylation activity of Cj0031MTase, *cj0031*-ON variants (methylated) and *cj0031*-OFF (unmethylated) variants were obtained. The problem associated with this analysis was phase variation. The ON or OFF variants were subject to phase variation events each time they were grown on media. This means that at any stage of growth there was a potential for the presence of an ON>>OFF mixture of variants in the case of predominantly ON-variants or
OFF>>ON phase variants mixture in the case of predominantly OFF variants. However, measurements of the phase variation rates using the \textit{cj0031lacZ} reporter constructs indicate that the proportion of variants at any stage was small (0.01%). So, during analysis of the methylation activity of Cj0031MTase, it was assumed that ON and OFF variants were not an absolutely pure variant in terms of their ON or OFF status of methylation activity. The presence of minor proportion of these impurities gave rise to the faint signals on Southern blots during the Southern blotting analysis of \textit{cj0031-ON} and \textit{cj0031-OFF} variants. For example, in a Southern blot, a 1.425 kb faint band can be observed along with the main 1kb band in a \textit{cj0031-OFF} variant while conversely a faint 1kb band can be found along with the main 1.425kb band for \textit{cj0031-ON} variant (Figure 8). A second possibility may be related to the incapacity of methyltransferases of Type II RM systems to methylate all sites in the genomic DNA sufficiently (Gerome \textit{et al.}, 2006). It might be speculated that the amount of the Cj0031 methyltransferase could not methylate all recognition sites effectively; rendering a fraction of sites unmethylated which caused the appearance of both bands on blots for WT-ON. This explanation does not predict the presence of faint bands on blots for WT-OFF. Hence, the first possibility seems more plausible than the second one.

To detect methylation activity, the strategy of inhibition of digestion by restriction enzymes was used. This strategy has been useful in identification of the methylation activity and subsequent determination of recognitions sequence of methyltransferases of a type III RM system (Srikhanta \textit{et al.}, 2009).

A total of 19 methylation sensitive (adenine and cytosine) restriction enzymes, covering 25 restriction sites, were utilized during the analysis. Some of the enzymes like DraI, Hind III, BglII, MboII were located in all probes; so they were used for all of three probes in order to test restriction patterns at their restriction sites occurring in different
sequence contexts in the *C. jejuni* strain 11168 genome. Only one site, a HindIII site in Probe III, was observed to exhibit a differential pattern of digestion between ON and OFF variants of *cj0031*, indicating that it overlapped with the recognition site of Cj0031 methyltransferase. Sequencing of probes III showed that the HindIII site was intact in probe III, indicating that the difference observed for ON and OFF variants did not result from mutation at the region supposed to be having the Hind III site.

The inhibition of HindIII site in probe III was verified by using the mutant and complemented mutant of *cj0031*. Thus, the Southern blot analysis of WT-ON, WT-OFF, 11168Δ*cj0031::kan* and complement11168Δ*cj0031::kan* revealed that restriction was inhibited at the HindIII site in probe III for the ON variants of the *cj0031* gene, in which it is assumed that the MTase has methylated the genomic DNA. In contrast, the HindIII site of probe III was not protected in the case of WT-OFF phase variants of *cj0031* and a knock-out mutant of this gene, 11168Δ*cj0031::kan*. This phenotype was partially restored by complementation of 11168Δ*cj0031::kan*, suggesting that the methylation phenotype observed for the *cj0031*-ON variant was not caused by a mutation in some other gene. The partial restoration of the phenotype may be due to the fact that the Cj0031 enzyme might have been produced in lower quantities under the CAT promoter in the pCfdxA cloning vector used to clone *cj0031*. This lower expression may not be sufficient to methylate all recognition sites of the Cj0031 methyltransferase hence the partial phenotype observed in the Southern blots. To note, the ON repeat tract (G9) of *cj0031* in the complemented mutant also undergoes phase variation and can generate the OFF phenotype at rate 0.01% (Bayliss et al., 2012). So the minor impurity of OFF variants in complemented mutant could lead to generation of strong 1 kb band characteristic of restriction digestion of OFF variants by HindIII in a Southern blot for probe III.
In short, it can be concluded that Cj0031 methylated the adenine residue in a HindIII site within probe III leading to inhibition of restriction by Hind III, clearly demonstrating that Cj0031 possessed methyltransferase activity.

4.2.7.6: Potential recognition sequence for Cj0031 methyltransferase: 5’-CCCGAA-3’ or 5’-CCCGA-3’

The HindIII recognition site is 5’-AAGCTT-3’, which is only blocked if the terminal adenine residue is methylated. If this site was the putative recognition site of the Cj0031 methyltransferase, then it should be blocked at all loci in the genome of an ON-variant. Probe I and Probe III contained the HindIII restriction sites, but digestion was only inhibited in probe III, suggesting that not all HindIII sites in the genome correspond to Cj0031 MTase recognition sites.

On inspection of flanking sequences of both HindIII sites at these loci, it was found that the flanking sequences were different. This led to the hypothesis that the recognition site of Cj0031 methyltransferase overlapped with the HindIII site either to the upstream 5’-end or downstream 5’-end and that a terminal A of the HindIII site in Probe III was methylated by Cj0031 methyltransferase, which caused the inhibition of HindIII restriction. Murray et al., (2012) characterized the cjeFV RM system and identified its recognition sequence to be CCCGA. As this cjeFV gene in 81176 has homology (99%) homology with cj0031, the possibility of overlap to the 5’-end (upstream) of the HindIII site in probe III was taken into consideration and various potential recognition sites were proposed for further tests. For example, the potential recognition sites proposed for Cj0031 methyltransferase were 5’-CCCGAA-3’, 5’-CCGAA-3’, 5’-CCGAAG-3’, 5’-CGAAG-3’ or 5’-CCCGA-3’. The PSR-HindIII-CCG probe was designed
containing the HindIII site 5’-CCGAAGCTT-3’. The Southern blotting analysis showed that it was not protected in PSR4-HindIII-CGG probe (Figure 18). This excluded the inclusion of 5’-CCGAA-3’, 5’-CCGAAGA-3’, 5’-CGAAG-3’ as potential recognition sites for Cj0031 methyltransferase. Thus the potential recognition site may be either 5’-CCGAA-3’ or 5’-CCCGA-3’.

A bioinformatics analysis revealed a similarity of Cj0031 encoded type IIG RM system with other RM systems of H. pylori and C. jejuni, which were yet to be characterized, excluding any possibility of comparison of Cj0031 proposed recognition site with those of its orthologs. However, the proposed recognition sequences of Cj0031 matched with those of two putative type IIG RM systems-cjeFV and HP99XIII- which show 81% and 38% identities, respectively. CjeFV present in C. jejuni strain 81176 has specificity GGGCA, while HP99XIII belonging to H. pylori and has sequence specificity for GCCTA (REBASE database). Intriguingly, the recognition sites of cjeFV and HP99XIII have GC rich contents like that of Cj0031, emphasizing that the proposed recognition sequence should be one of two options (5’-CCGAA-3’ or 5’-CCCGA-3’).

Eco57I from E.coli, a type IIG RM system, shared similarity (35%) with Cj0031, and this was previously characterised by Janulaitis et al. (1992) who determined its recognition site to be 5’-CTGAAG-3’ containing 50% GC contents. Lepikhov et al. (2001) characterized another type IIG RM system-BspLU11M - in Bacillus and demonstrated empirically that it contains 5’-GGGAC-3’. BspLU11M (530aa) shared 60% similarity with the adenine methylation domain of Cj0031. Jurenaite-Urbanaviciene et al. (2001) characterized BseMI-a type IIG RM system- and discovered 5’-CTGAG-3’ as it recognition site. From these examples, this is noted that recognition sites of type IIG RM systems are GC rich and comprise of 5 base pairs. So this data of type IIG RM system having similarity with Cj0031 methyltransferase can be used to
postulate that the recognition site for Cj0031 may be 5’-CCCGA-3’. Thus further work is needed to prove the exact recognition site of Cj0031MTase. A search for the CCCGAA/CCCGA sites was performed on lambda phage DNA, and its frequency was 3 and 10 in its genome. The CP8 and CP30 *Campylobacter* phage genomes possessed two and three sites, respectively, of CCCGAA. The presence of the Cj0031 proposed site provided an opportunity to determine the restriction activity of Cj0031 by carrying out the phage assay.
Chapter 5. Phenotypic Assays for Cj0031 Methyltransferase

5.1: Introduction:

In the previous chapter, empirical evidence has been provided about the methylation activity of the Cj0031 methyltransferase. The bioinformatics analysis had shown that Cj0031 belongs to type IIG restriction modification systems. Type IIG restriction modification systems, such as Eco57I, are similar to type III RM systems in terms of recognizing assemymetric recognition sequences, partial cleavage of the substrate and a requirement of AdoMet (Janulaitis et al., 1992).

Restriction modification systems play an important role in the protection of host bacteria. However some of the systems also hold a pivotal position in regulation of gene expression by switching genes ON/OFF through methylation of recognition sequences located in their promoter regions. The latter role of restriction modification systems is similar to the epigenetic control of the Dam methyltransferase in *E. coli*. DNA methylation mediated by the methyltransferase acts as an important signal for different DNA binding proteins to interact with each other around the cognate recognition sequences, thereby affecting the bacterial pathogenesis and other physiological processes of the host bacteria such as fitness and survival under stress conditions (Heusipp *et al*, 2007; Casadesus and Low, 2006; Wion and Casadesus, 2006).

Phase variable mod methyltransferases in some strains of *N. gonorrhoeae*, *N. meningitidis*, *H. pylori* and *H. influenzae* have been reported to coordinate the expression of a group of genes through differential methylation of their genomes. They are termed as “phasevarions (phase-variable regulons)”. The phase variable methylation
mediates random switches in expression and results in the generation of two kinds of cells, one type with a methylated promoter (ON variants) and another with an unmethylated promoter (OFF variants) [Srikhanta et al., 2009, Srikhanta et al., 2005, Srikhanta et al., 2012]. The studies also revealed that some of genes regulated by phasevarions are involved in pathogenesis and bacterial fitness by affecting key phenotypes such as adhesion and invasion of the bacterial cells to the host tissues, motility, biofilm formation, resistance to antimicrobial agents, and oxidative stress phenotype (Seib et al., 2011; Srikhanta et al., 2009; Kim et al., 2008).

Based on the empirical evidence from these studies, it was postulated that a phase variable Cj0031 methyltransferase might act as a phasevarion, and might be responsible for coordinated switching of gene expression by randomly methylating the genes involved in pathogenesis and bacterial fitness of C. jejuni strain 11168. In order to test this hypothesis, three phenotypes - adhesion and invasion, biofilm formation capability and motility - were selected for investigation.

### 5.1.2: Adhesion and Invasion:

Adhesion mechanisms and factors involved in this process are studied extensively by microbiologists in view of their importance for colonization and subsequent infection of cognate hosts. Adhesion is mediated by the surface appendages (pili) located on the outer membranes of Gram-negative and Gram-positive bacteria. The pilus or pilus like structures have not been identified in annotated open reading frames for many strains of C. jejuni. However, flagellar motility is considered an important determinant for C. jejuni’s adherence to the epithelium. It has been showed that non-motile and motile C. jejuni strains adhere to the epithelial cells in the same way if centrifugation-mediated
contact is allowed between them, suggesting that flagella likely steers the *C. jejuni* cells to the epithelium and allows the non-flagellar adhesins to initiate adherence to the epithelial surfaces (Golden *et al.*, 2000; Grant *et al.*, 1993).

Yao *et al.* (1994) reported that a *flaA* mutant reduces the motility and adherence to the INT407 cells supporting the evidence provided by Golden *et al.* (2002) and Grant *et al.* (1993), referring to the importance of structural integrity of flagella in the adherence process. These contradictory findings lead one to the conclusion that flagellar role in mediating adherence is not clear and it seems to be strain-specific. Nevertheless, flagellar secretory apparatus secretes some invasion antigens such as *CiaB* and *flaC* which are not required for motility but their effects on the invasion phenotype has been demonstrated by Konkel *et al.* (1999) and Song *et al.* (2004).

Despite the lack of identifiable structures for adhesion, many proteins have been identified in *C. jejuni*, which contribute to adhesion of bacteria to the host tissues in eukaryotic cells. A membrane channel forming proteins, CadF, has been demonstrated to increase adhesion to and invasion of eukaryotic cells by maximizing the binding capability of *C. jejuni* to fibronectin on epithelial cells (Konkel *et al.*, 1997; Moteville and Konkel, 2002; Moteville *et al.*, 2003). JlpA, a surface lipoprotein, represents another factor involved in the adherence of *C. jejuni* to the eukaryotic cells. Jin *et al.* (2003) provided direct evidence about the involvement of JlpA in enhancing adhesion of *C. jejuni* to HEp-2 cells. The JlpA-dependent adherence is suggested to play a role in causing the inflammation during pathogenesis of *C. jejuni* (Young *et al.*, 2006).

In addition, CapA, an autotransporter, has been reported to increase the adherence capability of *C. jejuni* to the host. Asghar *et al.* (2007) documented that mutation in
capA gene resulted in the reduction of *C. jejuni* capability to adhere to Caco-2 cells. It also decreased the colonization of *C. jejuni* in a chick model (Asghar *et al.*, 2007).

Hence, a complex network of proteins is involved in the adherence and invasion of bacteria to the cells, which might be subject to certain epigenetic control by the adenine methytransferase. Kim *et al.* (2008) showed that *cj1426*, a putative adenine methyltransferase, affected the adhesion and invasion. They showed that ∆*cj1426* reduced the invasion of Caco-2 epithelial cells but increased the adherence of *C. jejuni* cells to epithelial cells. The Cj0031 methyltransferase is a phase variable adenine methyltransferase which might act as a phasevarion affecting adhesion and invasion by *C. jejuni* cells. For this reason, it would be interesting to investigate the effect of Cj0031 mediated methylation on the expression of other genes.

### 5.1.3: Motility:

Motility holds a critical position in the pathogenesis of *C. jejuni*. The non-motile strains of *C. jejuni* were shown to be non pathogenic and non-colonizing in their respective hosts, revealing the importance of motility for colonization and subsequent spread of disease within a host. Motility in *C. jejuni* is mediated by the bipolar flagella which are also organs for attachment and/or leads the bacteria to sites of attachment in the epithelium. Mutations that change the glycosylation of flagella was shown to affect the motility and subsequent attachment to and invasion of host cells. Besides, aflagellated versions of *C. jejuni* strains demonstrated a decrease in the colonization of the chick and mice models, suggesting the influence of flagella on the virulence of bacteria (Guerry *et al.*, 1992, Yao *et al.*, 1994). Wassenaar and Blaser (1999) provided evidence about the markedly reduced invasion capability of aflagellated *C. jejuni*.
Biosynthesis of flagella is mainly regulated by a two-component system (FlgS and FlgR). Other protein determinants are also implicated in the regulation of flagella and motility. For example, *ciaB*, *fliA* (Sigma 28), sigma 54 (*rpoN*) and sigma70 (*rpoD*) are involved in the expression of many genes which exert their influence on motility of *C. jejuni* (Hendrixson, 2006; Carrillo *et al*., 2004; Wosten *et al*., 2004; Hendrixson and DiRita, 2003; Jagannathan *et al*., 2001). Some other regulators such as FlhC and FlhD are found in other flagellated bacterial species but they are nonexistent in *C. jejuni*, raising the possibility of the presence of some other control mechanism like epigenetic control in regulation of flagella and motility (Parkhill *et al*., 2000).

Flagella of *C. jejuni* also determines its chemotaxis behaviour (movement of an organism to or from the target location). Many genes regulate the chemotaxis behaviour such as *cheA*, *cheW*, *cheV*, *cheY*, *cheR*, and *cheB* which are known as chemotaxis genes. Mutations in these genes not only reduced the commensal colonization of *C. jejuni* in the chick model, but it also decreased the adherence to and invasion of INT407 epithelial cells (Hendrixson and DiRita, 2004; Yao *et al*., 1994).

Thus, due to the prime importance of motility in the colonization and virulence of *C. jejuni*, it was selected as a potential trait on which *cj0031* could have an effect. Previously, Kim *et al*. (2004) demonstrated that *cj1426*, an adenine methyltransferase, affected the motility phenotype. This provided a further stimulus to test the influence of *cj0031*, an adenine methyltransferase, on motility of *C. jejuni*.

### 5.1.4: Biofilm formation phenotype:
Slime like polymers are secreted by many microorganisms in response to the harsh environmental conditions, these polymers are collectively called glycocalyx or biofilm (Costerton et al., 1995). Biofilm performs important functions for the bacteria by protecting the cells from the unfavourable environment within or outside the host. The extracellular polymers are resistant to the bacteriophages, surfactants, host immune responses and certain antibiotic therapies (May et al., 1991; Chaurd et al., 1991; Nichols, 1991). In addition to this, biofilm forming bacteria are able to attach themselves to the host tissues with greater strength than non-biofilm forming bacteria, indicating their importance for virulence and pathogenicity of pathogenic bacteria (Moe et al., 2010). Joshua et al. (2006) proposed that the high incidence of C. jejuni infection in poultry and humans might partly have resulted from their high level of persistence in the environment in the form of matrix-enclosed bacterial populations (biofilm mass).

The monospecies biofilm has been reported for many pathogenic bacteria such as C. jejuni (Joshua et al., 2006), Pseudomonas aeruginosa (Whiteley et al., 2001), Staphylococcus epidermidis (Mack et al., 2000), Burkholderia cepacia (Huber et al., 2001), Salmonella enteritidis (Solano et al., 2002), showing the importance of biofilms for the particular life-cycle of these pathogens. Reuter et al. (2010) reported that C. jejuni formed biofilm mass under stressful aerobic condition as an adaptation to the zoonotic life-style. They concluded that biofilm formation masses of C. jejuni under aerobic conditions acted as reservoir for viable planktonic cells (Reuter et al., 2010)

As far as the understanding of biofilm formation at a molecular level, this area is still in its infancy. The flagellar regulatory apparatus have been documented to be an important determinant in controlling the biofilm formation phenotype. Kalmokoff et al. (2006) demonstrated the increased level of FlaA, FlaB, FliD, FlgG and FlgG2 during a proteomic study of bio-film forming cells, which are protein determinants regulating the
flagellar mediated motility. The *flaAB* mutant strain showed a decreased level of biofilm mass as compared to wild type strains. Similarly, the genes defective in flagellar modification (*cj1337*) and flagellar assembly (*fliS*) were defective in adhering to the glass surfaces, showing the role of flagellar gene regulation in development of biofilm formation in *C. jejuni*.

Srikhanta *et al.* (2010) showed that an ON version of the *H. influenzae* type III mod methyltransferase showed a higher biofilm mass than that of mutant and OFF versions of *mod*. This prompted the hypothesis that *cj0031*, an adenine specific type IIG methyltransferase, might affect the biofilm formation of *C. jejuni* and can contribute indirectly to the survival of bacterial cells under stressful conditions by methylating the promoters of genes, and thereby altering expression, implicated in the regulation of biofilm formation.
5.2: Results:

5.2.1: Natural transformation of *C. jejuni* chicken strain with genomic DNA of *C. jejuni* Lab strain:

Initially, the lab strain NCTC11168 was employed for screening WT-*cj0031*, WT-*cj0031*-OFF, and for construction of Δ*cj0031*:kan and complemented *cj0031* mutant. The screened strains were used for carrying out the restriction assays in order to investigate methyltransferase activity. The lab strain NCTC11168 was non-motile in terms of lacking swarming ability but it could swim, indicating that there was some defect in its flagella. So this strain was not useful for carrying out adhesion and invasion assays and other phenotypic assays which required the fully motile strain. Thus, the hypermutile chicken strain of *C. jejuni* (named CH11168) for adhesion and invasion assays was used (Cogan *et al*., 2007). CH11168 strain was therefore naturally transformed with genomic DNA extracted from lab strains: WT-*cj0031*, WT-*cj0031*-OFF, Δ*cj0031*:kan and complemented mutant. The transformants were grown on their respective antibiotics for selection. The confirmatory PCRs were performed in order to verify the nature of the transformants. Thus CH11168-*cj0031*-ON, CH11168-*cj0031*-OFF, CH11168Δ*cj0031*:kan, CH11168CΔ*cj0031* were obtained, which were used for carrying out the phenotypic assays.

5.2.2: Investigation of the impact of the Cj0031 methyltransferase on the adherence and invasion phenotypes of *C. jejuni*:

5.2.2.1: *cj0031* mutant causes hyporadherence to Caco-2 cells:
Four variants or strains of *C. jejuni* were used to examine the efficiencies of adherence, these were CH11168-cj0031-ON, CH11168-cj0031-OFF, CH11168Δcj0031::kan and a version of this mutant complemented with full-length *cj0031*.

Adherence and invasion assays were conducted as described previously by MacCallum *et al.* (2005) and as detailed in Materials and Methods chapter 2. Briefly, strains were grown up overnight in MHB and then the *C. jejuni* cells were resuspended in MHB. An inoculum was prepared from this suspension and then a total of $10^7$ cells from each strain were added to 12 wells of a tissue culture plate. The each well of 24-well plate (1.77cm$^2$) had previously been seeded with 4.8 x $10^5$ Caco2 cells. The experiment was performed in triplicate for each strain. The plates were incubated at 37°C in 5% CO2 incubator for 4 hours. After a washing step, serial dilutions were prepared and then three of these dilutions were plated out on MHA plates and incubated in a VAIN cabinet. After 3 days, the number of adherent bacteria was calculated. As the number of counted bacteria was quite low, so CFUs/ml were considered and the number of CFUs per dilution was calculated back to the volume of inoculum added. The cells of CH11168-cj0031-OFF and CH11168Δcj0031::kan strains were found to be twofold more adherent to Caco-2 cells than those of CH11168-cj0031-ON (Figure 1)
Figure 1: Adherence of *C. jejuni* strains to Caco-2 cells: Panel A and B, representing two separate experiments, shows the level of adherence of *C. jejuni* strains (CH11168-cj0031-ON, CH11168-cj0031-OFF, CH11168-Δcj0031::kan, and CH11168Δcj0031c) to Caco-2 cells. Adherence is expressed as the number of cfu which either adhered or invaded cultured Caco-2 cells following a 4-hours incubation. The triplicate readings were utilized in order to calculate SEM (Standard error of mean). Unpaired Student’s t-test was used to calculate the p values, p values for WT-ON versus WT-OFF and Δcj0031::kan were P<0.001; and p<0.01 for WT-ON versus CΔcj0031. Error bars were calculated from three measurements.
The complementation with the native \textit{cj0031} gene restored the adherent phenotype of WT-\textit{cj0031}-ON, confirming that hyperadherence of \textit{C. jejuni} cells to Caco-2 cells only resulted from the presence of \textit{cj0031} in an ON state.

On a separate occasion, the experiment was repeated in triplicate for each strain to test the reproducibility of the results obtained from the above experiment. In the second experiment, the CH11168-\textit{cj0031}-ON strain again showed hyperadherence to Caco-2 cells in contrast to the CH11168-\textit{cj0031}-OFF and CH11168\textit{Δcj0031::kan} strains (p<0.001). The increase in the level of adherence of WT-ON to Caco-2 cells was again found to be increased twofold. The complement of mutant of \textit{cj0031} restored the hyperadherence phenotype to the mutant strain, indicating that the Cj0031 methyltransferase had caused the hyperadherence to Caco-2 cells (Figure 2).

From each experiment, means and standard deviations (SD) were calculated from triplicate readings for each strain. The mean of CH11168-\textit{cj0031}-ON was compared with those of CH11168-\textit{cj0031}-OFF and CH11168\textit{Δcj0031::kan} strains. The p<0.001 was found, indicating that the difference in adherence ability of \textit{C. jejuni} cells was significant. The statistical analysis was performed by using the Graphpad statistical software. The raw data from these adherence experiments can be viewed in Appendix.

\textbf{5.2.2.2: Invasion of Caco-2 cells by \textit{C. jejuni} decreased drastically by \textit{cj0031} mutation:}

In order to investigate the influence of \textit{cj0031} on the invasion efficiency of \textit{C. jejuni} to the Caco-2 cells, an invasion assay was performed. The same batch of tissue culture cells were utilized for invasion and adherence studies at each time point of study. The experiment was repeated twice in triplicate on separate occasions. The samples for all
four strains: CH11168-cj0031-ON, CH11168-cj0031-OFF, CH11168Δcj0031::kan, complemented mutant of cj0031 were prepared by setting their OD at 0.5. The samples (10^7/well) from each strain were added in triplicate to the wells of a 12-well plate seeded with Caco-2 cells. The plates were incubated at 37°C in a 5% CO2 incubator for 4 hours. After four hours incubation, gentamycin (250ug/ml) was added into each well in order to kill the C. jejuni cells bound to the extracellular surfaces of the Caco-2 cells. The C. jejuni cells were harvested and quantified after an additional 4 hours incubation. Invasion efficiency was calculated by counting the number of colonies on the 10^-2 dilution and using this number to estimate CFUs/mL.

The results from both experiments revealed that the invasion efficiencies of CH11168-cj0031-OFF and CH11168Δcj0031::kan were significantly reduced. In the first experiment, the cj0031 mutation caused the reduction in invasion of Caco-2 cells by two and half fold compared to the wild type cj0031-ON. Similarly, WT-OFF invasion of Caco-2 cells was decreased by two fold in contrast to the WT-ON. Complementation of cj0031 mutant with native cj0031 restored the invasion phenotype of CH11168-cj0031-ON partially but significantly (p<0.01; obtained by comparing CH11168Δcj0031c with CH11168Δcj0031::kan and CH11168-cj0031-OFF, separately, and by using their means from triplicate readings in Graphpad statistical software), indicating that invasion of Caco-2 cells by CH11168-cj0031-ON was enhanced by the cj0031-ON phenotype (Figure 3A). In a second independent invasion assay, the CH11168-cj0031-OFF and CH11168Δcj0031::kan again showed a reduction in invasion of Caco-2 cells by 2.5-fold and 3-fold respectively compared to CH11168-cj0031-ON, which was similar to the results obtained in the first invasion assay. This indicated that the results of the invasion assay were reproducible. Again, complementation of the cj0031 mutant
partially but significantly (p < 0.01) restored the invasion phenotype of CH11168-\textit{cj0031}-ON (Figure 3B).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2a}
\caption{	extbf{Invasion of Caco-2 cells:} Panels A and B show the results of two independent gentamycin protection assays. The amount of \textit{C. jejuni} was calculated as the number of cfu/ml after 4 hours of gentamycin treatment. Triplicate readings were utilised to calculate SEM (standard error of the mean), and representative of two independent experiments. Each experiment was performed with \textit{CH11168-}cj0031-ON, \textit{CH11168-cj0031-OFF}, \textit{CH11168-Δcj0031::kan} and \textit{CH11168-Δcj0031c}. Unpaired Student’s t-test was used to calculate the p values, p values for WT-ON versus WT-OFF and}
\end{figure}
Δcj0031::kan were found P<0.001; and p<0.01 for WT-ON versus CΔcj0031. Error bars were calculated from three measurements.

5.2.3: Investigation of the impact of the Cj0031 methyltransferase on biofilm formation phenotype by C. jejuni:

The functional role of the Cj0031 methyltransferase in biofilm formation by C. jejuni was determined by employing the crystal violet (CV) staining method as described by (Joshua et al., 2006). For the quantification of biofilm formation by C. jejuni, The cultures of four strains, CH11168-cj0031-ON, CH11168-cj0031-OFF, CH11168Δcj0031::kan, complemented mutant of cj0031, were grown overnight in MHB. The culture OD was measured and adjusted to 0.025. Samples were then dispensed into 24-well, flat-bottomed polystyrene tissue culture dishes. A 1ml sample was added into each well in quadruplicate for each strain. The dishes were incubated statically at 37°C in a VAIN cabinet (10% CO2). After 48-72 hours incubation times, dishes were dried at 55°C for half an hour and stained with crystal violet. CV stained dishes were washed with water, followed by addition of a 20% acetone-80% ethanol solution to the wells in order to solubilise the crystal violet staining the biofilm mass attached to the bottom of wells. The crystal violet binding was determined by measuring the OD570 for samples in 96-well plates for 48 hours and 72 hours. This experiment was repeated twice on separate occasions. Quantification of CV binding for WT-cj0031-OFF and CH11168-Δ0031::kan showed a large decrease in their capability for biofilm formation compared to the WT-cj0031-ON (p<0.0001) after 48 and 72 hours (Figure A and B). The complementation of the mutant with native cj0031 restored the phenotype of WT-ON partially but significantly (p<0.001), indicating that absence of cj0031 caused the defect in the biofilm formation phenotype of C. jejuni. WT-cj0031-OFF and
CH11168-Δ0031::kan demonstrated 7-fold and 4-fold decreases in their capability of biofilm formation compared to WT-cj0031-ON after 48 hours and 72 hours, respectively. The raw data has been put in Appendix.

Figure 3: Defect in cj0031 caused reduction in the capability of C. jejuni strains in making biofilm formation. Assessment of biofilm formation in all four strains of C. jejuni ((CH11168-cj0031-ON, CH11168-cj0031-OFF, CH11168Δcj0031::kan, and CH11168CΔcj0031) was made by using the crystal violet staining of standing cultures of respective strain in wells of 24-well polystyrene plate and quantification.
of dissolved CV was made at 570nm. Unpaired Student’s t-test was used to calculate the p values, p values for WT-ON versus WT-OFF and Δcj0031::kan were found P<0.0001; and p<0.01 for WT-ON versus CΔcj0031. A: Triplicate readings from two independent experiments following 48 hours incubation at 37°C in VAIN, standard error of mean and error bars were calculated from 6 measurements. B: Triplicate readings from two independent experiments following 72 hours incubation at 37°C in VAIN, standard error of mean and error bars were calculated from 6 measurements.

5.2.4: Investigation of the impact of mutation of the cj0031 on motility of C. jejuni:

In order to investigate the functional consequences of loss of cj0031 function for motility of C. jejuni, a Mueller-Hinton-soft-agar assay was performed. The three strains, CH11168-cj0031-ON, CH11168-cj0031-OFF, CH11168Δcj0031::kan, were grown on Mueller-Hinton-soft-agar (MHB media supplemented with 0.4% MHA) and incubated in a VAIN cabinet at 37°C for 24 hours, 48 hours and 72 hours. The diameters of halos representing the distance moved by respective strains in soft agar were measured. No obvious difference in motilities of the cj0031 mutant, WT-OFF or WT-ON was found, indicating that the Cj0031 methyltransferase does not affect the motility of C. jejuni (Figure 4). The complement of cj0031 mutant was not used for this study because when the assay was conducted the construction of the complement of cj0031 mutant was in progress. Later on, no attempt was made to conduct the motility assay with the complement because previous experiments with CH11168-cj0031-ON, CH11168-cj0031-OFF, CH11168Δcj0031::kan revealed no effect of cj0031 on the motility phenotype.
Figure 4: Effect of mutation in *cj0031* on motility of *C. jejuni*:

Motility (diameters of halos) at 37°C of CH11168-*cj0031*-ON, CH11168-*cj0031*-OFF and CH11168-*Δcj0031::kan* cells. Data are the means calculated from duplicate readings from three independent experiments. Bars represent standard deviation.
5.3: Discussion:

*C. jejuni* strain NCTC11168 does not harbour a Dam homolog but does possess a unique phase variable methyltransferase encoded by *cj0031*. Phase variation of this methyltransferase was shown to be involved in the regulation of adhesion, invasion and biofilm formation of *C. jejuni*. However, there was no effect on motility. Due to the affect of *cj0031* on these phenotypes and by extrapolation on the expression of other genes, it is tempting to place the Cj0031 methyltransferase in the category of a phasevarion similar to those found in *N. meningitidis*, *N. gonorrhoeae* and *H. pylori*. This discussion will focus on the evidence to support this conclusion and on a putative mechanism of gene regulation.

5.3.1: Evidence for a phasevarion from motility, adhesion and invasion experiments

The present study has shown that the Cj0031 methyltransferase does not affect motility (Figure 4). This key finding suggests that CH11168-*cj0031*-ON, CH11168-*cj0031*-OFF, CH11168-Δ*cj0031* and the complement of the *cj0031* mutant were equally motile and that the observed alterations in the adhesion and invasion phenotypes of the *cj0031* mutant could not be attributed to motility related genes. However, this finding stands in contrast with the findings of Kim *et al.* (2008) who reported that the putative methyltransferase *cj1461* affected the motility phenotype of *C. jejuni* as well as the adhesion and invasion phenotypes. Moreover, studies have revealed that defects in the genes encoding the flagella structural proteins caused reductions in motility which ultimately attenuated the adhesion to and invasion of eukaryotic cells by *C. jejuni* (
For example, Szymanski et al. (1995) reported that flagella mediated motility is a key factor in binding of *C. jejuni* to Caco-2 epithelial cells. A contrasting finding was reported by Joshua and Thompson (2008) for Csr, a global post-transcriptional regulator, revealing that it was involved in enhancing the invasion by *C. jejuni* but that it did not affect the motility of *C. jejuni*. They proposed that CsrA affected invasion specific genes. This suggest that there are two sets of genes with one set of genes controlling motility via synthesis of the flagellar apparatus and a different set of genes affecting the adherence and invasion of *C. jejuni* into host cells. The effect of *cj0031* on motility suggests that this methyltransferase might be controlling adherence and invasion by altering the expression of genes specific to these processes in *C. jejuni*.

Apart from flagella, the role of other adhesins has been implicated in facilitating the adherence and internalization of *C. jejuni* such as ciaB (Konkel et al., 1999), cadF (Konkel et al., 1997, Ziprin et al., 1999), and peb1A (Pei et al., 1998). The regulation of these genes has not been characterised and it is possible that the Cj0031 methyltransferase might be involved in their regulation via epigenetic control just like Dam in *E. coli* and some other pathogenic bacteria. Thus, it can be postulated that the Cj0031 methyltransferase might act as a phasevarion affecting the expression of one or more of the adhesins of *C. jejuni*. This warrants further investigation by conducting a microarray analysis to detect the genes affected by the Cj0031 methyltransferase.

### 5.3.2 Evidence for a phasevarion from biofilm experiments

Due to the heavy public health cost incurred by *C. jejuni* infections, it is paramount to unravel the survival strategies employed by *C. jejuni* not only in the host but also in the food chain. There are many pressures acting on the organisms in an macro-environment
and microenvironment ranging from the fluctuations in temperature, pH, resistance to phages and antimicrobial agents and to the immune responses exacted by hosts to the invading microbes (Hall-Stoodley et al., 2004). *C. jejuni* is a zoonotic pathogen whose success in propagating the disease lies in its ability for survival during transmission between one host to another host and under the aerobic conditions, albeit it is an obligate microaerophilic bacteria. The survival of *C. jejuni* under stress conditions and its successful transmission in the environment as well as in the food chain has long been thought to be due to its capability of biofilm formation (Reuter et al., 2010).

The affect of methyltransferases on the biofilm formation in *C. jejuni* is not well understood. However, the previous reports showing the affect of mod component of a type III RM system on the biofilm forming capability of *N. meningitidis* and *N. gonorrhoeae* were documented (Srikhanta et al., 2010). In this study, the affect of the phase variable Cj0031 methyltransferase on biofilm formation has been reported. The mutation in *cj0031* decreased its biofilm mass. CH11168-*cj0031*-OFF and CH11168-*Δcj0031* exhibited a 3-fold reduction in their biofilm mass compared to wild type CH11168-*cj0031*-ON. Complementation partially but significantly restored the phenotype to wild type CH11168-*cj0031*-ON levels, showing that biofilm formation was under the influence of *cj0031* and was not the consequence of a mutation in any other gene.

The current study used the crystal violet technique (CV) by using polystyrene flat-bottomed 96-well microtiter plate as the attachment of *C. jejuni* with polystyrene (Moe et al., 2010) and glass surfaces (Joshua et al., 2006) were reported. This method is a powerful tool not only to study the early stages of biofilm formation in *C. jejuni*, *P. aerugena*, and *E. coli*, but it is also utilized to determine the later developmental stages of biofilms. Safranin or Red Congo dyes may also be used instead of CV, which
cast no impact on the results (Reuter et al., 2010). However, the present study only used CV due to its easy handling and availability. The modifications of this technique were successfully applied by many other researchers to screen the mutants defective for biofilm formation (Reeser et al., 2007; Teh et al., 2010; Reuter et al., 2010). Joshua et al. (2006) employed CV technique in order to determine the effect of mutations in Flis (a putative flagellar protein), Cj0688 (a phosphate acetyltransferase) and kpsM (capsular polysaccharide) genes on the biofilm formation of C. jejuni strains 33106, 32799 and 33084. Apart from CV method, Moe et al. (2010) also used scanning microscopy techniques - transmission electron microscopy and scanning electron microscopy - in order to visualize the effect of flagellar genes on the biofilm formation ability of C. jejuni. This method is useful particularly if the aim is to visualize the development stages of biofilm following periodic time intervals. This method may be employed to obtain the detailed picture of biofilm stages after 24, 48 or 72 hours.

Previous studies have revealed that motile and flagellate strains of C. jejuni showed increased biofilm formation compared to non-motile and aflagellate strains of C. jejuni, suggesting that flagellar expression is required for formation of C. jejuni biofilms. However, the phenotype of biofilm formation did not disappear completely from non-motile and aflagellated strains (Reuter et al., 2010). All strains of C. jejuni employed for the present study were hypermotile. This means that they were all flagellate and alterations in the capability of biofilm were not the consequence of lack of flagella.

An alternate theory is that cj0031 might have impacted on the biofilm formation of C. jejuni by influencing the genes involved in adherence. In this case biofilm formation is reduced due to weak adherence of cj0031 mutant cells of C. jejuni cells to solid surfaces (wells of polystyrenes plates). This suggestion reinforces the data from the adherence assay displaying decreased adherence to Caco-2 cells, suggesting a role of
methyltransferase in regulation of adherence specific genes epigenetically. Similarly, the genes involved in the formation of biofilm matrix are presumably affected by regulation of \textit{cj0031} mediated methylation patterns on the promoters of these genes. Thus it is intriguing to postulate that a phase variable Cj0031 methyltransferase acts as a phasevarion in \textit{C. jejuni}. This warrants further investigation by conducting further studies in order to gain more insight into its mechanism and genes affected by performing a microarray analysis.

5.3.3: Evidence for a phasevarion from in vivo experiments

Bayliss \textit{et al.} (2012) showed that during in vivo passage of CH11168-\textit{cj0031}-ON, the level of \textit{cj0031}-ON variants was increased to 85\% in the output population compared to 19 \% in the inoculum. Jerome \textit{et al.} (2011) and Bayliss \textit{et al.} (2012) conducted in vivo studies of colonization of \textit{C. jejuni} in the intestinal tract of mice and chickens, respectively, in which they determined the levels of various phase variable genes during in vivo passage. Jerome \textit{et al.} (2011) also concluded that \textit{cj0031}-ON variants were increased during in vivo passage through chickens and mice.

These findings about the increased levels of \textit{cj0031}-ON variants in vivo, supports the hypothesis that switching of \textit{cj0031} from OFF to ON expression might be an important factor for colonization by \textit{C. jejuni} in experimental animal models. The demonstration of 2.5 and 3 fold decreases in degrees of adhesion to and invasion of Caco-2 cells, respectively, by \textit{cj0031}-OFF and \textit{Δcj0031::kan} as compared to \textit{cj0031}-ON variant supports a hypothesis that a \textit{cj0031}-ON expression status was required for the adhesion to and invasion of epithelial cells in vivo. Moreover, it also points to the potential
importance of \textit{cj0031} as an important genetic determinant for pathogenesis and colonization by \textit{C. jejuni}

A previous study has also implied a role for a methyltransferase in \textit{C. jejuni} infections. There was evidence showing that \textit{cj1461} (a putative methyltransferase) is involved in the adherence and invasion phenotypes of \textit{C. jejuni} (Kim \textit{et al}., 2008). But the findings derived from the present study differed from those of Kim \textit{et al}.
(2008) in respect of the adherence phenotype, the former revealed the reduction of adherence with a \textit{cj0031} deficient strain while the latter showed an increase in adherence to eukaryotic cells by a \textit{cj1461} deficient strain NCTC11168. However, invasion was decreased for both \textit{cj0031} (present study) and \textit{cj1461} deficient mutants of NCTC11168 (Kim \textit{et al}., 2008), suggesting that both methyltransferases might act in concert in controlling the invasion by \textit{C. jejuni}.

\textbf{5.3.4 Putative mechanism for a role of \textit{cj0031} in regulation of other genes}

The results presented in this chapter clearly show that the efficiency of adherence to and invasion of Caco-2 cells by \textit{cj0031} deficient strains (CH11168-\textit{cj0031}-OFF and CH11168-\textit{Δcj0031}) was reduced, however the underlying mechanism is not exactly known. The previous chapter demonstrated that \textit{cj0031} is a methyltransferase and there is no evidence to support a direct role of this protein in adhesion or invasion. This leads to a hypothesis that Cj0031 methyltransferase-mediated methylation may be the key player in the regulation of genes required for adhesion and invasion, thereby influencing these processes in \textit{C. jejuni}. The methylation of specific sequences in the promoter region of genes is known to regulate gene expression, which is often termed as epigenetic control of gene regulation (Casadesus and Low, 2006; Wion and Casadesus,
Most of the reviews have focussed on the epigenetic control of gene regulation by Dam methyltransferase. Dam is reported to be essential in enhancing the virulence of pathogens within the host by regulating virulence-specific genes. For instance, a Dam mutant strain of *Salmonella* showed attenuation in virulence due to modulation of genes involved in motility and adhesion and invasion of bacteria to host epithelial cells (Heithoff *et al.*, 1999; Heithoff *et al.*, 2001; Garcia-Del Portillo *et al.*, 1999). Dam mediated attenuation in virulence was also reported for some other pathogens such as *Yersinia pseudo tuberculosis*, *Vibrio cholera* and *H. influenzae* (Taylor *et al.*, 2005; Julio *et al.*, 2001; Watson *et al.*, 2004). However, the mechanism of attenuation remained obscure in these studies, though in *E. coli* phase variation of the *pap* operon and Ag43 mediated by Dam methylation are known examples of mechanism of action of Dam (Blyn *et al.*, 1990; Haagmans and van der Woude, 2000; Nicholson and Low, 2000). The main point to note in these Dam-methylation dependent phase variable systems is that methylation alters the binding of regulatory proteins implicated in the transcription of these genes. Nevertheless, Dam itself is neither phase variable nor regulated by environmental factors, so this can not be termed a phasevarion, but it serves as model system showing how DNA methylation regulates gene expression via epigenetic control. Many strains of *N. meningitidis* and *N. gonorrhoeae* were found to be Dam negative, but they contained phase variable type III RM systems (*mod* components) to regulate the global gene expression, which are termed as phasevarions (Srikhanta *et al.*, 2010, Srikhanta *et al.*, 2009, Srikhanta *et al.*, 2005). Genetic regulation by Dam through methylation of target recognition sequences is known to be common phenomenon (Casadesus and Low, 2006; Marinus and Casadesus, 2009). Dam mediated phase variation events are characteristic of controlling the virulence in *Salmonella typhimurium* (Haagman and van der Woude, 2000; Nicholson and Low, 2000). Dam
competes with a particular regulatory protein at its target sites in the promoters of genes, and constitutes a specific methylation patterns, thereby controlling the expression of these genes. Alternately, the regulatory protein can bind at non-methylated Dam target sites (GATC) in the promoters, thus blocking the binding of Dam and preventing the subsequent methylation (Julio et al., 2001; Mahan and Low, 2001; Heithoff et al., 2001). The well studied example of the mechanism of control of global expression of genes by Dam methylation were reported for pap operon in uropathogenic E. coli and antigen43 involved in biofilm formation in E. coli. It is presumed that phase variable type IIG Cj0031 methyltransferase may affect the global gene expression of genes involved in bacteria-host interactions and biofilm formation through methylation of Cj0031 specific target sequences located in promoter regions. These target sequence may naturally occur in the promoter or can arise from a few point mutations leading to generation of the specific target sequence in key positions in promoter regions, and thereby controlling the gene expression through methylation by modulating the affinity of regulatory proteins for specific promoters. The prevalence of phase variable type III RM system in other pathogenic bacteria (N. menigitidis, and H. pylori) and their role in coordinating the multiple switching of genes (Seib et al, 2002, Srikhanta et al., 2009) suggests that phasevarion mediated regulation of gene expression is a commonly used mechanism for switching of multiple genes in pathogens.
Chapter 6. Analysis of restriction activity of *C. jejuni* type IIG Cj0031 RM system

6.1: Introduction

The restriction components of the restriction-modification systems in prokaryotes play an important role in protection of bacteria against infection by phages or preventing the penetration of foreign DNA into bacterial cells, thereby protecting the genetic integrity of specific bacterial strains. The restriction endonucleases act on the non-methylated foreign/phage DNA containing their recognition sites, and cause DNA hydrolysis by restricting DNA within the recognition sequence or in their vicinity. Prokaryotic RM systems contain another mandatory component the methyltransferase which recognizes the same sequence/site as their counterpart restriction endonuclease in the host genome and mediates protection of the host genome from restriction endonuclease accelerated hydrolysis by methylating these sites (Kobayashi, 2001; Furuta *et al.*, 2011).

In 1950, Luria and Human (1952) and Bertani and Weigle (1953) discovered the phenomenon of restriction and modification of phages invading different strains of a bacterial species. They observed that the efficiency of plating (EOP) of phages was equal to or almost equal to one on the restriction-deficient bacterial strains. EOP represents the proportion of phages having the capability to productively infect the host and form plaques. However, EOP of phages was decreased by five orders of magnitude upon infection of restriction-proficient bacterial strains. Some of the phage progeny recovered from the rare productive infections of restriction-proficient hosts demonstrated EOP values equal to one on both restricting and non-restricting hosts (Enikeeva *et al.*, 2010). The probability of failure of an RM endonuclease to attack an
unmodified target DNA lies in the range of $10^{-2}$ to $10^{-6}$ (Korona et al., 1993; Wilson and Murray, 1991). As a result of an escape from an endonuclease, the phage genome becomes methylated by the cognate methyltransferase component of the RM system just like the host genome and is thereby protected from further attacks of the endonuclease.

However, after a single passage on restriction-deficient strains, the “modified” phages had lost the ability to further infect the restriction-proficient host showing that modified phages did not carry the heritable trait of modification. Arber reviewed in 1978 that, in fact, the restriction-proficient bacterial strains contain two enzymatic activities-restriction and modification. The modification enzyme - DNA methyltransferase - of the RM system of a restricting strain may methylate the phage DNA containing its recognition sites and confer immunity to phages against restriction mediated by the cognate restriction endonuclease. In further rounds of replication, phages replicate without invoking a destructive response from the relevant restriction endonuclease of the host bacterium. The survival of a bacterial cell after phage invasion hinges upon the competition between MTase and RE for the recognition sites on the genome of the phage. If the phage genome is methylated before RE acts on it, then the phage should survive and cause a successful infection. Thus the presence of an active restriction endonuclease of an RM system is an important signpost for maintaining the bacterial genomic integrity by decreasing the chances of penetration of DNA from the external milieu.

The prokaryotic RM systems provide a first-line defense mechanism against phages. Their efficiency against phages can be appraised by the fact that phages have developed various anti-restriction strategies to cope with them. For example, the acquisition of multispecific methyltransferase genes and a decrease in the number of
bacterial RM recognition sites in the genomes of many phages (Bickle and Kruger, 1993; Tock and Dryden, 2005). Taken together, this highlights that evolutionary developments in the phages are related to their survival in the bacterial cells; and it also favours the hypothesis of RM systems serving as a defense tool against phages. Indeed Joltsch et al (2003) have argued that the diversity of the prokaryotic RM systems partly resulted from the diversity of phages invading the bacterial species.

Restriction enzymes are classified into four groups based on their requirements for ATP, sequence specificity and DNA hydrolysis mechanism and sub-unit architecture (Roberts et al., 2003). The detailed mechanism of their action was described in Chapter 1 section 1.6.1. The basic function of evolution of RM systems and new specificities of RM systems is to confer defense to the bacterial hosts against foreign DNA or phage infections. If the bacterium harbouring the RM system is attacked by a phage sensitive to that particular RM system, the endonuclease component of the RM system will recognize and cut the phage genome. This causes subsequent abortion of infection. Endonucleases act on double-stranded phage DNA. Once, the replication of a phage with a single stranded DNA intermediate kicks off and a second strand is synthesized, it becomes susceptible to restriction by the host RM system (Murray, 2002).

The search for active RM systems led to the discovery of different types of restriction endonucleases and methyltransferase from diverse bacteria, having novel specificities and the potential to be used in molecular cloning (REBASE, http://rebase.neb.com). More than 200 different RM systems have been discovered in various bacterial species (Roberts and Macelis, 2001). Moreover, a single bacterial species may serve as reservoir multiple different RM systems. For example, *N. gonorrhoeae* contains 16 different RM systems (Stein et al., 1995). *H. pylori* and *Methanococcus jannaschii*
harbour 30 and 13 genes encoding DNA MTases in their genomes, which can potentially be part of an RM system (Jelstch, 2002). This diversity of RM systems is consistent with the large diversity of phages invading bacteria. The multiple RM systems residing in the genome of a single species of bacteria confer the ability to their hosts to counteract the various phages containing corresponding specifies for these RM systems, subsequently restricting the phage genomes and preventing their replication within the host.

Like other pathogenic bacteria, *C. jejuni* also contains multiple RM systems, although their variety and number is limited compared to other bacteria (Parkhill et al., 2000). For example, *C. jejuni* strain 81176 contains two type I and four type II restriction modifications systems. Similarly, *C. jejuni* strain 11168 possesses one type I RM and four type II RM systems. The type I RM of strain 11168 is essentially identical with the CjeFIV system in *C. jejuni* strain 81176. M.CjeNI and M.CjeNII have been characterized by Vitor *et al.* (unpublished data) and Roberts *et al.* (2010) as type IIG RM systems. Cjo690 is another type IIG RM system whose restriction activity was shown by Murray *et al.* (2012). Cj0031 is also a type IIG RM system as shown by the bioinformatics analysis carried out in the previous chapter. Type IIG RM systems, as a rule, possess both endonuclease and MTase activity fused in the same polypeptide. The active status of Cj0031 MTase had been identified in *cj0031*-ON variants of *C. jejuni* strain 11168. It was postulated that Cj0031 type IIG RM system possessed the restriction activity fused with MTase polypeptide.

The present study intends to explore whether *cj0031* encodes an active restriction endonuclease. These studies will not only pertain to the active status of Cj0031 as a type IIG RM system but it will also open new avenues of investigation into characterization of Cj0031 restriction endonuclease activity. This will also be a good
addition to a growing collection of type IIG RM systems prevalent in prokaryotes.

6.2: Results

6.2.1: The Plaque Forming Assay

A total of 14 phages propagating on their respective strains were used for this study (Table 1). PT14 and HPC5, 11168 and 11168H (hypermotile version) are *C. jejuni* strains with an ability to infect broiler chickens. The genome sequence of PT14 was determined by the Ian Connerton group (unpublished). PT14, 11168 and HPC5 strains were used routinely for typing of *C. jejuni* phages (Frost et al., 1999; Scott et al., 2007). The CP8 and CP30 phages were originally isolated from the caecum of chickens (UK) by Loc Carrillo et al. (2005) and selected from the lab stock for the present study, as they were well-characterized and were known to generate different lytic spectra following infection of *C. jejuni* (Scott et al., 2007). The phages CP8PL and CP30PL are pseudolysogenic (PL) phages. PT14CP8PL and PT14CP30PL were clones of PT14 containing the CP8PL and CP30PL phages.
Table 1: List of Phages and their propagating strains used during the study

<table>
<thead>
<tr>
<th>Phages</th>
<th>Propagating strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP8</td>
<td>PT14</td>
</tr>
<tr>
<td>CP8</td>
<td>HPC5</td>
</tr>
<tr>
<td>CP8</td>
<td>11168</td>
</tr>
<tr>
<td>CP8</td>
<td>11168H</td>
</tr>
<tr>
<td>CP8PL</td>
<td>PT14CP8PL</td>
</tr>
<tr>
<td>CP8PL</td>
<td>HPC5CP8PL</td>
</tr>
<tr>
<td>CP30</td>
<td>PT14</td>
</tr>
<tr>
<td>CP30</td>
<td>HPC5</td>
</tr>
<tr>
<td>CP30</td>
<td>11168</td>
</tr>
<tr>
<td>CP30</td>
<td>11168H</td>
</tr>
<tr>
<td>CP30PL</td>
<td>PT14CP30PL</td>
</tr>
<tr>
<td>CP30PL</td>
<td>HPC5CP30PL</td>
</tr>
<tr>
<td>CP220</td>
<td>HPC5</td>
</tr>
<tr>
<td>Cp220</td>
<td>PT14</td>
</tr>
</tbody>
</table>

6.2.2: CP8 and CP30 phages demonstrated sensitivity to *cj0031*

The plaque forming ability of the phages (as listed in Table 1) on the *cj0031* wild-type and mutant strains plus the respective propagating strains of *C. jejuni* were determined
by following the Plaque assay protocol described in the Materials and Methods chapter. Briefly, C. jejuni strains-11168-cj0031-ON, 11168-cj0031-OFF, 11168Δcj0031::kan and 11168Δcj0031c- and the strains listed in table (1) were grown on blood agar (BA) plates. The bacterial cells were harvested into 10ml of 10mM Mg2SO4. A 500 ul sample of this cell suspension was added into a 4ml of NZCYM overlay agar, and then poured onto NZCYM agar plates. The concentrations of phages listed in Table 2 were diluted to 10⁶ PFU/ml. A 10ul sample of each dilution from each phage was spotted onto the surface of a plate covered with a lawn of bacteria. After 24 hours incubation in the VAIN, the plaque forming ability of the phages on the C. jejuni hosts were determined. A comparison of the plaque forming abilities of sensitive phages on the reference propagating host (PT14 and PT14PL) was made with those on strains of interest in order to find out the proportions of phages infecting C. jejuni strains: 11168-cj0031-ON, 11168-cj0031-OFF, and 11168-Δcj0031::kan.

Differences in the plaquing ability of CP30/PT14PL; CP30/PT14, CP8/PT14 and CP8/PT14PL were noticed on 11168-cj0031-ON (where plaquing ability was lower) versus 11168-cj0031-OFF and 11168-Δcj0031::kan (where plaquing ability was higher), as demonstrated by their highlighted lytic spectra [Table 2].
Table 2: Lytic spectrum for 14 phages propagated on reference strains and *C. jejuni* strains of interest

<table>
<thead>
<tr>
<th>Phage</th>
<th>Host</th>
<th>PT14</th>
<th>HPC5</th>
<th>11168</th>
<th>11168H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CP8</td>
<td>PT14</td>
<td>CL</td>
<td>CL</td>
<td>SCL</td>
</tr>
<tr>
<td>2</td>
<td>CP8</td>
<td>HPC5</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>3</td>
<td>CP8</td>
<td>11168</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>4</td>
<td>CP8</td>
<td>11168H</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>5</td>
<td>CP30</td>
<td>PT14</td>
<td>CL</td>
<td>&lt;SCL</td>
<td>&lt;SCL</td>
</tr>
<tr>
<td>6</td>
<td>CP30</td>
<td>HPC5</td>
<td>CL</td>
<td>CL</td>
<td>&lt;SCL</td>
</tr>
<tr>
<td>7</td>
<td>CP30</td>
<td>11168</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>8</td>
<td>CP30</td>
<td>11168H</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>9</td>
<td>CP8</td>
<td>PT14PL</td>
<td>CL</td>
<td>CL</td>
<td>&lt;SCL</td>
</tr>
<tr>
<td>10</td>
<td>CP8</td>
<td>HPC5PL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>11</td>
<td>CP30</td>
<td>PT14PL</td>
<td>CL</td>
<td>++</td>
<td>±</td>
</tr>
<tr>
<td>12</td>
<td>CP30</td>
<td>HPC5PL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>13</td>
<td>CP220</td>
<td>PT14</td>
<td>OL</td>
<td>OL</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>CP220</td>
<td>HPC5</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phage</th>
<th>Host</th>
<th>11168-cj0031-ON</th>
<th>11168-cj0031-OFF (1)</th>
<th>11168-Δcj0031::kan</th>
<th>11168-cj0031-OFF (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CP8</td>
<td>++</td>
<td>+++</td>
<td>&lt;SCL</td>
<td>&lt;SCL</td>
</tr>
<tr>
<td>2</td>
<td>CP8</td>
<td>HPC5</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>3</td>
<td>CP8</td>
<td>11168</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>4</td>
<td>CP8</td>
<td>11168H</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>5</td>
<td>CP30</td>
<td>PT14</td>
<td>±</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>6</td>
<td>CP30</td>
<td>HPC5</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>7</td>
<td>CP30</td>
<td>11168</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>8</td>
<td>CP30</td>
<td>11168H</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>9</td>
<td>CP8</td>
<td>PT14PL</td>
<td>-</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>10</td>
<td>CP8</td>
<td>HPC5PL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>11</td>
<td>CP30</td>
<td>PT14PL</td>
<td>-</td>
<td>±</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td>CP30</td>
<td>HPC5PL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>13</td>
<td>CP220</td>
<td>PT14</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>CP220</td>
<td>HPC5</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
</tbody>
</table>
Key – CL: (confluent) Clear lysis; OL: (confluent) Opaque lysis; SCL: Semi-confluent clear lysis; <SCL: Less than semi-confluent clear lysis; +++: >100 discrete plaques; ++: 50-99 discrete plaques; +: 20-49 discrete plaques; ±: 1-19 discrete plaques; -: No lysis (Method of reporting plaquing abilities taken from Frost et al., 1999)

EOP values were determined from the above experiment for the sensitive phages by counting the number of plaque forming units and dividing by the PFU/ml by mutant (Table 3). It was discovered that CP30/PT14PL was restricted on cj0031-ON, showing an EOP of 0.29. Similarly, 11168-cj0031-ON strain restricted CP30/PT14 by decreasing its EOP to 0.36 compared to mutant. CP8/PT14PL could not be propagated on 11168-cj0031-ON but would replicate on the mutant. Likewise, the efficiency of plating for CP8/PT14 was decreased to 0.46 on 11168-cj0031-ON. The data obtained from this experiment suggested that CP30/PT14, CP30/PT14PL, CP8/PT14 and CP8/PT14PL demonstrated decreased efficiency of plaquing (EOP) on 11168-cj0031-ON compared to a cj0031 deficient C. jejuni strain, 11168-Δcj0031. Plaques on both OFF variants were not clear and so their EOPs values were not included here. However, a difference was observable between less plaques and higher plaques on ON and OFF variants, respectively [Table 2].
Table 3: Relative EOP values for the sensitive phages

<table>
<thead>
<tr>
<th>Phages</th>
<th>Relative EOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>11168-cj0031-ON</td>
<td>11168-Δcj0031::kan</td>
</tr>
<tr>
<td>CP8/PT14</td>
<td>0.44</td>
</tr>
<tr>
<td>CP8/PT14PL</td>
<td>-</td>
</tr>
<tr>
<td>CP30/PT14</td>
<td>0.36</td>
</tr>
<tr>
<td>CP30/PT14PL</td>
<td>0.29</td>
</tr>
</tbody>
</table>

- - No lysis

This suggests that unmodified phages liberated from PT14 (r’m’) and PT14PL (r’m’) strains were restricted by Cj0031 when grown on 11168-cj0031-ON (r+m+). But in the case of strains in an OFF state or mutant (r’m) for cj0031 then these sensitive phages were propagated and showed immunity of restriction due to the lack of Cj0031 restriction endonuclease.

6.2.3. Isolation of sensitive phage strains from 11168-Δcj0031::kan

The sensitive phage strains were recovered from 11168-Δcj0031 by following the phage recovery protocol (Materials and Methods chapter section 2.15.3). As the phages liberated from 11168-Δcj0031::kan (r’m’) remained unmodified, it was hypothesized that if 11168-cj0031-ON strain produced an active Cj0031 restriction endonuclease as part of the Cj0031 RM system, then phage DNA could be recognized as foreign DNA and digested by it. Hence, the next step was to spot these phage strains back on the 11168-cj0031-ON, 11168-cj0031-OFF, 11168-Δcj0031::kan and 11168-Δcj0031c. The
sensitive phage strains were recovered, titrated and spotted on the host strains by following the protocol of phage plaque forming assay (as described in Materials and Methods). The plaque forming abilities of the sensitive phages were reported by following the method described by Frost et al (1999) [Table 4].

Table 4: Lytic spectrum of four sensitive phages

<table>
<thead>
<tr>
<th>Phage</th>
<th>Host</th>
<th>11168-(cj0031)-ON</th>
<th>11168-(cj0031)-OFF</th>
<th>11168-(\Delta cj0031)::kan</th>
<th>11168-(\Delta cj0031)c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CP30</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>2</td>
<td>CP8</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>3</td>
<td>CP30</td>
<td>CL</td>
<td>SCL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>4</td>
<td>CP8</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
</tbody>
</table>

Key: CL: - Confluent clear lysis; SCL: - semi-confluent clear lysis

Clear lysis was observed in the case of 11168-\(cj0031\)-OFF and 11168-\(\Delta cj0031\)::kan, which was the expected outcome of experiment due to the absence of the putative Cj0031 restriction enzyme. However, the observation of clear lysis of 11168-\(cj0031\)-ON was an unexpected outcome as it was predicted that Cj0031 restriction enzyme would be active and would restrict the phage. This suggests that either the Cj0031 restriction enzyme was inactive or the phage titer was too high to be restricted by
Cj0031 resulting in the successful lysis of host cells.

To test the hypothesis that the titer of sensitive phages was too high, dilutions ($10^{-1}$ and $10^{-2}$) were prepared and spotted on the NCZYM plates. The plaques were counted and recorded for all the *C. jejuni* strains under investigation (Table 5). The dilutions presented a clearer picture of the restriction and protection of phages on the host strains of interests, indicating that high phage titre on previous experiment caused the clear lysis event on all *C. jejuni* strains.

**Table 5: Lytic spectrum of four sensitive phages by using dilution factors**

<table>
<thead>
<tr>
<th>Phage</th>
<th>Host</th>
<th>11168-cj0031-</th>
<th>11168-cj0031-</th>
<th>11168-Δcj0031::kan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ON</td>
<td>OFF</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$10^{-1}$</td>
<td>$10^{-2}$</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>1</td>
<td>CP30</td>
<td>SCL</td>
<td>++</td>
<td>CL</td>
</tr>
<tr>
<td>2</td>
<td>CP8</td>
<td>SCL</td>
<td>+++</td>
<td>OL</td>
</tr>
<tr>
<td>3</td>
<td>CP30</td>
<td>++</td>
<td>±</td>
<td>+++</td>
</tr>
<tr>
<td>4</td>
<td>CP8</td>
<td>SCL</td>
<td>++</td>
<td>CL</td>
</tr>
</tbody>
</table>
Key – CL: (confluent) Clear lysis; OL: (confluent) Opaque lysis; SCL: Semi-confluent clear lysis;  
<SCL: Less than semi-confluent clear lysis; +++ : >100 discrete plaques; ++: 50-99 discrete plaques; +: 
20-49 discrete plaques; ±: 1-19 discrete plaques; -: No lysis

Decreased plaquing efficiencies were observed for CP30/PT14, CP8/PT14, CP8/PT14PL on 11168-cj0031-ON and 11168-Δcj0031c as compared to those on 11168-cj0031-OFF and 11168-Δcj0031::kan strains of C. jejuni for which semi-confluent or clear lysis was observed. An exception was for CP30/PT14 which showed more than 100 plaques on OFF phase variants and the mutant of cj0031.

6.2.4. Sensitivity of Typing Phages to the restriction activity of Cj0031

A panel of 46 phage strains including 16 typing phages was used to screen for sensitivity to Cj0031 restriction activity. The supporting information for these phages was provided in (Appendix…). Phage stocks, already titrated to 10^6 pfu/ml, were provided by a PhD student of Ian Connerton, University of Nottingham. The plaque forming assay was conducted as described in Materials and Methods chapter. The bacterial strains were grown on blood agar plates and harvested into 10ml of 10mM MgSO4. A mixture including 500ul of cell suspension and 4ml of the overlay agar was poured onto the surface of NZCYM plates. A 10ul sample of each phage (10^6 PFU/ml) was spotted onto these plates. After 24-hours incubation in the VAIN, the plaques were counted for each phage strain and recorded (Table 6)

Table 6: Lytic spectra of 46 phages on cj0031-ON and mutant strains
<table>
<thead>
<tr>
<th>Code</th>
<th>11168-cj0031-ON</th>
<th>11168-Δcj0031</th>
<th>Code</th>
<th>11168-cj0031-ON</th>
<th>11168-Δcj0031</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;SCL</td>
<td>SCL</td>
<td>Φ1</td>
<td>OL</td>
<td>± (= 9 plaque)</td>
</tr>
<tr>
<td>2</td>
<td>CL</td>
<td>CL</td>
<td>Φ2</td>
<td>OL</td>
<td>+++ (= 55 plaques)</td>
</tr>
<tr>
<td>3</td>
<td>CL</td>
<td>CL</td>
<td>Φ3</td>
<td>OL</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>&lt;SCL</td>
<td>SCL</td>
<td>Φ4</td>
<td>OL</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>CL</td>
<td>CL</td>
<td>Φ5</td>
<td>± (= 1 plaque)</td>
<td>± (= 8 plaque)</td>
</tr>
<tr>
<td>6</td>
<td>OL</td>
<td>OL</td>
<td>Φ6</td>
<td>+++</td>
<td>CL</td>
</tr>
<tr>
<td>7</td>
<td>CL</td>
<td>CL</td>
<td>Φ7</td>
<td>OL</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>OL</td>
<td>OL</td>
<td>Φ8</td>
<td>OL</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>OL</td>
<td>OL</td>
<td>Φ9</td>
<td>OL</td>
<td>&lt;SCL</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>-</td>
<td>Φ10</td>
<td>OL</td>
<td>OL with ring</td>
</tr>
<tr>
<td>11</td>
<td>+</td>
<td>+</td>
<td>Φ11</td>
<td>OL</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>-</td>
<td>Φ12</td>
<td>OL</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>SCL</td>
<td>CL</td>
<td>Φ13</td>
<td>OL</td>
<td>OL</td>
</tr>
<tr>
<td>14</td>
<td>&lt;SCL</td>
<td>SCL</td>
<td>Φ14</td>
<td>OL</td>
<td>OL</td>
</tr>
<tr>
<td>15</td>
<td>± (= 1 plaque)</td>
<td>± (= 2 plaque)</td>
<td>Φ15</td>
<td>OL</td>
<td>OL</td>
</tr>
<tr>
<td>16</td>
<td>OL</td>
<td>OL</td>
<td>Φ16</td>
<td>OL</td>
<td>SCL</td>
</tr>
<tr>
<td>17</td>
<td>± (= 16 plaques)</td>
<td>+++</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>&lt;OL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>SCL</td>
<td>SCL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>CL</td>
<td>CL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>CL</td>
<td>CL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>&lt;OL</td>
<td>&lt;OL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>-</td>
<td>&lt;OL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-</td>
<td>&lt;OL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>-</td>
<td>&lt;OL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>CL</td>
<td>SCL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>-</td>
<td>&lt;OL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>&lt;OL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>&gt;OL with ring</td>
<td>&gt;OL with ring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>-</td>
<td>&lt;OL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Eleven phage strains including six typing phage strains showed a difference in their plaque forming abilities on 11168-\textit{cj0031}-ON and 11168-\textit{Δcj0031}::kan strains. The plaques were more numerous on 11168-\textit{Δcj0031}::kan than on 11168-\textit{cj0031}-ON, suggesting that they were likely sensitive to Cj0031 restriction endonuclease. However, four of the typing phages – Φ3, Φ4, Φ11, Φ12 - in green squares showed the opposite behaviour compared to those sensitive to \textit{cj0031} mediated restriction. They were found to have a higher level of propagation on Cj0031 proficient strains than \textit{cj0031} deficient strains, indicating that their growth was facilitated in the presence of \textit{cj0031}. It was not clear why they were propagating better on 11168-\textit{cj0031}-ON than the mutant, but it might be due to \textit{cj0031}-mediated alterations in the phage receptors. This finding highlighted that some of the typing are adapted for \textit{cj0031} expression while some were identified to be sensitive to \textit{cj0031}.

\textbf{6.2.5. Isolation of sensitive typing phages from \textit{cj0031} mutant strain}

The 11 \textit{cj0031} sensitive phages were recovered from 11168-\textit{Δcj0031}::kan mutant plates, titrated and spotted on 11168-\textit{cj0031}-ON, 11168-\textit{cj0031}-OFF, 11168-\textit{Δcj0031}::kan and 11168-\textit{Δcj0031}c. In contrast to the data presented in Table 6, the same pattern of plaque forming abilities was detected for 7 of these phages on mutant, ON, and OFF and complemented mutant strains (Table 7), however four phages were found to be sensitive to the restriction activity of Cj0031 restriction endonuclease.
## Table 7: Lytic spectrum for phages sensitive to Cj0031

<table>
<thead>
<tr>
<th>Phage code</th>
<th>Host</th>
<th>11168-cj0031-ON</th>
<th>11168-cj0031-OFF</th>
<th>11168-Δcj0031::kan</th>
<th>11168-Δcj0031c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>Φ1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>Φ2</td>
<td>-</td>
<td>CL</td>
<td>OL</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>Φ5</td>
<td>SCL</td>
<td>CL</td>
<td>SCL</td>
<td>&lt;SCL</td>
</tr>
<tr>
<td>9</td>
<td>Φ6</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
<td>CL</td>
</tr>
<tr>
<td>10</td>
<td>Φ9</td>
<td>OL</td>
<td>CL</td>
<td>CL</td>
<td>OL</td>
</tr>
<tr>
<td>11</td>
<td>Φ16</td>
<td>SCL</td>
<td>CL</td>
<td>SCL</td>
<td>+++</td>
</tr>
</tbody>
</table>

Key – CL: Confluent clear lysis; OL: (Confluent) opaque lysis; SCL: Semi-confluent lysis; <SCL: Less than semi-confluent lysis; +++: >100 discrete plaques; ++: 50-99 discrete plaques; +: 20-49 discrete plaques; ±: 1-19 discrete plaques; -: No lysis
These phages were Φ2, Φ5, Φ9 and Φ16. The NCTC codes for these sensitive phages - Φ2, Φ5, Φ9, Φ16 - were NCTC12674, NCTC12678, NCTC12669 and NCTC12670, respectively. In the subsequent sections, only NCTC codes will be used. These phages seemed to be restricted on 0031-ON and complemented mutant and to be propagated on the cj0031- OFF strain and mutant. This outcome was in line with our hypothesis of restriction and protection on 0031-ON and OFF strains respectively.

In order to determine the proportions of infective phage particles (EOP), the sensitive phages were recovered from the 11168-Δcj0031::kan plate, titrated and serial dilutions of order 10^{-1} to 10^{-6} were spotted back on the 11168-Δcj0031c and 11168-cj0031-OFF and 11168-Δcj0031::kan plate. The 11168-cj0031-ON variant was not selected for this experiment, because complemented mutant showed greater degree of restriction on phages than an ON variant of cj0031 as it is clear from the lytic spectrum (table 7). The dilutions having countable plaques were selected. From mean values of three of dilutions, PFU/ml was enumerated. The EOP was calculated by dividing the PFU/ml with PFU/ml values derived from propagation of phages on C. jejuni strain 11168-Δcj0031::kan. Hence the mutant values all equal one. EOP values of 11168-cj0031-OFF were also found to be greater than one or almost one. This indicates that phage replication was similar on the wild-type and mutant strains. However, C. jejuni strain 11168-Δcj0031c (r^+m^+) restricts phages strains NCTC12678.Δcj0031 and NCTC12670.Δcj0031 by factors of 5 fold and 2.2 fold respectively, suggesting the effective restriction of these phages by R.Cj0031 (Table 7). This observation demonstrated that Cj0031 restriction endonuclease was active against these phages. However, 11168-Δcj0031c (r^+m^+) showed weak restriction and no restriction at all on phages NCTC12674.Δcj0031 and NCTC12669.Δcj0031, respectively. This furnished
further evidence of the presence of endonuclease activity of Cj0031.

Table 8: EOP values showing the degree of sensitivity of phages to R. Cj0031

<table>
<thead>
<tr>
<th>Phages</th>
<th>Relative EOP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11168-Δcj0031c</td>
</tr>
<tr>
<td>NCTC12674.Δcj0031</td>
<td>1</td>
</tr>
<tr>
<td>NCTC12678.Δcj0031</td>
<td>0.20</td>
</tr>
<tr>
<td>NCTC12669.Δcj0031</td>
<td>0.71</td>
</tr>
<tr>
<td>NCTC12670.Δcj0031</td>
<td>0.46</td>
</tr>
</tbody>
</table>

6.3. Discussion

RM systems are prevalent in pathogenic bacterial strains, and are involved in maintaining the integrity of bacterial genomic DNA by preventing the entry of foreign DNA, especially from phages. A typical RM system recognizes the foreign unmethylated DNA and restricts it with the assistance of its cognate restriction endonuclease component. *C. jejuni* possesses a phase variable cj0031 gene encoding a putative type IIG restriction-modification system. Empirical evidence of an active status of Cj0031 methyltransferase was provided in chapter 4, but its restriction activity was
yet to be characterized. In this study, I have shown that Cj0031 restriction endonuclease is also active, by carrying out phage assays which indicate that this putative enzyme restricts the unmodified phage DNA entering the host strain.

6.3.1. CP8 and CP30 phages are sensitive to Cj0031

In order to analyse the restriction activity of Cj0031, a total of 60 phages have been propagated on 11168-cj0031-ON, 11168-cj0031-OFF and mutant strains of C. jejuni. A lower number of plaques were found for some of these phages on 11168-cj0031-ON compared to the 11168-cj0031-OFF and 11168-△cj0031, suggesting that plaque forming abilities of these phages were decreased due to the potential restriction of phage DNA by Cj0031 restriction endonuclease which recognizes the same recognition sequence as is recognized by Cj0031 methyltransferase. The lower plaque forming abilities of the sensitive phages on 11168-cj0031-ON and higher plaques on mutant and OFF versions of C. jejuni also suggested that these phages may possess restriction sites for the Cj0031 restriction endonuclease. More detailed explanations can be drawn out by examining the phenotypes of the strains of interest used during the experiment.

6.3.2. Phenotypes of C. jejuni PT14, HPC5 and 11168 host strains and their effects on phage propagation

A homologue of cj0031 was located in the PT14 genome sequenced by Ian Connerton group (unpublished). A 75% identity (2800aa) was found between the N-terminal end of cj0031 from 11168 and its homologue in PT14 (Figure 1), indicating that PT14-cj0031 might be capable of restriction/methyltransferase activity. The CP8 and CP30 phages propagate on PT14, exhibiting confluent clear lysis. This indicates that PT14 might not
contain any PT14-cj0031 or any other RM system mediated restriction activity against these phages. Furthermore, it also serves as evidence for PT14 being a restriction deficient strain against at least these phages. This might be an explanation for its application for typing of *C. jejuni* phages.

**Figure 1: Homology between Cj0031 genes in *C. jejuni* 11168 and PT14 strains**

The green highlighted bar area represents the identical amino acid sequences between Cj0031 in 11168 and cj0031 in PT14; the red area of the bar shows non-matching amino acid sequence.

In contrast, when phages- CP8 and CP30- were propagated on PT14 and PT14PL strains there was restriction by an active restriction modification system in the second host (11168-cj0031-ON). Thus, when 11168-cj0031-ON and 11168-cj0031-OFF and 11168-Δcj0031 were challenged with C8/PT14, CP8/PT14PL, CP8/PT14 and CP8/PT14PL, the unmodified phage genomes were restricted by 11168-cj0031-ON but infected OFF and mutant strains producing higher numbers of plaques on them (Table 2). This indicates that phages liberated by the PT14 host strains were sensitive to the 11168-Cj0031 endonuclease. This observation led to the conclusion that the homologue
of *cj0031* in PT14 does not modify the same recognition sites as 11168-Cj0031, suggesting that Cj0031 in PT14 has a different specificity compared to Cj0031 in strain 11168.

6.3.3. **Unmodified phages isolated from a 11168 mutant are restricted by Cj0031:**

This preliminary experiment with strain PT14 suggested that the reduction in plaque forming abilities of sensitive phages was caused by Cj0031 mediated restriction of phage genomic DNA. It could be subjected to a further test by showing that unmodified phages recovered from the strain 11168 *cj0031* mutant could be also be restricted in the presence of Cj0031. In order to demonstrate this, the recovered phages from strain 11168 *cj0031* mutant strain were propagated on 11168-*cj0031*-ON and complemented mutant of *cj0031* by using the standard titer of phages (10^6 PFU/ml). Unexpectedly, confluent lysis of bacterial cells was observed on the lawn of plates. However, when the standard titer of phages were diluted further to 1.0 x 10^5 PFU/ml and 1.0 x 10^4 PFU/ml, it showed a difference between lytic spectra generated by sensitive phages on 11168-*cj0031*-ON, 11168-Δcj0031c and mutant of *cj0031*, suggesting that the titer of phages during infection of RM variants of bacteria is very important. By using the increased titer of phages, it increases the probability of phages escaping the restriction activity of the restriction component of an RM system and becoming modified by the cognate methyltransferase, because of competition between the modification and restriction components of an RM system. Subsequently, the modified phages become immune to the restriction enzyme and would start infecting the surrounding bacteria. This is why it is necessary to test a range of dilutions of phages for each experiment. Previously, Frost *et al.* (1999) described that a dilution phages causing less than clear lysis (<CL) should be regarded as a standard routine dilution,
which they determined to be in the range of $10^{-6}$ to $10^{-7}$ for typing phages. This provided the logic for why the high titer experiments described herein produced lytic spectra showing clear lysis on all *C. jejuni* strains under investigation.

### 6.3.4. General applicability of results to other phages

A panel of 46 phages was tested for sensitivity to Cj0031 and it was discovered that EOPs of two typing phages named NCTC12678 and NCTC12670 were decreased by two to five orders of magnitude on 11168-Δcj0031c compared to *C. jejuni* strains 11168-cj0031-OFF, 11168-Δcj0031. This also provided further evidence in the support of a hypothesis that Cj0031 restriction endonuclease is active and functional when it is expressed under a strong promoter.

Sensitive phages have shown various degrees of sensitivity to the Cj0031. For example, NCTC12678 demonstrated five-fold sensitivity to Cj0031 compared to the NCTC12670 which showed only 2.2-fold sensitivity. This endonuclease activity of Cj0031 on these phages was lesser than restriction activities of endonucleases possessed by *E. coli*, which is as high as 50 fold of Cj0031 mediated restriction activity (Arber, 1966). This indicates that restriction activity of Cj0031 is marginal in comparison with restriction activities of *E-coli* based endonucleases.

Furthermore, This is indicative of fact that Cj0031 performs restriction on NCTC12678 in a more efficient way than on NCTC12670. This experimental observation can be explained by taking into consideration the number of sites of action of Cj0031 in the genome of invading phages. It was reported that the efficiency of restriction by restriction endonucleases is directly proportional to the number of recognition sites located in the genome of invading foreign DNA (Wilson and Murray, 1991; Franklin...
and Dove, 1969; Murray et al., 1973), suggesting that genomes of phages having less restriction sites for the host RM system would have a greater probability to escape its restrictive action. This absence or reduce number of restriction sites allows these phages to perpetuate in their bacterial host and ultimately increase their fitness as an invader.

6.3.5. The genomes of sensitive phages contain restriction site for Cj0031

The number of specificity sites on phage genomes for Cj0031, and probably their distribution may exert some influence on the strength of restriction of phages (Arber 1966). The genomes of most of these phages are yet to be sequenced so it is not possible to search for the proposed restriction sites for Cj0031. However, CP30 was sequenced and so the genome sequence was retrieved from the NCBI database and searched for proposed sites for Cj0031. There were three 5’-CCCGA-3’ and five 5’-CCGAA-3’ sites in the genome of CP30, supporting the experimental observation of restriction of CP30 phage by Cj0031. Thus it is proposed that the genomes of other sensitive phages might also contain the restriction sites for Cj0031, whilst the phages whose lytic spectrum was not influenced by Cj0031, may be lacking recognition sites for Cj0031. These findings have implications for when C. jejuni is invading chickens by offering resistance to phages already residing in the guts of chickens.

6.3.6. Cj0031 switching in vivo may confer resistance to C. jejuni against phages:

The phage tested for their sensitivity to Cj0031 constitute part of the gut flora (Hansen et al., 2007). Carvalho et al. (2010) and Carrillo et al. (2005) reported that C. jejuni developed transient reversible resistance against the phages during phage therapy used
to reduce the *C. jejuni* population, but they did not provide any empirical evidence for that resistance. I hypothesize that this resistance in vivo may be linked to SSR mediated phase variation of *cj0031*. Phase variation of *cj0031* during in vivo passage through chickens was demonstrated by Bayliss *et al.* (2012). They further showed that *cj0031* was found in an ON phase during the in vivo passage through chicken. This implicates that the resistance phenotype of *C. jejuni* strain NCTC11168 against invading phages may perhaps be connected to the restriction activity exerted by Cj0031 on them. This hypothesis needs to be proved empirically, but might have important implications for in vivo phage therapy.

It is interesting to note that homologues of Cj0031 were also present in other *C. jejuni* strains as indicated by the phylogenetic tree in chapter 4 section... But it is not phase variable in other strains due to the lack of a repeat tract within its ORF in other strains. However other mechanisms of phage resistance are present in *C. jejuni* strains. For example, *C. jejuni* strain HPC5 employs recombination mediated inversion mechanism in order to generate a phage resistance phenotype against CP34 phage. Ahmed *et al.* (2002) detected an abortive phage protein in *C. jejuni* strain 81116, which was absent in 11168 strain, suggesting that it might be a potential candidate to offer resistance against phage infection. In other strains such as *C. jejuni* RM1221, *C. upsaliensis* RM 3195 and *C. coli* RM2228, there are several potential type I RM systems which may act as an immune system to guard them against phages (Fouts *et al.*, 2005).

In short, this study demonstrated that Cj0031 possesses restriction activity against a selected group of phages. This restriction activity is enhanced when the gene is expressed under a strong promoter. In the previous chapter, I have shown empirically that it also performs a modification function. Together, all this data shows that Cj0031 is a functional type IIG phase variable RM system in *C. jejuni*. 
Chapter 7. General Discussion

Campylobacteriosis, a food-borne disease, has posed a major health threat to many developed countries. The efforts to control *C. jejuni* infections in humans are mainly focused around the understanding of virulence factors and mechanisms by which they contribute directly to the virulence and pathogenesis of *C. jejuni*. Though major breakthroughs and developments towards understanding biological facets of *C. jejuni* have happened in the last two decades following the sequencing of a complete genome of *Campylobacter* in the year 2000 (Parkhill *et al.*, 2000). Nevertheless, the knowledge of such virulence factors and their mechanism of action are yet to be fully elucidated. Furthermore, the discovery of homopolymeric tracts prone to slipped-strand mispairing in the ORFs of virulent genes, led to the identification of phase variation as an important phenomenon in *C. jejuni*. The variability in these homopolymeric tracts led to the conclusion that phase variable genes are major players in controlling the survival and virulence of different strains of *C. jejuni* (Bourke, 2002). In recent years, another novel mechanism of gene regulation has been identified involving methylation dependent phase variation mediated by phase variable type III restriction-modification systems. These systems were found to be acting as phasevarions of many virulent unlinked genes in the genome of host-adapted pathogens (Srikhanta *et al.*, 2005; 2009; 2011).

The current project identified a phase variable type IIG RM system-*cj0031*- in *C. jejuni* strain NCTC11168. This study found that *cj0031* switched from an OFF to ON phase during passage through chickens, followed by determination of in vitro phase variation rates by the construction of a *cj0031*-LacZ reporter constructs (Bayliss *et al.*, 2012). The major shifts in “OFF/ON” proportions of *cj0031* during in vivo passage through
chickens and high mutation rates of in vitro suggested that \textit{cj0031} might be playing an important role for the virulence and pathogenesis of \textit{C. jejuni}. The role of \textit{cj0031} in regulation of global gene expression in \textit{C. jejuni}, like a type III mod phasevarion was postulated.

The project investigated the methylation activity of \textit{cj0031} by performing southern blots. This approach demonstrated that Cj0031 does possess methyltransferase activity. However, the investigation could not reach the final conclusion of drawing out the recognition sequence. Two potential recognition motifs -CCCGA and CCCGAA- were identified for Cj0031 by the present study. Following identification of its methylation activity, some phenotypic assays were performed in order to determine whether \textit{cj0031} influences virulence of \textit{C. jejuni}. An effect of \textit{cj0031} on motility of \textit{C. jejuni} was not found. However, mutations in \textit{cj0031} affected adhesion and invasion by \textit{C. jejuni}. Biofilm formation was also influenced by the mutation of \textit{cj0031}. Of note, \textit{cj0031} is a type IIG restriction modification system, which means that this enzyme is not directly involved in controlling these virulence-associated phenotypes. These findings suggested that \textit{cj0031} likely behaves as a phasevarion (\textit{phase variable regulon}) like the \textit{mod} component of type III RM system in many other pathogenic bacteria such as \textit{H. influenzae}, \textit{N. menigitidis} and \textit{H. pylori}. The most of the putative sites of Cj0031 were found to be located in the intergenic region near promoters of genes

Against the backdrop of findings of the present study, a model has been proposed here to elucidate the potential mechanism of regulation of gene expression by a phase variable \textit{cj0031} as a phasevarion of genes. The model will be supported by similar data
from a limited number of studies describing the phasevarion phenomenon ascribed to mod gene in other host-adapted pathogens.

The phase variation of *cj0031* generates two set of populations-11168-*cj0031*-ON and 11168-*cj0031*-OFF. Both sets of populations of *C. jejuni* are phenotypically different. The 11168-*cj0031*-ON population contains cells with an active Cj0031 methyltransferase which methylates its recognition sites in the genome. This differential methylation by Cj0031 might affect the expression of multiple genes carrying Cj0031 recognition sites in their upstream region or promoter regions. The changes in expression of affected genes probably control the survival under stress and adaptation to different niches, enabling *C. jejuni* to become a successful host-adapted pathogen. On the other hand, in the 11168-*cj0031*-OFF population, the gene is frameshifted, and Cj0031 methyltransferase is not available to differentially methylate the set of genes affected by Cj0031. Consequently, this leads to the creation of two phenotypically different populations. The phasevarion phenomenon related to phase variable type IIG Cj0031 RM system stands in contrast to the simple sequence repeat mediated phase variation phenomenon associated with genes containing repeats, because the former affects the expression of multiple genes while the latter only affects the expression of specific genes.

The important question is which genes are likely to be regulated by Cj0031. The project provided empirical evidence of an influence of Cj0031 on the adhesion and invasion of *C. jejuni* to Caco-2 epithelial cells. There are several bacterial factors involved in the adhesion and invasion of *C. jejuni* to the host which are likely to influenced by phase variable *cj0031*. The model proposes the impact of Cj0031 on previously characterized adhesion and invasion factors whose mechanism of controlling these processes are still under investigation. For instance, CadF protein (Outer membrane protein binding to
Fibronectin) [Konkel et al., 1997, PEB1 (periplasmic binding protein) [Pei and Blaser, 1993], JlpA (surface exposed lipoprotein) [Jin et al., 2001], pldA (outer membrane phospholipase) [Ziprin et al., 2001] and fIP (fibronectin-like protein A) [Flanagan et al., 2009; Leon et al., 2006] are involved in the adhesion and binding of C. jejuni cells to the host epithelial tissues. Some of these and or all factors are likely to be regulated by differential methylation mediated by Cj0031. Similarly, the factors which are essential for invasion and anticipated to be regulated to Cj0031 include CiaB (invasive antigens, secreted through the flagellar export apparatus) [Rivera-Amill et al., 2001; Konkel et al., 2004), lipooligosaccharide (Louwen et al., 2008), CPS (capsular polysaccharide) [Karlyshev and Wren, 2001] and CadF (activation of Rac1 and Cdc42) [Krause-Gruszczynska et al., 2007], docC/docB (prutative MCP protein) [Vegge et al., 2009; Hendrixson et al., 2004], and tlp1 (chemoreceptor-transducer-like protein) [Vegge et al., 2009; Hartley-Tassel et al., 2010]. The putative recognition sites could not be found in the promoter regions of these genes but they were located in the intergenic region of many of these genes.

Although, flagellum is a well-characterized factor whose impact has been demonstrated on adhesion and invasion (Wassenaar and Blaser, 1999), the current study concluded that Cj0031 does not control the adhesion and invasion of the C. jejuni to Caco-2 cells by affecting the flagellum. It is interesting to note that most of these factors belong to outer membrane proteins and some of them are excreted through flagellar secretory apparatus. The model suggests that recognition site of Cj0031 might be located in the upstream intergenic region or promoter of these genes, thereby affecting the expression of genes through differential methylation of their sites. This proposed model of Cj0031 has been based on the observations from studies which suggested that mod A13 dependent upregulated operons and genes in N. gonorrhoeae strain FA 1090 contain
modA13 methylation motifs in the upstream intergenic regions. However, how methylation of these sites controls the expression of these genes and operons is still waiting for experimental evidence. Indeed, the mechanism of controlling the gene expression of phasevarions by methylating intergenic recognition sites has not been tested empirically for any of the current systems (Srikhanta et al., 2010).

There is a set of 17 genes which are subjected to modA13 dependent regulation of expression in *N. gonorrhoeae*. Five of these genes were related to virulence of *N. gonorrhoeae*: one gene performed antimicrobial activity and four genes were implicated in equipping the bacteria to cope with oxidative stress. Similarly, modA11, a phase variable type III RM system, up-and down-regulated the expression of 80 genes, five of which encoded surface exposed proteins directly involved in virulence of *N. meningitidis* (Srikhanta et al., 2009). Moreover, the functional studies conducted by Srikhanta et al (2009) on the ON and OFF variants of modA13 confirmed the role of this gene in regulation of biofilm formation, adhesion and invasion of human epithelial cells and antimicrobial activities. These findings were consistent with a role of phase-variable mod alleles in generating different cells types specialized for adaptation to different niches. They suggested that phasevarions of genes via phase variable mod alleles in *N. gonorrhoeae* and *N. meningitidis* seems to be a common strategy for randomly switching between “differentiated” cell types to mediate the adaptation to the new host.

In *H. pylori*, the restriction function (res) of many type II and Type III RM systems was found to be inactive, indicating that these RM systems might have involved in performing some other function such as gene regulation, other than restriction (Skoglund et al., 2007; Fox et al., 2007). In *H. pylori*, the regulation of katA gene expression by M. HpyAIV, a putative phase-variable mod gene linked with type II RM
system was observed by Skoglund et al. (2007). Furthermore, its involvement in gene regulation during mice colonization experiments also suggested by Bjorkholm et al. (2002) who reported that mutation in M. HpyAIV caused the induction of a more robust host response for invading bacteria. Srikhanta et al (2011) reported by conducting microarray analysis of phase-variable modHS (ON) and mutant that flagellar genes (HP0609 and HP0870) were downregulated and an outer membrane protein (HP0253) were upregulated in the H. pylori strain P12 modHS mutant, suggesting that differential methylation caused by modHS affects the global gene expression in H. pylori.

In H. influenzae, phase variable mod gene (part of a type III RM system) was shown to be involved in downregulating outer-membrane proteins which perform transport functions and up-regulating of heat-shock proteins involved in coping with environmental and physiological sources of stress (Srikhanta et al., 2005). The mod-dependent phasevarion of genes in H. influenzae strain Rd generated two distinct cell types which could survive in two different physical environments.

Together, these studies suggest that genes regulated by the phasevarion effect of mod genes in different pathogenic strains seem to be encoding heat-shock proteins, surface exposed lipooligosaccharide biosynthesis, and transport-surface exposed proteins. Based on these findings in other model pathogenic bacteria, it is anticipated that there may be several genes controlling the phenotypes of adhesion and invasion, biofilm formation and survival under stress in C. jejuni under the phasevarion influence of Cj0031. The alterations in the and types of genes affected by Cj0031 can be determined by conducting further studies using epigenetic microarray technique or RNAseq analysis in C. jejuni, similar to those carried by Srikhanta et al. (2005; 2009; 2011) in H. influenza, N. gonorrhoeae, N. meningitides and H. pylori.
The \textit{cj0031} gene is subject to ON/OFF switching. How can the switching of \textit{cj0031} between ON and OFF phases be beneficial to the host strain? The switching of \textit{cj0031} into an ON phase may facilitate the host strain to offer resistance against infections caused by selected groups of phages. This resistance is offered at two levels: phage adsorption and phage restriction. In the first instance, \textit{cj0031} as a defense system can restrict the unmodified attacking phage and reduce the infection events of the host bacteria. Restriction by Cj0031 may prevent the post-adsorption events of phage propagation within the host. However, Cj0031 as a phasevarion may affect the expression of unlinked surface genes encoding the potential receptors for adsorption of phages, thereby preventing adsorption of phages. Contrastingly, the OFF phase of \textit{cj0031} may aid the acquisition of beneficial phages conferring a fitness advantage to the host. Moreover, switching of \textit{cj0031} to off state may prevent the phages from evolving a counter-defense mechanism in response to the host defense system, especially by weakening the selective pressure for resistant phages. The postulates of this section of the model can be supported by the following studies.

The phase variable type I RM system (Hind1) in \textit{H. influenzae} confers resistance against phages (Zaleski \textit{et al.}, 2005, De Bolle \textit{et al.}, 2000). Hind1 confers resistance against HP1c1 phages due to phase variation caused by a pentanucleotide repeat tract within the ORF of the \textit{hsdM} gene. Another example concerns the phase variable type III RM systems in \textit{meningococci}. An active prophage integrated chromosomally within invasive meningococcal isolates, is thought to be required for successful transmission of these isolates from one host to another host (Bille \textit{et al.}, 2005). Moxon and Jansen (2005) argued that acquisition of this prophage might be aided by switching the type III RM system (mod gene) into an OFF state. \textit{pglX}, a phase variable methyltransferase system in \textit{S. coelicolor}, switches off to induce the expression of adjacent \textit{pglS} gene.
which confers resistance to host strains against different groups of phages (Laity et al., 1993; Sumby and Smith, 2003). Finally, Sorensen et al. (2012) demonstrated that modification of capsular polysaccharides structures in NCTC11168 reduced the plaque forming abilities of four phages - F198, F287, F303, and F326 – but they did not show the source of that modification. So it might be possible that modification of CPS accrued due to Cj0031. Thus cj0031 might enhance the resistance phenotype of the host strain by altering the structure of receptors.

In short, it can be concluded that phase variation of cj0031 might offer phage resistance to and acquisition of phages for host strain 11168. The defense may be partly due to a restriction-based approach and partly due to a receptor based modification mechanism. These hypotheses can be proved by conducting further studies of the interactions between phages and C. jejuni strain 11168.
Bibliography


that strongly associates with colonization and disease in a mouse model. *Microbiology*, 158 (5): 1304-16


pneumoniae on other inhabitants of the upper respiratory tract. *Infect. Immun.*, **68**, 3990-3997.


Appendix 1

First Experiment

Adhesion Assay

<table>
<thead>
<tr>
<th>strains</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Average</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-ON</td>
<td>462000</td>
<td>492000</td>
<td>622000</td>
<td>525333.33</td>
<td>85049.01</td>
</tr>
<tr>
<td>WT-OFF</td>
<td>266000</td>
<td>230000</td>
<td>224000</td>
<td>240000.00</td>
<td>22715.63</td>
</tr>
<tr>
<td>Δ0031::Kan</td>
<td>280000</td>
<td>266000</td>
<td>220000</td>
<td>255333.33</td>
<td>31390.02</td>
</tr>
<tr>
<td>CΔ0031</td>
<td>500000</td>
<td>480000</td>
<td>456000</td>
<td>478666.67</td>
<td>22030.28</td>
</tr>
</tbody>
</table>

Second Experiment

Adhesion Assay

<table>
<thead>
<tr>
<th>strains</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Average</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-ON</td>
<td>500000</td>
<td>460000</td>
<td>520000</td>
<td>493333.33</td>
<td>30550.5</td>
</tr>
<tr>
<td>WT-OFF</td>
<td>260000</td>
<td>240000</td>
<td>180000</td>
<td>226666.67</td>
<td>41633.32</td>
</tr>
<tr>
<td>Δ0031::Kan</td>
<td>200000</td>
<td>240000</td>
<td>20000</td>
<td>153333.33</td>
<td>117189.3</td>
</tr>
<tr>
<td>CΔ0031</td>
<td>440000</td>
<td>440000</td>
<td>380000</td>
<td>420000</td>
<td>34641.02</td>
</tr>
</tbody>
</table>

Second experiment

Invasion Assay

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Average</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-ON</td>
<td>32000</td>
<td>28000</td>
<td>38500</td>
<td>32833.33</td>
<td>5299.371</td>
</tr>
<tr>
<td>WT-OFF</td>
<td>12000</td>
<td>20000</td>
<td>16000</td>
<td>16000</td>
<td>4000</td>
</tr>
<tr>
<td>Δ0031::Kan</td>
<td>14000</td>
<td>12000</td>
<td>12000</td>
<td>12666.67</td>
<td>1154.701</td>
</tr>
<tr>
<td>CΔ0031</td>
<td>26000</td>
<td>28000</td>
<td>34000</td>
<td>29333.33</td>
<td>4163.332</td>
</tr>
</tbody>
</table>

Second experiment

Invasion Assay

<table>
<thead>
<tr>
<th>strains</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-ON</td>
<td>58000</td>
<td>68000</td>
<td>46000</td>
<td>57333.33</td>
<td>11015.14</td>
</tr>
<tr>
<td>WT-OFF</td>
<td>18000</td>
<td>26000</td>
<td>23000</td>
<td>22333.33</td>
<td>4041.452</td>
</tr>
<tr>
<td>Δ0031::Kan</td>
<td>16000</td>
<td>22000</td>
<td>16000</td>
<td>18000</td>
<td>3464.102</td>
</tr>
<tr>
<td>CΔ0031</td>
<td>42000</td>
<td>36000</td>
<td>54000</td>
<td>44000</td>
<td>9165.151</td>
</tr>
<tr>
<td></td>
<td>48 Hours</td>
<td></td>
<td>72 hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>-----</td>
<td>----------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WT-ON</td>
<td>WT-OFF</td>
<td>Δ0031</td>
<td>CΔ0031</td>
<td>WT-ON</td>
</tr>
<tr>
<td>3.5</td>
<td>0.79</td>
<td>1.609</td>
<td>2.5</td>
<td>4.5</td>
<td>1.3</td>
</tr>
<tr>
<td>3.5</td>
<td>0.61</td>
<td>0.892</td>
<td>1.09</td>
<td>4.8</td>
<td>1.78</td>
</tr>
<tr>
<td>3.5</td>
<td>0.75</td>
<td>0.857</td>
<td>1.09</td>
<td>4.3</td>
<td>1.23</td>
</tr>
<tr>
<td>3.1</td>
<td>0.8</td>
<td>0.709</td>
<td>2.21</td>
<td>3.99</td>
<td>2.1</td>
</tr>
<tr>
<td>3.4</td>
<td>0.15</td>
<td>0.5</td>
<td>1.57</td>
<td>4.5</td>
<td>1.88</td>
</tr>
<tr>
<td>3.5</td>
<td>0.59</td>
<td>0.13</td>
<td>1.89</td>
<td>5.2</td>
<td>1.21</td>
</tr>
<tr>
<td>Mean</td>
<td>3.41667</td>
<td>0.616</td>
<td>0.782833</td>
<td>1.725</td>
<td>4.548333</td>
</tr>
<tr>
<td>SD</td>
<td>0.160208</td>
<td>0.273825</td>
<td>0.492261</td>
<td>0.581988</td>
<td>0.416193</td>
</tr>
</tbody>
</table>

6-fold   5-fold  2

p=0.0001
p<0.0001