Species wide analysis of ***Escherichia coli*** identifies a negative regulatory input that controls *flhDC* transcription

Senan Almashta M.Sc., B.Sc.

Thesis submitted in partial fulfilment of the requirements of the regulations for the degree of Doctor of Philosophy

Centre for Bacterial Cell Biology
Institute for Cell and Molecular Biosciences
Faculty of Medical Sciences
June 2018

Abstract

Escherichia coli is a species of bacteria that can be found almost everywhere. It belongs to the family *Enterobacteriaceae* which, in turn, belongs to gamma subdivision of the phylum Proteobacteria. Genomic plasticity and the ability of genes to transfer across different strains contributed to the division of *E. coli* into six major phylogenetic lineages: A, B1, B2, D, E and F. The motility organelle of bacteria is the flagellum which consists of: basal body, hook and filament. The flagellar system of *E. coli* organised into a transcriptional hierarchy strictly dependent on the expression of the flagellar master regulator, $FlhD_4C_2$.

Bioinformatic analysis of the *yecG-flhDC* intergenic region showed that the DNA sequence of this region could be grouped into a phylogeny that reflects the phylogenetic clades of *E. coli*. However, three dominant sequence types of the *yecG-flhDC* intergenic region mapped to clades: A/B1, B2 and D/E/F. We illustrated the impact of replacing the *yecG-flhDC* intergenic region of the weakly motile strain MG1655 with the same region of model strains on motility. Our data revealed dramatic increase in motility in the newly created strains. We also dissected the *yecG-flhDC* intergenic region to investigate the effect of each part on motility. Our analysis revealed that the *yecG* is a negative regulator of *flhDC* expression.

We tested the impact of host niche environments on the motility phenotype of *E. coli*. For this reason, we sequenced a collection of bovine strains and aligned them to their phylogenetic groups and compared them to another collection isolated from UTIs. The data showed that the bovine isolates lie mostly in clades B1 and A, whereas UTI isolates mostly occupy clades B2 and D/F. The bovine strains exhibited robust motility compared to the UTI strains. We screened the *yecG-flhDC* intergenic region for IS elements, and found no isolate has the IS element.

Dedication

This work is dedicated to the memory of my mother and father.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Phillip Aldridge for providing me with this great opportunity to study a PhD in a well accredited university which has an advanced global ranking. His assistance has been vitally important in supervising me and helping me to maintain focus in order to achieve the findings for this project. I would also acknowledge my assessors Dr. David Bolam and Dr. Henrik Strahl for their advice throughout my work. I would like to express my gratitude to Prof. Mark Fielder at Kingston University in London for providing the bovine strains. A special thanks to my postgraduate tutor Dr. Timothy Cheek for his continuous support and kind understanding.

My thanks to everyone who helped me in the Centre for Bacterial Cell Biology and the Medical School at Newcastle University. I would also like to include my heartfelt appreciation for my friends in Iraq and the UK who exhibited real support to me in my life, during times of both happiness and sadness. A big thanks to my sponsor: The Higher Committee for Education Development in Iraq for sending me to study a PhD in the UK, and for their funding during my scholarship. Notable thanks to the Ministry of Higher Education and Scientific Research in Iraq and the University of Babylon for their approval in allowing me to gain this doctorate.

Senan Almashta, May 2018

Table of Contents

Table of Figures

Table of Tables

List of Abbreviations

˚C	Degree Celsius
µg	Microgram
µl	Microlitre
µm	Micrometer
A.U.	Arbitrary unit
A/E	Attaching and effacing
A˚	Angstrom
AAF	Aggregative adherence fimbria
ABU	Asymptomatic bacteriuria
AIEC	Adherent invasive *E. coli*
Amp	Ampicillin
Ap58	Aggregative protein 58
APEC	Avian pathogenic *E. coli*
ATP	Adenosine triphosphate
bp	base pair
cAMP	Cyclic Adenosine monophosphate
cAMP-CAP	cAMP catabolite activator protein
Cas9	CRISPR associated protein 9
c-di-GMP	Cyclic di-guanosine-monophosphate
CEACAM	Carcinoembryonic antigen-related cell adhesion molecule
cm	Centimetre
Cm	Chloramphenicol
CNF	Cytotoxic necrotizing factor
CRISPR	Clustered regularly interspaced short palindromic repeats
CRP	cAMP receptor protein
crRNA	CRISPR RNA
C-terminus	Peptide carboxy terminus
cTet	Chlortetracycline
CUP	Chaperone-usher pathway
DAEC	Diffusely adherent *E. coli*
DAF	Decay accelerating factor
DGC	Diguanylate cyclase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid

dNTP	deoxyribonucleotide triphosphate
DSB	Double-strand break
E.	*Escherichia*
EAEC	Enteroaggregative *E. coli*
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extraintestinal pathogenic *E. coli*
FLP	Flippase
FRT	Flippase recognition target
g/l	gram per litre
Gb3s	Globotriaosylceramides
GEI	Genomic island
GFP	Green fluorescent protein
H	Flagellar antigen
HAP	Hook associated protein
HBB	Hook basal body
HGT	Horizontal gene transfer
H-NS	Histone-like nucleoid-structuring protein
IBCs	Intracellular bacterial communities
IM	Inner membrane
IPEC	Intestinal pathogenic *E. coli*
IR	Inverted repeat
IS	Insertion sequence
K	Capsular antigen
Kan	Kanamycin
kb	kilobase pair
KDa	kilodalton
L	Litre
LB	Luria Bertani
LBP	Lucifirin binding protein
LDC	Lysine decarboxylase
LEE	Locus of enterocyte effacement
LPS	Lipopolysaccharide
LT	Heat-labile enterotoxin

M	Molar
mg	Milligram
MGE	Mobile genetic element
MinE	Minimal E salt growth medium
ml	Millilitre
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
mM	Millimolar
mRNA	Messenger ribonucleic acid
ms	millisecond
ng	Nanogram
nm	nanometre
NMEC	Neonatal meningitis *E. coli*
N-terminus	Peptide amino terminus
O	Somatic antigen
OD	Optical density
OM	Outer membrane
OMV	Outer membrane vesicles
P	Promoter
PAI	Pathogenicity island
PCR	Polymerase chain reaction
PDE	Phosphodiesterase
PG	Peptidoglycan
PGA	Poly-β-1,6-N-acetyl-D-glucosamine
pH	Potential of hydrogen
Pic	Protein involved in colonization
PMN	Polymorphonuclear leukocyte
pmole	Picomole
REG	Regulatory zone
RNA	Ribonucleic acid
rpm	Revolutions per minute
rUTI	Recurrent UTI
s	Second
Sat	Secreted autotransporter toxin
SCR	SCAR zone
SDS	Sodium dodecyl sulfate

sgRNA	Single synthetic guide RNA
ShET	*Shigella* enterotoxin
SIDD	Stress-induced DNA duplex destabilization
SNP	Single Nucleotide polymorphism
ST	Heat-stable enterotoxin
ST	Sequence type
Stx	Shiga toxin
T	Time
T3SS	Type three secretion system
TAE	Tris Acetate-EDTA
TB	Tryptone broth
Tet, tet	Tetracycline
TLR	Toll-like receptor
TPA	Type Phillip Aldridge
tracrRNA	Trans activating crRNA
U	Unit
UPEC	Uropathogenic *E. coli*
UTI	Urinary tract infection
UTR	Untranslated zone
UV	Ultra violet
WT	Wild type
X-gal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside
YE	Yeast extract
YFP	Yellow fluorescent protein
Ypet	Yellow fluorescent protein for energy transfer
λ	Lambda
σ	Sigma factor
$\boldsymbol{\alpha}$	Alpha

Chapter One: Introduction

1.1. *Escherichia coli* basic characteristics

Escherichia coli is a Gram-negative, rod-shaped, non-spore forming, facultative anaerobic, ferment glucose, flagellated bacterium (Berg, 1998). The genus *Escherichia* belongs to the family *Enterobacteriaceae* that taxonomically is part of the gamma subdivision of the phylum Proteobacteria (Lukjancenko *et al.*, 2010). *E. coli* is an omnipresent bacterium that commonly inhabits the intestine of healthy individuals, other warm-blooded animals and reptiles within their commensal microflora. However, this species includes many pathogenic strains which can cause severe diseases in a wide range of animals including humans (Kaper *et al.*, 2004; Lukjancenko *et al.*, 2010; Tenaillon *et al.*, 2010). This project will deal in some detail with a variety of strains of this species known to be both pathogenic and non-pathogenic.

1.2. Strain diversity of the species *E. coli*

1.2.1. E. coli and its pangenome

E. coli is a versatile species that can be found almost everywhere. It may exist in abiotic environmental niches such as water, food, soil and sediment (Diaz *et al.*, 2001). Furthermore, *E. coli* is equipped with the required metabolic abilities to colonize the mammalian intestine, successfully to live alongside more than 500 other bacterial species (Tenaillon *et al.*, 2010). *E. coli* strains that cause illnesses are divided into intestinal pathogenic *E. coli* (IPEC) [also known as InPEC (Rojas-Lopez *et al.*, 2018)], and extraintestinal pathogenic *E. coli* (ExPEC). These two divisions can subsequently be subdivided into pathovars which are strains of the same species sharing defined pathogenic characteristics. Moreover, the classification of the pathovars depends on the clinical symptoms of the disease, virulence factors, epidemiology and the phylogenetic background. The main examples of IPEC pathovars are enteroaggregative *E. coli* (EAEC), enterohaemorrhagic *E. coli* (EHEC),

enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), diffusely adherent *E. coli* (DAEC) and adherent invasive *E. coli* (AIEC). On the other hand, uropathogenic *E. coli* (UPEC), neonatal meningitis *E. coli* (NMEC also called MNEC) and avian pathogenic *E. coli* (APEC) are well known examples of the ExPEC pathovars (Crossman *et al.*, 2010; Croxen and Finlay, 2010; Croxen *et al.*, 2013).

The genetic plasticity of the *E. coli* genome led to the adaptation of *E. coli* to both biotic and abiotic habitats, and also aided in the emergence of diversified lifestyles such as being mutual, commensal and pathogenic (Ochman *et al.*, 2000). Species pangenomes are defined as the complete genetic content of all isolates within a bacterial species that is often divided into two classes: the core genome and the flexible or dispensable genome (Medini *et al.*, 2005). The core genome is the genomic backbone representing the genes found in all strains of the species. It consists mainly of the housekeeping genes that are essential for replication, transcription, translation, metabolism and motility. Additionally, the flexible or dispensable genome refers to the accessory genes found only in a few strains or a group of isolates. These accessory genes such as antibiotic resistance genes are responsible for conferring different strains with various characteristics allowing them to adapt to a wide range of environments and conditions (Dobrindt *et al.*, 2004; Medini *et al.*, 2005). The presence of accessory genes is the result of the actions of the mobile genetic elements (MGEs) such as plasmids, bacteriophages, genomic islands (GEIs), integrons, transposons and insertion sequence (IS) elements (Figure. 1.1) (Dobrindt *et al.*, 2004; Wiedenbeck and Cohan, 2011). The genomic material of isolates is not fixed, but rather can be subjected to alteration through horizontal gene transfer (HGT), gene loss or insertion, DNA rearrangements, point mutations and the movement of IS elements (Figure. 1.1) (Goldenfeld and Woese, 2007; Tenaillon *et*

al., 2010; Jackson *et al.*, 2011; Leimbach *et al.*, 2013). Mobile genetic elements can transfer genes that encode virulence factors such as hemolysin A (*hlyA*) among other which can be found in ExPEC strains (Kohler and Dobrindt, 2011). GEIs are distinct mobile or nonmobile DNA fragments which can be transferred by transduction or conjugation. HGT has an important role in the evolution of a bacterial species through the transmission of novel genes such as antibiotic resistance, virulence and catabolic genes (Juhas *et al.*, 2009). Pathogenicity islands (PAIs) which are part of GEIs originally evolved from bacteriophages and plasmids (Hacker *et al.*, 2003; Dobrindt *et al.*, 2004; Ho Sui *et al.*, 2009). However, other non PAI genomic islands provide non-pathogenic environmental strains with the genetic material which impact their own evolution. In contrast, reduced genomes through deletion mechanisms are common for intracellular bacteria. For example, EIEC lacks the *cadA* gene that encodes for lysine decarboxylase (LDC) through a deletion process. A reaction catalyzed by LDC produces cadaverine which inhibits the enterotoxin activity and, thus, weakens virulence of these strains (Maurelli *et al.*, 1998). Plasmids are another way of transferring genetic material especially in pathogenic strains such as UPEC. For example, some strains of UPEC isolated from humans produce high levels of plasmid encoded microcin H47 and colicin E1 which enhance the fitness of these strains (Smajs *et al.*, 2010).

It has been demonstrated that GEIs and other transferrable genetic elements are more dynamic in intercellular communities such as commensal *E. coli* which share a niche with other bacterial species rather than an isolated bacterial community (Dobrindt *et al.*, 2004). Thus, because *E. coli* is originally a commensal bacterium that lives in the intestine with hundreds of other species, it is believed that it has acquired new traits through HGT and homologous recombination leading to new pathogenic strains emerging. Therefore, it has been assumed that commensal

strains are the origin of pathogenic isolates (Hacker *et al.*, 2003; Dobrindt *et al.*, 2004; Ren *et al.*, 2004; Tenaillon *et al.*, 2010). To conclude, *E. coli* is a versatile species of bacteria. Its genome plasticity has conferred individual strains a wide range of metabolic abilities and phenotypic diversity. Moreover, the common commensal intestinal strains are the potential source for the repertoire of pathovar emergence leading to their definition according to the site of infection as intestinal or extra-intestinal *E. coli*.

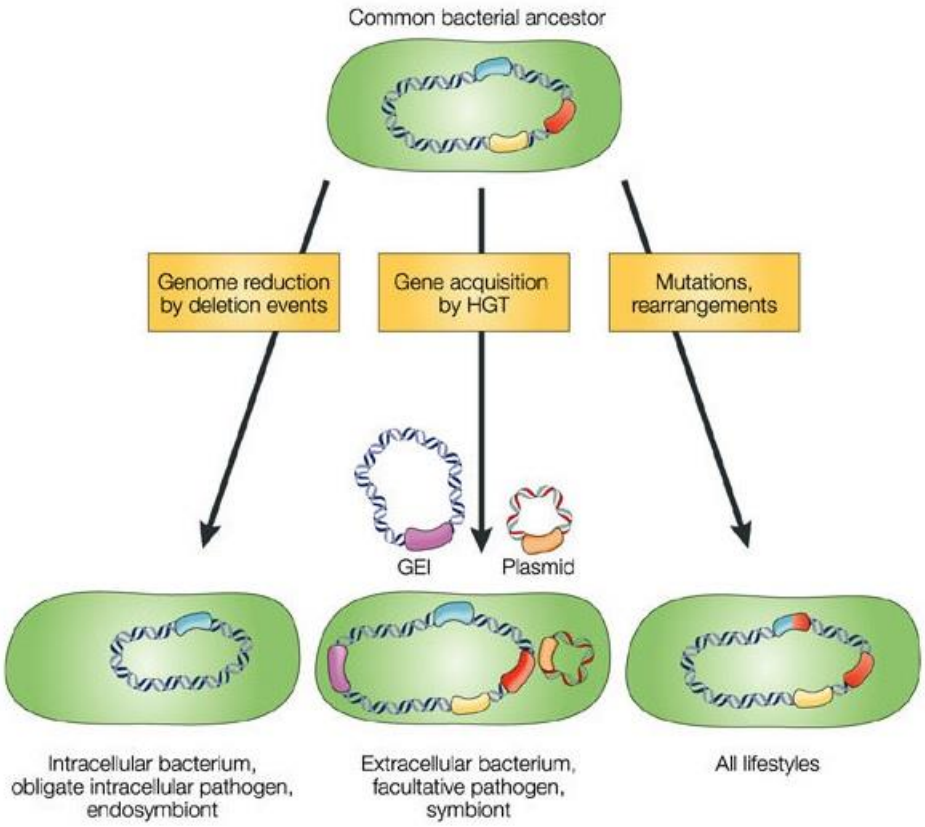


Figure. 1.1. The emergence of new strains by the change in the genetic materials.

The bacterial genome may undergo genome reduction. This type of gene alteration is common in intracellular bacteria, obligate intracellular pathogens and endosymbionts. Horizontal gene transfer (HGT) is another method of gene alteration by the acquisition of new genes. This type is mediated by genomic islands (GEI), plasmids and bacteriophages. Mutations and DNA rearrangements are also contributing to the mixing of the bacterial gene pool. New strains may emerge as a result of a mix of all three methods described above. Figure taken from (Dobrindt *et al.*, 2004).

1.2.2. Early molecular approaches for bacterial classification

The flexible genome of *E. coli* has enabled the species to adapt to a range of different host and external environments. MGEs contributed to the transfer of genetic elements across different strains with a strong evidence to suggest they played a key role in the emergence of pathogenic strains (Dobrindt *et al.*, 2004). In order to easily diagnose strains during epidemiological outbreaks, scientists are always keen to try to find a quick, easy and reliable approach to identify strains and attribute them to their phylogenetic groups. Moreover, faecal contamination of water is a serious issue. Therefore, convenient bacterial source tracking in water systems is important (Field and Samadpour, 2007).

The traditional biochemical tests such as indole, oxidase, lactose fermentation on MacConkey agar are used to regularly identify *E .coli* (Trepeta and Edberg, 1984). Serological tests for capsular (K), somatic (O) and flagellar (H) antigens are also useful in *E. coli* diagnosis (Orskov *et al.*, 1977). Polynucleotide sequence variation has also been used through DNA hybridisation to distinguish different strains (Brenner *et al.*, 1972). Furthermore, multilocus enzyme electrophoresis (MLEE) was used to characterize the microorganisms according to a number of their intracellular enzymes. MLEE exploits the theoretical aspect that genes have different alleles that potentially encode different amino acids for a specific enzyme and, thus, alter the electro charges of the protein. These differences will show variable mobilities of a given protein on a gel which can then be assigned to that specific allele (Stanley and Wilson, 2001). A study using MLEE was performed on 829 *E. coli* strains from 156 natural sources, and the variation in the enzymatic mobility due to the expression of five loci was examined (Milkman, 1973). Later studies continued to perform this procedure, but on a bigger number of loci and more divergent groups of *E. coli* were discovered (Selander and Levin, 1980). Later, Ochman and Selander defined the

ECOR strain collection consisting of 72 *E. coli* strains from different hosts and geographical areas based on MLEE (Ochman and Selander, 1984). It has been thought that this study included *E. coli* strains from different groups. However, it has since been recognised that the ECOR collection represents only part of the genetic diversity found among strains of *E. coli* (Wirth *et al.*, 2006). The MLEE method, however, proved to have limitations to draw phylogenetic groups as different enzymes expressed by genes with little similarity may show the same mobility (Bisercic *et al.*, 1991). However, MLEE analysis of the ECOR collection led to Selander, Whittam and their colleagues between 1987 and 1990 to derive phylogenetic trees based on the mobility of the alleles (Saitou and Nei, 1987; Selander *et al.*, 1987; Herzer *et al.*, 1990). These trees provide the earliest representation of the clade structure of *E. coli* reproduced by Chaudhuri and Henderson (2012) (Figure. 1.2). It is these historic trees and the subsequent analysis of more molecular based studies that define the six clades as A, B1, B2, D, E and F.

Another method used the presence of multicopy single stranded DNA produced by reverse transcription to classify *E. coli* strains as it considered evolutionary tool for new emerging genotypes (Herzer *et al.*, 1990). An additional study applied the pathogenicity determinants of the ECOR strains and the diarrheagenic *E. coli* collection as an evolution indicator. The strains classified again into the five phylogenetic groups A, B1, B2, D and E with the tendency of virulence genes to clades B2 and D (Boyd and Hartl, 1998). A triplex PCR method was developed for a rapid and simple examination of different *E. coli* strains (Clermont *et al.*, 2000). It uses a combination of two genes (*chuA* and *yjaA*) and an anonymous DNA fragment on the basis that genes or fragments of DNA might be considered markers for certain phylogenetic clades. Clermont *et al.* (2000) analysed these markers for 230 strains which produced very similar results to those using MLEE.

This method attributed *E. coli* strains into four phylogenetic groups: A, B1, B2 and D, and strains of these groups differ in their virulence factors and environmental niche. The rationale behind choosing these genes argued that they showed no recombination through HGT (Clermont *et al.*, 2000).

A further study used the Clermont approach to assign *E. coli* strains isolated from humans, chicken, cows, goat, pigs and sheep. Researchers of this study found that clade B1 strains belong to cow, goat and sheep, whereas B2 strains were found only in human samples. They also found similar genetic content of strains isolated from humans and pigs and also between strains isolated from cows, goats and sheep (Carlos *et al.*, 2010). They suggested that this method therefore could be used as a bacterial source tracking tool around the world. The triplex PCR method was then modified by adding *arpA* to the previous two genes and the DNA fragment and became quadruplex PCR method (Clermont *et al.*, 2013). The new method has refined the classification of *E. coli* strains into eight groups by the addition of C, E, F and I to the previously mentioned ones (Clermont *et al.*, 2013).

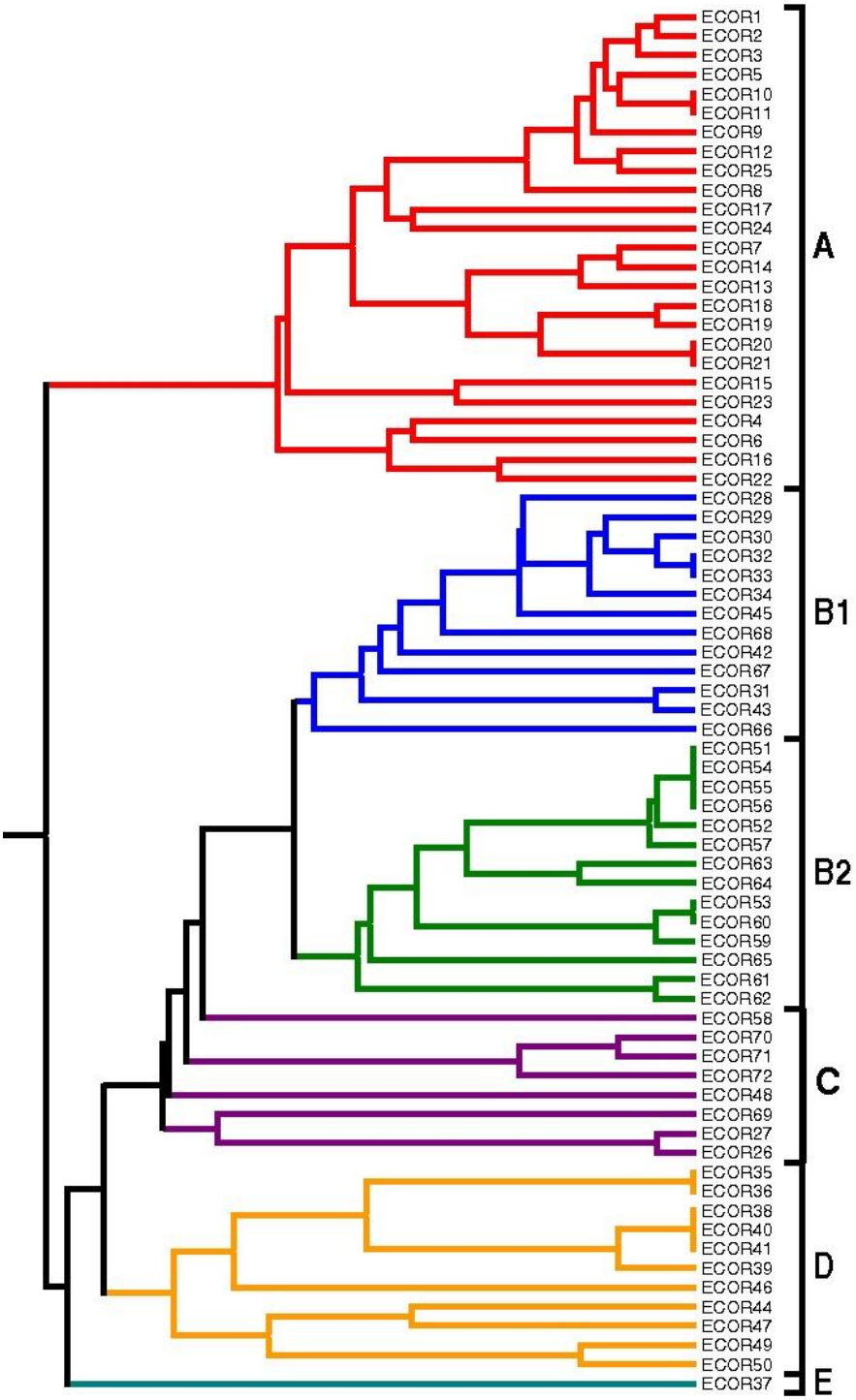


Figure. 1.2. Phylogenetic groups of the ECOR collection.

The ECOR collection which consists of 72 *E. coli* strains is classified into six phylogenetic groups: A, B1, B2, C, D and E based on the MLEE method. Figure taken from (Chaudhuri and Henderson, 2012).

1.2.3. Multilocus sequence typing

Multilocus sequence typing (MLST) is another method used to classify bacteria which was applied as an alternative to the previously described methods. It can unify strain data from all laboratories around the world and facilitate its exchange during epidemic diseases to help diagnosing pathogenic strains quickly (Maiden *et al.*, 1998). In addition to its importance in the epidemiological purposes, MLST proved as a reliable tool in evolutionary studies (Maiden, 2006). This method analyses a number of housekeeping genes and assigns allele numbers according to their DNA sequence. By combining allele numbers together, the strain can be allocated a Sequence Type number for identification and attributed to a phylogenetic group. MLST was first developed by Maiden *et al.* (1998) as a typing method for the human pathogen *Neisseria meningitidis* which causes meningitis and/or septicemia. They analysed 11 housekeeping genes of 107 isolates and managed to identify the major meningococcal lineages of this pathogen (Maiden *et al.*, 1998). The scheme of Maiden *et al.* (1998) was highly successful and, therefore, was accepted by a wide group of researchers working in the field of *Neisseria* (Belén *et al.*, 2009). Consequently, MLST has been applied to many bacterial species and eukaryotic organisms (Maiden, 2006).

Researchers in different laboratories have used different *E. coli* based MLST schemes (Reid *et al.*, 2000; Escobar-Paramo *et al.*, 2004; Turner *et al.*, 2006a; Wirth *et al.*, 2006). There are three well known *E. coli* MLST projects using three distinct housekeeping gene combinations. The first one hosted by Michigan State University in the US uses *aspC*, *clpX*, *fadD*, *icd*, *lysP*, *mdh* and *uidA* (Reid *et al.*, 2000). While the scheme hosted by Pasteur Institute in France uses *dinB*, *icd*, *pabB*, *polB*, *putP*, *trpA*, *trpB* and *uidA* (Jaureguy *et al.*, 2008). The last scheme is hosted by Warwick Medical School (UK) utilizes *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA* in *E. coli*

strains classification (Wirth *et al.*, 2006). The reason behind using these gene combinations is because they are housekeeping genes exhibiting the least frequency of HGT among different strains. McNally *et al.* (2013) analysed the genome of 62 *E. coli* strains available online using whole genome sequencing technique. To give the data a better resolution, they also applied MLST on these DNA sequences. They included 50 strains ranging from commensal to IPEC and ExPEC, and also 12 antibiotic resistant ST131 strains which gave this study a further dimension compared to previous researches (McNally *et al.*, 2013). The phylogenetic tree generated by McNally *et al.* (2013) still identified the six distinct groups: A, B1, B2, D, E and F (Figure. 1.3). The figure shows the known lab model strain MG1655 assigned to clade A, strain CFT073 which causes urinary tract infection (UTI) in clade B2 while EHEC in clade E. This work concluded that the emergence of pathogenic strains is due to reduction in genome recombination not as previously thought (McNally *et al.*, 2013).

Whole genome sequencing will eventually replace all other conventional techniques used to classify bacteria (Brodrick *et al.*, 2016). However, the data of whole genome analysis revealed a great match with MLST data hosted by Warwick Medical School (Sahl *et al.*, 2012) and, therefore, this technique will remain valid compared to others (Clermont *et al.*, 2015).

1.2.4. Model strains of E. coli

The use of comprehensive classification techniques and computer softwares such as Mugsy to align the genome sequences; Mothur to diagnose the core genome; RAxML to determine the phylogeny and BAPS was also used to determine the different groups of *E. coli* strains representing all the species culminated in drawing the phylogenetic tree with six distinct groups: A, B1, B2, D, E and F (Figure. 1.3) (McNally *et al.*, 2013). The figure shows the known lab model strain MG1655

assigned to clade A. The strain MG1655 is a K12 derivative (as it has a capsular polysaccharide K12 antigen) with known sequenced genome that reveals the least genetic manipulation (Blattner *et al.*, 1997). Additionally, the strain DH10B which is another derivative of K12 also belongs to clade A. It was constructed in the lab for DNA transformation purposes and maintenance of large plasmids (Durfee *et al.*, 2008). The genome of DH10B is altered by series of genetic recombination steps found to have 226 mutations due to the high DNA genetic manipulation. The high frequency of mutation rate is attributed to the activities of IS elements (Durfee *et al.*, 2008). Moreover, DH5 α is another strain belongs to clade A has *endA* mutation which makes it highly suitable for DNA transformation. The gene *endA* encodes endonuclease which degrades the DNA and reduces the transformation efficiency (Phue *et al.*, 2008). On the other hand, the strain CFT073 which causes urinary tract infection (UTI) is considered as UTI model strain belongs to clade B2. The complete genome sequence of CFT073 was revealed in 2002 (Welch *et al.*, 2002). Strain UTI89 is also a well-known UTI strain belonging to B2 with a genome sequencing completed in 2006 (Chen *et al.*, 2006). The strain EDL933 is associated with two focused outbreaks of haemorrhagic colitis infected 47 people in the US in 1982 (Riley *et al.*, 1983). The genome of this strain is completely sequenced (Perna *et al.*, 2001). This well-known strain belongs to clade E as it appears in figure 1.3.

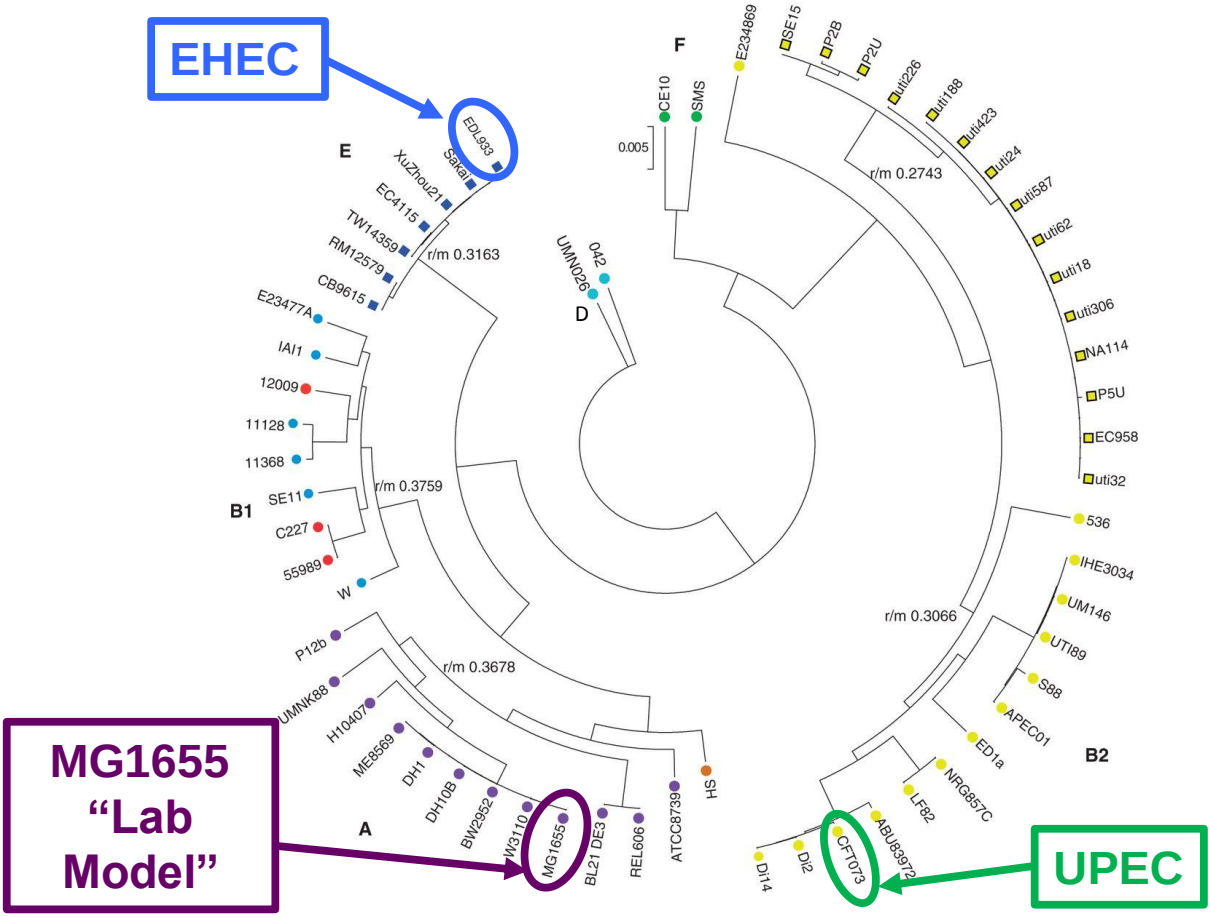


Figure. 1.3. The phylogenetic tree of *Escherichia coli* based on whole genome alignment.

The whole genome analysis of *E. coli* reveals six major groups: A, B1, B2, D, E and F. The well-known lab model strain MG1655 lies in clade A, whereas enterohaemorrhagic *E. coli* (EHEC)strain belong to clade E. Strain CFT073 is a model strain that causes urinary tract infections (UTIs) belonging to clade B2. Figure adapted from (McNally *et al.*, 2013).

1.3. Pathogenicity and virulence factors of *E. coli* pathovars

E. coli has many pathogenic strains that implicate in different illnesses to humans and animals. Pathovars can be divided into intestinal (IPEC) and extraintestinal (ExPEC). These two groups can be subdivided into enteroaggregative *E. coli* (EAEC), enterohaemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), diffusely adherent *E. coli* (DAEC), and adherent invasive *E. coli* (AIEC). While uropathogenic *E. coli* (UPEC), neonatal meningitis *E. coli* (NMEC), and avian pathogenic *E. coli* (APEC) (Crossman *et al.*, 2010; Croxen and Finlay, 2010; Croxen *et al.*, 2013). These pathovars possess many virulence factors including toxins (Table. 1.1).

1.3.1. Enteropathogenic E. coli (EPEC)

The strains of this pathovar cause attaching and effacing (A/E) lesions to the epithelial cells of the intestine leading to a potential deadly diarrhoea in infants in developing countries (Kaper *et al.*, 2004). A 35 kb PAI called locus of enterocyte effacement (LEE) encodes a type three secretion system (T3SS). T3SS is a nanomachine responsible for the attachment and translocation of the bacterial effector proteins which have different physiological roles into the host cells cytoplasm (McDaniel *et al.*, 1995; Dean and Kenny, 2009). EPEC uses this secretion system to translocate the intimin receptor (Tir) into the epithelial cell cytoplasm and then it is displayed on the surface facilitating the binding of bacterial outer-membrane protein intimin to it (Table. 1.1) (Kenny *et al.*, 1997). Tir has another function to suppress NF-kB to protect the bacterial cells from the immune response (Ruchaud-Sparagano *et al.*, 2011). Moreover, the EAF plasmid encodes bundle-forming pili which facilitates the attachment to the host cells (Hyland *et al.*, 2008). Additionally, EPEC expresses a surface protein called lymphocyte inhibitory factor (Efa1/LifA) which also helps in adherence (Badea *et al.*, 2003). Some strains of EPEC encodes 23 non-LEE effector

proteins participating in immune response inhibition (Iguchi *et al.*, 2009). EspC is another effector protein secreted by EPEC helps the bacteria to survive longer inside the host cells by conferring lysozyme resistance to bacterial cells among other functions (Table. 1.1) (Salinger *et al.*, 2009).

1.3.2. Enterohaemorrhagic E. coli (EHEC)

The bacterial cells infect the humans' ileum and large intestine leading to severe gastroenteritis outbreaks in developed countries. Infected individuals experience bloody diarrhoea and may develop haemolytic uraemic syndrome (Kaper *et al.*, 2004). The attachment of EHEC to the host cells is similar to EPEC in that it includes the binding of the outer-membrane protein intimin to its Tir receptors (DeVinney *et al.*, 1999). Moreover, the haemorrhagic coli pilus (type 4 pilus) contributes to the adhesion to the host cells (Xicohtencatl-Cortes *et al.*, 2009). Flagella are another virulence factor especially antigen 21 provide more support for bacterial adherence (Rogers *et al.*, 2012). Furthermore, the autotransporter protein Sab also aids the adhesion and biofilm formation of EHEC (Herold *et al.*, 2009). EHEC strain secretes Shiga toxin (Stx) which can be subdivided into Stx1 and Stx2 encoded by prophages (Table. 1.1) (Neely and Friedman, 1998). Stxs can bind to globotriaosylceramides (Gb3s) receptors on Paneth cells of the human intestinal mucosa and human kidney epithelial cells (Nataro and Kaper, 1998). In addition to its role in facilitating adhesion, Stx is also known to suppress the inflammatory response by inactivating NF-kB in Gb3 negative epithelial cells (Gobert *et al.*, 2007). Alternative to Stx, in Stx-negative strains the toxin hemolysin causes death to the host cells (Table. 1.1) (Aldick *et al.*, 2007).

1.3.3. Enterotoxigenic E. coli (ETEC)

This pathovar is considered as the most common cause of travellers' diarrhoea and lethal diarrhoea in piglets (Nataro and Kaper, 1998). ETEC attaches to the epithelial

cells of the small intestine by a number of colonization factors also called coli surface antigens especially CFA/II and CFA/IV expressed from virulence plasmids (Qadri *et al.*, 2005). Flagella with the secreted adhesin EtpA are also important in this process (Roy *et al.*, 2009). Moreover, the outer-membrane proteins Tia and TibA are other important factors in bacterial attachment to the host cells (Turner *et al.*, 2006b). TibA also has a role in autoaggregation of bacteria and biofilm formation in addition to its role in adhesion (Cote and Mourez, 2011) . Additionally, this pathovar secretes two types of toxins: heat-stable enterotoxin (ST) and heat-labile enterotoxin (LT) (Table. 1.1) which can function in different virulence pathways (Croxen and Finlay, 2010). Other virulence factors include EatA which is a serine protease autotransporter produced through type V bacterial secretion pathway that can degrade EtpA and delivers LT (Roy *et al.*, 2011). ETEC also secretes another toxin (CylA) which is a pore-forming cytotoxin, and *E. coli* ST1 (EAST1) which has similar function as ST (Turner *et al.*, 2006b). The three previously mentioned pathovars: EPEC, EHEC and ETEC share the same feature in that they adhere to the host cells rather than invade the epithelia.

1.3.4. Enteroinvasive E. coli (EIEC)

These bacteria share the same pathogenicity mechanisms with *Shigella*. This pathovar causes watery diarrhoea, dysentery and invasive inflammatory colitis (Kaper *et al.*, 2004). The bacterial cells infect the host cells via a 220 kb plasmid which has a Mxi-Spa locus expressing a T3SS (Schroeder and Hilbi, 2008). The bacteria invade macrophages and induce apoptosis to these cells (Schroeder and Hilbi, 2008). Then, it migrates again to submucosa with the aid of IpaC secreted through T3SS as it leads to Src tyrosine kinase dependent actin polymerisation resulting in host cell invasion (Table. 1.1) (Mounier *et al.*, 2009). IpaB and IpaBCD are translocators, encoded by the same virulence plasmid that bind to CD44 (Skoudy

et al., 2000) and $\alpha_5\beta_1$ integrin receptor (Watarai M *et al.*, 1996) respectively, to help bacteria attach to the host cells . Moreover, protein effectors such as OspB, OspI, OspH, OspZ and IpaH participate in the inhibition of the host immune responses (Ashida *et al.*, 2011). As these bacteria are nonmotile, the movement is achieved by the action of IcsA and the subsequent production of IcsA/N-WASP/ARP2/3 complex actin tail on EIEC cell surface (Agaisse, 2016). Then, the bacterial cell enters the adjacent host cell surrounded by the plasma membrane and translocates inside a vacuole for further movement (Agaisse, 2016). Additionally, EIEC produce *Shigella* enterotoxin 2 (ShET-2) (Table. 1.1) encoded from the bacterial chromosome which induces inflammation of the epithelial cells (Farfan *et al.*, 2011).

Factor	Pathotype	Toxin class	Target	Activity/Effect
Heat-labile enterotoxin (LT)	ETEC	AB subunit, type II effector	G_s	ADP ribosylates and activates adenylate cyclase resulting in ion secretion
Shiga toxin (Stx)	EHEC	AB subunit	rRNA	Depurinates rRNA, inhibiting protein synthesis; induces apoptosis
Cytolethal distending toxin (CDT)	Various	ABC subunit	DNA	DNaseI activity, blocks mitosis in G2/M phase
Shigella enterotoxin 1 (ShET1)	EAEC, EIEC*	AB subunit	–	Ion secretion
Urease	EHEC	ABC subunit	Urea	Cleaves urea to NH_3 and CO_2
EspC	EPEC	Autotransporter	?	Serine protease; ion secretion
EspP	EHEC	Autotransporter	?	Serine protease; cleaves coagulation factor V
Haemoglobin-binding protease (Tsh)	ExPEC, APEC	Autotransporter	Haem	Degrades haemoglobin to release haem/iron
Pet	EAEC	Autotransporter	Spectrin	Serine protease; ion secretion; cytotoxicity
Pic	UPEC, EAEC, EIEC*	Autotransporter	?	Protease, mucinase
Sat	UPEC	Autotransporter	?	Vacuolation
SepA	EIEC*	Autotransporter	?	Serine protease
SigA	EIEC*	Autotransporter	?	Ion secretion
Cycle-inhibiting factor (Cif)	EPEC, EHEC	Type III effector	?	Blocks mitosis in G2/M phase; results in inactivation of Cdk1
EspF	EPEC, EHEC	Type III effector	?	Opens tight junctions, induces apoptosis
EspH	EPEC, EHEC	Type III effector	?	Modulates filopodia and pedestal formation
Map	EPEC, EHEC	Type III effector	Mitochondria	Disrupts mitochondrial membrane potential
Tir	EPEC, EHEC	Type III effector	Nck	Nucleation of cytoskeletal proteins, loss of microvilli, GAP-like activity
IpaA	EIEC	Type III effector	Vinculin	Actin depolymerization
IpaB	EIEC	Type III effector	Caspase 1	Apoptosis, IL-1 release; membrane insertion
IpaC	EIEC	Type III effector	Actin	Actin polymerization, activation of Cdc42 and Rac
IpaH	EIEC	Type III effector	Nucleus	Modulates inflammation (?)
IpgD	EIEC	Type III effector	PtdIns $(4,5)P_2$	Inositol 4-phosphatase, membrane blebbing
VirA	EIEC	Type III effector	Tubulin	Microtubule destabilization, membrane ruffling
StcE	EHEC	Type II effector	C1-esterase inhibitor (C1-INH)	Cleaves C1-INH, disrupts complement cascade
HlyA	UPEC	RTX toxins	Erythrocytes, Leukocytes	Cell lysis
Ehx	EHEC	RTX toxins	Erythrocytes, Leukocytes	Cell lysis
Cytotoxic necrotizing factors (CNF-1,-2)	MNEC, UPEC, NTEC		RhoA, Cdc42, Rac	Altered cytoskeleton, necrosis
LifA/Efa	EPEC, EHEC		Lymphocytes	Inhibits lymphocyte activation, adhesion
Shigella enterotoxin 2 (ShET2)	EIEC, ETEC		?	Ion secretion
Heat-stable enterotoxin a (STa)	ETEC	Heat-stable enterotoxins	Guanylate cyclase	Activates guanylate cyclase resulting in ion secretion
Heat-stable enterotoxin b (STb)	ETEC	Heat-stable enterotoxins	?	Increase intracellular calcium resulting in ion secretion
EAST	Various	Heat-stable enterotoxins	Guanylate cyclase	Activates guanylate cyclase resulting in ion secretion

Table. 1.1. Toxins and their mode of action of *E. coli* pathovars, (Kaper *et al.*, 2004).

1.3.5. Enteroaggregative E. coli (EAEC)

These bacteria infect the mucosa layer of small and large intestine leading to watery, mucoid or bloody diarrhoea (Nataro and Kaper, 1998). The main virulence factor is encoded by a 100 kb pAA plasmid to form aggregative adherence fimbriae (AAFs) which belong to the Dr family of adhesins (Croxen and Finlay, 2010) as well as type IV pili encoded by a IncI1 plasmid (Dudley *et al.*, 2006). Some strains can also produce afimbrial adhesins such as aggregative protein 58 (Ap58) on the outer-membrane layer of the bacterial cells (Monteiro-Neto *et al.*, 2003). As a consequence, the bacterial cells aggregate and form biofilms on the surface of epithelial cells which have a stacked-brick appearance (Figure. 1.4) (Estrada-Garcia and Navarro-Garcia, 2012). The presence of EAEC virulence factors *fliC*, *aggR*, *aafA*, and *aap* enhances the levels inflammatory materials in faeces of humans with diarrhea (Mercado *et al.*, 2011). EAEC produce multiple types of toxins one of them called plasmid encoded toxin (Pet) (Table. 1.1) which causes host cell damage and exfoliation due to the breakdown of α-fodrin (Navarro-Garcia and Elias, 2011). The heat stable toxin EAST-1 and ShET-1 (Table. 1.1) are also produced by EAEC, yet their role in pathogenicity is unclear (Croxen and Finlay, 2010). Additionally, a protein involved in colonization (Pic) which has a mucolytic activity can aid the bacteria to penetrate biofilms mucus layer (Table. 1.1) (Henderson *et al.*, 1999). The protein AggR regulates AAF formation, dispersin protein and type VI secretion system and is positively regulated by FIS protein and negatively regulated by H-NS (Morin *et al.*, 2010).

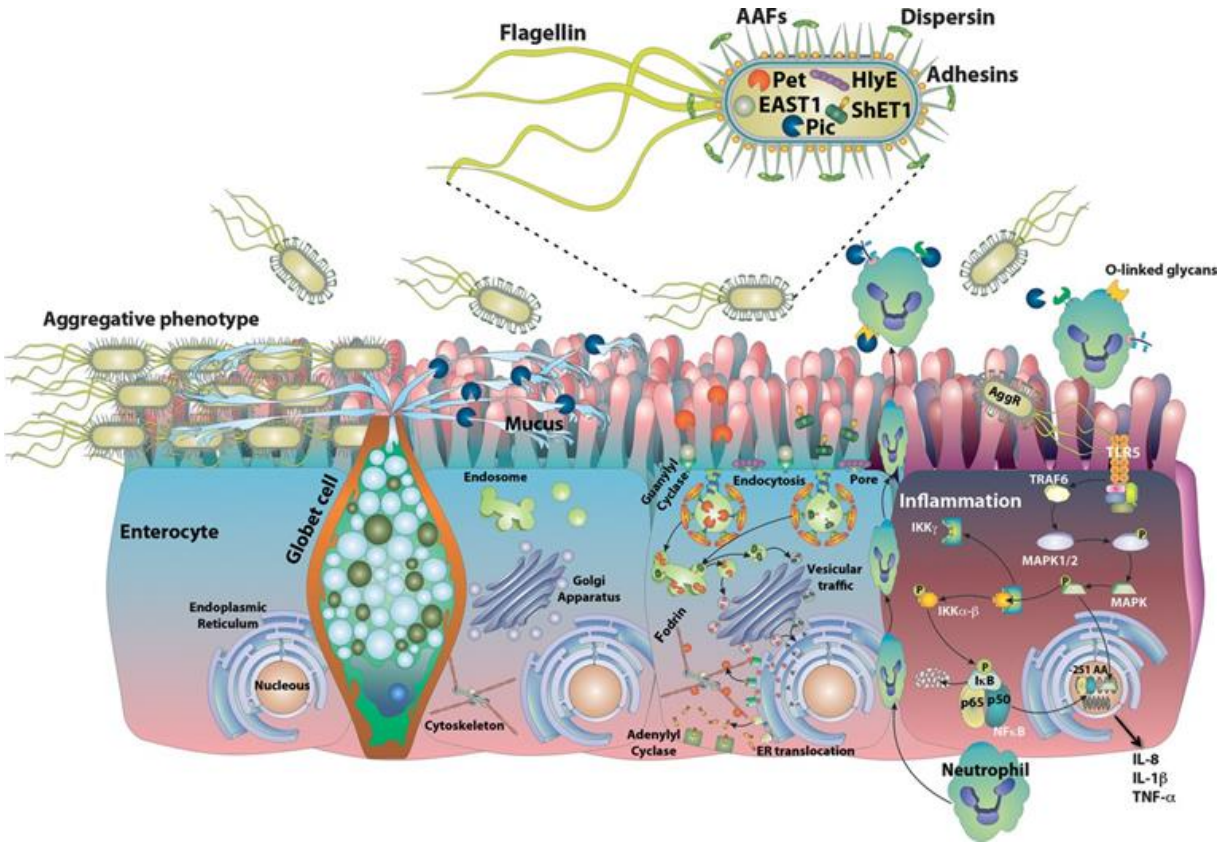


Figure. 1.4. Virulence factors and mode of infection of Enteroaggregative *E. coli*.

The drawing shows the bacterial infection of the mucosal layer. The cells adhere to the mucosa by secreting adhesins, and form stacked-brick biofilms. Toxins help the bacteria to invade the epithelial cells. This pathovar causes watery, mucoid or bloody diarrhea. Figure taken from (Estrada-Garcia and Navarro-Garcia, 2012).

1.3.6. Diffusely adherent E. coli (DAEC)

This pathogen infects the small intestine and can cause diarrhoea in children and recurrent UTI in adults (Nowicki *et al.*, 2001). The bacterial cells attach the host cells by the Afa-Dr group of adhesins such as fimbrial and afimbrial adhesins (Servin, 2005). These adhesins are able to adhere to host cells either through decay accelerating factors (DAF) receptors or through carcinoembryonic antigen-related cell adhesion molecule (CEACAM) receptors leading to changes in host cells pathways (Le Bouguenec and Servin, 2006). Dr adhesins can bind to type IV collagen to cause DAEC based chronic pyelonephritis (kidney infection) (Selvarangan *et al.*, 2004) . Alongside with these changes, secreted autotransporter toxin (Sat) which belong to SPATE family is also involved in the processes of bacterial uptake and survival inside host cells (Guignot *et al.*, 2007). The adherence to the host cells induces IL-8 production which in turn recruits other immune elements such as polymorphonuclear leukocytes (PMNs) to the site of infection leading to more DAF (CD55) expression (Betis *et al.*, 2003).

1.3.7. Adherent invasive E. coli (AIEC)

The bacterial cells can attach to the epithelial cells of the small intestine especially the ileum of Crohn patients by the means of type 1 pili to CEACAM6 receptors on the host cells (Barnich *et al.*, 2007). This adherence leads to the activation of interferon-γ and tumor necrosis factor-α which in turn leads to more CEACAM6 expression (Barnich *et al.*, 2007). Following adherence, AIEC invades epithelial cells by expressing outer membrane vesicles (OMVs) which translocate effector proteins such as OmpA that binds endoplasmic reticulum stress response chaperone Gp96 (Rolhion *et al.*, 2010). Furthermore, AIEC can migrate to lamina propria (beneath the epithelium) by synthesizing long polar fimbriae to interact with Peyer's patches, translocate across M cells (Chassaing *et al.*, 2011) and subsequently replicate inside

the phagolysosomes and macrophages throughout the expression of stress protein HtrA (Bringer *et al.*, 2005).

1.3.8. Neonatal meningitis E. coli (NMEC)

This pathovar inhabits the gastrointestinal tract, and is acquired perinatally from the mother (Croxen and Finlay, 2010). The bacterial cells then must enter the blood stream and survive inside the serum (Kaper *et al.*, 2004). NMEC has a capsule (especially K1) that consists mainly of sialic acid which prevents phagocytosis (Wooster *et al.*, 2006). It also has OmpA which protects the bacteria from neutrophils (Belaaouaj *et al.*, 2000), and also attaches the bacteria to the blood brain barrier (Kim, 2008). The adherence of the bacteria to the blood brain barrier is also achieved by type 1 fimbriae (FimH adhesin) binding to glycosylphosphatidylinositol-anchored receptor (CD48) (Khan *et al.*, 2007). The invasion of the central nervous system is mediated by IbeABC proteins, type 1 pili, OmpA and cytotoxic necrotizing factor 1 (Kim, 2008). The outcome of these mechanisms is oedema, inflammation and neural damage (Croxen and Finlay, 2010).

1.3.9. Uropathogenic E. coli (UPEC)

UPEC is the most common causative agent of UTIs (Kaper *et al.*, 2004). In the past, research was performed on the virulence factors of UTI isolates in order to find a vaccine, they were concentrated on the lipopolysaccharide (LPS) side chain (O) antigen (Uehling and Wolf, 1969). Antibodies were shown to have antiadhesive power against these antigens (Lahiri *et al.*, 2008). However, heterogeneity is the main problem that faces the researchers who try to develop a vaccine (Sivick and Mobley, 2010). P fimbriae have an adherence role in the colonization of bacteria in kidneys of mice and humans (O'Hanley *et al.*, 1985). On the other hand, type1 fimbria is another virulence factor that enables the pathogen to colonize the bladder (Gunther *et al.*, 2002). Vaccinations administered to animals made by using different

parts of type1 fimbria revealed increased levels of specific antibodies and decreased colonization (Poggio *et al.*, 2006). Many Gram-negative pathogens such as *E. coli*, *Klebsiella* spp., *Proteus* spp., *Pseudomonas* spp., *Haemophilus* spp., *Salmonella* spp., and *Yersinia* spp., have a group of adhesive fibres named type 1 pili or chaperone-usher pathway (CUP) pili (Wurpel *et al.*, 2013). *E. coli* as a species has 38 CUP pilus operons, and every UPEC strains can potentially produce up to twelve CUP pili or more (Wurpel *et al.*, 2013). Different CUP pili produced by UPEC have different adhesins at their tips leading to tissue and epithelial cells tropism (Wright and Hultgren, 2006). For example, type1 pili which have the adhesin FimH can adhere to specific cell structures defined as mannosylated uroplakins and $\alpha_3\beta_1$ integrins mediating the subsequent invasion of umbrella cells (Hannan *et al.*, 2012). The adhesion of these pili to umbrella cells promotes a series of activities ending with the uptake of pathogens by the host cells (Martinez and Hultgren, 2002). This mechanism protects the pathogens from the immune system and treatment with antibiotics. Another type of pili called P pili which has the adhesin PapG can adhere to human kidney cells facilitating their invasion (Wright and Hultgren, 2006). Pathogens can secrete several proteins and toxins that kill the bladder epithelial host cells to free their food. For example, UPEC produce HlyA that can fuse to regions in the umbrella host cell membranes causing their lysis, and bacteria can use iron and nutrients for their survival. The HlyA also causes epithelial cells to detach and float in the urinary tract facilitating the infection of new cells (Dhakal and Mulvey, 2012). The toxin and P fimbriae are necessary to colonize the kidney (O'Hanley *et al.*, 1991). Some UPEC isolates also have the ability to evade the immune system by synthesizing TcpC. This protein is homologous to part of toll-like receptor 1 (TLR1) that fuses with MyD88 blocking cytokine production (Cirl *et al.*, 2008). Another virulence factor CNF-1 (Table. 1.1) produced by UPEC has a role in pathogen

endocytosis into epithelial host cells (Garcia *et al.*, 2013). Moreover, siderophore systems which are molecules that bind to free or host bound iron, can be used by uropathogens to gain iron from the very iron limited region of bladder (Caza and Kronstad, 2013). Aerobactin and yersiniabactin are two siderophore systems secreted by UPEC important in UTI (Garcia *et al.*, 2011). In contrast, the innate immunity has an expulsion mechanism which relies on TLR4 produced by the host cells. Cyclic AMP (cAMP) is produced following the activation of adenylyl cyclase3 (AC3) by TLR4. The cAMP catalyses the expulsion of the pathogens outside the epithelial cells (Song *et al.*, 2009).

Strong evidence has shown the role of flagella of UPEC in the ascension process of the pathogen to the upper parts of the urinary tract (Lane *et al.*, 2007). The ascending process of pathogens from the periurethral area to urethra and then to the bladder results in the infection of the bladder (cystitis). Moreover, bacteria may ascend to the kidneys by the aid of flagella, produce toxins leading to tissue damage causing pyelonephritis. If left without treatment, the bacteria can cross the epithelial barrier to the blood causing bacteraemia (Figure. 1.5) (Flores-Mireles *et al.*, 2015). After the ascension, bacteria can invade the apical cells of the bladder leading to the formation of intracellular bacterial communities (IBCs) which have the same characters of biofilm (Anderson *et al.*, 2003). These aggregations prevent bacteria from being recognized by the host immune responses, and also protect the pathogens from medications and urination (Mulvey *et al.*, 2001). Researchers succeeded in viewing the IBCs dissemination using a high-resolution time lapse video microscopy of mice bladders (Justice *et al.*, 2004). The phenomenon of UPEC motility was studied by using UPEC isolates producing green fluorescent protein (GFP) when the dissemination of bacteria from mature IBCs was detected after 16 hours of infection (Wright *et al.*, 2005). In mice, many experiments performed on the

wild-type (WT) and *ΔfliC* UPEC strains mutants using reporter fusions have been shown that there is a shut down in the flagellar gene expression during the IBC formation, maturation, and dispersal. As these studies showed that there was no function for flagella in IBCs different stages, they introduced evidences on the importance of flagella in the fitness by the increase of the infection period two more weeks compared to *ΔfliC* mutants (Wright *et al.*, 2005). Another study has also demonstrated enhanced fitness for the wild-type at 3 days postinfection compared to *ΔfliC* mutants of another UPEC strains (Lane *et al.*, 2005).

In addition to various virulence factors listed above, the emergence of antibiotic resistant bacteria enhanced the fitness of these different pathovars to cause infections. Cephalosporins, fluoroquinolones and trimethoprim-sulfamethoxazole are common antibiotics used to treat *E. coli* UTIs (Foxman, 2010). However, the increase in *E. coli* strains encoding β-lactamases complicated the treatment. These strains can produce plasmid encoded AmpC β-lactamases such as CMY types, ESBL such as CTX-M types and KPC carbapenemases (Livermore and Woodford, 2006). The *E. coli* ST131 which belong to phylogenetic group B2 is a well-known example of an antibiotic resistance strain (McNally *et al.*, 2013). ST131 can carry CTX-M-15 which enables it to persist and cause serious complications including septicemia when treatment involves extended spectrum β-lactams such as cefalexin (Peirano and Pitout, 2010).

To sum up, pathogenic *E. coli* strains have multiple virulence factors that increase the fitness of the bacteria to cause infections. Moreover, the continuous emergence of antibiotic resistance strains makes it difficult to find the effective treatments. The emergence of pathogenic bacteria is not inactive, but rather a dynamic process throughout the species driven by the gain or loss of new genetic materials as in the case of antibiotic resistant strains.

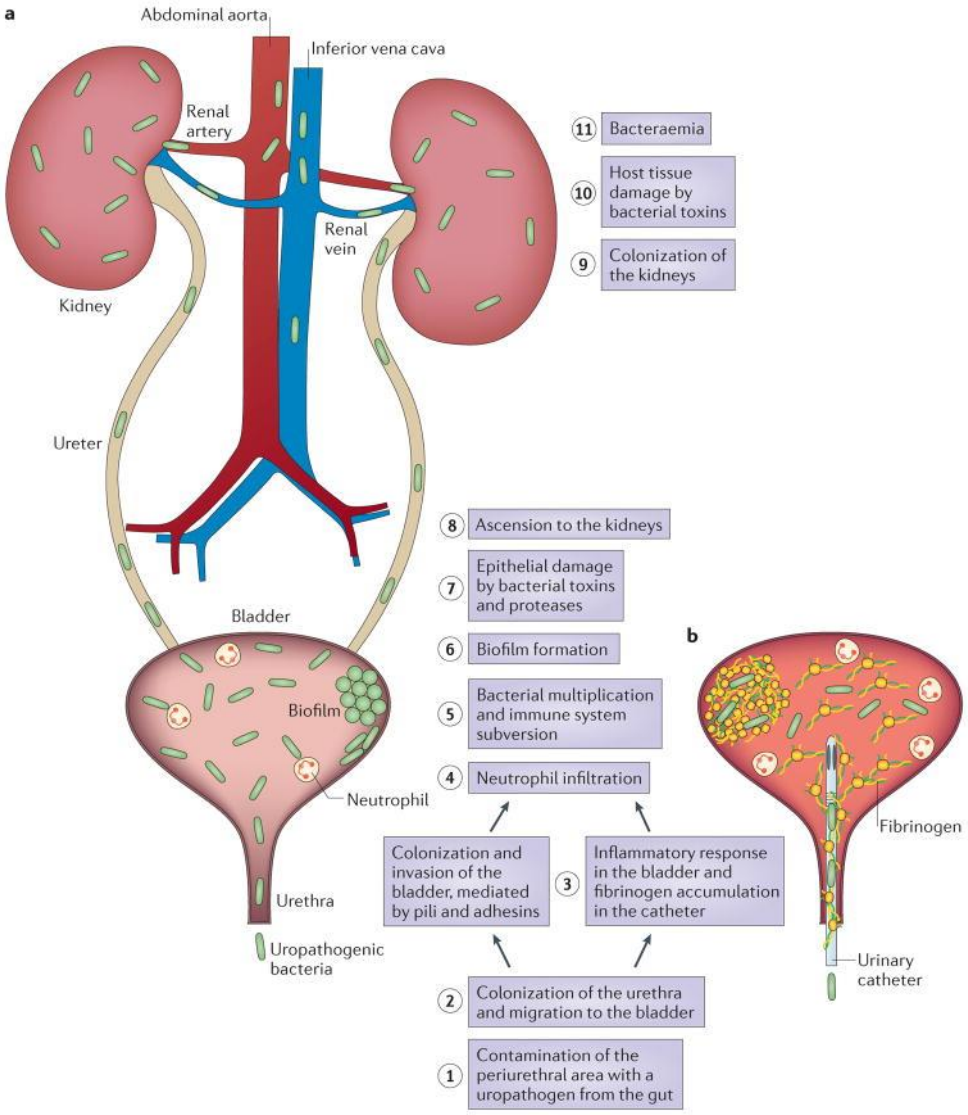


Figure. 1.5. Pathogenecity of urinary tract infection by uropathogenic *E. coli*.

a) The cells migrate through the urethra to the bladder and adhere to the epithelial cells causing cystitis. The bacteria can also ascend to the kidney causing pyelonephritis. If left untreated, the cells can enter the blood and cause septicaemia. **b)** The bladder could also infected by contaminated catheters. In this case, the cells will follow the same steps of pathogenicity. Figure taken from (Flores-Mireles *et al.*, 2015).

1.4. Key methods in bacterial recombineering

Many revolutionary techniques have been developed to manipulate DNA in specific ways. These techniques allow researchers to avoid the need of using random mutagenesis and the subsequent complicated screen procedures to identify mutants. They also reduced the requirement of molecular cloning steps.

1.4.1. Datsenko and Wanner

Free DNA uptake by bacterial cells and the subsequent recombination is difficult due to the presence of intracellular DNAses which degrade linear DNA fragments (Lorenz and Wackernagel, 1994). Therefore, λ phages are successful in recombining in bacterial chromosome as they possess the λ red system (Murphy, 1998). This system includes Gam protein which can bind to bacterial RecBCD which inhibits its exonucleases and endonucleases enzymatic activities (Murphy, 1991). While the other two components (Bet protein and Exo nuclease) facilitate DNA recombination (Muniyappa and Radding, 1986). Thus, in order to promote DNA recombination, the λ red system should be available inside the bacterial cells. The Datsenko and Wanner method is a developed way of introducing recombination modified from Murphy (1998). The latter method used multicopy plasmids that carry λ red system alongside with PCR amplified DNA fragments (Murphy, 1998). However, the use of these multicopy plasmids can reduce the rate of recombinant cells as they might act as competitive inhibitors (Datsenko and Wanner, 2000). This method is using low copy number plasmids that carry λ red recombinase system under arabinose inducible promoter (Figure. 1.6).

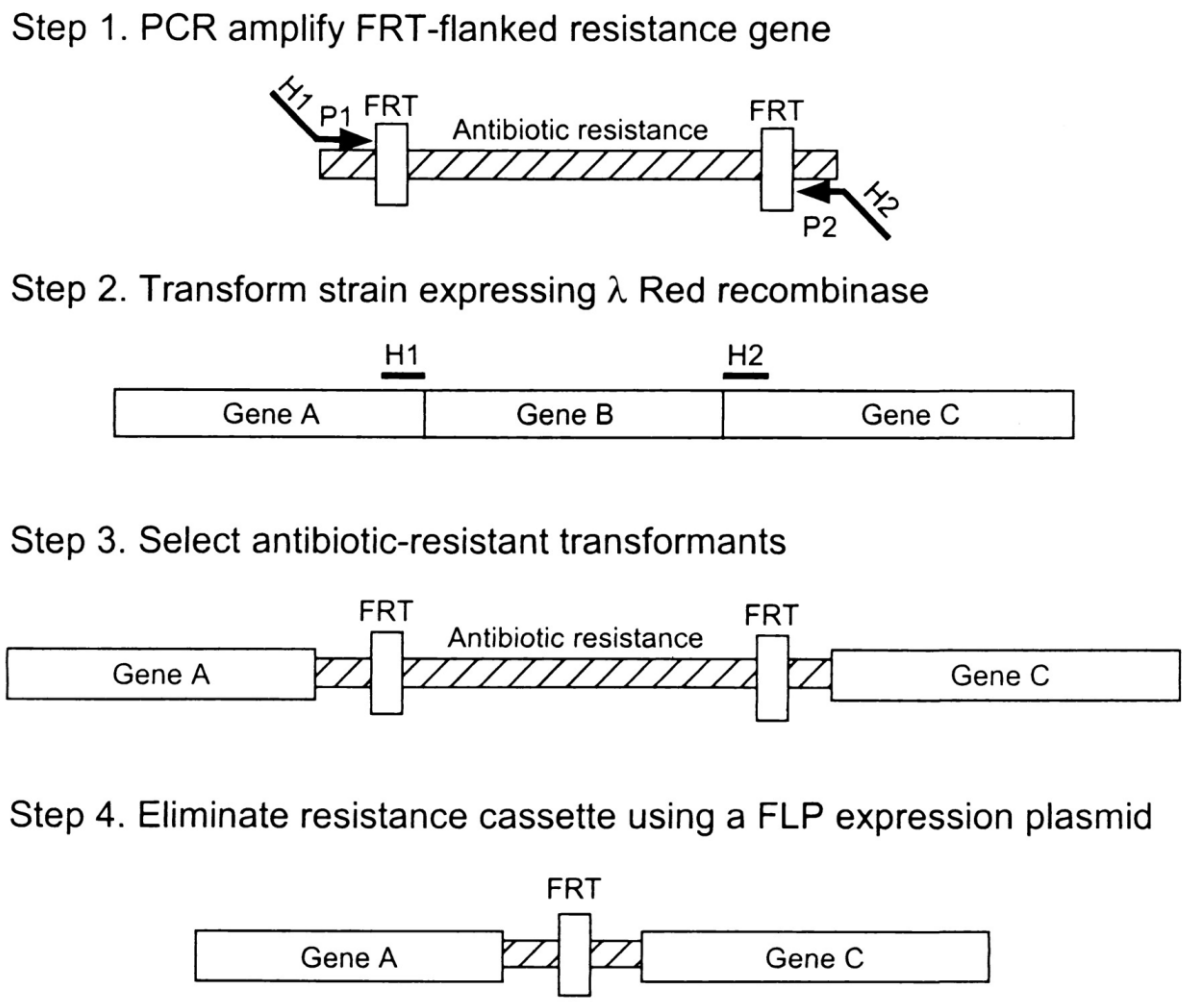


Figure. 1.6. Datsenko and Wanner method of gene removal.

The method includes the amplification of a DNA fragment that carry an antibiotic cassette. This fragment which has a sequence homology (H1 and H2) to the gene of interest will then transform to the bacterial cells. With the aid of λ red expressing plasmid, the antibiotic cassette will replace the gene of interest. Cells will be selected on antibiotic selective media. P1 and P2 are primers start sites. Figure taken from (Datsenko and Wanner, 2000).

1.4.2. Blank method

The Datsenko and Wanner method could be followed by an extra step to insert a DNA fragment into the bacterial chromosome to study gene function and expression. Thus, studies continued to find a proper way to achieve this. The first attempt is included the replacement of genes by electroporating a PCR DNA fragment with an antibiotic resistance gene into cells expressing λ red from plasmid pKD46 (Karlinsey and Hughes, 2006). The second use of this method is performed by introducing a tetracycline resistance gene DNA fragment into the bacterial cells. Then, this fragment was replaced by electroporating mutagenic PCR products into cells expressing λ red (pKD46) to introduce point mutations into a given gene (Aldridge *et al.*, 2006c). A study for Gerlach *et al.* (2009) was able to perform this by using λ red recombinase system in a two-step method. The first step is to insert a tetracycline resistant cassette into the required locus on the chromosome. The second step is to replace this cassette by an oligonucleotide using the λ red system for both steps (Gerlach *et al.*, 2009). However, the limitation in the successful selection of Tet^s clones on Bochner-Maloy plates in this method was the cause for further investigations and improvements (Blank *et al.*, 2011). The method of Blank had developed the selection parameters by adding I-SceI recognition site to the template vector pKD3 to become pWRG100. Moreover, the plasmid pKD46 was modified to include I-SceI endonuclease to become pWRG99 (Blank *et al.*, 2011).

The Datsenko and Wanner method has a disadvantage in that it leaves a genetic scar. Thus, Gerlach, Blank and others including our lab have modified the Datsenko and Wanner method to generate scarless mutants. The first step of scarless targeted mutagenesis of bacterial chromosome includes the amplification of DNA fragment from plasmid pWRG100 that contains chloramphenicol resistance gene with I-SceI site. This fragment is introduced to the bacterial cells to target the

gene of interest with the help of pKD46. The second step includes the replacement of this cassette with a DNA fragment that carry the required gene with the help of pWRG99 that has the λ red system that mediates the recombination. This plasmid also has I-SceI endonuclease activity that recognizes I-SceI sites and selects for successful recombinants (Figure. 1.7).

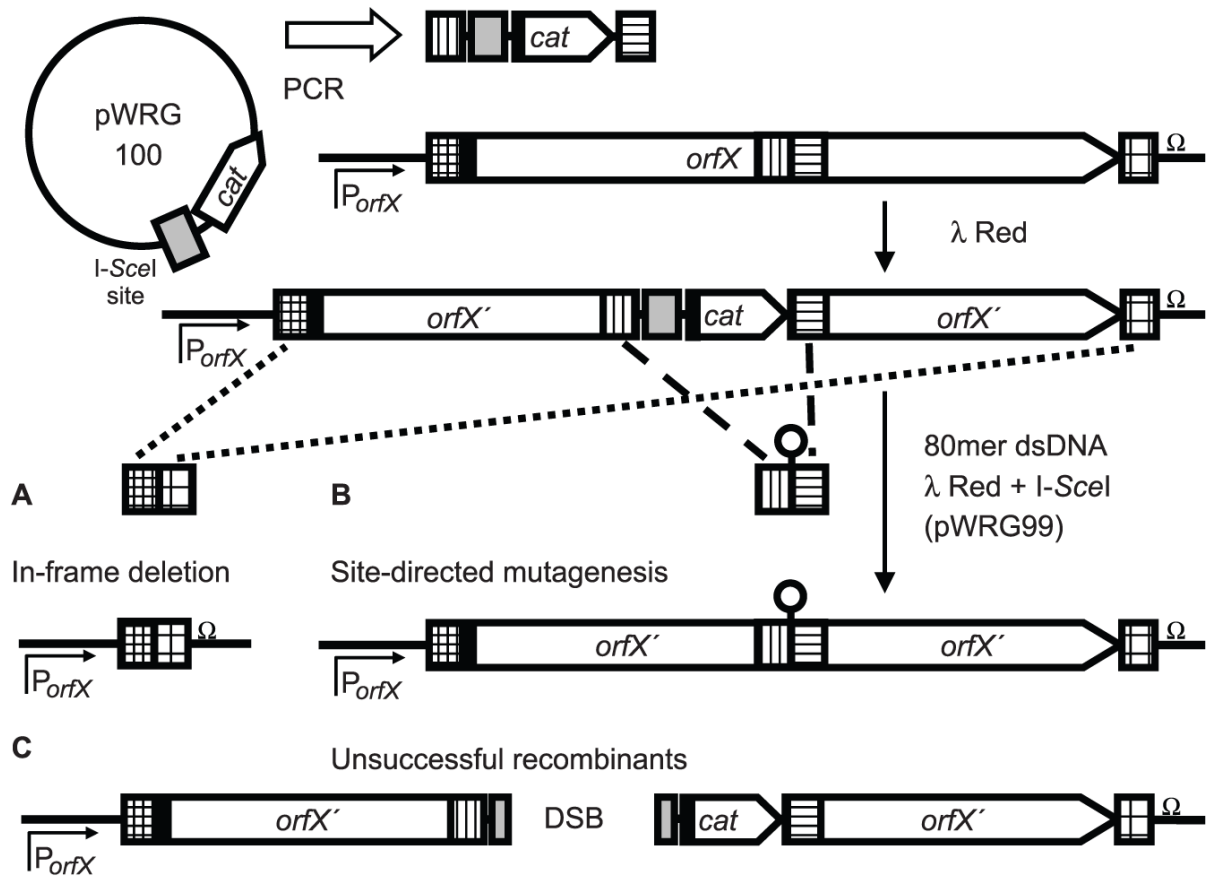


Figure. 1.7. Blank method of DNA recombineering.

This method includes the amplification of an I-SceI site together with a chloramphenicol resistance cassette (*cat*) from the plasmid pWRG100. This fragment is introduced to the cells expressing λ red to replace the gene of interest. The second step includes the replacement of the antibiotic resistance gene with the DNA fragment that carry the required gene. This step is mediated by the plasmid pWRG99 which carry the λ red system and the I-SceI recognizing endonuclease. This method could be used to create **A)** In-frame deletion. **B)** Site-directed mutagenesis. Or DNA insertion within the chromosome (not shown). **C)** Unsuccessful recombination due to the recognition of the I-SceI site by the endonuclease which lead to a double strand break (DSB). Figure taken from (Blank *et al.*, 2011).

1.4.3. CRISPR-Cas9 system

Clustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats technique is considered as a key innovation in bacterial genetics. This natural system found in prokaryotic genomes is a means of adaptive immune response against the invasion of bacteriophages and other foreign DNAs. Spacers from phage DNA integrate into the bacterial chromosome, and the CRISPR-Cas system then recognizes these DNA pieces and provides resistance specific to this type of phage (Barrangou *et al.*, 2007). CRISPR systems come in different formats, however, for this study two plasmids, pCas and pTarget will be used. pCas carries the gene encoding Cas9, *repA101* for self-curing from pKD46, λ red genes which are arabinose inducible and kanamycin resistance gene. While pTarget bears a single synthetic guide RNA (sgRNA) which consists of CRISPR RNA (crRNA) and trans activating crRNA (tracrRNA) and spectinomycin resistance gene (Jiang *et al.*, 2015). The first step is to transform pCas into the cells, then, pTarget is also transformed alongside with synthetic DNA of interest. The previously designed sgRNA that have complimentary fragment will bind to Cas9 and guide it to the target DNA. Then, the Cas9 protein which has endonuclease activity will cause a double strand break excising this DNA and allowing to the synthetic DNA to recombine in (Figure. 1.8).

In 2012, CRISPR was first mentioned as a potential programmed method for genome editing of any organism including humans (Jinek *et al.*, 2012). Since then, this highly efficient method has been distributed widely across the world as it equals the invention of PCR in 1985 (Ledford, 2015). This method is unique compared to other known ways of gene editing in that it can perform multiple gene editing compared to one at a time achieved by other traditional methods including the Datsenko and Wanner method and Blank method. Another important feature of this

system is the flexibility of targeting any DNA sequence required to be replaced (Jiang *et al.*, 2015).

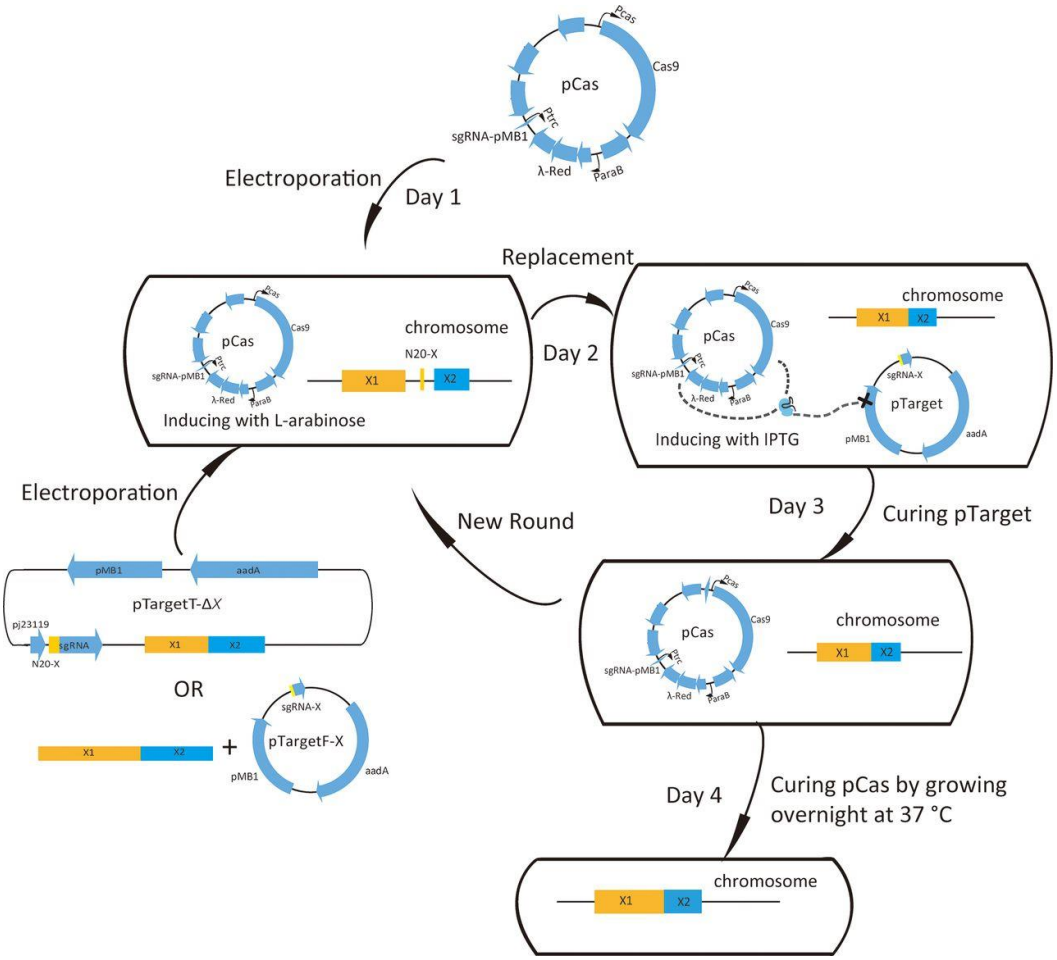


Figure. 1.8. Schematic representation of genome editing using CRISPR-Cas system.

This state of art technique in genetic recombineering uses a two-plasmid system for gene removal and addition at the same time. The plasmid pCas which carry the λ red system and endonuclease is electroporated into the competent cells. Another plasmid pTarget which has the single synthetic guide RNA (sgRNA) is also electroporated into the same cells together with an amplified fragment of DNA. The sgRNA will bind to Cas endonuclease and guide it to the target DNA. Then, Cas will excise the DNA and allow the amplified DNA to recombine. Figure taken from (Jiang *et al.*, 2015).

1.4.4. Bioluminescence and fluorescent reporters

Bioluminescence is a chemical reaction that can be performed by a wide range of organisms both prokaryotes and eukaryotes. It functions in the protection from predators, for predation, and also for communication. The luciferases are enzymes working on luciferins are responsible for this phenomenon. Importantly, the high sensitivity and the easy dealing with the *lux* genes encouraged researchers to harness bioluminescence in the field of molecular biology to quantify gene and protein expression in real-time (Noguchi and Golden, 2017). Moreover, fluorescent protein fusions are indispensable tools in the localization of protein molecules in the subcellular level.

Many living organisms around the world such as dinoflagellates, bacteria, fungi, fish, insects, shrimp, and squid can emit light as a result of the enzyme luciferase interacting on a substrate called luciferins (Meighen, 1991). Moreover, strains belong to more than thirty species of bacteria have *lux* genes responsible for the bioluminescence. These Gram-negative bacteria are members of three families of Gammaproteobacteria: Vibrionaceae (*Aliivibrio*, *Photobacterium*, and *Vibrio*), Enterobacteriaceae (*Photorhabdus*), and Shewanellaceae (*Shewanella*) (Dunlap, 2014). The bioluminescence is a phenomenon used by these organisms as a defense mechanism, for predation, or communication (Hastings *et al.*, 1985). Fish, squids and other animals emit light by their own or as a result of endosymbiosis with certain organisms in a phenomenon called counterillumination (Haddock *et al.*, 2010). Insects like the firefly were studied for bioluminescence, and luciferase was isolated (Marques and Esteves da Silva, 2009). In Dinoflagellates there is a decrease in the pH following stimuli or stress that free the lucifirin binding protein (LBP) and luminescence is triggered to threaten the predators (Valiadi and Iglesias-Rodriguez, 2013). In luminous bacteria, bioluminescence plays an important role when the

oxygen level is low. Bioluminescence has been thought to be emerged in bacterial evolution as an oxygen coping strategy (Dunlap, 2014). In agreement with that, luciferase as an oxidase may act as a secondary respiratory chain and bacteria can resist anaerobic environments (Dunlap, 2014). Additionally, bioluminescence is also important in the dissemination of bacteria through the attraction of the predators to feed on the dead animals that have endosymbiosis relationship with luminous bacteria (Dunlap, 2014). The chemical reaction of bioluminescence in bacteria is induced by luciferase which is heterodimeric protein of 80 KDa in the molecular weight. It consists of two subunits $\boldsymbol{\alpha}$ (40 KDa) and β (37 KDa), and catalyses the oxidation of reduced Flavin mononucleotide ($FMNH_2$) and the aldehyde (RCHO) by the oxygen to emit the light.

$$FMNH_2 + O_2 + RCHO \longrightarrow FMN + H_2O + RCOOH + light\ (490nm)$$

On the other hand, eukaryotes have different luciferases and alternative chemical reactions. The interaction of $FMNH_2$ with luciferase and oxygen will form Flavin-4a-hydroperoxide which-in turn- binds to aldehyde. Then, the oxidation of the $FMNH_2$ and aldehyde will occur producing the light (Hastings *et al.*, 1985). As a result of easy dealing and high sensitivity of light emission detection, researchers performed a wide range of experiments in the field of molecular biology using the *lux* system. They used the *lux* genes system as a reporter for gene expression and gene regulation, applications related to cloned DNA, and as detectors for metabolic processes. Different types of promoters can be cloned to the *lux* genes, and their strength and transcriptional activities can be detected by the light emission (Meighen, 1991). The study of bacterial infection and pathogenicity inside live animals as well as host cells response is another application of luciferase reporters (Waidmann *et al.*, 2011). In 1998, the first *luxCDABE* operon from the bacterium *Photorhabdus luminescens* that quantifies temporal gene expression was cloned into different

vectors (Winson *et al.*, 1998a; Winson *et al.*, 1998b). These vectors have different replicons to make them suitable to use with a wide range of cells as well as antibiotic resistance genes and restriction sites to ease their use. The *luxCDE* are essential for the synthesis of the substrate luciferin for the enzyme luciferase that is expressed from *luxAB* (Craney *et al.*, 2007). However, due to their brightness limit and detection methods luciferase reporters are inefficient in protein localisation studies (Noguchi and Golden, 2017).

Unlike bioluminescence, fluorescent proteins are other tools in molecular biology to detect gene expression and protein synthesis in parts of the cell. They are different from bioluminescent reporters in that they need an external light source for protein excitation and subsequent emission of fluorescence (Noguchi and Golden, 2017). Green fluorescent protein was extracted and purified from the jellyfish *Aequorea victoria* (Shimomura *et al.*, 1962). Since then, efforts continued to increase the resolution of these proteins in the field of molecular biology and other derivatives such as blue, red, cyan and yellow were synthesized (Shaner *et al.*, 2007). The yellow fluorescent protein for energy transfer (YPet) proved to be an example of a highly efficient glowing variant that can be used as fusion with the gene of interest (Day and Davidson, 2009). Instead of just detecting activity, fluorescent proteins could visualize their location. This has led to the major shift in appreciation of bacterial cell biology. They can be visualized easily, their detection is easier than detecting bioluminescence and we can see their location inside the cell (Delalez *et al.*, 2010).

1.5. Movement in bacteria

The movement of bacterial cells in different liquid and semi liquid environments is crucial for the survival of these cells. It aids the bacteria to translocate from harmful environments to other favorable conditions. By this means, the cells may trace

nutrients, infect host cells, drive cell to cell communication and instigate biofilm formation. However, not all bacterial species are motile.

1.5.1. Flagella independent motility

Bacterial motility without flagella has been extensively studied, and researchers defined three major types of flagella independent motility: twitching, gliding and sliding. All three types are systems that drive movement on a surface. Twitching motility includes irregular motility with a jerky or twitchy appearance with a speed rate of 0.06-0.3 $\mu m.s^{-1}$. This motility occurs by the type IV pili which are 6 nm in diameter and 4 µm in length (Harshey, 2003). Many species extend their type IV pili from the cell pole away from the cell body to the semi solid surface, adhere to the surface and then retract pulling the cell forward (Skerker and Berg, 2001). Type IV pili play other important roles during infection of host cells, biofilm formation and fruiting body development (Mattick, 2002). Around 40 genes have been found in *Pseudomonas aeruginosa* responsible for the assembly and regulation of type IV pili. Controlled by chemosensory system (Mattick, 2002), the retraction force exceeds 100 piconewton and is dependent widely on the amount of PilT motor protein (Ghosh *et al.*, 2014).

Gliding is another means of smooth movement by many bacterial species with a speed rate of 0.025-10 $\mu m.s^{-1}$ (Harshey, 2003). Although many mechanisms have been suggested for this type of motility, the exact system involved is not fully understood in different species of bacteria (Nan and Zusman, 2016). However, the focal-adhesion systems theory is still a prominent model (Mignot, 2007). This model involves the secretion of multiple types of proteins and lipoproteins to adhere the cell into the surface. Then, motor molecules distributed around the cell cytoskeleton push against the adhesion complexes pulling the cell forward. Yet another surface motility type, is sliding or spreading motility which has a speed rate of 0.03-6 $\mu m.s^{-1}$ (Harshey, 2003). This type, unlike other types of motility, does not need any special

compound or motor, but it relies on the growth of the cells which results in pushing the adjacent cells away. Surfactants produced by cells reduce the friction between the cells and the surface to facilitate this type of motility (Henrichsen, 1972).

1.5.2. Flagella dependent motility

Bacterial cells can move with the aid of flagella operated by a motor at the base. They either swim in the liquid media or swarm on the surfaces. Swimming of individual cells requires a motor which can rotate the flagellum counterclockwise forming a flagellar bundle and push the cell forward at a speed rate between 10-90 $\mu m.s^{-1}$. In contrast, when the flagellum rotates in a clockwise position, it reduces the speed of the cell and makes the cell tumble and change orientation (Darnton *et al.*, 2010). The switching between these two modes is driven by the chemotaxis system which senses the level of attractants or repellents in the environment changing the phosphorylation state of the response regulator protein, CheY (Darnton *et al.*, 2010). In *E. coli*, this can be detected by transmembrane receptors (methyl-accepting chemotaxis proteins) such as Tsr, Tar, Trg and Tap (Burkart *et al.*, 1998). Unlike swimming, swarming is a mode of bacterial motility that includes the migration of crowds of cells across a slippy surface. Swarming cells synthesize more flagella than swimming cells and become elongated. Although swarming cells form more flagella than swimmers, swarming speed rate is lower than that of swimming with 2-10 $\mu m.s^{-1}$ probably due to cell-surface tension (Harshey, 2003). Only Tsr and Tar are found to induce cells to swarm in *E. coli* (Burkart *et al.*, 1998).

1.6. The structure and assembly of the flagellum

More than fifty genes are responsible for flagellar assembly and protein synthesis necessary for regulatory functions distributed among 17 operons (Chilcott and Hughes, 2000). Moreover, the flagellum is a complex molecular nanomachine consisting of about 25 different protein subunits (Macnab, 2003). The key

components of the flagellum include the basal body, motor, flagellar protein secretion apparatus, hook, a junction zone and the filament (Figure. 1.9) (Macnab, 2004).

1.6.1. The basal body

The basal body is a passive component in that it does not rotate by itself, but it rather rotates as a result of the motor rotation. It transfers the torque from the motor to the hook and filament (Macnab, 2003). The basal body consists of the MS ring which integrates into the cytoplasmic membrane as seen by the electron microscope (Katayama *et al.*, 1996). The MS ring is synthesized from *fliF.* It plays an important role in supporting the rotor switch, connects the motor to the hook and filament (Figure. 1.9) (Macnab, 2003). Then, FliE will form on the MS ring as a rod junction protein (Minamino *et al.*, 2000b). The second component of the basal body is the rod which conveys the motor rotation to the filament. The rod is synthesized by four protein subunits belonging to the same family: FlgB, FlgC, FlgF and FlgG (Homma *et al.*, 1990). As the rod must grow through the peptidoglycan layer, there should be something facilitating this process. It has been demonstrated that FlgJ is a bifunctional protein, the C terminus has a muramidase activity making a hole in the peptidoglycan layer allowing rod growth. At the same time, the N terminus of FlgJ from position 1 to 151 acts as a cap aiding in the folding of the rod structural subunits (Hirano *et al.*, 2001). However, the mechanism which controls the length of the rod is still unclear (Macnab, 2003). In addition to the MS ring and the rod, the basal body also consists of the L-ring and P-ring (Figure. 1.9). The two substrates FlgI and FlgH build the P-ring and the L-ring, respectively. FlgI and FlgH are exported to the periplasm by the Sec secretion system as they are assembled around the rod (Macnab, 2003). FlgI is located in the periplasmic space which is known to contain proteases. Therefore, to protect FlgI from the degradation and to aid in its polymerization, it must bind to its chaperone FlgA in the periplasm (Nambu and

Kutsukake, 2000). The P ring holds a hole in the peptidoglycan layer, whereas the L-ring holds a hole in the outer membrane layer (Macnab, 2003).

1.6.2. The motor

This part of the flagellum represents an important part in powering the bidirectional rotation (Figure. 1.9). The motor can be subdivided into two parts: the rotor/switch or C ring and the stator or Mot complexes (Macnab, 2004). The C ring which is attached to the cytoplasmic side of the MS ring is assembled from three protein subunits: FliG, FliM and FliN. The C ring has a cylindrical shape with a length of 170 A˚ and a diameter of 450 A˚ (Francis *et al.*, 1994). On the other hand, the stator is formed by two types of proteins: MotA and MotB. They are integrated into the cytoplasmic membrane surrounding the MS ring and C ring (Macnab, 2003). The interaction between FliG and MotA causes the torque power in the flagellum (Lloyd *et al.*, 1996). Moreover, FliM has been found to interact with the phosphorylated chemotactic sensor CheY to adjust the direction of rotation (Park *et al.*, 2006).

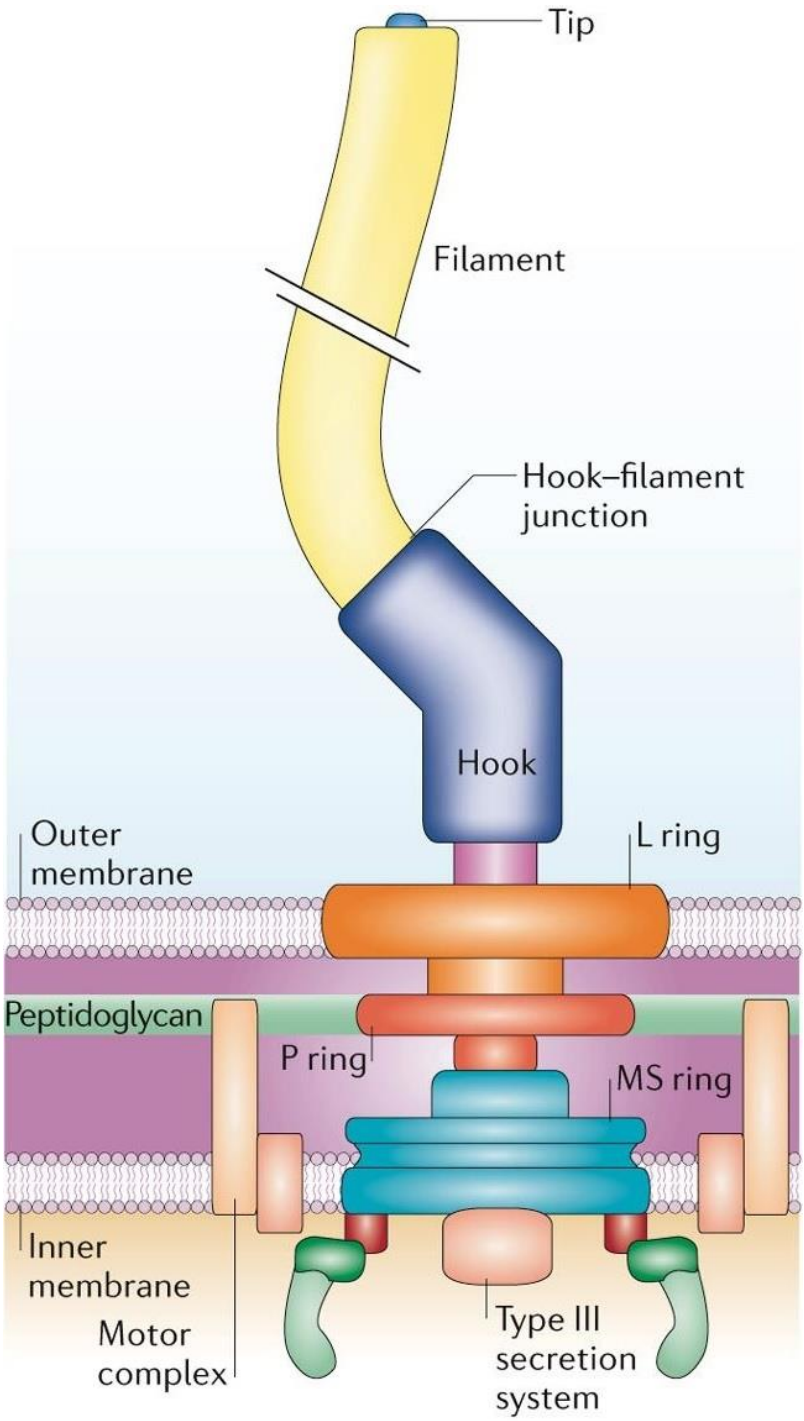


Figure. 1.9. The main structural components of the flagellum.

The MS ring is integrated into the cytoplasmic membrane. It binds the motor to the hook. The P-ring is integrated into the peptidoglycan layer, whereas the L-ring is integrated into the outer membrane. The growing hook ends up with the hook-filament junction zone from where the filament is extended. Figure taken from (Pallen and Matzke, 2006).

1.6.3. The flagellar protein secretion apparatus

Extracellular components of the flagellum are exported by the action of flagellum specific type three secretion system (T3SS). It is located in the cytoplasmic side of the basal body and consists of six integral membrane proteins: FlhA, FlhB, FliO, FliP, FliQ and FliR and three soluble proteins: FliI, FliH and FliJ (Figure. 1.10) (Minamino and Macnab, 1999). The integral membrane proteins form a gate at the base of the MS ring (Minamino *et al.*, 2008). Moreover, FlhA and FlhB have large C-terminal domains extending into the cytoplasm (Figure. 1.10). FliI has been known to have ATPase activity to generate the power for secretion system function, whereas FliH is a negative regulator for FliI (Minamino and Macnab, 2000). Additionally, FliJ has been known as a general chaperone to translocate extra-flagellar subunits to the export apparatus gate (Minamino *et al.*, 2000a).

The mechanism of action of the apparatus includes the formation of a $FliH_2$FliI-FliJ-substrate complex and the following translocation of this complex to the FlhA-FlhB gate (Minamino *et al.*, 2008). The switch between exporting rod/hook subunits and filament subunits is determined by a cleavage in the C terminal domain of FlhB. Upon the completion of the hook, FlhB undergoes conformational changes in the region between Asn-269 and Pro-270 in order to switch substrate specificity (Fraser *et al.*, 2003). The protein FliK has been known to give the signal for such change. Furthermore, the inner diameter of the flagellum is 20 A˚. Thus, the substrate units must translocate inside this channel in an unfolded manner (Yonekura *et al.*, 2003). In addition to FliJ, three substrate specific chaperones have been identified. FlgN, FliS and FliT bind to FlgK and FlgL, FliC and FliD, respectively, and guide them in an unfolded state preventing their degradation (Macnab, 2004). Additionally, the protein σ^{28} that induce the late gene transcription is also considered as a chaperone for FlgM facilitating it's secretion upon the completion of the hook-

basal body (HBB) (Aldridge *et al.*, 2006c). Aldridge *et al.* (2006) demonstrated that *ΔfliA* mutants are unable to secrete FlgM through HBB structures.

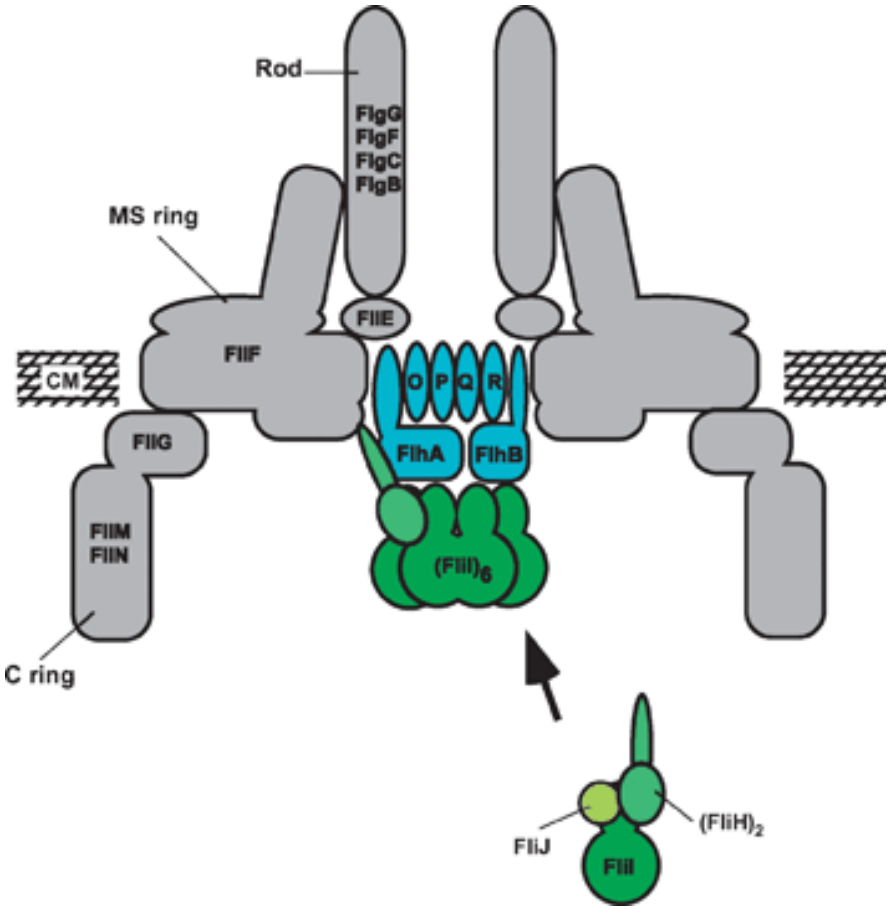


Figure. 1.10. Schematic model of the flagellar protein secretion apparatus.

The integral membrane proteins FlhA, FlhB, FliO, FliP, FliQ and FliR sit in the opening of the MS ring. FlhB is responsible for substrate specificity switch upon the completion of the hook-basal body. In addition to the integral proteins, three soluble proteins exist: FliI, FliH and FliJ. FliI has the ATPase activity needed for the function of this apparatus, whereas FliH is a negative regulator for FliI. FliJ is a general chaperone translocate flagellar subunits to the export apparatus gate. CM: cytoplasmic membrane Figure taken from (Minamino *et al.*, 2008).

1.6.4. The hook

The hook is a cylindrical structure linking the basal body and the filament (Figure. 1.9). It is built up by FlgE protein subunits exported by the export apparatus system (Macnab, 2003). Upon the completion of the rod, the rod cap will be replaced by the hook cap which is synthesized by FlgD. The hook cap is acting as a scaffold, aiding the polymerization of FlgE into the growing hook (Ohnishi *et al.*, 1994). The length of the complete hook is about 55 nm. This length has been studied to be controlled by FliK (Hirano *et al.*, 1994). FliK binds to the T3SS gate protein, FlhB, leading to its proteolytic cleavage which, in turn, changes FlhB substrate specificity to flagellar filament subunits (Fraser *et al.*, 2003). Upon the completion of hook assembly, the hook cap is replaced by three zones of proteins that work as structural adaptors. They are the first hook-filament junction, the second hook-filament junction and the filament cap. These structures are synthesized by FlgK, FlgL and FliD, respectively (Macnab, 2003). These structures are known to take part in filament formation (Homma *et al.*, 1984).

1.6.5. The filament

The filament has a long cylindrical shape which consists of 11 fibrils that form a helical structure that works as a propeller to aid the bacteria to move (Figure. 1.9). It is synthesized by around 20000 flagellin subunits of the FliC protein (Macnab, 2003). However, some species of bacteria have more than one type of flagellin subunit. For example, *Salmonella* has up to three flagellin types, FliC, FljB and FlpA (triphasic) which can be generated as a result of lateral transfer or flagellin genes recombination allowing the emergence of new serovars (Smith and Selander, 1991). While in *Caulobacter crescentus* there are six flagellin types: FljJ, FljK, FljL, FljM, FljN and FljO (Faulds-Pain *et al.*, 2011). Nevertheless, not all flagellins in *Caulobacter crescentus* are required to build the filament. Different combinations of the

Caulobacter crescentus subunits lead to different lengths of filament which, in turn, caused variations in motility speeds (Faulds-Pain *et al.*, 2011).

Upon the initiation of filament growth, the two junction proteins FlgK and FlgL remain in their place, whereas the filament cap FliD moves up as new flagellin units are added. As with the hook cap, the pentameric filament cap is thought to support the polymerization of flagellin units in the growing filament and permits insertion of one flagellin unit as it rotates (Yonekura *et al.*, 2000). Unlike the hook, the filament does not have length control mechanisms (Macnab, 2003). Furthermore, broken filaments can be regenerated again as the filament cap and the hook associated proteins are synthesized continuously (Homma and Iino, 1985).

To sum up, flagella are the motility organelles of many bacteria which serve in the movement from poor to enriched environments. The process of flagellar assembly is a complicated mechanism that requires over 50 genes and 25 protein subunits. Additionally, they are assembled starting from the basal body and the subsequent transport of different structural subunits to form the hook and the filament.

1.7. The regulation of flagellar genes

Flagella help bacteria to move from bad conditions to more favourable ones. Flagella also helps the bacteria to adhere to host cells and in biofilm formation (Soutourina and Bertin, 2003). Because flagella consume high energy for their assembly and function, flagellar synthesis should be highly regulated to reflect the real need of the cells (Soutourina and Bertin, 2003). Flagellar gene promoters which are divided into class 1, class 2 and class 3 initiate transcription of early, middle and late genes, respectively (Chilcott and Hughes, 2000). Following external and internal cues, the class 1 promoter initiates transcription hierarchy of *flhD* and *flhC* which, in turn, express FlhD and FlhC to form the flagellar master regulator, $FlhD_4C_2$. The master

regulator with σ^{70} acts on class 2 promoters to express middle genes. As a result, the structural units of basal body, hook, secretion apparatus and some regulatory proteins will be synthesized. FlgM and FliA (σ^{28}) which are encoded by the middle genes will combine together until the basal body and the hook will be synthesized. Upon the completion of the HBB, FlgM will be secreted out of the cell by the export apparatus freeing σ^{28} to act on class 3 promoters (Figure. 1.11 and 1.15). The class 3 promoters will activate the expression of the late genes (subunits of the filament, motor and chemotactic system) (Chilcott and Hughes, 2000; Aldridge and Hughes, 2002).

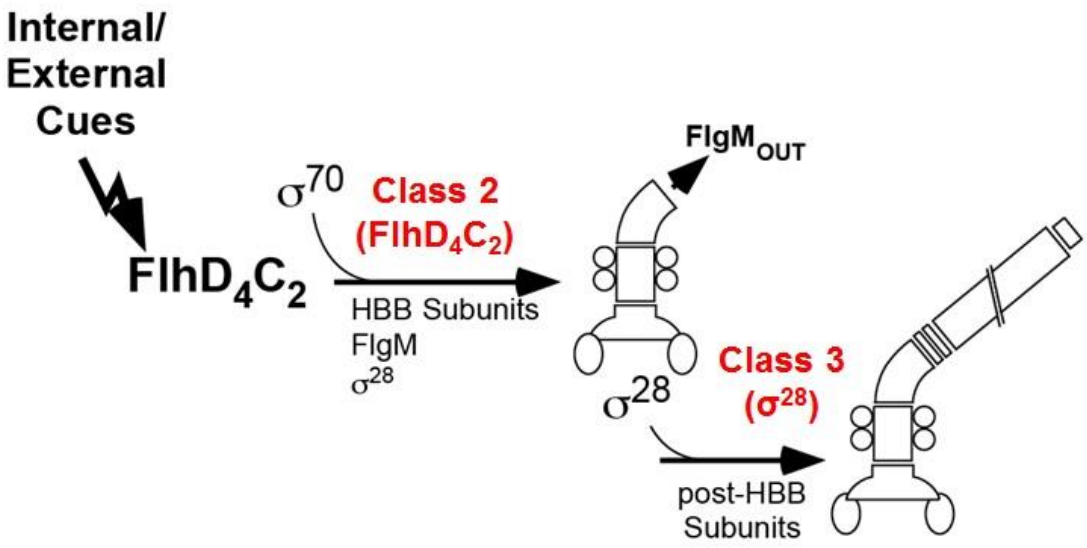


Figure. 1.11. Schematic representation of the transcription hierarchy of flagellar promoter classes.

Following internal and external signals, the class 1 promoter initiates transcription of *flhDC* to form $FlhD_4C_2$. $FlhD_4C_2$ together with σ^{70} activate class 2 promoters and structural units such as hook-basal body (HBB) will be synthesized. Upon the completion of HBB, FlgM will be secreted out of the cell freeing σ^{28} to activate the class 3 promoters. The class 3 promoters will activate the expression of the late genes to synthesize other structural units such as FliC. Figure taken and modified from (Brown *et al.*, 2009).

1.7.1. The flagellar master reulator: $FlhD_4C_2$

The flagellar master regulator in the enterobacteria is $FlhD_4C_2$ expressed from the *flhDC* operon. Initially, it has been thought that this master regulator is a heterotetrameric molecule with the structure $FlhD_2C_2$ (Liu and Matsumura, 1994). However, a crystallographic study showed that two FlhC molecules bind to two FlhD dimeric molecules. Therefore, together FlhD and FlhC form a hexameric molecule $FlhD_4C_2$ which has a molecular mass of 96.4 kDa (Figure. 1.12) (Wang *et al.*, 2006). Wang *et al.* (2006) also diagnosed a zing binding site in FlhC. When $FlhD_4C_2$ binds to the DNA, it is proposed to bend the DNA by ~111 ˚ (Wang *et al.*, 2006).

The protein FlhC alone is able to bind its specific promoters. However, when FlhD links to it, FlhC will undergo conformational changes and the DNA binding specificity and stability of $FlhD_4C_2$-DNA complex will increase (Claret and Hughes, 2000a). In order to bind efficiently to the DNA, $FlhD_4C_2$ requires a minimum of 48 bp (Liu and Matsumura, 1994). This region consists of 17-18 bp inverted repeats separated by a spacer of 10-11 bp (Claret and Hughes, 2002). In strain MG1655, the $FlhD_4C_2$ consensus sequence is AATGGCAGAAATAGCG and CGCTATTTCTGCCATT separated by 10-12 bp (Stafford *et al.*, 2005). In *E. coli*, the binding sites of $FlhD_4C_2$ overlap the promoter -35 region in *flhB* and *fliL*, whereas it is near -40 of *fliA* (Liu and Matsumura, 1994).

The transcription initiation of $FlhD_4C_2$ dependent promoters requires RNA polymerase complex and σ^{70}. The σ^{70} binds to the core RNA polymerase complex to form a holoenzyme and guide the RNA polymerase to the specific promoters (Paget and Helmann, 2003). After the binding of transcriptional factor $FlhD_4C_2$ to DNA, it interacts with the holoenzyme facilitating its binding to the DNA to initiate the transcription apparatus (Ishihama, 1993). It has been demonstrated that $FlhD_4C_2$

interacts with the RNA polymerase through the RNA polymerase C-terminal region of the ***α*** subunit (Liu *et al.*, 1995).

FlhD has been found to have another role in addition to regulating flagellar genes: as it can potentially regulate cell division. The *flhD* mutants continue dividing and they are much smaller than wild type cells in stationary phase (Prüss and Matsumura, 1996). Moreover, Stafford *et al.* (2005) screened the *E. coli* genome for putative $FlhD_4C_2$ binding sites and found 39 non-flagellar gene promoters can be activated by the master regulator. The non-flagellar genes described in this study involved in different biological activities such as global regulation and membrane transport (Stafford *et al.*, 2005).

1.7.2. Transcription of flhDC

Many environmental factors are demonstrated to affect the expression of flagellar genes. Heat shock proteins DnaK, DnaJ and GrpE enhance *flhDC* and *fliA* expression following the change in growth temperature (Shi *et al.*, 1992). Additionally, Sim *et al.* (2017) studied the effect of growth rate on flagellar assembly in *E. coli* by growing the cells in steady state chemostat culture. They found that flagellar number is proportionally linked to growth rate and is controlled by *flhDC* expression. The chemostat cultures provide continuous nutrients to the cells, nevertheless, the cells continued to build flagella. Thus, this data provides evidence that motility in *E. coli* is required for other reasons rather than only the scarcity of nutrients (Sim *et al.*, 2017).

Moreover, phosphorylated OmpR binds to the *flhDC* promoter and negatively regulates its expression following the change in osmolarity or level of acetyl phosphate *in vivo* (Shin and Park, 1995). Quorum sensing which include the secretion of autoinducers to the media and the following concentration recognition by the cells to modify gene expression also impacts *flhDC* expression (Sperandio *et al.*,

2002). The quorum sensing *E. coli* regulator QseBC which is considered as a transcriptional regulator of the flagellar genes positively regulate *flhDC* expression leading to increased motility (Sperandio *et al.*, 2002). In low pH environments, flagellar gene expression is repressed. This repression is mediated by the effect of Histone-like nucleoid-structuring protein (H-NS) on the *flhDC* promoter region as well as on the 5' *flhDC* mRNA region (Soutourina *et al.*, 2002). Inversely, the cyclic AMP catabolite activator protein (cAMP-CAP) has been found to positively regulate flagellar gene expression by binding to class 1 promoters (Soutourina *et al.*, 1999). This is true when the bacteria grow on a poor carbon source as high amount of cAMP-CAP will synthesize and the cells become increasingly motile. The global regulator, CsrA, represses a number of genes by binding to their mRNA to prevent them from translation, is found to activate flagellar genes expression. It protects the 5' of *flhDC* transcript from the action of RNases (Yakhnin *et al.*, 2013). The cross talk of host and pathogenic bacteria is another way of regulating different virulence factors including flagella.

In *salmonella*, the expression of flagellar genes, pathogenicity island 1 and type 1 fimbrial genes is coordinated in a sequential manner (Saini *et al.*, 2010). Moreover, Clegg and hughes (2002) found an inverse mechanism between flagella and fimbriae formation. FimZ which activates fimbriae formation is found to negatively regulate *flhDC* expression leading to a nonmotile phenotype (Clegg and Hughes, 2002). MatA which is considered common activator of the *mat* fimbria operon is a negative regulator of *flhDC* by interacting with its promoter. MatA, therefore, found to change *E. coli* life style from planktonic cells to biofilm formers or vice versa (Lehti *et al.*, 2012). On the other hand, LrhA was also found to bind *flhDC* promoter and negatively regulate its expression (Lehnen *et al.*, 2002). In a similar fashion, the His-Asp phosphorelay system RcsCDB is considered as a negative regulator of *flhDC*

(Francez-Charlot *et al.*, 2003). DksA and ppGpp which bind to RNA polymerases are found to negatively regulate *flhDC* by inactivating class 1 and class 2 promoters in stationary phase (Lemke *et al.*, 2009). Additionally, *Salmonella* pathogenicity island 1 master regulator HilD binds to the *Salmonella flhDC* promoter and positively activate *flhDC* expression (Singer *et al.*, 2014). Unlike *Salmonella* which has six transcription start sites within its *flhDC* promoter, *E. coli flhDC* promoter has only one (Yanagihara *et al.*, 1999).

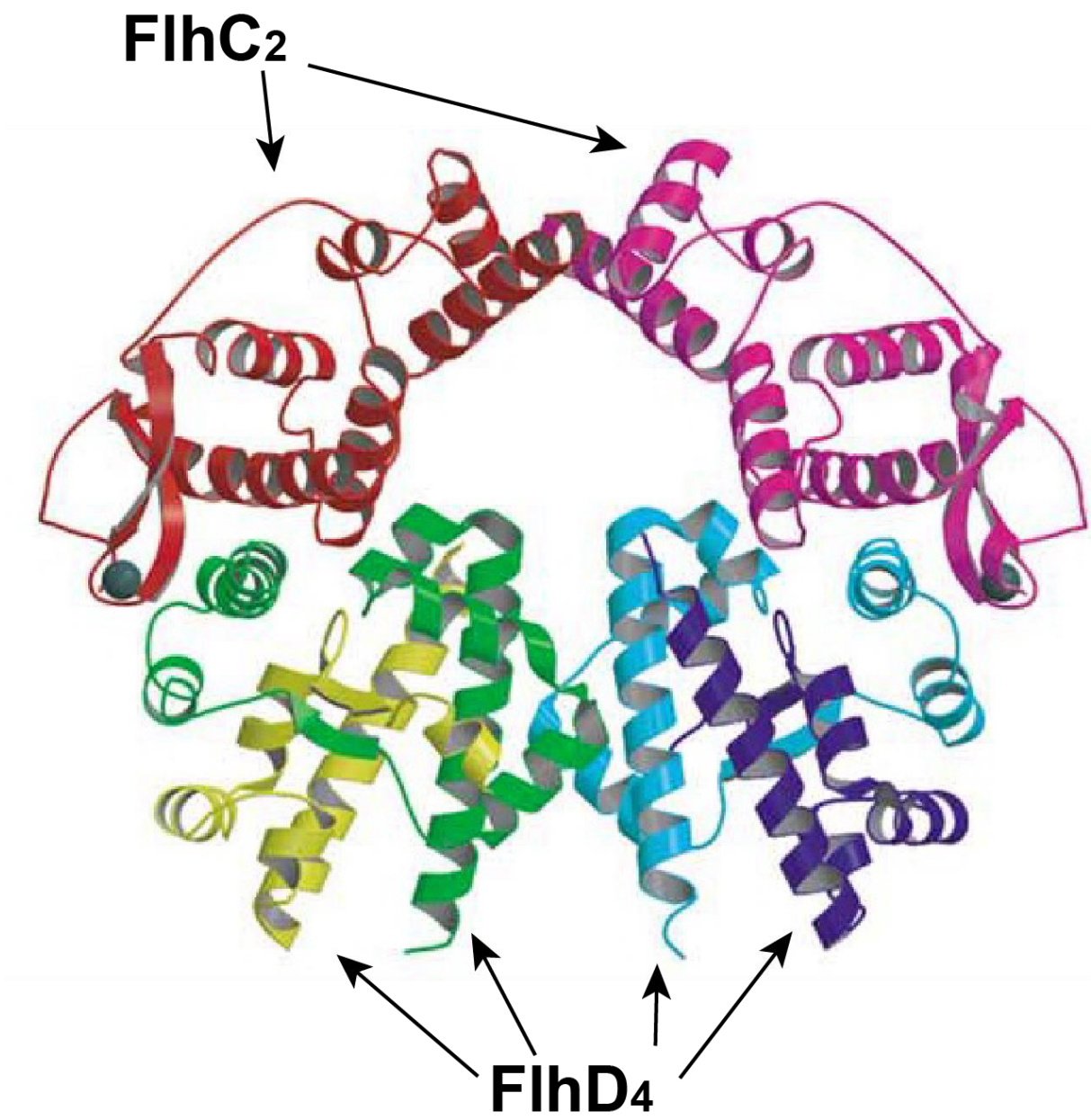


Figure. 1.12. The crystallographic structure of the flagellar master regulator, $FlhD_4C_2$.

Two FlhC molecules bind to two FlhD dimeric molecules to form the $FlhD_4C_2$ hexameric complex. The two FlhC helices are in red and pink, whereas the four FlhD helices are in yellow, green, blue and purple. Figure taken from (Wang *et al.*, 2006).

1.7.3. IS elements

IS elements are part of the mobile genetic events that alter the cellular genetic material through the association or dissociation of DNA fragments. IS elements are segments of DNA that can move from one position on a chromosome to a different position on the same chromosome or horizontally as parts of bacteriophages or plasmids (Siguier *et al.*, 2014). They are generally small in size (< 3 kb) flanked by short terminal inverted repeats (IR) and contain genes expressing transposon-specific enzymes, the transposase, which is required for their mobility (Figure 1.13) (Mahillon and Chandler, 1998; Siguier *et al.*, 2006; Siguier *et al.*, 2014). A relatively old procedure used to capture IS elements from Gram-negative bacteria exploits the *sacB* gene of *Bacillus subtilis*. The *sacB* gene was cloned into a plasmid and transformed to *Agrobacterium tumefaciens* and *E. coli*. The transformed bacterial cells were then plated on LB agar containing 5 % sucrose, and a few colonies survived. Analysis showed many colonies had captured an IS element (Gay *et al.*, 1985). The idea of this research was to inactivate *sacB* which expresses levansucrase in the presence of sucrose by the IS elements. Bacterial cells without IS elements in the presence of levansucrase resulted in the death of the cells (Gay *et al.*, 1985). Therefore, this method facilitated the isolation and characterisation of IS elements as they often do not possess genes with observable phenotypes.

Barker *et al.* (2004) studied IS elements specific for the *yecG-flhDC* intergenic region of a collection of K-12 *E. coli* strains. They found that strains with no IS events were poorly motile, while increased motility was observed in strains containing IS elements in the *yecG-flhDC* region (Figure. 1.14A) (Barker *et al.*, 2004). Motility has been noticed after prolonged incubation in motility agar, and the motile population were taken for subsequent analysis by performing PCR reactions comparing the band size of the products (Figure 1.14B). The *yecG-flhDC* intergenic region with no

IS elements has a DNA fragment size of ~ 1 kb, while the same region with an IS element will become bigger (approximately 2-3 kb). Two types of IS elements (IS1 and IS5) were isolated from this region with a molecular size of 768 bp and 1195 bp, respectively. Barker *et al.* (2004) concluded from their study that the *yecG-flhDC* intergenic region is vulnerable to IS elements and spontaneous mutation is high due to *E. coli* genome plasticity.

Another study confirmed the impact of IS elements in the *yecG-flhDC* intergenic region on motility (Fahrner and Berg, 2015). The study demonstrated that the insertion of IS elements hundreds of base pairs upstream of the *flhDC* promoter increases the transcription of the flagellar master regulator genes, *flhDC*. In addition to IS1 and IS5 discovered in Barker's study, Fahrner and Berg (2015) found two more types: IS2 and IS30. They found IS elements inserted between -100 bp to –476 bp upstream of the *flhDC* transcription start site, activating transcription regardless of the direction. The motility extent, however, was proportional with the distance of these elements from the *flhDC* promoter region. They concluded that the *yecG-flhDC* intergenic region has a high frequency for IS insertion as a means of selective pressure enabling cells to cope with environmental changes.

Yet another study confirmed the presence of IS1, IS3 and IS5 in the *yecG-flhDC* intergenic region and their correlation with increased motility. The researchers confirmed the insertion sites of IS elements upstream the *flhDC* promoter to take place between -100 bp and -476 bp (Zhang *et al.*, 2017). Moreover, Zhang *et al.* (2017) have explored this region and found that it has multiple stress-induced DNA duplex destabilization (SIDD) sequences. These DNA sequences are susceptible to stress conditions, and they undergo strand separation allowing IS elements to insert. They tested the effect of the environment on the frequency of IS elements insertion by growing the bacteria on different concentrations of agar media (different pore

size). The optimal agar concentration to promote IS insertion was 0.24 % as it was the least concentration which allowed the cells to migrate through interparticle channels. Higher agar concentrations did not permit the bacterial cells to move, while agar concentrations less 0.24 % allowed the cells to migrate freely without the need to synthesize flagella and less mutation rates occurred. Thus, Zhang *et al.* (2017) concluded that IS elements can be inserted into SIDD DNA sequences only when it is needed by the cells as a method for selective pressure. When the agar concentrations where more or less than the typical one, less frequent IS events have been noticed as the bacteria will not be able to use flagella, and thus saving energy (Zhang *et al.*, 2017).

Humayun *et al.* (2017) have confirmed this hypothesis, and added that the DNA changes at SIDD locations is mediated by RcsAB, GlpR, Crp and H-NS proteins following environmental cues resulting in obscuring IS elements specific binding sites (Humayun *et al.*, 2017). The alteration of the genetic content by IS elements is not specific to *flhDC* transcription. IS elements inserted in locations of chromosomes can detach taking other functional genes from the chromosome. Studies on systems such as *glpFK* (glycerol utilization), *bglGFP* (β-glucoside utilization), *flhDC*, *fucAOPIK* (propanediol utilization) and *nfsAB* (nitroreductases) showed that the process of IS insertion is reversible providing fitness advantages and disadvantages for the cells (Humayun *et al.*, 2017; Plague *et al.*, 2017). The mechanism behind interrupting genes through IS elements is influenced by the existence of more than one IS copy on the same DNA molecule and the transposition of these elements (He *et al.*, 2015).

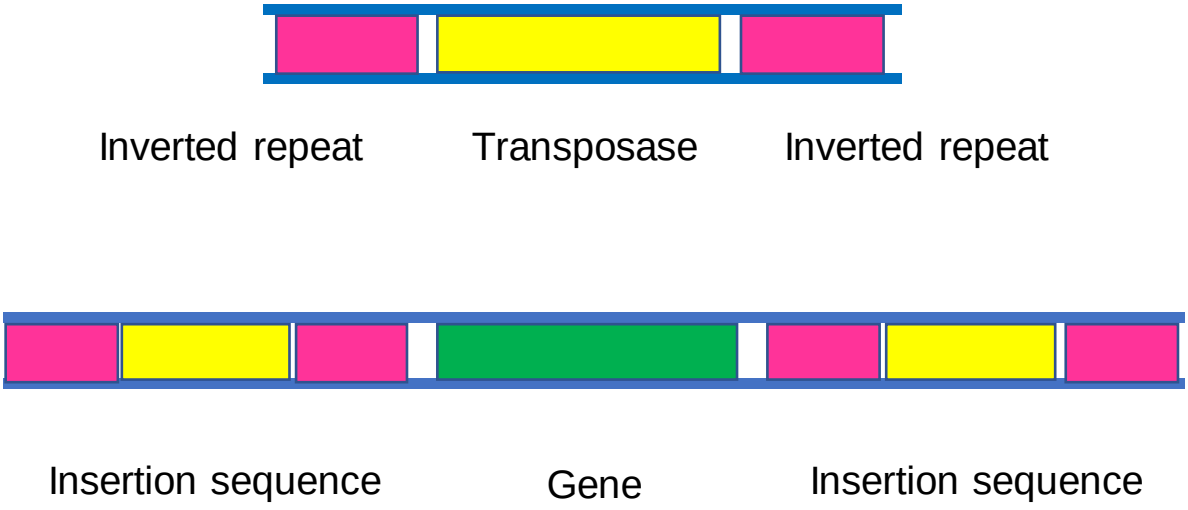


Figure. 1.13. Paradigm of the insertion sequence element structure.

IS elements are DNA fragments flanked by short terminal inverted repeats and include a transposase. Genes could be flanked by two IS elements which they can move together causing loss of DNA. These elements contribute to the transfer of genes between different strains in a species or different species. IS elements can also interrupt gene regulatory regions lead to altered gene expression.

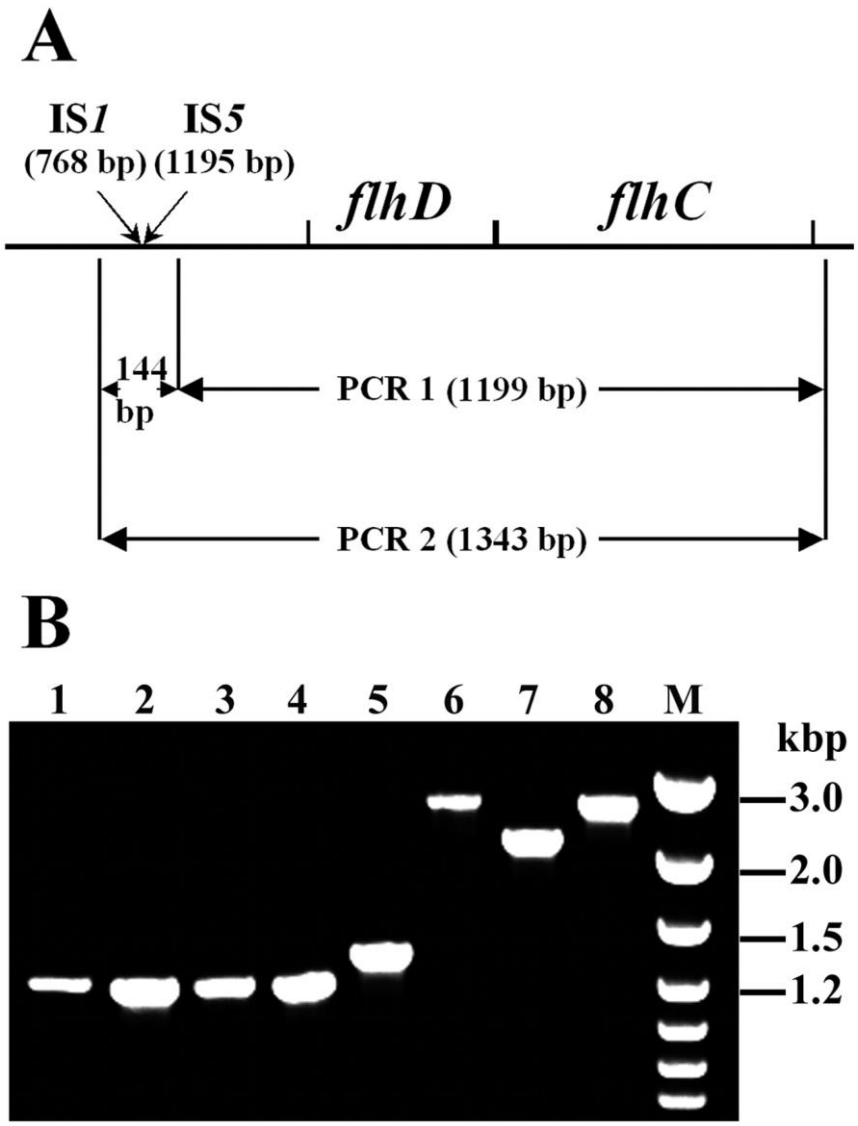


Figure. 1.14. Screening for IS elements in the *flhDC* promoter region of *E. coli*.

A) Two PCR reactions were made: PCR1 performed downstream of the IS elements hot spot, whereas PCR2 performed upstream of the IS elements hot spot. **B)** Gel image shows the PCR products of both reactions (lanes 1 to 4 for PCR1, whereas lanes 5 to 8 for PCR2). If the strain does not have the IS element in the *flhDC* promoter region, the band size will be ~1 kb. The band size will be between 2-3 kb if there was an IS element in the *flhDC* promoter region. Figure taken from (Barker *et al.*, 2004).

1.7.4. Degradation of $FlhD_4C_2$

The proteolytic degradation of damaged or misfolded cytoplasmic proteins by proteases is necessary to the viability of the cell especially under conditions of severe stress (Gottesman, 1999). Proteases can also degrade regulatory proteins to maintain their normal intracellular level (Gottesman, 1999). In prokaryotes, ATP dependent protein degradation is performed by Clp, Lon and FtsH proteases (Tsilibaris *et al.*, 2006). These proteases belong to the AAA^+ family (ATPases associated with a variety of cellular activities) (Neuwald *et al.*, 1999). The mode of action of these proteases includes the recognition of a protein by the protease and the subsequent unfolding which facilitates its translocation to the proteolytic domain. After entering the proteolytic chamber, the unfolded protein undergoes peptide bond cleavage and converted to short peptides (Wickner *et al.*, 1999).

The protease ClpXP or ClpAP consists of a ClpX or ClpA ATPase chamber which recognises substrates with a certain sequence specificity or SsrA tag (a short peptide sequence) (Sauer *et al.*, 2004). The ATPase domain then unfolds the protein and transfer the polypeptide to the ClpP peptidase domain for protein breakdown (Sauer *et al.*, 2004). ClpXP, is a negative regulator for flagellar genes expression. It controls FlhDC master regulator in the post-transcriptional and post-translational levels resulting in the suppression of class 2 and class 3 promoters (Tomoyasu *et al.*, 2002). Tomoyasu *et al.* (2002) showed that *ΔclpXP* mutants of *Salmonella enterica* exhibit an increased rate of flagellar protein production and a higher number of flagella. A study has demonstrated that FlhC in *ΔclpXP* mutants was more stable with a five times longer half life compared to the wild type (Tomoyasu *et al.*, 2003). Interestingly, the study of Tomoyasu *et al.* (2003) revealed that ClpXP does not have the ability to act on FlhC and FlhD separately, but can degrade both subunits when they formed $FlhD_4C_2$ complex.

Lon is another ATP dependent protease that consists of multiple identical subunits. Each of these subunits contains three domains: the N-terminal domain, the ATPase domain and C-terminal domain (Tsilibaris *et al.*, 2006). Lon protease has different biological roles in *E. coli* such as cell division and capsule production (Ebel *et al.*, 1999). It has been also shown that Lon degrades $FlhD_4C_2$ in *Proteus* and *E. coli* (Claret and Hughes, 2000b). Claret and Hughes (2000) showed that *Proteus* and *E.coli Δlon* mutants reveal more $FlhD_4C_2$ stability than that of the wild type.

Yet another protease which regulates unfolded protein intracellular levels is FtsH. This protease is anchored in the cytoplasmic membrane by the N-terminal domain. It also has the ATPase domain and the metalloprotease domain (Ito and Akiyama, 2005). FtsH contributes in the quality control of membrane proteins by degrading damaged ones which are not assembled into complexes (Akiyama, 2009). Although FtsH does not participate in $FlhD_4C_2$ degradation, it regulates motility by influencing FliC levels. The aconitase AcnB which catalyses citrate and isocitrate in the citric acid and glyoxylate cycles controls FliC synthesis by binding to *ftsH* transcript (Tang *et al.*, 2004). Tang *et al.* (2004) showed that the binding of AcnB to *ftsH* transcript will reduce FtsH level and, therefore, enhances σ^{32} level. High levels of σ^{32} will increase DnaK intracellular amount which leads to elevated production of FliC and, thus, motility (Tang *et al.*, 2004).

1.7.5. Other aspects of regulation in the flagellar system

As we have mentioned earlier, FliS is a specific chaperone for Flagellin FliC subunit, whereas FliT is specific chaperone for the filament cap, FliD. These proteins which are expressed from both middle and late genes are found to negatively regulate the export of FlgM through T3SS. Mutants of *fliDST* operon have been found to synthesize more flagella due to enhanced FlgM secretion which lead to the activation of late genes expression (Yokoseki *et al.*, 1996). The crystallographic studies showed

that FliT consists of four α-helices, and that the C-terminus of FliT binds to $FlhD_4C_2$, FliI and FliJ (Imada *et al.*, 2010). Thus, FliT has an important function in regulating flagellar gene expression by binding to free $FlhD_4C_2$ complex (not DNA bound) and preventing it from binding to the DNA of the class 2 promoters (Aldridge *et al.*, 2010). This mode of action has been shown to depend on the relative interactions of FliT/FliD, FliT/$FlhD_4C_2$ and $FlhD_4C_2$-DNA complex and, thus, modulating the level of $FlhD_4C_2$ available to induce flagellar expression, accordingly. Aldridge *et al.* (2010) have proved that this mechanism is important for the cells to better respond to external cues with regard to motility by regulating the flagellar system. Altogether, FliT negatively regulates the flagellar assembly by binding to free $FlhD_4C_2$, whereas FliD acts as anti FliT factor before HBB completion (Aldridge *et al.*, 2010).

YdiV is another protein that regulates the amount of $FlhD_4C_2$ according to the nutrient availability by binding to $FlhD_4C_2$ and preventing it from activating other promoters. This regulator also facilitates the degradation of $FlhD_4C_2$ by ClpXP (Takaya *et al.*, 2012). Moreover, YdiV can bind to DNA-bound $FlhD_4C_2$ causing its release from the DNA and stopping it from subsequent activation cycles (Takaya *et al.*, 2012). Unlike *E. coli* which is motile under low nutrients conditions, *Salmonella enterica* serovar Typhimurium becomes motile under rich media conditions as proposed during host cell colonization (Koirala *et al.*, 2014). Interestingly, it has been found that *Salmonella* can form motile and nonmotile cells in the same media and at the same time by using YdiV/FliZ switch to regulate its response to nutrients (Koirala *et al.*, 2014). High amount of nutrients supresses the *flhDC* negative regulator, YdiV leads to the expression of *flhDC* to form flagella. As a result, class 2 genes such as FliA will amplify the expression of FliZ which, in turn, represses the action of YdiV. This study revealed the effect of nutrients on flagellar gene regulation and adaptability to different environments (Koirala *et al.*, 2014). FliZ has also found to

couple class 2 promoter activation to flagellar assembly by binding to $FlhD_4C_2$ mRNA to positively regulate its level (Saini *et al.*, 2008). It has also found that FliZ activates class 2 promoters allowing the cell to regulate flagellar number according to the external conditions (Saini *et al.*, 2008). The *sirA* gene in *Salmonella enterica* serovar Typhimurium has orthologs in other species of bacteria such as *uvrY* in *E. coli*. This gene is found to negatively regulate flagellar gene regulation, though the mechanism is not well understood (Goodier and Ahmer, 2001). The secretion systems in bacteria use a secretion signal found in the amino acid sequence of the secreted substrates, the 5' region of the mRNA and the ability of a secretion chaperone to bind and translocate substrate (Aldridge and Hughes, 2001). In regard to these signals, a signal was found in the 5' untranslated region (5' UTR) of *fliC* transcript which is required in the translation to FliC in the growing filament. The whole 62 bases upstream of *fliC* was found to be required to make a stem loop in the *fliC* transcript required for this signal (Aldridge *et al.*, 2006a).

1.7.6. FlgM-σ^{28} regulatory checkpoint

The sigma factor σ^{28} must bind to the RNA polymerase in order to activate the class 3 promoter and initiate transcription of late flagellar genes (Ohnishi *et al.*, 1992). The expression of the late genes results in the synthesis of some flagellar structures such as flagellin subunits, motor and chemosensory system (Chilcott and Hughes, 2000). The σ^{28}, however, is inactivated as a result of its binding to its anti sigma factor: FlgM, prior to the HBB completion (Chadsey *et al.*, 1998). Chadsey *et al.* (1998) demonstrated that FlgM is not only capable of binding free σ^{28} but also polymerase bound σ^{28} which causes holoenzyme destabilization. Upon the completion of HBB, T3SS switches substrate specificity from the hook protein to the late protein subunits by the interaction of FliK with FlhB (Minamino *et al.*, 1999). Coincidentally, FlgM will

be secreted out of the cell through HBB structures freeing σ^{28} to activate late flagellar genes (Figure. 1.15) (Hughes *et al.*, 1993).

Yet another mechanism to control FlgM secretion is by a cytoplasmic facing protein Flk which is anchored to the inner membrane by its C terminal (Figure. 1.15) (Aldridge *et al.*, 2006b). Aldridge *et al.* (2006) studied *flk* mutants and found that FlgM will be secreted to and degraded in the periplasm. Therefore, Flk is another mechanism to monitor the hook length and is interacting with FlhB to regulate substrate specificity switch which allows shorter hooks to grow to the normal length (Aldridge *et al.*, 2006b).

Moreover, Karlinsey *et al.* (2000) showed that FlgM translation is coupled to its secretion by the chaperones. At the start of the HBB synthesis, FlgM translation is suppressed by the chaperone FlgN due to interaction with FlgN specific substrates, FlgK and FlgL. Upon completion of the hook-basal body, however, FlgN will activate FlgM translation as a result of the mRNA and amino acid secretion signal (Karlinsey *et al.*, 2000). Additionally, Aldridge *et al.* (2003) studied this theory further by measuring the level of FlgK and FlgL and its relation to FlgM translation. Aldridge *et al.* (2003) demonstrated that upon the completion of flagellar assembly, the secretion level of FlgKL will decrease leading to increased level of FlgN-FlgKL complexes and thus FlgM translation will be reduced (Aldridge *et al.*, 2003). Therefore, this negative feedback loop of suppressing FlgM translation by FlgN (when FlgK and FlgL level is adequate inside the cell) is considered as a signal to regulate different stages of flagellar assembly (Aldridge *et al.*, 2003).

The σ^{28}-FlgM control mechanism of coupling late gene expression to the completion of HBB structures exists in the cells which start flagella synthesis (Aldridge and Hughes, 2002). The cells which already have synthesized flagella, the σ^{28}-FlgM mechanism will not be remarkable (Brown *et al.*, 2008). Brown *et al.* (2008)

found that the activation of the different promoter classes is continuous and it depends on the number of flagella. Additionally, the expression of these promoter classes will be determined by the temporal secretion rates of the late proteins especially FlgM which control flagellar gene expression (Brown *et al.*, 2008).

The genes *fliA* and *flgM* are not only considered as middle genes, but also as late genes transcribed from both class 2 and class 3 promoters (Chevance and Hughes, 2008). When FlgM is secreted outside the cell, σ^{28} will be free to bind to the RNA polymerase and activate class 3 promoters. On the other hand, when FlgM is available in the cell it will bind to σ^{28} and negatively regulate the class 3 promoter (Chevance and Hughes, 2008). After the transcription initiation of the class 3 promoter, σ^{28} will split from the RNA polymerase and degraded in the absence of FlgM (Barembruch and Hengge, 2007). This degradation of σ^{28} by proteases mainly Lon will help in the maintenance of balanced level of σ^{28}-FlgM and, thus, regulating class 3 promoter activation (Barembruch and Hengge, 2007).

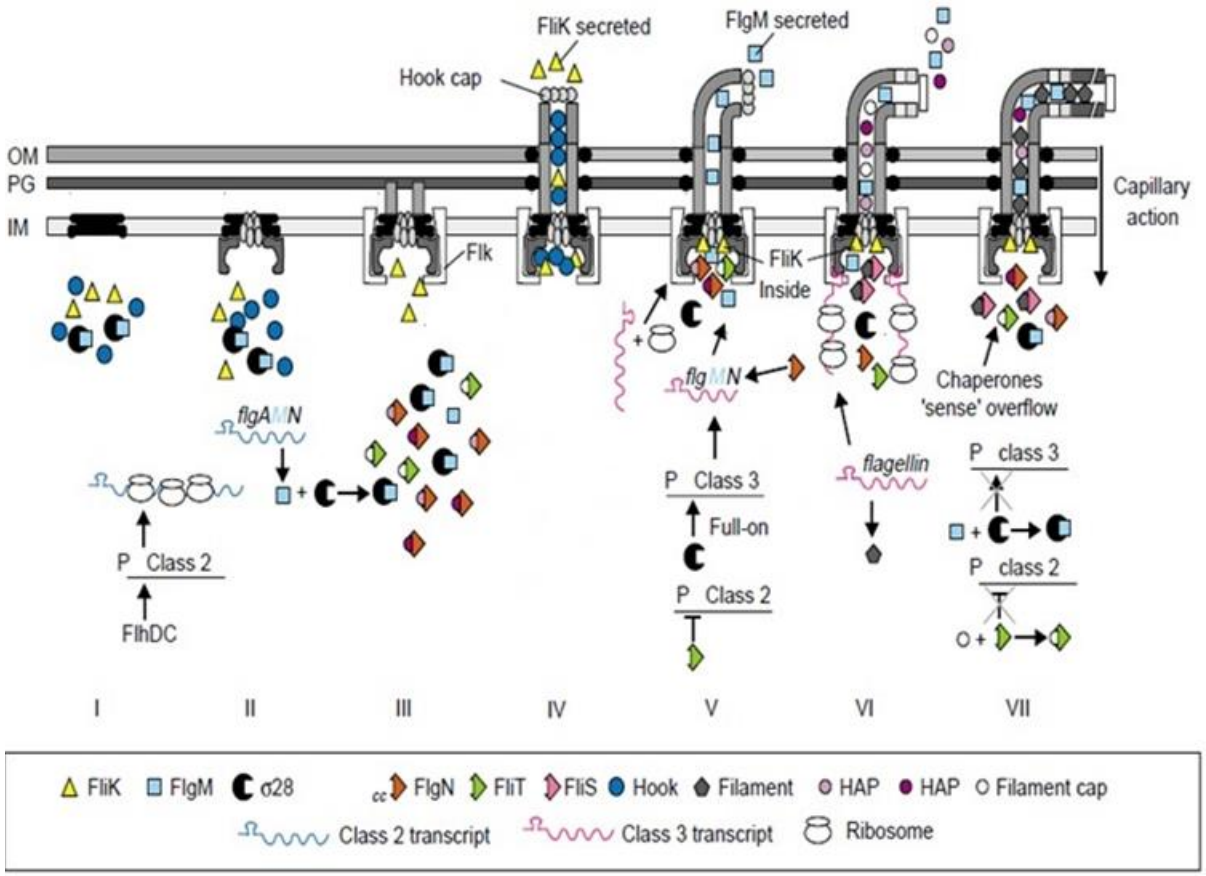


Figure. 1.15. Paradigm of the regulatory network of flagellar assembly.

Following internal and external cues, the class 1 promoter activates the expression of *flhD* and *flhC* to form the master regulator, $FlhD_4C_2$. The master regulator will activate the class 2 promoter and the middle genes expression. This will lead to the synthesis of the basal body, hook, secretion apparatus subunits and other regulatory proteins. Upon the completion of the hook and the basal body, the secretion apparatus undergoes substrate specificity, and FlgM will be secreted out of the cell. As a result, σ^{28} will activate the class 3 promoter and late genes will be expressed. The subunits of filament, motor and the chemotactic system will form. IM: inner membrane. PG: peptidoglycan. OM: outer membrane. Figure taken and modified from (Aldridge and Hughes, 2002).

1.8. Biofilm formation: The other life style

Bacteria can move in the aqueous media from one location to another using flagella. This motility help the bacterial cells to avoid detrimental environments and migrate to more favourable conditions. In certain conditions, however, the bacterial cells give up motility and enter to another life style: biofilm formation on surfaces. The bacterial biofilm is a structured community of cells encased in a polymeric matrix synthesized by these cells to adhere to inactive or living surfaces (Costerton *et al.*, 1999). The biofilm has many advantages to bacterial cell survival. For example, bacteria that form biofilm on implanted medical devices and tissues can resist antibiotics due to the inability of the antibiotics to penetrate the biofilm layer (Stewart and Costerton, 2001). Strains of *E. coli* are common in causing chronic and recurrent UTIs mainly because of biofilm formation. Unlike the virulent UPEC which causes symptomatic UTIs, the non-virulent asymptomatic bacteriuria (ABU) were demonstrated as a better biofilm formers and they can out-compete other UTI pathogens (Ferrieres *et al.*, 2007). This bacterial interaction and the variation in ability to form biofilms has a potential importance in treating patients vulnerable to uropathogenic strains (Ferrieres *et al.*, 2007). Moreover, bacteria can tolerate extreme dryness conditions and high osmolarity by forming biofilm, thus, aiding the cells to survive under these severe conditions (Chang *et al.*, 2009).

In drinking water distribution systems, *E. coli* can survive by forming biofilms and the presence of *E. coli* is considered as an indicator for fecal contamination (Abberton *et al.*, 2016). Additionally, enteric bacteria such as *E. coli* are able to synthesize biofilms outside the host which protect the bacteria from being killed by soil nematodes and other predatory bacteria (DePas *et al.*, 2014). Biofilms can act as nutritional reservoirs having more than one species of bacteria. In this regard, a species of bacteria can locate with another species that produce nutritional elements

as by products of their metabolism such as nitrifying bacteria in wastewater biofilms (Satoh *et al.*, 2006). The oral bacteria that form biofilms on the enamel and gingival tissues is another example of adhering to nutrition rich media (Kolenbrander *et al.*, 2002). Furthermore, horizontal gene transfer occurs more frequently inside biofilms bacterial communities. Plasmids and free DNA can be transferred to foreign cells leading to the emergence of new antibiotic resistance strains and strains having new catabolism plasmids (Molin and Tolker-Nielsen, 2003).

E. coli has been extensively studied for its biofilm formation. Upon the transition from postexponential phase characterized by suboptimal nutrients to stationary phase when nutrients are further reduced, cells will stop building flagella, change to be ovoid and start to form a biofilm (Serra *et al.*, 2013b). The extracellular matrix of the biofilm consists of protein components known as curli (amyloid fimbria) expressed by *csgBAC* and *csgDEFG* operons. Biofilms also consist of exopolysaccharide mainly as cellulose which is synthesized by cellulose synthase (BcsA and BcsB). Other exopolysaccharide components may include capsules, BapA and poly-β-1,6-N-acetyl-D-glucosamine (PGA) (Yaron and Romling, 2014). The formation of biofilm and the accompanied cellular changes depends on the stationary phase sigma subunit of RNA polymerase RpoS (σ^s), cAMP, ppGpp, cyclic-di-GMP (c-di-GMP) and the biofilm transcription factor (CsgD) (Serra *et al.*, 2013b). The formation of a biofilm is regulated by an intricate and complicated regulatory network in *E. coli* (Figure. 1.16). The two master regulators FlhDC and σ^s are regulated by cAMP and ppGpp, respectively in this network (Serra *et al.*, 2013b). The protein FliZ which is expressed under the control of FlhDC plays a crucial role in switching between flagellar synthesis or biofilm formation (Figure. 1.16) (Pesavento and Hengge, 2012). When bacteria still producing flagella, FlhDC master regulator induces the class 2 flagellar promoters. As a result, FliA (σ^{28}), the anti FliA factor

FlgM and other hook-basal body components will be synthesized. At the same time, FliZ will also be synthesized and in turn will represses σ^S to prevent cells from forming biofilms (Pesavento and Hengge, 2012). The secondary messenger signalling system c-di-GMP is produced by multiple diguanylate cyclases (DGCs), characterized by GGDEF domains and is degraded by phosphodiesterases (PDEs), characterized by HD-GYP or EAL domains. These DGCs and PDEs have been demonstrated as big players in the regulatory process of biofilm formation (Hengge, 2010). The accumulation of c-di-GMP has been found to encourage biofilm formation and inhibiting planktonic lifestyle (Wolfe and Visick, 2008). The PDE encoding *yhjH* gene belongs to the flagellar late genes which require FlhDC/FliA for their expression is important in keeping the levels of c-di-GMP low to encourage motility. This, in turn, will keep other late flagellar genes such as *ycgR* inactive which also keep flagella working (Pesavento *et al.*, 2008). However, upon the entry of the cells into stationary phase, FlhDC expression will be inactivated and σ^S will induce the expression of multiple genes. As a result, YegE and YedR will outcompete the action of YhjH, increasing the level of c-di-GMP which, in turn, leads to the activation of YcgR, resulting in the inhibition of flagellar rotation (Pesavento *et al.*, 2008). At the same time, CsgD will be activated and the production of curli and cellulose will increase (Figure. 1.16).

In addition to YegE/YhjH system, the DGC YdaM and the PDE YciR is another system that regulates the level of c-di-GMP. This system is under the control of σ^S increases the level of c-di-GMP which leads to further *csgD* activation and curli production (Sommerfeldt *et al.*, 2009). Moreover, there is a third c-di-GMP system that leads to the production of cellulose, YaiC/YoaD. The DGC YaiC (AdrA in *Salmonella*) catalyze the c-di-GMP level necessary to synthesize BcsA and thus cellulose, whereas PDE YoaD has a negative effect. The *yaiC* gene expression is

controlled by σ^s, YdaM, MlrA and CsgD, and it occurs only in the stationary phase under 30 ˚C (Brombacher *et al.*, 2006). While *yoaD* is simultanuously expressed under the same effect of σ^s, it seems that it is negatively regulated by CsgD. PGA is another exopolysaccharide element which is synthesized by *pgaABCD* operon. Unlike cellulose which is secreted in biofilms outside the host under 30 ˚C, PGA is secreted inside the host cells at 37 ˚C (Cerca *et al.*, 2007). CsrA is found to negatively regulate PGA at the posttranscriptional level (Wang *et al.*, 2005). In contrast, NhaR is a positive regulator of PGA under high salt conditions and alkaline pH (Goller *et al.*, 2006). Again, c-di-GMP was found to contribute to the PGA production by the action of two DGC proteins, YcdT and YdeH (Jonas *et al.*, 2008). These DGC proteins are negatively regulated by CsrA in the posttranscriptional level. The regulatory network of biofilm formation is similar in *Salmonella*. However, there are a few differences especially regarding c-di-GMP catalyzing proteins. Instead of the protein YdaM which increases the level of c-di-GMP in *E. coli*, *Salmonella* has a functionally alternative protein, STM3388 (Kader *et al.*, 2006). Moreover, the PDE protein STM4264 in *Salmonella* has been found to regulate CsgD instead of YhjH (Simm *et al.*, 2007).

Altough flagella are considered as the motility organelle in bacteria, it has been confirmed that they have architectural function within biofilms. In the outer rim and bottom layer of the biofilm regions, cells are in contact with nutrients on the surface and still producing flagella. These flagella, however, will form a lattice which is used to bind cells together and to aid in organizing the cells' position during growth. This cellular structure and binding with the help of flagella supports the growth of the upper layer cells when bacteria enter the stationary phase and start to form the wrinkled structures of the biofilms (Serra *et al.*, 2013b).

To summarise, *E. coli* is a ubiquitous species of bacteria. Its genome plasticity lead to the emergence of new strains, both pathogenic and nonpathogenic. There are different models of classifying the species of *E. coli*. However, the well-known tree of McNally classified *E. coli* into six phylogenetic groups which differ in the content of their DNA. Many molecular methods aided the researchers to manipulate the DNA of *E. coli*. Most of *E. coli* strains are motile by the means of flagella. Flagella are complexed structures which synthesized in a hierarchical manner as a result of three promoter classes. Moreover, FlhDC is the master regulator of flagellar synthesis which synthesized following different regulatory inputs. Additionaly, *E. coli* has another life style which is the biofilm. Biofilms have many functions in the life of bacteria. It has an intricate regulatory network involving many factors such as c-di-GMP.

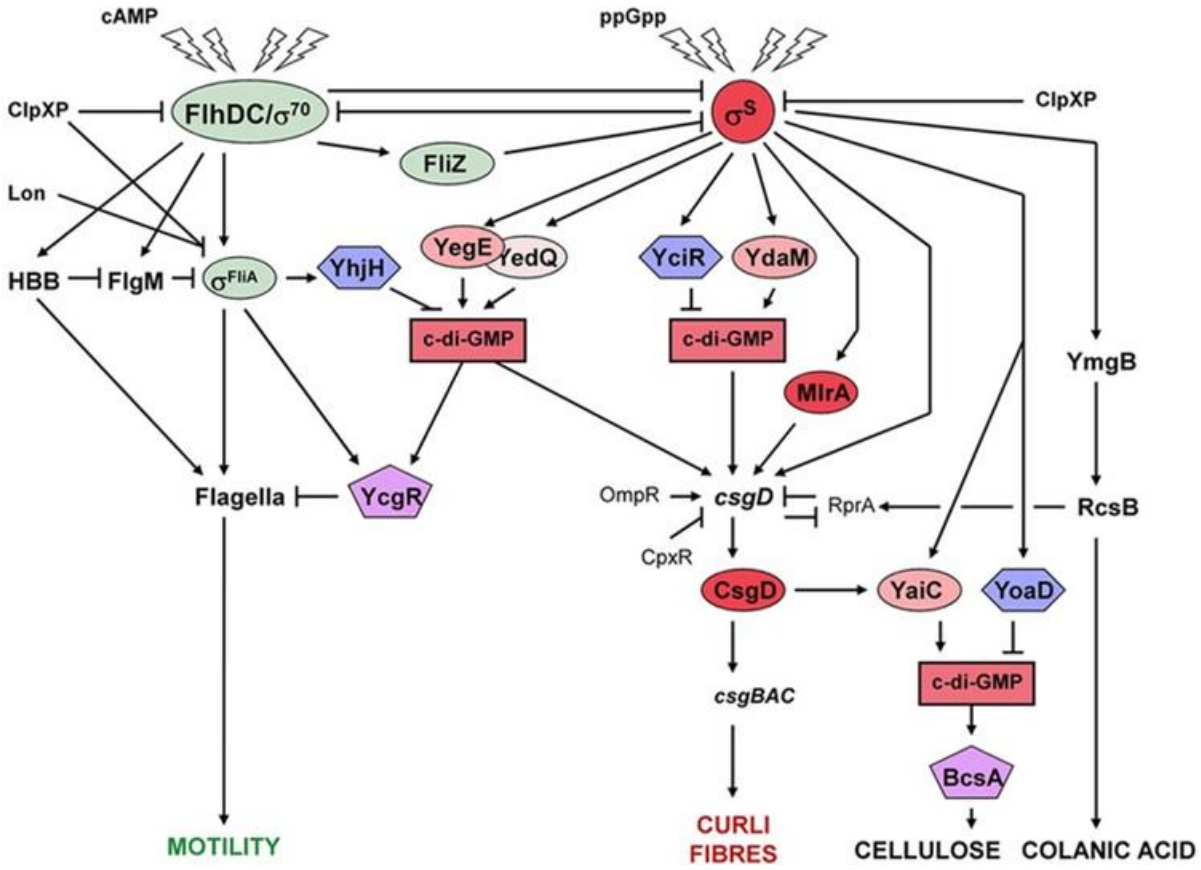


Figure. 1.16. Schematic representation of the regulatory network of biofilms.

Upon entering the stationary phase, cells FliZ has the greatest role in switching from flagella synthesis to biofilm formation. The stationary phase sigma subunit of RNA polymerase RpoS (σ^S) will be activated. As a result, σ^S will activate multiple diguanylate cyclases (pink color oval shapes) and phosphodiesterases (blue color oval shapes). These compounds control the level of c-di-GMP required for different activities. The increased level of c-di-GMP suppresses flagella synthesis and activates the production of curli fibres and cellulose required for biofilm formation. Figure taken from (Serra *et al.*, 2013b).

Chapter Two: Aims of Study

2.1. Aims of study

The literature revealed that whole genome analysis of the species of *E. coli* divides this bacterial species into six phylogenetic groups, A, B1, B2, D, E and F. This classification could be reflected in the *yecG-flhDC* intergenic region which dictates motility. Thus, the alignment of the *yecG-flhDC* intergenic region leads to three main categories, A/B1, B2 and E/D/F. The motility in *E. coli* is driven by the flagellar system which synthesized in a hierarchical process includes the activity of three promoter classes. Following external and internal cues, the *flhDC* master regulator is expressed and lead to the subsequent steps of flagellar assembly. The literature argues that motility in *E. coli* responds to IS events upstream of the *flhDC* operon. IS events activates the *flhDC* expression in strain MG1655. Thus, the central question for this thesis is: Is IS insertion a strain wide or a species wide mechanism to control motility in *E. coli*?

The specific aims of this study to address the central question include:

1. Replace the *yecG-flhDC* intergenic region of strain MG1655 with the same region of other strains of *E. coli* from different phylogenetic groups to study the impact of *flhDC* promoter on motility.
2. Clone the *yecG-flhDC* intergenic region on pSB401 to measure *flhDC* promoter activity.
3. Study the effect of different *flhDC* promoters on the abundance of flagella.
4. Dissect the *yecG-flhDC* intergenic region to define its impact on motility.
5. Characterise *E. coli* strains from different biological niches to demonstrate the effect of the strain phylogenetic location on motility, biofilm formation and IS events.

In our chapters of results, we will indicate that the *flhDC* promoter activity is different in the different strains of *E. coli* representing the species. Our analysis to the *yecG-flhDC* intergenic region will reveal a novel function of *yecG* in repressing flagellar

gene activation. Additionally, our characterisation to two different collections of *E. coli* strains will show different motility phenotypes which reflect the life style of the strains in the chosen environment.

Chapter Three: Materials and Methods

3.1. Bacterial Strains Activation and Growth Conditions

3.1.1. Standard growth conditions

All bacterial strains are assigned a TPA number and frozen at -80 ˚C in 10 % Dimethyl sulfoxide (DMSO). Strains were activated on Luria-Bertani (LB) solid media at 37 ˚C overnight, and LB plates with the appropriate antibiotic were used when needed to maintain plasmid selection. A single colony was taken to inoculate 5 ml LB at 37 ˚C overnight as pre-cultures. Growth curves were made by using glass Erlenmeyer flasks (VWR).

3.1.2. Growth curve in minimal media

The procedure is used when minimal media (minimal E-salts or MinE) is added in growth experiments. Yeast extract (YE) was added to a base of MinE (2x) and glucose to improve growth and alter nutrient availability used in this study at 2 temperatures 37 ˚C and 30 ˚C. Strains were grown in 5 ml of chosen media overnight. Next day, the OD_{600} of the cultures was measured and the volume to be added to 30 ml of media was calculated to give a starting OD_{600} of 0.05. T_0 was set after the addition and the glass Erlenmeyer flasks (VWR) were placed in a water bath at the correct temperature. The OD_{600} was recorded every 30 minutes until either 300-400 minutes or OD_{600} of 3 was reached. The steps were used in conjunction with the protocol for measuring bioluminescence.

3.2. Antibiotic concentrations

Antibiotics used in this study have the following concentrations:

Antibiotic	Stock Concentration (mg/ml)	Working Concentration (µg/ml)
Ampicillin	20	100
Chloramphenicol	2.5	12.5
Kanamycin	10	50
Spectinomycin	50	50
Tetracycline	2.5	12.5

Table. 3.1. Antibiotic solutions used with *E. coli*.

3.3. Motility assay

Strains were checked for their motility by streaking bacteria onto TB or LB agar and incubating overnight at 37 ˚C. Then, motility semi-solid agar was stabbed with a single colony using sterile wooden stick, plates were incubated for six to eight hours at 37 ˚C before scoring. A Syngene Bioimaging cabinet and GeneSnap software at 75 ms was used to capture images of all plates. Images were analysed by Image J software using a ruler photographed with the plates to analyse the swarming size according to the number of pixels. All motility assays were performed in triplicate with one batch of motility agar. Batch experiments prevent variation in chemotaxis ability influencing data comparison.

3.4. Polymerase chain reaction

PCR reactions were performed in the thermocycler (Biometra T3000), and primers were designed using Clone Manager or Serial Cloner softwares. The reaction used Taq polymerases from Promega or NEB or Q5 High Fidelity polymerase from (NEB). The volume of each reaction was 50 µl consists of the following:

Component	Volume (µl)
10 X or 5 X Reaction buffer	5 or 10 respectively
dNTPs (Final concentration of 250 µM)	5
Forward primer (Final concentration of 20 pmole/µl)	2.5
Reverse primer (Final concentration of 20 pmole/µl)	2.5
Template DNA	2
DNA polymerase (Final concentration of 1 U/µl)	1
PCR grade water	To a final volume of 50

Table. 3.2. Standard PCR reaction composition

Thermocycler temperature was set according to the polymerase manufacturer's protocol and primer design. PCR products were then purified by using PCR clean up kit and/or gel extraction kit. To check all PCR reactions, 5 ul was run on agarose gel and the remaining 45 ul was cleaned if necessary.

3.5. Agarose gel DNA electrophoresis

Samples of DNA were tested for their quality and size by running on agarose gel from peqGold Universal. Gel powder was dissolved in TAE running buffer (Sigma) to a concentration of 1 %, heated in a microwave, and stored in an oven at 60 ˚C to avoid solidification. Gels were formed using casting trays, and located in the electrophoresis tank which contains TAE running buffer. Typically, 2 µl of 6X loading dye were added into 5 µl of DNA sample, and 7 µl were run per lane alongside 5 µl 1 kb or 100 bp DNA ladder. Power parameters used to run gels were 110 volts and 400 milliampere for 50 minutes. Images of the gels were taken using Syngene Ingenius Transilluminator and Genesnap software at 150 ms exposure time. When DNA bands had to be extracted from gels, they were exposed to ultra violet light using Syngene GVM20 equipment.

3.6. DNA Sequencing

Samples of successful preliminary confirmation were sent to Source Bioscience DNA sequencing company. Sample requirements for Sanger services sequencing were 100 ng/µl for plasmids and 10 ng/µl for PCR products. The samples sent off for sequencing and primers were either sent with the samples or synthesized by the company. Tubes or plates were prepared and packaged according to company's instructions.

3.7. Isolation of plasmid DNA

3.7.1. Crude mini method

This method is an alternative to the expensive miniprep kit, used to isolate plasmids from the bacterial cell and modified by Phillip Aldridge (Sambrook *et al.*, 1989). Bacterial cultures of 3-5 ml were grown overnight, centrifuged and treated with three solutions: solution I (50 mM Glucose, 25 mM Tris HCl, and 10 mM EDTA with 250 µl RNase for each 50 ml of solution I), solution II (2 ml NaOH 1 M, 1 ml 10 % SDS, and

7 ml water), and solution III (5 M potassium acetate, glacial acetic acid, and water) to destroy cells and extract the plasmids. Tubes were centrifuged for 20 minutes. Plasmids were precipitated by isopropanol by spinning for 15 minutes and aspiration. Pellets were washed with 500 ul 70 % ethanol and centrifuged for 5 minutes. The tubes were centrifuged and speed vacuumed and re-suspended in 50 ul PCR-grade water. All centrifugation steps were performed under 12000 rpm.

3.7.2. Miniprep kit purification of DNA

To extract plasmid DNA from bacterial cultures, Sigma Aldrich or NEB kits manufacturer's protocol was used. A 5 ml overnight culture used to isolate the plasmid, and the DNA eluted in 50 µl PCR grade water. The concentration of the DNA was measured by using Nanodrop NA1000 spectrophotometer, and was directly used or kept at -20 ˚C for subsequent experiments.

3.7.3. Midiprep kit purification of DNA

If a bigger quantity of plasmid DNA was required, the midiprep method using the Sigma Aldrich kit according to the manufacturer's instruction was performed. A 50 ml overnight bacterial culture was prepared, and the plasmid DNA was eluted into 1 ml using PCR grade water. Sample DNA was quantified using Nanodrop NA1000 spectrophotometer, and was used directly or stored at -20 ˚C for future experiments.

3.8. Isolation of bacterial genomic DNA

3.8.1. Colony-PCR

A part of a colony was picked up from an agar plate, and re-suspended in 30 µl PCR grade water. Then, the suspension was placed in a heat block at 100 ˚C for 10 minutes. After vortexing the tube, 2 µl was used as a template for the PCR reaction.

3.8.2. Culture-PCR

Another successful method to extract crude genomic DNA was to use a 2-ml bacterial culture grown overnight. Twenty microlitres of the culture was added to 180

µl PCR grade water in a micro centrifuge tube. After centrifugation at 12,000 g for 1 minute, the supernatant was discarded, and the pellet was re-suspended in 100 µl PCR grade water. Then, the tube was heated at 100 ˚C for 10 minutes. After vortexing for a few seconds, the tube was put on ice and 2 µl was used as a template in the PCR reaction.

3.8.3. Bacterial genomic DNA isolation kit

Bacterial genomic DNA was extracted using a GenElute Bacterial Genomic DNA kit (Sigma Aldrich) according to the manufacturer's protocol. A 5 ml overnight bacterial culture was used in this method, and the DNA was eluted into 100 µl PCR grade water. Samples were used immediately or stored at -20 ˚C for future experiments.

3.9. DNA extraction from a gel

Using the GenElute Gel Extraction kit (Sigma Aldrich), the required DNA band was cut from Agarose gel with multiple bands according to the manufacturer's protocol. The DNA band with the correct size was identified using a UV- light transilluminator, and cut using a sterilized scalpel. DNA was eluted into 50 µl PCR grade water, and was immediately used or stored at -20 ˚C for future experiments.

3.10. Purification of PCR products

Following successful Agarose gel image, amplified DNA was cleaned up using a Sigma Aldrich PCR Clean Up kit or NEB PCR Clean Up kit to remove unwanted substances that may interfere with downstream steps according to the manufacturer's protocol. The DNA was eluted into a specific amount of PCR grade water, and was used immediately or stored at -20 ˚C for future experiments.

3.11. Concentrating DNA by ethanol precipitation

In order to get more concentrated DNA and to remove salts that may affect electroporation, ethanol precipitation was used. To a 50 µl PCR product, 5 µl of 3 M NaAc (pH= 5.2), and 140 µl of 100 % ethanol was added. The solution was left at

room temperature for 30 minutes for the DNA to precipitate, and centrifuged at 12,000 g for 15 minutes. Supernatant was discarded, and the pellet was washed with 500 µl of 70 % ethanol and centrifuged at 12,000 g for 10 minutes. Then, again the supernatant was discarded and the pellet dried using a speed vacuum at a temperature of -100 ˚C. The pellet was re-suspended by adding 15 µl PCR grade water.

3.12. Molecular cloning

3.12.1. Restriction Digest

Plasmid DNA and PCR products were digested using Promega restriction enzymes according to the manufacturer's protocol to prepare them for ligation in classical cloning experiments. Solutions were incubated for three hours at 37 ˚C, and insert DNA was purified using the PCR clean up kit (Sigma Aldrich). Digested vectors were run on agarose gel and the required bands were cut and extracted using a gel extraction kit (Sigma Aldrich). Vectors were also dephosphorylated by treating with Thermosensitive Alkaline Phosphatase (TSAP) from Promega for 30 minutes at 37 ˚C, and then inactivated for 15 minutes at 74 ˚C. Then, the DNA was ethanol precipitated and 5 µl sample was run on 1 % Agarose gel for confirmation before proceeding to the ligation step.

Restriction digest was also used to check successful recombinants after cloning. In this case, plasmid DNA was incubated with the restriction enzyme, specific buffer and bovine serum albumin and water was added to a final volume of 20 µl for one hour at 37 ˚C. Then, the whole sample was run on 1 % Agarose gel.

3.12.2. DNA ligation

Dephosphorylated vector DNA was ligated with insert DNA using T4 DNA Ligase from Promega in a 20 µl mixture. Phosphorylated vector and dephosphorylated

vector's DNA without insert were also used in ligation step as controls. The mixture was incubated overnight at 4 ˚C before transforming it into competent cells.

3.12.3. Gibson assembly

Another method of DNA cloning was used to ligate fragments of DNA by using the Gibson Assembly Cloning Kit from NEB according to the manufacturer's protocol. PCR reactions were made to synthesize DNA for inserts and vectors. Vector DNA was treated with DpnI for 30 minutes at 37 ˚C to destroy background template. Then, all PCR products were cleaned up, and DNA was quantified by Nanodrop NA-1000. Different factors including the size of DNA fragments, DNA concentrations and molarity were taken into consideration when calculating the volume of each DNA fragment to be added to the assembly reaction. The tubes were incubated later in the thermocycler for 15-60 minutes at 50 ˚C. After incubation, 10 µl of the mix were transformed chemically into competent cells provided with the kit, 200 µl and 200 µl of 1:50 dilution were plated out onto selection plates and incubated overnight at 37 ˚C.

As a preliminary confirmation step, cultures were made, and plasmids were extracted using the extraction kit. Plasmids were digested using restriction enzymes, and agarose gel analysis was performed to check the obtained fragments size. Sometimes, PCR reactions using checking primers were made as another way of confirming cloning success. The final step included sending samples off for sequencing to Source Bioscience. Following successful cloning, plasmids were transformed into host cells to measure biochemical activities of different strains.

3.13. DNA Transformation

Plasmids and DNA segments were transformed in two ways: chemical transformation and electroporation. In chemical transformation, DNA was transformed into calcium chloride treated *E. coli* competent cells by exposing the cells to heat shock at 42 ˚C

for 50 seconds. While in electroporation, DNA was transformed into *E. coli* competent cells by shocking them with a high voltage.

3.13.1. Preparation of competent E. coli cells for heat-shock

A 5 ml LB bacterial pre-culture was inoculated with a colony taken from a freshly streaked agar plate, and grown overnight on an orbital shaker at 37 ˚C. Next day, a 200 ml LB culture was inoculated from pre-culture to a starting OD_{600} of 0.05 and incubated at 37 ˚C until they reach OD_{600} of 0.1-0.2. Then, the culture was distributed into four 50 ml sterile falcon tubes of 50 ml each and centrifuged at 4 ˚C for 10 minutes at 4500 rpm. The supernatant was discarded and the pellet was re-suspended with 25 ml pre-chilled 0.1 M calcium chloride ($CaCl_2$), and located on ice for 40 minutes. The tubes were centrifuged again under the same conditions, and supernatant was discarded. The pellets were re-suspended with the left over liquid and pooled together in one pre-chilled tube. To the cell concentrate, 100 % glycerol was added to give a final concentration of 10 %. Aliquots were made of 100 µl portion into 1.5 ml sterile microfuge tubes, and shock freezed in liquid nitrogen for 2 minutes. Samples were stored immediately at -80 ˚C until use.

The same method was used to prepare chemically competent cells to use on the day. The bacterial cultures were made of 25 ml LB instead of 200 ml, and cells were harnessed directly for transforming DNA without proceeding to the freezing step.

A modification was applied to this method when transforming plasmids to clinical isolates competent cells needed. An extra step was performed before harvesting the cells was the incubation of cells at 50 ˚C for 15 minutes.

3.13.2. Chemical transformation of DNA (heat shock)

Chemically competent cells were either brought from -80 ˚C freezer or prepared on the day were located on ice for 30 minutes. A DNA volume of 1-20 µl was added to the cells and the liquid was gently mixed by pipetting up and down and left on ice for

30 minutes to settle down. Negative controls were treated in the same way, but without adding DNA. Then, the cells were heat shocked by placing the tubes in a water bath at 42-45 ˚C for 50 seconds and transferred back on the ice for 2 minutes before adding 900 µl of LB. Later, the tubes were incubated at 37 ˚C for 45-60 minutes on an orbital shaker. Culture volumes of 200 µl and 200 µl of 1:50 dilution were plated out onto LB agar containing the appropriate antibiotic or X-gal for blue-white screening, and plates were incubated overnight.

In Gibson cloning, *E. coli* competent cells provided with Gibson Cloning kit from NEB were used. Tubes of cells were brought from -80 ˚C freezer and thawed on ice, and heat shock transformation was performed according to the manufacturer's protocol.

3.13.3. Preparation of competent E. coli cells for electroporation

Bacterial pre-cultures of 5 ml LB -/+ antibiotic were incubated overnight at the correct temperature. Next day, 25 ml LB were inoculated from the pre-culture to a starting OD_{600} of 0.05 and incubated at the correct temperature on an orbital shaker. When the culture reached an OD_{600} of 0.6-0.8 (incubation at 50 ˚C for 15 minutes was performed when electroporating a plasmid to a clinical isolate), it was transferred to a 50 ml falcon tube and centrifuged at 4 ˚C for 10 minutes at 4500 rpm. Subsequently, the supernatant was discarded and the pellet was re-suspended with 25 ml pre-chilled sterile water and centrifuged under the same conditions. Another washing step was performed, and the pellet was re-suspended in the left over liquid. When competent cells needed to be frozen at -80 ˚C in 50 µl portions, cells were washed with pre-chilled sterile water containing 10 % glycerol.

3.13.4. Transforming DNA by electroporation

A portion of 40 µl cell suspension was transferred to a 1.5 ml microfuge tube, and PCR or plasmid DNA was added to the cells. The solution was mixed by pipetting up and down and left on ice to settle for 5 minutes. Then, the mixture was transferred to

a pre-chilled sterile electroporation cuvette (Flowgen Bioscience) and located in the electroporator (BIO-RAD MicroPulser) using programme 1 (1.8 kV, one pulse for 4.8 ms). Then, the cells were pulsed with electricity and 1 ml of LB was added immediately to the mixture which was transferred back to a sterile microfuge tube. Later, the tubes were incubated at the correct temperature for 45-60 minutes and volumes of 200 µl and 200 µl of 1:50 dilution were plated out onto LB agar containing the selective antibiotic, and plates were incubated overnight.

3.14. DNA recombination

3.14.1. Datsenko and Wanner method

The Datsenko and Wanner method was used to delete targeted genes performed according to their one-step inactivation of chromosomal gene method (Datsenko and Wanner, 2000). Primers were designed to amplify regions in the plasmid pKD3 containing a chloramphenicol resistance gene that have overlapping ends of homology to the chromosomal sites upstream and downstream of the genes to be deleted. Bacterial cells that contain the gene to be deleted were prepared for electroporation, and the plasmid pKD46 was transformed to these cells. Successful selection was made by growing the cells on LB+Amp agar plates at 30 ˚C overnight. Cultures of pKD46 transformed strains were made from a single colony, and 0.1 % arabinose was added to LB+Amp broth when the OD_{600}= 0.1. Then, 3 µl of PCR product amplified from pKD3 were electroporated. Following 45-60 minutes incubation at 30 ˚C and overnight incubation at 37 ˚C, recombinants were selected on LB+Cm agar plates. Checking primers were used to confirm the successful outcome.

3.14.2. Blank method

Targeted mutagenesis of bacterial chromosome was performed on bacterial DNA using the λ red recombinase system with I-SceI endonuclease. The method of Blank *et al*., 2011 was used to replace DNA fragments. The first step involved the PCR amplification of a region from the plasmid pWRG100 using oligonucleotides to amplify a DNA cassette that contains a I-SceI recognition site and a chloramphenicol resistance gene. Bacterial cells containing pKD46 were prepared for electroporation of the PCR product. This generated an in-frame deletion with the help of the temperature-sensitive plasmid pKD46. The resulting colonies were then tested for their chloramphenicol resistance by growing them on LB plates containing this antibiotic at 37 ˚C overnight to select successful recombinants.

The temperature-sensitive, ampicillin resistant plasmid pWRG99 was electroporated to these Cm^R strains, and cells incubated at 30 ˚C. The plasmid pWRG99 has the same features of pKD46, but with I-SceI endonuclease activity which can be induced by tetracycline addition. Therefore, colonies were screened on LB+Amp and LB+Amp+chlorTet plates to identify transformants.

In the second step, the I-SceI selection process was maintained by growing bacteria on media containing ampicillin and arabinose. Then, new PCR products were electroporated into these cells and the bacteria were plated out onto LB+Amp+ cTet for plasmids selection. Later, cells were grown onto LB, LB+Amp and LB+Cm, and only bacteria that were sensitive to both ampicillin and chloramphenicol were picked from corresponding colonies on LB plates. PCR products encompassing the replaced regions were also sent off for sequencing to Source Bioscience for final verification.

3.14.3. CRISPR-Cas9 method

The clustered regularly interspaced short palindromic repeats-CRISPR-associated system (CRISPR-Cas system) was used as precise genome engineering technology to delete and insert genes in *E. coli* (Jiang *et al.*, 2015). The first heat sensitive helper plasmid pCas9 was transformed to the bacterial strain with a DNA Cm-resistant cassette, and transformants were selected on LB+Kan agar plates. The pCas9 plasmid has the endonuclease activity and lambda red recombinase system which can be induced by adding 0.1 % arabinose to the culture. The genes to be inserted were amplified using PCR. importantly, PCR products were cleaned up using Sigma clean up kit, and quantified using nanodrop.

A DNA of 300-400 ng of PCR product was electroporated along with 100 ng of a second helper plasmid pTrg-Cm to the cells with pCas9. The second helper plasmid encodes for the short guide RNA to guide Cas9 protein to the target DNA sequence, in this case the chloramphenicol resistance gene *cat*. After incubation for 60 minutes at 30 ˚C, cells were grown on LB+Kanamycin/ Spectinomycin agar plates overnight. The next step included checking colonies on LB and LB+Cm agar plates, and Cm sensitive colonies were chosen for colony PCR verification.

3.15. Measuring bioluminescence activity

Microplate assay was used to detect the expression of bioluminescence of the vector pSB401 that has a promoter that initiates the transcription of *lux* genes. In this procedure, 3 ml LB cultures of each of the transformed strains were made in triplicates with tetracycline, and placed on an orbital shaker at 37 ˚C overnight. Next day, new 3 ml cultures with tetracycline were made from pre-cultures to a starting OD_{600} of 0.05. New cultures were incubated at 37 ˚C until they reached an OD_{600} of 0.5-0.8. Then, 200 µl of each culture were transferred to a well of the 96 well plate, and 3-5 samples of fresh LB were added as a blank control. After inserting the plate

in the microplate reader FLUOstar OPTIMA (BMG LABTECH), bioluminescence as well as optical density were measured independently using Optima software and relative values were obtained. The average activity of each strain and the standard deviation were calculated, the bar chart was drawn in excel.

3.16. Fluorescence microscopy

Bacterial pre-cultures of 3 ml TB (-/+antibiotic) were made in triplicates and incubated on an orbital shaker at 30 ˚C overnight. Next day, new 3 ml TB (-/+antibiotic) were made in triplicates to a starting OD_{600} of 0.05 and incubated at 30 ˚C until they reached an OD_{600} of 0.5-0.8.

Agarose was added to sterile milliQ water to a concentration of 1 % and dissolved by heating in a microwave. A volume of 500 µl of this motility buffer was pipetted onto a Hendley- Essex multispot microscope slide and covered by a plain microscope slide (VWR). The cover was pressed gently to remove any air bubbles and to spread the agarose evenly on the slide. The slides were left for 30 minutes to allow the agarose to solidify, and then the cover was removed by sliding on the bottom one. A 2 µl of the bacterial culture was added to a spot of the slide and put in an oven at 42 ˚C for 2 minutes to dry, and covered with a cover slip. Images were taken using 100x 1.30 oil objective lenses of Nikon ECLIPSE Ti and acquired using the MetaMorph v7.7.80 software. A total of 3-5 images were taken for each culture with exposure time of 100 ms for the phase contrast wavelength, and 1000 ms for GFP 507 and YFP wavelengths.

3.17. Image analysis

Fluorescent microscope images have been analysed using oufti.org software from oufti.org. Cells defined in the phase contrast image using *E. coli* LB parameters adjusted to the following values that come in oufti.org:

Threshold factor M	0.96
Threshold Min level	0.7
Edge sigma L	1.46
Edge mode	Loa

Table. 3.3. Parameters of oufti.org software.

While spots were detected with parameters set as the following:

Wavelet scale	1
Low pass	2
Spot radius	3
Int. threshold	0.6
Min region size	8
Min height	0.001
Min width	0.5
Max width	10
Adjusted square error	0.45

Table. 3.4. Parameters for spot detection of oufti.org.

If needed, spot detection was then manually assessed using spot viewer. Data was then processed in excel.

3.18. Multilocus sequence typing (MLST)

A collection of 186 strains of *E. coli* was given by Mark Fielder (Kingston University/ London). The isolates were obtained at a depth of 30-50 cm from rectal examination of cows in several herds across the South of England. Culture-PCRs were performed on these isolates to extract the genomic DNA to use in the PCR reactions. Sets of 7 pairs of primers were designed according to The University of Warwick Medical School website:

(http://mlst.warwick.ac.uk/mlst/dbs/Ecoli/documents/primersColi_html) to amplify internal fragments of 7 house-keeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*) for each strain. PCR reactions were made in 96-well plates covered by sticky aluminum sheets using the thermocycler TECHNE TC-512. PCR conditions were 2 minutes at 95 ˚C, 30 cycles of 1 minute at 95 ˚C, 1 minute at the gene's annealing temperature, 2 minutes at 72 ˚C then 5 minutes at 72 ˚C and Taq polymerase was used. Subsequently, 5 µl of each sample were run on 1 % agarose gel to verify the

correct size. Successful PCR products were purified using the PCR clean up kit (NEB), and failed samples were repeated using the same conditions. Samples were then sent off for sequencing with either forward or reverse primer to Source Bioscience at a final concentration of 5-10 ng/µl. Upon receiving sequencing data, the sequences were arranged in a word file. By using the website (pubmlst.org), the data was pasted in the specified field in the batch query page and allele numbers were recorded for each of the fragments. Then, exact allele numbers for the 7 genes of each strain were entered into: (Search MLST locus/sequence definitions database by combinations of loci) in the same website to obtain the sequence type of each strain.

3.19. MLST alignment

Using the defined sequence types identified from the MLST sequencing data, pubmlst.org was used to download the consensus, concatenated sequence of each sequence type. Each sequence type was added to one FASTA file that included the sequence type data from the genomes used by McNally *et al.* (2013). These control sequences are included to provide focal points in the alignment analysis. The FASTA formatted sequence data was then aligned using CLUSTAL-X generating an alignment output. The .aln clustal file was then imported into Jalview (jalview.org) to allow for the generation of a Nexus .tre file not available in CLUSTAL-X. Each .tre file was then imported into the statistical maths platform R, and the work package “ape” used to generate a radial phylogenetic tree of the MLST data. For presentation purposes, the phylogenetic tree was exported as a PDF and colours added to the tree diagram using adobe illustrator.

3.20. Biofilm formation

Cow strains as well as controls were grown on LB agar at 37 ˚C overnight. Next day, 2 ml LBnoNaCl cultures were made and incubated at 37 ˚C overnight. A volume of 10

µl of each culture was pipetted onto LBnoNaCl agar and incubated at 28 ˚C for 3-5 days in a box with a water beaker to prevent drying out. Images were taken for the plates using the Syngene Bioimaging cabinet and GeneSnap software at 75 ms, and phenotypes were scored based on a 1 to 6 scale.

3.21. Statistical analysis

Statistical analysis performed on the data of this work using t-test and analysis of variance (ANOVA).

Chapter Four: Species Wide Analysis of *Escherichia coli flhDC* Promoter Activity and Regulation

4.1. Introduction

This study has used the McNally's data set as the foundation of *E. coli* phylogeny as the genomes used were accurately declared. Whole genome analysis of *E. coli* revealed six phylogenetic lineages: A, B1, B2, D, E and F according to McNally *et al.* (2013) (Figure. 1.3). The initial idea of this project emerged from the fact that the vast majority of the literature on the *E. coli* flagellar system referred to one strain only, MG1655. Strain MG1655 is a well-established lab model strain, belonging to clade A. We were interested in working with other clades of *E. coli* regarding the analysis of a 780 bp fragment of the genome that leads to motility of this species of bacteria: the *yecG-flhDC* intergenic region.

The master regulator ($FlhD_4C_2$) is expressed upon internal and external cues such as food, temperature and pH. As a result, $FlhD_4C_2$ binds with sigma factor σ^{70} to activate class 2 promoters. The class 2 dependent genes are responsible for the assembly of the hook-basal body and the regulators FlgM and FliA (σ^{28}). Upon the completion of the hook-basal body, FlgM will be secreted out of the cell freeing σ^{28} to bind to the RNA polymerase at the class 3 promoters. Class 3 promoters are responsible for the expression of the filament and motor genes (Figure. 1.15).

We have asked: what impact replacing the *yecG-flhDC* intergenic region with model sequence types has upon motility in the *E. coli* strain MG1655? We know that transcription is hierarchical with *flhDC* at the top. Therefore, to investigate flagellar gene expression and answer the above question relating to motility we were appreciating *flhDC* expression across *E. coli* strains. We showed that 1) the different *flhDC* promoters sustained motility in MG1655; 2) we measured *flhDC* transcription; 3) we measured the effect of different media on these promoters; 4) we tested the impact of known global regulators on the motility; and 5) we determined the distribution of flagella on *E. coli* cells.

This list was itself a significant task to complete and was a central component of the project. However, by developing the necessary tools to explore the above aspects and generating data in model isolates of *E. coli* we have created strains and reporters that can be exploited to fulfil this task. The lab has a long track record in measuring flagellar gene expression and visualising flagellar assembly in bacterial cells. We harnessed a range of molecular based techniques that complement bacterial cell biology such as advanced bacterial genetics and bioinformatics. Our primary directions to do so included the use of a luciferase-based system to measure flagella gene expression. We also utilized fluorescent protein tags to flagellar subunits to visualise and count where flagella are on the cell.

4.2. Results and discussion

4.2.1. Types of the flhDC promoter in E. coli

We were initially interested in studying the *flhDC* promoter as it represents the first step in flagellar assembly and regulation hierarchy or class 1 promoter. Therefore, we sequenced and aligned the *yecG-flhDC* intergenic region of 51 clinical isolates in the Aldridge's lab (Picton and Aldridge, unpublished). This, then, led us to analyse the *yecG-flhDC* intergenic region taken from the *E. coli* genomes used by McNally *et al.* (2013) (Figure. 4.1A). Bioinformatics showed that this region reflected the phylogenetic groups of *E. coli*. However, three dominant sequence types of the *yecG-flhDC* intergenic region have been mapped into three groups, A/B1, B2 and D/E/F (Figure. 4.1B).

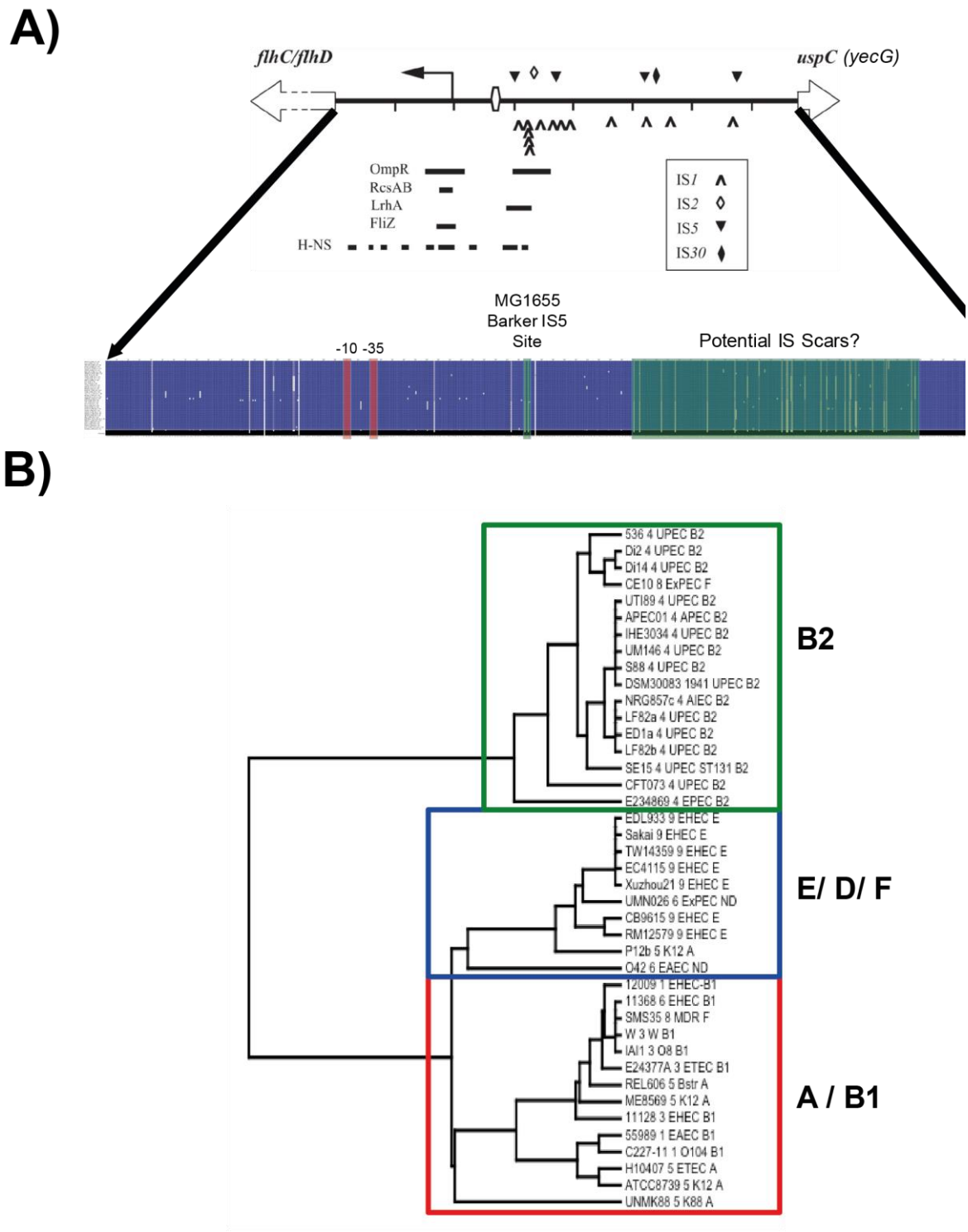


Figure. 4.1. DNA sequence alignment of the *yecG-flhDC* intergenic region in *E. coli*.

A) Alignment of the *yecG-flhDC* intergenic region of *E. coli* strains based on Fahrner and Berg (2015). **B)** The alignment of the *yecG-flhDC* intergenic region of McNally's *E. coli* genomes reflected three dominant types of the *yecG-flhDC* intergenic region: A/B1, B2 and E/D/F.

4.2.2. Replacing the flhDC promoter of strain MG1655

Our lab is specialised in motility aspects of bacteria with a huge collection which consists of thousands of strains. We used strain MG1655 from this collection which is known to be very weakly motile. We wanted to know the impact of replacing the *flhDC* promoter in MG1655 which belongs to clade A with other *flhDC* promoters from strains known to be motile. For this reason, we have chosen three motile isolates descending from different phylogenetic groups B2, B1 and D of *E. coli* (Figure. 4.2A). These model sequence types were picked from our strain collection, strain CFT073 is a model strain isolated in the UK from a case of pyelonephritis used in UTI experiments, belonging to clade B2. The other two strains of clades B1 and D are isolated from acute UTI and asymptomatic bacteriuria cases, respectively.

Manipulation was performed on chromosomal bacterial DNA using Blank *et al.* (2011) method (Figure. 4.2B). This technique has been used initially to create recombinant strains in *Salmonella enterica* proving to be efficient and time saving when used to create new *E. coli* strains. The λ red recombinase system with I-SceI endonuclease of the heat sensitive plasmids pKD46 and pWRG99 were used for this purpose with a PCR product. The in-frame deletion was achieved by adding arabinose to induce the λ-red system, and resulting strains were nonmotile (Figure. 4.2C). Then, the DNA cassette from the first step was replaced with *flhDC* promoters from three isolates belonging to groups B2, B1 and D and motile isolates were obtained (Figure. 4.2D).

In order to quantify motility of newly created strains and compare them to the motility phenotype of strain MG1655, these strains were grown on semi-solid motility agar for 8 hours at 37 ˚C. This medium, which is low in carbon source, allows cells to swim away from the stabbing point as they are induced to form flagella, but strain MG1655 was still weakly motile (Figure. 4.3A). However, newly created strains which

have the same background as MG1655 with only the *yecG-flhDC* intergenic region replaced were highly motile with swarming diameters of around 3 cm (Figure. 4.3B). This experiment led us to conclude that it is possible to replace the *flhDC* promoter between different clades and maintain flagellar formation. It also encouraged us to measure the activity of the different types of the *flhDC* promoters on the chromosome using the same isogenic background strain.

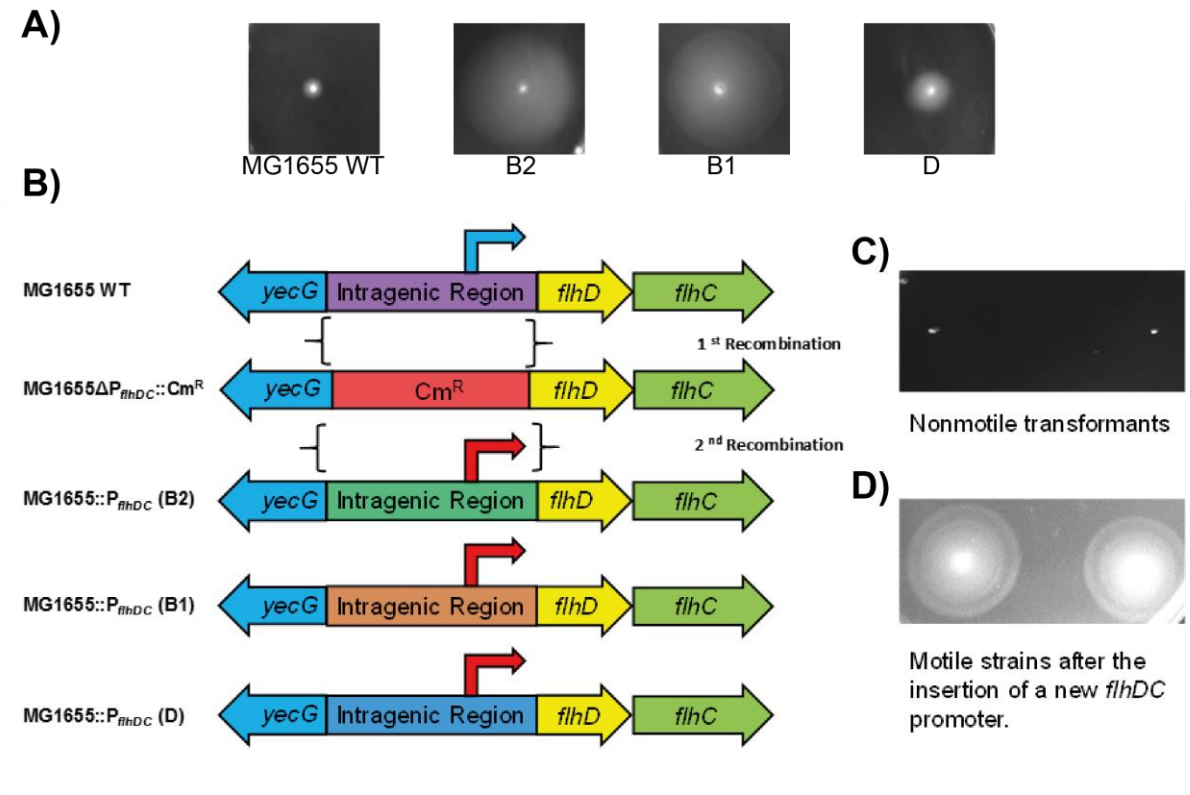


Figure. 4.2. Method of Blank *et al.* (2011) for the *yecG-flhDC* intergenic region replacement in strain MG1655.

A) Motility phenotype of strain MG1655 and the original *flhDC* promoter donor strains that belong to different clades**. B)** Schematic representation showing the two-step DNA manipulation. First step included the replacement of the *yecG-flhDC* intergenic region of strain MG1655 with chloramphenicol resistance (Cm^R) cassette. Second step was performed to replace the Cm^R cassette with *yecG-flhDC* intergenic regions amplified from three *E. coli* strains representing B2, B1 and D clades. **C)** Nonmotile transformants obtained after the first step by stabbing single colonies in motility plates containing 0.3 % agar for 8 hours at 37 ˚C. **D)** Motile strains obtained after the second step by stabbing single colonies in motility plates containing 0.3 % agar for 8 hours at 37 ˚C.

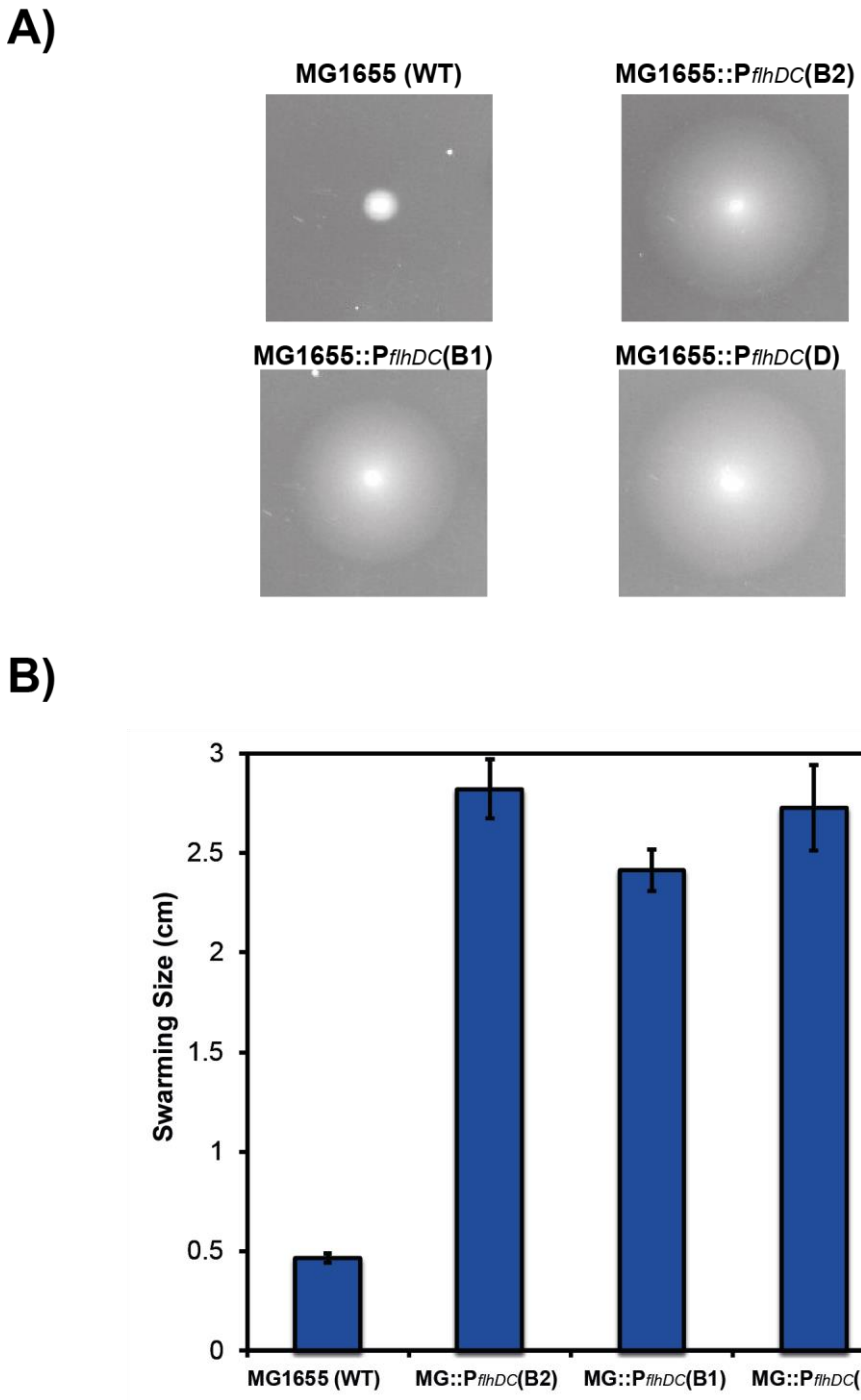


Figure. 4.3. Motility quantification of strain MG1655 before and after *yecG-flhDC* replacement.

A) Motility phenotypes of strain MG1655 (WT), and after replacement of its *yecG-flhDC* intergenic region with three regions from B2, B1 and D clades. Single colonies were stabbed into motility plates containing 0.3 % agar for 8 hours at 37 ˚C. **B)** The average of swarming diameter in centimeter of strain MG1655 wild type, and after its *yecG-flhDC* intergenic region replacement. The values are mean ± standard deviation of three independent experiments.

4.2.3. Catabolite repression

This experiment was performed in order to detect if there was any change in *flhDC* expression following the alteration in the global regulatory protein cAMP-CAP (catabolite gene activator protein) concentration. The cAMP-CAP complex binds to the flagellar class 1 promoter and positively regulates the process of flagellar assembly (Soutourina *et al.*, 1999). Four different strains of *E. coli* that have different P_{flhDC} were grown on LB medium and a poor carbon source TB medium (Tryptone as a sole source of micronutrients and lacking yeast extract) in four different manners as per the following:

1- LB ⟶ LB

2- LB ⟶ TB

3- TB ⟶ LB

4- TB ⟶ TB

Theoretically, when the bacteria grow on a poor carbon source as a secondary growth medium, high cAMP amount will be synthesized, and cAMP-CAP will be formed (Notley-McRobb *et al.*, 1997). As a result, bacteria will increase flagellar synthesis and become increasingly motile (Soutourina *et al.*, 1999). However, no change in motility phenotype was observed when the bacteria examined under the microscope (Table. 4.1). This experiment suggests that our strain MG1665 is unable to alter its *flhDC* expression under different growth conditions. The reason is probably due to its DNA topology upstream of the *flhDC* promoter.

Strain	Primary Growth Medium	Secondary Growth Medium	Motility Phenotype
MG1655ΔP*flhDC*::P*flhDC*(B2)	**LB**	**TB**	**+**
	TB	**LB**	**+**
	LB	**LB**	**+**
	TB	**TB**	**+**
MG1655ΔP*flhDC*::P*flhDC*(B1)	**LB**	**TB**	**+**
	TB	**LB**	**+**
	LB	**LB**	**+**
	TB	**TB**	**+**
MG1655ΔP*flhDC*::P*flhDC*(D)	**LB**	**TB**	**+**
	TB	**LB**	**+**
	LB	**LB**	**+**
	TB	**TB**	**+**
MG1655 (WT)	**LB**	**TB**	**+/-**
	TB	**LB**	**+/-**
	LB	**LB**	**+/-**
	TB	**TB**	**+/-**

Table. 4.1. The motility phenotype of strain MG1655 and its derivatives after growing on a different combination of LB and TB media.

4.2.4. The flhDC promoters on plasmid pSB401

Inserting the *flhDC* promoters in the chromosome of MG1655 by replacing its promoter resulted in the alteration of *flhDC* transcription regulation. It was clear that these different promoters influenced motility in a different way from the wild type strain MG1655 when we quantified motility. However, the motility phenotype output of this experiment was due to the whole motility system not the *flhDC* promoter alone. In order to measure *flhDC* promoter strength, we cloned these four different promoters

into the plasmid pSB401. The plasmid pSB401 has been successfully used to determine gene expression in many bacterial systems and in our lab (Winson *et al.*, 1998a; Brown *et al.*, 2008).

This 10.5 kb plasmid has *luxCDABE* reporter genes which synthesize luminescent light when they are expressed (Figure. 4.4) (Winson *et al.*, 1998a). The intensity of the luminescence can be measured in light units. pSB401 also has a tetracycline resistant gene which helps in selecting successful recombinants on the agar plate. Another important feature of pSB401 is the presence of two EcoRI sites ~1 kb apart, flanking the original *luxR* fragment which make them a suitable site for cloning promoter regions (Winson *et al.*, 1998a).

Classical cloning techniques were used to integrate the four *flhDC* promoters into pSB401. The plasmid was treated with EcoRI restriction enzyme to cut out the *luxR* fragment and linearizing the pSB401 vector for subsequent steps. On the other hand, the *flhDC* promoters on the chromosome of MG1655, CFT073, acute UTI and ABU strains were amplified and digested with EcoRI to create complementary ends with the linearized vector. Then, the vector and inserts were ligated together to form new plasmids (Figure. 4.4). The same technique was used to clone the *flhDC* promoters into plasmid pBlueKS ~3.5 kb to the multiple cloning site located within *lacZ* fragment as a control cloning experiment. This vector which has an ampicillin resistant gene was used as an indicator for the success of the cloning experiments through blue-white screening on LB+Amp plates containing XGal. The cells will produce β-galactosidase if the cloning was unsuccessful and colonies pigmented with blue will grow, while in the case of working cloning experiment the *lacZ* region will be interrupted by the insert (*flhDC* promoter in our experiment) preventing the production of β-galactosidase, and white colonies will grow, instead. By comparing the number of the blue colonies to the number of the white ones in the plate, we were

able to predict the success or the failure of the pSB401 cloning. All *flhDC* promoters were cloned to pSB401 and transformed to DH5α cells. When digesting plasmids after the cloning experiment with EcoRI, the correct size fragments were obtained. Still another challenge was to insert the promoters in the correct direction into pSB401. To verify this, further steps were performed to determine the orientation of the insert. This involved the amplification of the promoter fragment from the ends outside EcoRI sites with a pair of checking primers, and running a gel to detect any band in the right size (Figure. 4.5). This considered as one disadvantage of classical cloning compared to Gibson cloning is that both ends in EcoRI sites were complementary, whereas by using Gibson cloning different sites can be created according to the primers synthesized for the vector and insert(s).

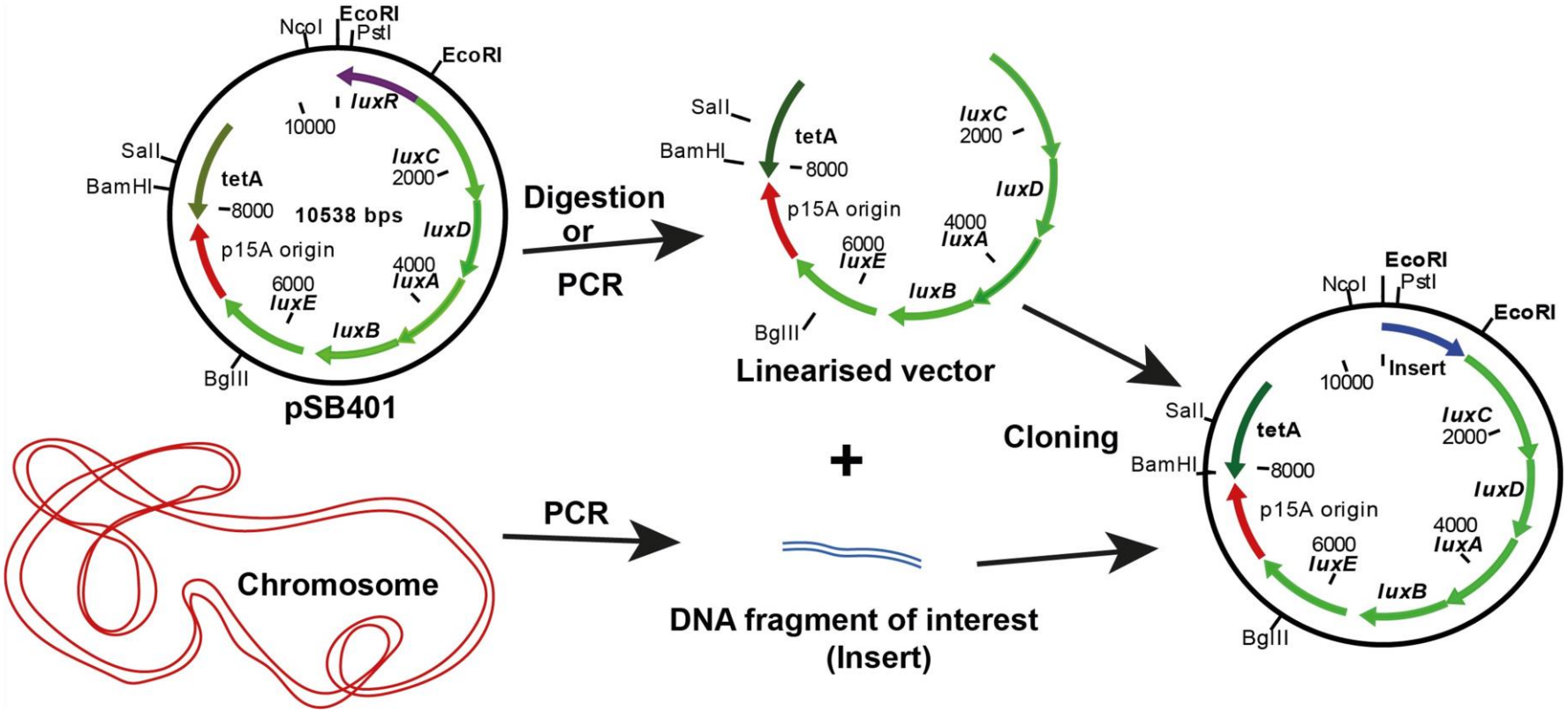


Figure. 4.4. Schematic representation of traditional and Gibson cloning method used for the plasmid pSB401.

The plasmid pSB401 was treated with the restriction enzyme EcoRI to remove *luxR* DNA fragment. On the other hand, the gene of interest was amplified from the chromosome. In the case of the traditional cloning, PCR amplified fragment was treated with EcoRI to create complementary ends with the linearised vector. Linearised vector and amplified DNA were ligated together according to the manufacturer's protocol to form a new plasmid. In Gibson cloning, the assembly master mix and water will be added to the quantified amounts of the vector and insert(s), incubated and transformed according to the manufacturer's protocol.

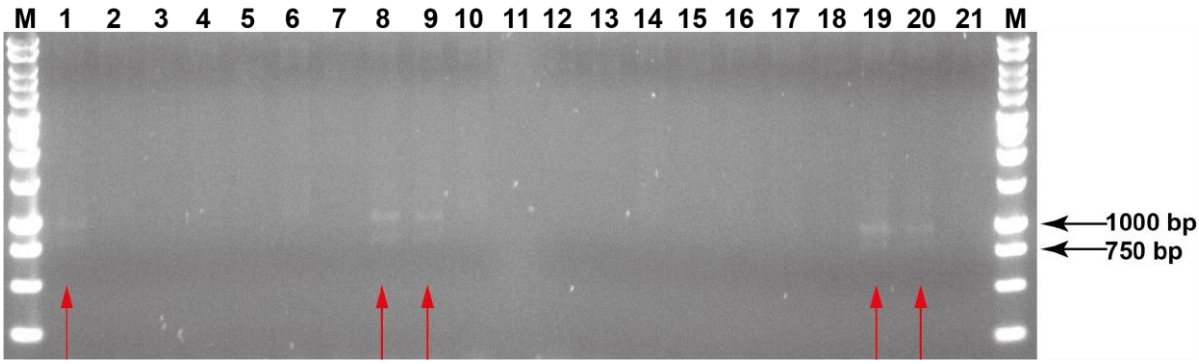


Figure. 4.5. Gel image showing the success of the cloning experiments.

After performing cloning experiments, strains were picked up from the selective plates and the plasmid pSB401 was extracted. Extracted plasmids were digested with EcoRI and samples were run on a 1 % agarose gel to detect the correct band size. Further steps were performed after this procedure to confirm the orientation of the *flhDC* promoter by using a pair of checking primers and running samples on the gel. The bands with the correct orientation are referred to them in the figure with the red arrows (lanes 1, 8, 9, 19 and 20) will have the size of 1 kb. M: DNA ladder.

4.2.5. Measuring cloned flhDC promoters' activity in panel strains

The four different *flhDC* promoters cloned on plasmid pSB401 were transformed into the original three strains created earlier that have the *flhDC* promoters on the chromosome (MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2, B1, D) and MG1655 in two different temperatures 30 ˚C and 37 ˚C. This experiment let us know the feedback of these panel strains on the behaviour of *flhDC* promoters and the impact of the temperature on their activity.

The plasmid pSB-P$_{flhDC}$ (B2) shows strong activity in MG1655::P$_{flhDC}$ (B2), but not in other strains ([all data] P= 0.0001, [MG1655::P$_{flhDC}$ (B1):MG1655::P$_{flhDC}$ (D):MG1655:MG1655ΔP$_{flhDC}$] P= 0.467). The plasmids pSB-P$_{flhDC}$ (B1) and (D) show very little variation in all strains tested ([pSB-P$_{flhDC}$ (B1) in all strains] P= 0.218, [pSB-P$_{flhDC}$ (D) in all strains] P= 0.134). On the other hand, pSB-P$_{flhDC}$ (MG1655) exhibited the greatest level of variation in all strains, whereas it showing the weakest activity in its parent and even versus MG1655ΔP$_{flhDC}$ ([pSB-P$_{flhDC}$ (MG1655) in all strains] P= 0.006, [in strains B1:D:MG1655:MG1655ΔP$_{flhDC}$] P= 0.0297, [MG1655:MG1655ΔP$_{flhDC}$] P= 0.0286, but [strains B1:D] P= 0.117 (NS)) (Figure 4.6). The activity of the promoters in the panel strains at 37 ˚C were low in general (Figure. 4.7). All data for pSB-P$_{flhDC}$ (B2), (MG1655) and (D) are not significant (P= 0.183, P= 0.824 and P= 0.147, respectively). The data for pSB-P$_{flhDC}$ (B1) is significant (P= 0.03) due to the activity of this promoter in strain MG1655ΔP$_{flhDC}$ is higher than the activity in other strains. When we exclude strain MG1655ΔP$_{flhDC}$ the difference will not be significant (P= 0.263) (Figure. 4.7). However, the drop from 30 ˚C to 37 ˚C is significant (pSB-P$_{flhDC}$ (B2)30 ˚C:37 ˚C is significant (P= 0.0009)). Generally, the data from this experiment reflected a link between promoters' activity and motility phenotype of the panel strains. The experiment showed evident positive regulation of

the *flhDC* promoters at 30 ˚C. Thus, it suggests that thermoregulation is conserved across the clades of *E. coli*.

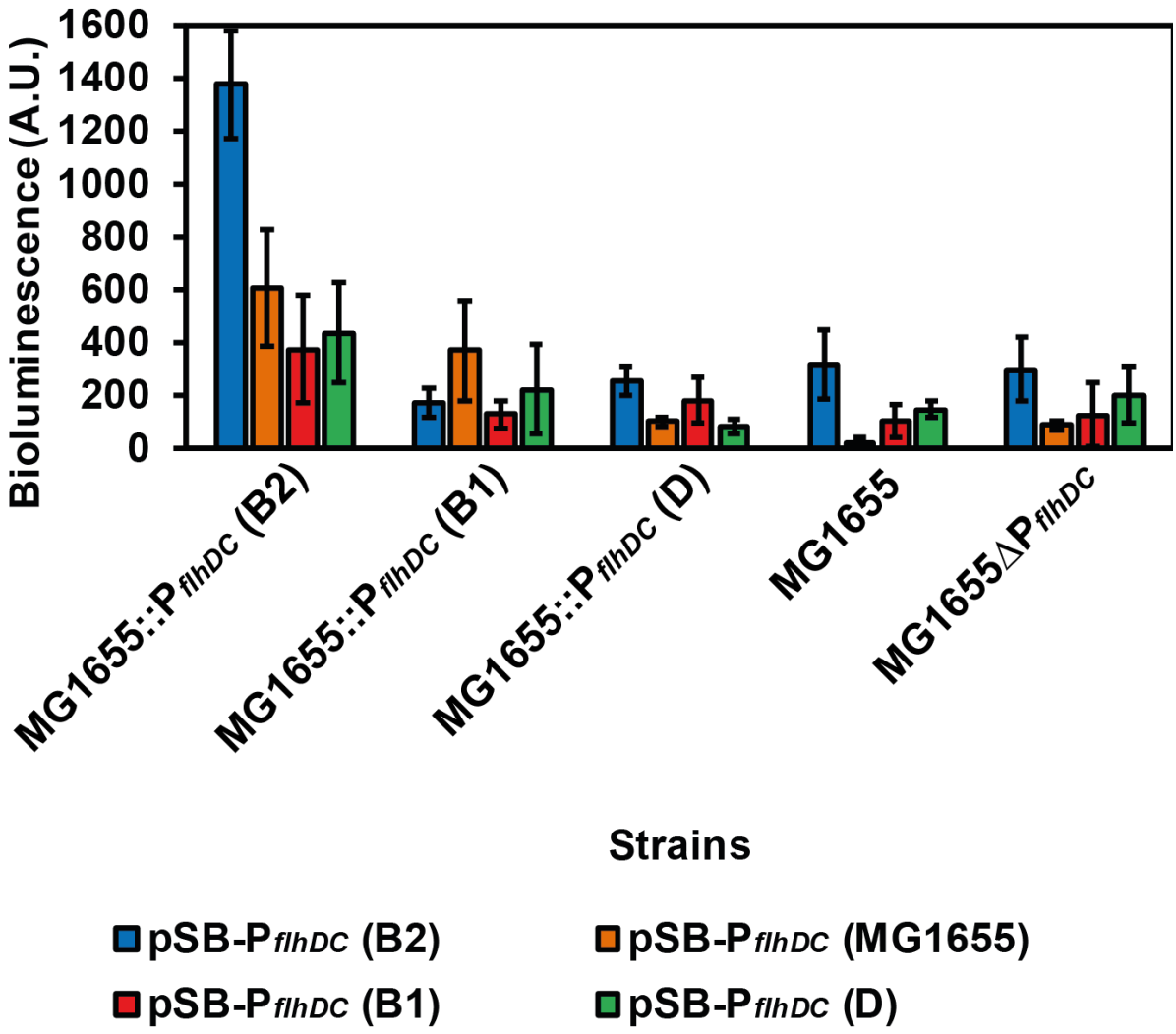


Figure. 4.6. Activity of the *flhDC* promoters from the original strains in MG1655 derivatives at 30 ˚C.

The four *flhDC* promoters representing A, B2, B1 and D clades cloned into pSB401 were transformed into strains MG1655 and MG1655ΔP_{flhDC}::P_{flhDC}-B2, B1, D. Colonies were grown in LB cultures until OD_{600}= 0.6-0.8. The activity of *flhDC* promoters was measured by microplate reader. The *flhDC* promoter from clade B2 scored the highest values among other promoters. The data reflected a link between *flhDC* promoter activity and motility phenotype of the panel strains. The values are mean ± standard deviation of three independent experiments.

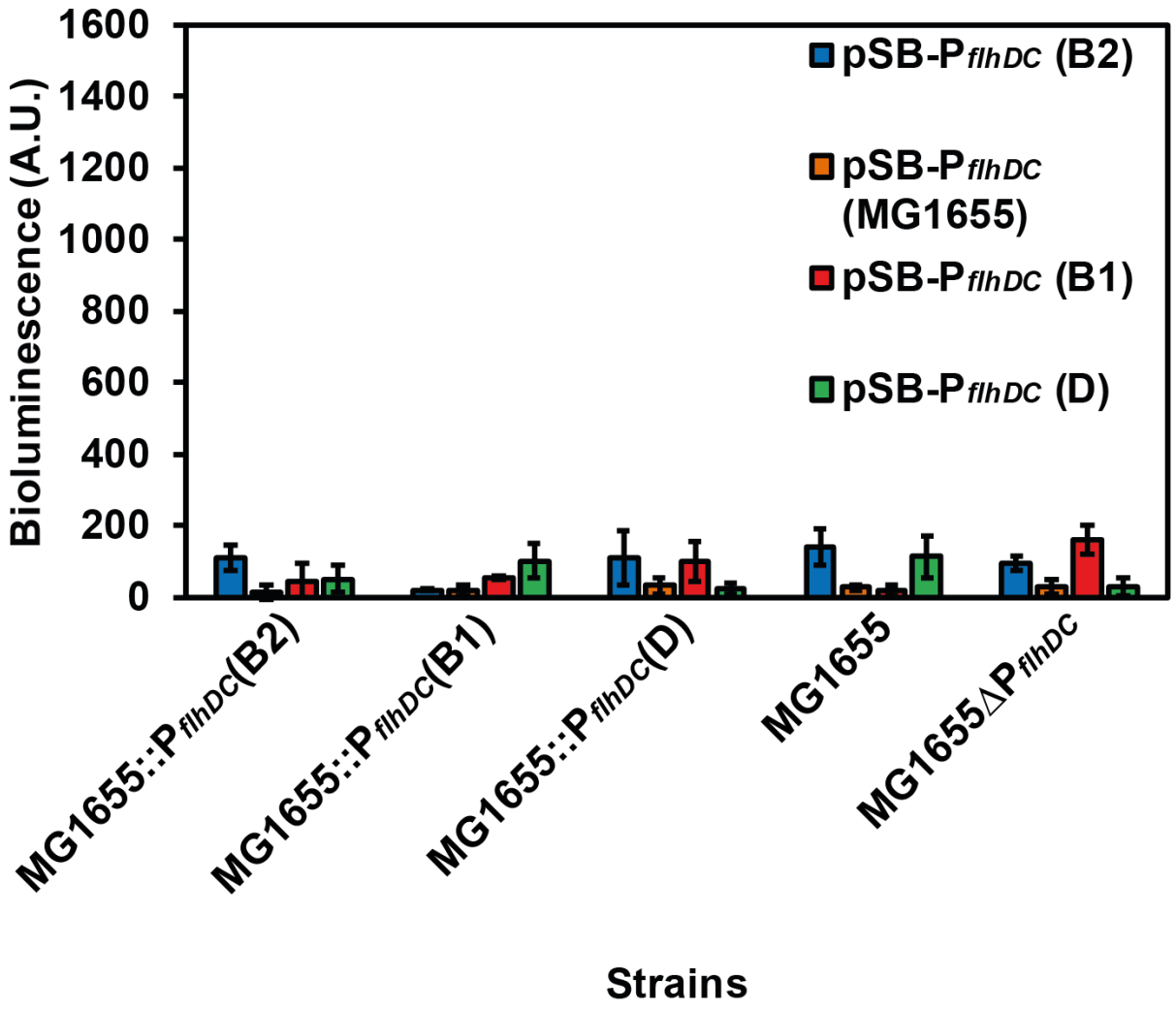


Figure. 4.7. Activity of the *flhDC* promoters from the original strains in MG1655 derivatives at 37 ˚C.

The four *flhDC* promoters representing A, B2, B1 and D clades cloned into pSB401 were transformed into strains MG1655 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2, B1, D. Colonies were grown in LB cultures until OD$_{600}$= 0.6-0.8. The activity of *flhDC* promoters was measured by microplate reader. In general, the *flhDC* promoter from clade B2 scored the highest values among other promoters. The data showed low level of activity for the tested promoters. The values are mean ± standard deviation of three independent experiments.

In order to study the impact of the motility phenotype on *flhDC* promoters' activity further, we chemically transformed pSB::P_{flhDC}_B2, B1, D, A into other panel strains. The members of this panel are NCTC10418 which is a well-known strain used in antibiotic susceptibility tests, MG1655 is the lab model strain and CFT073 is the typical uropathogenic *Escherichia coli* strain. These three strains have been chosen because they belong to three different phylogenetic groups of *E. coli* (B1, A and B2, respectively) and also they have different motility phenotype(s) (Figure. 4.8A). Growth curve experiments were performed to see if there is any variation in growth pattern between the members of this panel. The data show that these strains grew in a similar way which make them convenient to use in subsequent experiments (Figure. 4.8B).

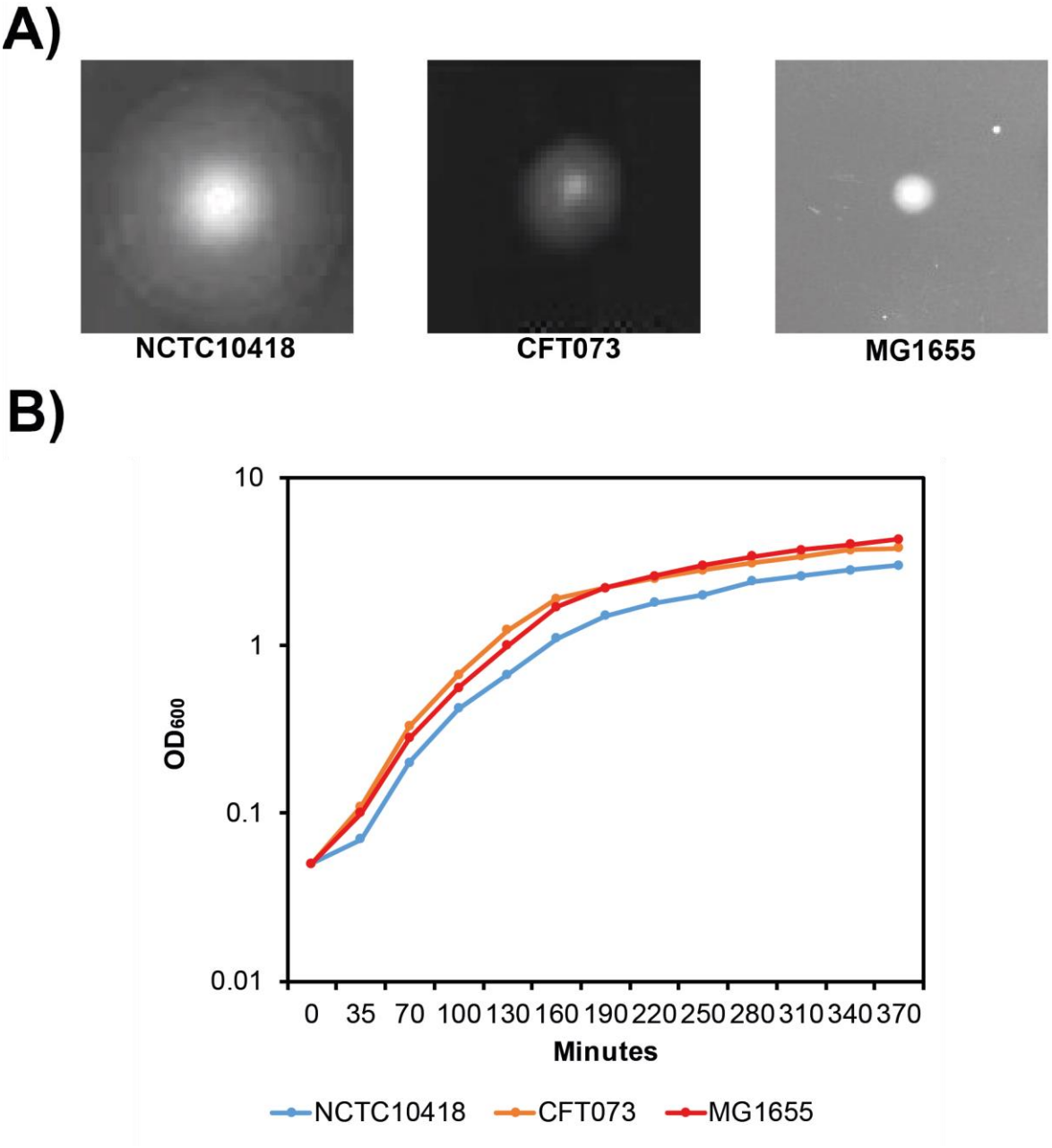


Figure. 4.8. Motility phenotype and growth curve of the panel strains.

A) Motility of the panel strains NCTC10418, CFT073 and MG1655. Colonies were independently stabbed onto motility plates containing 0.3 % agar for 8 hours at 37 ˚C. NCTC10418 exhibits high motility phenotype, whereas strain CFT073 is moderately motile compared to the low motility of MG1655. **B)** The growth curve experiment was performed by growing three independent LB cultures starting from OD_{600}= 0.05 at 37 ˚C. Optical density was measured at 30 minutes intervals. Data shows similar growth pattern for these panel strains.

Using LB cultures in triplicate for the panel strains with transformed plasmids, we measured both the absorbance at OD_{600} and the bioluminescence to determine the strength of different *flhDC* promoters through the expression of *luxCDABE* genes and light production. The variations between different strains and promoters were noticeable. All *flhDC* promoters in strain NCTC10418 showed the highest bioluminescence compared to the other two strains with P_{flhDC}-B2 on the top (~11000 light unit) (NCTC10418 data are significant). Consistently, all the promoters in strain CFT073 came below NCTC10418, but higher than the activity in MG1655 (CFT073 data are significant) (in strain CFT073 [P_{flhDC} (B2):P_{flhDC} (D)] P= 0.77, [P_{flhDC} (B1):P_{flhDC} (A)] P= 0.175, [P_{flhDC} (B2):P_{flhDC} (B1)] P= 0.00197, [P_{flhDC} (D):P_{flhDC} (A)] P= 0.00299). Again, P_{flhDC}-B2 which showed the highest activity among other promoters, it showed low or almost no activity when transformed into strain MG1655. All *flhDC* promoters in strain MG1655 showed very low activity, and the group of strains with P_{flhDC}-MG1655 exhibited the lowest bioluminescence values compared to other groups (no significant variation in the MG1655 data [P= 0.33]) (Figure. 4.9).

This experiment showed that the activity of the *flhDC* promoters reflected the motility extent of their original strains. For example P_{flhDC}-B2 was originally from CFT073 which showed the highest motility phenotype in motility assays, P_{flhDC}-B1 and D showed less motility, and P_{flhDC} of strain MG1655 was on the bottom of the list (Figure. 4.3). The experiment has also shown that these promoters' activity was consistent with the motility phenotype of the panel strains (NCTC10418, CFT073 and MG1655). Therefore, we conclude that the activity of the *flhDC* promoter depends on the *flhDC* promoter type and other factor(s) in the cell dictated by clade location and strain variation amongst the species *E. coli*.

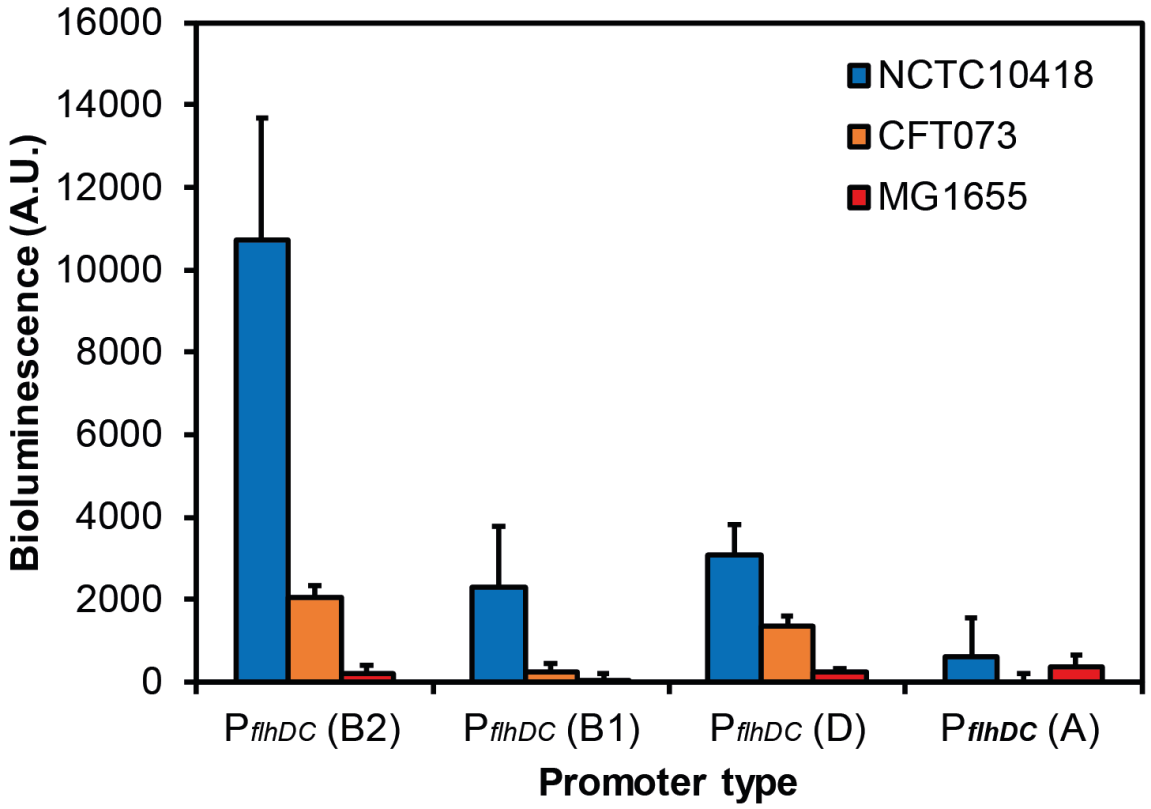


Figure. 4.9. Activity of the *flhDC* promoters in panel strains NCTC10418, CFT073 and MG1655.

The four *flhDC* promoters representing A, B2, B1 and D clades cloned into pSB401 were transformed into strains NCTC10418, CFT073 and MG1655. Strains were grown in LB cultures at 37 ˚C until OD_{600}= 0.6-0.8. All *flhDC* promoters transformed into strain NCTC10418 scored the highest bioluminescence values among other promoters. In the same regard, P_{flhDC}-B2 revealed the highest activity among other promoters. The data reflected a link between *flhDC* promoter activity and motility phenotype of the panel strains. The values are mean ± standard deviation of three independent experiments.

4.2.6. Effect of media and temperature on the flhDC promoters

We have known from the literature that bacteria form flagella under certain conditions such as lack of food, change in temperature and pH so that bacteria can move from an undesirable location to another more favourable atmosphere (Soutourina *et al.*, 1999). Here, we were interested to know if changes in culture media and temperature have an effect on our *flhDC* promoters which represent four phylogenetic groups of *E. coli*.

In this experiment, we used the four *flhDC* promoters in the panel strains (Figure. 4.8). Strains were grown in different nutritional conditions based on a minimal media base containing 0.2 % glucose. Nutrients were varied by adding increasing amount of yeast extract (0.04 g/l, 0.2 g/l, 1 g/l and 3 g/l), and bioluminescence was measured at two temperatures 30 ˚C and 37 ˚C (Figure. 4.11). A consistent pattern was observed in the figures 4.10 and 4.11 in comparison to the data from LB growth (Figure. 4.9). P_{flhDC} (B1) exhibited only strong activity in the highly motile NCTC10418 strain. In contrast, P_{flhDC} (A) was again the weakest promoter. When we perform statistics on figure 4.10, we notice significant variations in NCTC10418 and CFT073 data at 30 ˚C and 37 ˚C. MG1655 data are only significant at 37 ˚C, but not significant at 30 ˚C. Complication is when to begin to assess 30 ˚C versus 37 ˚C or one promoter against another one. For example, in CFT073 P_{flhDC} (B1) 30 ˚C:P_{flhDC} (B1) 37 ˚C (P= 0.141). In NCTC10418, P_{flhDC} (B2) 30 ˚C:P_{flhDC} (B2) 37 ˚C (P= 0.86). Similarly, a lot of the pairwise comparisons are not significant. Interestingly, no clear thermoregulation was observed here in contrast to previous MG1655 based experiments (Figure. 4.6 and 4.7). In figures 4.6 and 4.7, a rich complex medium was used, whereas a minimal medium with added nutrients was used in figure. 4.10. This suggests a link between nutrition and temperature

regulation (high nutrients lead to temperature regulation, while reduced nutrients lead to the absence of temperature regulation).

However, when we took the whole picture into consideration we noticed that the different promoters had variable activity under the two temperatures with changes of nutrients in media. We notice in figure 4.10 that *flhDC* promoter (B1) in strain CFT073 has low activity, this is not due to nutrition change because we see the same in all tested conditions in figure 4.11. Overall, these promoters showed that they have a very regular response to changes, and the bioluminescence activity was consistent with the motility phenotype of the promoters and panel strains. For example, all four promoters in strain NCTC10418 revealed the highest activity while the same promoters in MG1655 had very low activity consistent with the weakly motile strain. Moreover, P_{flhDC}-B2 had the greatest values compared to other sets of this experiment (Figure 4.11).

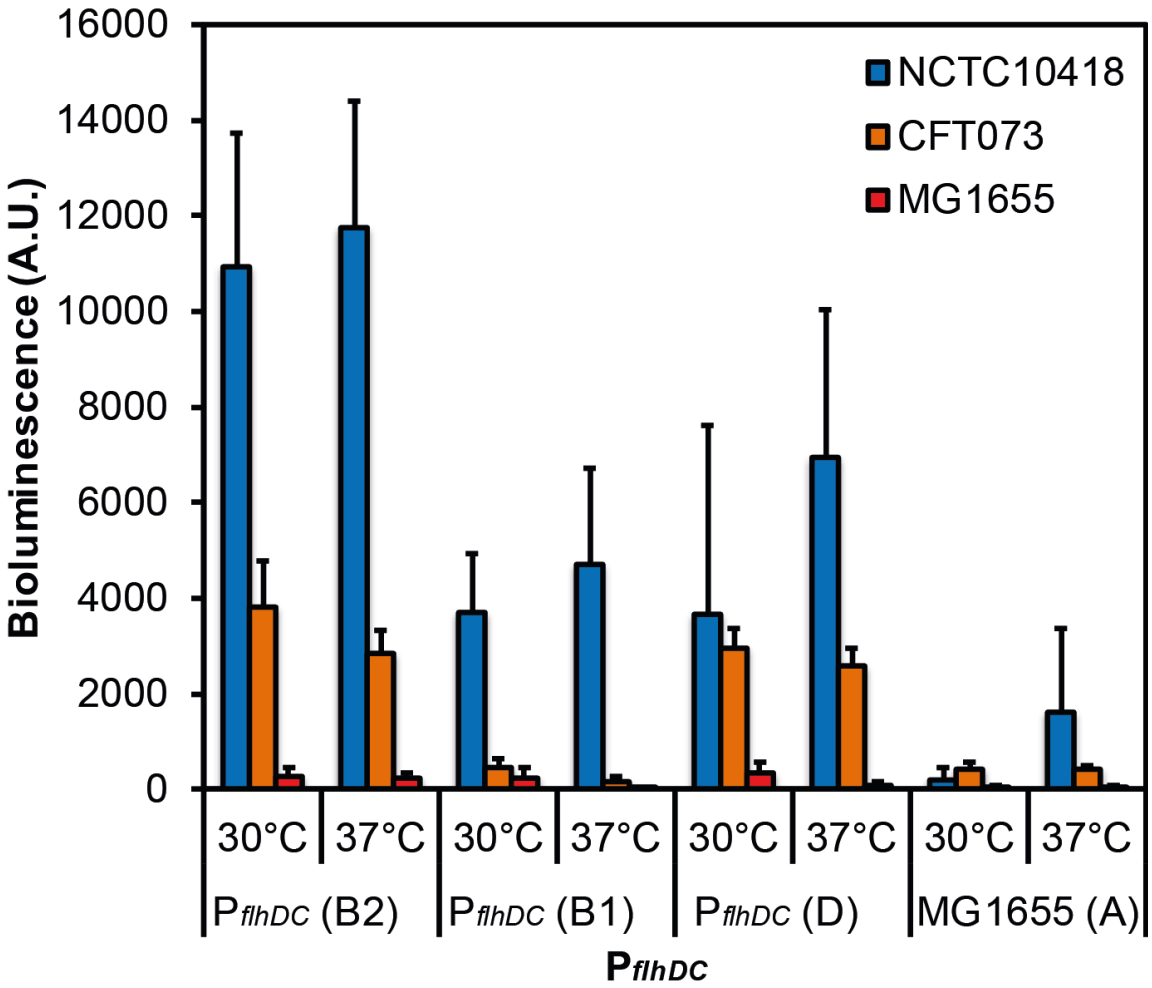


Figure. 4.10. Activity of the *flhDC* promoters in panel strains in minimal media with 1 g/l yeast extract at 30 ˚C and 37 ˚C.

The four *flhDC* promoters representing A, B2, B1 and D clades cloned into pSB401 were transformed into strains NCTC10418, CFT073 and MG1655. Strains were grown in cultures containing minimal media with 1 g/l yeast extract until OD_{600}= 0.6-0.8. The activity of *flhDC* promoters was measured by microplate reader. All *flhDC* promoters transformed into strain NCTC10418 scored the highest bioluminescence values among other promoters, whereas these promoters manifested the lowest activity in MG1655 at 30 ˚C and 37 ˚C. In the same regard, P_{flhDC}-B2 in NCTC10418 revealed the highest activity among other promoters. The values are mean ± standard deviation of three independent experiments.

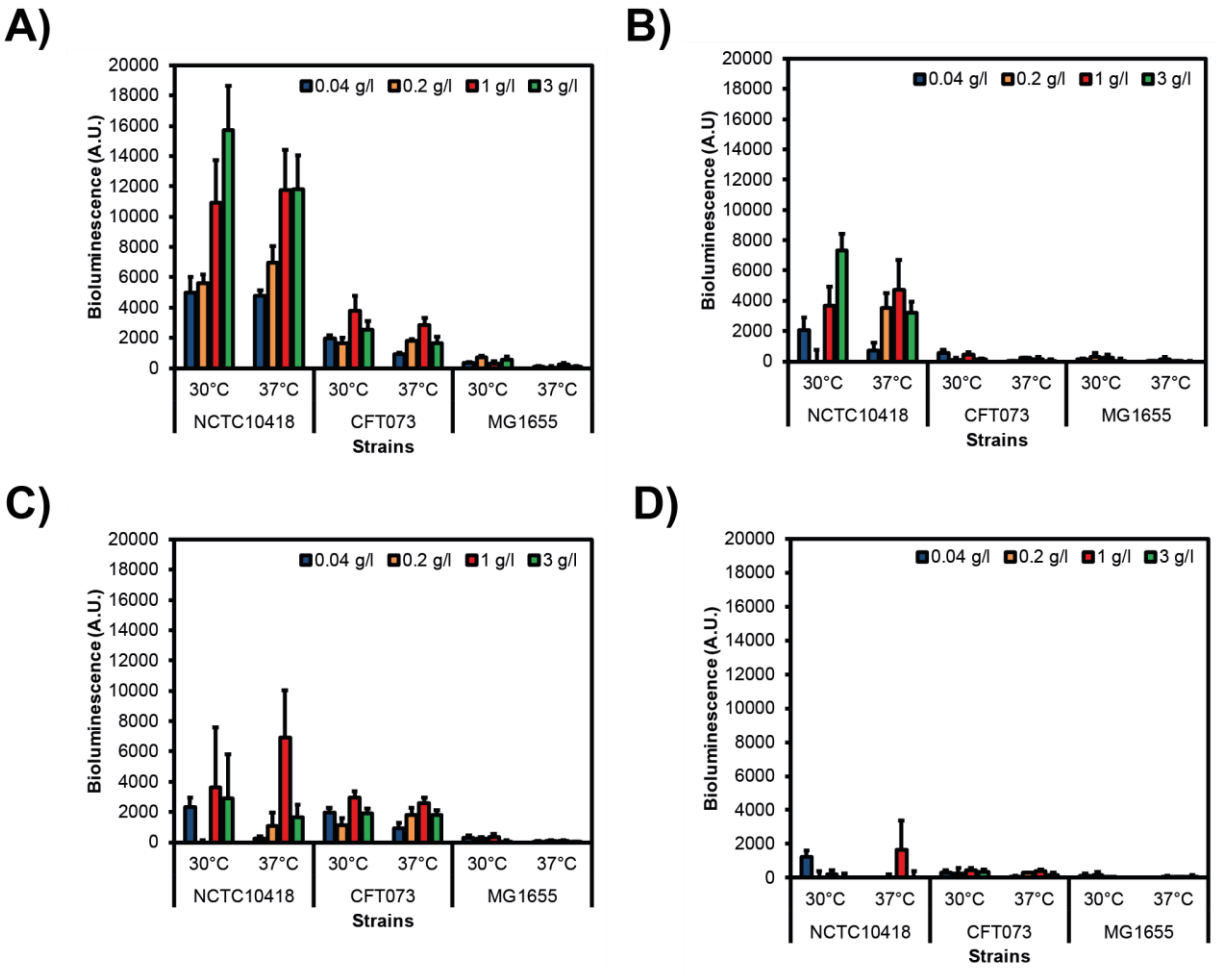


Figure. 4.11. Activity of the *flhDC* promoters in panel strains in minimal media with 0.04, 0.2, 1, 3 g/l yeast extract at 30 ˚C and 37 ˚C.

The four *flhDC* promoters representing different clades **A)** B2. **B)** B1. **C)** D. **D)** A cloned into pSB401 were transformed into strains NCTC10418, CFT073 and MG1655. Strains were grown in cultures containing minimal media with 0.04, 0.2, 1 and 3 g/l yeast extract until OD_{600}= 0.6-0.8. Promoters showed very regular responses to changes in media. Generally, the activity of these promoters was consistent with the strength of the promoter and motility phenotype of panel strains. The values are mean ± standard deviation of three independent experiments. g/l: gram per litre.

4.2.7. Impact of global regulation on the flhDC promoters

There are many regulators in the cells effecting gene expression, biochemical processes and protein-protein interactions. These regulators are especially important for the vitality of the cells in such severe conditions like stress. Others may also bind to promoters to alter gene expression to keep the cell viable in hard and good times. Here, we tested the effect of five known global regulators (*ompR, ftsH, lon, crp and clpP*) on the motility phenotype of strain MG1655 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$_B2, B1 and D strains (Figure. 4.12).

These experiments included the deletion of each of these regulators from the four strains, replacing the genes with DNA fragments containing chloramphenicol resistance cassette. This cassette was amplified from the plasmid pKD3. Deletions exploited λ-red activity of the heat sensitive plasmid pKD46. Confirmation of the deletion was performed by PCR reactions and testing the motility by stabbing inocula on semi-solid agar motility medium (Figure. 4.12A). By performing these experiments, we attempted to know the effect of these regulators on *flhDC* promoters other than the one of strain MG1655.

One of the stress proteins which belong to the *clp* family is ClpP. It has proteolytic activity against cytoplasmic misfolded or damaged proteins (Weichart *et al.*, 2003). This protease can degrade the flagellar master regulon complex FlhD$_4$C$_2$ with the help of YdiV (Takaya *et al.*, 2012). Our results signified significant increase in motility of MG1655 and MG1655::P$_{flhDC}$-B2, B1 and D *ΔclpP* mutants compared to their wild type strains (P= 0.01, P= 0.01, P= 0.02 and P= 0.005, respectively) (Figure. 4.12). This data was consistent with other researches in that ClpP acts as a negative regulator to FlhD$_4$C$_2$ activity.

The gene *ompR* regulates the expression of the outer membrane porin proteins OmpF and OmpC in association with EnvZ during changes in osmolarity (Cai and

Inouye, 2002). It also has been shown that phosphorylated OmpR binds to the *flhDC* promoter and downregulates flagellar synthesis (Shin and Park, 1995). However, significant decrease has been observed in strains MG1655::P_{flhDC}_B2, B1 and D Δ*ompR* mutants compared to their wild type (P= 0.04, P= 0.02 and P= 0.03, respectively), while strain MG1655 exhibited no significant increase in motility (P= 0.07) (Figure. 4.12B). This data contradicts the literature which state that deleting *ompR* gene leads to increased motility.

The protease FtsH has been shown to degrade membrane proteins that have not been assembled into complexes (Akiyama, 2009). The *ftsH* transcript associates with AcnB to reduce FtsH protease activity which maintain the levels of σ^{32}, DnaK and subsequently FliC which is the subunit of flagellar filament (Tang *et al.*, 2004). Our data revealed no significant change in motility in strains MG1655::P_{flhDC}_B1, D and in MG1655 *ΔftsH* mutants compared to the wild type (P=0.49, P=0.11 and P=0.67, respectively). These data again showed opposite information to what we found in the literature. However, other proteases may compensate for the role of FtsH in this process. Thus, further experiments to generate protease deficient combinations by deleting *lon* and *clpP* in the *ΔftsH* mutants should be done in the future. The decrease in motility in strain MG1655::P_{flhDC}_B2 is significant (P= 0.02) (Figure. 4.12B) which is not consistent with Tang *et al.* (2004).

Yet another protease expressed from *lon* gene. Lon in *E. coli* contributes to different physiological processes in the cell such as cell division and capsule biosynthesis. It also degrades RcsA and SulA which lead to capsular polysaccharide accumulation (Ebel *et al.*, 1999). This increased production of polysaccharide in our *lon* mutants changed the colony phenotype consistency to mucoid when they grew on an agar plate. We considered this change as a primary step in addition to using PCR and antibiotic resistance as extra steps of confirmation. It has been also shown

in the literature that *lon* has a possible role in the degradation of FlhD and FlhC (Claret and Hughes, 2000b). Moreover, Lon also plays a role in the degradation of σ^{28} (Barembruch and Hengge, 2007). Thus, we wanted to test the effect of deleting the *lon* gene on the activity of *flhDC* promoters in our strains. No significant change in motility phenotype was observed in strain MG1655 (P= 0.07) which indicate no role for *lon* in the process of flagellar synthesis. However, all other strains (MG1655ΔP_{flhDC}::P_{flhDC}_B2, B1, D) *Δlon* mutants showed significant decrease in motility compared to the wild type (P= 0.02, P= 0.01 and P= 0.01, respectively) (Figure. 4.12B). As the deletion of *lon* did not enhance the swimming diameter, we believed that a possible reason for this is that the deletion makes the cell longer and thus hinders the motility. We also attribute the slow motility to the high rate of polysaccharide secretions which increase the friction of the cell with the growth medium surface impeding the motility.

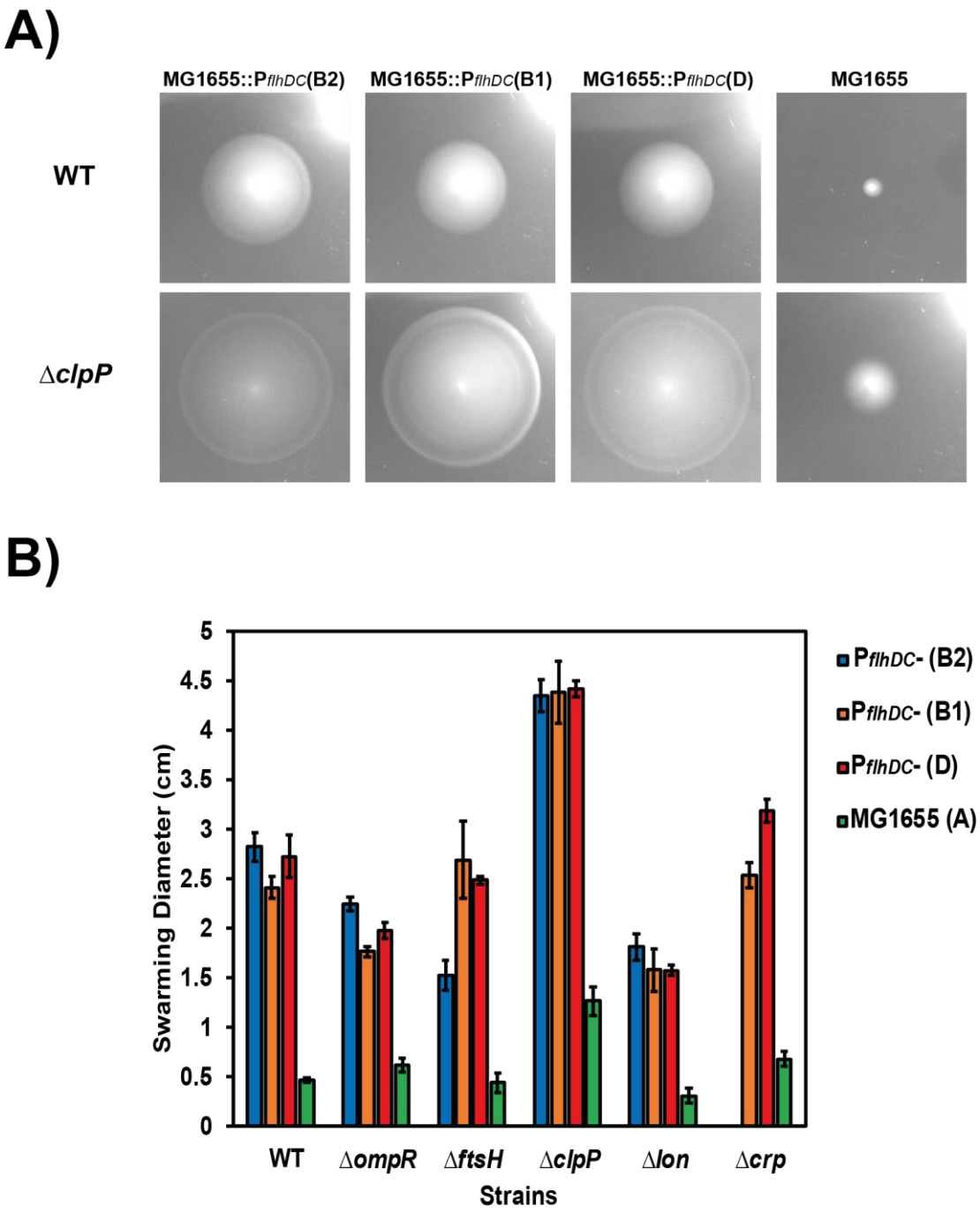


Figure. 4.12. Impact of global regulators on motility of species *E. coli*.

A) Motility phenotype of MG1655 and MG1655ΔP_{flhDC}::P_{flhDC}-B2, B1, D. Colonies were stabbed onto motility plates containing 0.3 % agar for 8 hours at 37 ˚C. *ΔclpP* mutants revealed higher motility compared to their corresponding wild types. **B)** Motility quantification of global regulators mutants. Average of swarming diameter of each strain is measured in centimeter. *ΔclpP* mutants showed the highest motility among other deletion mutants compared to the wild types. MG1655ΔP_{flhDC}::P_{flhDC}-B2 *crp* mutant did not grow on TB agar, thus, we were unable to perform motility assay. The values are mean ± standard deviation of three independent experiments.

The CRP (cyclic-AMP receptor protein) is known to activate a number of genes for utilization of carbon source other than glucose (Shimada *et al.*, 2011). The literature showed that *crp* mutants are nonmotile as a result of reduced *fliC* expression (Soutourina *et al.*, 1999). Reduced FliC production in *crp* mutants is caused by low *flhDC* expression due to unavailability of CRP to bind to its promoter (Soutourina *et al.*, 1999). Interestingly, our strains MG1655, MG1655ΔP$_{flhDC}$::P$_{flhDC}$_B1 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$_D *Δcrp* mutants showed no significant change in their motility phenotype (P= 0.06, P= 0.17 and P= 0.16, respectively) (Figure. 4.12B). We attributed these results to the fact that we are using MG1655 strain that has no IS elements upstream *flhDC* promoter which makes this strain very weakly motile as we have shown earlier.

Although *crp* mutants grew slowly in TB medium as a poor carbon source, we were unable to grow MG1655ΔP$_{flhDC}$::P$_{flhDC}$_B2 *crp* mutant in TB agar medium to perform motility assay. We therefore, as an extra step of confirmation, streaked wild type strains and *crp* mutants on MacConkey agar, and plates were incubated overnight at 37 ˚C. wild type strains were able to grow normally as they utilized lactose in media, while *crp* mutants produced small and colourless colonies (Figure 4.13). We concluded that *Δcrp* mutants have other phenotypes consistent with this mutation. Future research could explore CRP regulation of *flhDC* transcription across clades of *E. coli*.

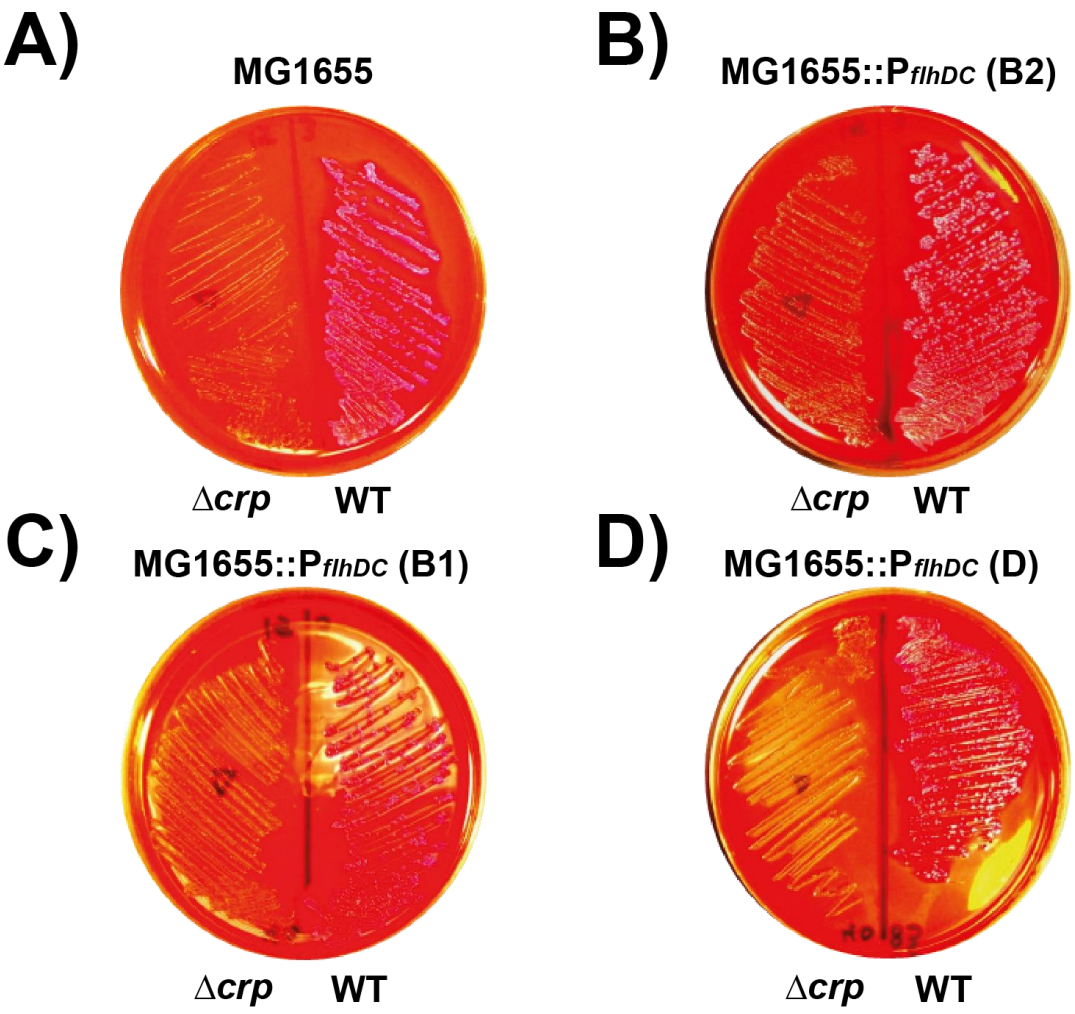


Figure. 4.13. Confirmation step of *Δcrp* mutants on MacConkey agar.

This step performed by streaking **A)** MG1655. **B)** MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2. **C)** MG1655ΔP$_{flhDC}$::P$_{flhDC}$- B1. **D)** MG1655ΔP$_{flhDC}$::P$_{flhDC}$-D wild types and *Δcrp* mutants on MacConkey agar plates. Plates were incubated overnight at 37 ˚C. Wild types were able to grow as normal colonies, whereas *Δcrp* mutants grew as small and colorless colonies.

4.2.8. Replacing the flhDC promoter of MG1655 and UTI89

CRISPR-Cas9 technology was used for the first time in this project to generate flagellar mutants and strain variants. It is considered as an efficient and reliable way to precisely target the gene of interest and replace it with another that may carry a reporter or mutation. The literature suggest that this multi-gene editing system is more precise in genetic recombination than the ordinary methods with efficiency up to 100 %. The system used in this project was based on a CRISPR-Cas system described by Jiang *et al.* (2015). After investigating the variation of the *flhDC* promoter in the newly created strains, we wanted to perform a control experiment to confirm our findings. For this purpose, we choose the strain MG1655 and the strain UTI89 which belonging to clade B2 represents another model UTI strain. This experiment included the replacement of the *flhDC* promoter of strain MG1655 with that of UTI89, and the replacement of the *flhDC* promoter of UTI89 with that of strain MG1655. Furthermore, the CRISPR technique was used rather than the Blank *et al.* (2011) method by electroporating the helper plasmids pCas (which has λ-red recombinase system) and chloramphenicol target plasmid (pTrg-Cm) to the MG1655ΔP$_{flhDC::}$Cm with an amplified DNA fragment that includes the *flhDC* promoter of strain UTI89. The Datsenko and Wanner method was performed on strain UTI89 to replace its *flhDC* promoter with the chloramphenicol resistance cassette. Then, the CRISPR method was applied on UTI89ΔP$_{flhDC}$::Cm to add the *flhDC* promoter from strain MG1655 instead of the antibiotic cassette. The motility, then, was quantified by growing the bacteria on the motility agar for 8 hours at 37 ˚C (Figure. 4.14A).

Our data showed a decrease in motility of 85.43 % in strain UTI89::P$_{flhDC}$ (MG1655) compared to the wild type strain UTI89. On the other hand, the strain MG1655::P$_{flhDC}$ (UTI89) showed a high increase in motility of 465 % compared to the wild type strain MG1655 (Figure 4.14B). There is a significant variation in motility

between MG1655 and UTI89::P_{flhDC} (MG1655) (P= 0.02). A significant variation in motility is also observed between UTI89 and MG1655::P_{flhDC} (UTI89) (P= 0.02). This control experiment confirmed our previous *flhDC* promoter replacement shown at the start of this chapter when we replaced the weak *flhDC* promoter of strain MG1655 with other *flhDC* promoters (Figure. 4.2B).

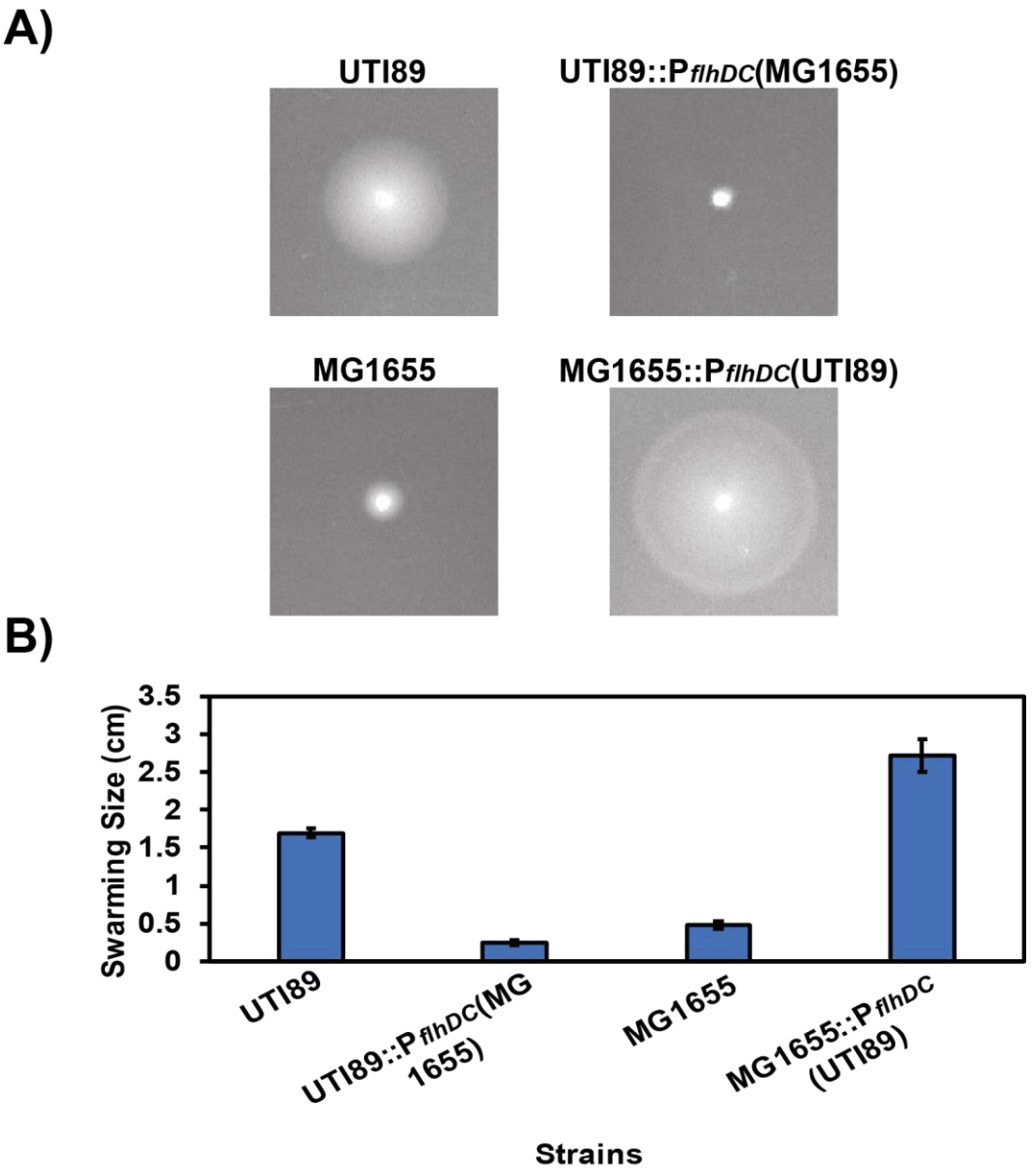


Figure. 4.14. Control experiment of the MG1655 and UTI89 *flhDC* promoter replacement.

The experiment included the replacement of the *flhDC* promoter of strain MG1655 with that of strain UTI89, and the replacement of the *flhDC* promoter of strain UTI89 with that of strain MG1655. **A)** Motility phenotypes by stabbing colonies of each strain onto motility plates containing 0.3 % agar for 8 hours at 37 ˚C. Strain MG1655::P_{flhDC} (UTI89) showed increased motility compared to the wild type, whereas strain UTI89::P_{flhDC} (MG1655) exhibited suppressed motility compared to its wild type. **B)** Motility quantification. Average of swarming diameter of each strain measured in centimetre. The values are mean ± standard deviation of three independent experiments.

4.2.9. Replacing of fliM with fliM-yPet

Amplified DNA cassette contains chloramphenicol resistance gene from plasmid pKD3 was used to replace *fliM* in strain MG1655 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$_B2, B1, D. Then, by using the two helper plasmids pTrg-Cm and pCas, we managed to replace this DNA cassette with another *fliM* gene that carry a gene capable of expressing a yellow fluorescent protein (*fliM-yPet*). We chose this system over how Jiang originally intended to exploit it as it allowed myself and the lab to use chloramphenicol resistance mutants in flagellar genes we have already generated. This was facilitated by using a pTrg-Cm helper plasmid that targets the chloramphenicol cassette.

Our lab is interested in flagellar cell biology relating flagellar abundance to flagellar expression. We now know that if *flhDC* activity is below a given level, flagellar promoter activity will not relate to the flagella number per cell and swarming diameter on motility plates (Albanna, 2017). Furthermore, questions were being raised by the flagellar field whether FliM-XFP fusions are good indicators of functional flagella. Therefore, we were interested in using a FliM-YPet fusions to investigate flagellar number in strain MG1655 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2, B1, D by counting flagellar fusion foci (Figure. 4.15). We were also aiming in the long term to transfer CRISPR technology to use such fusions in a wider range of *E. coli* isolates.

However, further experiments are to be done to investigate the ratio of what we see inside a cell to what we can visualize on the cell surface using other cell biology based method such as *flgEA240C* in these strains (Sim *et al.*, 2017). The analysis of FliM-YPet foci revealed that over 75 % of the population in strains MG1655 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B1 lack the flagella, whereas over 40 % of the population in strains MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$-D have an

average of one FliM foci (Figure. 4.16). The data allowed us to compare motility phenotype, the *flhDC* promoter activity with the flagellar number. The replacement of the *flhDC* promoter of strain MG1655 with the same promoter from B2, B1 and D groups led to the activation of flagellar system in MG1655. This increased activation was observed through the motility phenotype and the *flhDC* promoter activity. However, the behaviour of the *flhDC* promoter of clade B1 is confusing as this strain is highly motile though the lack of FliM foci. We attribute this phenomenon to a possible mutation developed in the strain MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B1.

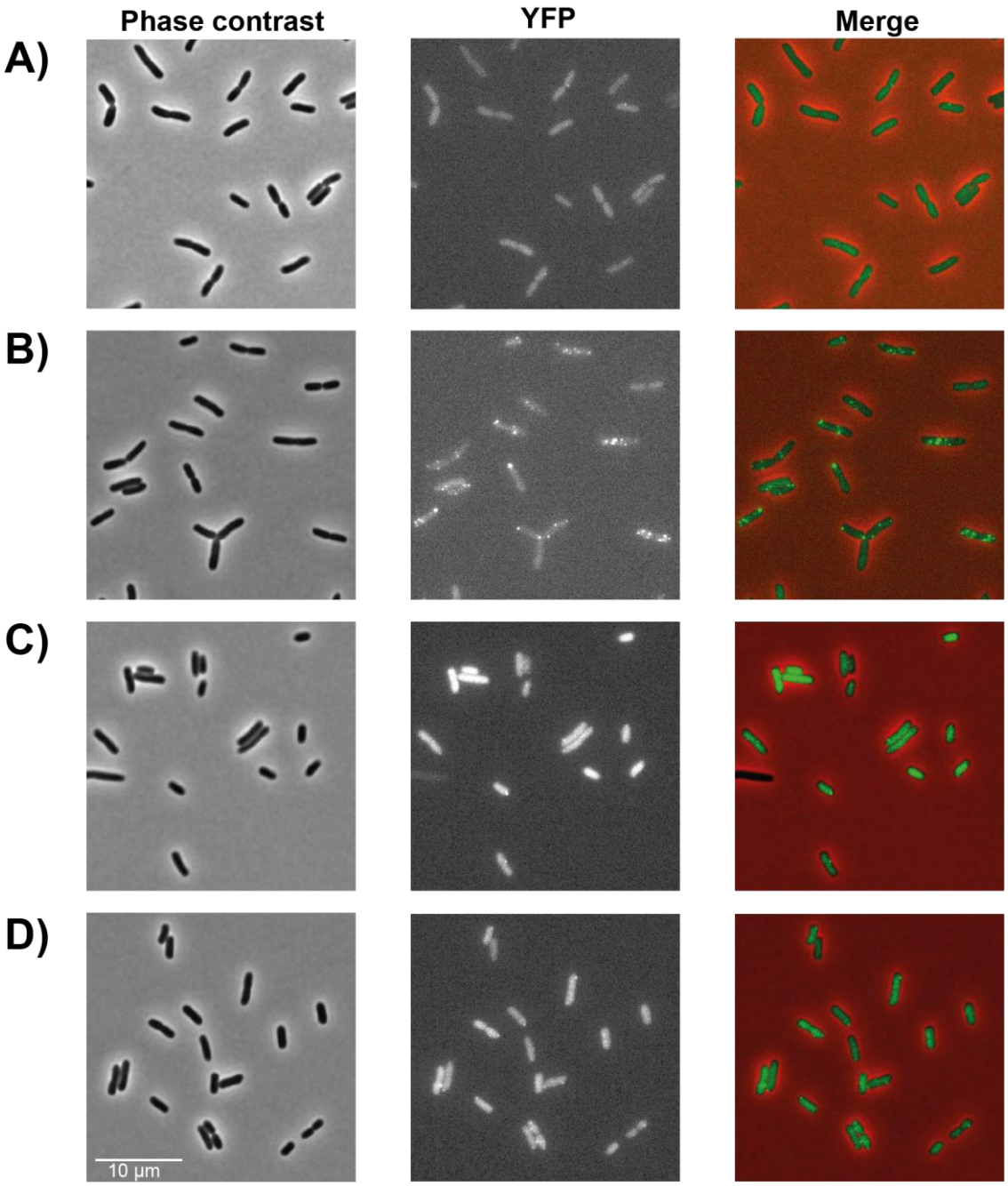


Figure. 4.15. Fluorescent microscope images of MG1655 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2, B1, D with *fliM-yPet*.

Three TB cultures of each strain were started from OD$_{600}$= 0.05 and incubated at 30 ˚C until they reach OD$_{600}$= 0.6-0.8. Slides for each culture were prepared and cells were examined under fluorescent microscope. Fifteen images were taken for each strain. Strains used in this experiment are: **A)** MG1655::*fliM-yPet*. **B)** MG1655::P$_{flhDC}$-B2::*fliM-yPet*. **C)** MG1655::P$_{flhDC}$-B1::*fliM-yPet*. **D)** MG1655::P$_{flhDC}$-D::*fliM-yPet*. Images are representative of the experimental data obtained.

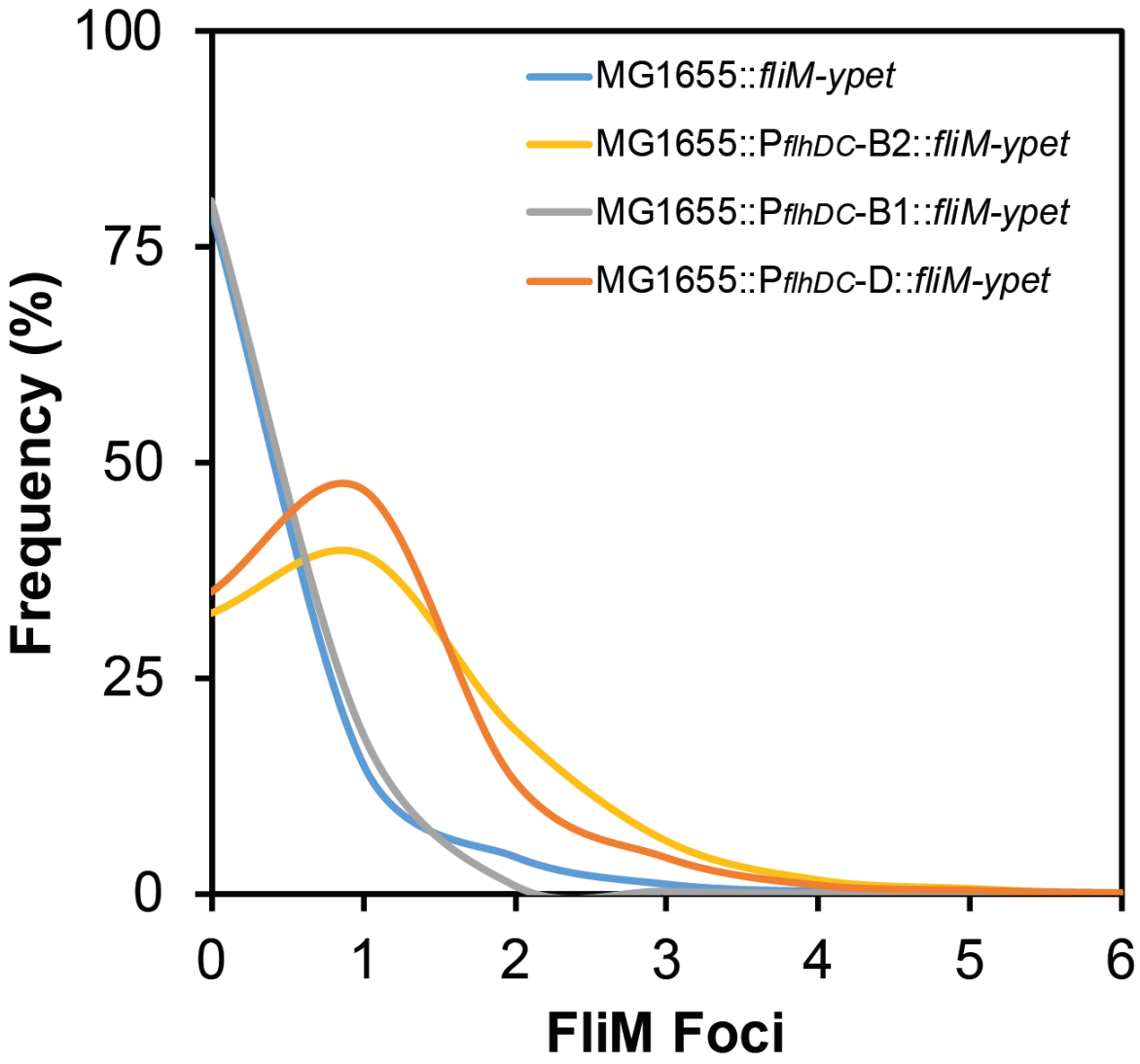


Figure. 4.16. Quantification of *fliM-yPet* foci in MG1655::*fliM-yPet*, MG1655::P_{flhDC}-B2::*fliM-yPet,* MG1655::P_{flhDC}-B1::*fliM-yPet* and MG1655::P_{flhDC}-D::*fliM-yPet*.

The fluorescent images were analysed using oufti.org software. The *fliM-yPet* foci were quantified using parameters available in this software. Figure was created by Excel to show the relevant number of foci per strain. The vast majority of cells in strains MG1655::P_{flhDC}-B2::*fliM-yPet* and MG1655::P_{flhDC}-D::*fliM-yPet* possess 1-2 *fliM-yPet* foci. Most of the cells in strains MG1655::*fliM-yPet* and MG1655::P_{flhDC}-B1::*fliM-yPet* were lacking such foci. Data represents 700-1000 cells analysed from 5 of 15 images taken from three independent cultures of each strain. Figure created in the same way used in Sim *et al.* (2017).

4.3. Summary

This chapter provides a strong foundation, and data gathered here will form the backbone to other chapters in this project. We started this work from defining different groups of *E. coli*, and we defined the phylogenetic tree of McNally who allocated *E. coli* strains into six groups (A, B1, B2, D, E and F) based on the whole genome analysis. In order for an *E. coli* strain to be motile, it must express the master regulator complex $FlhD_4C_2$. Flagellar synthesis is also a hierarchical process includes the activation of three classes of genes (Class 1, 2 and 3) with *flhDC* on the top. Regarding *E. coli* motility, the majority of the literature focus on one lineage of strains, clade A. Depending on the history of these strains, they are either highly or weakly motile (Barker *et al.*, 2004). The literature state that the extent of motility of this lineage is sensitive to insertion sequence elements bind to their *flhDC* promoter region. Therefore, we started to perform comprehensive investigation of motility in *E. coli*, and we took into consideration the wider picture of the species not only a specific strain.

As a result, we aligned the *flhDC* promoter region of *E. coli* genomes, and we found that they can be divided into three distinct groups. The first experiment we attempted to make is to replace the *yecG-flhDC* intergenic region of MG1655 with other *yecG-flhDC* region belong to different clades to examine the effect on motility of our weakly motile MG1655 strain. Data has shown significant increase in motility of MG1655 with new promoters inserted. The *yecG-flhDC* intergenic regions used did not have an IS element. However, the motility phenotype reflected not only the activity of the *flhDC* promoter region but also class 2 and σ^{28} dependent class 3 promoter activity. We also tested these promoters for catabolite repression by growing the strains in LB and TB media combination and found no difference occured

in motility phenotype regardless to the combination used which indicate no change in *flhDC* expression.

In order to measure the strength of the *flhDC* promoter, we cloned this promoter of MG1655 and its derivatives into a reporter plasmid and we used pSB401 which carry *lux* genes for this purpose. Our data showed clearly that the activity of the promoter reflect whether panel strains descending from three different phylogenetic groups are motile or not. These promoters also exhibited very similar response in different growth media and in two temperatures. However, we found discrepancy in the effect of the different temperatures used based on the panel strains. There was higher activity of the promoters tested under 30 ˚C compared to their activity under 37 ˚C when promoters transformed into strain MG1655 and its derivatives (Figures 4.6 and 4.7). No clear variation in the activity has been observed for the *flhDC* promoters under 30 ˚C and 37 ˚C when they transformed into NCTC10418, CFT073 and MG1655 in minimal media with 1 g/l yeast extract (Figure. 4.10).

We also tested the effect of deleting five genes known to be regulating the *flhDC* promoter, and we examined the impact on motility of MG1655 and its derivatives. Our results did not always reflect the gene function with motility phenotypes with respect to what we understand from the literature. However, we attributed this to some physical changes in the cells in case of *lon* mutants. It may also caused by the fact that our MG1655 strain lack the insertion element which can insert in the *flhDC* promoter and enhance motility. Being this region is susceptible to mutations is also another possible interpretation of our data. Further experiments using another MG1655 with IS element need to be done to confirm motility phenotype in our mutants.

The new CRISPR-Cas technology had been used successfully to generate YFP tagged FliM reporter to study flagellar abundance and distribution of foci in the basal body region. However, one limitation accompanied this experiment after prolonged incubation of the motility plates of negative control using pTrg-Cm alone inoculated with cells after performing final step of CRISPR method. False positive swarming phenotype was observed sometimes due to contamination of plasmid with chromosomal DNA in miniprep preparations as we had to sequence *fliM* amplified from these cells. Nevertheless, we were able to obtain the correct transformants from the first attempt from LB/ Spectinomycin-Kanamycin agar plates or from positive motility plates. The number of foci seen in our experiments is low compared to what the literature says. We argue that there is a good chance that the *E. coli* strains used in the literature have the IS element in the *yecG-flhDC* intergenic region and, thus, have a higher number of foci.

The data introduced in this chapter showed that replacing the *yecG-flhDC* intergenic region of our strain MG1655 with other regions belong to highly motile *E. coli* strains dramatically enhanced its motility. It has also shown that cloning this region into *lux* plasmid expressed high bioluminescence activity by all strains across *E. coli* as a species. This leads to the question: Which part of the *yecG-flhDC* region is responsible for the *flhDC* promoter activity seen here? We will be dealing with this question in the next chapter which will analyse this region looking for tenable recourses for the observed phenomena.

Chapter Five: Analytical Study of the *Escherichia coli yecG-flhDC* Intergenic Region

5.1. Introduction

In the preceding chapter we have asked what impact replacing the *yecG-flhDC* intergenic region with model sequence types has upon motility in the model *E. coli* strain MG1655. We demonstrated that this weakly motile strain became highly motile after swapping its *flhDC* promoter region with the same region descended from other clades which represent the species. The data suggested that other factors dictate the regulation of *flhDC* transcription across *E. coli* as a species, as it was clear the different regions influenced motility differently from the original region in MG1655 reflected by quantified motility and the *flhDC* promoter activity. Thus, we wanted to analyse the *yecG-flhDC* intergenic region to see the effect of its different parts of this region on the motility phenotype.

We know from experimental results achieved in our lab on *Salmonella* serovars that flagellar gene expression and swarming diameter on an agar plate do not always reflect each other (Albanna, 2017). We also know from our early findings that *E. coli* exhibits a dramatic level of variation with respect to motility. Therefore, we wanted to also ask how do changes to the *yecG-flhDC* intergenic region architecture impact other flagellar gene expression? We, therefore, cloned a class 2 *flhDC* dependant promoter and a class 3 σ^{28} dependant promoter onto plasmid pSB401 that carries *lux* genes. We chose P_{flgB} and P_{fliC} to represent these classes. This chapter demonstrates our analysis of the *yecG-flhDC* intergenic region in two distinct *E. coli* strains looking for the part of the DNA that dictates motility, and also the activity of the other flagellar promoter classes (class 2 and class 3) with respect to motility.

5.2. Results and discussion

5.2.1. Dissection of the yecG-flhDC intergenic region

Having observed an alteration in motility in strain MG1655 we wanted to study this phenomenon further to determine which part of the *yecG-flhDC* intergenic region influences *flhDC* expression. By using bioinformatics, we looked closely at this region with the help of the most recent paper on IS element activation of motility in MG1655 (Figure. 5.1) (Fahrner and Berg, 2015). We noticed we could split the 780 bp *yecG-flhDC* intergenic region into zones including: 1) a long untranslated region downstream of the known transcriptional start site which many regulators can bind to (Fahrner and Berg, 2015), 2) a *flhDC* promoter which is conserved in *E. coli* as a species, 3) a regulatory region upstream of the *flhDC* promoter which is defined by Barker *et al.* (2004) and 4) a potential IS SCAR region which shows variation among different strains of *E. coli* (Figures 4.1A and 5.1).

For this purpose, we analysed this region in two *E. coli* strains MG1655 and MG1655ΔP_{flhDC}::P_{flhDC}-B2. Our selection of these two strains was based on the fact that the *yecG-flhDC* intergenic region descended from two totally different clades of *E. coli* A and B2, respectively. As we mentioned earlier, MG1655 is a well-known lab model studied exclusively by researchers, whereas the other *flhDC* promoter belongs to strain CFT073 which is considered as a model uropathogenic strain that causes urinary tract infection in humans.

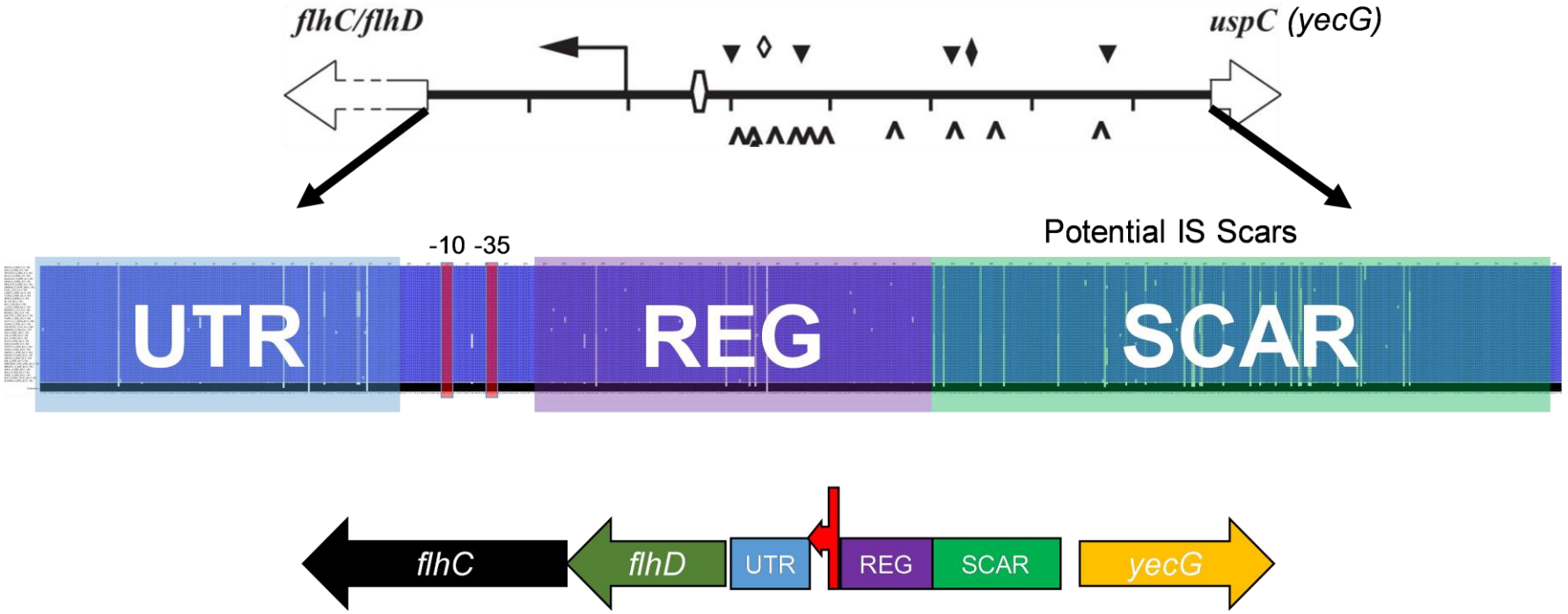


Figure. 5.1. Dissection of the *yecG-flhDC* intergenic region showing different zones.

Bioinformatics used to align 51 *E. coli* isolates to dissect a 780 bp *yecG-flhDC* intergenic region based on (Fahrner and Berg, 2015). The alignment revealed four different zones lie in this region. They are: the untranslated region (UTR), the *flhDC* promoter region, the regulatory region (REG) and the IS SCAR region (SCAR).

5.2.2. Effect of different parts of the yecG-flhDC intergenic region on motility

In order to study this region in more detail, our aim was to generate a collection of strains using CFT073/MG1655 *flhDC* promoter region as templates to create the desired amplified DNA fragments with different sets of primers (Figure. 5.2A). Then, we used the CRISPR system to replace the *yecG-flhDC* intergenic region of each of the strains with a defined deletion. As we noticed earlier, we could convert strain MG1655 to a highly motile strain by replacing just this intergenic region (Figure. 4.3). Hence, we wished to perform this work to answer the following question: Which zone of the *yecG-flhDC* intergenic region dictates the motility phenotype?

As this region includes multiple zones, we tried to delete one zone at a time and also in combinations to test the impact on motility. Importantly, all constructs were designed to leave the *flhDC* promoter intact (Figure. 5.2A). Thus, the following deletion mutants were created: 1) untranslated region downstream of the *flhDC* promoter (ΔUTR), 2) the SCAR fragment (ΔSCAR), 3) regulatory and SCAR fragments (ΔRG_SCR), 4) untranslated region and SCAR region (ΔUTR_SCR) and 5) deleting the whole intergenic region (ΔU_R_S), whereas the wild type strains were used as a control in this series of experiments (Figure. 5.2A).

After performing the confirmation steps, we tested the motility of these mutants. The deletion mutants exhibited a range of motility phenotypes on motility agar plates ranging from nonmotile to highly motile (Figure. 5.2B).

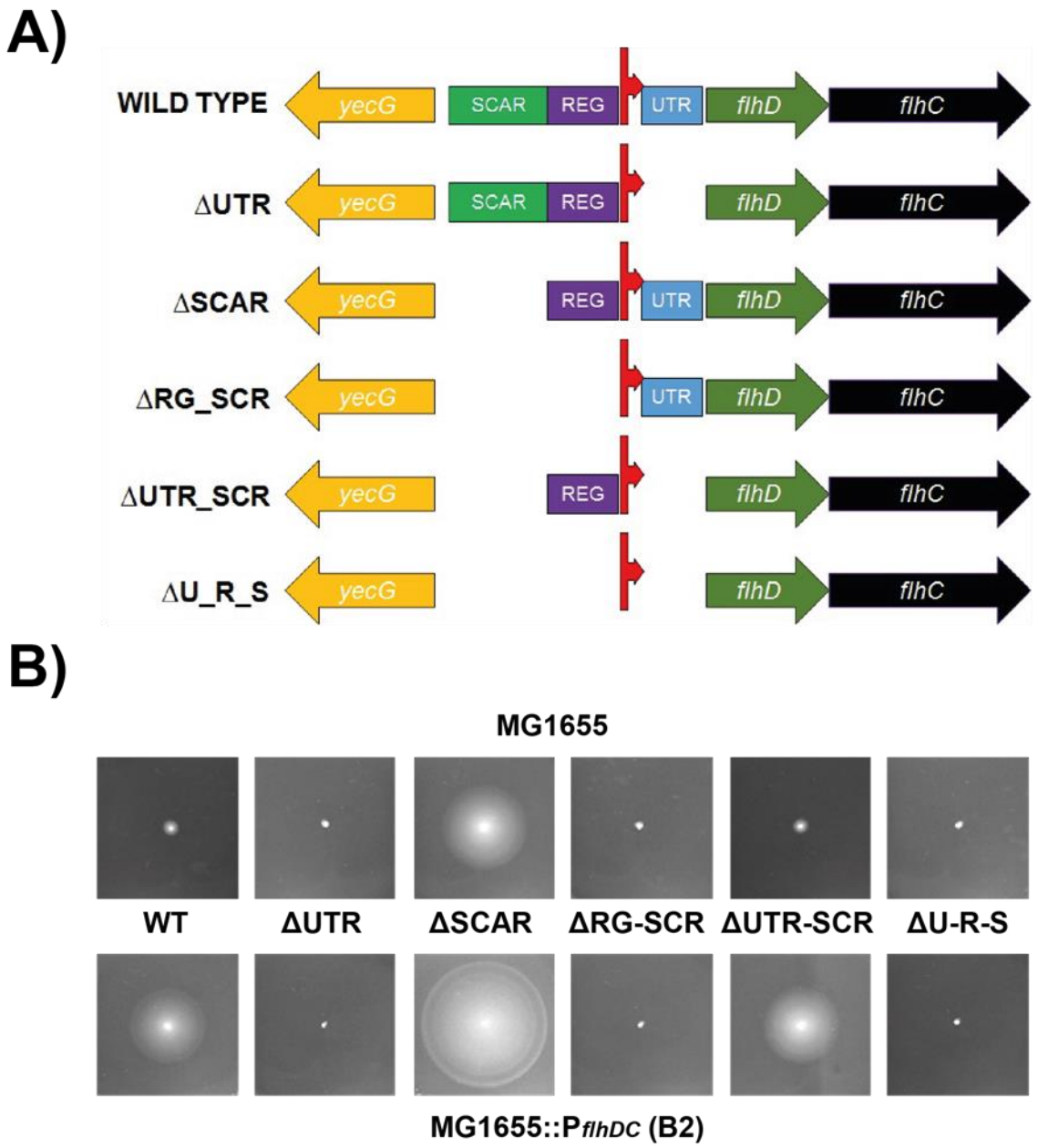


Figure. 5.2. Impact of deleting the *yecG-flhDC* intergenic region zones of strains MG1655 and MG1655::P_{flhDC}(B2) on motility.

A) Schematic drawing shows different deletion mutants performed within the *yecG-flhDC* intergenic region. Deletion mutants created are: untranslated region downstream of the *flhDC* promoter (ΔUTR), the SCAR fragment (ΔSCAR), regulatory and SCAR fragments (ΔRG_SCR), untranslated region and SCAR region (ΔUTR_SCR) and deleting the whole intergenic region (ΔU_R_S). **B)** Motility phenotypes of deletion mutants. Three colonies of each strain were stabbed onto motility plates containing 0.3 % agar, incubated for 8 hours at 37 ˚C. Different mutants exhibited various motility phenotypes with ΔSCAR mutants was the highest among other mutants.

Our data revealed that the ΔUTR, the ΔRG_SCR and the ΔU_R_S mutants turned the two strains MG1655 and MG1655ΔP_{flhDC}::P_{flhDC}-B2 into completely nonmotile strains (Figure. 5.2B). This suggests the importance of these zones in the process of flagellar assembly and regulation. Indeed, the *flhDC* operon is under control of different levels of regulation, and is influenced by transcriptional and posttranscriptional regulators that respond to environmental cues such as nutrients, osmotic pressure, oxygen, pH and temperature (Soutourina and Bertin, 2003).

The literature showed that many post-transcriptional regulators do positively or negatively impact flagellar synthesis in *E. coli*. The 5' untranslated region (5' UTR) of the *flhDC* mRNA acts as a binding site for small non-coding regulatory RNAs (sRNAs). ArcZ, OmrA, OmrB and OxyS negatively regulate flagellar synthesis while McaS positively impact the process following a variety of environmental prompts (De Lay and Gottesman, 2012). Moreover, the *flhDC* mRNA leader sequence that has a 197 nucleotide in the 5' UTR zone acts as a binding site for CsrA. Binding does dramatically impact both *flhDC* mRNA stability and translation initiation and therefore activates *flhDC* expression (Wei *et al.*, 2001). Furthermore, the regulatory region has multiple binding sites for OmpR, RcsAB, LrhA and FliZ. These proteins are other regulators that impact the process of *flhDC* gene expression (Fahrner and Berg, 2015). Histone-like nucleoid (H-NS) protein and cyclic AMP-catabolite activator protein (CAP) have multiple binding sites across the *yecG-flhDC* intergenic region and they positively regulate flagellar synthesis (Fahrner and Berg, 2015). In accordance with this information, we concluded that nonmotile strains obtained from deleting the UTR and REG parts were a normal result from removing many important regulatory inputs from *flhDC* operon expression.

On the other hand, the data showed a weighty shift in motility in strains MG1655 and MG1655ΔP*flhDC*::P*flhDC*-B2 following the deletion of the SCAR region (ΔSCAR). The motility of strain MG1655ΔSCAR has been increased 6-fold compared to the wild type (P< 0.0001). No significant variation between MG1655 WT and UTR-SCR (P= 1), but the variation of SCAR:UTR-SCR is significant (P< 0.0001). RG-SCR:UTR:U-R-S has significant variation (P= 0.01). Both MG1655 WT:RG-SCR (P= 0.01) and MG1655 WT:U-R-S (P= 0.0009) are significant (Cutoff is close to 0.3 cm which later we use to define the nonmotile strains (see chapter 6) (Figure. 5.3). The data of strain MG1655::P*flhDC*-B2 as a whole is significant. The motility of strain MG1655ΔP*flhDC*::P*flhDC*-B2ΔSCAR has been increased 1.8-fold compared to its wild type (P< 0.0001) (Figure. 5.3). No significant variation between WT and UTR-SCR (P= 0.22), whereas there is significant variation between SCAR and UTR-SCR (P< 0.0001). No difference among RG-SCR:UTR:U-R-S (P= 0.82), but still significant against WT (Figure. 5.3). The SCAR fragment of the *yecG-flhDC* intergenic region of K-12 strains such as MG1655 has been known in the literature to be a hot spot region which has many sites for IS elements insertions (Fahrner and Berg, 2015). A study showed poor motility of K-12 strains of *E. coli* lacking IS elements. While other K-12 strains were highly motile due to IS elements inserted in the SCAR region of their *flhDC* promoters (Figure. 1.14) (Barker *et al.*, 2004). This study suggests that IS elements disorganise repressor binding sites leading to its inactivation and thus, positively regulate *flhDC* expression. However, we used strain K-12 MG1655 that has no IS elements, therefore, Barker's paper does not apply to our work as we deleted the SCAR region regardless to the prescence or absence of IS elements.

Interestingly, deleting both the UTR and SCAR from MG1655 and MG1655ΔP*flhDC*::P*flhDC*-B2 (ΔUTR-SCR) maintained the motility of these two strains

as their wild type strain phenotype (Figure. 5.2B). We propose that because deleting SCAR increased *flhDC* expression, this deletion overcame regulatory inputs via UTR.

Throughout this experiment, we tried to delete the regulatory region (REG) from the *yecG-flhDC* intergenic region to see what effect it has over motility. We amplified two DNA fragments and attempted to sew them together to create ΔREG strains. However, there was a limitation preventing us from doing so as one primer designed over a regulatory inverted repeat in the *flhDC* promoter region, generating hairpins preventing PCR amplification (Figures 5.4 and 8.2). We demonstrated by performing this set of experiments the importance of different parts of the *yecG-flhDC* intergenic region on motility of two strains. Because the SCAR region showed suppressive activity upon flagellar gene expression, we wished to study this part further to explore how does this zone repress motility.

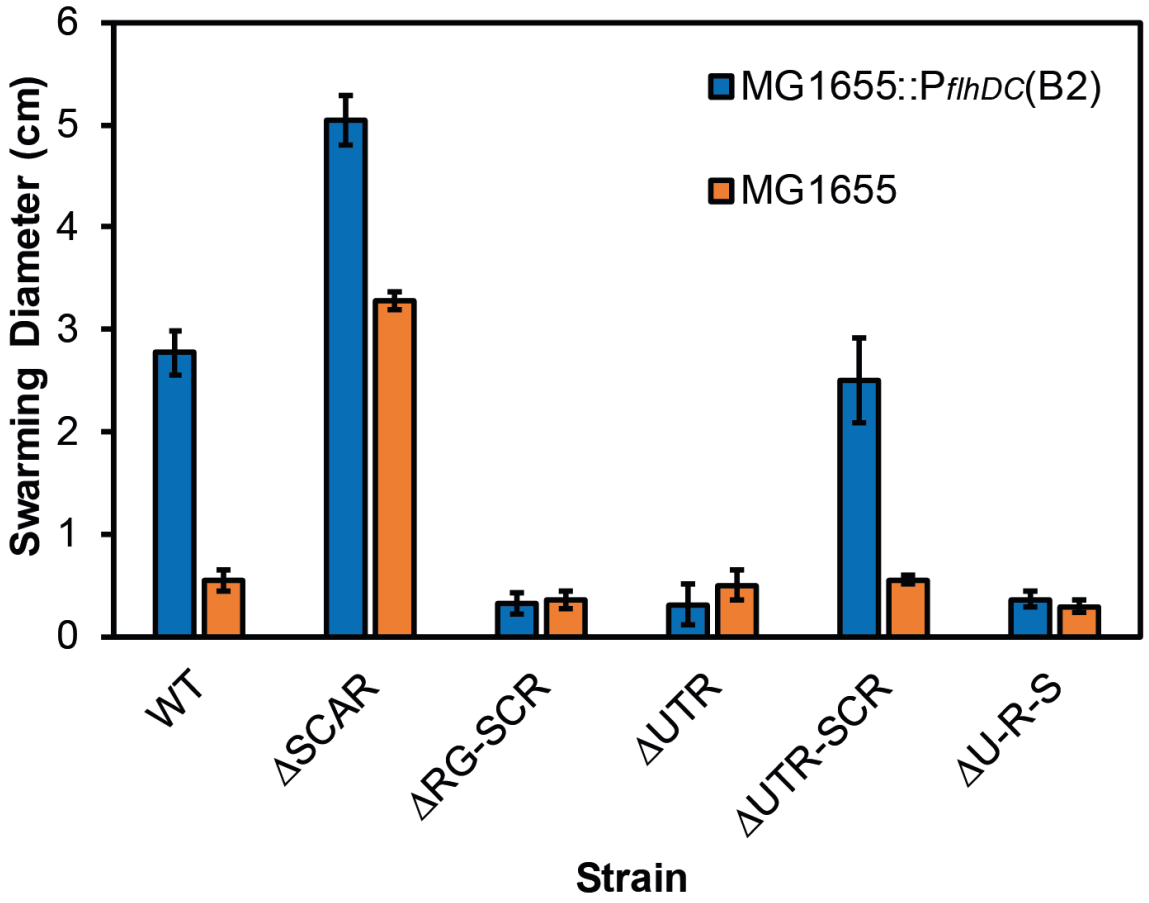


Figure. 5.3. Quantification of motility of the *yecG-flhDC* intergenic region deletion mutants of strains MG1655 and MG1655::P_{flhDC}(B2).

The average of swarming diameter in centimeter of each strain was measured from motility plates. Data showed that ΔSCAR mutants revealed the highest score among other mutants compared to the wild type, whereas ΔUTR-SCR mutants maintained the same motility observed in the wild types. All other deletion mutants were nonmotile. The values are mean ± standard deviation of three independent experiments.

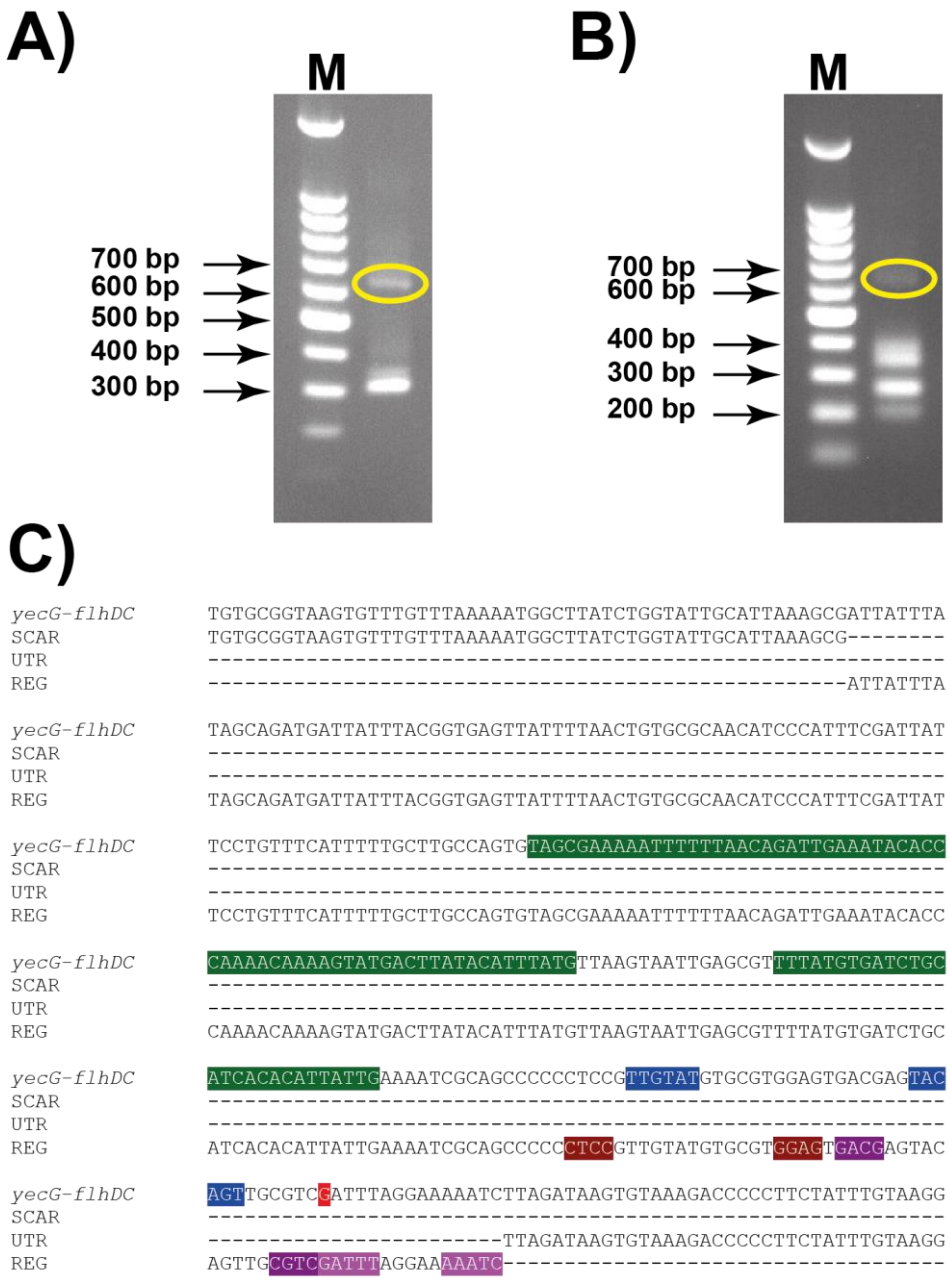


Figure. 5.4. Gel images showing the multiple bands during the creation of ΔREG mutants and DNA alignment showing the hairpins.

Multiple bands in a 1 % agarose gel after the sewing of two DNA fragments to create ΔREG mutants in **A)** MG1655::P_{flhDC}(B2) and **B)** MG1655. The correct band size of the expected PCR product is 630 bp encircled with the yellow oval shapes. M: DNA ladder. **C)** An alignment showing the REG region and the *flhDC* promoter defined in figure 5.1. The green shaded boxes show the CRP binding site upstream of the *flhD* promoter region. The -35 and -10 promoter sequences are shaded in blue. The red highlighted "G" is the transcription start site. At this end of the REG region there are three sequential hairpins (pink, purple and brown) which hindered the successful use of PCR primers. All three hairpins were confirmed using the DNA version of mFold (http://unafold.rna.albany.edu/) with all settings set as default.

5.2.3. Dissection of the yecG-flhDC SCAR region

The *flhDC* operon of *E. coli* is under multiplex adjustment by different proteins in which the SCAR region plays an important role in upregulating or downregulating the process of flagellar synthesis. As we have described earlier, regulators such as OmpR, LrhA, H-NS and cAMP-CRP do bind to specific sites in the regulatory region and therefore regulate flagellar gene expression following a variety of environmental cues. In our analysis of the *yecG-flhDC* intergenic region, we have noticed increased motility in SCAR deletion mutants, and we, therefore, proposed that the SCAR region has a repression impact in the wild type strains. Thus, we wished to analyse the SCAR region further to determine what and which part of this region is suppressing motility. We divided the SCAR region into six zones (A, B, C, D, E and F) each of which is 40 bp long, and we aimed to generate a series of deletions (Figure. 5.5A). We designed a group of primers for this purpose and we utilized genomic DNA of strain MG1655 and strain MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2 as templates to amplify the SCAR zone fragments and to subsequently electroporate them to strain MG1655ΔP$_{flhDC}$::Cm using the CRISPR-Cas technique.

We then confirmed deletion mutants by culture PCR and by running the fragments on the agarose gel. The gel image showed different DNA fragments as a ladder with size increasing by 40 bp for both MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2 and strain MG1655 (we were unable to produce ΔSCAR-C in the B2 version) (Figure. 5.5B and C). We have also confirmed the right transformation by sending the DNA samples off for sequencing.

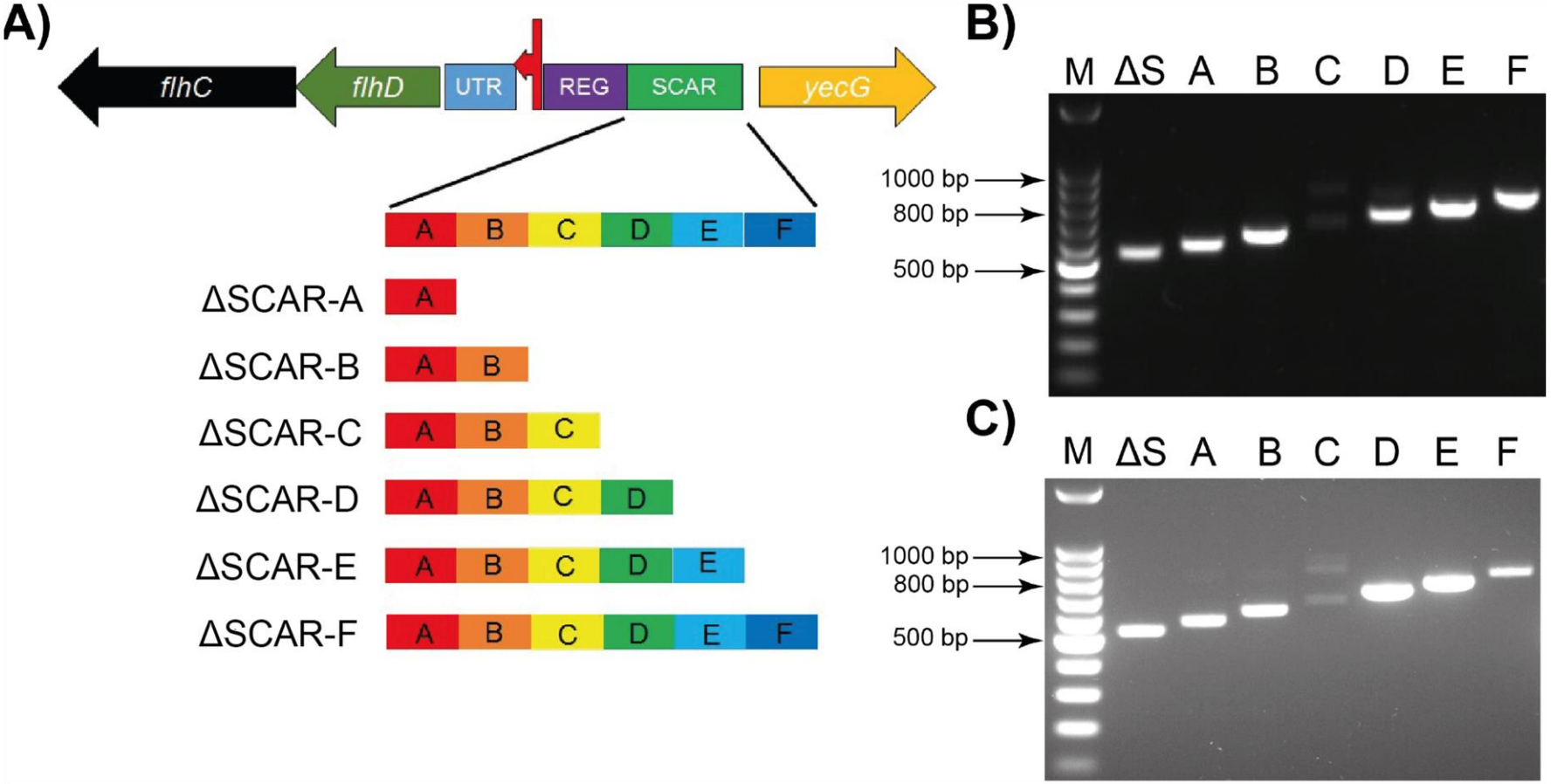


Figure. 5.5. Dissection of the *yecG-flhDC* SCAR region in MG1655 and MG1655::P_{flhDC}(B2).

A) SCAR zone of strains MG1655 and MG1655::P_{flhDC}(B2) was dissected into six fragments (A to F). a group of primers was designed to create different SCAR deletion mutants (ΔSCAR-A to ΔSCAR-F) by adding 40 bp at a time. **B)** A 1 % agarose gel showing different SCAR deletion DNA fragments in strain MG1655::P_{flhDC}(B2). **C)** A 1 % agarose gel showing the same DNA fragments in strain MG1655. SCAR deletion mutants (ΔS) were included as controls. M: DNA ladder.

We wanted then to quantify motility of ΔSCAR-A through ΔSCAR-F and therefore we used soft agar media for this reason. Our data of strain MG1655ΔP_{flhDC}::P_{flhDC}-B2 showed an increase in swarming diameter of 2-fold, 2-fold, 2-fold, 2-fold and 1.5-fold for ΔSCAR-A, ΔSCAR-B, ΔSCAR-D, ΔSCAR-E and ΔSCAR-F, respectively compared to the wild type strain (Figure. 5.6). The data set in figure. 5.6 is significant ($P< 0.0001$). The difference is not significant for ΔSCAR-A:B:D:E ($P= 0.149$) and is not significant for ΔSCAR:ΔSCAR-F ($P= 0.286$). In the same way, strain MG1655 exhibited an increase in its SCAR region fragments mutants' motility. The motility increased 3.4-fold, 5-fold, 2.3-fold, 8-fold, 6.6-fold and 1.7-fold for ΔSCAR-A, ΔSCAR-B, ΔSCAR-C, ΔSCAR-D, ΔSCAR-E and ΔSCAR-F, respectively in comparison to the wild type strain (Figure. 5.7). All of the data in figure. 5.7 is significant ($P< 0.0001$), whereas in pairs ($P< 0.001$).

These experiments showed consistent response to deletion for both strain MG1655 which belong to clade A, and strain MG1655ΔP_{flhDC}::P_{flhDC}-B2 which its promoter belong to clade B2. Since deleting the SCAR fragments from A through E displayed the highest shift in motility compared to fragment F, we proposed that something included in this fragment is repressing the motility in both strains. This was clear as the motility decreased when we added fragment F to the SCAR region. Based on the proximity of fragment F to the *yecG* promoter, we propose this zone to impact the *yecG* promoter. Therefore, we conclude that *yecG* is responsible for this motility phenotype.

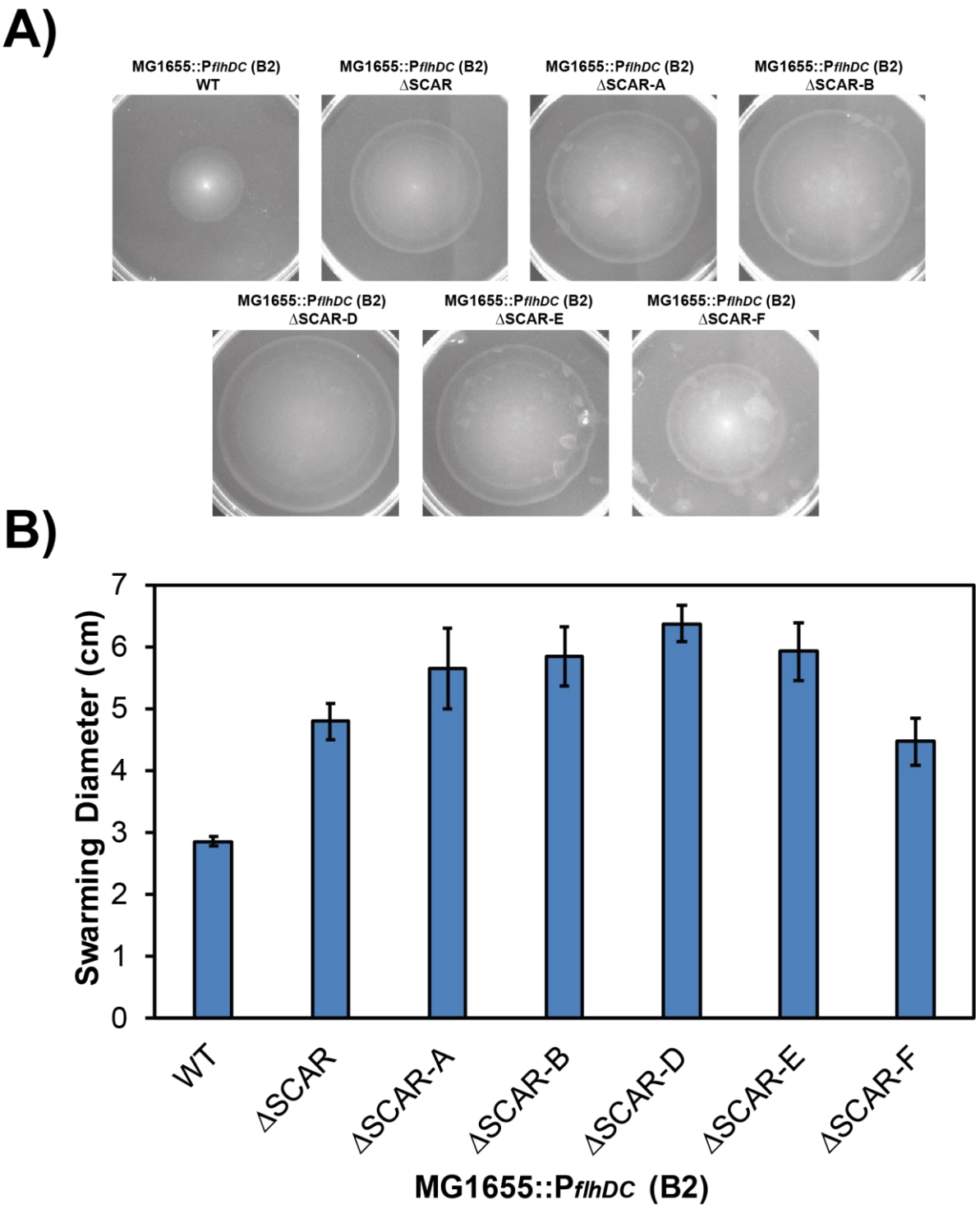


Figure. 5.6. Motility quantification of SCAR fragments deletion mutants of strain MG1655::P*flhDC* (B2).

A) Motility phenotypes. Colonies of each strain were stabbed onto motility plates containing 0.3 % agar, incubated for 8 hours at 37 ˚C. **B)** The average of swarming diameter in centimeter. ΔSCAR-F scored the lowest value compared to other deletion fragments. ΔSCAR included as a control. The values are mean ± standard deviation of three independent experiments.

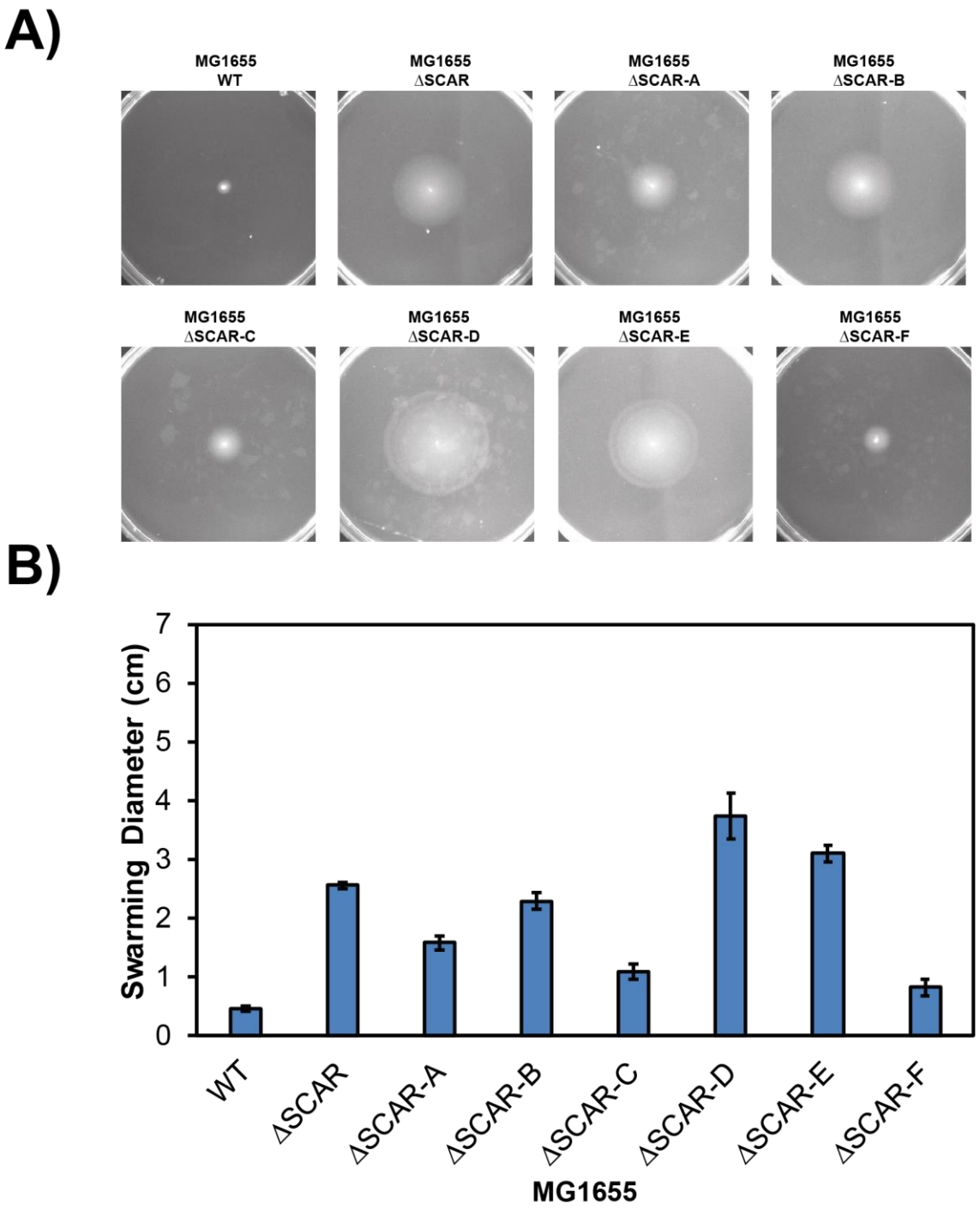


Figure. 5.7. Motility quantification of SCAR fragments deletion mutants of strain MG1655.

A) Motility phenotypes. Colonies of each strain were stabbed onto motility plates containing 0.3 % agar, incubated for 8 hours at 37 ˚C. **B)** The average of swarming diameter in centimeter. ΔSCAR-F scored the lowest value compared to other deletion fragments. ΔSCAR included as a control. The values are mean ± standard deviation of three independent experiments.

5.2.4. Measuring the activity of flgB and fliC promoters

We mentioned earlier that *E. coli* has three flagellar genes classes: early, middle and late according to the sequence of their expression. These genes are expressed as a result of the activity of their class 1, class 2 and class 3 promoters, respectively (Chilcott and Hughes, 2000). We also stated that the master regulator FlhDC is expressed following internal and external cues. This protein acts in conjunction with σ^{70} to induce the class 2 promoter which expresses genes needed for hook and basal body assembly, in addition to FlgM and FliA (σ^{28}). Upon completion of hook-basal body, σ^{28} becomes free and activates class 3 promoters which can lead to the activation of late flagellar genes, and the filament will be synthesized.

Here, we wished to examine the other two classes of flagellar promoters by measuring their activity in our deletion mutants with regard to motility. We have chosen the *flgB* promoter and the *fliC* promoter as representatives for class 2 and class 3 promoters, respectively. To monitor these promoter activities, we used the same previously described pSB401 which carry luciferase operon *luxCDABE* as gene expression reporters. As these promoters are conserved in *E. coli* as a species, we amplified them from strain MG1655, and we mounted them onto their vector using Gibson cloning. We then chemically transformed pSB::P_{flgB} and pSB::P_{fliC} plasmids into WT, ΔSCAR, ΔRG-SCR, ΔUTR, ΔUTR-SCR and ΔU-R-S of strains MG1655 and MG1655ΔP_{flhDC}::P_{flhDC}-B2. We also transformed these two plasmids into the other set of our SCAR region deletion mutants (ΔSCAR-A through F) for strain MG1655 and its B2 version. The strain MG1655ΔP_{flhDC}::Cm was our negative control to subtract the background bioluminescence from the data collected.

Our data indicated that the activities of P_{flgB} and P_{fliC} promoters in strains ΔRG-SCR, ΔUTR and ΔU-R-S were very low in both strains MG1655 and its B2 version compared to their wild type strains activity. Furthermore, higher activity of UTR-SCR

mutants was recorded for both strains with values exceeded, to some extent, the activity of promoters in their wild type strains (Figure. 5.8). Interestingly, SCAR region mutants documented the highest values for P_{flgB} and for P_{fliC} for strains MG1655ΔP_{flhDC}::P_{flhDC}-B2 and MG1655 (Figure. 5.8). The data for A and B is significant ($P< 0.0001$) and this is not surprising as this matches the motility data for these mutants. The impact of the UTR mutants is also seen significant ($P< 0.0001$), except in B2 for P_{flgB} is not significant ($P= 0.08$). In addition to that, SCAR fragment mutants (A to F) scored high bioluminescence values for both strains compared to their wild types. However, SCAR-F mutants showed the lowest activity among other mutants in the two strains compared to the wild types (Figure. 5.9). All data in figure. 5.9 is significant. In strain MG1655, D:F is significant ($P= 0.00047$), E:F is not significant ($P= 0.176$), WT:F is significant ($P= 0.05$), SCAR:F is not significant ($P= 0.236$) and WT:SCAR is significant ($P= 0.0007$) (Figure. 5.9A).

We have managed, by measuring bioluminescence, to quantitatively diagnose *flgB* and *fliC* promoter activities using these reporter plasmids and *lux* genes. As we recall from earlier data (Figure. 5.3), UTR, RG-SCR and U-R-S mutants of strain MG1655 and its B2 version were nonmotile. While UTR-SCR mutants exhibited similar motility phenotype as their wild types, SCAR region mutants were displayed the highest motility. Moreover, the figures 5.6 and 5.7 showed dramatic shift in motility with lowest activity of SCAR-F. Here, the data successfully proved consistent behaviour between the activity of promoters in the deletion mutants and motility phenotype of the strains.

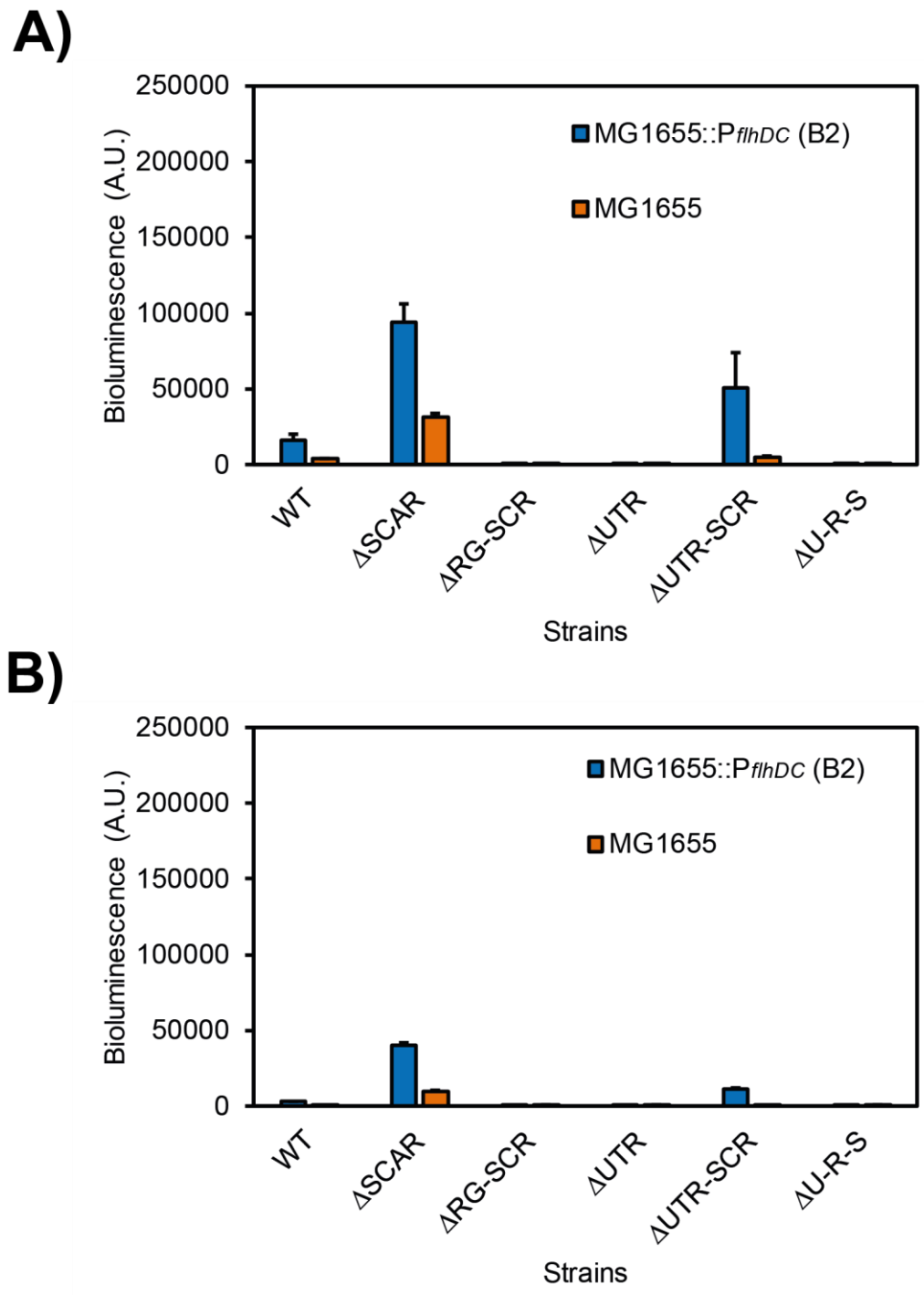


Figure. 5.8. Activity of P_{flgB} and P_{fliC} in *yecG-flhDC* zones deletion mutants.

A) P_{flgB} and **B)** P_{fliC} cloned to pSB401, and plasmids transformed into MG1655::P_{flhDC}(B2) and MG1655 that have different deletion mutants in *yecG-flhDC* intergenic region. Cultures were made and started from OD_{600}=0.05 until they reached OD_{600}=0.6-0.8, and measured in microplate reader. Promoter activity reflected motility phenotypes. The values are mean ± standard deviation of three independent experiments.

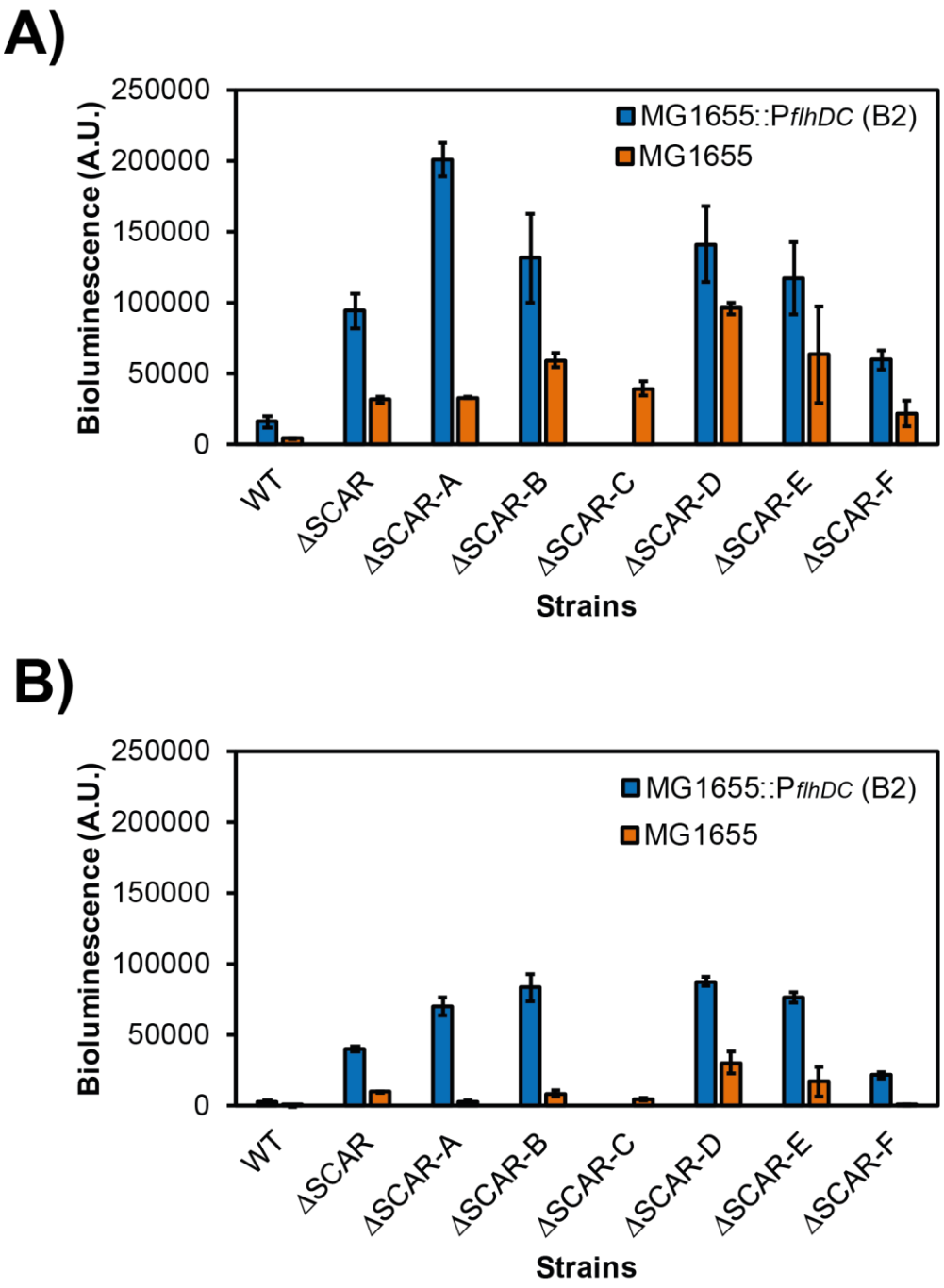


Figure. 5.9. Activity of P_{flgB} and P_{fliC} in SCAR fragments deletion mutants.

A) P_{flgB} and **B)** P_{fliC} cloned to pSB401, and plasmids transformed into MG1655::P_{flhDC}(B2) and MG1655 that have different deletion mutants in SCAR zone. Cultures were made for each strain and started from OD_{600}=0.05 until they reached OD_{600}=0.6-0.8, and measured in the microplate reader. Promoters activity reflected motility phenotypes. The values are mean ± standard deviation of three independent experiments.

5.2.5. Does yecG repress flagellar gene expression?

The Usp universal stress protein superfamily is an ancient and sustained class of proteins that are found in bacteria, archaea, fungi, flies and plants (Kvint *et al.*, 2003). *E. coli* has six Usp proteins which can be split into two subfamilies based on sequence similarities of the Usp domain. UspA, UspC (*yecG*) and UspD (*yiiT*) belong to one subfamily while UspF (*ynaF*) and UspG (*ybdQ*) belong to the other subfamily (Gustavsson *et al.*, 2002). Moreover, Gustavsson *et al.* (2002) allocated the sixth protein UspE (*ydaA*) in between these two subfamilies due to shared characteristics, and they suggested that these proteins have a role in the DNA damage control. The protein YecG (UspC) plays an important role in stress conditions of *E. coli* (Gustavsson *et al.*, 2002). Under salt stress especially potassium ion (K^+) deficiency, YecG acts as a stabilizing scaffold for the sensor kinase KdpD and the response regulator KdpE by binding to a Usp domain in KdpD. As a result, the *kdpFABC* operon will be expressed and the KdpFABC K^+ transport system will uptake K+ for the cell from other transporters (Heermann *et al.*, 2009). Notwithstanding, Heermann *et al.* (2009) mentioned that the exact function of these proteins is unknown. A previous work has dealt with the deletion of *uspC* in strain MG1655, and proved that mutants were nonmotile. The electron microscope showed images for *uspC* deletion mutants lacking flagella (Nachin *et al.*, 2005).

Although all SCAR deletion mutants displayed higher motility phenotype compared to the wild type, fragment F presented the lowest value among the others. Therefore, our observation suggested negative regulatory effect on motility after adding fragment F to the SCAR region. For the reason that this fragment is very close to the *yecG* promoter (Conway *et al.*, 2014), we turned our attention to the opposite end of the *yecG-flhDC* intergenic region which is *yecG*. We, therefore, replaced *yecG* with a DNA cassette which has a chloramphenicol resistance gene,

amplified from plasmid pWRG100 using Datsenko and Wanner method. We then electroporated the cassette into the strains with the help of pKD46 that has λ-red activity by adding arabinose to the culture media. Motility assays were then performed using soft agar media. Our data showed that *ΔyecG* mutants displayed enhanced swimming capacity (Figure. 5.10). The increase in motility was 1.4-fold in strain MG1655ΔP_{flhDC}::P_{flhDC}-B2 compared to the wild type. Motility in strain MG1655 *ΔyecG* mutant increased 4.3-fold compared to the wild type (Figure. 5.10B). Thus, the impact of the *yecG* deletion is more visible in strain MG1655 (P= 0.0026), whereas the difference is not significant for B2 version (P= 0.136). The data for all is significant (P= 0.00026).

We, therefore, demonstrate that *yecG* is a negative regulator of flagellar assembly in our MG1655 strain and also in a strain which has the *yecG-flhDC* intergenic region belonging to clade B2. This data therefore does not seem to correlate with the previous work of Nachin *et al.* (2005) that *yecG* mutants were completely nonmotile due to lack of flagella. However, it is not known whether YecG directly supresses *flhDC* expression or it has indirect effect on flagellar assembly. Additionally, it is not recognised what causes this enhancement in motility whether it was due to increase in flagellar activity or increased number of functional flagella per cell.

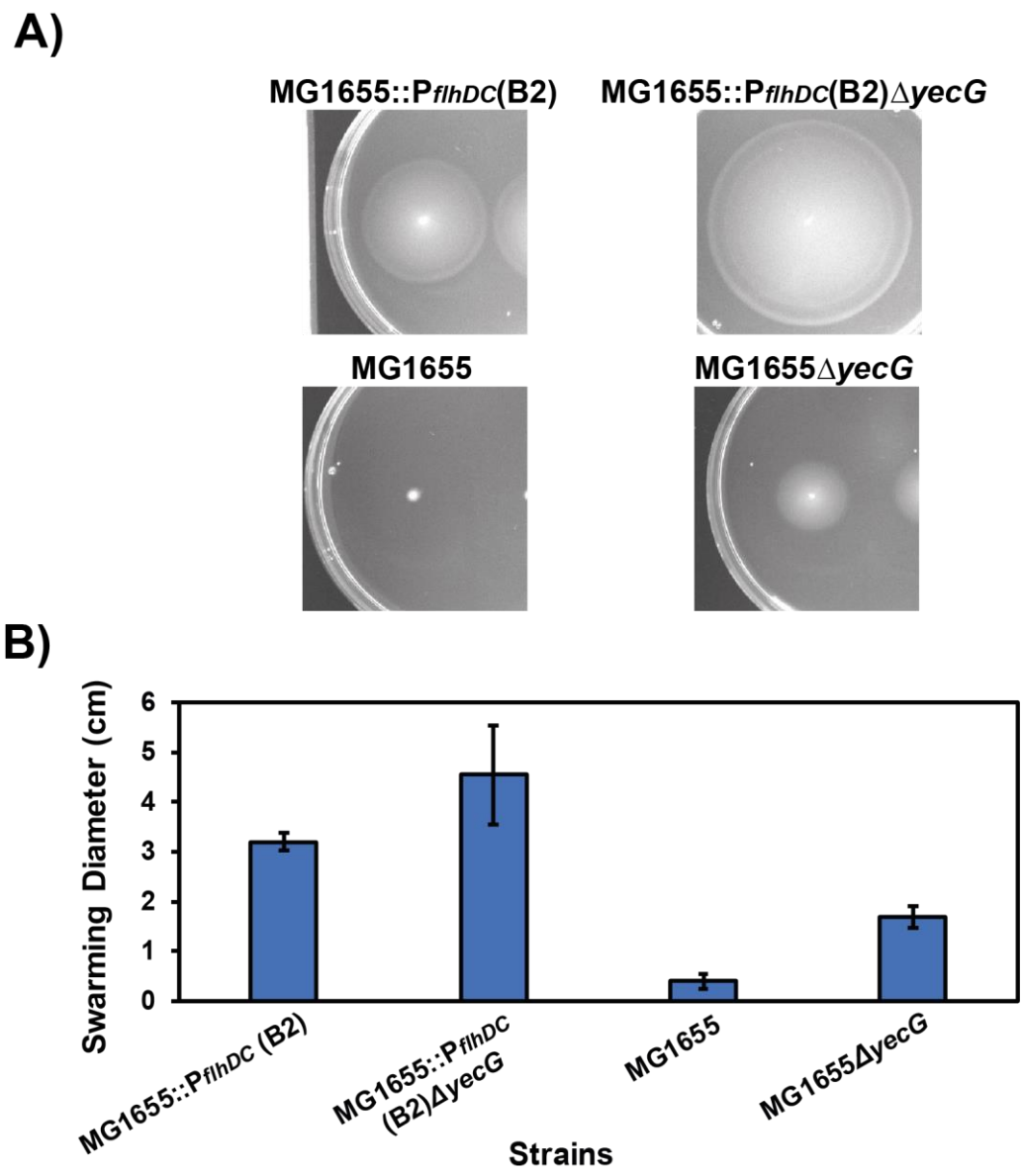


Figure. 5.10. Motility quantification of *yecG* deletion mutants in strains MG1655::P*flhDC*(B2) and MG1655.

A) Motility phenotypes. Colonies of each strain were stabbed onto motility plates containing 0.3 % agar, incubated for 8 hours at 37 ˚C. **B)** The average of swarming diameter in centimeter. Data show increased motility in *ΔyecG* mutants compared to the wild type strains. The values are mean ± standard deviation of three independent experiments.

5.2.6. Replacing the SCAR region of strain MG1655

Our data revealed that strains MG1655 and MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2 *yecG* mutants showed increased motility compared to their wild type strains. We also discussed in the preceding section the impact of different mutants including *yecG* upon flagellar gene expression summarised in (Figure. 5.11). As the *yecG* promoter is very close to the SCAR region of the *yecG-flhDC* intergenic region, we wished to test the impact of switching the SCAR region of strain MG1655 with the SCAR region of strain CFT073 on motility.

For this purpose, we used our MG1655 strains lacking the *yecG-flhDC* intergenic region that we created in the early steps of this project. We designed two sets of primers to create PCR products of the *flhDC* promoter region without SCAR for strain MG1655. While the second set of primers was to create a PCR product includes the SCAR region of strain CFT073. Then, we sewed the two amplified DNA to create a segment which is SCAR-CFT073_MG_*flhDC*. The CRISPR technique was used here to electroporate this product into MG1655ΔP$_{flhDC}$::Cm. After performing confirmation steps, the newly created strain was tested in the motility assay. The strains MG1655, MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2 and MG1655ΔSCAR were used as controls for this assay (Figure. 5.12).

The output of our data regarding the control strains correlates with previous experiments. The strain MG1655 exhibited poor motility on the soft agar plate with swarming diameter of 0.481 cm, while MG1655ΔSCAR unveiled 427.14 % increase in motility with comparison to the wild type strain. Moreover, the strain MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2 showed 475.71 % motility increase compared to the wild type with a swarming diameter of ~ 2.77 cm. Interestingly, our newly created strain MG1655-*flhDC*ΔSCAR::SCAR-CFT073 was completely nonmotile (Figure. 5.11 and 5.12).

We assumed that adding the SCAR region of CFT073 to strain MG1655ΔSCAR will turn the newly created strain to have the same motility phenotype as strain CFT073 or strain MG1655ΔP_{flhDC}::P_{flhDC}-B2. Surprisingly, the new strain had swarming diameter of 0.28 which is the diameter of the woody tooth pick used to stab the bacteria in the motility agar plate. This outcome led us to suppose that the suppression of flagellar gene expression in the SCAR region of strain CFT073 is higher than the SCAR region in strain MG1655. We believe that the DNA sequence dissimilarities occur in the SCAR region of strains MG1655 and CFT073 could affect the binding of regulators to this region. Could these changes reflect this negative regulation? We also know that the SCAR region is very close to the *yecG* promoter which has a negative impact on flagellar gene regulation. Does the *yecG* promoter of strain CFT073 have more strength than that of strain MG1655? If the answer was yes, then this will explain why our new strain turned out to be completely nonmotile. To take this analysis further, we had to measure the *yecG* promoter activity of strains MG1655 and CFT073.

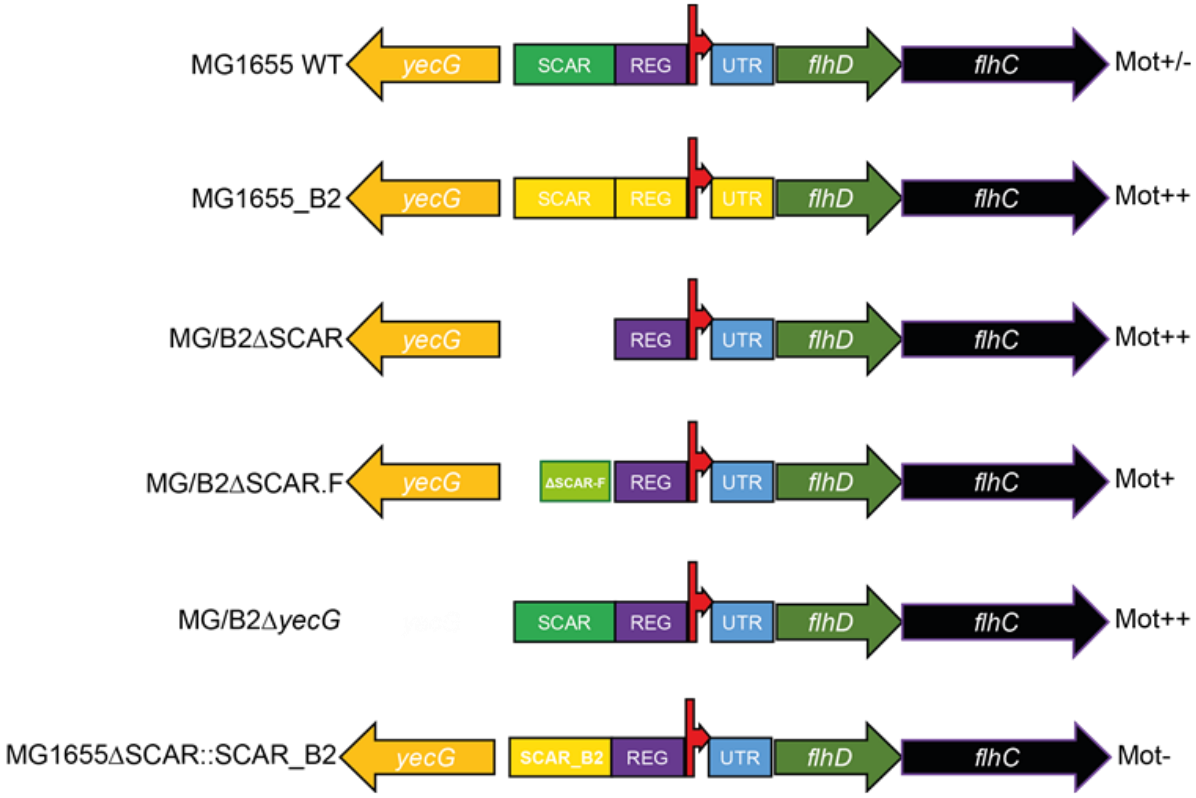


Figure. 5.11. Schematic representation review of motility phenotypes of miscellaneous strains.

The drawing shows the weak motility of our strain MG1655 wild type (Mot+/-), and the high motility after replacing its *flhDC* promoter with that of strain CFT073 which belongs to clade B2 (Mot++). High motility was also observed when we deleted the SCAR zone in strains MG1655 and MG1655::P_{flhDC} (B2) (Mot++). Then, we identified the part in the SCAR zone which exhibits the lowest motility among other parts (fragment F which is very close to the *yecG* promoter) (Mot+). As a result, we concluded that *yecG* may negatively impact the *flhDC* promoter. Thus, we quantified motility of *ΔyecG* mutants and confirmed the repression effect of *yecG* upon motility (Mot++). In the same regard, we replaced the SCAR zone of MG1655 with the SCAR zone of MG1655::P_{flhDC} (B2) and found that this strain turned out to be nonmotile (Mot-).

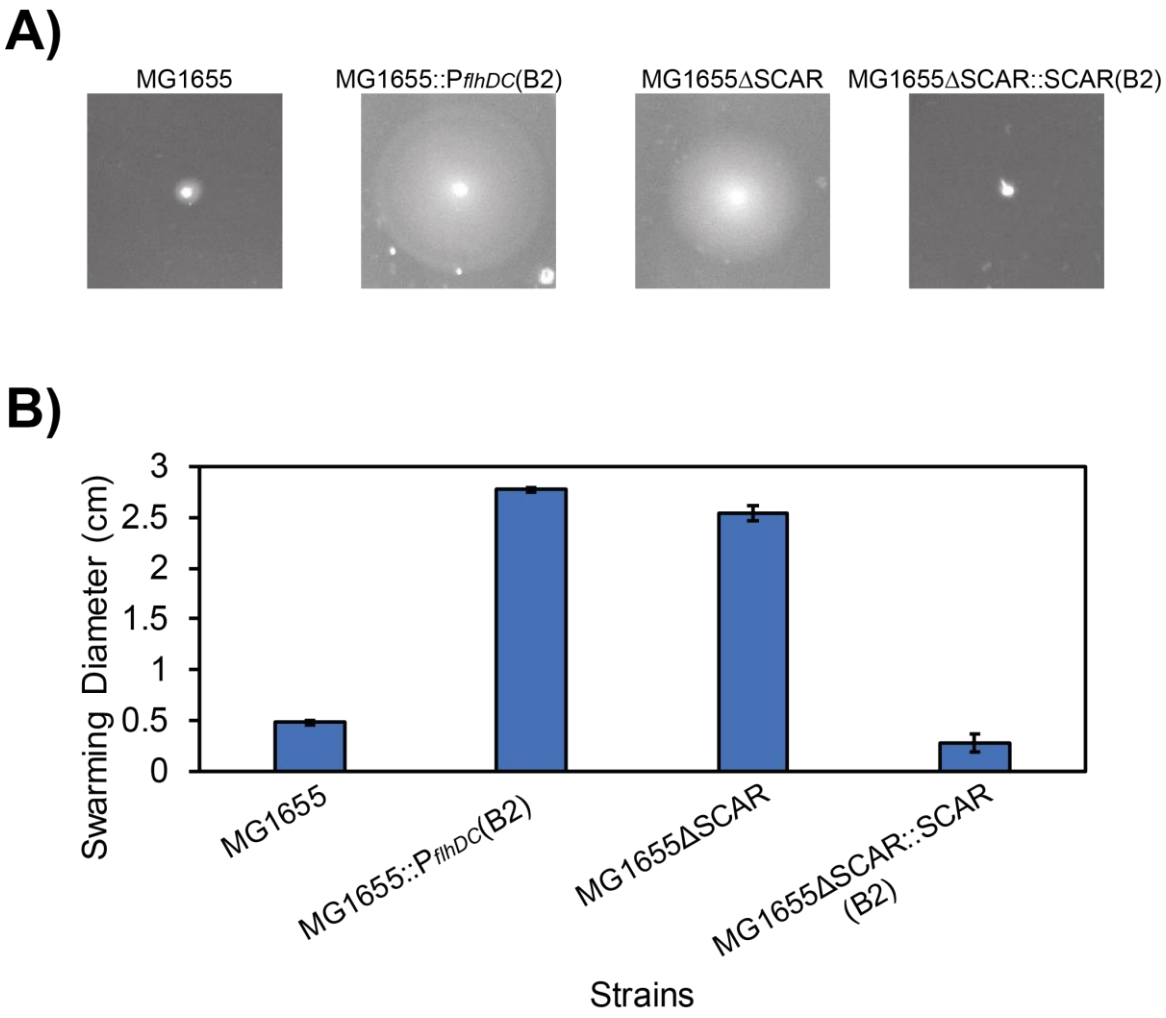


Figure. 5.12. Impact of replacing the SCAR region in strain MG1655 with the SCAR region of strain MG1655::P_{flhDC} (B2) on motility.

A) Motility phenotypes. Colonies of each strain were stabbed onto motility plates containing 0.3 % agar, incubated for 8 hours at 37 ˚C. **B)** Average of swarming diameter in centimeter. Data reveal that strain MG1655ΔSCAR::SCAR(B2) turned out to be completely nonmotile. Strains MG1655, MG1655::P_{flhDC} (B2) and MG1655ΔSCAR were included in this experiment as controls. The values are mean ± standard deviation of three independent experiments.

5.2.7. Strains MG1655 and CFT073 P_{yecG} activity

The previous section (Figure. 5.12) showed that replacing the SCAR region of strain MG1655 with the SCAR region of strain CFT073 inhibited the motility of the former strain. We stated that *yecG* could be the reason behind this scenario as its promoter lies near the SCAR region and acts as a possible suppressor for motility. Therefore, we wanted to measure the activity of the *yecG* promoter in both MG1655 and CFT073. We used the plasmid pSB401 that carry *lux* genes for this purpose.

By using Gibson cloning, we inserted the *yecG* promoter of each of MG1655 and CFT073 into pSB401. The cloning details were the same as what we did with the cloning of the *yecG-flhDC* intergenic region (Figure 4.4), but in the opposite direction. The two plasmids pSB401::P_{yecG}-MG1655 and pSB401::P_{yecG}-CFT073 were then introduced to MG1655 (WT, ΔSCAR and ΔSCAR-F) and also to MG1655ΔP_{flhDC}::P_{flhDC}-B2 (WT, ΔSCAR and ΔSCAR-F). While the strains MG1655ΔP_{flhDC} and MG1655::pSBΔER were used as controls. The promoter activity was measured by the bioluminescence expressed as light units (Figure. 5.13).

The *yecG* promoter activity of strain MG1655 ranged between 1773 to 3724 light units (Figure. 5.13). However, the activity in strain MG1655 derivatives was lower than that in MG1655ΔP_{flhDC}::P_{flhDC}-B2 versions. On the other hand, the activity of the *yecG* promoter of strain CFT073 was higher in the same panel strains regardless to the strain genotype with bioluminescence ranging from 4338 to 6411 light units. The activity of the two promoterless strains containing pSBΔER was extremely low as expected. MG1655WT:MG1655ΔSCAR:MG1655ΔSCR.F is significant (P= 0.045), MG-B2:MG-B2ΔSCAR:MG-B2ΔSCR.F is not significant (P= 0.116) and MG1655WT:MG-B2WT:MGΔP_{flhDC} is significant (P= 0.00001) which suggests some level of P_{flhDC} feedback (Figure. 5.13A). MG1655WT:MG1655ΔSCAR:MG1655ΔSCR.F, MG-B2:MG-B2ΔSCAR:MG-

B2ΔSCR.F and MG1655WT:MG-B2WT:MGΔP$_{flhDC}$ are not significant (P= 0.686, P= 0.332 and P= 0.076, respectively) (Figure. 5.13B). Importantly, there is a significant difference between MG1655::pSB-P$_{yecG}$-MG1655:MG1655::pSB-P$_{yecG}$-CFT073 (P= 0.00014). From the above data, we concluded that the *yecG* promoter strength of strain CFT073 is higher than that in strain MG1655.

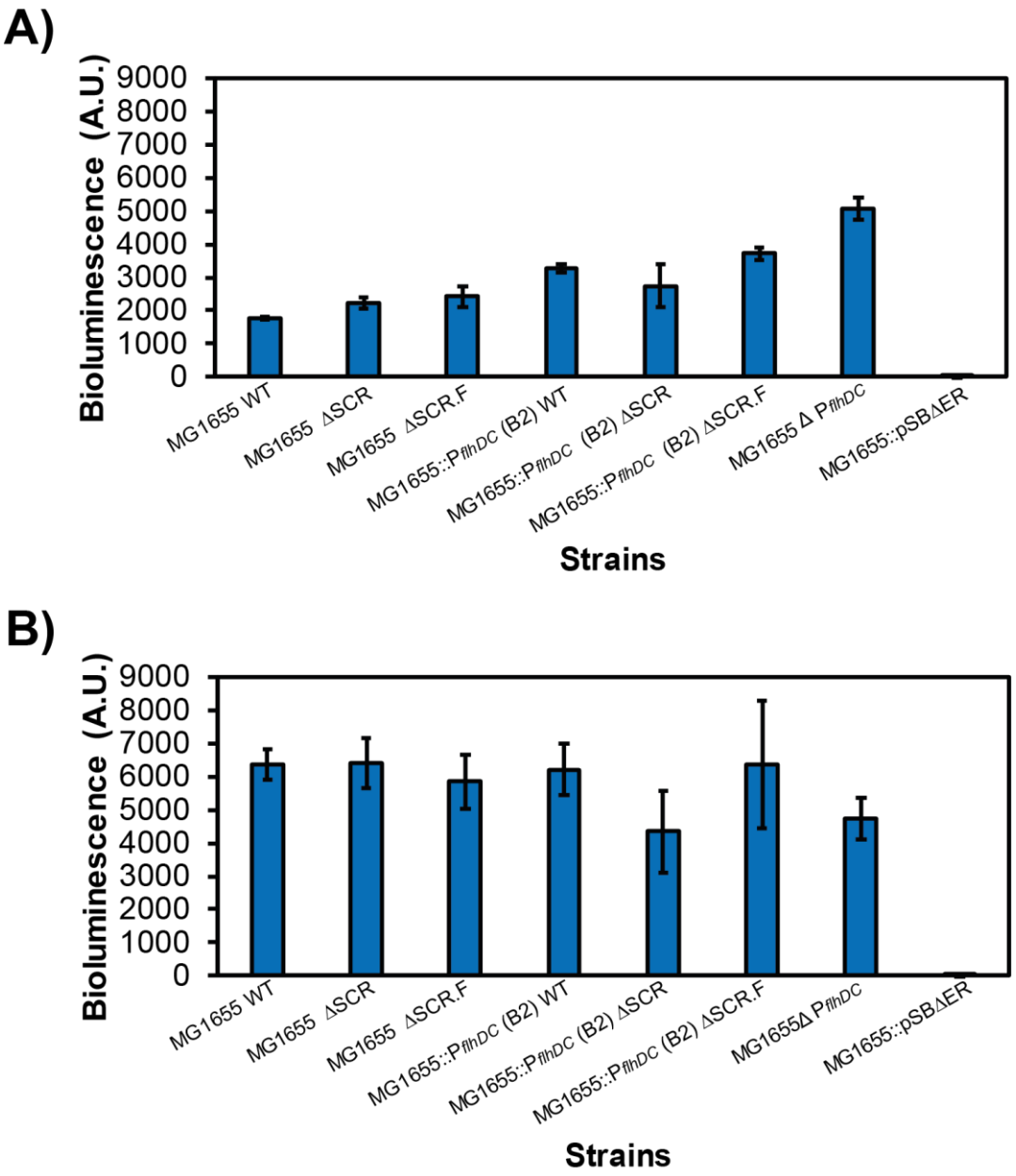


Figure. 5.13. Activity of the *yecG* promoter of strains MG1655 and CFT073.

A) Plasmid pSB401::P_{yecG}-MG1655 and **B)** Plasmid pSB401::P_{yecG}-CFT073 were transformed into MG1655 (WT, ΔSCAR and ΔSCAR-F) and MG1655ΔP_{flhDC}::P_{flhDC}-B2 (WT, ΔSCAR and ΔSCAR-F). Cultures were made and started from OD_{600}=0.05 until they reached OD_{600}=0.6-0.8, and measured in the microplate reader. Data show that *yecG* promoter activity of strain CFT073 is higher than that in strain MG1655 in all panel strains. Strains MG1655ΔP_{flhDC} and MG1655::pSBΔER used as controls. The values are mean ± standard deviation of three independent experiments. pSBΔER: plasmid pSB401 with *luxR* deleted.

5.3. Summary

In our previous chapter, we focussed on the master regulator protein which is expressed due to the activity of the *flhDC* promoter. All the enteric bacterial species must have the ability to express $FlhD_4C_2$ complex to be motile. Therefore, *flhDC* expression in *E. coli* and the resulting $FlhD_4C_2$ protein were our focal points for regulating the decision to be motile. Most of the literature are dealing with one strain of *E. coli* which is strain MG1655. However, *E. coli* is a divergent species consisting of six phylogenic groups which could cause numerous diseases to humans. Thus, these facts suggest a more comprehensive investigation of motility in *E. coli* taking into our consideration the wider image of this species not only one strain. We placed the *yecG-flhDC* intergenic region of highly motile strains from different clades instead of the one of our MG1655. The poor motility of our MG1655 strain is attributed to the lack of transposons in its SCAR region, and that was consistent with the literature which showed the importance of IS elements for increasing motility (Barker *et al.*, 2004). The swarming ability of this strain, though, enhanced after replacing its *yecG-flhDC* intergenic region with that of highly motile strains suggesting different *flhDC* promoter regulatory input. We believe that the DNA sequence dissimilarities in this region have an impact on motility.

In this chapter, we continued to utilize our genetic recombination technologies to explore the various aspects of *flhDC* expression and regulation in two recognizable *E. coli* strains MG1655 and its B2 version. The *yecG-flhDC* intergenic region is composed of the untranslated region, the *flhDC* promoter, the regulatory region and the SCAR region. Therefore, we wished to dissect this region looking for the role of each zone related to the motility of strain MG1655 and its B2 version. For this reason, we created different mutants lacking one region and also in combinations. Our data showed different motility phenotypes for these mutants due to

transcriptional and post-transcriptional reasons. Importantly, the SCAR mutants displayed the highest swarming rates compared to the wild type. We attributed this phenomenon to the removal of the binding sites for regulators of the *yecG* promoter. We verified our hypothesis about the role of *yecG* in motility when we studied the effect of various parts of the SCAR region, and we explained the lower motility after adding fragment F. To shed light into the SCAR region, the different parts of this region were investigated in more detail. The region was cut into six zones (A through F) with the size of 40 bp for each. We then created a group of strains that have one zone added to the SCAR region. After the motility assay being performed, we did indeed observe that the swarming diameter was diminished in fragment-F mutants. As a consequence of the presence of the *yecG* promoter very close to this zone, we predicted that this promoter could play a significant role in the motility of our bacterial strains. In order to confirm this further, we have deleted *yecG* from strains MG1655 and its B2 version. As a result, *ΔyecG* mutants turned to be highly motile in comparison to their wild types.

To test this hypothesis further, we swapped the SCAR region of strain MG1655 with the SCAR region of strain CFT073 which is considered highly motile. Our prediction was increased motility of strain MG1655 as its *flhDC* promoter will be under the effect of less suppression. However, data showed loss of motility of strain MG1655 (Figure. 5.12). We proposed that the SCAR region of strain CFT073 has more suppression effect than that of MG1655. To overcome this ambiguity, we wished to measure the *yecG* promoter of strain MG1655 and its B2 version to identify the activity of each promoter. The data did show higher activity of the *yecG* promoter of strain MG1655ΔP$_{flhDC}$::P$_{flhDC}$-B2 than the *yecG* promoter of strain MG1655 (Figure. 5.13). We therefore concluded that motility loss of strain MG1655ΔSCAR::SCAR-B2 is due to adding more repression power on its *flhDC* promoter.

We suppose that different motility extent in different strains is due to not only the strength of the *yecG* promoter, but also the power of the *flhDC* promoter and its regulatory region. To support this hypothesis, we have seen the higher activity of the *yecG* promoter of strain CFT073 compared to a lower luminescence value of strain MG1655 *yecG* promoter. If the *flhDC* promoters of both strains CFT073 and MG1655 and their regulatory regions have the same strength, then the motility of strain MG1655 will be higher than the motility of its B2 version as it will have less suppressive impact on its *flhDC* promoter. Nonetheless, we have recorded the opposite motility phenotypes for these two strains as mentioned earlier in our data.

Chapter Six: Characterisation of Bovine and UTI Isolates of *Escherichia coli*

6.1. Introduction

The project began when we asked: what impact did the *yecG-flhDC* intergenic region replacement in strain MG1655 have upon motility? The study has shown that *flhDC* expression can drive motility in strain MG1655 after switching the *yecG-flhDC* intergenic region. The strength of the *yecG-flhDC* intergenic region of different clades of *E. coli* reported variable *flhDC* promoter activity which can reflect the original motility phenotype of the wild strains. Moreover, we have examined the impact of global regulators on the *flhDC* promoter activity.

Later, we dissected the *yecG-flhDC* intergenic region of two *E. coli* strains belonging to two phylogenetic groups, A and B2. By doing so, we confirmed the effect of each region on motility in these strains. We also further examined the SCAR region, and reported that the area upstream of the *flhDC* promoter region has a suppressive effect on motility. We diagnosed that the other end of the *yecG-flhDC* intergenic region, *yecG*, is responsible for the regulation of flagellar gene expression. Other flagellar promoter classes (class 2 and class 3) were also investigated, and experimental data showed that activity was consistent with motility phenotype. This led to a conclusion that YecG is a negative regulator of motility.

Previous studies have shown that transposons may target a region upstream of the *flhDC* promoter, but our *yecG* data argues it is more likely that they impact the expression of *yecG* that is leading to the noted phenotypes in strains such as MG1655. Strain MG1655 is often seen as the ‘’model’’ *E. coli* with many groups worldwide focussing all their research on it. Barker *et al.* (2004) isolated IS elements by diagnostic PCR of the *flhDC* region and distinguished the presence or absence of IS elements by looking at the size of the gel image (Figure. 1.14). However, these findings cannot be extrapolated to all *E. coli* strains as a species. We have generously been given a collection of *E. coli* strains from Prof. Mark Fielder at

Kingston University. These bacterial isolates were taken at a specific depth (30-50 cm) from rectal examination of cows in several herds across the south of England. This collection of bovine isolates was pre-screened for the serotype O157:H7 to exclude EHEC strains. On the other hand, we have a collection of *E. coli* strains isolated from the urine of UTI patients over six months which are already sequence typed and assigned to the *E. coli* phylogenetic tree.

Our aim in this chapter is to shine new light on the impact of two different niches of *E. coli* on the isolation of transposons in the *yecG-flhDC* intergenic region. Moreover, we will unravel some of the mysteries surrounding the phenotypical and genetic characterisation of these strains to define them for motility, biofilm formation and assign them to the *E. coli* phylogeny by performing multilocus sequence typing (MLST).

6.2. Results and discussion

6.2.1. MLST analysis

MLST is a molecular biology technique that types multiple genetic loci. This approach uses DNA sequences of internal fragments of seven housekeeping genes in order to type microbial isolates (Wirth *et al.*, 2006). Even though conserved housekeeping gene DNA sequences are variable across a species. Moreover, each of these sequences allocated as a recognisable allele can be conferred to a certain sequence type (ST). The species of *E. coli* is highly heterogeneous with six major phylogenetic groups (McNally *et al.*, 2013). It causes a wide range of diseases to humans and animals alike, and its presence in drinking water or food indicates faecal contamination (Clermont *et al.*, 2015). Many techniques in molecular biology have been developed to assign the bacterial isolates to their clades. This strategy is especially important because the tendency of the isolates to cause diseases, it's

environmental niche and behaviour differs depending on the phylogenetic emergence (Clermont *et al.*, 2015).

The characterisation of *E. coli* isolates and assigning them into their corresponding clades is a convenient bacterial source tracking tool for faecal contamination (Carlos *et al.*, 2010). Three MLST schemes for *E. coli* are available carried by Michigan State University, Pasteur Institute and Warwick Medical School (Clermont *et al.*, 2015). These schemes are established on different gene combinations which share the gene *icd*. Despite the ambiguity of choosing such genes, they are all known as housekeeping genes (Clermont *et al.*, 2015). It is also worth mentioning that the match noticed in ST nomenclature when compared to whole genome-based phylogeny with these different MLST schemes was the highest for Warwick MLST tool (Sahl *et al.*, 2012).

In this project, we implemented the Warwick MLST tool to assign 94 bovine *E. coli* isolates to their phylogenetic groups based on their chosen genetic markers. By aligning DNA sequences of the bovine isolates of *E. coli* of this project to those of *E. coli* strains studied by McNally *et al.* (2013) using Clustal X software, we were able to pinpoint the ST numbers and clades of our bovine isolates. Our MLST data showed well grounded distribution of these bovine isolates between different clades (Figure. 6.1). Most of the sequence types isolated were associated with phylogenetic group B1. The data was consistent with literature as clade B1 is the home of *E. coli* strains that live in the gut of animals as normal microflora (McNally *et al.*, 2013). Moreover, *E. coli* strains isolated from the intestine of humans, chicken, cows, goats, pigs and sheep have been used to track the source of faecal contamination (Carlos *et al.*, 2010). The study of Carlos *et al.* (2010) indicated that all these hosts had isolates belong to B1 clade, but these isolates were widely spread in cows, goats and sheep specimens (Carlos *et al.*, 2010). Other STs were allocated to clade A and clade E as

in figure 6.1. Clade A is considered as a sister to group B1 as they have high similarities in common. Intestinal commensal *E. coli* strains such as the well-known lab model K-12 MG1655 belong to clade A (McNally *et al.*, 2013). A few isolates were sitting in clade E which considered as the home of EHEC *E. coli* strains. These strains could exist in the gut of ruminants and they can infect humans via eating or drinking contaminated food or water (Ferens and Hovde, 2011). In addition to the distribution of these isolates on the major phylogenetic groups, one sequence type (ST7207) was undefined.

A recent study carried out in our lab focussed on *E. coli* isolates that cause UTI allowed us to gain an interesting perspective for the distribution of these isolates on the phylogenetic groups. This study reflected previous studies in that what we find associated with a human bladder is biased towards clade B2 (McNally *et al.*, 2013). However, other phylogenetic group members such as B1, E, D/F can be easily identified. When we use McNally's strains alongside our bovine isolates in comparison to rUTI isolates, the resulting tree highlights the impact of environmental source (Figure. 6.2). When we compared *E. coli* isolates from these two different host environments known to be preferential sites of colonisation, bovine isolates mostly occupied B1 and A clades with 56.4 % and 35.1 %, respectively. The percentage of the bovine isolates belonging to clade E was 6.4 %, whereas the percentage of the undefined isolates was 2.1 %. On the other hand, the majority of UTI isolates were located in clade B2. Therefore, this study makes a major contribution to research on *E. coli* by demonstrating that different phylogenetic groups reflect their environmental sources.

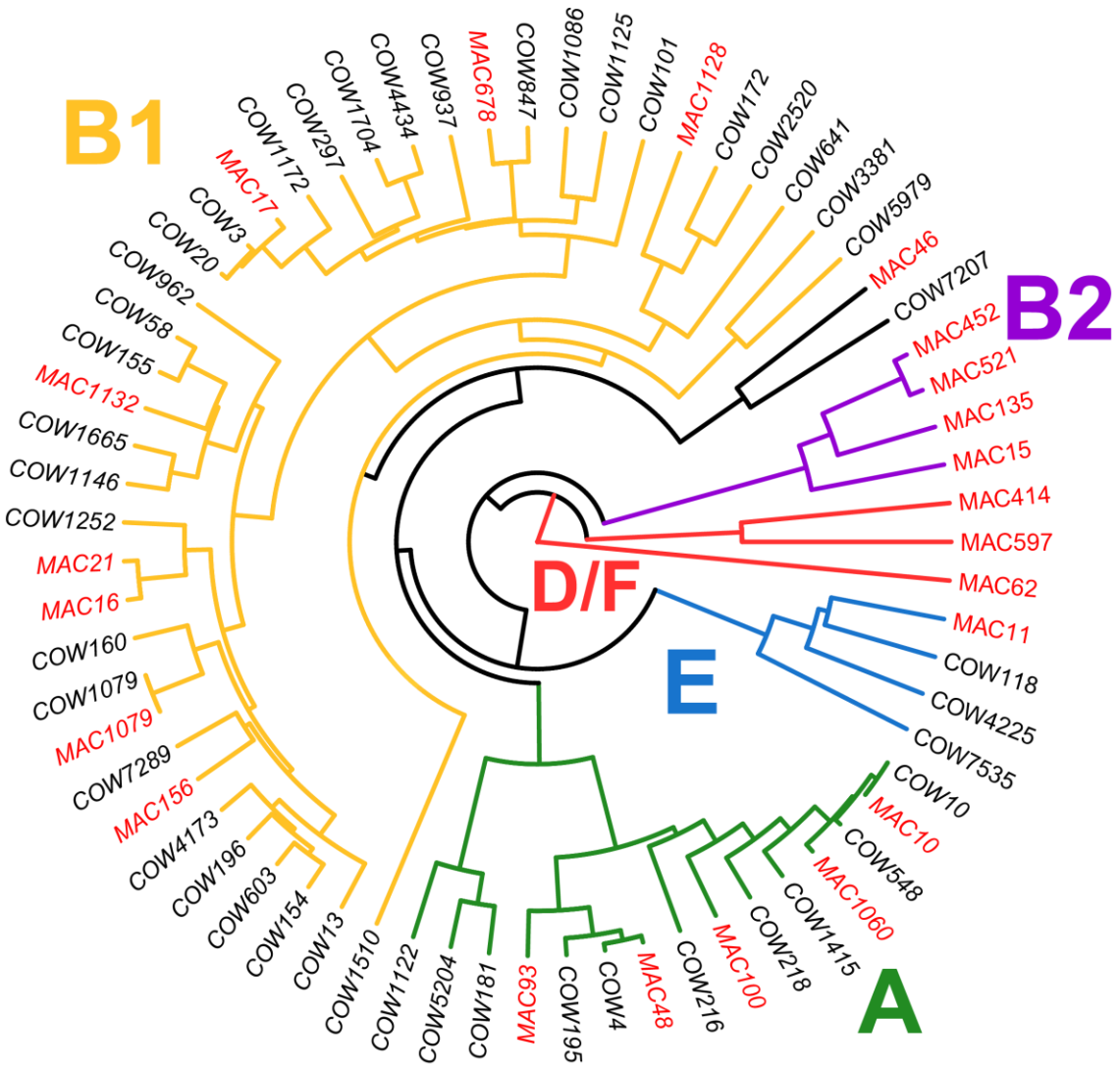


Figure. 6.1. Distribution of bovine isolates on the phylogenetic groups of McNally *et al.* (2013).

Tree shows the six phylogenetic groups of *E. coli*: B2, B1, A, D, E and F. The bovine isolates (black color) lie mostly in clades B1 and A. The original strains of McNally *et al.* (2013) are in the red color.

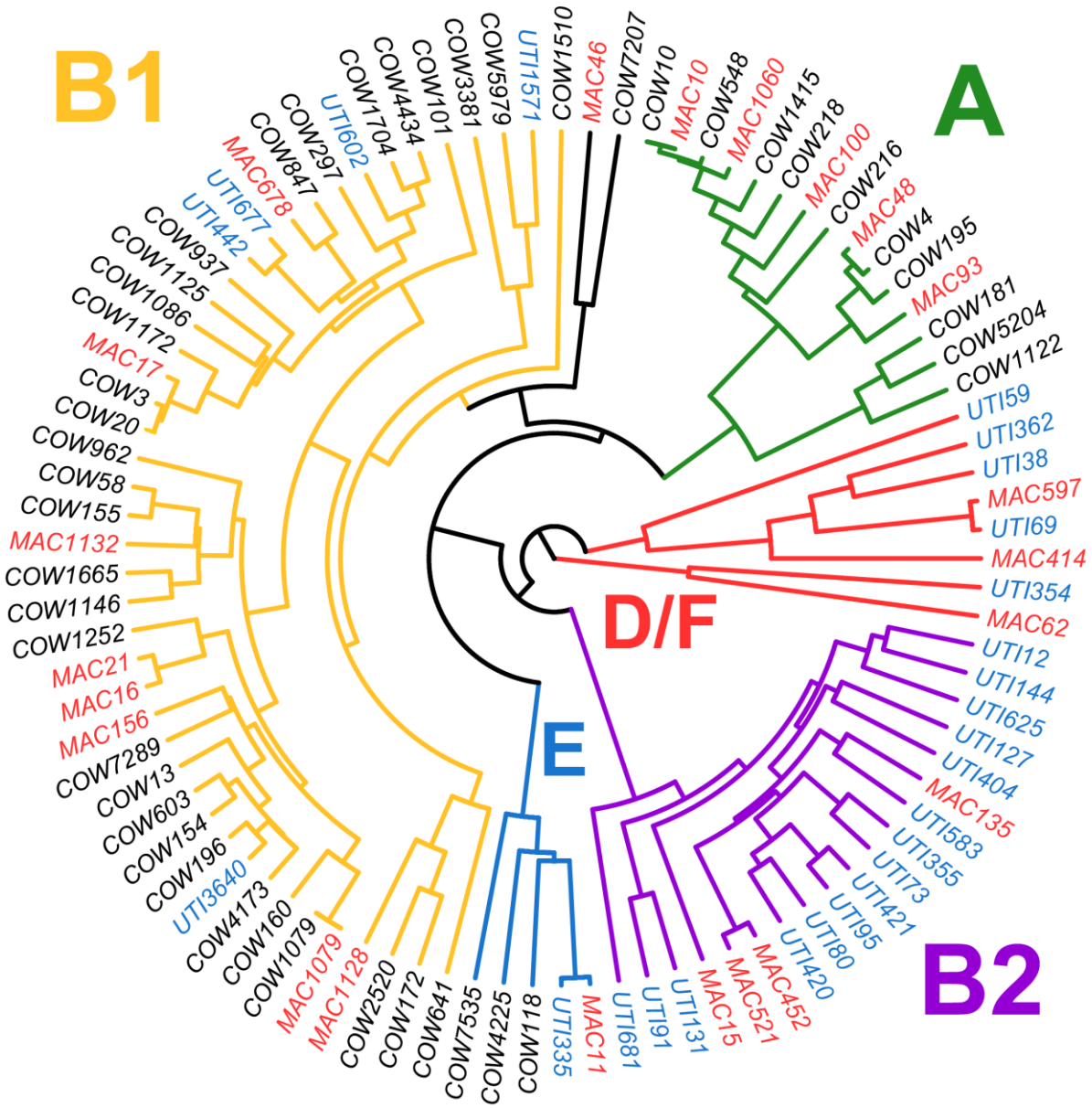


Figure. 6.2. Distribution of bovine and UTI isolates on phylogenetic groups of McNally *et al.* (2013).

Tree shows the six phylogenetic groups of *E. coli*: B2, B1, A, D, E and F. The bovine isolates (black color) lie mostly in clades B1 and A, whereas the majority of UTI isolates occupy clade B2 (blue color). The original strains of McNally *et al.* (2013) are in the red color.

6.2.2. Drawbacks associated with MLST

As a powerful tool to analyse and track the origin of bacterial isolates and assign them to their phylogenetic groups, MLST works by aligning DNA sequences of specific genetic markers (Maiden *et al.*, 1998). In order to give a wide picture with high resolution, there should be a big number of bacterial isolates available for the MLST analysis. Thus, a number of caveats need to be noted regarding the present study. The big drawback is that researchers argue that this method requires both DNA strands to be sequenced to gain accurate data. In our case, we argue that the quality of data obtained from sequencing one DNA strand is very high to get what we needed. However, it means our phylogeny can only be defined as an estimate.

Another issue arose during working on MLST was the cost. To complement such type of experiments, we generated ~ 700 PCR products to send off for sequencing beside other consumable materials required. As a result, we were unable to involve all the strains we have collected for MLST. The cost issue also affected how we extracted genomic DNA to run PCR reactions. We used culture PCR method instead of proper genome preparation using kits to get DNA from the cells. Another limitation arose when we sent hundreds of DNA samples off for sequencing company. We adjusted DNA concentration to be sent off for sequencing according to their sample requirements using our Nanodrop ND-1000 spectrophotometer. Nevertheless, the sequencing company informed us that they are reading much higher concentrations than ours. This variation in equipment's sensitivity in reading DNA sample concentration led us to dilute our samples again 7-10 folds and send the samples again off for sequencing.

Yet another limitation commenced throughout creating PCR products for two genetic markers, *gyrB* and *purA*. Running the products of these fragments on agarose gel showed more than one band (Figure. 6.3). These findings should

doubtless be much scrutinized, but there are some immediately dependable conclusions for this phenomenon. The primers designed to generate *gyrB* and *purA* inadvertently amplify a 400 bp and 200 bp extra band, respectively. We sequenced the *gyrB* contaminant and found it's a specific region in a range of avian strains (Ma *et al.*, 2013). It has been found in our analysis that there are DNA sequence similarities between different *E. coli* strains (Figure. 6.4). To investigate this issue further, new primers should have been designed to tackle this problem. Nonetheless, time constraint prevented us from exploring this issue to a greater extent. It has to be noted that this is the first time we declare this issue to the scientific field as we did not, to our knowledge, find something similar in the literature. This was something that caused a hassle to us when we were analysing our data. The reason behind that was the need for the exact allelic profile in order to obtain the sequence type for a given isolate through the combination of multiple loci. These limitations, however, did not affect the concatenated data we gained as we believe we processed enough isolates for our experiments, and through knowing the exact and possible STs for the isolates in this project.

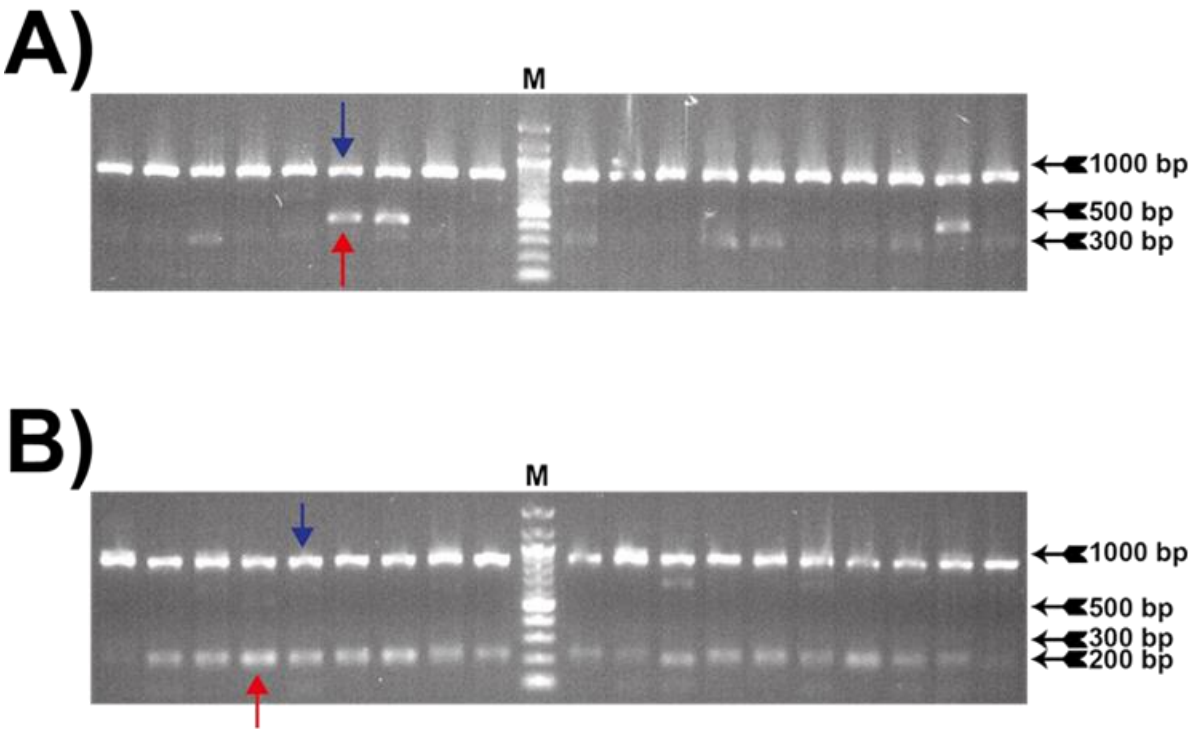


Figure. 6.3. Gel images showing the multiple bands of *gyrB* and *purA* in MLST.

A) A 1 % agarose gel image show more than one *gyrB* band. The correct band size of *gyrB* gene is indicated by the blue arrow. Another 400 bp band (indicated by the red arrow) was amplified at the same time. The 400 bp band is a specific region in a range of avian strains, and is amplified due to primer sequence similarity of *gyrB* and this region. **B)** Anonymous 200 bp band in a 1 % agarose gel through the amplification of *purA* housekeeping gene in MLST procedure (red arrow), whereas the bands indicated by the blue arrow are in the right size. M: DNA ladder.

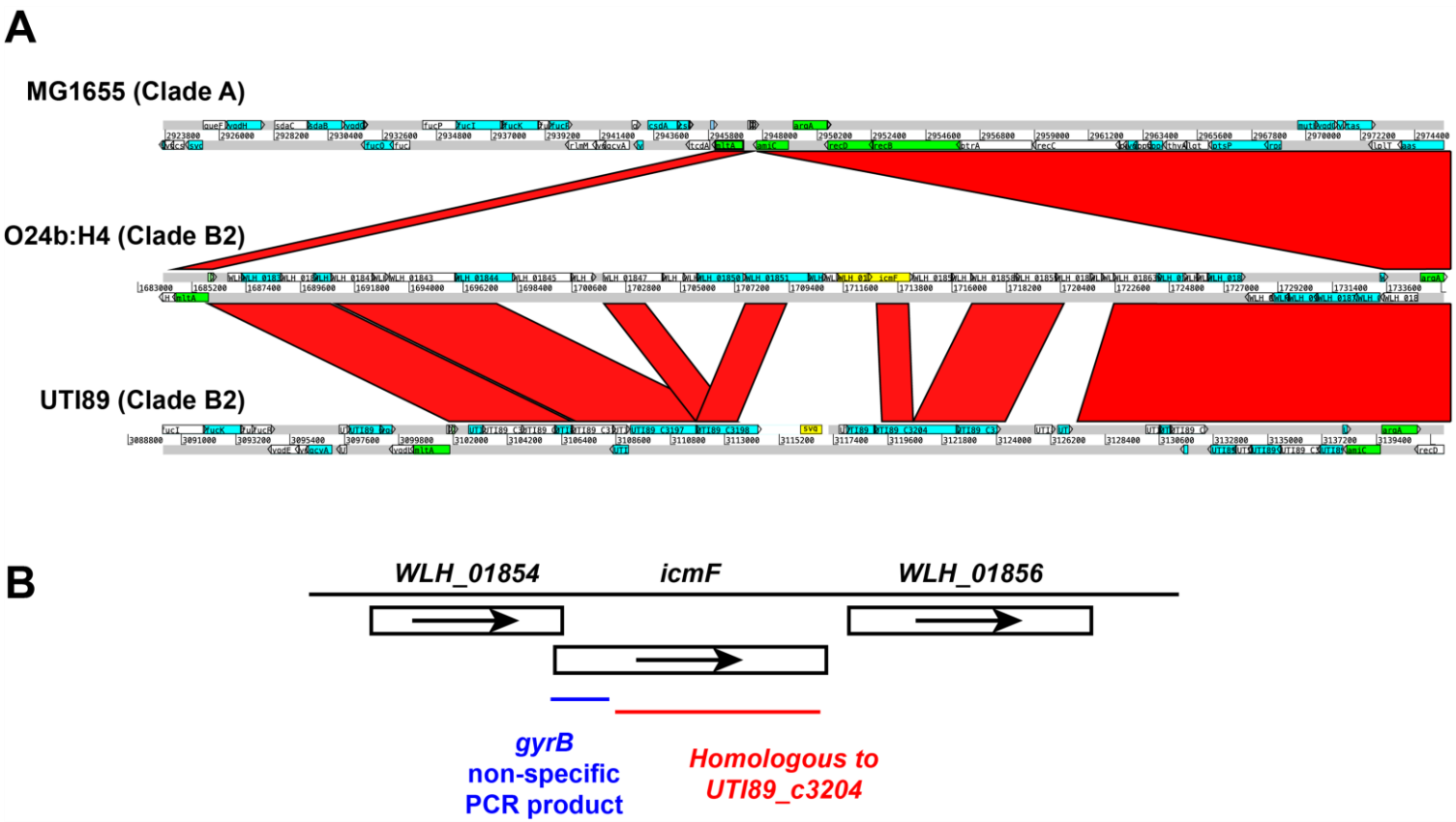


Figure 6.4. Schematic representation of the DNA sequence similarities in MG1655, O24b:H4 and UTI89 genomes in the vicinity of the nonspecific band produced during *gyrB* MLST PCR reactions.

A) Artemis comparison tool output when the region around the non-specific PCR fragment are aligned using the Artemis genome package. Each red box indicates a region of homology between each genome identified via a whole genome BLAST pairwise alignment. MG1655 is missing the region between *mltA* and *amiC* that leads to the non-specific amplification product. In contrast, both O24b:H4 and UTI89 have inserted DNA in this region, but only O24b:H4 has the potential to generate the amplification product. **B)** Schematic diagram of the three genes adjacent to the *gyrB* nonspecific product if O24b:H4 is used as a DNA template. Only if *icmF* is intact, the genetic ability for this non-specific fragment is generated.

6.2.3. Sequence type (ST) frequencies

As we have shown in figure 6.1, the bovine isolates are distributed mostly on phylogenetic groups B1 and A. Of the 94 isolates used in MLST, 53 isolates were included in 31 STs allocated to the B1 clade. ST58 was prevalent among other STs with 7 isolates, followed by ST1510 and ST603 for 4 isolates for each (Figure. 6.5). The frequency was lower for STs 3381 and 1252 with 3 isolates for each. The STs 1146, 1079, 101, 2520, 847 and 962 had 2 isolates for each. While the STs 1125, 155, 937, 641, 196, 20, 297, 1665, 7289, 5979, 3, 1704, 154, 13, 1086, 1172, 172, 4434, 160 and 4173 had 1 isolate for each. On the other hand, clade A had 33 isolates distributed across ten STs with ST10 on the top as 17 isolates associated with this ST. ST195 had 4 isolates followed by ST1415 with 3 isolates, while ST1122 and ST5204 had 2 for each. The STs 181, 218, 4, 216 and 548 had 1 isolate for each. Only six isolates were found belonging to clade E distributed across 3 STs. The ST7535 had the highest frequency with 4 isolates while ST118 and ST4225 had one isolate each.

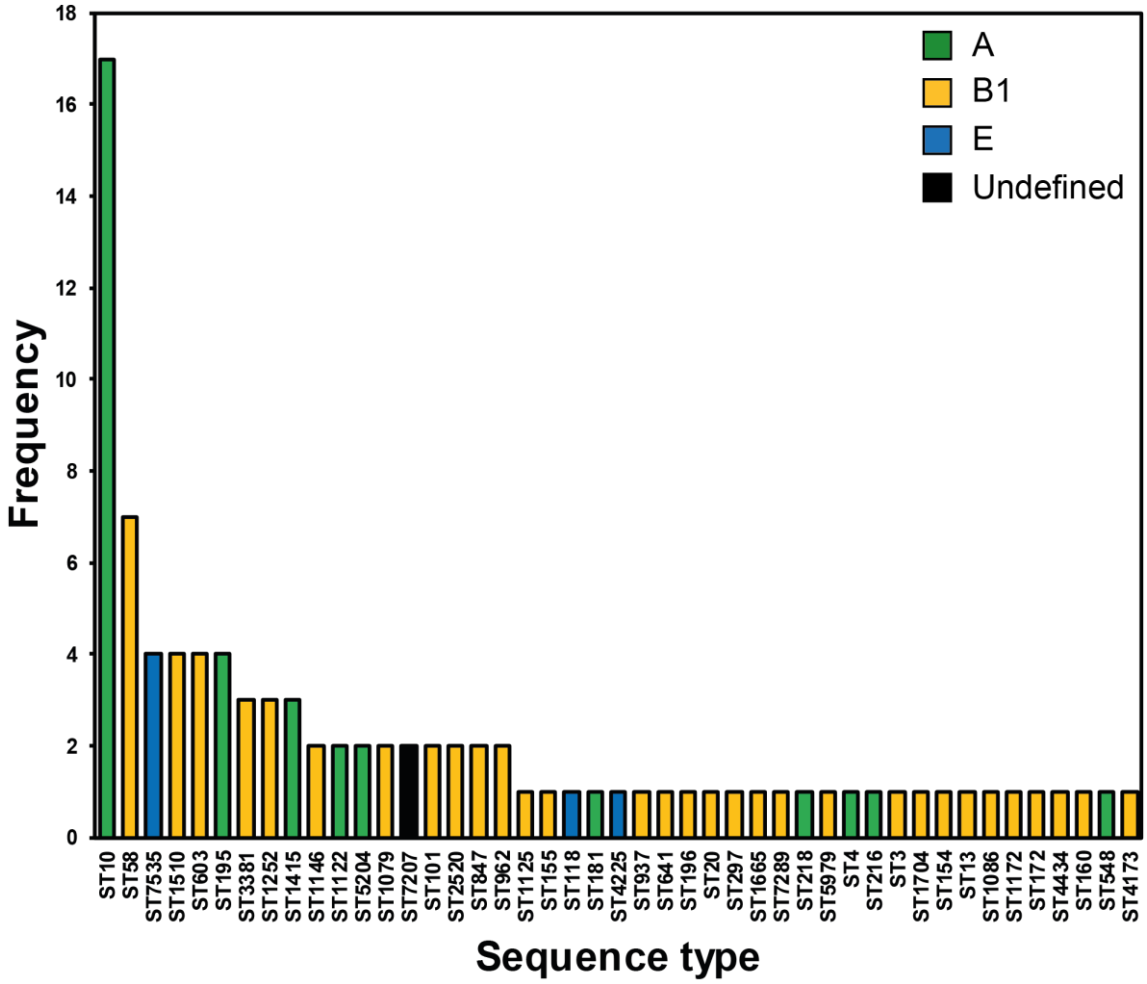


Figure. 6.5. Number of each sequence type of bovine isolates.

Figure shows the frequency (number) of each sequence type of the bovine isolates. Sequence types belong to clade A are in green color, sequence types of clade B1 are in yellow color, sequence types of clade E are in blue and there was one undefined sequence type in black.

6.2.4. Motility phenotype of bovine isolates vs UTI isolates

The flagellum is the motility organelle in motile bacteria, it helps the bacterium to avoid the undesirable conditions and reach favourable and nutrition enriched environments (Soutourina and Bertin, 2003). the flagellum also aids in adhesion, biofilm formation and as a virulence factor (Soutourina and Bertin, 2003) . Therefore, we wished to assess the bovine isolates for motility with that of UTI isolates as they descended from two different *E. coli* favourite niches. Motility assay was performed on 186 bovine isolates using the soft motility agar. Our data showed that the majority of bovine *E. coli* strains (~ 95 %) were motile, whereas 9 isolates representing the remaining 5 % were nonmotile, after 7-8 hours incubation at 37 ˚C on motility agar (Figure. 6.6A). This could suggest that there is no selective pressure on these isolates inside the cows' body preventing them from being motile. From the previous study on 184 UTI associated *E. coli* strains, we know that 67.4 % of these isolates were nonmotile (Figure. 6.6B) (Drage, 2016). It has been suggested that the discrepancy in motility between bovine and UTI isolates is to allow UTI isolates to avoid the human immune system via the TLR5 pathway (unpublished data). We can prove from this data that the majority of *E. coli* strains belonging to clades B1 and A are motile, whereas strains isolated from clades B2 and D tend to be nonmotile.

We then wished to know the extent of motility phenotype of our bovine isolates taking into consideration the swarming diameter threshold of a nonmotile isolate being 0.3 cm. The data revealed most of our bovine strains were highly motile with swarming diameter ranging between 2.5-5 cm (Figure. 6.7).

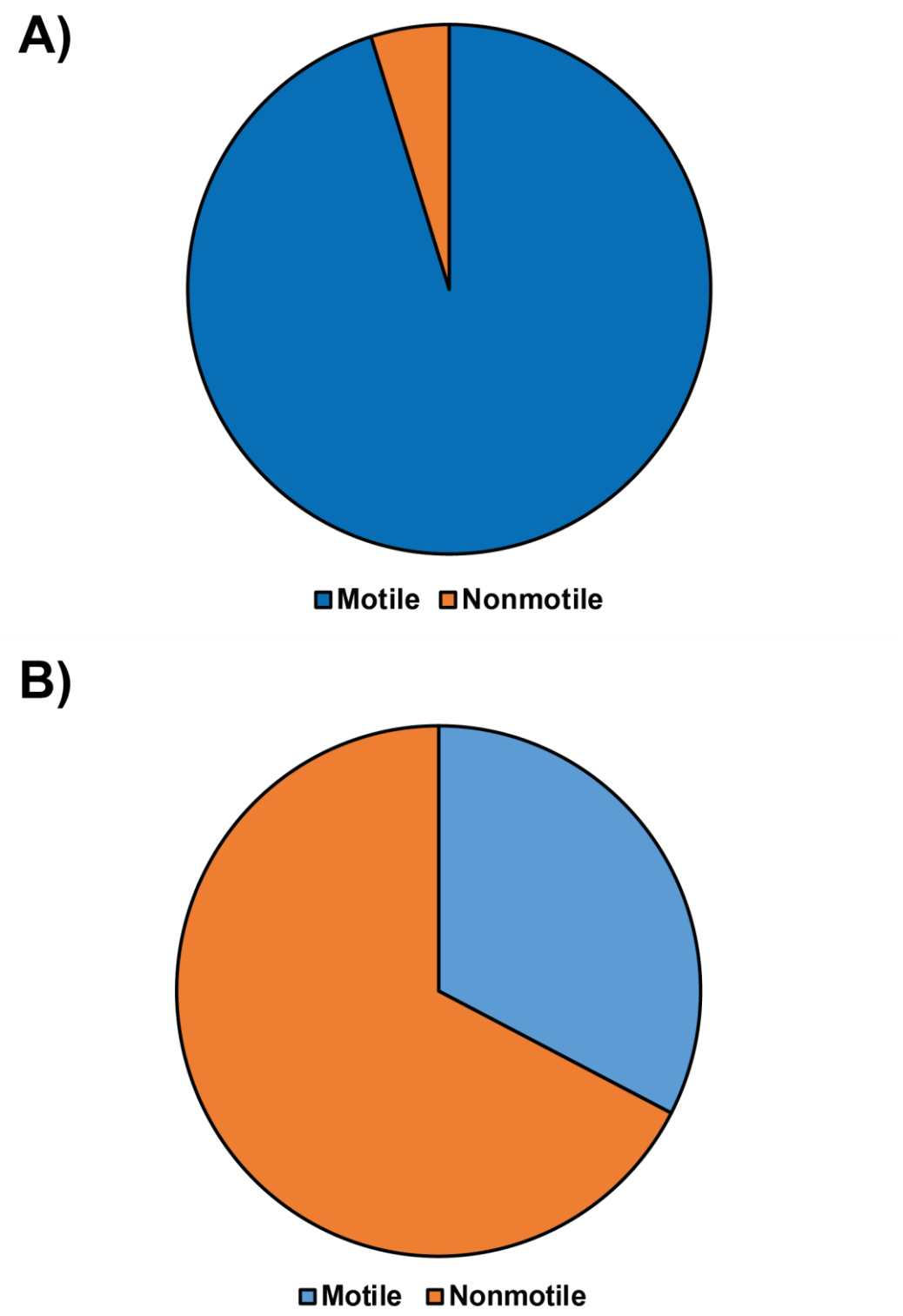


Figure. 6.6. The percentage of motile and nonmotile bovine and UTI isolates.

The pie charts represent the percentage of motile and nonmotile bovine and UTI isolates tested in this study. **A)** The percentage of motile and nonmotile bovine isolates. Of the 186 bovine isolates, ~ 95 % were motile. **B)** The percentage of motile and nonmotile UTI isolates. Of 184 UTI isolates, ~ 67 % were nonmotile (Drage, 2016).

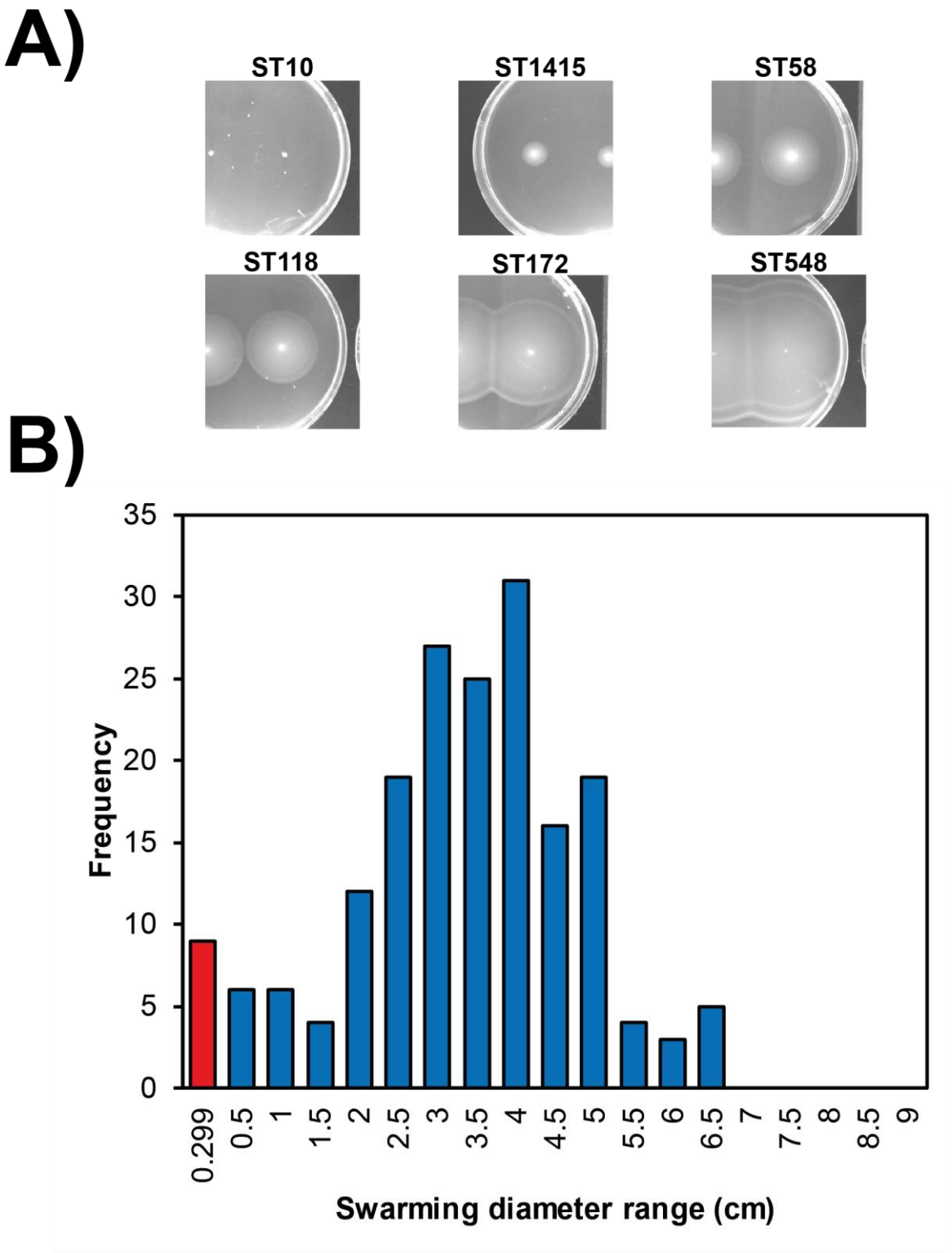


Figure. 6.7. Frequency of motility extent of the bovine isolates.

A) Different motility phenotypes on motility plates after 8 hours at 37 ˚C. **B)** Frequency of motility extent. Data show most of the isolates have swarming diameter between 2.5-5 cm. The red bar represents the number of nonmotile isolates.

The most common ST isolated throughout the study performed on patients with UTI was ST73. The isolates of ST73 has been noticed to be both motile and nonmotile. In our analysis to bovine isolates, we noticed something similar to that occurred with UTI isolates. ST10 which includes strain MG1655 was prevalent with 17 isolates from the total of 186. This ST belonging to clade A contained the majority of nonmotile isolates, with ST1510 of clade B1 have been noticed to be motile and nonmotile while all other STs tended to be either motile or nonmotile (Figure. 6.8). This study is therefore consistent with the complicated motility phenotype seen in the model *E. coli* strain.

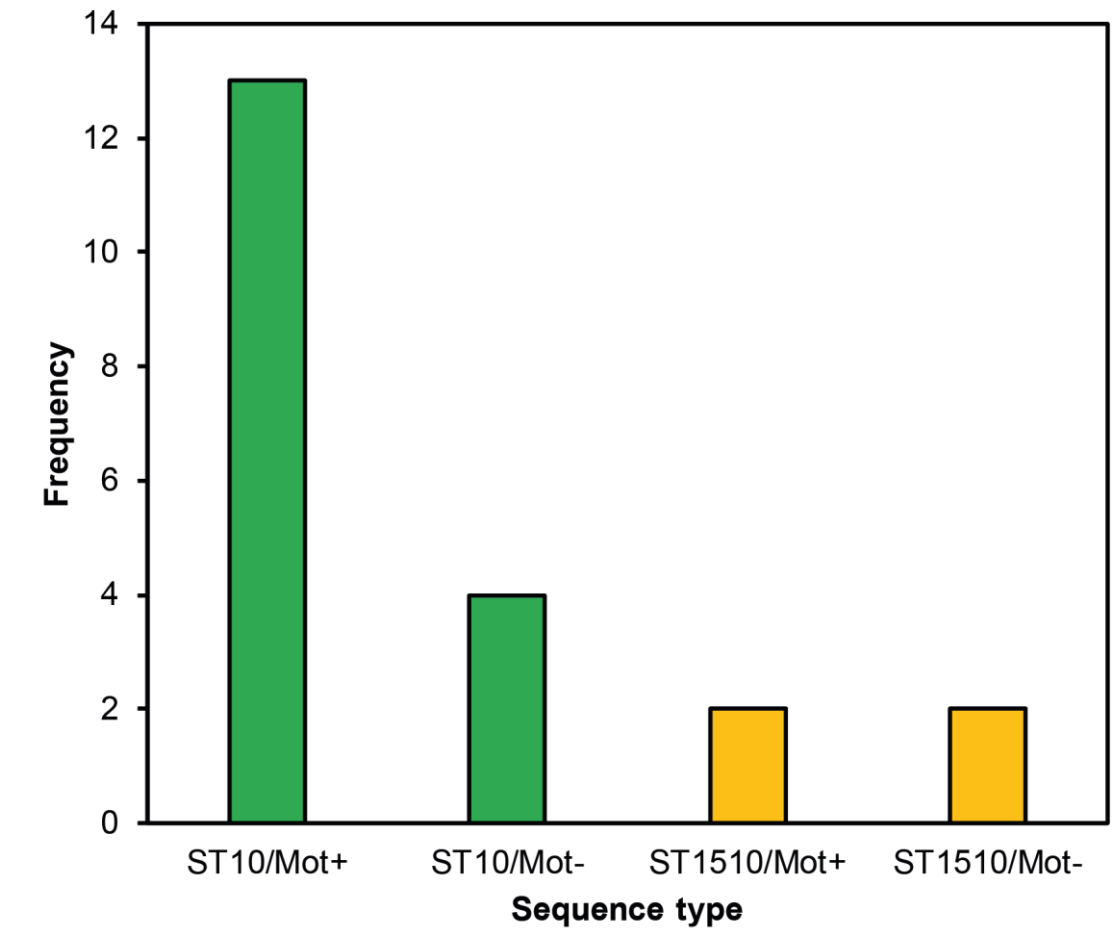


Figure. 6.8. Frequency of bovine isolates sequence types that are motile and nonmotile.

Two sequence types (ST10 and ST1510) exhibited two motility states (motile and nonmotile). Total isolates of ST10 which belong to clade A were 17 with 13 motile isolates and 4 nonmotile isolates. ST1510 which belong to clade B1 isolated 4 times, two of them were motile.

6.2.5. Screening isolates for IS elements

As we pointed out in the introduction to this chapter, bovine and UTI *E. coli* isolates have been studied for IS elements. The literature studied IS elements integrated into the regulatory region of the *flhDC* operon of *E. coli* extensively. A study has dealt with K-12 strains has found a link between the existence of these jumping genes and motility (Barker *et al.*, 2004). Barker *et al.* (2004) screened for the IS elements by running the *flhDC* operon on the gel and checking the size of the bands introduced (Figure. 1.14). It has been known that the topology of the region upstream of the *flhDC* promoter is highly vulnerable to IS elements, and dramatically impacts motility (Fahrner and Berg, 2015). In our lab, we screened 51 clinical isolates for IS elements and we found only one had an IS30 insertion (data not shown). Therefore, we sequenced the *yecG-flhDC* intergenic region and aligned the DNA, we found what we believe to be potential IS scars (Figure. 5.1).

Another study showed three types of IS elements (IS1, IS3 and IS5) can integrate in the *flhDC* promoter region under stress conditions. Stressed cells can suffer SIDD leading to specific pathways permitting IS elements insertion (Zhang *et al.*, 2017). These facts have been confirmed by other researchers who showed that regulators in the *yecG-flhDC* region could make changes under stress conditions that facilitate IS insertion (Humayun *et al.*, 2017). These DNA structural changes which are leading to increased swarming speed up as the cells enter the stationary phase (Humayun *et al.*, 2017). Moreover, one paper revealed the importance of such elements to the bacterial cells in maintaining the resources needed to synthesize flagella as they contribute to the loss of flagellar genes after entering the stationary phase (Plague *et al.*, 2017).

However, these research groups investigated IS elements in K-12 strains which, as we know, belong to only one clade of *E. coli*, group A. As we stated earlier,

our UTI and bovine isolates belong to diverse phylogenetic groups of *E. coli* rather than on a specific clade as studied in the literature. As a result, we asked is the variability in motility phenotype due to IS insertion events? To check this, we amplified the *yecG-flhDC* intergenic region of our UTI and bovine isolates. Our procedure showed no existence of IS elements for all the isolates (Figure. 6.9 and 6.10). These figures show the band size of the *yecG-flhDC* intergenic region of 780 bp while the controls which have IS elements had a band size ~ 2 kb. We conclude from our observations that transposons are not important elements for the motility of *E. coli* strains belonging to all clades. We also suggest that other mutations in the *yecG-flhDC* intergenic regions or elsewhere on the chromosome could increase or decrease motility.

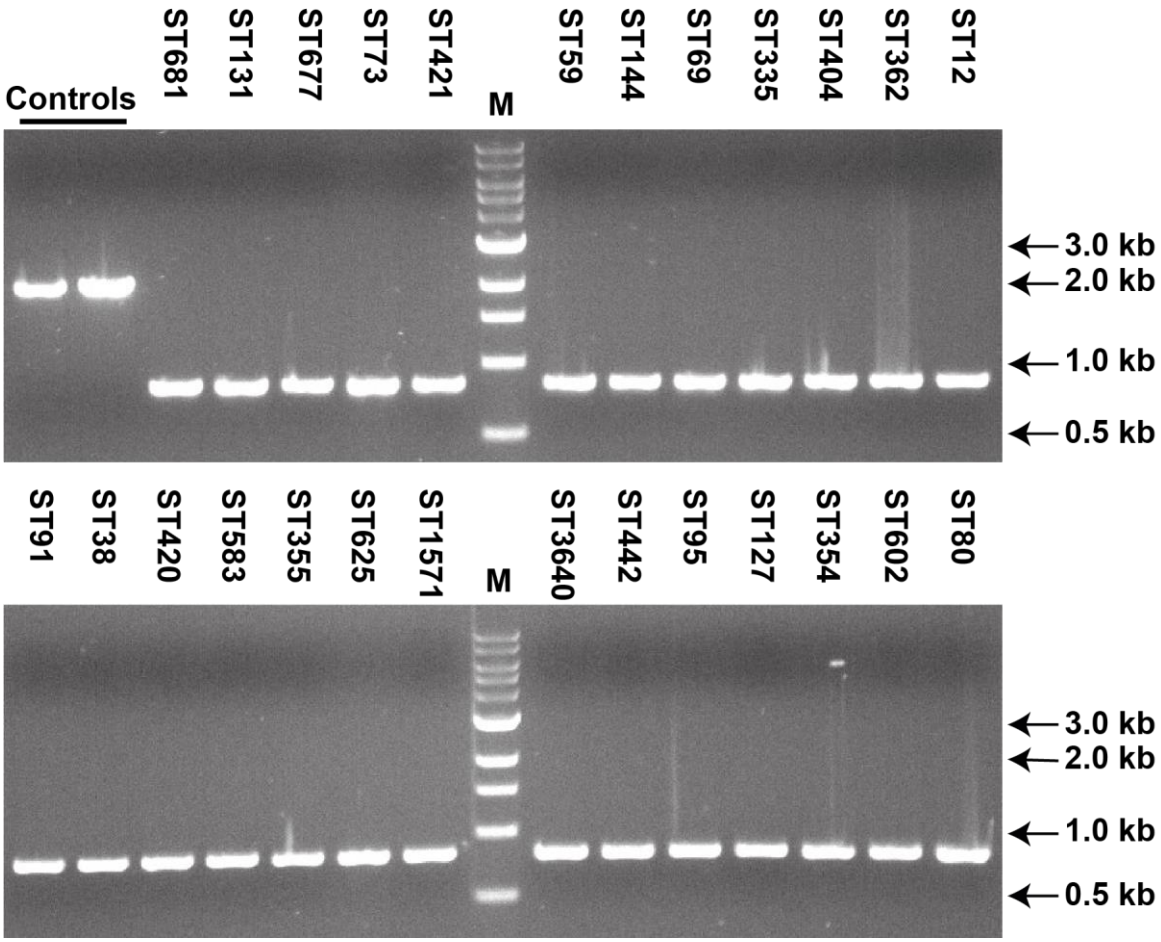


Figure. 6.9. Screening for IS elements in the *yecG-flhDC* intergenic region of UTI isolates.

The *yecG-flhDC* intergenic region of representative UTI isolates was amplified, and PCR products were run on a 1 % agarose gel. Gel images show the 780 bp *yecG-flhDC* intergenic regions without IS element, whereas the same region with IS element in the controls have a size of 2 kb. M: DNA ladder. All the 184 UTI isolates were screened for this experiment.

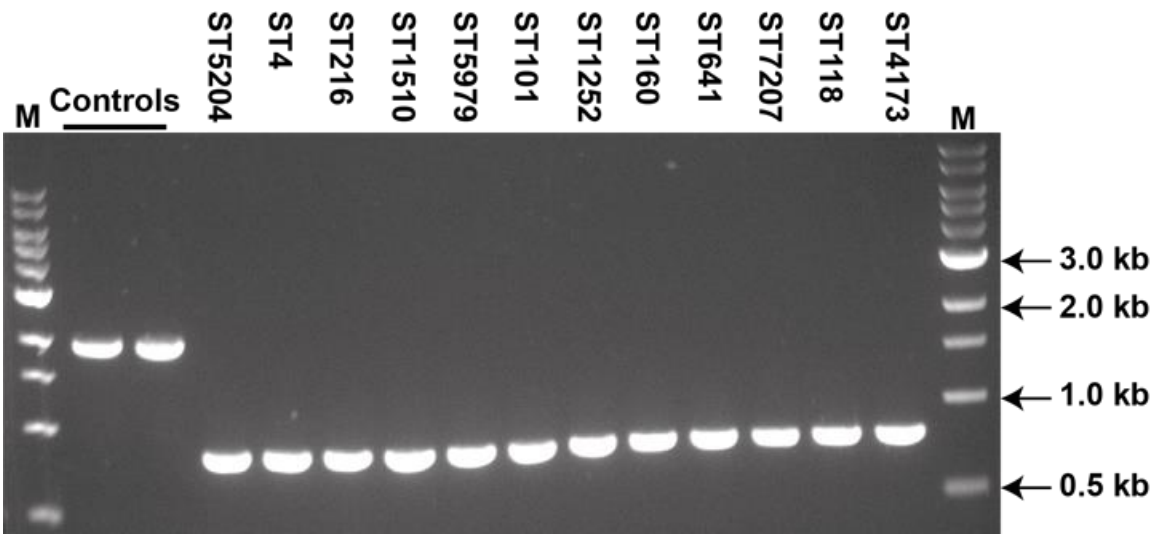


Figure. 6.10. Screening for IS elements in the *yecG-flhDC* intergenic region of bovine isolates.

The *yecG-flhDC* intergenic regions of representative bovine isolates were amplified, and PCR products were run on a 1 % agarose gel. The gel image shows the 780 bp *yecG-flhDC* intergenic regions without an IS element, whereas the same region with an IS element in the controls has a size of 2 kb. M: DNA ladder. All the 186 bovine isolates were screened for this purpose.

6.2.6. Biofilm formation of bovine and UTI isolates

Bacterial biofilms are complex three-dimensional multicellular communities which can take different shapes. They are synthesized as a result of a very complicated regulatory cascade including the induction of the biofilm master regulator σ^S (RpoS). This induction results in the activation of the *csg* loci and subsequent curli fimbriae fibres secretion and *bcs* which leads to cellulose production (Serra *et al.*, 2013b). Cells produce biofilms under stress conditions when survival become the first focus. The biofilm adheres cells to surfaces and protects them from antibiotics (Costerton *et al.*, 1999), desiccation, UV radiation, plant immune system (Yaron and Romling, 2014), stress conditions inside drinking water pipes (Abberton *et al.*, 2016), killing by soil nematodes (DePas *et al.*, 2014). We wanted to test the effect of the biological niches and the location of the isolates in the phylogenetic tree on biofilm formation. In order to study biofilm formation of our *E. coli* isolates, we grew the bacteria under stressful conditions on LB agar with no NaCl and 28 ˚C as this temperature is typical for many *E. coli* strains to express CsgD (Bokranz *et al.*, 2005). Although phenotypic variation of biofilm morphology has been indicated, we could not find defined classes in the literature. However, we classified biofilm heterogeneity into six grades according to their level of complexity with grade 6 representing the most complex one (Figure. 6.11A).

Our data showed variation in biofilm formation by bovine isolates (Figure. 6.11B). Of the 186 bovine isolates, the frequency of grade 1 was 38 followed by 53 for grade 2. Grade 3 and grade 4 were very similar and were diagnosed 13 and 12 times, respectively. Moreover, Grade 5 phenotype scored 48 times while the frequency of grade 6 was 22. Similar results have been demonstrated through studying 184 *E. coli* isolates from UTI patients (Bright and Aldridge, unpublished). The data showed all six biofilm phenotypes distributed all over the place with

tendency towards grade 2 which suggest there is no selective pressure to form complex biofilm phenotypes in UTIs (Bright and Aldridge, unpublished). In the bovine isolates with defined clades, we found a tendency for isolates belong to clades B1 and E to form grade 5 biofilms, whereas clade A isolates commonly form grade 2 biofilms (Figure. 6.12). In contrast to clade B2 UTI isolates which exhibited grade 2 biofilms, we suggest selective pressure of B1 and E bovine isolates to form elaborate biofilms. One thing to consider in this regard is the cow immune system which may try to eradicate the infection with these strains. Clade A bovine isolates consists domesticated K-12 *E. coli* strains. The findings observed here are consistent with the literature in that clade A isolates are nonmotile and weak biofilm formers (Serra *et al.*, 2013a). On the other hand, clade B1 isolates have a good motility and they are good biofilm formers.

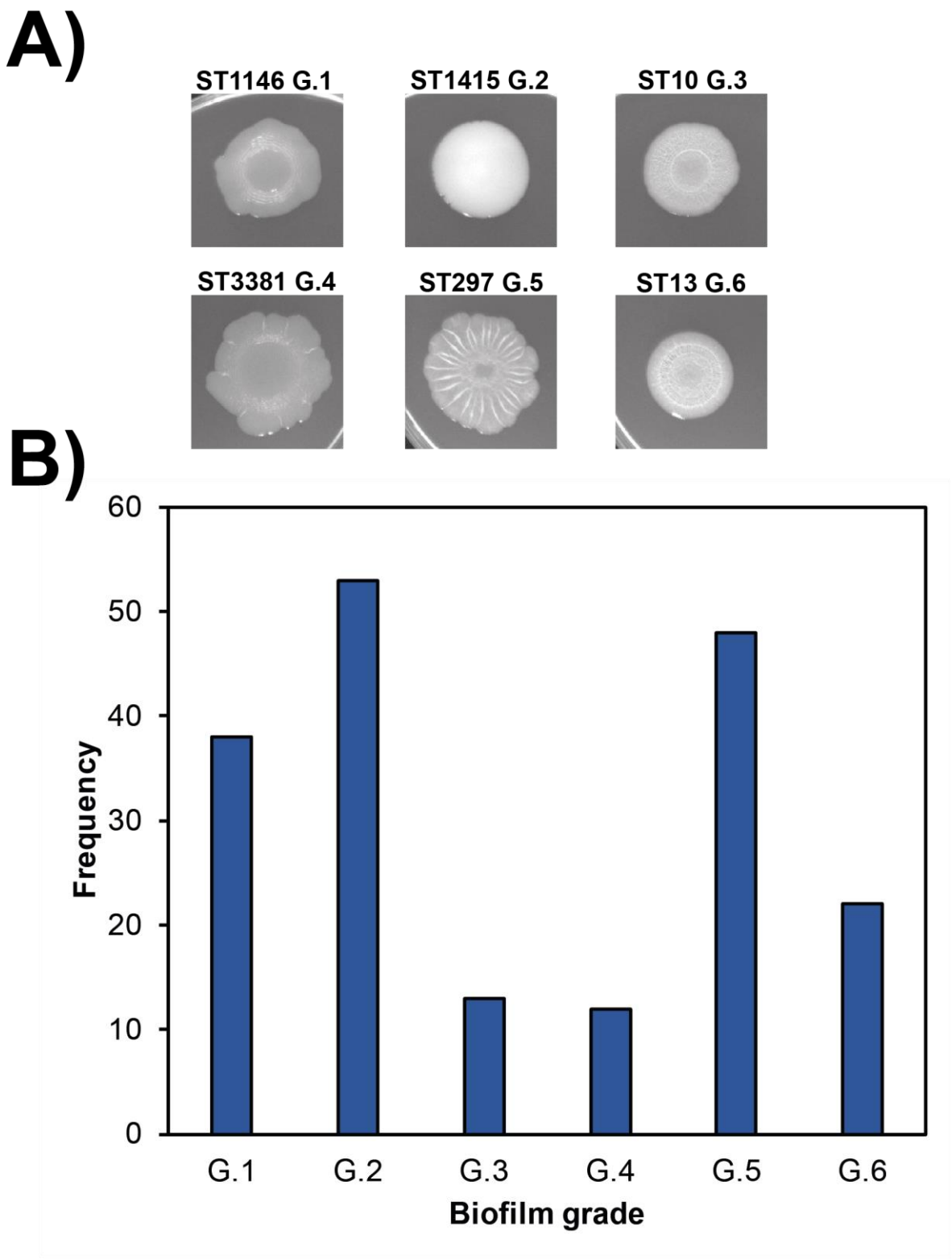


Figure. 6.11. Heterogeneity and frequency of biofilms of bovine isolates.

A) Grades of biofilms. Cells grown in LB agar with no NaCl for 3-5 days at 28 ˚C. G.1: flat, G.2: flat and glossy, G.3: rippled but flat, G.4: rigged but flat, G.5: very rigged, G.6: complex. **B)** Number of each biofilm grade of the bovine isolates.

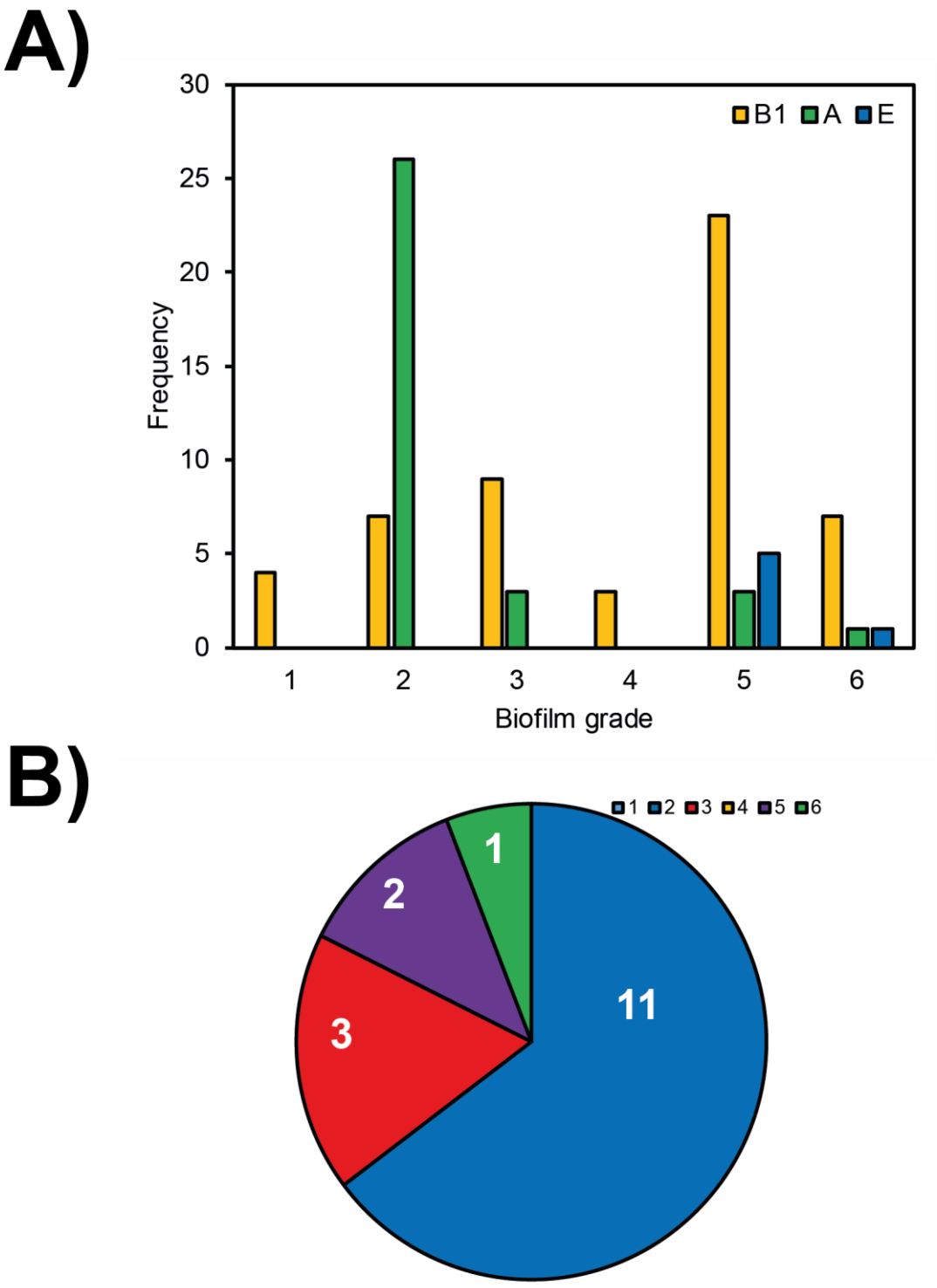


Figure. 6.12. Frequency of biofilm grades of different clades of bovine isolates.

A) Number of biofilm grades of each clade of bovine isolates. Clades B1 and E isolates tend to form complex biofilms, whereas isolates of clade A commonly form simple biofilms. **B)** Frequency of biofilm grades of ST10. This ST which belongs to clade A tends to form grade 2 biofilms. The pie chart shows 11 isolates of ST10 form grade 2 biofilms, whereas 6 other isolates distributed between 3, 5 and 6 grades.

6.3. Summary

The strength of the worldwide research carried on *E. coli* is mitigated in some way by dealing with K-12 strains only. We have studied the motility in *E. coli* as a species and have shown in an explicit empirical procedure that a 780 bp of DNA dictates the motility in *E. coli*. We confirmed that by replacing the *yecG-flhDC* intergenic region of strain MG1655 with the same region of three various *E. coli* strains belong to different phylogenetic groups. Additionally, we switched this region of UTI89 with the one of strain MG1655. We also demonstrated the impact of *yecG* as a negative regulator on *flhDC* expression.

In this chapter, we followed these experiments by analysing *E. coli* strains isolated from UTI patients and cows. By doing so, we wanted to study the behaviour of these isolates in two different preferential niches colonised by *E. coli*. We started our analysis by performing MLST on our bovine isolates as a reliable way to track the origin of these isolates and compared them to UTI isolates. We have chosen MLST of Warwick Medical School as it proved to be the closest to the whole-genome analysis method. Together these results provide important insights into the distribution of *E. coli* isolates from the human bladder and bovine guts on the phylogenetic tree. When we aligned our DNA sequences with those strains from McNally *et al.* (2013), we have succeeded to assign our bovine isolates to mainly clade B1 and A. At the same time, UTI isolates were mainly belonging to clade B2 and D. We followed this by investigating the motility of our bovine isolates compared to UTI strains. Our data showed that ~95 % of the isolates were motile compared to ~33 % motile UTI isolates. This suggests no selective pressure on bovine isolates to avoid motility, unlike UTI isolates which hide from the human immune system and cause recurrent infections. We then wished to screen our 186 bovine isolates and 184 UTI isolates for transposable elements known in the literature to be common in

the *yecG-flhDC* intergenic region. The results revealed no existence for IS elements in bovine and UTI isolates. The evidence presented in this chapter indicates that IS elements are not necessary for the motility of *E. coli*, and what is stated in the literature is related to only one strain, MG1655.

The tendency of bovine and UTI isolates to form biofilms was also studied. Although it is not defined in the literature, we classified the phenotypic morphology of biofilms into six grades according to their complexity. We have demonstrated that UTI isolates form heterogenous biofilms, but commonly of grade 2. This indicates no selective pressure on UTI isolates to form elaborate biofilms. However, our bovine isolates which belong to clade B1 and E tend to form grade 5 biofilms while clade A isolates commonly formed grade 2 biofilms. The inability of clade A isolates to form complex biofilms due to SNPs in their *bcs* operon which prevent them from producing cellulose is worth looking in our strain collection (Serra *et al.*, 2013a).

Chapter Seven: General Discussion

7.1. Introduction

Motility of bacteria is an important feature that helps the cells to avoid detrimental conditions and move toward favourable environments (Soutourina and Bertin, 2003). The process of flagellar assembly is hierarchical and depends on three promoter classes (class 1, class 2 and class 3). The *flhDC* operon which is expressed as a result of the activity of a class 1 promoter is considered as the master regulator of the subsequent steps of flagellar assembly (Chilcott and Hughes, 2000; Aldridge and Hughes, 2002). We analysed in this project the activity of the *flhDC* promoter of *E. coli* in context of the species. We also dissected the *yecG-flhDC* intergenic region to understand the part of this region that dictates motility. In order to achieve these goals, we exploited the cutting-edge tools in DNA recombineering in such experiments. We then studied a collection of *E. coli* strains from two different biological niches to understand the impact of the variation of environmental hosts on motility. The main information discovered in this study was that motility phenotype is dependent on the strength of the *flhDC* promoter, and that different *flhDC* promoter types reflect various motility magnitudes. We also detected a negative regulatory effect of YecG on the flagellar system. Another main finding of this project was that the variation in the biological niches of *E. coli* strains impacts the motility phenotype.

7.2. Analysis of the *flhDC* promoter across the species of *E. coli*

The literature revealed that *E. coli* as a species can be divided into six major phylogenetic groups: B2, B1, A, D, E and F (McNally *et al.*, 2013). We wanted to know whether this classification impacts the *flhDC* promoter in *E. coli* as a species. Thus, we aligned the *yecG-flhDC* intergenic region of a collection of *E. coli* strains, and found that the *flhDC* promoter reflects the phylogenetic tree (Figure. 4.1). We identified three types of *flhDC* promoters in *E. coli*: B2, A/B1 and E/D/F which is similar to the clade boundaries defined by Wirth *et al.* (2006). Wirth *et al.* (2006)

argued a classification similar to A, B1, B2 and D which we approximately see for the *flhDC* promoter region.

Then, we studied the effect of replacing the *flhDC* promoter on motility phenotype in *E. coli*. Using the Blank *et al.* (2011) method, we replaced the *yecG-flhDC* intergenic region of strain MG1655 with the same region of three *E. coli* strains belonging to three different phylogenetic groups according to our analysis (Figure. 4.1B). These three *E. coli* strains were characterised by their high motility phenotypes compared to our strain MG1655 which is weakly motile due to the lack of IS element (Fahrner and Berg, 2015). When we performed the motility assay for the newly created strains, we found that our strain MG1655 with the new promoters were highly motile. We also replaced the *yecG-flhDC* intergenic region of strain MG1655 with that of strain UTI89, and replaced the *yecG-flhDC* intergenic region of strain UTI89 with that of strain MG1655 as a control experiment. The data obtained from the control experiment was consistent with our first experiment. The motility phenotype on motility plates was our primary assessment of the tested *flhDC* promoters. Therefore, we wanted to study the *flhDC* promoters further by measuring their activity through the production of light. For this reason, we cloned the four *flhDC* promoters we used in our first experiment on the plasmid pSB401 which has the *lux* genes that produce bioluminescence. We then transformed the plasmids into panel strains known to have different motility phenotypes. Our data showed that the different *flhDC* promoters behaved in the same manner that we noticed on the motility plates. Furthermore, the data revealed that these *flhDC* promoters reflected the motility phenotypes of the panel strains.

The effect of nutrients and temperature was also studied to detect the possible change of activity toward the changes. However, our results indicated that these *flhDC* promoters maintained similar behavioural activity regardless to the tested

conditions. The impact of five global regulators was examined on the different *flhDC* promoters representing the species of *E. coli*. The quantification of motility of the deletion mutants showed that these *flhDC* promoters were consistent in their behaviour, with *ΔclpP* mutants scored the highest motility extent for all *flhDC* promoters. The data was consistent with the literature in that ClpP degrades $FlhD_4C_2$ preventing flagellar assembly (Takaya *et al.*, 2012).

Then, we wanted to test the relation between the activity of our *flhDC* promoters on the abundance of flagella. We used for this reason yellow fluorescent protein fusion conjugated to *fliM* (Delalez *et al.*, 2010). The data showed that the main population of strain MG1655 was lacking any flagella, whereas 10-15 % of the cells have 1 or 2 flagella. On the other hand, the *flhDC* promoters from clade B2 and D produced the average of 1 spot of *fliM-yPet* which led to the high motility. The main message we can take from this group of experiments is that the type of the *flhDC* promoter is responsible for the observed motility phenotypes in the *E. coli* strains representing the species. Data argues that promoter type will lead to different *flhDC* transcription and, thus, variable regulation across *E. coli* strains. Moreover, the lack of IS elements in the *yecG-flhDC* intergenic region of strain MG1655 leads to very low *flhDC* transcription.

7.3. The dissection of the *yecG-flhDC* intergenic region in *E. coli*

Further analysis to the preceding experiments was performed to investigate each part in the *yecG-flhDC* intergenic region on motility. Based on a recent paper of Fahrner and Berg (2015) who studied the regulatory aspects in this region, we divided the *yecG-flhDC* intergenic region into four zones: 1) the untranslated region downstream of the *flhDC* promoter which is a binding site of carbohydrate metabolism regulator CsrA , 2) a conserved *flhDC* promoter region across *E. coli*, 3) the regulatory region which acts as a binding site for many regulators such as OmpR

and H-NS and 4) the IS SCAR region which represents the hot spot for IS elements binding. We created a combination of deletion mutants in this region in two *yecG-flhDC* regions representing the clades B2 and A, and quantified the motility for each deletion mutant. Our data revealed different motility phenotypes on the motility plates. However, ΔSCAR mutants exhibited the highest motility extent among the other mutants. We concluded from these experiments that something lies within the SCAR region is repressing motility. Therefore, we took the data further and dissected the SCAR region of the *yecG-flhDC* intergenic region of B2 and A clades. For this purpose, we divided the SCAR region into six fragments, and created six deletion mutants. When we quantified the motility, we observed that the addition of the distant fragment suppressed the motility. We know that the distant fragment of the SCAR region is very close to the *yecG* promoter according to our dissection to the *yecG-flhDC* intergenic region. Thus, our interest directed towards the analysis of *yecG* impact on motility. As a result, we deleted *yecG* from B2 and A strains to see the impact of this deletion on motility. Data showed that ***ΔyecG*** mutants revealed increased motility compared to their wildtypes. Our findings regarding ***ΔyecG*** mutants are not consistent with the literature which mentioned the various functions of *yecG* (Nachin *et al.*, 2005; Heermann *et al.*, 2009). Thus, we declare a novel role of *yecG* in regulating the motility in *E. coli*.

We then replaced the SCAR region of strain MG1655 with that of strain CFT073 which belongs to clade B2 to detect the effect of replacement on motility. We argued that as strain CFT073 was motile, our strain MG1655 will tend to have the same motility as strain CFT073. However, the data revealed the opposite in that the new strain MG1655 with the SCAR region from strain CFT073 converted to a completely nonmotile strain. This nonmotile phenotype led as to generate a new hypothesis: CFT073 SCAR region has multiple factors that allow greater flagellar

gene suppression. We know that the SCAR region is very close to the *yecG* promoter, therefore, we asked: Is the observed phenotype due to increased *yecG* transcription? To confirm this phenomenon, we cloned the *yecG* promoter of strains CFT073 and MG1655 on the plasmid pSB401 to measure its activity. Our data revealed that indeed the *yecG* promoter in strain CFT073 is more active than that in strain MG1655 (3-4 fold) which is consistent with what we had on the motility plates.

Flagellar synthesis is a hierarchical process that includes the activity of three promoter classes (class 1, class 2 and class 3). In order to study the other two promoter systems, we cloned the *flgB* promoter and the *fliC* promoter into the plasmid pSB401 to measure the activity. We then transformed these two plasmids to our deletion mutants' collection which have different motility phenotypes. As with the *flhDC* promoter, we noticed that the *flgB* and *fliC* promoters showed consistent activity with the motility phenotypes of the strains. We concluded from this set of experiments that flagellar gene expression is a sequential process that depends on three types of promoters activated in the same manner according to the expression of the master regulator. The key point that we concluded from performing these experiments is that YecG is a negative regulator of the flagellar system in *E. coli*. We recommend that future work may include the study of YecG mechanism of action on the flagellar system. This can be done by testing the interaction capability of YecG with the master regulator in both transcriptional and translational levels. However, the architecture of the *yecG-flhDC* intergenic region argues that YecG interacts with the *flhD* and *flhC* genes and regulates their transcription.

7.4. Impact of host niche environments on the motility phenotype of *E. coli*

The classification of phylogroups of *E. coli* strains according to McNally *et al.* (2013) is based on the genomic structure of the strains. Strains belonging to a specific phylogroup usually share similar characteristics, especially with respect to their

source of isolation. In this project, we wanted to test the impact of host niche condition on the motility phenotype observed for the tested strains. For this purpose, we used a collection of bovine *E. coli* strains isolated from several herds across the south of England. We compared this collection with another collection of *E. coli* strains isolated from human UTI cases (Drage, 2016). We performed the MLST analysis based on the scheme of Warwick Medical School by analysing the DNA sequence of seven housekeeping genes. Our data showed that the bovine-derived *E. coli* strains can be found mostly in phylogroup B1 and less frequently in A and E, but they are absent from B2 and D/F. On the other hand, human-derived UTI strains can be found across all groups, with a significant cluster around phylogroups B2 and D/F (Drage, 2016).

We compared the motility phenotype of the bovine and the UTI *E. coli* strains. The data revealed that the bovine strains exhibited robust motility compared to the motility of UTI isolates (Figure. 6.6). The literature suggests that the *yecG-flhDC* intergenic region is a hot spot for IS element insertion (Barker *et al.*, 2004; Fahrner and Berg, 2015). We, therefore, argued that the diverse phenotypes seen amongst this strain collection was due to IS attack. Thus, we screened up to 450 *E. coli* strains for IS elements across the period of this project. However, we found only a single UTI isolate has an IS element in the *yecG-flhDC* intergenic region. The frequency of 0.2 % is within the natural frequency of IS transposition events (Sousa *et al.*, 2013). Thus, we conclude, in contrast to the literature, that strains can still be motile even with the absence of IS elements in the *E. coli* species.

We also investigated the biofilm formation among the different clades of bovine and UTI isolates. We found a tendency in bovine isolates belonging to clades B1 and E to form intricate biofilms, whereas bovine isolates lie in clade A formed simple biofilms. Data of UTI isolates biofilms revealed tendency to form simple

biofilms which suggest no selective pressure to form complicated biofilms. The data observed regarding to biofilm formation was consistent with the literature in showing clade A isolates as nonmotile and weak biofilm formers. On the other hand, clade B1 isolates exhibited good motility and good biofilm formation.

The *E. coli* research community should appreciate from this work the strong evidence that IS elements impact *yecG* expression, but not *flhDC* expression. The presence of IS elements may negatively regulate *yecG* expression which leads to enhanced *flhDC* expression and, thus, greater motility. Because of the low incidence of IS insertions amongst *E. coli* strains, we argue that the motility phenotype is driven by alternative regulatory mechanisms. Our main conclusion from this set of experiments is that the motility phenotype is different under different conditions to facilitate survival in various host niche environments and is dependent on the strain location within the *E. coli* species.

Chapter Eight: Appendix

8.1. Growth media

Ingredients of growth media were dissolved in MilliQ water and sent for sterilization according to the standard autoclave conditions. Antibiotics were added to the cooled sterilized growth media.

Table. 8.1. Luria-Bertani liquid medium

Luria-Bertani liquid medium (1 L)	
Bacto Tryptone	10 g
Bacto Yeast Extract	5 g
NaCl	5 g

Table. 8.2. Luria-Bertani solid medium

Luria-Bertani soid medium (1 L)	
Bacto Tryptone	10 g
Bacto Yeast Extract	5 g
NaCl	5 g
Bacto Agar	15 g

Table. 8.3. Tryptone Broth liquid medium

Tryptone Broth liquid medium (1 L)	
Bacto Tryptone	10 g
NaCl	5 g

Table. 8.4. Tryptone Broth solid medium

Tryptone Broth solid medium (1 L)	
Bacto Tryptone	10 g
NaCl	5 g
Bacto Agar	15 g

Table. 8.5. Motility medium

Motility medium (1 L)	
Bacto Tryptone	10 g
NaCl	5 g
Bacto Agar	3 g

Table. 8.6. 50x Minimal E salts

50x Minimal E salts (1 L)	
$MgSO_4.7H_2O$	10 g
Citric acid.H_2O	100 g
K_2HPO_4 (anhydrous)	500 g
$NaNH_4HPO_4.4H_2O$	175 g
Water	Dissolve components and adjust the final volume to 1 L.

Table. 8.7. 2X Minimal E salts

2X Minimal E salts (1 L)	
50X	40 ml
Water	960 ml
Distribute into bottles and autoclave	

Table. 8.8. Minimal media 3 g

Minimal media 3 g (100 ml)	
2x Minimal E salts	50 ml
50 % Glucose	0.4 ml
Bacto Yeast Extract 25 g/l	12 ml
Water	37.6 ml

Table. 8.9. Minimal media 1 g

Minimal media 1 g (100 ml)	
2x Minimal E salts	50 ml
50 % Glucose	0.4 ml
Bacto Yeast Extract 25 g/l	4 ml
Water	45.6 ml

Table. 8.10. Minimal media 0.2 g

Minimal media 0.2 g (100 ml)	
2x Minimal E salts	50 ml

50 % Glucose	0.4 ml
Bacto Yeast Extract 25 g/l	0.8 ml
Water	48.8 ml

Table. 8.11. Minimal media 0.04 g

Minimal media 0.04 g (100 ml)	
2x Minimal E salts	50 ml
50 % Glucose	0.4 ml
Bacto Yeast Extract 25 g/l	0.16 ml
Water	49.44 ml

Table. 8.12. Biofilm formation liquid medium

Biofilm formation liquid medium (1 L)	
Bacto Tryptone	10 g
Bacto Yeast Extract	5 g

Table. 8.13. Biofilm formation solid medium

Biofilm formation solid medium (1 L)	
Bacto Tryptone	10 g
Bacto Yeast Extract	5 g
Bacto Agar	15 g

8.2. Solutions

Table. 8.14. Alkaline lysis solution I

Alkaline Lysis Solution I (100 ml)	
Glucose	0.9 g
1 M Tris HCl (pH=8)	2.5 ml
0.5 M EDTA (pH=8)	2 ml
RNase A	500 ul
Add sterile water to a final volume of 100 ml, filter sterilise and store at 4 ˚C	

Table. 8.15. Alkaline lysis solution II

Alkaline Lysis Solution II (100 ml)	
1 M NaOH	20 ml
10 % SDS	10 ml
Sterile Water	70 ml

Table. 8.16. Alkaline lysis solution III

Alkaline Lysis Solution III (100 ml)	
3 M KAc	29.44 g
Dissolve in 50 ml deionised water	
Glacial Acetic Acid	11.5 ml
Add deionised water up to 100 ml	

Table. 8.17. Agarose gel

Agarose Gel 1 %	
Agarose	4 g
Add 1X TAE buffer up to 400 ml, heat until powder melts and store at 65 ˚C until use	

Table. 8.18. 10X DNA loading buffer

10X DNA Loading Buffer	
Tris Acetate	200 mM
EDTA (pH=8)	5 mM
Glycerol	50 %
Bromophenol Blue	0.1 %
Xylene Cyanole FF	0.1 %
Orange G	0.1 %
Sterile Water	Up to 50 ml

8.3. Oligonucleotides

Table. 8.19. Primers used in this study

Primer Number	Primer Name	Sequence
849	CFT073_FliCP1	ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAGTGTAGGCTGGAGCTGCTTC
873	*ompR*P1	atgcaagagaactacaagattctggtggtcgatgacgacatgcgc
874	*ompR*P2	tcatgctttagagccgtccggtacaaagacgtagcccagacccccaCATATGAATATCCTCCTTA
875	*ftsH*P1	atggcgaaaaacctaatactctggctggtcattgccgttgtgctg
876	*ftsH*P2	ttacttgtcgcctaactgctctgacatggtgttacccgggttcggCATATGAATATCCTCCTTA
877	*clpP*1	atgtcatacagcggcgaacgagataactttgcaccccatatggcgGTGTAGGCTGGAGCTGCTTC
878	*clpP*P2	tcaattacgatgggtcagaatcgaatcgaccagaccgtattccacCATATGAATATCCTCCTTA

879	*lon*P1	atgaatcctgagcgttctgaacgcattgaaatccccgtattgccg GTGTAGGCTGGAGCTGCTTC
880	*lon*P2	ctattttgcagtcacaacctgcataccagacggttcattttgcagC ATATGAATATCCTCCTTA
893	*clpP*_CFT073-212chk	ggcccgtcaccgccaggtggtggg
894	*clpP*_CFT073+848chk	ggatggaccggcaatcagcttgcg
895	*lon*_CFT073-199chk	tctggtgaataattaaccattccc
896	*lon*_CFT073+2529chk	gaatccttcaaggtacgaacgcgc
899	*ftsH*_CFT073-276chk	cgtgatgtattagcgccaggtggc
900	*ftsH*_CFT073+2101chk	tcaggcgtgacgttgaggatcccc
901	*ompR*_CFT073-258chk	acccttcccgggtaaccaggggcg
902	*ompR*_CFT073+995chk	atctcccgacggaaagcgggaggc
1034	*fliM*+760P1	CGCGTAATGAAGATCAGAACTGGCGCGATA ACCTGGTGCGCCAGGGTGTAGGCTGGAGCT GCTTC
1035	*fliN*+168P2	CGTACGGCCCAGCTCGACGGTCAGCTTGAC CGGAATATCCATAATCATATGAATATCCTCC TTA
1036	*fliN*+268R	GGTTCGCCCGCCAGACCGTCCAGC
1037	*fliM*+498F	GACGCCTGGAAGGCGATTAATCCG
1038	*crp*+1P1	atggtgcttggcaaaccgcaaacagacccgactctcgaatggtt cGTGTAGGCTGGAGCTGCTTC
1039	*crp*+632P2	ttaacgagtgccgtaaacgacgatggttttaccgtgtgcggagat CATATGAATATCCTCCTTA
1042	*crp*-361Fchk	cccttcgacccacttcactcgcgc
1148	RP_*flhD*+26R	TGTTTCAGCAACTCGGAGGTATGC
1149	RP_*yecG*+28R	CAGCGACAAGAATATTGCTATAGC
1150	*adk*F1	TCATCATCTGCACTTTCCGC
1151	*adk*R1	CCAGATCAGCGCGAACTTCA
1152	*fumC*R1	TCCCGGCAGATAAGCTGTGG
1153	*fumC*F	TCACAGGTCGCCAGCGCTTC
1154	*gyrB*F	TCGGCGACACGGATGACGGC
1155	*gyrB*R1	GTCCATGTAGGCGTTCAGGG
1156	*icd*F	ATGGAAAGTAAAGTAGTTGTTCCGGCACA
1157	*icd*R	GGACGCAGCAGGATCTGTT
1158	*mdh*F1	AGCGCGTTCTGTTCAAATGC
1159	*mdh*R1	CAGGTTCAGAACTCTCTCTGT
1160	*purA*F1	TCGGTAACGGTGTTGTGCTG
1161	*purA*R	CATACGGTAAGCCACGCAGA
1162	*recA*R1	AGCGTGAAGGTAAAACCTGTG
1163	*recA*F1	ACCTTTGTAGCTGTACCACG
1165	RP_*yecG*-40_DELF	TCATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGCCGCCTTACGCCCCGCCC TGC

1178	RP_*flhD*-11_DELR	CATAAATGTGTTTCAGCAACTCGGAGGTATG CATTATTCCCACCCCTAGACTATATTACCCT GTT
1206	*flhD*+26R_ERI	GCGCGAATTCTGTTTCAGCAACTCGGAGGT ATGC
1207	*yecG*+28R_ERI	GGGCGAATTCCAGCGACAAGAATATTGCTAT AGC
1241	pSB401+9706F	ATCTTGCGAATATATGTG
1242	pSB401_*luxC*+86R	GTAAACACTATTATCACC
1250	RP*yecG*+45ER1	GGG CGA ATT CCT TCC GGT GTA ACC GCA ACA GCG AC
1279	*flgB*_pSB_F	GATACCTGCGGAGGAGATATGGaattcaggcttg gaggatac
1280	*flgB*_pSB_R	gtatcctccaagcctgaattCCATATCTCCTCCGCAG GTATC
1285	*fliC*_pSB_F	CAATATAGGATAACGAATCATGGaattcaggcttg gaggatac
1286	*fliC*_pSB_R	gtatcctccaagcctgaattCCATGATTCGTTATCCTA TATTG
1315	Ec*fliC*+1F	ATGGCACAAGTCATTAATACC
1316	Ec*fliC*+1497R	TTAACCCTGCAGCAGAGACAG
1317	delDCUTR_R	AAATGTGTTTCAGCAACTCGGAGGTATGCAT TATTCCCACCCGATTTTTCCTAAATCGACGC AAC
1318	delDCREG_F	GCTTATCTGGTATTGCATTAAAGCGAAATCG CAGCCCCCCTCCGTTG
1319	delDCREG_R	CAACGGAGGGGGGCTGCGATTTCGCTTTAA TGCAATACCAGATAAGC
1320	delDCSCAR_F	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGCATTATTTATAGCAGATGA TTATTTACG
1321	delDCRGSC_F	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGCAAATCGCAGCCCCCCTC CGTTG
1336	pSB*flgA*_newF	caccgtctttcattgccatacggaattcGTTGATTGATATC CAGCACCGTACG
1337	pSB*flgA*_newR	CGTACGGTGCTGGATATCAATCAACgaattccgt atggcaatgaaagacggtg
1338	pSB*fliD*_newF	caccgtctttcattgccatacggaattcGCTTAACGATGA GTTAGCCGCGCTG
1339	pSB*fliD*_newR	CAGCGCGGCTAACTCATCGTTAAGCgaattccg tatggcaatgaaagacggtg
1375	delSCAR_A	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGtttgtttaaaaatggcttatctgg
1376	delSCAR_B	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGgcctttctttgtagttaattagtttg
1377	delSCAR_C	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGaatatatgtaaaacattttattaac
1378	delSCAR_D	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGagaaaataatgtactgatttcc

1379	delSCAR_E	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGaattgctcttggttatcatatg
1380	delSCAR_F	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGaggtaaatattagtctggtcattag
1395	delSCAR_MG_A	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGtctgtttaaaaatagcttatctgg
1396	delSCAR_MG_B	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGgcctttctttgtatttaattagtttg
1397	delSCAR_MG_C	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGaatatatgtaaaacatttcattaac
1398	delSCAR_MG_D	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGagaaaatagtgtactgattctcc
1399	delSCAR_MG_E	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGaatcactcccggtgatcatataatttc
1400	delSCAR_MG_F	CATAACCTGTTCCTTATTCTGTGAACTTCAG GTGACATTAAAGaggtaaatattagtctgattattag
1403	*yecG*+6WRGR	tgtcacctgaagttcacagaataaggaacaggttatgagcCG CCTTACGCCCCGCCCTGC
1404	*yecG*+463WRGF	agttgctacctttcctaagcttgcgtaggtgcggttaatcCTAGA CTATATTACCCTGTT
1415	*yecG*+546R	GATATTAACCACTGGCAGGAG
1432	CFT_SCRREG_F	GGCTTATCTGGTATTGCATTAAAGCGATTAT TTATAGCAGATGATTATTTACGG
1433	CFT_SCRREG_R	CCGTAAATAATCATCTGCTATAAATAATCGCT TTAATGCAATACCAGATAAGCC
1434	pSB*flhD*F	caccgtctttcattgccatacggaattcTGTTTCAGCAACT CGGAGGTATGC
1435	*flhD*pSBR	GCATACCTCCGAGTTGCTGAAACAgaattccgta tggcaatgaaagacggtg
1436	*yecG*pSBR	GCTATAGCAATATTCTTGTCGCCGaattcaggctt ggaggatac
1437	pSB*yecG*F	gtatcctccaagcctgaattCGGCGACAAGAATATTG CTATAGC

8.4. Strains Background

Table. 8.20. Strain genotype

Strain Number	Genotype	Source
2	DH5a (PH1-80dLACDM15 ENDA1 RECA1 HSDR17 SUPE44 TTH-1 GYRA96 RELA12 DLACU169	Aldridge's lab
22	pkD3 in *E.coli* (strain unknown PA)	Aldridge's lab
728	pSB401 in *E. coli*	Aldridge's lab
1944	DH5a / pBluescriptKSII	Aldridge's lab
3066	*E. coli* NCTC10418	Aldridge's lab
3373	CFT073	Aldridge's lab
3405	IS in *yecG-flhDC*/ *E. coli* Clinical Isolate 106205X	Aldridge's lab
3406	*E. coli* Clinical Isolate 105264X	Aldridge's lab

3408	*E. coli* Clinical Isolate 100273D	Aldridge's lab
3426	IS in *yecG-flhDC*/ RP437	Aldridge's lab
3428	JPA945 (FLIM-YPET)	Aldridge's lab
3439	MG1655	Aldridge's lab
3459	MG1655 / pKD46	Aldridge's lab
3799	pWRG99 / DH5a [AmpR grow at 30°C]	Aldridge's lab
3800	pWRG100 / DH5a	Aldridge's lab
4066	Δ*yecG-flhDC*(3439)::FCF in TPA3459	This study
4069	Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+)	This study
4074	Δ*yecG-flhDC*3439::*yecG-flhDC*3408 (mot+)	This study
4083	Δ*yecG-flhDC*3439::*yecG-flhDC*3406 (mot+)	This study
4102	pSB401ΔER [pSB401 with 1kb EcoRI insert deleted] / DH5a	This study
4195	pBluescriptKSII::P_{flhDC}(CFT073)/ DH5a	This study
4196	pBluescriptKSII::P_{flhDC}(MG1655)/ DH5a	This study
4197	pSB401::P_{flhDC}(CFT073)/ DH5a	This study
4198	pSB401::P_{flhDC}(MG1655)/ DH5a	This study
4238	pSB401::P_{flhDC}(CFT073)/ *E. coli* NCTC10418	This study
4239	pSB401::P_{flhDC}(MG1655)/ *E. coli* NCTC10418	This study
4240	pSB401ΔER/ *E. coli* NCTC10418	This study
4241	pSB401::P_{flhDC}(CFT073)/ MG1655	This study
4242	pSB401::P_{flhDC}(MG1655)/ MG1655	This study
4243	pSB401ΔER/ MG1655	This study
4256	pSB401::P_{flhDC}(CFT073)/ CFT073	This study
4257	pSB401::P_{flhDC}(MG1655)/ CFT073	This study
4258	pSB401ΔER/ CFT073	This study
4259	pBluescriptKSII::P_{flhDC}(*E. coli* Clinical Isolate 100273D)/ DH5a	This study
4260	pBluescriptKSII::P_{flhDC}(*E. coli* Clinical Isolate 105264X)/ DH5a	This study
4425	pSB401::P_{flhDC}(*E. coli* Clinical Isolate 105264X)/ DH5a	This study
4460	pSB401::P_{flhDC}(*E. coli* Clinical Isolate 105264X)/ *E. coli* NCTC10418	This study
4461	pSB401::P_{flhDC}(*E. coli* Clinical Isolate 105264X/ CFT073	This study
4462	pSB401::P_{flhDC}(*E. coli* Clinical Isolate 105264X/ MG1655	This study
4465	pSB401::P_{flhDC}(*E. coli* Clinical Isolate 100273D)/ DH5a	This study
4488	pSB401::P_{flhDC}(*E. coli* Clinical Isolate 100273D)/ *E. coli* NCTC10418	This study
4489	pSB401::P_{flhDC}(*E. coli* Clinical Isolate 100273D)/ CFT073	This study
4490	pSB401::P_{flhDC}(*E. coli* Clinical Isolate 100273D)/ MG1655	This study
4571	pCas/ DH5a [KanR grow at 30°C]	This study
4686	*ΔompR*::FCF in 4069	This study
4687	*ΔftsH*::FCF in 4069	This study
4688	*ΔclpP*::FCF in 4069	This study

4689	*Δlon*::FCF in 4069	This study
4690	*Δcrp*::FCF in 4069	This study
4691	*ΔompR*::FCF in 4074	This study
4692	*ΔftsH*::FCF in 4074	This study
4693	*ΔclpP*::FCF in 4074	This study
4694	*Δlon*::FCF in 4074	This study
4695	*Δcrp*::FCF in 4074	This study
4696	*ΔompR*::FCF in 4083	This study
4697	*ΔftsH*::FCF in 4083	This study
4698	*ΔclpP*::FCF in 4083	This study
4699	*Δlon*::FCF in 4083	This study
4700	*Δcrp*::FCF in 4083	This study
4701	*ΔompR*::FCF in 3459	This study
4702	*ΔftsH*::FCF in 3459	This study
4703	*ΔclpP*::FCF in 3459	This study
4704	*Δlon*::FCF in 3459	This study
4705	*Δcrp*::FCF in 3459	This study
4718	ΔfliM::FCF in 4069	This study
4720	UTI89	Aldridge's lab
4723	pTrg-Cm/ DH5a [SpectinomycinR grow at 30 ˚C]	This study
4724	*ΔfliM*::FCF ΔFCF::*fliM-ypeT* in 4069	This study
4725	*ΔfliM*::FCF ΔFCF::*fliM-ypeT* in 4069	This study
4728	*ΔfliM*::FCF in 3459	This study
4729	*ΔfliM*::FCF in 4074	This study
4730	*ΔfliM*::FCF in 4083	This study
4759	*ΔfliM*::FCF ΔFCF::*fliM-ypeT* in 4074	This study
4760	*ΔfliM*::FCF ΔFCF::*fliM-ypeT* in 4074	This study
4761	*ΔfliM*::FCF ΔFCF::*fliM-ypeT* in 4083	This study
4762	*ΔfliM*::FCF ΔFCF::*fliM-ypeT* in 4083	This study
4763	*ΔfliM*::FCF ΔFCF::*fliM-ypeT* in 3459	This study
4764	*ΔfliM*::FCF ΔFCF::*fliM-ypeT* in 3459	This study
4785	pKD46/ DH5a	Aldridge's lab
4818	4066::*yecG-flhDC*(CFT073)	This study
4819	4066::*yecG-flhDC*(CFT073)	This study
4820	4066::*yecG-flhDC*ΔSCAR(CFT073)	This study
4821	4066::*yecG-flhDC*ΔSCAR(CFT073)	This study
4822	4066::*yecG-flhDC*ΔRG-SCR(CFT073)	This study
4823	4066::*yecG-flhDC*ΔRG-SCR(CFT073)	This study
4824	4066::*yecG-flhDC*ΔUTR(CFT073)	This study
4825	4066::*yecG-flhDC*ΔUTR(CFT073)	This study
4826	4066::*yecG-flhDC*ΔUTR-SCR(CFT073)	This study
4827	4066::*yecG-flhDC*ΔUTR-SCR(CFT073)	This study
4828	4066::*yecG-flhDC*ΔU-R-S(CFT073)	This study
4829	4066::*yecG-flhDC*ΔU-R-S(CFT073)	This study
4830	4066::*yecG-flhDC*(MG1655)	This study
4831	4066::*yecG-flhDC*(MG1655)	This study
4832	4066::*yecG-flhDC*ΔSCAR(MG1655)	This study
4833	4066::*yecG-flhDC*ΔSCAR(MG1655)	This study
4834	4066::*yecG-flhDC*ΔRG-SCR(MG1655)	This study
4835	4066::*yecG-flhDC*ΔRG-SCR(MG1655)	This study

4836	4066::*yecG-flhDC*ΔUTR(MG1655)	This study
4837	4066::*yecG-flhDC*ΔUTR(MG1655)	This study
4838	4066::*yecG-flhDC*ΔUTR-SCR(MG1655)	This study
4839	4066::*yecG-flhDC*ΔUTR-SCR(MG1655)	This study
4840	4066::*yecG-flhDC*ΔU-R-S(MG1655)	This study
4841	4066::*yecG-flhDC*ΔU-R-S(MG1655)	This study
4876	4066::*yecG-flhDC*ΔSCAR(CFT073) round 2	This study
4877	4066::*yecG-flhDC*ΔSCAR(CFT073) round 2	This study
4878	4066::*yecG-flhDC*ΔSCAR-A(CFT073)	This study
4879	4066::*yecG-flhDC*ΔSCAR-A(CFT073)	This study
4880	4066::*yecG-flhDC*ΔSCAR-B(CFT073)	This study
4881	4066::*yecG-flhDC*ΔSCAR-B(CFT073)	This study
4882	4066::*yecG-flhDC*ΔSCAR-D(CFT073)	This study
4883	4066::*yecG-flhDC*ΔSCAR-D(CFT073)	This study
4884	4066::*yecG-flhDC*ΔSCAR-E(CFT073)	This study
4885	4066::*yecG-flhDC*ΔSCAR-E(CFT073)	This study
4886	4066::*yecG-flhDC*ΔSCAR-F(CFT073)	This study
4887	4066::*yecG-flhDC*ΔSCAR-F(CFT073)	This study
4888	pSB401::P_{flgB}/ DH5a	This study
4889	pSB401::P_{fliC}/ DH5a	This study
4890	pSB401::P_{flgB}/ 4066::*yecG-flhDC*(CFT073)	This study
4891	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR(CFT073)	This study
4892	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔRG-SCR(CFT073)	This study
4893	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔUTR(CFT073)	This study
4894	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔUTR-SCR(CFT073)	This study
4895	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔU-R-S(CFT073)	This study
4896	pSB401::P_{flgB}/ 4066::*yecG-flhDC*(MG1655)	This study
4897	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR(MG1655)	This study
4898	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔRG-SCR(MG1655)	This study
4899	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔUTR(MG1655)	This study
4900	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔUTR-SCR(MG1655)	This study
4901	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔU-R-S(MG1655)	This study
4902	pSB401::P_{flgB}/ MG1655	This study
4903	pSB401::P_{flgB}/ Δ*yecG-flhDC*(3439)::FCF in TPA3459 [PCR1165/1178]	This study
4904	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR(CFT073) round 2	This study
4905	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-A(CFT073)	This study
4906	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-B(CFT073)	This study

4907	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-D(CFT073)	This study
4908	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-E(CFT073)	This study
4909	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-F(CFT073)	This study
4910	4066::*yecG-flhDC*ΔSCAR(MG1655) round 2	This study
4911	4066::*yecG-flhDC*ΔSCAR(MG1655) round 2	This study
4912	4066::*yecG-flhDC*ΔSCAR-A(MG1655)	This study
4913	4066::*yecG-flhDC*ΔSCAR-A(MG1655)	This study
4914	4066::*yecG-flhDC*ΔSCAR-B(MG1655)	This study
4915	4066::*yecG-flhDC*ΔSCAR-B(MG1655)	This study
4916	4066::*yecG-flhDC*ΔSCAR-C(MG1655)	This study
4917	4066::*yecG-flhDC*ΔSCAR-C(MG1655)	This study
4918	4066::*yecG-flhDC*ΔSCAR-D(MG1655)	This study
4919	4066::*yecG-flhDC*ΔSCAR-D(MG1655)	This study
4920	4066::*yecG-flhDC*ΔSCAR-E(MG1655)	This study
4921	4066::*yecG-flhDC*ΔSCAR-E(MG1655)	This study
4922	4066::*yecG-flhDC*ΔSCAR-F(MG1655)	This study
4923	4066::*yecG-flhDC*ΔSCAR-F(MG1655)	This study
4924	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR(MG1655) round 2	This study
4925	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-A(MG1655)	This study
4926	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-B(MG1655)	This study
4927	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-C(MG1655)	This study
4928	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-D(MG1655)	This study
4929	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-E(MG1655)	This study
4930	pSB401::P_{flgB}/ 4066::*yecG-flhDC*ΔSCAR-F(MG1655)	This study
4969	pSB401::P_{fliC}/ 4066::*yecG-flhDC*(CFT073)	This study
4970	pSB401::P_{fliC}/ 4066::*yecG-flhDC*ΔSCAR1(CFT073)	This study
4971	pSB401::P_{fliC}/ 4066::*yecG-flhDC*ΔSCAR2(CFT073)	
4972	pSB401::P_{fliC}/ 4066::*yecG-flhDC*ΔRG-SCR(CFT073)	This study
4973	pSB401::P_{fliC}/ 4066::*yecG-flhDC*ΔUTR(CFT073)	This study
4974	pSB401::P_{fliC}/ 4066::*yecG-flhDC*ΔUTR-SCR(CFT073)	This study
4975	pSB401::P_{fliC}/ 4066::*yecG-flhDC*ΔU-R-S(CFT073)	This study
4976	pSB401::P_{fliC}/ 4066::*yecG-flhDC*ΔSCAR-A(CFT073)	This study
4977	pSB401::P_{fliC}/ 4066::*yecG-flhDC*ΔSCAR-B(CFT073)	This study

4978	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-D(CFT073)	This study
4979	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-E(CFT073)	This study
4980	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-F(CFT073)	This study
4981	pSB401::P*fliC*/ 4066::*yecG-flhDC*(MG1655)	This study
4982	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR1(MG1655)	This study
4983	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR2(MG1655)	
4984	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔRG-SCR(MG1655)	This study
4985	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔUTR(MG1655)	This study
4986	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔUTR-SCR(MG1655)	This study
4987	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔU-R-S(MG1655)	This study
4988	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-A(MG1655)	This study
4989	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-B(MG1655)	This study
4990	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-C(MG1655)	This study
4991	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-D(MG1655)	This study
4992	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-E(MG1655)	This study
4993	pSB401::P*fliC*/ 4066::*yecG-flhDC*ΔSCAR-F(MG1655)	This study
4994	pSB401::P*fliC*/ MG1655	This study
4995	pSB401::P*fliC*/ Δ*yecG-flhDC*(3439)::FCF in TPA3459	This study
4996	pSB401::P*flhDC*(CFT073)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+)	This study
4997	pSB401::P*flhDC*(MG1655)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+)	This study
4998	pSB401::P*flhDC*(*E. coli* Clinical Isolate 100273D)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+)	This study
4999	pSB401::P*flhDC*(*E. coli* Clinical Isolate 105264X)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+)	This study
5000	pSB401::P*flhDC*(CFT073)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3408 (mot+)	This study
5001	pSB401::P*flhDC*(MG1655)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3408 (mot+)	This study
5002	pSB401::P*flhDC*(*E. coli* Clinical Isolate 100273D)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3408 (mot+)	This study
5003	pSB401::P*flhDC*(*E. coli* Clinical Isolate 105264X)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3408 (mot+)	This study
5004	pSB401::P*flhDC*(CFT073)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3406 (mot+)	This study

5005	pSB401::P*flhDC*(MG1655)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3406 (mot+)	This study
5006	pSB401::P*flhDC*(*E. coli* Clinical Isolate 100273D)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3406 (mot+)	This study
5007	pSB401::P*flhDC*(*E. coli* Clinical Isolate 105264X)/ Δ*yecG-flhDC*3439::*yecG-flhDC*3406 (mot+)	This study
5008	pSB401::P*flhDC*(CFT073)/ MG1655	This study
5009	pSB401::P*flhDC*(MG1655)/ MG1655	This study
5010	pSB401::P*flhDC*(*E. coli* Clinical Isolate 100273D)/ MG1655	This study
5011	pSB401::P*flhDC*(*E. coli* Clinical Isolate 105264X)/ MG1655	This study
5012	pSB401::P*flhDC*(CFT073)/ Δ*yecG-flhDC*(3439)::FCF in TPA3459	This study
5013	pSB401::P*flhDC*(MG1655)/ Δ*yecG-flhDC*(3439)::FCF in TPA3459	This study
5014	pSB401::P*flhDC*(*E. coli* Clinical Isolate 100273D)/ Δ*yecG-flhDC*(3439)::FCF in TPA3459	This study
5015	pSB401::P*flhDC*(*E. coli* Clinical Isolate 105264X)/ Δ*yecG-flhDC*(3439)::FCF in TPA3459	This study
5022	UTI89ΔP*flhDC*::FCF	This study
5023	Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+) Δ*yecG*::FCF	This study
5024	Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+) Δ*yecG*::FCF	This study
5025	Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+) Δ*yecG*::FCF	This study
5026	Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+) Δ*yecG*::FCF	This study
5027	Δ*yecG-flhDC*3439::*yecG-flhDC*3373 (mot+) Δ*yecG*::FCF	This study
5120	UTI89ΔP*flhDC*::P*flhDC*(MG1655)	This study
5121	MG1655Δ*yecG*::FCF	This study
5435	MG1655ΔSCAR::SCAR(CFT073)	This study
5436	MG1655ΔSCAR::SCAR(CFT073)	This study
5437	pSB401::P*yecG*(MG1655)/ MG1655	This study
5438	pSB401::P*yecG*(MG1655)/ 4066::*yecG-flhDC*ΔSCAR(MG1655)	This study
5439	pSB401::P*yecG*(MG1655)/ 4066::*yecG-flhDC*ΔSCAR-F(MG1655)	This study
5440	pSB401::P*yecG*(MG1655)/ 4066::*yecG-flhDC*(CFT073)	This study
5441	pSB401::P*yecG*(MG1655)/ 4066::*yecG-flhDC*ΔSCAR(CFT073)	This study
5442	pSB401::P*yecG*(MG1655)/ 4066::*yecG-flhDC*ΔSCAR-F(CFT073)	This study
5443	pSB401::P*yecG*(CFT073)/ MG1655	This study
5444	pSB401::P*yecG*(CFT073)/ 4066::*yecG-flhDC*ΔSCAR(MG1655)	This study
5445	pSB401::P*yecG*(CFT073)/ 4066::*yecG-flhDC*ΔSCAR-F(MG1655)	This study

5446	pSB401::P_{yecG}(CFT073)/ 4066::*yecG-flhDC*(CFT073)	This study
5447	pSB401::P_{yecG}(CFT073)/ 4066::*yecG-flhDC*ΔSCAR(CFT073)	This study
5448	pSB401::P_{yecG}(CFT073)/ 4066::*yecG-flhDC*ΔSCAR-F(CFT073)	This study
5449	pSB401::P_{yecG}(MG1655)/ DH5a	This study
5450	pSB401::P_{yecG}(CFT073)/ DH5a	This study
5451	pSB401::P_{yecG}(MG1655)/ DH5a/ Δ*yecG-flhDC*(3439)::FCF in TPA3459 [PCR1165/1178]	This study
5452	pSB401::P_{yecG}(CFT073)/ DH5a/ Δ*yecG-flhDC*(3439)::FCF in TPA3459 [PCR1165/1178]	This study
5456	MG1655ΔP_{flhDC}::P_{flhDC}(UTI89)	This study
5457	MG1655ΔP_{flhDC}::P_{flhDC}(UTI89)	This study

Table. 8.21. Background of bovine isolates used in the characterization

TPA	Site	Date	Kingston no.
5201	DWV1	15/12/2015	17D2
5202	DWV1	15/12/2015	12D
5203	DWV1	15/12/2015	13E
5204	DWV1	15/12/2015	4A1
5205	DWV1	15/12/2015	1DE
5405	DWV1	15/12/2015	4A2
5206	DWV1	15/12/2015	1B
5207	DWV1	15/12/2015	9E
5208	DWV1	15/12/2015	5D
5209	DWV1	15/12/2015	12B
5210	DWV1	15/12/2015	12F
5211	DWV1	15/12/2015	6B1
5212	DWV1	15/12/2015	1C
5213	DWV1	15/12/2015	12C
5214	DWV1	15/12/2015	10B
5215	DWV1	15/12/2015	10D
5216	DWV1	15/12/2015	14E
5217	DWV1	15/12/2015	3B
5218	DWV1	15/12/2015	10A
5407	DWV1	15/12/2015	3E2
5219	DWV1	15/12/2015	1DE
5220	DWV1	15/12/2015	2B
5221	DWV1	15/12/2015	2A
5409	DWV1	15/12/2015	16F
5222	DWV1	15/12/2015	3D
5223	DWV1	15/12/2015	79/7G
5224	DWV1	15/12/2015	17212
5225	DWV1	15/12/2015	7F
5411	DWV1	15/12/2015	11B1
5226	DWV1	15/12/2015	6B2
5227	DWV1	15/12/2015	10C

5228	DWV1	15/12/2015	1E
5229	DWV1	15/12/2015	13A
5230	DWV1	15/12/2015	6A
5231	DWV1	15/12/2015	5E
5232	DWV1	15/12/2015	7A
5234	DWV1	15/12/2015	6D2
5235	DWV1	15/12/2015	3A
5236	DWV1	15/12/2015	9B
5237	DWV1	15/12/2015	7C2
5238	DWV1	12/12/2015	17D1
5239	DWV1	12/12/2015	2C
5403	DWV1	12/12/2015	12F
5400	DWV1	12/12/2015	11B2
5242	DWV1	12/12/2015	15D
5243	DWV1	12/12/2015	6D1
5244	DWV1	12/12/2015	170
5245	DWV1	12/12/2015	13D
5246	DWV1	12/12/2015	300
5247	DWV1	12/12/2015	12E
5248	DWV1	12/12/2015	5C
5249	DWV1	12/12/2015	15L
5250	DWV1	12/12/2015	6A2
5251	DWV1	15/12/2015	7E
5252	DWV1	15/12/2015	1DF
5253	DWV1	12/12/2015	16B
5254	???	16/05/2016	1L
5255	???	16/05/2016	13L
5256	???	16/05/2016	10B
5257	DW	18/12/2015	3F
5258	DW	18/12/2015	17C2
5259	DW	18/12/2015	9A
5260	DW	18/12/2015	7C1
5398	DW	18/12/2015	18A
5261	DW	18/12/2015	4E
5262	Binnington	16/05/2016	19B
5265	Binnington	16/05/2016	3L
5266	Binnington	16/05/2016	8L
5267	Binnington	16/05/2016	7L
5268	Binnington	16/05/2016	10L
5269	Binnington	16/05/2016	4L?
5270	Binnington	16/05/2016	5c
5271	Binnington	16/05/2016	20B
5272	Binnington	16/05/2016	1D
5274	Binnington	16/05/2016	14L
5275	Binnington	16/05/2016	12C
5276	Binnington	16/05/2016	23C
5277	???	16/05/2016	9L
5278	???	16/05/2016	12b
5279	???	16/05/2016	14b
5280	???	16/05/2016	1B

5281	???	16/05/2016	15L
5283	???	16/05/2016	2L
5284	LynwoodV3	06/04/2015	15c
5285	LynwoodV3	06/04/2015	17D
5286	LynwoodV3	06/04/2015	3D
5287	LynwoodV3	06/04/2015	25b
5288	LynwoodV3	06/04/2015	6d
5289	LynwoodV3	06/04/2015	6C
5290	LynwoodV3	06/04/2015	21B
5291	LynwoodV3	06/04/2015	27D
5292	LynwoodV3	06/04/2015	4B
5293	LynwoodV3	06/04/2015	16B
5294	LynwoodV3	06/04/2015	25C
5396	LynwoodV3	06/04/2015	16E
5295	LynwoodV3	06/04/2015	24b
5296	LynwoodV3	06/04/2015	4C
5415	LynwoodV3	06/04/2015	12b
5297	LynwoodV3	06/04/2015	17e
5298	LynwoodV3	06/04/2015	1D1
5299	LynwoodV3	06/04/2015	27C
5300	LynwoodV3	06/04/2015	25D
5301	LynwoodV3	06/04/2015	1D2
5303	F2	14/12/2015	14E
5304	F2	14/12/2015	300
5305	F2	14/12/2015	5D
5306	F2	14/12/2015	6E
5307	F2	14/12/2015	8C
5394	F2	14/12/2015	6C
5308	F2	14/12/2015	7L
5309	F2	14/12/2015	19B
5310	F2	14/12/2015	18c
5311	F2	14/12/2015	12A
5312	F2	14/12/2015	16B
5313	F2	14/12/2015	6D
5314	F2	14/12/2015	13?
5315	F2	10/12/2015	12D
5316	F2	10/12/2015	1F
5317	F2	10/12/2015	4D
5318	F2	10/12/2015	5F
5319	F2	10/12/2015	4F
5320	F2	10/12/2015	10D
5321	F2	10/12/2015	20A
5322	F2	10/12/2015	16E
5324	F2	10/12/2015	12C
5325	F2	10/12/2015	7A
5326	F2	10/12/2015	14A
5327	F2	10/12/2015	1E
5328	F2	10/12/2015	13C
5329	F2	10/12/2015	9B
5330	F2	10/12/2015	15C

5331	F2	10/12/2015	5E
5332	F2	10/12/2015	10B
5333	F2	10/12/2015	2B
5334	BV3	27/04/2016	9B
5335	BV3	27/04/2016	7B
5336	BV3	27/04/2016	21B
5337	BV3	27/04/2016	24B
5393	BV3	27/04/2016	23C
5339	BV3	27/04/2016	15b
5340	BV3	27/04/2016	17C
5341	BV3	27/04/2016	21C
5342	BV3	27/04/2016	6C
5343	BV3	27/04/2016	13C
5344	BV3	27/04/2016	5b
5345	BV3	27/04/2016	17b
5346	BV3	27/04/2016	16d
5347	BV3	27/04/2016	176
5348	BV3	27/04/2016	24C
5349	BV3	27/04/2016	13C
5350	BV3	27/04/2016	14b
5351	BV3	27/04/2016	6e
5353	BV3	27/04/2016	1B
5354	BV3	27/04/2016	15A
5355	BV3	27/04/2016	14C
5356	BV3	27/04/2016	9C
5358	BV3	27/04/2016	12C
5359	BV3	27/04/2016	11b
5360	BV3	27/04/2016	19b
5361	BV3	27/04/2016	24d
5362	BV3	27/04/2016	4d
5363	BV3	27/04/2016	8C
5364	BV3	27/04/2016	20C
5365	BV3	27/04/2016	7b
5366	BV3	27/04/2016	4C
5367	BV3	27/04/2016	11C
5368	BV3	27/04/2016	15C
5370	BV3	27/04/2016	2e
5371	BV3	27/04/2016	23b
5372	BV3	27/04/2016	10C
5373	BV4	16/05/2016	11C
5374	BV4	16/05/2016	4C
5376	BV4	16/05/2016	17C
5377	BV4	16/05/2016	19C
5379	BV4	16/05/2016	11C
5380	BV4	16/05/2016	10C
5381	BV4	16/05/2016	18C
5382	BV4	16/05/2016	16B
5384	BV4	16/05/2016	6C
5387	Binnington	13/12/2016	13D
5388	Binnington	14/12/2015	2b1

5389	Binnington	14/12/2015	2F
5390	Binnington	14/12/2015	20B
5391	Binnington	V1	15C
5392	Binnington	14/12/2015	15B
5385	Binnington	14/12/2015	9A

Table. 8.22. Background of UTI isolates used in the characterization

*UTI number	**Donation number	***MLST
UTI115	3	ST131
UTI115	6	ST681
UTI115	7	ST131
UTI115	11	ST131
UTI115	12	ST681
UTI138	1	ST131
UTI138	2	ST131
UTI138	3	ST131
UTI138	4	ST677
UTI138	5	ST677
UTI138	6	ST131
UTI138	7	ST677
UTI138	8	ST677
UTI138	9	ST677
UTI138	10	ST677
UTI138	11	ST677
UTI138	12	ST677
UTI139	5	ST131
UTI139	8	ST131
UTI139	9	ST131
UTI218	2	ST73
UTI218	5	ST73
UTI236	4	ST421
UTI337	1	ST73
UTI337	2	ST73
UTI337	3	ST73
UTI337	4	ST73
UTI337	5	ST73
UTI337	6	ST59
UTI337	7	ST73
UTI337	8	ST73
UTI337	9	ST73
UTI337	10	ST73
UTI337	11	ST73
UTI337	12	ST73
UTI343	1	ST144
UTI343	2	ST144

UTI343	3	ST144
UTI343	4	ST144
UTI343	5	ST144
UTI343	7	ST12
UTI343	8	ST12
UTI343	9	ST144
UTI343	10	ST144
UTI343	11	ST69
UTI365	2	ST335
UTI365	4	ST404
UTI365	5	ST404
UTI365	6	ST404
UTI365	7	ST404
UTI365	8	ST404
UTI365	9	ST404
UTI365	10	ST404
UTI365	12	ST404
UTI376	9	ST362
UTI376	12	ST362
UTI383	1	ST12
UTI383	2	ST12
UTI383	3	ST12
UTI383	4	ST12
UTI383	5	ST12
UTI383	6	ST12
UTI383	7	ST12
UTI383	8	ST12
UTI383	9	ST12
UTI383	10	ST12
UTI383	12	ST12
UTI414	1	ST91
UTI414	5	ST91
UTI414	7	ST91
UTI414	8	ST91
UTI414	10	ST91
UTI468	1	ST73
UTI468	2	ST73
UTI468	3	ST73
UTI468	4	ST73
UTI468	5	ST73
UTI468	6	ST73
UTI468	7	ST73
UTI468	8	ST73
UTI468	9	ST73
UTI468	10	ST73
UTI468	11	ST73

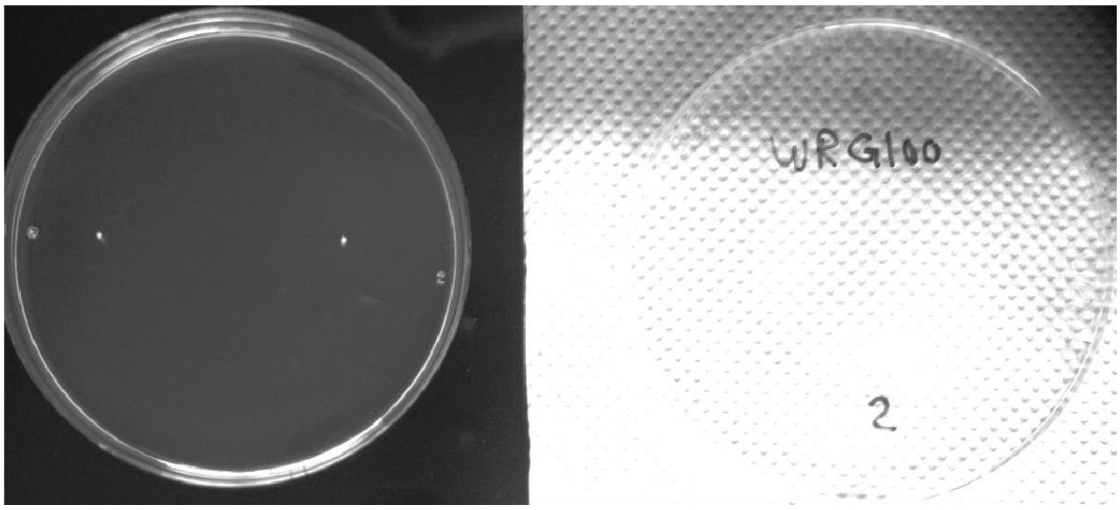
UTI468	12	ST73
UTI524	2	ST73
UTI524	5	ST73
UTI524	7	ST73
UTI524	8	ST38
UTI524	11	ST73
UTI524	12	ST73
UTI531	1	ST420
UTI531	2	ST420
UTI531	3	ST420
UTI531	4	ST420
UTI531	5	ST420
UTI531	6	ST420
UTI531	7	ST583
UTI531	8	ST420
UTI531	9	ST420
UTI531	10	ST420
UTI531	11	ST420
UTI531	12	ST420
UTI536	1	ST73
UTI536	2	ST355
UTI536	3	ST355
UTI536	5	ST73
UTI536	6	ST355
UTI536	7	ST355
UTI536	8	ST355
UTI536	10	ST69
UTI536	12	ST355
UTI562	1	ST625
UTI562	2	ST625
UTI562	11	ST1571
UTI569	5	ST3640
UTI569	6	ST442
UTI569	8	ST3640
UTI569	9	ST442
UTI569	10	ST3640
UTI569	11	ST3640
UTI569	12	ST3640
UTI675	1	ST69
UTI675	2	ST69
UTI675	3	ST69
UTI675	6	ST69
UTI675	7	ST69
UTI675	10	ST69
UTI675	12	ST69
UTI726	2	ST69

UTI726	3	ST73
UTI726	5	ST73
UTI726	6	ST73
UTI726	8	ST73
UTI726	9	ST73
UTI726	10	ST73
UTI726	11	ST73
UTI726	12	ST73
UTI755	6	ST95
UTI755	7	ST95
UTI755	8	ST95
UTI755	9	ST95
UTI755	12	ST95
UTI781	1	ST127
UTI781	2	ST127
UTI781	3	ST127
UTI781	4	ST127
UTI781	5	ST127
UTI781	6	ST127
UTI781	7	ST127
UTI781	8	ST127
UTI781	9	ST127
UTI781	12	ST127
UTI840	3	ST354
UTI840	4	ST354
UTI840	5	ST354
UTI840	6	ST354
UTI840	7	ST354
UTI840	8	ST354
UTI840	9	ST354
UTI840	10	ST73
UTI840	11	ST354
UTI840	12	ST354
UTI891	2	ST127
UTI899	1	ST69
UTI899	2	ST602
UTI899	3	ST73
UTI899	4	ST73
UTI899	5	ST73
UTI899	6	ST73
UTI899	7	ST73
UTI899	8	ST73
UTI899	9	ST73
UTI899	10	ST73
UTI899	11	ST73
UTI899	12	ST73

UTI924	1	ST12
UTI924	2	ST12
UTI924	3	ST73
UTI924	4	ST73
UTI924	7	ST80
UTI924	10	ST677
UTI966	5	ST69
UTI966	7	ST95
UTI966	10	ST73

* Anonymised patient identifier used during UTI study
** Donation *E. coli* isolated from. Each patient provided 12 donations over 6 months
*** Sequence type defined by Warwick scheme

A)



B)

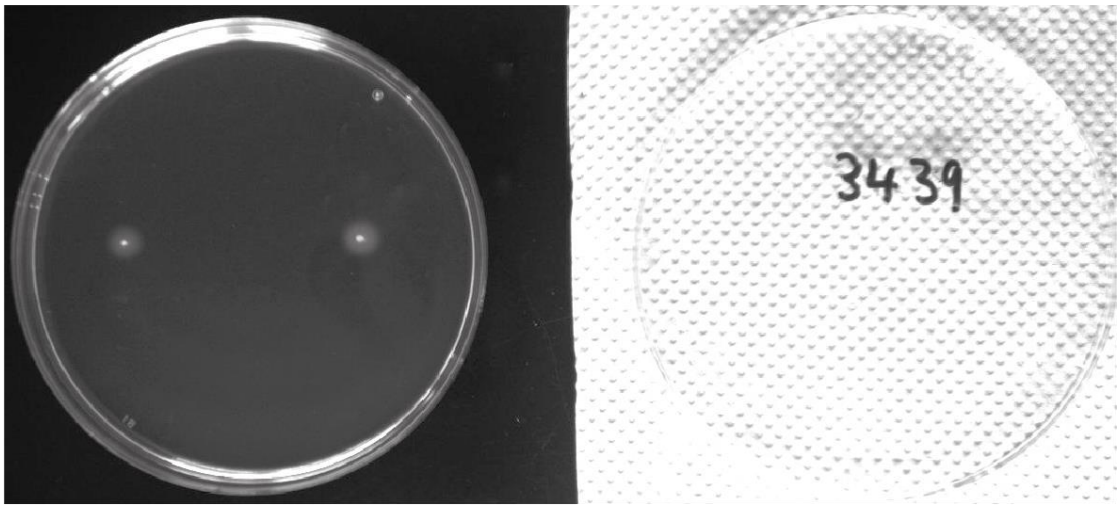


Figure. 8.1. Image showing an example of full motility plates used during the *flhDC* promoter replacement of strain MG1655.

A) An example of a full motility plate cropped in figure 4.2C. **B)** Motility phenotype of strain MG1655 wild type.

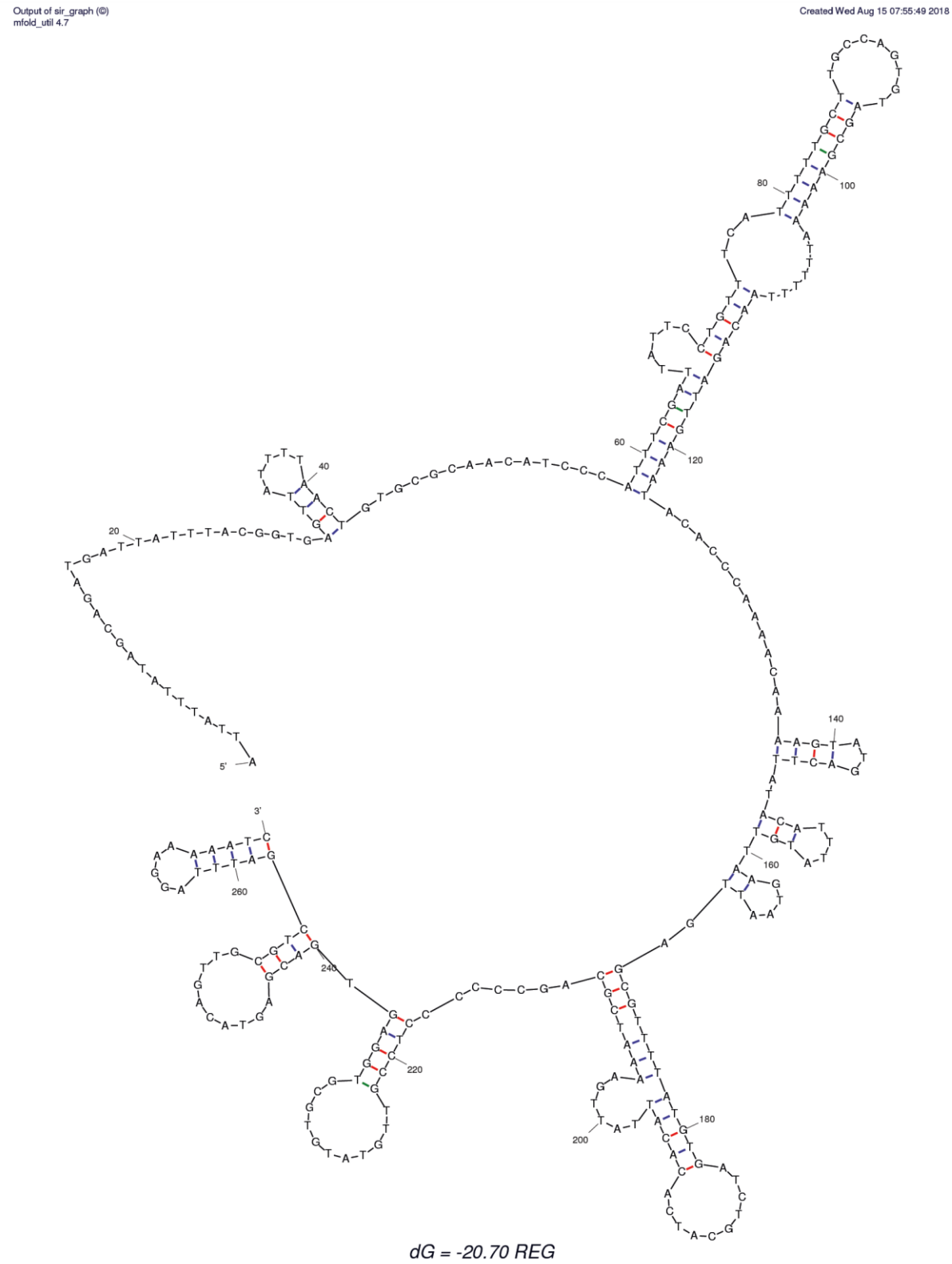
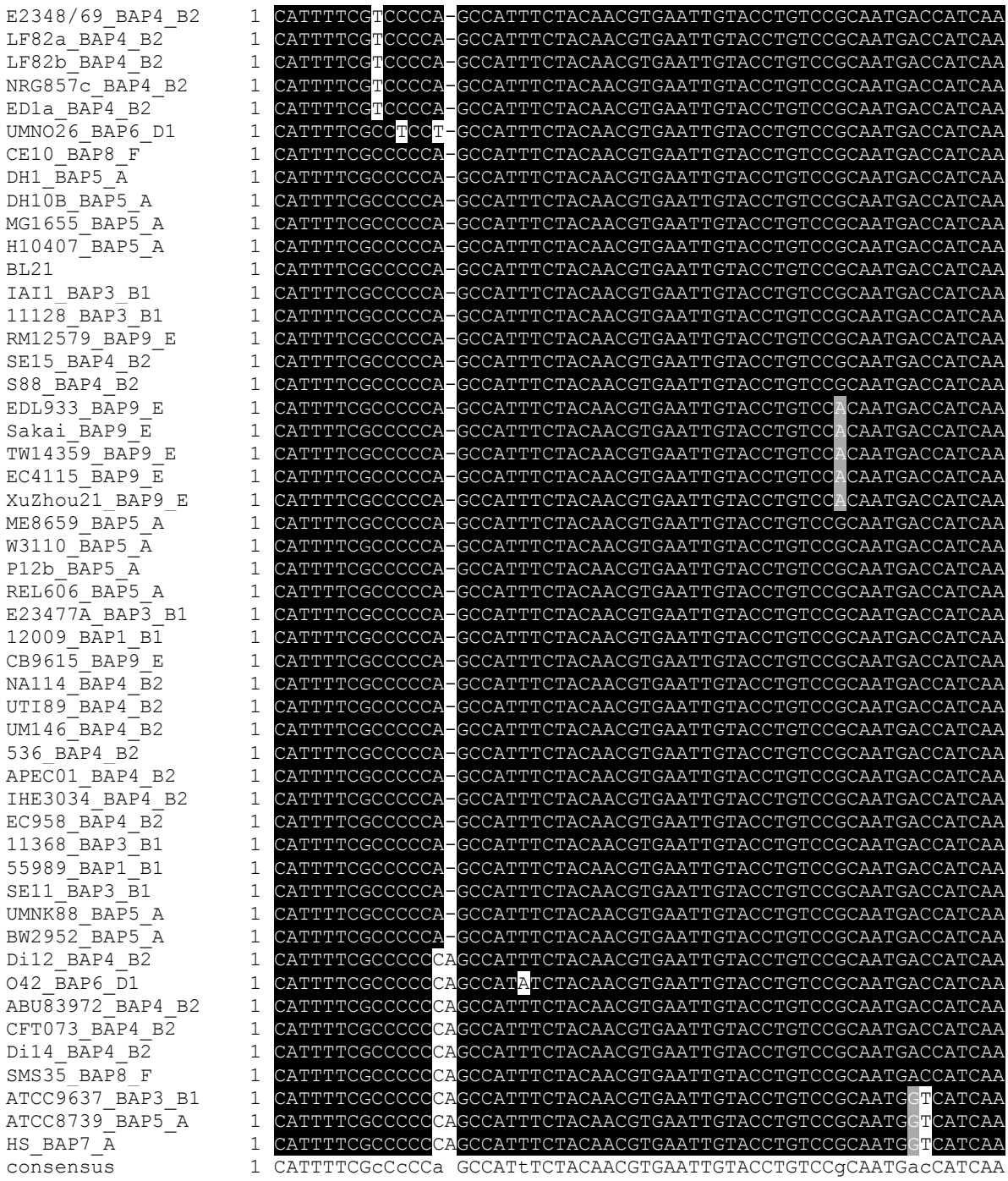


Figure. 8.2. The secondary structure of the DNA sequence in figure. 5.4.

The output of mFold and shows the secondary structure of the DNA sequence in figure. 5.4 if it was single stranded DNA (which impacts PCR reactions).


E2348/69_BAP4_B2 1 CATTTTCGTCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
LF82a_BAP4_B2 1 CATTTTCGTCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
LF82b_BAP4_B2 1 CATTTTCGTCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
NRG857c_BAP4_B2 1 CATTTTCGTCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
ED1a_BAP4_B2 1 CATTTTCGTCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
UMNO26_BAP6_D1 1 CATTTTCGCCTCCT-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
CE10_BAP8_F 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
DH1_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
DH10B_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
MG1655_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
H10407_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
BL21 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
IAI1_BAP3_B1 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
11128_BAP3_B1 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
RM12579_BAP9_E 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
SE15_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
S88_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
EDL933_BAP9_E 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCACAATGACCATCAA
Sakai_BAP9_E 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCACAATGACCATCAA
TW14359_BAP9_E 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCACAATGACCATCAA
EC4115_BAP9_E 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCACAATGACCATCAA
XuZhou21_BAP9_E 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCACAATGACCATCAA
ME8659_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
W3110_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
P12b_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
REL606_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
E23477A_BAP3_B1 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
12009_BAP1_B1 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
CB9615_BAP9_E 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
NA114_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
UTI89_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
UM146_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
536_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
APEC01_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
IHE3034_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
EC958_BAP4_B2 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
11368_BAP3_B1 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
55989_BAP1_B1 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
SE11_BAP3_B1 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
UMNK88_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
BW2952_BAP5_A 1 CATTTTCGCCCCCA-GCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
Di12_BAP4_B2 1 CATTTTCGCCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
O42_BAP6_D1 1 CATTTTCGCCCCCAGCCATATCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
ABU83972_BAP4_B2 1 CATTTTCGCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
CFT073_BAP4_B2 1 CATTTTCGCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
Di14_BAP4_B2 1 CATTTTCGCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
SMS35_BAP8_F 1 CATTTTCGCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAA
ATCC9637_BAP3_B1 1 CATTTTCGCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGGTCATCAA
ATCC8739_BAP5_A 1 CATTTTCGCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGGTCATCAA
HS_BAP7_A 1 CATTTTCGCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGGTCATCAA
consensus 1 CATTTTCGcCcCCa GCCATtTCTACAACGTGAATTGTACCTGTCCgCAATGacCATCAA

```
E2348/69_BAP4_B2    60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
LF82a_BAP4_B2       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
LF82b_BAP4_B2       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
NRG857c_BAP4_B2     60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
ED1a_BAP4_B2        60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
UMNO26_BAP6_D1      60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
CE10_BAP8_F         60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
DH1_BAP5_A          60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
DH10B_BAP5_A        60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
MG1655_BAP5_A       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
H10407_BAP5_A       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
BL21                60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
IAI1_BAP3_B1        60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
11128_BAP3_B1       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
RM12579_BAP9_E      60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
SE15_BAP4_B2        60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
S88_BAP4_B2         60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
EDL933_BAP9_E       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
Sakai_BAP9_E        60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
TW14359_BAP9_E      60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
EC4115_BAP9_E       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
XuZhou21_BAP9_E     60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
ME8569_BAP5_A       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
W3110_BAP5_A        60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
P12b_BAP5_A         60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
REL606_BAP5_A       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
E23477A_BAP3_B1     60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
12009_BAP1_B1       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
CB9615_BAP9_E       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
NA114_BAP4_B2       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
UTI89_BAP4_B2       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
UM146_BAP4_B2       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
536_BAP4_B2         60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
APEC01_BAP4_B2      60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
IHE3034_BAP4_B2     60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
EC958_BAP4_B2       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
11368_BAP3_B1       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
55989_BAP1_B1       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
SE11_BAP3_B1        60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
UMNK88_BAP5_A       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
BW2952_BAP5_A       60 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
Di12_BAP4_B2        61 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
O42_BAP6_D1         61 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
ABU83972_BAP4_B2    61 CGGCATAAATAG-GACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
CFT073_BAP4_B2      61 CGGCATAAATAG-GACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
Di14_BAP4_B2        61 CGGCATAAATAG-GACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
SMS35_BAP8_F        61 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
ATCC9637_BAP3_B1    61 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
ATCC8739_BAP5_A     61 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
HS_BAP7_A           61 CGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
consensus           61 CGGCATAAATAGcGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATT
```

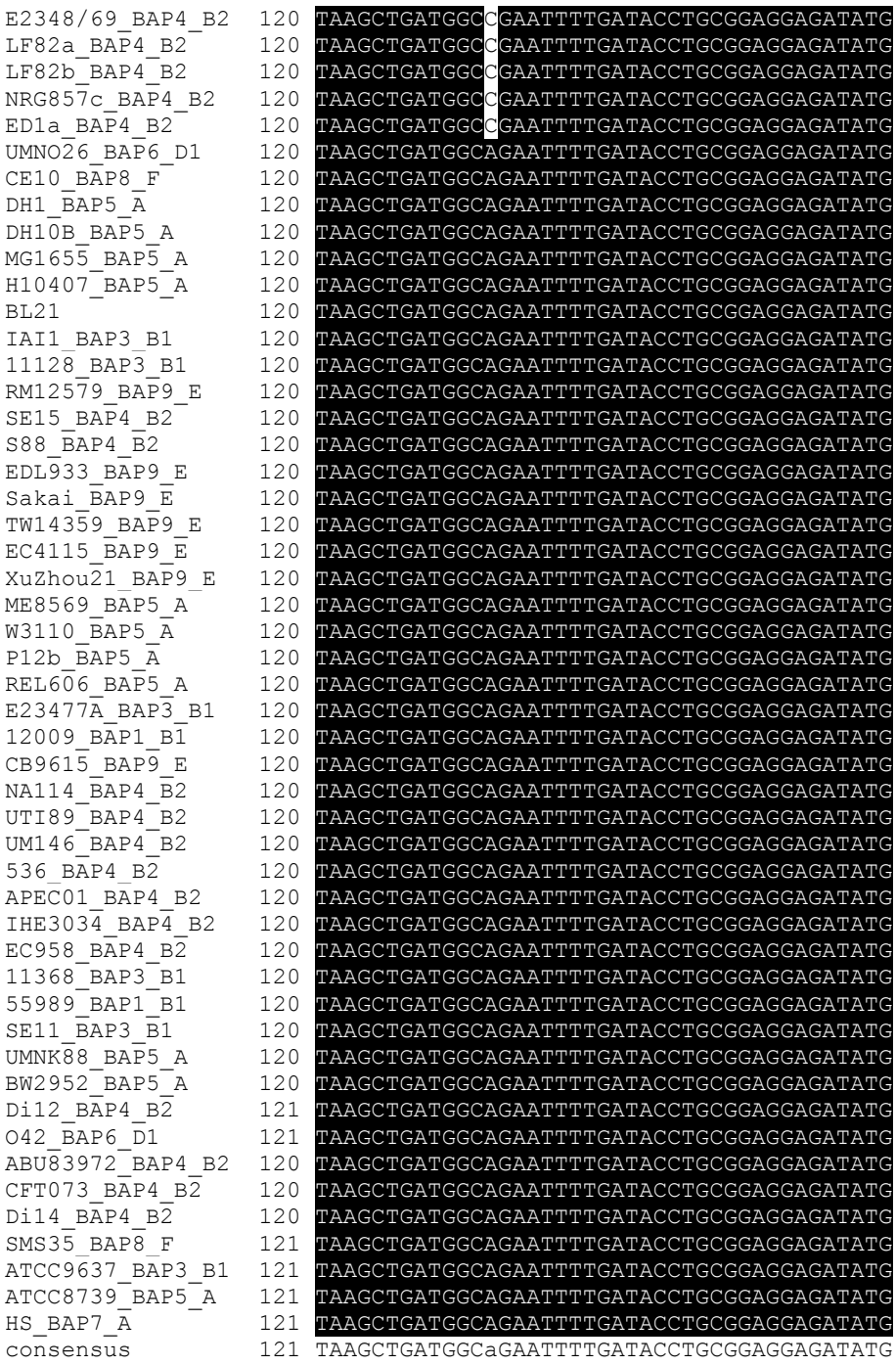


Figure. 8.3. Alignment of the *flgB* promoter of *E. coli* strains used in the work of McNally *et al.* (2013).

Chapter Nine: Bibliography

Abberton, C.L., Bereschenko, L., van der Wielen, P.W. and Smith, C.J. (2016) 'Survival, Biofilm Formation, and Growth Potential of Environmental and Enteric *Escherichia coli* Strains in Drinking Water Microcosms', *Appl Environ Microbiol*, 82(17), pp. 5320-31.

Agaisse, H. (2016) 'Molecular and Cellular Mechanisms of *Shigella flexneri* Dissemination', *Frontiers in Cellular and Infection Microbiology*, 6, p. 29.

Akiyama, Y. (2009) 'Quality control of cytoplasmic membrane proteins in *Escherichia coli*', *J Biochem*, 146(4), pp. 449-54.

Albanna, A. (2017) *Regulation of flagellar mediated motility in the species Salmonella enterica*. Ph.D. thesis. Newcastle University.

Aldick, T., Bielaszewska, M., Zhang, W., Brockmeyer, J., Schmidt, H., Friedrich, A.W., Kim, K.S., Schmidt, M.A. and Karch, H. (2007) 'Hemolysin from Shiga toxin-negative *Escherichia coli* O26 strains injures microvascular endothelium', *Microbes Infect*, 9(3), pp. 282-90.

Aldridge, C., Poonchareon, K., Saini, S., Ewen, T., Soloyva, A., Rao, C.V., Imada, K., Minamino, T. and Aldridge, P.D. (2010) 'The interaction dynamics of a negative feedback loop regulates flagellar number in *Salmonella enterica* serovar Typhimurium', *Mol Microbiol*, 78(6), pp. 1416-30.

Aldridge, P., Gnerer, J., Karlinsey, J.E. and Hughes, K.T. (2006a) 'Transcriptional and translational control of the *Salmonella fliC* gene', *J Bacteriol*, 188(12), pp. 4487-96.

Aldridge, P. and Hughes, K.T. (2001) 'How and when are substrates selected for type III secretion?', *Trends Microbiol*, 9(5), pp. 209-14.

Aldridge, P. and Hughes, K.T. (2002) 'Regulation of flagellar assembly', *Curr Opin Microbiol*, 5(2), pp. 160-5.

Aldridge, P., Karlinsey, J. and Hughes, K.T. (2003) 'The type III secretion chaperone FlgN regulates flagellar assembly via a negative feedback loop containing its chaperone substrates FlgK and FlgL', *Mol Microbiol*, 49(5), pp. 1333-45.

Aldridge, P., Karlinsey, J.E., Becker, E., Chevance, F.F.V. and Hughes, K.T. (2006b) 'Flk prevents premature secretion of the anti-σ factor FlgM into the periplasm', *Molecular Microbiology*, 60(3), pp. 630-643.

Aldridge, P.D., Karlinsey, J.E., Aldridge, C., Birchall, C., Thompson, D., Yagasaki, J. and Hughes, K.T. (2006c) 'The flagellar-specific transcription factor, σ(28), is the Type III secretion chaperone for the flagellar-specific anti-σ(28) factor FlgM', *Genes & Development*, 20(16), pp. 2315-2326.

Anderson, G.G., Palermo, J.J., Schilling, J.D., Roth, R., Heuser, J. and Hultgren, S.J. (2003) 'Intracellular bacterial biofilm-like pods in urinary tract infections', *Science*, 301(5629), pp. 105-7.

Ashida, H., Ogawa, M., Kim, M., Suzuki, S., Sanada, T., Punginelli, C., Mimuro, H. and Sasakawa, C. (2011) '*Shigella* deploy multiple countermeasures against host innate immune responses', *Curr Opin Microbiol*, 14(1), pp. 16-23.

Badea, L., Doughty, S., Nicholls, L., Sloan, J., Robins-Browne, R.M. and Hartland, E.L. (2003) 'Contribution of Efa1/LifA to the adherence of enteropathogenic *Escherichia coli* to epithelial cells', *Microb Pathog*, 34(5), pp. 205-15.

Barembruch, C. and Hengge, R. (2007) 'Cellular levels and activity of the flagellar sigma factor FliA of *Escherichia coli* are controlled by FlgM-modulated proteolysis', *Mol Microbiol*, 65(1), pp. 76-89.

Barker, C.S., Pruss, B.M. and Matsumura, P. (2004) 'Increased motility of *Escherichia coli* by insertion sequence element integration into the regulatory region of the *flhD* operon', *J Bacteriol*, 186(22), pp. 7529-37.

Barnich, N., Carvalho, F.A., Glasser, A.L., Darcha, C., Jantscheff, P., Allez, M., Peeters, H., Bommelaer, G., Desreumaux, P., Colombel, J.F. and Darfeuille-Michaud, A. (2007) 'CEACAM6 acts as a receptor for adherent-invasive *E. coli*, supporting ileal mucosa colonization in Crohn disease', *J Clin Invest*, 117(6), pp. 1566-74.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science*, 315(5819), pp. 1709-12.

Belaaouaj, A., Kim, K.S. and Shapiro, S.D. (2000) 'Degradation of outer membrane protein A in *Escherichia coli* killing by neutrophil elastase', *Science*, 289(5482), pp. 1185-8.

Belén, A., Pavón, I. and Maiden, M.C.J. (2009) 'Multilocus Sequence Typing', *Methods in molecular biology (Clifton, N.J.)*, 551, pp. 129-140.

Berg, R.D. (1998) '*Escherichia coli*, infection and immunity', in Roitt, I.M. and Delves, P.J. (eds.) *Encyclopedia of immunology*. 2nd edn. London: Academic Press, pp. 842-845.

Betis, F., Brest, P., Hofman, V., Guignot, J., Kansau, I., Rossi, B., Servin, A. and Hofman, P. (2003) 'Afa/Dr diffusely adhering *Escherichia coli* infection in T84 cell monolayers induces increased neutrophil transepithelial migration, which in turn promotes cytokine-dependent upregulation of decay-accelerating factor (CD55), the receptor for Afa/Dr adhesins', *Infect Immun*, 71(4), pp. 1774-83.

Bisercic, M., Feutrier, J.Y. and Reeves, P.R. (1991) 'Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic gnd locus', *J Bacteriol*, 173(12), pp. 3894-900.

Blank, K., Hensel, M. and Gerlach, R.G. (2011) 'Rapid and highly efficient method for scarless mutagenesis within the *Salmonella enterica* chromosome', *PLoS One*, 6(1), p. e15763.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) 'The complete genome sequence of *Escherichia coli* K-12', *Science*, 277(5331), pp. 1453-62.

Bokranz, W., Wang, X., Tschape, H. and Romling, U. (2005) 'Expression of cellulose and curli fimbriae by *Escherichia coli* isolated from the gastrointestinal tract', *J Med Microbiol*, 54(Pt 12), pp. 1171-82.

Boyd, E.F. and Hartl, D.L. (1998) 'Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution', *J Bacteriol*, 180(5), pp. 1159-65.

Brenner, D.J., Fanning, G.R., Skerman, F.J. and Falkow, S. (1972) 'Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms', *J Bacteriol*, 109(3), pp. 953-65.

Bringer, M.A., Barnich, N., Glasser, A.L., Bardot, O. and Darfeuille-Michaud, A. (2005) 'HtrA stress protein is involved in intramacrophagic replication of adherent and invasive *Escherichia coli* strain LF82 isolated from a patient with Crohn's disease', *Infect Immun*, 73(2), pp. 712-21.

Brodrick, H.J., Raven, K.E., Harrison, E.M., Blane, B., Reuter, S., Torok, M.E., Parkhill, J. and Peacock, S.J. (2016) 'Whole-genome sequencing reveals transmission of vancomycin-resistant *Enterococcus faecium* in a healthcare network', *Genome Med*, 8(1), p. 4.

Brombacher, E., Baratto, A., Dorel, C. and Landini, P. (2006) 'Gene expression regulation by the Curli activator CsgD protein: modulation of cellulose biosynthesis and control of negative determinants for microbial adhesion', *J Bacteriol*, 188(6), pp. 2027-37.

Brown, J., Faulds-Pain, A. and Aldridge, P. (2009) 'The coordination of flagellar gene expression and the flagellar assembly pathway', in Jarrell, K.F. (ed.) *Pili and flagella: Current research and future trends*. Norfolk: Caister Academic Press, pp. 99-120.

Brown, J.D., Saini, S., Aldridge, C., Herbert, J., Rao, C.V. and Aldridge, P.D. (2008) 'The rate of protein secretion dictates the temporal dynamics of flagellar gene expression', *Mol Microbiol*, 70(4), pp. 924-37.

Burkart, M., Toguchi, A. and Harshey, R.M. (1998) 'The chemotaxis system, but not chemotaxis, is essential for swarming motility in *Escherichia coli*', *Proceedings of the National Academy of Sciences*, 95(5), pp. 2568-2573.

Cai, S.J. and Inouye, M. (2002) 'EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*', *J Biol Chem*, 277(27), pp. 24155-61.

Carlos, C., Pires, M.M., Stoppe, N.C., Hachich, E.M., Sato, M.I., Gomes, T.A., Amaral, L.A. and Ottoboni, L.M. (2010) '*Escherichia coli* phylogenetic group determination and its application in the identification of the major animal source of fecal contamination', *BMC Microbiol*, 10, p. 161.

Caza, M. and Kronstad, J.W. (2013) 'Shared and distinct mechanisms of iron acquisition by bacterial and fungal pathogens of humans', *Front Cell Infect Microbiol*, 3, p. 80.

Cerca, N., Maira-Litran, T., Jefferson, K.K., Grout, M., Goldmann, D.A. and Pier, G.B. (2007) 'Protection against *Escherichia coli* infection by antibody to the

Staphylococcus aureus poly-N-acetylglucosamine surface polysaccharide', *Proc Natl Acad Sci U S A*, 104(18), pp. 7528-33.

Chadsey, M.S., Karlinsey, J.E. and Hughes, K.T. (1998) 'The flagellar anti-sigma factor FlgM actively dissociates *Salmonella typhimurium* sigma28 RNA polymerase holoenzyme', *Genes Dev*, 12(19), pp. 3123-36.

Chang, W.S., Li, X. and Halverson, L.J. (2009) 'Influence of water limitation on endogenous oxidative stress and cell death within unsaturated *Pseudomonas putida* biofilms', *Environ Microbiol*, 11(6), pp. 1482-92.

Chassaing, B., Rolhion, N., de Vallee, A., Salim, S.Y., Prorok-Hamon, M., Neut, C., Campbell, B.J., Soderholm, J.D., Hugot, J.P., Colombel, J.F. and Darfeuille-Michaud, A. (2011) 'Crohn disease--associated adherent-invasive *E. coli* bacteria target mouse and human Peyer's patches via long polar fimbriae', *J Clin Invest*, 121(3), pp. 966-75.

Chaudhuri, R.R. and Henderson, I.R. (2012) 'The evolution of the *Escherichia coli* phylogeny', *Infect Genet Evol*, 12(2), pp. 214-26.

Chen, S.L., Hung, C.-S., Xu, J., Reigstad, C.S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R.R., Ozersky, P., Armstrong, J.R., Fulton, R.S., Latreille, J.P., Spieth, J., Hooton, T.M., Mardis, E.R., Hultgren, S.J. and Gordon, J.I. (2006) 'Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: A comparative genomics approach', *Proceedings of the National Academy of Sciences*, 103(15), pp. 5977-5982.

Chevance, F.F. and Hughes, K.T. (2008) 'Coordinating assembly of a bacterial macromolecular machine', *Nat Rev Microbiol*, 6(6), pp. 455-65.

Chilcott, G.S. and Hughes, K.T. (2000) 'Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar typhimurium and *Escherichia coli*', *Microbiol Mol Biol Rev*, 64(4), pp. 694-708.

Cirl, C., Wieser, A., Yadav, M., Duerr, S., Schubert, S., Fischer, H., Stappert, D., Wantia, N., Rodriguez, N., Wagner, H., Svanborg, C. and Miethke, T. (2008) 'Subversion of Toll-like receptor signaling by a unique family of bacterial Toll/interleukin-1 receptor domain-containing proteins', *Nat Med*, 14(4), pp. 399-406.

Claret, L. and Hughes, C. (2000a) 'Functions of the subunits in the $FlhD_2C_2$ transcriptional master regulator of bacterial flagellum biogenesis and swarming', *Journal of Molecular Biology*, 303(4), pp. 467-478.

Claret, L. and Hughes, C. (2000b) 'Rapid turnover of FlhD and FlhC, the flagellar regulon transcriptional activator proteins, during *Proteus* swarming', *J Bacteriol*, 182(3), pp. 833-6.

Claret, L. and Hughes, C. (2002) 'Interaction of the Atypical Prokaryotic Transcription Activator $FlhD_2C_2$ with Early Promoters of the Flagellar Gene Hierarchy', *Journal of Molecular Biology*, 321(2), pp. 185-199.

Clegg, S. and Hughes, K.T. (2002) 'FimZ is a molecular link between sticking and swimming in *Salmonella enterica* serovar Typhimurium', *J Bacteriol*, 184(4), pp. 1209-13.

Clermont, O., Bonacorsi, S. and Bingen, E. (2000) 'Rapid and simple determination of the *Escherichia coli* phylogenetic group', *Appl Environ Microbiol*, 66(10), pp. 4555-8.

Clermont, O., Christenson, J.K., Denamur, E. and Gordon, D.M. (2013) 'The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups', *Environ Microbiol Rep*, 5(1), pp. 58-65.

Clermont, O., Gordon, D. and Denamur, E. (2015) 'Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes', *Microbiology*, 161(Pt 5), pp. 980-8.

Conway, T., Creecy, J.P., Maddox, S.M., Grissom, J.E., Conkle, T.L., Shadid, T.M., Teramoto, J., San Miguel, P., Shimada, T., Ishihama, A., Mori, H. and Wanner, B.L. (2014) 'Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing', *MBio*, 5(4), pp. e01442-14.

Costerton, J.W., Stewart, P.S. and Greenberg, E.P. (1999) 'Bacterial biofilms: a common cause of persistent infections', *Science*, 284(5418), pp. 1318-22.

Cote, J.P. and Mourez, M. (2011) 'Structure-function analysis of the TibA self-associating autotransporter reveals a modular organization', *Infect Immun*, 79(5), pp. 1826-32.

Craney, A., Hohenauer, T., Xu, Y., Navani, N.K., Li, Y. and Nodwell, J. (2007) 'A synthetic *luxCDABE* gene cluster optimized for expression in high-GC bacteria', *Nucleic Acids Research*, 35(6), pp. e46-e46.

Crossman, L.C., Chaudhuri, R.R., Beatson, S.A., Wells, T.J., Desvaux, M., Cunningham, A.F., Petty, N.K., Mahon, V., Brinkley, C., Hobman, J.L., Savarino, S.J., Turner, S.M., Pallen, M.J., Penn, C.W., Parkhill, J., Turner, A.K., Johnson, T.J., Thomson, N.R., Smith, S.G. and Henderson, I.R. (2010) 'A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407', *J Bacteriol*, 192(21), pp. 5822-31.

Croxen, M.A. and Finlay, B.B. (2010) 'Molecular mechanisms of *Escherichia coli* pathogenicity', *Nat Rev Microbiol*, 8(1), pp. 26-38.

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M. and Finlay, B.B. (2013) 'Recent Advances in Understanding Enteric Pathogenic *Escherichia coli*', *Clinical Microbiology Reviews*, 26(4), pp. 822-880.

Darnton, N.C., Turner, L., Rojevsky, S. and Berg, H.C. (2010) 'Dynamics of Bacterial Swarming', *Biophysical Journal*, 98(10), pp. 2082-2090.

Datsenko, K.A. and Wanner, B.L. (2000) 'One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products', *Proc Natl Acad Sci U S A*, 97(12), pp. 6640-5.

Day, R.N. and Davidson, M.W. (2009) 'The fluorescent protein palette: tools for cellular imaging', *Chemical Society reviews*, 38(10), pp. 2887-2921.

De Lay, N. and Gottesman, S. (2012) 'A complex network of small non-coding RNAs regulate motility in *Escherichia coli*', *Mol Microbiol*, 86(3), pp. 524-38.

Dean, P. and Kenny, B. (2009) 'The effector repertoire of enteropathogenic *E. coli*: ganging up on the host cell', *Current Opinion in Microbiology*, 12(1-3), pp. 101-109.

Delalez, N.J., Wadhams, G.H., Rosser, G., Xue, Q., Brown, M.T., Dobbie, I.M., Berry, R.M., Leake, M.C. and Armitage, J.P. (2010) 'Signal-dependent turnover of the bacterial flagellar switch protein FliM', *Proc Natl Acad Sci U S A*, 107(25), pp. 11347-51.

DePas, W.H., Syed, A.K., Sifuentes, M., Lee, J.S., Warshaw, D., Saggar, V., Csankovszki, G., Boles, B.R. and Chapman, M.R. (2014) 'Biofilm formation protects *Escherichia coli* against killing by *Caenorhabditis elegans* and *Myxococcus xanthus*', *Appl Environ Microbiol*, 80(22), pp. 7079-87.

DeVinney, R., Stein, M., Reinscheid, D., Abe, A., Ruschkowski, S. and Finlay, B.B. (1999) 'Enterohemorrhagic *Escherichia coli* O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated', *Infect Immun*, 67(5), pp. 2389-98.

Dhakal, B.K. and Mulvey, M.A. (2012) 'The UPEC Pore Forming Toxin α-Hemolysin Triggers Proteolysis of Host Proteins to Disrupt Cell Adhesion, Inflammatory and Survival Pathways', *Cell host & microbe*, 11(1), pp. 58-69.

Diaz, E., Ferrandez, A., Prieto, M.A. and Garcia, J.L. (2001) 'Biodegradation of aromatic compounds by *Escherichia coli*', *Microbiol Mol Biol Rev*, 65(4), pp. 523-69, table of contents.

Dobrindt, U., Hochhut, B., Hentschel, U. and Hacker, J. (2004) 'Genomic islands in pathogenic and environmental microorganisms', *Nat Rev Microbiol*, 2(5), pp. 414-24.

Drage, L. (2016) *A longitudinal study of long-term Escherichia coli colonisation of the elderly bladder*. Ph.D. thesis. Newcastle University.

Dudley, E.G., Abe, C., Ghigo, J.-M., Latour-Lambert, P., Hormazabal, J.C. and Nataro, J.P. (2006) 'An IncI1 Plasmid Contributes to the Adherence of the Atypical Enteroaggregative *Escherichia coli* Strain C1096 to Cultured Cells and Abiotic Surfaces', *Infection and Immunity*, 74(4), pp. 2102-2114.

Dunlap, P. (2014) 'Biochemistry and genetics of bacterial bioluminescence', *Adv Biochem Eng Biotechnol*, 144, pp. 37-64.

Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, Burland, V., Mau, B., Petrosino, J.F., Qin, X., Muzny, D.M., Ayele, M., Gibbs, R.A., Csorgo, B., Posfai, G., Weinstock, G.M. and Blattner, F.R. (2008) 'The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse', *J Bacteriol*, 190(7), pp. 2597-606.

Ebel, W., Skinner, M., Dierksen, K., Scott, J. and Trempy, J. (1999) 'A Conserved Domain in *Escherichia coli* Lon Protease Is Involved in Substrate Discriminator Activity', *J Bacteriol*, 181(7), pp. 2236-2243.

Escobar-Paramo, P., Clermont, O., Blanc-Potard, A.B., Bui, H., Le Bouguenec, C. and Denamur, E. (2004) 'A specific genetic background is required for acquisition

and expression of virulence factors in *Escherichia coli*', *Mol Biol Evol*, 21(6), pp. 1085-94.

Estrada-Garcia, T. and Navarro-Garcia, F. (2012) 'Enteroaggregative *Escherichia coli* pathotype: a genetically heterogeneous emerging foodborne enteropathogen', *FEMS Immunol Med Microbiol*, 66(3), pp. 281-98.

Fahrner, K.A. and Berg, H.C. (2015) 'Mutations That Stimulate *flhDC* Expression in *Escherichia coli* K-12', *J Bacteriol*, 197(19), pp. 3087-96.

Farfan, M.J., Toro, C.S., Barry, E.M. and Nataro, J.P. (2011) '*Shigella* enterotoxin-2 is a type III effector that participates in *Shigella*-induced interleukin 8 secretion by epithelial cells', *FEMS Immunol Med Microbiol*, 61(3), pp. 332-9.

Faulds-Pain, A., Birchall, C., Aldridge, C., Smith, W.D., Grimaldi, G., Nakamura, S., Miyata, T., Gray, J., Li, G., Tang, J.X., Namba, K., Minamino, T. and Aldridge, P.D. (2011) 'Flagellin redundancy in *Caulobacter crescentus* and its implications for flagellar filament assembly', *J Bacteriol*, 193(11), pp. 2695-707.

Ferens, W.A. and Hovde, C.J. (2011) '*Escherichia coli* O157:H7: animal reservoir and sources of human infection', *Foodborne Pathog Dis*, 8(4), pp. 465-87.

Ferrieres, L., Hancock, V. and Klemm, P. (2007) 'Biofilm exclusion of uropathogenic bacteria by selected asymptomatic bacteriuria *Escherichia coli* strains', *Microbiology*, 153(Pt 6), pp. 1711-9.

Field, K.G. and Samadpour, M. (2007) 'Fecal source tracking, the indicator paradigm, and managing water quality', *Water Res*, 41(16), pp. 3517-38.

Flores-Mireles, A.L., Walker, J.N., Caparon, M. and Hultgren, S.J. (2015) 'Urinary tract infections: epidemiology, mechanisms of infection and treatment options', *Nat Rev Microbiol*, 13(5), pp. 269-84.

Foxman, B. (2010) 'The epidemiology of urinary tract infection', *Nat Rev Urol*, 7(12), pp. 653-60.

Francez-Charlot, A., Laugel, B., Van Gemert, A., Dubarry, N., Wiorowski, F., Castanie-Cornet, M.P., Gutierrez, C. and Cam, K. (2003) 'RcsCDB His-Asp phosphorelay system negatively regulates the *flhDC* operon in *Escherichia coli*', *Mol Microbiol*, 49(3), pp. 823-32.

Francis, N.R., Sosinsky, G.E., Thomas, D. and DeRosier, D.J. (1994) 'Isolation, characterization and structure of bacterial flagellar motors containing the switch complex', *J Mol Biol*, 235(4), pp. 1261-70.

Fraser, G.M., Hirano, T., Ferris, H.U., Devgan, L.L., Kihara, M. and Macnab, R.M. (2003) 'Substrate specificity of type III flagellar protein export in *Salmonella* is controlled by subdomain interactions in FlhB', *Mol Microbiol*, 48(4), pp. 1043-57.

Garcia, E.C., Brumbaugh, A.R. and Mobley, H.L. (2011) 'Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection', *Infect Immun*, 79(3), pp. 1225-35.

Garcia, T.A., Ventura, C.L., Smith, M.A., Merrell, D.S. and O'Brien, A.D. (2013) 'Cytotoxic necrotizing factor 1 and hemolysin from uropathogenic *Escherichia coli* elicit different host responses in the murine bladder', *Infect Immun*, 81(1), pp. 99-109.

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. and Kado, C.I. (1985) 'Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria', *J Bacteriol*, 164(2), pp. 918-21.

Gerlach, R.G., Jäckel, D., Hölzer, S.U. and Hensel, M. (2009) 'Rapid Oligonucleotide-Based Recombineering of the Chromosome of *Salmonella enterica*', *Applied and Environmental Microbiology*, 75(6), pp. 1575-1580.

Ghosh, R., Kumar, A. and Vaziri, A. (2014) 'Type-IV Pilus Deformation Can Explain Retraction Behavior', *PLoS ONE*, 9(12), p. e114613.

Gobert, A.P., Vareille, M., Glasser, A.L., Hindre, T., de Sablet, T. and Martin, C. (2007) 'Shiga toxin produced by enterohemorrhagic *Escherichia coli* inhibits PI3K/NF-kappaB signaling pathway in globotriaosylceramide-3-negative human intestinal epithelial cells', *J Immunol*, 178(12), pp. 8168-74.

Goldenfeld, N. and Woese, C. (2007) 'Biology's next revolution', *Nature*, 445(7126), p. 369.

Goller, C., Wang, X., Itoh, Y. and Romeo, T. (2006) 'The cation-responsive protein NhaR of *Escherichia coli* activates *pgaABCD* transcription, required for production of the biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine', *J Bacteriol*, 188(23), pp. 8022-32.

Goodier, R.I. and Ahmer, B.M.M. (2001) 'SirA Orthologs Affect both Motility and Virulence', *Journal of Bacteriology*, 183(7), pp. 2249-2258.

Gottesman, S. (1999) 'Regulation by proteolysis: developmental switches', *Curr Opin Microbiol*, 2(2), pp. 142-7.

Guignot, J., Chaplais, C., Coconnier-Polter, M.H. and Servin, A.L. (2007) 'The secreted autotransporter toxin, Sat, functions as a virulence factor in Afa/Dr diffusely adhering *Escherichia coli* by promoting lesions in tight junction of polarized epithelial cells', *Cell Microbiol*, 9(1), pp. 204-21.

Gunther, N.W.t., Snyder, J.A., Lockatell, V., Blomfield, I., Johnson, D.E. and Mobley, H.L. (2002) 'Assessment of virulence of uropathogenic *Escherichia coli* type 1 fimbrial mutants in which the invertible element is phase-locked on or off', *Infect Immun*, 70(7), pp. 3344-54.

Gustavsson, N., Diez, A. and Nystrom, T. (2002) 'The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage', *Mol Microbiol*, 43(1), pp. 107-17.

Hacker, J., Hentschel, U. and Dobrindt, U. (2003) 'Prokaryotic chromosomes and disease', *Science*, 301(5634), pp. 790-3.

Haddock, S.H., Moline, M.A. and Case, J.F. (2010) 'Bioluminescence in the sea', *Ann Rev Mar Sci*, 2, pp. 443-93.

Hannan, T.J., Totsika, M., Mansfield, K.J., Moore, K.H., Schembri, M.A. and Hultgren, S.J. (2012) 'Host-pathogen checkpoints and population bottlenecks in persistent and intracellular uropathogenic *Escherichia coli* bladder infection', *FEMS Microbiol Rev*, 36(3), pp. 616-48.

Harshey, R.M. (2003) 'Bacterial motility on a surface: many ways to a common goal', *Annu Rev Microbiol*, 57, pp. 249-73.

Hastings, J.W., Potrikus, C.J., Gupta, S.C., Kurfurst, M. and Makemson, J.C. (1985) 'Biochemistry and physiology of bioluminescent bacteria', *Adv Microb Physiol*, 26, pp. 235-91.

He, S., Hickman, A.B., Varani, A.M., Siguier, P., Chandler, M., Dekker, J.P. and Dyda, F. (2015) 'Insertion Sequence IS26 Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria by Replicative Transposition', *mBio*, 6(3), pp. e00762-15.

Heermann, R., Weber, A., Mayer, B., Ott, M., Hauser, E., Gabriel, G., Pirch, T. and Jung, K. (2009) 'The universal stress protein UspC scaffolds the KdpD/KdpE signaling cascade of *Escherichia coli* under salt stress', *J Mol Biol*, 386(1), pp. 134-48.

Henderson, I.R., Czeczulin, J., Eslava, C., Noriega, F. and Nataro, J.P. (1999) 'Characterization of Pic, a Secreted Protease of *Shigella flexneri* and Enteroaggregative *Escherichia coli*', *Infection and Immunity*, 67(11), pp. 5587-5596.

Hengge, R. (2010) 'Role of Cyclic Di-GMP in the Regulatory Networks of *Escherichia coli*', in *The Second Messenger Cyclic Di-GMP*. American Society of Microbiology.

Henrichsen, J. (1972) 'Bacterial surface translocation: a survey and a classification', *Bacteriological Reviews*, 36(4), pp. 478-503.

Herold, S., Paton, J.C. and Paton, A.W. (2009) 'Sab, a novel autotransporter of locus of enterocyte effacement-negative shiga-toxigenic *Escherichia coli* O113:H21, contributes to adherence and biofilm formation', *Infect Immun*, 77(8), pp. 3234-43.

Herzer, P.J., Inouye, S., Inouye, M. and Whittam, T.S. (1990) 'Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*', *J Bacteriol*, 172(11), pp. 6175-81.

Hirano, T., Minamino, T. and Macnab, R.M. (2001) 'The role in flagellar rod assembly of the N-terminal domain of *Salmonella* FlgJ, a flagellum-specific muramidase', *J Mol Biol*, 312(2), pp. 359-69.

Hirano, T., Yamaguchi, S., Oosawa, K. and Aizawa, S. (1994) 'Roles of FliK and FlhB in determination of flagellar hook length in *Salmonella typhimurium*', *J Bacteriol*, 176(17), pp. 5439-49.

Ho Sui, S.J., Fedynak, A., Hsiao, W.W., Langille, M.G. and Brinkman, F.S. (2009) 'The association of virulence factors with genomic islands', *PLoS One*, 4(12), p. e8094.

Homma, M. and Iino, T. (1985) 'Excretion of unassembled hook-associated proteins by *Salmonella typhimurium*', *J Bacteriol*, 164(3), pp. 1370-2.

Homma, M., Kutsukake, K., Hasebe, M., Iino, T. and Macnab, R.M. (1990) 'FlgB, FlgC, FlgF and FlgG. A family of structurally related proteins in the flagellar basal body of *Salmonella typhimurium*', *J Mol Biol*, 211(2), pp. 465-77.

Homma, M., Kutsukake, K., Iino, T. and Yamaguchi, S. (1984) 'Hook-associated proteins essential for flagellar filament formation in *Salmonella typhimurium*', *Journal of Bacteriology*, 157(1), pp. 100-108.

Hughes, K.T., Gillen, K.L., Semon, M.J. and Karlinsey, J.E. (1993) 'Sensing structural intermediates in bacterial flagellar assembly by export of a negative regulator', *Science*, 262(5137), pp. 1277-80.

Humayun, M.Z., Zhang, Z., Butcher, A.M., Moshayedi, A. and Saier, M.H., Jr. (2017) 'Hopping into a hot seat: Role of DNA structural features on IS5-mediated gene activation and inactivation under stress', *PLOS ONE*, 12(6), p. e0180156.

Hyland, R.M., Sun, J., Griener, T.P., Mulvey, G.L., Klassen, J.S., Donnenberg, M.S. and Armstrong, G.D. (2008) 'The bundlin pilin protein of enteropathogenic *Escherichia coli* is an N-acetyllactosamine-specific lectin', *Cellular microbiology*, 10(1), pp. 10.1111/j.1462-5822.2007.01028.x.

Iguchi, A., Thomson, N.R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I.R., Harris, D., Asadulghani, M., Kurokawa, K., Dean, P., Kenny, B., Quail, M.A., Thurston, S., Dougan, G., Hayashi, T., Parkhill, J. and Frankel, G. (2009) 'Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69', *J Bacteriol*, 191(1), pp. 347-54.

Imada, K., Minamino, T., Kinoshita, M., Furukawa, Y. and Namba, K. (2010) 'Structural insight into the regulatory mechanisms of interactions of the flagellar type III chaperone FliT with its binding partners', *Proc Natl Acad Sci U S A*, 107(19), pp. 8812-7.

Ishihama, A. (1993) 'Protein-protein communication within the transcription apparatus', *J Bacteriol*, 175(9), pp. 2483-9.

Ito, K. and Akiyama, Y. (2005) 'Cellular functions, mechanism of action, and regulation of FtsH protease', *Annu Rev Microbiol*, 59, pp. 211-31.

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

Jaureguy, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G., Carbonnelle, E., Lortholary, O., Clermont, O., Denamur, E., Picard, B., Nassif, X. and Brisse, S. (2008) 'Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains', *BMC Genomics*, 9, p. 560.

Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J. and Yang, S. (2015) 'Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system', *Appl Environ Microbiol*, 81(7), pp. 2506-14.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*, 337(6096), pp. 816-21.

Jonas, K., Edwards, A.N., Simm, R., Romeo, T., Romling, U. and Melefors, O. (2008) 'The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins', *Mol Microbiol*, 70(1), pp. 236-57.

Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W. and Crook, D.W. (2009) 'Genomic islands: tools of bacterial horizontal gene transfer and evolution', *FEMS Microbiol Rev*, 33(2), pp. 376-93.

Justice, S.S., Hung, C., Theriot, J.A., Fletcher, D.A., Anderson, G.G., Footer, M.J. and Hultgren, S.J. (2004) 'Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis', *Proc Natl Acad Sci U S A*, 101(5), pp. 1333-8.

Kader, A., Simm, R., Gerstel, U., Morr, M. and Romling, U. (2006) 'Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of *Salmonella enterica* serovar Typhimurium', *Mol Microbiol*, 60(3), pp. 602-16.

Kaper, J.B., Nataro, J.P. and Mobley, H.L. (2004) 'Pathogenic *Escherichia coli*', *Nat Rev Microbiol*, 2(2), pp. 123-40.

Karlinsey, J.E. and Hughes, K.T. (2006) 'Genetic Transplantation: *Salmonella enterica* Serovar Typhimurium as a Host To Study Sigma Factor and Anti-Sigma Factor Interactions in Genetically Intractable Systems', *Journal of Bacteriology*, 188(1), pp. 103-114.

Karlinsey, J.E., Lonner, J., Brown, K.L. and Hughes, K.T. (2000) 'Translation/secretion coupling by type III secretion systems', *Cell*, 102(4), pp. 487-97.

Katayama, E., Shiraishi, T., Oosawa, K., Baba, N. and Aizawa, S. (1996) 'Geometry of the flagellar motor in the cytoplasmic membrane of *Salmonella typhimurium* as determined by stereo-photogrammetry of quick-freeze deep-etch replica images', *J Mol Biol*, 255(3), pp. 458-75.

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D.J., Frey, E.A. and Finlay, B.B. (1997) 'Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells', *Cell*, 91(4), pp. 511-20.

Khan, N.A., Kim, Y., Shin, S. and Kim, K.S. (2007) 'FimH-mediated *Escherichia coli* K1 invasion of human brain microvascular endothelial cells', *Cell Microbiol*, 9(1), pp. 169-78.

Kim, K.S. (2008) 'Mechanisms of microbial traversal of the blood-brain barrier', *Nat Rev Microbiol*, 6(8), pp. 625-34.

Kohler, C.D. and Dobrindt, U. (2011) 'What defines extraintestinal pathogenic *Escherichia coli*?', *Int J Med Microbiol*, 301(8), pp. 642-7.

Koirala, S., Mears, P., Sim, M., Golding, I., Chemla, Y.R., Aldridge, P.D. and Rao, C.V. (2014) 'A Nutrient-Tunable Bistable Switch Controls Motility in *Salmonella enterica* Serovar Typhimurium', *mBio*, 5(5), pp. e01611-14.

Kolenbrander, P.E., Andersen, R.N., Blehert, D.S., Egland, P.G., Foster, J.S. and Palmer, R.J., Jr. (2002) 'Communication among oral bacteria', *Microbiol Mol Biol Rev*, 66(3), pp. 486-505, table of contents.

Kvint, K., Nachin, L., Diez, A. and Nyström, T. (2003) 'The bacterial universal stress protein: function and regulation', *Current Opinion in Microbiology*, 6(2), pp. 140-145.

Lahiri, A., Das, P. and Chakravortty, D. (2008) 'Engagement of TLR signaling as adjuvant: towards smarter vaccine and beyond', *Vaccine*, 26(52), pp. 6777-83.

Lane, M.C., Alteri, C.J., Smith, S.N. and Mobley, H.L. (2007) 'Expression of flagella is coincident with uropathogenic *Escherichia coli* ascension to the upper urinary tract', *Proc Natl Acad Sci U S A*, 104(42), pp. 16669-74.

Lane, M.C., Lockatell, V., Monterosso, G., Lamphier, D., Weinert, J., Hebel, J.R., Johnson, D.E. and Mobley, H.L. (2005) 'Role of motility in the colonization of uropathogenic *Escherichia coli* in the urinary tract', *Infect Immun*, 73(11), pp. 7644-56.

Le Bouguenec, C. and Servin, A.L. (2006) 'Diffusely adherent *Escherichia coli* strains expressing Afa/Dr adhesins (Afa/Dr DAEC): hitherto unrecognized pathogens', *FEMS Microbiol Lett*, 256(2), pp. 185-94.

Ledford, H. (2015) 'CRISPR, the disruptor', *Nature*, 522(7554), pp. 20-4.

Lehnen, D., Blumer, C., Polen, T., Wackwitz, B., Wendisch, V.F. and Unden, G. (2002) 'LrhA as a new transcriptional key regulator of flagella, motility and chemotaxis genes in *Escherichia coli*', *Mol Microbiol*, 45(2), pp. 521-32.

Lehti, T.A., Bauchart, P., Dobrindt, U., Korhonen, T.K. and Westerlund-Wikstrom, B. (2012) 'The fimbriae activator MatA switches off motility in *Escherichia coli* by repression of the flagellar master operon *flhDC*', *Microbiology*, 158(Pt 6), pp. 1444-55.

Leimbach, A., Hacker, J. and Dobrindt, U. (2013) '*E. coli* as an all-rounder: the thin line between commensalism and pathogenicity', *Curr Top Microbiol Immunol*, 358, pp. 3-32.

Lemke, J.J., Durfee, T. and Gourse, R.L. (2009) 'DksA and ppGpp directly regulate transcription of the *Escherichia coli* flagellar cascade', *Mol Microbiol*, 74(6), pp. 1368-79.

Liu, X., Fujita, N., Ishihama, A. and Matsumura, P. (1995) 'The C-terminal region of the alpha subunit of *Escherichia coli* RNA polymerase is required for transcriptional activation of the flagellar level II operons by the FlhD/FlhC complex', *J Bacteriol*, 177(17), pp. 5186-8.

Liu, X. and Matsumura, P. (1994) 'The FlhD/FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons', *J Bacteriol*, 176(23), pp. 7345-51.

Livermore, D.M. and Woodford, N. (2006) 'The beta-lactamase threat in Enterobacteriaceae, *Pseudomonas* and *Acinetobacter*', *Trends Microbiol*, 14(9), pp. 413-20.

Lloyd, S.A., Tang, H., Wang, X., Billings, S. and Blair, D.F. (1996) 'Torque generation in the flagellar motor of *Escherichia coli*: evidence of a direct role for FliG but not for FliM or FliN', *J Bacteriol*, 178(1), pp. 223-31.

Lorenz, M.G. and Wackernagel, W. (1994) 'Bacterial gene transfer by natural genetic transformation in the environment', *Microbiol Rev*, 58(3), pp. 563-602.

Lukjancenko, O., Wassenaar, T.M. and Ussery, D.W. (2010) 'Comparison of 61 sequenced *Escherichia coli* genomes', *Microb Ecol*, 60(4), pp. 708-20.

Ma, J., Sun, M., Bao, Y., Pan, Z., Zhang, W., Lu, C. and Yao, H. (2013) 'Genetic diversity and features analysis of type VI secretion systems loci in avian pathogenic *Escherichia coli* by wide genomic scanning', *Infection, Genetics and Evolution*, 20, pp. 454-464.

Macnab, R.M. (2003) 'How bacteria assemble flagella', *Annu Rev Microbiol*, 57, pp. 77-100.

Macnab, R.M. (2004) 'Type III flagellar protein export and flagellar assembly', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1694(1), pp. 207-217.

Mahillon, J. and Chandler, M. (1998) 'Insertion Sequences', *Microbiology and Molecular Biology Reviews*, 62(3), pp. 725-774.

Maiden, M.C. (2006) 'Multilocus sequence typing of bacteria', *Annu Rev Microbiol*, 60, pp. 561-88.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. and Spratt, B.G. (1998) 'Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms', *Proc Natl Acad Sci U S A*, 95(6), pp. 3140-5.

Marques, S.M. and Esteves da Silva, J.C. (2009) 'Firefly bioluminescence: a mechanistic approach of luciferase catalyzed reactions', *IUBMB Life*, 61(1), pp. 6-17.

Martinez, J.J. and Hultgren, S.J. (2002) 'Requirement of Rho-family GTPases in the invasion of Type 1-piliated uropathogenic *Escherichia coli*', *Cell Microbiol*, 4(1), pp. 19-28.

Mattick, J.S. (2002) 'Type IV Pili and Twitching Motility', *Annual Review of Microbiology*, 56(1), pp. 289-314.

Maurelli, A.T., Fernández, R.E., Bloch, C.A., Rode, C.K. and Fasano, A. (1998) '"Black holes" and bacterial pathogenicity: A large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America*, 95(7), pp. 3943-3948.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S. and Kaper, J.B. (1995) 'A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens', *Proc Natl Acad Sci U S A*, 92(5), pp. 1664-8.

McNally, A., Cheng, L., Harris, S.R. and Corander, J. (2013) 'The evolutionary path to extraintestinal pathogenic, drug-resistant Escherichia coli is marked by drastic reduction in detectable recombination within the core genome', *Genome Biol Evol*, 5(4), pp. 699-710.

Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R. (2005) 'The microbial pan-genome', *Curr Opin Genet Dev*, 15(6), pp. 589-94.

Meighen, E.A. (1991) 'Molecular biology of bacterial bioluminescence', *Microbiol Rev*, 55(1), pp. 123-42.

Mercado, E.H., Ochoa, T.J., Ecker, L., Cabello, M., Durand, D., Barletta, F., Molina, M., Gil, A.I., Huicho, L., Lanata, C.F. and Cleary, T.G. (2011) 'Fecal Leukocytes in Children Infected with Diarrheagenic *Escherichia coli*', *Journal of Clinical Microbiology*, 49(4), pp. 1376-1381.

Mignot, T. (2007) 'The elusive engine in *Myxococcus xanthus* gliding motility', *Cell Mol Life Sci*, 64(21), pp. 2733-45.

Milkman, R. (1973) 'Electrophoretic variation in *Escherichia coli* from natural sources', *Science*, 182(4116), pp. 1024-6.

Minamino, T., Chu, R., Yamaguchi, S. and Macnab, R.M. (2000a) 'Role of FliJ in flagellar protein export in *Salmonella*', *J Bacteriol*, 182(15), pp. 4207-15.

Minamino, T., Gonzalez-Pedrajo, B., Yamaguchi, K., Aizawa, S.I. and Macnab, R.M. (1999) 'FliK, the protein responsible for flagellar hook length control in *Salmonella*, is exported during hook assembly', *Mol Microbiol*, 34(2), pp. 295-304.

Minamino, T., Imada, K. and Namba, K. (2008) 'Mechanisms of type III protein export for bacterial flagellar assembly', *Mol Biosyst*, 4(11), pp. 1105-15.

Minamino, T. and Macnab, R.M. (1999) 'Components of the *Salmonella* flagellar export apparatus and classification of export substrates', *J Bacteriol*, 181(5), pp. 1388-94.

Minamino, T. and Macnab, R.M. (2000) 'FliH, a soluble component of the type III flagellar export apparatus of *Salmonella*, forms a complex with FliI and inhibits its ATPase activity', *Mol Microbiol*, 37(6), pp. 1494-503.

Minamino, T., Yamaguchi, S. and Macnab, R.M. (2000b) 'Interaction between FliE and FlgB, a Proximal Rod Component of the Flagellar Basal Body of *Salmonella*', *Journal of Bacteriology*, 182(11), pp. 3029-3036.

Molin, S. and Tolker-Nielsen, T. (2003) 'Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure', *Curr Opin Biotechnol*, 14(3), pp. 255-61.

Monteiro-Neto, V., Bando, S.Y., Moreira-Filho, C.A. and Giron, J.A. (2003) 'Characterization of an outer membrane protein associated with haemagglutination

and adhesive properties of enteroaggregative *Escherichia coli* O111:H12', *Cell Microbiol*, 5(8), pp. 533-47.

Morin, N., Tirling, C., Ivison, S.M., Kaur, A.P., Nataro, J.P. and Steiner, T.S. (2010) 'Autoactivation of the AggR regulator of enteroaggregative *Escherichia coli* in vitro and in vivo', *FEMS Immunol Med Microbiol*, 58(3), pp. 344-55.

Mounier, J., Popoff, M.R., Enninga, J., Frame, M.C., Sansonetti, P.J. and Van Nhieu, G.T. (2009) 'The IpaC Carboxyterminal Effector Domain Mediates Src-Dependent Actin Polymerization during *Shigella* Invasion of Epithelial Cells', *PLoS Pathogens*, 5(1), p. e1000271.

Mulvey, M.A., Schilling, J.D. and Hultgren, S.J. (2001) 'Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection', *Infect Immun*, 69(7), pp. 4572-9.

Muniyappa, K. and Radding, C.M. (1986) 'The homologous recombination system of phage lambda. Pairing activities of beta protein', *J Biol Chem*, 261(16), pp. 7472-8.

Murphy, K.C. (1991) 'Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of *Escherichia coli* RecBCD enzyme', *J Bacteriol*, 173(18), pp. 5808-21.

Murphy, K.C. (1998) 'Use of Bacteriophage λ Recombination Functions To Promote Gene Replacement in *Escherichia coli*', *Journal of Bacteriology*, 180(8), pp. 2063-2071.

Nachin, L., Nannmark, U. and Nystrom, T. (2005) 'Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility', *J Bacteriol*, 187(18), pp. 6265-72.

Nambu, T. and Kutsukake, K. (2000) 'The *Salmonella* FlgA protein, a putativeve periplasmic chaperone essential for flagellar P ring formation', *Microbiology*, 146 (Pt 5), pp. 1171-8.

Nan, B. and Zusman, D.R. (2016) 'Novel mechanisms power bacterial gliding motility', *Molecular microbiology*, 101(2), pp. 186-193.

Nataro, J.P. and Kaper, J.B. (1998) 'Diarrheagenic *Escherichia coli*', *Clin Microbiol Rev*, 11(1), pp. 142-201.

Navarro-Garcia, F. and Elias, W.P. (2011) 'Autotransporters and virulence of enteroaggregative *E. coli*', *Gut Microbes*, 2(1), pp. 13-24.

Neely, M.N. and Friedman, D.I. (1998) 'Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release', *Mol Microbiol*, 28(6), pp. 1255-67.

Neuwald, A.F., Aravind, L., Spouge, J.L. and Koonin, E.V. (1999) 'AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes', *Genome Res*, 9(1), pp. 27-43.

Noguchi, T. and Golden, S. (2017) *Bioluminescent and fluorescent reporters in circadian rhythm studies*.

Notley-McRobb, L., Death, A. and Ferenci, T. (1997) 'The relationship between external glucose concentration and cAMP levels inside *Escherichia coli*: implications for models of phosphotransferase-mediated regulation of adenylate cyclase', *Microbiology*, 143(6), pp. 1909-1918.

Nowicki, B., Selvarangan, R. and Nowicki, S. (2001) 'Family of *Escherichia coli* Dr adhesins: decay-accelerating factor receptor recognition and invasiveness', *J Infect Dis*, 183 Suppl 1, pp. S24-7.

O'Hanley, P., Lalonde, G. and Ji, G. (1991) 'Alpha-hemolysin contributes to the pathogenicity of piliated digalactoside-binding *Escherichia coli* in the kidney: efficacy of an alpha-hemolysin vaccine in preventing renal injury in the BALB/c mouse model of pyelonephritis', *Infection and Immunity*, 59(3), pp. 1153-1161.

O'Hanley, P., Lark, D., Falkow, S. and Schoolnik, G. (1985) 'Molecular basis of *Escherichia coli* colonization of the upper urinary tract in BALB/c mice. Gal-Gal pili immunization prevents *Escherichia coli* pyelonephritis in the BALB/c mouse model of human pyelonephritis', *J Clin Invest*, 75(2), pp. 347-60.

Ochman, H., Lawrence, J.G. and Groisman, E.A. (2000) 'Lateral gene transfer and the nature of bacterial innovation', *Nature*, 405(6784), pp. 299-304.

Ochman, H. and Selander, R.K. (1984) 'Standard reference strains of *Escherichia coli* from natural populations', *Journal of Bacteriology*, 157(2), pp. 690-693.

Ohnishi, K., Kutsukake, K., Suzuki, H. and Lino, T. (1992) 'A novel transcriptional regulation mechanism in the flagellar regulon of *Salmonella typhimurium*: an antisigma factor inhibits the activity of the flagellum-specific sigma factor, sigma F', *Mol Microbiol*, 6(21), pp. 3149-57.

Ohnishi, K., Ohto, Y., Aizawa, S., Macnab, R.M. and Iino, T. (1994) 'FlgD is a scaffolding protein needed for flagellar hook assembly in *Salmonella typhimurium*', *J Bacteriol*, 176(8), pp. 2272-81.

Orskov, I., Orskov, F., Jann, B. and Jann, K. (1977) 'Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*', *Bacteriological Reviews*, 41(3), pp. 667-710.

Paget, M.S.B. and Helmann, J.D. (2003) 'The σ(70) family of sigma factors', *Genome Biology*, 4(1), pp. 203-203.

Pallen, M.J. and Matzke, N.J. (2006) 'From The Origin of Species to the origin of bacterial flagella', *Nature Reviews Microbiology*, 4, p. 784.

Park, S.Y., Lowder, B., Bilwes, A.M., Blair, D.F. and Crane, B.R. (2006) 'Structure of FliM provides insight into assembly of the switch complex in the bacterial flagella motor', *Proc Natl Acad Sci U S A*, 103(32), pp. 11886-91.

Peirano, G. and Pitout, J.D. (2010) 'Molecular epidemiology of *Escherichia coli* producing CTX-M beta-lactamases: the worldwide emergence of clone ST131 O25:H4', *Int J Antimicrob Agents*, 35(4), pp. 316-21.

Perna, N.T., Plunkett, G., 3rd, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J.,

Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R. (2001) 'Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7', *Nature*, 409(6819), pp. 529-33.

Pesavento, C., Becker, G., Sommerfeldt, N., Possling, A., Tschowri, N., Mehlis, A. and Hengge, R. (2008) 'Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*', *Genes Dev*, 22(17), pp. 2434-46.

Pesavento, C. and Hengge, R. (2012) 'The global repressor FliZ antagonizes gene expression by sigmaS-containing RNA polymerase due to overlapping DNA binding specificity', *Nucleic Acids Res*, 40(11), pp. 4783-93.

Phue, J.N., Lee, S.J., Trinh, L. and Shiloach, J. (2008) 'Modified *Escherichia coli* B (BL21), a superior producer of plasmid DNA compared with *Escherichia coli* K (DH5alpha)', *Biotechnol Bioeng*, 101(4), pp. 831-6.

Plague, G.R., Boodram, K.S., Dougherty, K.M., Bregg, S., Gilbert, D.P., Bakshi, H. and Costa, D. (2017) 'Transposable Elements Mediate Adaptive Debilitation of Flagella in Experimental *Escherichia coli* Populations', *J Mol Evol*, 84(5-6), pp. 279-284.

Poggio, T.V., La Torre, J.L. and Scodeller, E.A. (2006) 'Intranasal immunization with a recombinant truncated FimH adhesin adjuvanted with CpG oligodeoxynucleotides protects mice against uropathogenic *Escherichia coli* challenge', *Can J Microbiol*, 52(11), pp. 1093-102.

Prüss, B.M. and Matsumura, P. (1996) 'A regulator of the flagellar regulon of *Escherichia coli*, *flhD*, also affects cell division', *Journal of Bacteriology*, 178(3), pp. 668-674.

Qadri, F., Svennerholm, A.M., Faruque, A.S. and Sack, R.B. (2005) 'Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention', *Clin Microbiol Rev*, 18(3), pp. 465-83.

Reid, S.D., Herbelin, C.J., Bumbaugh, A.C., Selander, R.K. and Whittam, T.S. (2000) 'Parallel evolution of virulence in pathogenic *Escherichia coli*', *Nature*, 406(6791), pp. 64-7.

Ren, C.P., Chaudhuri, R.R., Fivian, A., Bailey, C.M., Antonio, M., Barnes, W.M. and Pallen, M.J. (2004) 'The ETT2 gene cluster, encoding a second type III secretion system from *Escherichia coli*, is present in the majority of strains but has undergone widespread mutational attrition', *J Bacteriol*, 186(11), pp. 3547-60.

Riley, L.W., Remis, R.S., Helgerson, S.D., McGee, H.B., Wells, J.G., Davis, B.R., Hebert, R.J., Olcott, E.S., Johnson, L.M., Hargrett, N.T., Blake, P.A. and Cohen, M.L. (1983) 'Hemorrhagic colitis associated with a rare *Escherichia coli* serotype', *N Engl J Med*, 308(12), pp. 681-5.

Rogers, T.J., Thorpe, C.M., Paton, A.W. and Paton, J.C. (2012) 'Role of lipid rafts and flagellin in invasion of colonic epithelial cells by Shiga-toxigenic *Escherichia coli* O113:H21', *Infect Immun*, 80(8), pp. 2858-67.

Rojas-Lopez, M., Monterio, R., Pizza, M., Desvaux, M. and Rosini, R. (2018) 'Intestinal Pathogenic *Escherichia coli*: Insights for Vaccine Development', *Frontiers in Microbiology*, 9, p. 440.

Rolhion, N., Barnich, N., Bringer, M.A., Glasser, A.L., Ranc, J., Hebuterne, X., Hofman, P. and Darfeuille-Michaud, A. (2010) 'Abnormally expressed ER stress response chaperone Gp96 in CD favours adherent-invasive *Escherichia coli* invasion', *Gut*, 59(10), pp. 1355-62.

Roy, K., Hilliard, G.M., Hamilton, D.J., Luo, J., Ostmann, M.M. and Fleckenstein, J.M. (2009) 'Enterotoxigenic *Escherichia coli* EtpA mediates adhesion between flagella and host cells', *Nature*, 457(7229), pp. 594-8.

Roy, K., Kansal, R., Bartels, S.R., Hamilton, D.J., Shaaban, S. and Fleckenstein, J.M. (2011) 'Adhesin degradation accelerates delivery of heat-labile toxin by enterotoxigenic *Escherichia coli*', *J Biol Chem*, 286(34), pp. 29771-9.

Ruchaud-Sparagano, M.H., Muhlen, S., Dean, P. and Kenny, B. (2011) 'The enteropathogenic *E. coli* (EPEC) Tir effector inhibits NF-kappaB activity by targeting TNFalpha receptor-associated factors', *PLoS Pathog*, 7(12), p. e1002414.

Sahl, J.W., Matalka, M.N. and Rasko, D.A. (2012) 'Phylomark, a tool to identify conserved phylogenetic markers from whole-genome alignments', *Appl Environ Microbiol*, 78(14), pp. 4884-92.

Saini, S., Brown, J.D., Aldridge, P.D. and Rao, C.V. (2008) 'FliZ Is a posttranslational activator of $FlhD_4C_2$-dependent flagellar gene expression', *J Bacteriol*, 190(14), pp. 4979-88.

Saini, S., Slauch, J.M., Aldridge, P.D. and Rao, C.V. (2010) 'Role of cross talk in regulating the dynamic expression of the flagellar *Salmonella* pathogenicity island 1 and type 1 fimbrial genes', *J Bacteriol*, 192(21), pp. 5767-77.

Saitou, N. and Nei, M. (1987) 'The neighbor-joining method: a new method for reconstructing phylogenetic trees', *Molecular Biology and Evolution*, 4(4), pp. 406-425.

Salinger, N., Kokona, B., Fairman, R. and Okeke, I.N. (2009) 'The plasmid-encoded regulator activates factors conferring lysozyme resistance on enteropathogenic *Escherichia coli* strains', *Appl Environ Microbiol*, 75(1), pp. 275-80.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: A laboratory manual*. 2nd edn. USA: Cold Spring Harbor Laboratory Press.

Satoh, H., Yamakawa, T., Kindaichi, T., Ito, T. and Okabe, S. (2006) 'Community structures and activities of nitrifying and denitrifying bacteria in industrial wastewater-treating biofilms', *Biotechnol Bioeng*, 94(4), pp. 762-72.

Sauer, R.T., Bolon, D.N., Burton, B.M., Burton, R.E., Flynn, J.M., Grant, R.A., Hersch, G.L., Joshi, S.A., Kenniston, J.A., Levchenko, I., Neher, S.B., Oakes, E.S.C., Siddiqui, S.M., Wah, D.A. and Baker, T.A. (2004) 'Sculpting the Proteome with AAA+ Proteases and Disassembly Machines', *Cell*, 119(1), pp. 9-18.

Schroeder, G.N. and Hilbi, H. (2008) 'Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion', *Clin Microbiol Rev*, 21(1), pp. 134-56.

Selander, R.K., Caugant, D.A. and Whittam, T.S. (1987) 'Genetic structure and variation in natural populations of *Escherichia coli. Escherichia coli* and *Salmonella typhimurium*', *Cell. Mol. Biol*, 2, pp. 1625-1648.

Selander, R.K. and Levin, B.R. (1980) 'Genetic diversity and structure in *Escherichia coli* populations', *Science*, 210(4469), pp. 545-7.

Selvarangan, R., Goluszko, P., Singhal, J., Carnoy, C., Moseley, S., Hudson, B., Nowicki, S. and Nowicki, B. (2004) 'Interaction of Dr Adhesin with Collagen Type IV Is a Critical Step in *Escherichia coli* Renal Persistence', *Infection and Immunity*, 72(8), pp. 4827-4835.

Serra, D.O., Richter, A.M. and Hengge, R. (2013a) 'Cellulose as an architectural element in spatially structured *Escherichia coli* biofilms', *J Bacteriol*, 195(24), pp. 5540-54.

Serra, D.O., Richter, A.M., Klauck, G., Mika, F. and Hengge, R. (2013b) 'Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm', *MBio*, 4(2), pp. e00103-13.

Servin, A.L. (2005) 'Pathogenesis of Afa/Dr Diffusely Adhering *Escherichia coli*', *Clinical Microbiology Reviews*, 18(2), pp. 264-292.

Shaner, N.C., Patterson, G.H. and Davidson, M.W. (2007) 'Advances in fluorescent protein technology', *Journal of Cell Science*, 120(24), pp. 4247-4260.

Shi, W., Zhou, Y., Wild, J., Adler, J. and Gross, C.A. (1992) 'DnaK, DnaJ, and GrpE are required for flagellum synthesis in *Escherichia coli*', *J Bacteriol*, 174(19), pp. 6256-63.

Shimada, T., Fujita, N., Yamamoto, K. and Ishihama, A. (2011) 'Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources', *PLoS One*, 6(6), p. e20081.

Shimomura, O., Johnson, F.H. and Saiga, Y. (1962) 'Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*', *J Cell Comp Physiol*, 59, pp. 223-39.

Shin, S. and Park, C. (1995) 'Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR', *J Bacteriol*, 177(16), pp. 4696-702.

Siguier, P., Filee, J. and Chandler, M. (2006) 'Insertion sequences in prokaryotic genomes', *Curr Opin Microbiol*, 9(5), pp. 526-31.

Siguier, P., Gourbeyre, E. and Chandler, M. (2014) 'Bacterial insertion sequences: their genomic impact and diversity', *FEMS Microbiol Rev*, 38(5), pp. 865-91.

Sim, M., Koirala, S., Picton, D., Strahl, H., Hoskisson, P.A., Rao, C.V., Gillespie, C.S. and Aldridge, P.D. (2017) 'Growth rate control of flagellar assembly in *Escherichia coli* strain RP437', *Scientific Reports*, 7, p. 41189.

Simm, R., Lusch, A., Kader, A., Andersson, M. and Romling, U. (2007) 'Role of EAL-containing proteins in multicellular behavior of *Salmonella enterica* serovar Typhimurium', *J Bacteriol*, 189(9), pp. 3613-23.

Singer, H.M., Kühne, C., Deditius, J.A., Hughes, K.T. and Erhardt, M. (2014) 'The *Salmonella* Spi1 Virulence Regulatory Protein HilD Directly Activates Transcription of the Flagellar Master Operon *flhDC*', *Journal of Bacteriology*, 196(7), pp. 1448-1457.

Sivick, K.E. and Mobley, H.L. (2010) 'Waging war against uropathogenic *Escherichia coli*: winning back the urinary tract', *Infect Immun*, 78(2), pp. 568-85.

Skerker, J.M. and Berg, H.C. (2001) 'Direct observation of extension and retraction of type IV pili', *Proceedings of the National Academy of Sciences*, 98(12), pp. 6901-6904.

Skoudy, A., Mounier, J., Aruffo, A., Ohayon, H., Gounon, P., Sansonetti, P. and Tran Van Nhieu, G. (2000) 'CD44 binds to the *Shigella* IpaB protein and participates in bacterial invasion of epithelial cells', *Cell Microbiol*, 2(1), pp. 19-33.

Smajs, D., Micenkova, L., Smarda, J., Vrba, M., Sevcikova, A., Valisova, Z. and Woznicova, V. (2010) 'Bacteriocin synthesis in uropathogenic and commensal *Escherichia coli*: colicin E1 is a potential virulence factor', *BMC Microbiol*, 10, p. 288.

Smith, N.H. and Selander, R.K. (1991) 'Molecular genetic basis for complex flagellar antigen expression in a triphasic serovar of *Salmonella*', *Proceedings of the National Academy of Sciences*, 88(3), pp. 956-960.

Sommerfeldt, N., Possling, A., Becker, G., Pesavento, C., Tschowri, N. and Hengge, R. (2009) 'Gene expression patterns and differential input into curli fimbriae regulation of all GGDEF/EAL domain proteins in *Escherichia coli*', *Microbiology*, 155(Pt 4), pp. 1318-31.

Song, J., Bishop, B.L., Li, G., Grady, R., Stapleton, A. and Abraham, S.N. (2009) 'TLR4-mediated expulsion of bacteria from infected bladder epithelial cells', *Proc Natl Acad Sci U S A*, 106(35), pp. 14966-71.

Sousa, A., Bourgard, C., Wahl, L.M. and Gordo, I. (2013) 'Rates of transposition in *Escherichia coli*', *Biology Letters*, 9(6), p. 20130838.

Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A. and Bertin, P. (1999) 'Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon', *J Bacteriol*, 181(24), pp. 7500-8.

Soutourina, O.A. and Bertin, P.N. (2003) 'Regulation cascade of flagellar expression in Gram-negative bacteria', *FEMS Microbiology Reviews*, 27(4), pp. 505-523.

Soutourina, O.A., Krin, E., Laurent-Winter, C., Hommais, F., Danchin, A. and Bertin, P.N. (2002) 'Regulation of bacterial motility in response to low pH in *Escherichia coli*: the role of H-NS protein', *Microbiology*, 148(Pt 5), pp. 1543-51.

Sperandio, V., Torres, A.G. and Kaper, J.B. (2002) 'Quorum sensing *Escherichia coli* regulators B and C (QseBC): a novel two-component regulatory system involved in

the regulation of flagella and motility by quorum sensing in *E. coli*', *Mol Microbiol*, 43(3), pp. 809-21.

Stafford, G.P., Ogi, T. and Hughes, C. (2005) 'Binding and transcriptional activation of non-flagellar genes by the *Escherichia coli* flagellar master regulator FlhD(2)C(2)', *Microbiology (Reading, England)*, 151(Pt 6), pp. 1779-1788.

Stanley, T. and Wilson, I.G. (2001) 'Multilocus Enzyme Electrophoresis', in Spencer, J.F.T. and de Ragout Spencer, A.L. (eds.) *Food Microbiology Protocols*. Totowa, NJ: Humana Press, pp. 369-393.

Stewart, P.S. and Costerton, J.W. (2001) 'Antibiotic resistance of bacteria in biofilms', *Lancet*, 358(9276), pp. 135-8.

Takaya, A., Erhardt, M., Karata, K., Winterberg, K., Yamamoto, T. and Hughes, K.T. (2012) 'YdiV: a dual function protein that targets FlhDC for ClpXP-dependent degradation by promoting release of DNA-bound FlhDC complex', *Mol Microbiol*, 83(6), pp. 1268-84.

Tang, Y., Guest, J.R., Artymiuk, P.J., Read, R.C. and Green, J. (2004) 'Post-transcriptional regulation of bacterial motility by aconitase proteins', *Molecular Microbiology*, 51(6), pp. 1817-1826.

Tenaillon, O., Skurnik, D., Picard, B. and Denamur, E. (2010) 'The population genetics of commensal *Escherichia coli*', *Nat Rev Microbiol*, 8(3), pp. 207-17.

Tomoyasu, T., Ohkishi, T., Ukyo, Y., Tokumitsu, A., Takaya, A., Suzuki, M., Sekiya, K., Matsui, H., Kutsukake, K. and Yamamoto, T. (2002) 'The ClpXP ATP-Dependent Protease Regulates Flagellum Synthesis in *Salmonella enterica* Serovar Typhimurium', *Journal of Bacteriology*, 184(3), pp. 645-653.

Tomoyasu, T., Takaya, A., Isogai, E. and Yamamoto, T. (2003) 'Turnover of FlhD and FlhC, master regulator proteins for *Salmonella* flagellum biogenesis, by the ATP-dependent ClpXP protease', *Mol Microbiol*, 48(2), pp. 443-52.

Trepeta, R.W. and Edberg, S.C. (1984) 'Methylumbelliferyl-beta-D-glucuronide-based medium for rapid isolation and identification of *Escherichia coli*', *J Clin Microbiol*, 19(2), pp. 172-4.

Tsilibaris, V., Maenhaut-Michel, G. and Van Melderen, L. (2006) 'Biological roles of the Lon ATP-dependent protease', *Res Microbiol*, 157(8), pp. 701-13.

Turner, S.M., Chaudhuri, R.R., Jiang, Z.D., DuPont, H., Gyles, C., Penn, C.W., Pallen, M.J. and Henderson, I.R. (2006a) 'Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages', *J Clin Microbiol*, 44(12), pp. 4528-36.

Turner, S.M., Scott-Tucker, A., Cooper, L.M. and Henderson, I.R. (2006b) 'Weapons of mass destruction: virulence factors of the global killer enterotoxigenic *Escherichia coli*', *FEMS Microbiol Lett*, 263(1), pp. 10-20.

Uehling, D.T. and Wolf, L. (1969) 'Enhancement of the bladder defense mechanism by immunization', *Invest Urol*, 6(5), pp. 520-6.

Valiadi, M. and Iglesias-Rodriguez, D. (2013) 'Understanding Bioluminescence in Dinoflagellates—How Far Have We Come?', *Microorganisms*, 1(1), pp. 3-25.

Waidmann, M.S., Bleichrodt, F.S., Laslo, T. and Riedel, C.U. (2011) 'Bacterial luciferase reporters: The Swiss army knife of molecular biology', *Bioengineered Bugs*, 2(1), pp. 8-16.

Wang, S., Fleming, R.T., Westbrook, E.M., Matsumura, P. and McKay, D.B. (2006) 'Structure of the *Escherichia coli* FlhDC complex, a prokaryotic heteromeric regulator of transcription', *J Mol Biol*, 355(4), pp. 798-808.

Wang, X., Dubey, A.K., Suzuki, K., Baker, C.S., Babitzke, P. and Romeo, T. (2005) 'CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*', *Mol Microbiol*, 56(6), pp. 1648-63.

Watarai M, Funato S and Sasakawa C (1996) 'Interaction of Ipa proteins of *Shigella flexneri* with alpha5beta1 integrin promotes entry of the bacteria into mammalian cells', *J Exp Med*, 183(3), pp. 991-9.

Wei, B.L., Brun-Zinkernagel, A.M., Simecka, J.W., Pruss, B.M., Babitzke, P. and Romeo, T. (2001) 'Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*', *Mol Microbiol*, 40(1), pp. 245-56.

Weichart, D., Querfurth, N., Dreger, M. and Hengge-Aronis, R. (2003) 'Global Role for ClpP-Containing Proteases in Stationary-Phase Adaptation of *Escherichia coli*', *Journal of Bacteriology*, 185(1), pp. 115-125.

Welch, R.A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L., Donnenberg, M.S. and Blattner, F.R. (2002) 'Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*', *Proc Natl Acad Sci U S A*, 99(26), pp. 17020-4.

Wickner, S., Maurizi, M.R. and Gottesman, S. (1999) 'Posttranslational quality control: folding, refolding, and degrading proteins', *Science*, 286(5446), pp. 1888-93.

Wiedenbeck, J. and Cohan, F.M. (2011) 'Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches', *FEMS Microbiol Rev*, 35(5), pp. 957-76.

Winson, M.K., Swift, S., Fish, L., Throup, J.P., Jorgensen, F., Chhabra, S.R., Bycroft, B.W., Williams, P. and Stewart, G.S. (1998a) 'Construction and analysis of *luxCDABE*-based plasmid sensors for investigating N-acyl homoserine lactone-mediated quorum sensing', *FEMS Microbiol Lett*, 163(2), pp. 185-92.

Winson, M.K., Swift, S., Hill, P.J., Sims, C.M., Griesmayr, G., Bycroft, B.W., Williams, P. and Stewart, G.S.A.B. (1998b) 'Engineering the *luxCDABE* genes from *Photorhabdus luminescens* to provide a bioluminescent reporter for constitutive and promoter probe plasmids and mini-Tn5 constructs', *FEMS Microbiology Letters*, 163(2), pp. 193-202.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden, M.C., Ochman, H. and Achtman, M. (2006) 'Sex and virulence in *Escherichia coli*: an evolutionary perspective', *Mol Microbiol*, 60(5), pp. 1136-51.

Wolfe, A.J. and Visick, K.L. (2008) 'Get the Message Out: Cyclic-Di-GMP Regulates Multiple Levels of Flagellum-Based Motility', *Journal of Bacteriology*, 190(2), pp. 463-475.

Wooster, D.G., Maruvada, R., Blom, A.M. and Prasadarao, N.V. (2006) 'Logarithmic phase *Escherichia coli* K1 efficiently avoids serum killing by promoting C4bp-mediated C3b and C4b degradation', *Immunology*, 117(4), pp. 482-493.

Wright, K.J. and Hultgren, S.J. (2006) 'Sticky fibers and uropathogenesis: bacterial adhesins in the urinary tract', *Future Microbiol*, 1(1), pp. 75-87.

Wright, K.J., Seed, P.C. and Hultgren, S.J. (2005) 'Uropathogenic *Escherichia coli* flagella aid in efficient urinary tract colonization', *Infect Immun*, 73(11), pp. 7657-68.

Wurpel, D.J., Beatson, S.A., Totsika, M., Petty, N.K. and Schembri, M.A. (2013) 'Chaperone-usher fimbriae of *Escherichia coli*', *PLoS One*, 8(1), p. e52835.

Xicohtencatl-Cortes, J., Monteiro-Neto, V., Saldaña, Z., Ledesma, M.A., Puente, J.L. and Girón, J.A. (2009) 'The Type 4 Pili of Enterohemorrhagic *Escherichia coli* O157:H7 Are Multipurpose Structures with Pathogenic Attributes', *Journal of Bacteriology*, 191(1), pp. 411-421.

Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T. and Babitzke, P. (2013) 'CsrA activates *flhDC* expression by protecting *flhDC* mRNA from RNase E-mediated cleavage', *Molecular microbiology*, 87(4), pp. 851-866.

Yanagihara, S., Iyoda, S., Ohnishi, K., Iino, T. and Kutsukake, K. (1999) 'Structure and transcriptional control of the flagellar master operon of *Salmonella typhimurium*', *Genes Genet Syst*, 74(3), pp. 105-11.

Yaron, S. and Romling, U. (2014) 'Biofilm formation by enteric pathogens and its role in plant colonization and persistence', *Microb Biotechnol*, 7(6), pp. 496-516.

Yokoseki, T., Iino, T. and Kutsukake, K. (1996) 'Negative regulation by *fliD*, *fliS*, and *fliT* of the export of the flagellum-specific anti-sigma factor, FlgM, in *Salmonella typhimurium*', *Journal of Bacteriology*, 178(3), pp. 899-901.

Yonekura, K., Maki-Yonekura, S. and Namba, K. (2003) 'Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy', *Nature*, 424, p. 643.

Yonekura, K., Maki, S., Morgan, D.G., DeRosier, D.J., Vonderviszt, F., Imada, K. and Namba, K. (2000) 'The bacterial flagellar cap as the rotary promoter of flagellin self-assembly', *Science*, 290(5499), pp. 2148-52.

Zhang, Z., Kukita, C., Humayun, M.Z. and Saier, M.H. (2017) 'Environment-directed activation of the *Escherichia coli flhDC* operon by transposons', *Microbiology*, 163(4), pp. 554-569.