UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Flagellin glycosylation in *Pantoea ananatis*

Submitted by

Angelique du Preez

Submitted in partial fulfilment of the requirements for the degree

Master Scientiae (MSc)

In the Faculty of Natural & Agricultural Sciences, Department of Microbiology,
Forestry and Agricultural Biotechnology Institute
University of Pretoria
Pretoria

April 2016

Supervisor: Dr. P. De Maayer

Co-Supervisor: Prof. T. Coutinho

DECLARATION

I, the undersigned, declare that the dissertation submitted herewith for the degree Master Scientiae to the University of Pretoria, contains my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

_____

Angelique du Preez

15 April 2016

ACKNOWLEDGEMENTS

- All praise, glory and thanks is given to God for the guidance, wisdom and strength in this long journey.

- My parents Anna and André du Preez, thank-you for always encouraging me to do more than I thought I could and for your constant support, love and staying up late to keep me company while I work. I will always cherish those moments and be forever grateful for them.

- My siblings Stephen and Kelly du Preez for your support and always pushing me to think critically so my work doesn't sound like "Science-Fiction".

- My best friend and boyfriend Ludwig Eksteen for always being there to support me with everything, as well as to challenge and help me with my work. I am truly forever grateful.

- Special thanks to my amazing supervisors Teresa Coutinho and Pieter De Maayer for their guidance, knowledge, support and tremendous amounts of patience. I know the journey was not always smooth sailing but we made it and I am thankful for all I learned from both of you.

- All my friends especially Gaby, Thumbiko and Gina for all their support, love, patience and knowledge.

- All my lab mates in FABI 1-48 for their knowledge, help and patience.

- The National Research Foundation (NRF) and the Centre of excellence for Tree Health Biotechnology (CTHB) for funding.

PREFACE

Pantoea ananatis, a member of the family *Enterobacteriaceae*, is distributed globally and has been isolated from a wide range of environmental sources. Strains of this species are frequently found in close association with a number of plant species and may cause a wide range of disease symptoms. However, little is known about how *P. ananatis* causes these symptoms and whether or not strains are host specific. Genomic analyses revealed the presence of flagellin glycosylation island (FGI) in this bacterium, which is involved in the posttranslational modification of bacterial flagella. This modification has been linked to a number of different functions, including host specificity and virulence in related phytopathogens. In this, comparative genomic and molecular genetic approaches will be used to analyse the FGIs in a number of *P. ananatis* strains isolated from different plant hosts.

In **Chapter 1** of this dissertation, literature pertinent to the topic of the dissertation will be reviewed and will include aspects of the genetics, regulation and structural components of flagellar motility, before providing an in depth discussion on flagellin glycosylation, with a specific focus on this phenomenon in phytobacteria. In the first research chapter, **Chapter 2**, the molecular genetics and evolution of the extensive and versatile flagellin glycosylation loci in the genomes of seventeen distinct *P. ananatis* strains for which genome sequences are available will be undertaken. In **Chapter 3**, genetic screening to determine the presence of FGIs in a large collection of *P. ananatis* strains, as well as the prevalence of each distinct FGI type will be employed. Finally, any linkage between the FGI type an individual *P. ananatis* strain contains and the particular host it infects, and geographical location from which it was isolated, will be elucidated.

CONTENTS

DECLARATION.....i

ACKNOWLEDGEMENTS.....ii

PREFACE.....iii

Chapter one: Flagella and flagellin glycosylation in bacteria

1.1 Abstract.....2

1.2 Introduction.....2

1.3 The Bacterial Flagellum.....4

1.3.1 Structure.....5

1.3.2 Assembly and regulation of the flagellum.....7

1.3.3 Genomic organisation.....8

1.3.4 The Chemotactic cascade.....9

1.3.5 Functions of flagellum.....9

1.3.6 Antigenicity of the flagellum.....11

1.4 Means of avoiding host detection of the flagellum.....13

1.5 Flagellin glycosylation.....15

1.5.1 Structure.....15

1.5.2 Genomic organisation of the flagellin glycosylation genes.....17

1.5.3 Assembly of the flagellin glycan chain.....17

1.5.4 The functions of flagellin glycosylation.....18

1.6 Conclusion.....21

1.7 References.....22

1.8 Figures and tables.....43

Chapter two: Comparative genomic analyses of the flagellin glycosylation islands of seventeen *Pantoea ananatis* strains

2.1 Abstract.....48

2.2 Introduction.....48

2.3 Materials and methods.....51

2.3.1 Identification of flagellin glycosylation islands in the *P. ananatis* genomes.....51

2.3.2 Ortholog identification between the P. ananatis FGI regions.....51

2.3.3 In silico analyses of the FGI CDS sets 51
2.3.4 Phylogenetic analyses and construction of an FGI typing dendrogram 52
2.4 Results and discussion 52
2.4.1 General properties of the *P. ananatis* strains FGI 52
2.4.2 The *P. ananatis* FGIs can be clustered into four distinct types 53
2.4.3 The *P. ananatis* strains FGI carry genes for several distinct sugar biosynthetic pathway 54
2.4.4 The *P. ananatis* FGIs show evidence of further glycan modifications 56
2.4.5 *P. ananatis* strains vary in both the number and type of flagellin genes 57
2.5 Conclusion 59
2.6 References 61
2.7 Tables and figures 70

Chapter three: Typing of *P. ananatis* strains on the basis of the variability in their flagellin glycosylation islands

3.1 Abstract 79
3.2 Introduction 79
3.3 Method and Materials 81
3.3.1 *P. ananatis* strain collection, growth and DNA extraction 81
3.3.2 Primer design for conserved FGI genes and variable genes 81
3.3.3 Polymerase chain reaction (PCR) screening protocols 82
3.3.4 Visualisation of PCR amplicons 83
3.4 Results and discussion 83
3.4.1 FGI loci are present in all screened *P. ananatis* strains 83
3.4.2 FGI Types I, II and III, but not Type IV, are prevalent among *P. ananatis* strains. 84
3.4.3 The *P. ananatis* flagellin glycosylation island may show greater genetic variability than the current scope of FGI types. 85
3.4.4 No correlation could be observed between FGI type and the host from which a *P. ananatis* strain was isolated. 85
3.4.5 There is poor correlation between *P. ananatis* FGI Type and geographic source of isolation 86
3.5 Conclusion 87
3.6 References 89

3.7 Tables and figures.....95

SUMMARY 105

CHAPTER ONE

FLAGELLA AND FLAGELLIN GLYCOSYLATION IN BACTERIA

1.1 Abstract

Bacteria are one of the most abundant living organisms that are capable of dwelling in a broad range of ecological niches that have numerous stresses. Their fitness is determined by the capabilities they possess to survive in a particular ecological niche, through the use of various mechanisms including flagella and flagellin glycosylation. Flagella and the post-translational modification, flagellin glycosylation provide the bacteria with several abilities to assist with motility, adhesion, invasion, infection and survival in its niche. This review focuses on the genetic, phylogenetic, physical assembly and functional aspects of both the flagellum and flagellin glycosylation. As well as how both flagella and flagellin glycosylation make the bacterium capable of avoiding host defence responses.

1.2 Introduction

Bacteria are among the dominant living organisms, with an estimated 9.2 to 31.7 x 10^{29} bacterial cells found on Earth (Kallmeyer *et al.*, 2012). Their dominance can be attributed to their capacity to survive and thrive in almost all of Earth's ecological niches, including soil (Geisseler *et al.*, 2010, Husen, 2013, Udikovic-Kolic *et al.*, 2014), water (Morohoshi *et al.*, 2007, Maleki *et al.*, 2015), the atmosphere (Maruthamuthu *et al.*, 2008, Hospodsky *et al.*, 2012), food (Gobbetti *et al.*, 2010, Ryan *et al.*, 2015), as well as in association with various vertebrate, insect and plant hosts (Coutinho & Venter, 2009, Rosenberg *et al.*, 2010, Feasey *et al.*, 2012, Eleftherianos *et al.*, 2013, McFall-Ngai *et al.*, 2013, Oliver *et al.*, 2014, Scharf *et al.*, 2016). In these diverse ecological niches, bacteria are exposed to a wide range and combination of biotic and abiotic stresses (Davey & O'toole, 2000, An *et al.*, 2006), including UV irradiation and associated DNA damage on light exposed surfaces (Davey & O'toole, 2000, Jacobs & Sundin, 2001), nutrient limitation (Redford *et al.*, 2010, Remus-Emsermann *et al.*, 2012), low water availability (Remus-Emsermann *et al.*, 2012), antibiotics and toxic compounds (Ramos *et al.*, 2004) and competition with other bacteria for limiting resources (Lindow, 1991, Hirano & Upper, 2000, Coleman, 2001, An *et al.*, 2006).

The ability of a bacterium to cope with environmental stress is referred to as their fitness (Hirano & Upper, 2000). This fitness can be increased in small increments by a combination of phenotypic traits expressed by the bacterium (Hirano & Upper, 2000). Some of these traits include the production of exopolysaccharides and biofilms that surrounds the bacteria and protects them from harsh environmental conditions (Lindow, 1991, Hirano & Upper, 2000), DNA repair mechanisms to repair UV-mediated mutagenic damage (Jacobs & Sundin, 2001), as well as producing of siderophores and antibiotics to outcompete other microorganisms sharing the same ecological niche (Hirano & Upper, 2000). Another important fitness factor among plant-associated bacteria is swimming motility which allows the bacteria to move in liquids, and towards nutrient rich areas and away from toxic compounds (Rashid & Kornberg, 2000, Tomich *et al.*, 2002). Furthermore, in host-associated bacteria, it assists in the infection process as well as evasion of host defences (Tomich *et al.*, 2002, Harshey, 2003, Takeuchi *et al.*, 2003, Merritt *et al.*, 2007).

Swimming motility occurs by means of a surface appendage, the flagellum, and has been observed in most known bacterial taxa (Aldridge & Hughes, 2002, Thomas *et al.*, 2007). As a result, the molecular genetics, mechanical functioning and biological roles they play in various bacteria have been well-documented. As such, the flagellum has been demonstrated to be a multifunctional organ of the bacterial cell, playing a role not only in motility, but also adhesion and biofilm formation, as well as host invasions and virulence in pathogenic bacteria (Josenhans & Suerbaum, 2002, Tomich *et al.*, 2002, Ramos *et al.*, 2004, Herrera *et al.*, 2008). However, flagella are also highly immunogenic structures and can trigger host defence responses (Felix *et al.*, 1999). Some bacteria have evolved several mechanisms to counteract this effect. These include phase variable expression of antigenically distinct flagellin proteins, the main structural subunit of the flagellum, and glycosylation of the flagellin protein (Felix *et al.*, 1999, Schmitt *et al.*, 2001, Ottemann & Lowenthal, 2002, Tomich *et al.*, 2002). The latter involves the posttranslational addition of a sugar chain to the flagellin protein, effectively masking its antigenic domains (Takeuchi *et al.*, 2003, Hayakawa & Ishizuka, 2012). This has been observed in a wide range of bacterial taxa and is particularly common among clinical and plant pathogens (Takeuchi *et al.*, 2003, Logan, 2006, Hayakawa & Ishizuka, 2012).

In this dissertation studies were performed, by means of comparative genomic and molecular analyses, flagellin glycosylation in the plant-associated bacterium *Pantoea ananatis*. This bacterium is frequently found as epi- or endophyte on a broad host range of plant hosts, but more significantly also causes plant diseases in a range of these hosts and we postulate that flagellin glycosylation is an important factor in *P. ananatis* plant host interactions. Here we highlight some of the pertinent literature about the flagellum, its structure, assembly, and genetics. Furthermore, we discuss means employed by bacteria to prevent detection by the host and the launching of host defences, with particular focus on flagellin glycosylation.

1.3 The Bacterial Flagellum

The flagellum is a filamentous organ located in the bacterial membrane protruding to the exterior of the cell (Aldridge & Hughes, 2002, Balaban & Hendrixson, 2011, Hayakawa & Ishizuka, 2012). It is typically between 15 and 20 nm in width and approximately 10 -15 µm in length (Namba & Vonderviszt, 1997, Harshey, 2003, Sowa & Berry, 2008, Parker *et al.*, 2014). Flagellation is a common phenomenon among bacteria and has been observed in more than 80% of all known bacterial species (Thomas *et al.* 2001). There is tremendous variability in terms of the number of flagella and their arrangement on the cell surface (Aldridge & Hughes, 2002, Balaban & Hendrixson, 2011, Hayakawa & Ishizuka, 2012). Some bacteria contain a single flagellum on one (polar) or both sides (bi-polar) of the cell (Takeuchi *et al.*, 2003, Amiel *et al.*, 2010, Hayakawa & Ishizuka, 2012). Others have many flagella on one side (lophotrichous), both sides (amphitrichous) or around the whole cell (peritrichous) (Figure 1). Due to the complexity of its structure, the flagellum is composed of various proteins, which all play a role in its structure or functioning (Aldridge & Hughes, 2002).

1.3.1 Structure

The flagellum typically comprises of three main components: a basal body, hook and a filament (Figure 2). The basal body comprises of several rings which anchor the flagellar filament to the bacterial cell envelope (Minamino & Macnab, 1999). The basal body in Gram-positive bacteria differs from that of Gram-negative bacteria, due to differences in their cell wall composition, and comprises of only two rings, with one in the periplasm and one in the cytoplasmic membrane (Courtney *et al.*, 2012, Schuhmacher *et al.*, 2015). In Gram-negative bacteria, the basal body consists of four rings (L, P, MS and C) located in different layers of the cell wall and the rod, which connects the hook to the basal body (Hayakawa & Ishizuka, 2012). The inner ring, the C-ring, is located in the cytoplasm (Hayakawa & Ishizuka, 2012). It consists of three proteins, FliG, FliM and FliN, which together make up components of the flagellar motor switch and energizing component (Pandini *et al.*, 2015). FliG directly assists in the rotation of the flagella as it is situated next to the motor complex (MotA/B), at the top of the C-ring (Pandini *et al.*, 2015)(Figure 2). The second ring complex, the MS-ring (a complex of two rings the M- and S-rings)(Ueno *et al.*, 1994) is located in the cytoplasmic membrane and is connected to the hook (Hayakawa & Ishizuka, 2012). The MS-ring consists of multiple copies of the membrane protein FliF (Ogawa *et al.*, 2015). The P- and L- rings are located in the peptidoglycan layer and the outer membrane, respectively (Jones *et al.*, 1989). These two rings consist of the structural protein FlgI (Hizukuri *et al.*, 2008) and the lipoprotein FlgH (Berg, 2003), respectively. Both the P- and L-rings form the outer cylinder of a large pore in the basal body, through which the rod passes (Stallmeyer *et al.*, 1989, Sowa & Berry, 2008). The combination of the two rings supports the hook like a "clamp" to keep the hook in place while it is able to rotate. This is known as bushing, which occurs between the rotor and the outer parts of the cell wall (Sowa & Berry, 2008). The proximal end of the rod passes through the MS-ring to the P- and L-rings (Figure 2), and consists of the FliE, FlgB, FlgC, and FlgF proteins (Dailey & Berg, 1993, Berg, 2003, Saijo-Hamano *et al.*, 2004). The distal end of the rod connects to the flagellar hook and is composed of the FlgG protein (Dailey & Berg, 1993, Berg, 2003, Saijo-Hamano *et al.*, 2004). The proteins of the rod are all structural proteins that assist in the rotation of the hook (Samatey *et al.*, 2004, Courtney *et al.*, 2012).

The flagellar hook is composed of multiple monomers of the FlgE protein and is connected, at its proximal end, to the rod and is known as the hollow linker domain (Bonifield *et al.*, 2000, Berg, 2003, Courtney *et al.*, 2012, Hayakawa & Ishizuka, 2012). The flagellar hook forms a hollow cylindrical structure, which serves as an export channel for the filament proteins (Logan, 2006). The hook component is also flexible in order to transfer torque from the rod to the filament and assists with change in angle that occurs in rotation (Samatey *et al.*, 2004, Courtney *et al.*, 2012). On the distal terminal of the hook there are two hook-filament junction proteins consisting of the hook-associated proteins FlgK and FlgL (Berg, 2003). These assist in the connection between the hook and the filament (Berg, 2003).

The filament represents the surface-exposed portion of the flagellum (Figure 2). The filament is polymerized from large numbers (between 20,000 – 30,000 subunits) of copies of a single protein, flagellin (FliC)(Namba & Vonderviszt, 1997). In some bacteria, there are several copies of the *fliC* gene, which together synthesize the filament. For instance, *Campylobacter* species produce two different flagellin proteins, FlaA and FlaB (Logan, 2006). FlaB has homology to the flagellin proteins found in other bacterial species, whereas the FlaA protein has no sequence similarities (Logan, 2006). At the distal end of the flagellar structure is the flagellar cap protein (FliD)(Berg, 2003). This cap protein assists with the polymerization and self-assembly of the flagellin monomers in the filament, ensuring that the flagellin monomers are incorporated into the protofilament and do not escape into the extracellular environment (Berg, 2003, Hayakawa & Ishizuka, 2012).

Flagellum rotation is driven by a motor (Berg, 2003),that consists of two components, the stator and rotor (Reid *et al.*, 2006). The stator is composed of the MotA and MotB proteins (Ridgway *et al.*, 1977, Lloyd *et al.*, 1999, Berg, 2003). These proteins span through the cytoplasmic membrane into the peptidoglycan layer, that assists in the anchoring of the protein complex (Berg, 2003). The rotor is composed of the C-ring containing, FliM, FliN and FliG, where the FliG protein is in direct contact with the Mot complex (Berg, 2003, Kojima, 2015, Pandini *et al.*, 2015). In the motor complex, MotAB, generates the force needed to rotate the flagellum through proton conduction (Hayakawa & Ishizuka, 2012). This proton conduction is created by the influx of monovalent cations from the periplasmic space to the cytoplasm across the inner membrane (Hayakawa & Ishizuka, 2012).

1.3.2 Assembly and regulation of the flagellum

The production of flagella is tightly regulated, as flagellar synthesis places a substantial metabolic burden on the bacterial cell (Bonifield *et al.*, 2000, Bonifield & Hughes, 2003, Kinoshita *et al.*, 2013). Hence, the flagellar genes are transcribed at three different stages and transcription occurs under the control of three distinct classes of promoters (Figure 3)(Kutsukake *et al.*, 1990, Kinoshita *et al.*, 2013). The first set of genes, Class I, consists of the master regulator *flh*CD, and is controlled by different σ factors depending on the bacterial species at an early stage of flagellum synthesis (Aldridge & Hughes, 2002). For example *Salmonella* sp. and *Vibrio* sp. use the σ^{70} promoter whereas *Caulobacter crescentus* utilises the σ^{73} promoter (Aldridge & Hughes, 2002).

The second class of genes in the cascade, Class II, has its promoters directly activated by FlhCD, and it incorporates genes encoding the basal body and export machinery, as well as the sigma factor *fliA* (σ^{28}), which positively regulates the Class III genes, while *flgM* negatively regulates the next genes in the cascade (Chilcott & Hughes, 2000, McCarter, 2001, Aldridge & Hughes, 2002). The Class III regulon incorporates those genes required for the synthesis of the hook complex, the filament, motor force generators and chemosensory machinery (Chilcott & Hughes, 2000, Aldridge & Hughes, 2002). Once the flagellin (FliC) proteins have been produced, the cytosolic chaperone FliS binds to a 40 amino acid region in the C-terminal of the former protein (Evdokimov *et al.*, 2003, Ozin *et al.*, 2003, Logan, 2006). This complex assists in the prevention of premature monomer interaction in the cytoplasm (Evdokimov *et al.*, 2003, Ozin *et al.*, 2003, Logan, 2006). Flagellin proteins are then driven to the export apparatus through the mediation of a specific N-terminus sequence recognition and is then secreted through a type III secretion system channel to the tip of the growing filament, where the flagellin monomers are assembled (Macnab, 2003, Logan, 2006). The flagellin monomers are assembled into 11 protofilaments, which are subsequently utilized to produce the filament (Hayakawa & Ishizuka, 2012), forming a long, hollow, cylindrical shape in the centre of the filament (Hayakawa & Ishizuka, 2012). This is known as the flagellar channel, which enables the transport of the proteins needed for filament assembly, post-translational modification and, in some cases, the secretion of virulence effectors (Macnab, 1999, Macnab, 2003, Logan, 2006, Abby & Rocha, 2012, Forster & Marquis, 2012, Hayakawa & Ishizuka, 2012). This has led to the theory that the flagellar

export system may have evolved from a common ancestor to the Type III Secretion Systems (T3SS) which are pivotal in the virulence of several clinical and plant pathogenic bacteria (Abby & Rocha, 2012, Forster & Marquis, 2012).

1.3.3 Genomic organisation

The flagellum is a complex structure and requires a large number of genes for assembly, structure, regulation, motor force generation and the chemosensory machinery (Aldridge & Hughes, 2002). In the model organism *Escherichia coli*, approximately 50 genes are involved in flagellum biosynthesis (Soutourina & Bertin, 2003), while the flagellum biosynthetic gene complement of *Vibrio parahaemolyticus* incorporates 57 genes (Kim & McCarter, 2000). These genes are localised in distinct loci on the chromosome, and are generally well conserved among different bacterial taxa (Chilcott & Hughes, 2000, Aldridge & Hughes, 2002, Dasgupta *et al.*, 2003, Takeuchi *et al.*, 2003). In most Gram-negative bacteria, including the motility model organisms *E. coli*, *Salmonella enterica* and *Pseudomonas aeruginosa*, the genes involved in flagellum biosynthesis, regulation and functioning are maintained in three non-contiguous loci (Silverman & Simon, 1973, Soutourina & Bertin, 2003). Locus I comprises mainly of flagellum structural genes, locus II codes for proteins involved in regulation of flagellum assembly, the motor proteins and the chemotactic proteins, while locus III incorporates the export apparatus genes, the flagellum-specific sigma factor σ^{28} (*fliA*) as well as several flagellum structural genes (Silverman & Simon, 1973, Soutourina & Bertin, 2003). Within each of the loci, several operons can be found, which further assist in the regulatory hierarchy of flagellum biosynthesis (Komeda *et al.*, 1980, Kutsukake *et al.*, 1988, Kutsukake *et al.*, 1990, Macnab, 1992).

1.3.4 The Chemotactic cascade

Flagellated bacterium are capable of directed motility in response to exposure to distinct chemical stimuli, towards nutrients and oxygen, and away from toxic substances (Harshey, 2003, Takeuchi *et al.*, 2003, Merritt *et al.*, 2007). This phenomenon occurs through the signal transduction known as the chemotaxis system (Macnab, 1992, Kojima & Blair, 2004, Olsen *et al.*, 2013). This system was first extensively researched in *E. coli* and *S. enterica* (Bischoff & Ordal, 1992). The chemotactic cascade consists of six major proteins CheA, CheY, CheW, CheB, CheR and CheZ (Bischoff & Ordal, 1992). Once an environmental stimulus is present, it causes a conformational change in the cell membrane-bound methyl-accepting chemotaxis proteins (MCPs)(Bischoff & Ordal, 1992, Ottemann & Miller, 1997, Maurer *et al.*, 2005, Hazelbauer *et al.*, 2008, Sourjik & Wingreen, 2012). These MCPs subsequently autophosphorylates the histine kinase protein CheA, which has a coupling protein, CheW, which autoregulates CheA (Bischoff & Ordal, 1992, Ottemann & Miller, 1997, Maurer *et al.*, 2005, Hazelbauer *et al.*, 2008, Sourjik & Wingreen, 2012). CheA then phosphorylates the CheY protein through transfer of a phosphate group (Bischoff & Ordal, 1992, Hazelbauer & Lai, 2010), and the phosphorylated CheY protein (chemotaxis related protein) subsequently binds to FliM in the C-ring, which has the capability of switching the rotation direction of the flagellum (Sockett *et al.*, 1992, Mathews *et al.*, 1998, Hayakawa & Ishizuka, 2012). When the phosphorylated CheY interacts with FliM, clockwise rotation occurs that causes the bacterium to tumble, and when there is no CheY to interact with, counter-clockwise rotation occurs (Irikura *et al.*, 1993, Mathews *et al.*, 1998, Kojima & Blair, 2004, Hayakawa & Ishizuka, 2012). The chemotaxis system is able to adapt to this change due to the covalent modifications that occur, with the methylation of CheR and the demethylation of CheB (Hazelbauer & Lai, 2010). FliN and CheZ are then capable of dephosphorylating the CheY protein, which reverts the latter protein to its original state (Bischoff & Ordal, 1992, Lowenthal *et al.*, 2009, Pandini *et al.*, 2015).

1.3.5 Functions of flagellum

A multitude of distinct functions have been ascribed to the flagellum. Notwithstanding, the best characterized flagellar function is two distinct forms of motility, namely swimming and swarming motility (Young *et al.*, 1999b, Harshey, 2003, Merritt *et al.*, 2007, Herrera *et al.*,

2008, Sourjik & Wingreen, 2012). Swimming motility involves the rapid forwards and backwards movement, as well as running and tumbling of an individual cell in response to chemotactic signals (Harshey, 2003, Merritt *et al.*, 2007, Sourjik & Wingreen, 2012). This motility is advantageous to the bacteria as they are able to get access to better nutrients, are able to avoid toxic substances and move to more favourable ecological niches (Rashid & Kornberg, 2000). Swarming motility occurs when bacterial cells are elongated, hyperflagellated, and moves in a coordinated manner as a group in a thin fluid layer on a surface (Young *et al.*, 1999b, Harshey, 2003, Herrera *et al.*, 2008). This coordinated movement plays a role in surface colonization and biofilm formation (Harshey, 2003). Furthermore swarming motility has been observed to assist in the invasion and virulence of the human pathogen *Helicobacter pylori* and phytopathogen *Ps. syringae* pv. *glycinea* (Hattermann & Ries, 1989, Ottemann & Lowenthal, 2002, Herrera *et al.*, 2008).

Flagella have also been demonstrated to play a role in the adhesion to surface and subsequent colonization (Ramos *et al.*, 2004, Merritt *et al.*, 2007). When the adhesion of a flagellated *Clostridium difficile* strain was compared to a non-flagellated strain it was observed that the adhesion to the cecal tissue of germ-free mice was 10-fold lower, emphasizing the importance flagella has in adherence in this opportunistic pathogen (Tasteyre *et al.*, 2001, Ramos *et al.*, 2004). In *Listeria monocytogenes* it was demonstrated that the flagella played a role in adhesion to stainless steel surfaces (Vatanyoopaisarn *et al.*, 2000). The plant-pathogen *Agrobacterium tumefaciens* and plant-growth promoting bacteria *Rhizobium leguminosarum* bv. *viciae* and *Azospirillum brasilense* have also been elucidated to utilize flagellum-mediated adhesion in their infection of plant cells and root surfaces, respectively (Rodríguez-Navarro *et al.*, 2007). In the cystic fibrosis pathogen *Ps. aeruginosa* the flagellum was observed to play a role in attachment to mucus in the airway lumen (Ramos *et al.*, 2004).

Flagellar motility is pivotal to host invasion and spread within host tissues among several pathogenic bacteria (Tomich *et al.*, 2002). Enhanced flagellum-mediated invasion has for example been demonstrated in *Campylobacter jejuni*, *Proteus mirabilis*, *Vibrio anguillarum*, *Salmonella enterica* and *Listeria monocytogenes* (Yao *et al.*, 1994, Mobley *et al.*, 1996, Ormonde *et al.*, 2000, Tomich *et al.*, 2002, van Asten *et al.*, 2004, O'Neil & Marquis, 2006). Flagellum-mediated motility enabled *Burkholderia cepacia* to penetrate host epithelial

barriers, therefore leading to formation of infection and systemic spread (Tomich *et al.*, 2002). Furthermore, flagellar motility has been observed to be essential for the effective virulence of many bacterial pathogens, and as such this form of motility is generally considered as a virulence factor. Moreover virulence factors have also been demonstrated to be secreted via the flagellar export system, including *Campylobacter jejuni* that uses the flagellar export system to secrete a set of invasive antigen proteins (Cia proteins)(Konkel *et al.*, 2004), *S. enterica* serovar *typhimurium* secretes several virulence proteins (Komoriya *et al.*, 1999) and *Y. enterocolitica* that secretes numerous extracellular toxins, including YplA (Young *et al.*, 1999a).

1.3.6 Antigenicity of the flagellum

While flagella are essential for moving to the site of entry, invasion, attachment, colonisation and virulence, the flagellum also serves as a means of recognition of an invading bacterium by the host, which can subsequently mount a concerted defence response (Hattermann & Ries, 1989, Schmitt *et al.*, 2001, Josenhans & Suerbaum, 2002, Ottemann & Lowenthal, 2002, Tomich *et al.*, 2002, Harshey, 2003, Takeuchi *et al.*, 2003, Ramos *et al.*, 2004, Merritt *et al.*, 2007, Herrera *et al.*, 2008). This can be linked to the highly immunogenic nature of the major flagellar structural protein flagellin (Jones & Dangl, 2006). Plant-associated and mammal-associated bacteria have flagellin proteins that contain a number of epitopes, which can serve as Microbe-Associated Molecular Patterns (MAMPs)(Akira & Takeda, 2004, Jones & Dangl, 2006). These elicitors are recognized by host pattern recognition receptors (PRRs) located in the transmembrane of the host (Jones & Dangl, 2006). In mammalian hosts the best characterised PRR are the Toll-like receptors (TLRs), while plants recognise MAMPs by means of receptor kinases (Akira & Takeda, 2004, Jones & Dangl, 2006, Dangl *et al.*, 2013). Recognition of MAMPs by these transmembrane PRRs triggers inducible innate immunity and MAMP triggered immunity and inducible innate immunity in mammalian and plant hosts, respectively (Jones & Dangl, 2006, Cai *et al.*, 2011, Dangl *et al.*, 2013). Recognition of flagellin by mammalian TLRs, in particular TLR5 (Gewirtz *et al.*, 2001a, Hayashi *et al.*, 2001, Akira & Takeda, 2004), activates the production of numerous defence proteins, including nuclear factor NF-κB, interleukin-1b, mitogen-activated protein kinases (MAPKS) (McDermott *et al.*, 2000, Gewirtz *et al.*, 2001a, Hayashi *et al.*, 2001), matrilysin, inducible nitric oxide synthase and nitric oxide, human β–defensin 2, tumour necrosis factor-α (TNF-

α)(McDermott *et al.*, 2000) and chemokines (McDermott *et al.*, 2000, Eaves-Pyles *et al.*, 2001, Gewirtz *et al.*, 2001a, Gewirtz *et al.*, 2001b, Sierro *et al.*, 2001, Zeng *et al.*, 2003, Ramos *et al.*, 2004). This immune response then serves to eradicate the pathogen from host tissues and prevent re-infection (Sierro *et al.*, 2001).

MAMP Triggered Immunity (MTI) in plants activates the salicylic acid (SA) signalling pathway, activating the production of abscisic acid (ABA), which results in the closure of the plants stomata to prevent further pathogen entry, contain the already internalised pathogen and potentially destroy the infecting pathogen (Felix *et al.*, 1999, Melotto *et al.*, 2006, Neill *et al.*, 2008, Tsuda *et al.*, 2008, Wang *et al.*, 2009, Zeng *et al.*, 2010). MTI also leads to the production of reactive oxygen species, phenolic esters, PR's: protein antimicrobial compounds, resistance (R) genes and the thickening of the plant cell wall, to try prevent and contain the bacteria from spreading (Dangl & Jones, 2001, Nimchuk *et al.*, 2003). General host defence responses will be activated regardless of whether the flagellated bacterial invader is pathogenic or non-pathogenic (Hutcheson, 1998). However, non-pathogenic bacteria, result in much quicker triggering of host responses and at higher levels than those induced in response to pathogenic bacteria (Hutcheson, 1998).

Characterized flagellin epitopes which serve as MAMPs include flgII 28, flg15 and flg22 in *Ps. syringae* species, *Agrobacterium tumefaciens*, *Rhizobium meliloti*, and *Ralstonia solanacearum* (Gómez-Gómez *et al.*, 1999, Meindl *et al.*, 2000, Jones & Dangl, 2006, Cai *et al.*, 2011, Veluchamy *et al.*, 2014). The antigenicity of the flagellin protein (FliC) has been linked to a highly conserved domain with a short amino acid sequence, namely the flg22 epitope (Felix *et al.*, 1999). This epitope is a 22 amino acid sequence which is located in the N-terminal of the flagellin protein secondary structure (Sun *et al.*, 2013b). In *Arabidopsis* plants, the FLS2 receptor is responsible for recognition of the bacterial-derived flagellin elicitor (Felix *et al.*, 1999). The FLS2 protein contains an extracellular leucine rich repeat (LRR), transmembrane domain and a cytoplasmic serine/threonine kinase domain (Gómez-Gómez & Boller, 2000). In a study performed on *Arabidopsis thaliana* seedlings it was observed that the flg22 and flg15 bacterial elicitors were able to cause callose deposition, a strong inhibition of seedling growth, the release of active oxygen species and the induction of genes that code for the pathogenesis–related proteins PR-1 and PR-5 that help with

pathogenesis (Gómez-Gómez *et al.*, 1999). However if there is a difference in the amino acid sequence of the receptor peptide or in the corresponding conserved domain of the flagellin, then there will be limited or no host defence response (Gómez-Gómez *et al.*, 1999, Meng *et al.*, 2013).

1.4 Means of avoiding host detection of the flagellum

In response to hosts detecting the flagellin of invading bacteria and initiating defence mechanisms, bacteria have evolved several means to avoid host detection of the flagellum (Bonifield & Hughes, 2003, Soutourina & Bertin, 2003, Akira & Takeda, 2004, Jones & Dangl, 2006, Liu *et al.*, 2012). One well-characterised means of avoiding host detection is through the phenomenon of flagella phase variation (Ikeda *et al.*, 2001, Aldridge *et al.*, 2006, Yamamoto & Kutsukake, 2006). Flagella phase variation has demonstrated in a number of distinct bacterial taxa include *E. coli*, *Pseudomonas* sp., *Salmonella* sp. and *Vibrio* sp. (Déziel *et al.*, 2001, Ikeda *et al.*, 2001, Achouak *et al.*, 2004, Liu *et al.*, 2012, Zhou *et al.*, 2015). Flagella phase variation is the stochastic variation in the production of the flagellum, by switching genes on and off to create a different flagellum (Ikeda *et al.*, 2001, Van Der Woude & Bäumler, 2004, Aldridge *et al.*, 2006). This variation in flagellum is specifically caused by the regulation of the production of flagellin (FliC), where several antigenically distinct copies may be produced by some taxa (Ikeda *et al.*, 2001, Bonifield & Hughes, 2003, Van Der Woude & Bäumler, 2004, Aldridge *et al.*, 2006, Liu *et al.*, 2012, Zhou *et al.*, 2015). For example, the genome of *Vibrio parahaemolyticus* incorporates six *fli*C genes whereas *V. cholera* contains five flagellin genes (McCarter, 2001, Soutourina & Bertin, 2003). The majority of the research undertaken on phase variation has been performed on *Salmonella enterica* which encodes two distinct flagellin proteins, FljB and FliC (Bonifield & Hughes, 2003, Liu *et al.*, 2012).

The stochastic variation of flagella phase variation is produced through the process where each bacterial flagellin has its own H phase, one H phase with a promoter, the flagellin and an inhibitory, while the other just consists of a promoter and flagellin (Zieg *et al.*, 1978, Ikeda *et al.*, 2001, Aldridge *et al.*, 2006, Yamamoto & Kutsukake, 2006, Feng *et al.*, 2008). *Salmonella enterica* serovar *typhimurium* is one of the first bacteria that flagellar phase

variation was observed in, with the switching between two flagellar types, through the alternative expression of two flagellin, FljB and FliC (Stocker, 1949, Ikeda *et al.*, 2001, Bonifield & Hughes, 2003, Aldridge *et al.*, 2006, Liu *et al.*, 2012, Zhou *et al.*, 2015). This switching is between the expression of a small operon, *flj*BA, flagellin *flj*B, and the inhibitory gene *flj*A that inhibits the synthesis of the second flagellin, *fli*C, and then of just *fli*C (Zieg *et al.*, 1978, Feng *et al.*, 2008, Liu *et al.*, 2012). This process occurs through the site-specific DNA inversion of a chromosomal region, the promoter (Ikeda *et al.*, 2001, Bonifield & Hughes, 2003, Liu *et al.*, 2012). Due to recombination sites that are 26bp invert repeats containing the *hin* gene that codes for a recombinase, which is needed to assist the switching, flanking the promoter and allowing it to be inverted (Silverman & Simon, 1980, Ikeda *et al.*, 2001, Bonifield & Hughes, 2003, Liu *et al.*, 2012). In the one orientation FljA and FliB are produced, while *fli*C is then inhibited by FljA, and in the other orientation FliC is produced while *flj*A and *flj*B are inhibited (Steege, 2000, Bonifield & Hughes, 2003, Feng *et al.*, 2008, Liu *et al.*, 2012).

More over this process of stochastic variation only occurs in *Salmonella enterica* serovars that are bilateral with two or more flagellin, unlike that of *S. enterica* serovar *typhi* that only possesses one flagellin FliC (Liu *et al.*, 2012). However, a study performed in 2007 observed that *S. enterica* serovar t*yphi* strains isolated from Indonesia produce two flagellin, FliC and a novel flagellin Z66 (Baker *et al.*, 2007). In general flagellin phase variation occurs in those bacterial taxa containing more than one flagellin gene. However, a distinct phase variation has been observed in some *E. coli* strains which contain only a single *fliC* gene (Feng *et al.*, 2008, Liu *et al.*, 2012). This has been linked to a flagellin-specifying gene *flkA* and a repressor gene, *flkB*, located on a genomic islet (*flk* GI)(Liu *et al.*, 2012). The presence of the *flk* GI results in repression of *fliC* (Liu *et al.*, 2012). However, once the *flk* GI has been removed from the chromosome, *flk*AB will be deleted and FliC will be produced (Liu *et al.*, 2012). Another means that some bacteria use to prevent host detection of the flagellar MAMP is through posttranslational modification of the flagellin with glycan chains, known as flagellin glycosylation.

1.5 Flagellin glycosylation

Flagellin glycosylation is a modification process whereby bacteria have glycan and polysaccharide chains that are post-translationally attached to the flagellin (Logan, 2006). Flagellin glycosylation has been observed in a number of different bacterial and archaeal taxa, and is particularly prevalent among plant and animal pathogens (Logan, 2006). The process of flagella glycosylation was first observed in the human pathogen *Campylobacter coli* (Alm *et al.*, 1992). It has subsequently been demonstrated in other clinical pathogens including *Helicobacter pylori*, *Aeromonas* spp. and *Shewanella* spp. as well as phytopathogens, including *Xanthomonas* spp. and *Pseudomonas* spp. (Schirm *et al.*, 2003, Takeuchi *et al.*, 2003, Logan, 2006, Taguchi *et al.*, 2010, Hayakawa & Ishizuka, 2012).

1.5.1 Structure

The monomeric flagellin proteins consist of four distinct domains: D0, D1, D2 and D3 (Figure 4)(Sun *et al.*, 2013a). The D0 and D1 domains are highly conserved regions and are found at the N- and C- terminals of the amino acid sequence, respectively. These two regions are composed of α-helices and form the core of the filament structure (Beatson *et al.*, 2006). The D2 and D3 domains are hypervariable regions and are composed of β-strands (Beatson *et al.*, 2006). These two regions form the central portion of the flagellin amino acid sequence, and are located on the surface of the flagellum filament (Figure 4). This hypervariable region represents the predominant site for flagellin glycosylation (Takeuchi *et al.*, 2003, Hayakawa & Ishizuka, 2012).

Two distinct conformations of flagellin glycosylation of the hypervariable regions of flagellin have been observed, namely *N*- and *O*- linkage (Takeuchi *et al.*, 2003, Logan, 2006). In *N*-linkage the glycan is attached to the nitrogen in either Asparagine or Arginine amino acid residues of the flagellin protein (Voisin *et al.*, 2005, Logan, 2006, Nothaft & Szymanski, 2010, Hayakawa & Ishizuka, 2012). *O*-linkage involves the attachment of the glycan to the hydroxyl oxygen of either Serine, Threonine or Tyrosine amino acid residues (Schirm *et al.*, 2003, Logan, 2006, Nothaft & Szymanski, 2010, Hayakawa & Ishizuka, 2012). *O*-linkage glycosylation is the more abundant form of glycosylation and has been observed in *Campylobacter* spp., *Aeromonas* spp., *Helicobacter* spp., and *Ps. aeruginosa* (Alm *et al.*,

1992, Thibault *et al.*, 2001, Schirm *et al.*, 2003, Schirm *et al.*, 2004, Schirm *et al.*, 2005, Logan, 2006). *N*-linked glycans have been observed for the flagellins of *Borrelia burgdorferi* and *Serpolina hyodysenteriae* (Li *et al.*, 1993, Ge *et al.*, 1998, Logan, 2006). Differences have also been observed in terms of the degree of glycosylation i.e. the number of glycans attached to a single flagellin monomer (Alm *et al.*, 1992, Thibault *et al.*, 2001, Zampronio *et al.*, 2011). The flagellins of *Campylobacter coli* VC167 are the most heavily glycosylated, with up to 19 Ser/Thr binding sites per flagellin monomer that are glycosylated (Alm *et al.*, 1992, Thibault *et al.*, 2001, Zampronio *et al.*, 2011). Furthermore, flagellin glycosylation is not universally present for all strains of a specific species. For example, *Ps. aeruginosa* PAK is referred to as having a glycosylated 'type-a' flagellum, whereas *Ps. aeruginosa* PA01 lacks genes for flagellin glycosylation, and its flagellum is hence referred to as a 'type-b' flagellum (Dasgupta *et al.*, 2003).

Extensive variability has also been observed in terms of the lengths of the glycan chains and the types of sugars incorporated (Thibault *et al.*, 2001, Schirm *et al.*, 2004, Logan, 2006, Zampronio *et al.*, 2011). The type-a flagella of *Ps. aeruginosa* PAK, have heterogeneous glycan chains consisting of eleven monosaccharide units, that are attached to the flagellin monomer at two sites by an *O*-linkage through a rhamnose residue (Schirm *et al.*, 2004, Logan, 2006). *Ps. syringae* pv *tabaci* 6605 is glycosylated by a trisaccharide at three residue sites (Takeuchi *et al.*, 2007). *Campylobacter jejuni* 81-176 was found to be glycosylated with pseudaminic acid, 5-acetamidino pseudaminic acid, and 5,7-N-(2,3-dihydroxyproprionyl-) pseudaminic acid attached to nineteen serine and threonine glycosylation sites (Thibault *et al.*, 2001, Zampronio *et al.*, 2011). *Campylobacter jejuni* 11168 was found to be glycosylated with two novel glycans of dimethylglyceric acid derivatives of pseudaminic acid and 7-acetamidino pseudaminic acid, I and II respectively (Zampronio *et al.*, 2011). These were attached by O-linkages, I to two serine sites and one threonine, and II to two serine sites (Zampronio *et al.*, 2011). The two flagellins (FlaA and FlaB) of *Helicobacter pylori* are glycosylated with 5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-L-manno-nonulosonic acid (pseudaminic acid)(Thibault *et al.*, 2001, Logan *et al.*, 2002, Schirm *et al.*, 2005) on seven and ten sites, respectively (Logan, 2006).

Further modifications of the glycan backbone chains have also been observed. For example, methyl, acetyl, acyl, amino and formyl groups are frequently found attached to the glycans (Power & Jennings, 2003, Hayakawa & Ishizuka, 2012). Initial observation of other post-translational modifications namely methylation was first discovered with the human and animal pathogen *Salmonella enterica* serovar *typhimurium* (Tronick & Martinez, 1971, Sun *et al.*, 2013a). These other modifications further contribute to the extensive variability of the glycan structures which are attached to flagellin proteins in distinct bacterial taxa.

1.5.2 Genomic organisation of the flagellin glycosylation genes

The genes responsible for flagellin glycosylation are typically found associated with the flagellum biosynthetic locus III (refer to Section 2.3), which contains the genes coding for the flagellin (FliC) proteins, capping protein (FliD), as well as the motor force generators (MotA and MotB), and the chemosensory machinery (Silverman & Simon, 1973, Soutourina & Bertin, 2003). Typically, the flagellin glycosylation genes are localised downstream of the *fliC* gene (Power & Jennings, 2003). The G+C content of the FGI frequently differs substantially from that of rest of the genome, so it is considered to be a genomic island, termed the flagellin glycosylation island (FGI), which may have been derived through horizontal gene transfer events (Power & Jennings, 2003, Parker *et al.*, 2014). The size and gene content of the FGIs also differs markedly among different flagellin-glycosylated species. For example, the *Ps. aeruginosa* PAK FGI is ~16 kb in size and contains fourteen genes (Arora *et al.*, 2001), whereas the ~50 kb FGI of *Campylobacter jejuni* 11168 encodes approximately 50 genes (Logan, 2006).

1.5.3 Assembly of the flagellin glycan chain

The flagellin glycosylation island is translated by the class III promoter, in the late stage along with the genes essential for the formation of the filament, motor force generator and the chemosensory machinery (section 2.3)(Silverman & Simon, 1973, Soutourina & Bertin, 2003). The translated protein products are all produced in the cytoplasm (Josenhans *et al.*, 2002). It was first assumed that the FGI encoded glycan chains would attach by *O*-linkage glycosylation to the surface-exposed hyper variable domain of the flagellin during the

filament assembly (Samatey *et al.*, 2000). However, it has since been proposed that *O*-glycosylation of the flagellin protein occurs either in the basal body environment, as the proteins and glycans are being exported through the type III secretion system channel, or at the cytoplasmic membrane close to the basal body (Figure 6)(Logan, 2006). This would allow for the assembly of multiple monosaccharides to form a glycan chain (Logan, 2006). These glycan chains are linked onto serine and threonine residues of the flagellin surface exposed domain by glycosyltransferases that are usually located within the FGI (Hayakawa & Ishizuka, 2012). Glycosylated flagellin proteins are then secreted via the type III secretion system channel, through the basal body and hook until they reach the filament tip, where they are arranged into the growing filament with the assistance of the filament cap protein (FliD)(Logan, 2006).

1.5.4 The functions of flagellin glycosylation

Flagellin glycosylation is of importance for bacteria dwelling in different niches including the phyllosphere or rhizosphere. It is critical for bacterial survival in harsh conditions, against other microorganisms, against plant defence mechanisms, invasion, and spread of infection (Hirano & Upper, 2000, Lindow & Brandl, 2003, Logan, 2006, Whipps *et al.*, 2008, Redford *et al.*, 2010). Various functional roles have been ascribed to flagellin glycosylation. In some bacterial taxa flagellin glycosylation is essential for flagellar gene expression, structural formation of the flagellum and flagellar motility. This has been observed, for example, in *Campylobacter* spp., (Logan *et al.*, 2002), *Helicobacter pylori* (Logan *et al.*, 2002, Power & Jennings, 2003) and *Caulobacter crescentus* (Leclerc *et al.*, 1998, Goon *et al.*, 2003). In *Aeromonas caviae*, the flagellin glycosylation genes are not only required for flagellum biosynthesis and motility, but also for the expression of the lipopolysaccharide O-antigen (Gryllos *et al.*, 2001, Hayakawa & Ishizuka, 2012). In other bacterial taxa, such as *Ps. aeruginosa* and *Ps. syringae*, however, flagellin glycosylation is not essential for flagellum biosynthesis and functioning, and as a result other functional roles have been demonstrated or proposed for flagellin glycosylation. These include adhesion, biofilm formation, virulence, host defence avoidance, as well as immune stimulation (Takeuchi *et al.*, 2003, Hayakawa & Ishizuka, 2012).

Binding of bacterium to a surface is ubiquitous in a diverse ecosystem, this gives the bacteria a greater chance for survival and a selective advantage over the rest (Costerton *et al.*, 1999, Dunne, 2002). Bacteria can become surface bound through the adhesion of the flagella (Dunne, 2002). In *Aeromonas caviae*, flagellin glycosylation was shown, by means of transposon mutagenesis, to play a key role in flagellar adhesion to surfaces (Schirm *et al.*, 2005, Logan, 2006). This microbial adhesion is furthermore essential in the formation of the biofilm (An *et al.*, 2000, Dunne, 2002, Merritt *et al.*, 2007). In *Campylobacter* autoagglutination is the initial steps in the production and formation of biofilms, and has been observed to be specifically mediated by the glycans on the flagellin (Jeon *et al.*, 2003, Guerry *et al.*, 2006).

In studies performed by Taguchi *et al.*, (2003) and Takeuchi *et al.*, (2003) it was observed that flagellin from *Ps. syringae* pv. *glycinea* and *Ps. syringae* pv. *tomato* were capable of producing hypersensitive cell death in a non-host plant (Taguchi *et al.*, 2003, Takeuchi *et al.*, 2003, Logan, 2006). This process was performed by the glycosylation of the flagellin which plays an important role not only in that process but also with host cell recognition. In *Ps. syringae* pv. *glycinea* it was observed that flagellin glycosylation is essential for virulence on its soybean host (Taguchi *et al.*, 2003, Takeuchi *et al.*, 2003, Logan, 2006). By contrast, the flagellin glycan resulted in a rapid hypersensitive response in non-host plants and hence flagellin glycosylation may be a determinant of host recognition and, concomitantly, host specificity (Takeuchi *et al.*, 2003).

Flagellin glycosylation was demonstrated to contribute to virulence in the opportunistic human pathogen *Ps. aeruginosa* PAK (Arora *et al.*, 2005, Hayakawa & Ishizuka, 2012). In *Campylobacter jejuni* it was reported that flagellin glycosylation was vital in virulence (Guerry, 2007, Ichinose *et al.*, 2013). In this latter pathogen it was demonstrated that glycosylation of the flagellin resulted in antigenic variation of this surface structure, and prevented host detection and concomitant triggering of host defence responses (Josenhans *et al.*, 2002). Similarly, in *Ps. syringae* pv. *glycinea* modification of a glycosyl moiety in the flagellin glycan also resulted in recognition by the host soybean, thereby further demonstrating a potential role for flagellin glycosylation in avoidance of host defences (Takeuchi *et al.*, 2003). In 2013, Ichinose *et al*., created a glycosylation defective mutant of *Xanthomonas campestris* pv. *campestris* and it was discovered that there was a decrease in the bacterium's motility as well as virulence (Ichinose *et al.*, 2013). Also *Campylobacter* spp. has been hypothesized to use flagellin glycosylation in antigenic variation on the bacterial cell surface to assist the bacterium in the avoidance of the host defence responses (Josenhans *et al.*, 2002). Given the extensive variability in terms of the flagellin glycan structure and different functions ascribed to flagellin glycans in different bacterial taxa, it can be expected that flagellin glycosylation makes a tailor-made contribution towards bacterial physiological functions and host interactions, and further work on the growing number of taxa in which flagellin glycosylation is observed will contribute towards our understanding of how it functions and what additional functional roles it plays in bacterial biology.

1.6 Conclusion

The literature review focused on the flagella and posttranslational modification of flagellin glycosylation in bacteria from different ecological niches. Flagella are complex and highly versatile structures, found in the majority of bacterial taxa. Aside from their role in motility, they play roles in biofilm formation, attachment to surfaces and colonization and infection of both plant and animal hosts. However, as flagellin, the major structural protein of the flagellum, is highly immunogenic, recognition of conserved domains in this protein can lead to the launching of host defence responses and disruption of the infection cycle. To counteract this effect, bacteria have developed various means to avoid detection of the flagellum, including phase variation and the posttranslational modification of flagellin with sugar chains. The latter mechanism is found mainly in plant and animal pathogens and is highly diverse, with extensive variability in the types and number of sugars incorporated in the glycan chain as well as the biological function associated with this trait. Furthermore, other posttranslational modification also occur to this glycan chain including methylation, acetylation, formylation and acylation. Thus producing a far more diverse flagellum, with glycosylated flagellin in the hypervariable domain, which can avoid detection by host receptors and can therefore serve as an innovative means to avoid host defence responses and contribute to other phenotypes, including biofilm formation, attachment and virulence. Possible limitations which impacted the review was the availability of the literature specifically on the flagellin glycosylation as little research has been undertaken. This limitation can be seen as an opportunity for future research.

1.7 References

Abby, S.S. & E.P. Rocha, (2012). The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *Plos Genetics.* **8**(9): e1002983

Achouak, W., S. Conrod, V. Cohen & T. Heulin, (2004). Phenotypic variation of *Pseudomonas brassicacearum* as a plant root-colonization strategy. *Molecular Plant-Microbe Interactions.* **17**(8): 872-879

Akira, S. & K. Takeda, (2004). Toll-like receptor signalling. *Nature Reviews Immunology.* **4**: 499-511

Aldridge, P. & K.T. Hughes, (2002). Regulation of flagellar assembly. *Current Opinion in Microbiology*. **5**: 160-165

Aldridge, P.D., C. Wu, J. Gnerer, J.E. Karlinsey, K.T. Hughes & M.S. Sachs, (2006). Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. *Proceedings of the National Academy of Sciences*. **103**(30): 11340-11345

Alm, R., P. Guerry & M. Power, (1992). Variation in antigenicity and molecular weight of *Campylobacter coli* VC167 flagellin in different genetic backgrounds. *Journal of Bacteriology.* **174**(13): 4230-4238

Amiel, E., R.R. Lovewell, G.A. O'Toole, D.A. Hogan & B. Berwin, (2010). *Pseudomonas aeruginosa* evasion of phagocytosis is mediated by loss of swimming motility and is independent of flagellum expression. *Infection and Immunity*. **78**(7): 2937-2945

An, D., T. Danhorn, C. Fuqua & M.R. Parsek, (2006). Quorum sensing and motility mediate interactions between *Pseudomonas aeruginosa* and *Agrobacterium tumefaciens* in biofilm cocultures. *Proceedings of the National Academy of Sciences of the United States of America.* **103**(10): 3828-3833

An, Y.H., R.B. Dickinson & R.J. Doyle, (2000). Mechanisms of bacterial adhesion and pathogenesis of implant and tissue infections. In: Handbook of Bacterial Adhesion. Springer. pp. 1-27

Arora, S.K., M. Bangera, S. Lory & R. Ramphal, (2001). A genomic island in *Pseudomonas aeruginosa* carries the determinants of flagellin glycosylation. *Proceedings of the National Academy of Sciences.* **98**(16): 9342-9347

Arora, S.K., A.N. Neely, B. Blair, S. Lory & R. Ramphal, (2005). Role of motility and flagellin glycosylation in the pathogenesis of *Pseudomonas aeruginosa* burn wound infections. *Infection and Immunity*. **73**(7): 4395-4398

Baker, S., J. Hardy, K.E. Sanderson, M. Quail, I. Goodhead, R.A. Kingsley, J. Parkhill, B. Stocker & G. Dougan, (2007). A novel linear plasmid mediates flagellar variation in *Salmonella typhi*. *PLoS Pathogens.* **3**(5) e59

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

Beatson, S.A., T. Minamino & M.J. Pallen, (2006). Variation in bacterial flagellins: from sequence to structure. *Trends in Microbiology.* **14**(4): 151-155

Berg, H.C., (2003). The rotary motor of bacterial flagella. *Biochemistry.* **72**: 19

Bischoff, D.S. & G.W. Ordal, (1992). Identification and characterization of FliY, a novel component of the *Bacillus subtilis* flagellar switch complex. *Molecular Microbiology*. **6**(18): 2715-2723

Bonifield, H.R. & K.T. Hughes, (2003). Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. *Journal of Bacteriology*.**185**(12): 3567-3574

Bonifield, H.R., S. Yamaguchi & K.T. Hughes, (2000). The flagellar hook protein, FlgE, of *Salmonella enterica* serovar *Typhimurium* is posttranscriptionally regulated in response to the stage of flagellar assembly. *Journal of Bacteriology*. **182**(14): 4044-4050

Cai, R., S. Yan, H. Liu & B.A. Vinatzer, (2011). Reconstructing host range evolution of bacterial plant pathogens using *Pseudomonas syringae* pv. *tomato* and its close relatives as a model. *Infection, Genetics and Evolution.* **11**: 1738-1751

Chevance, F.F. & K.T. Hughes, (2008). Coordinating assembly of a bacterial macromolecular machine. *Nature Reviews Microbiology.* **6**: 455-465

Chilcott, G.S. & K.T. Hughes, (2000). Coupling of Flagellar Gene Expression to Flagellar Assembly in *Salmonella enterica* Serovar *Typhimurium* and *Escherichia coli*. *Microbiology and Molecular Biology Reviews.* **64**(4): 694-708

Coleman, D., (2001). Cellular Microbiology. Bacteria–Host Interactions in Health and Disease:. Henderson B, Wilson M, McNab R, et al, eds. John Wiley and Sons, 1999. ISBN 0 471 98681 X. *Molecular Pathology*. **54**(1): 55

Costerton, J.W., P.S. Stewart & E. Greenberg, (1999). Bacterial biofilms: a common cause of persistent infections. *Science.* **284**: 1318-1322

Courtney, C.R., L.M. Cozy & D.B. Kearns, (2012). Molecular characterization of the flagellar hook in *Bacillus subtilis*. *Journal of Bacteriology*. **194**(17): 4619-4629

Coutinho, T.A. & S.N. Venter, (2009). *Pantoea ananatis*: an unconventional plant pathogen. *Molecular Plant Pathology*. **10**(3): 325-335

Dailey, F.E. & H.C. Berg, (1993). Mutants in disulfide bond formation that disrupt flagellar assembly in *Escherichia coli*. *Proceedings of the National Academy of Sciences*. **90**: 1043-1047

Dangl, J.L., D.M. Horvath & B.J. Staskawicz, (2013). Pivoting the plant immune system from dissection to deployment. *Science.* **341**: 746-751

Dangl, J.L. & J.D. Jones, (2001). Plant pathogens and integrated defence responses to infection. *Nature* .**411**: 826-833

Dasgupta, N., M.C. Wolfgang, A.L. Goodman, S.K. Arora, J. Jyot, S. Lory & R. Ramphal, (2003). A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Molecular Microbiology.* **50**(3): 809-824

Davey, M.E. & G.A. O'toole, (2000). Microbial biofilms: from ecology to molecular genetics. *Microbiology and Molecular Biology Reviews.* **64**(4): 847-867

Déziel, E., Y. Comeau & R. Villemur, (2001). Initiation of biofilm formation by *Pseudomonas aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. *Journal of Bacteriology*. **183**(4): 1195-1204

Dunne, W.M., (2002). Bacterial adhesion: seen any good biofilms lately? *Clinical Microbiology Reviews.* **15**(2): 155-166

Eaves-Pyles, T., K. Murthy, L. Liaudet, L. Virág, G. Ross, F.G. Soriano, C. Szabó & A.L. Salzman, (2001). Flagellin, a novel mediator of *Salmonella*-induced epithelial activation and systemic inflammation: IκBα degradation, induction of nitric oxide synthase, induction of proinflammatory mediators, and cardiovascular dysfunction. *The Journal of Immunology.* **166**: 1248-1260

Eleftherianos, I., J. Atri, J. Accetta & J.C. Castillo, (2013). Endosymbiotic bacteria in insects: guardians of the immune system. *Front Physiology.* **4**: 1-10

Evdokimov, A.G., J. Phan, J.E. Tropea, K.M. Routzahn, H.K. Peters, M. Pokross & D.S. Waugh, (2003). Similar modes of polypeptide recognition by export chaperones in flagellar biosynthesis and type III secretion. *Nature Structural & Molecular Biology.* **10**: 789-793

Feasey, N.A., G. Dougan, R.A. Kingsley, R.S. Heyderman & M.A. Gordon, (2012). Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *The Lancet.* **379**: 2489-2499

Felix, G., J.D. Duran, S. Volko & T. Boller, (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *The Plant Journal.* **18**(3): 265-276

Feng, L., B. Liu, Y. Liu, Y.A. Ratiner, B. Hu, D. Li, X. Zong, W. Xiong & L. Wang, (2008). A genomic islet mediates flagellar phase variation in *Escherichia coli* strains carrying the flagellin-specifying locus flk. *Journal of Bacteriology.* **190**(13): 4470-4477

Forster, B.M. & H. Marquis, (2012). Protein transport across the cell wall of monoderm Gram-positive bacteria. *Molecular Microbiology.* **84**(3): 405-413

Ge, Y., C. Li, L. Corum, C.A. Slaughter & N.W. Charon, (1998). Structure and expression of the FlaA periplasmic flagellar protein of *Borrelia burgdorferi*. *Journal of Bacteriology*. **180**(9): 2418-2425

Geisseler, D., W.R. Horwath, R.G. Joergensen & B. Ludwig, (2010). Pathways of nitrogen utilization by soil microorganisms–a review. *Soil Biology and Biochemistry*. **42**: 2058-2067

Gewirtz, A.T., T.A. Navas, S. Lyons, P.J. Godowski & J.L. Madara, (2001a). Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *The Journal of Immunology.* **167**: 1882-1885

Gewirtz, A.T., P.O. Simon Jr, C.K. Schmitt, L.J. Taylor, C.H. Hagedorn, A.D. O'Brien, A.S. Neish & J.L. Madara, (2001b). *Salmonella typhimurium* translocates flagellin across intestinal epithelia, inducing a proinflammatory response. *Journal of Clinical Investigation.* **107**(1): 99

Gobbetti, M., R.D. Cagno & M. De Angelis, (2010). Functional microorganisms for functional food quality. *Critical Reviews in Food Science and Nutrition* **50**(8): 716-727.

Gómez-Gómez, L. & T. Boller, (2000). FLS2: An LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Molecular Cell.* **5**: 1003-1011

Gómez-Gómez, L., G. Felix & T. Boller, (1999). A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*. *The Plant Journal.* **18**(3): 277-284

Goon, S., J.F. Kelly, S.M. Logan, C.P. Ewing & P. Guerry, (2003). Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the Cj1293 gene. *Molecular Microbiology.* **50**(2): 659-671

Gryllos, I., J.G. Shaw, R. Gavín, S. Merino & J.M. Tomás, (2001). Role of flm Locus in Mesophilic *Aeromonas* Species Adherence. *Infection and Immunity.* **69**(1): 65-74

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

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

Harshey, R.M., (2003). Bacterial motility on a surface: many ways to a common goal. *Annual Reviews in Microbiology*. **57**: 249-273

Hattermann, D. & S. Ries, (1989). Motility of *Pseudomonas syringae* pv. *glycinea* and its role in infection. *Phytopathology.* **79**: 284-289

Hayakawa, J. & M. Ishizuka, (2012). *Flagellar Glycosylation: Current Advances*. INTECH Open Access Publisher.

Hayashi, F., K.D. Smith, A. Ozinsky, T.R. Hawn, C.Y. Eugene, D.R. Goodlett, J.K. Eng, S. Akira, D.M. Underhill & A. Aderem, (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*. **410**: 1099-1103

Hazelbauer, G.L., J.J. Falke & J.S. Parkinson, (2008). Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends in Biochemical Sciences*. **33**(1): 9-19

Hazelbauer, G.L. & W.-C. Lai, (2010). Bacterial chemoreceptors: providing enhanced features to two-component signaling. *Current Opinion in Microbiology*. **13**(2): 124-132

Herrera, C.M., M.D. Koutsoudis, X. Wang & S.B. Von Bodman, (2008). *Pantoea stewartii* subsp. *stewartii* exhibits surface motility, which is a critical aspect of Stewart's wilt disease development on maize. *Molecular Plant-Microbe Interactions*. **21**(10): 1359-1370

Hirano, S.S. & C.D. Upper, (2000). Bacteria in the Leaf Ecosystem with Emphasis on *Pseudomonas syringae*—a Pathogen, Ice Nucleus, and Epiphyte. *Microbiology and Molecular Biology Reviews*. **64**(3): 624-653

Hizukuri, Y., S. Kojima, T. Yakushi, I. Kawagishi & M. Homma, (2008). Systematic Cys mutagenesis of FlgI, the flagellar P-ring component of *Escherichia coli*. *Microbiology*. **154**: 810-817

Hospodsky, D., J. Qian, W.W. Nazaroff, N. Yamamoto, K. Bibby, H. Rismani-Yazdi & J. Peccia, (2012). Human occupancy as a source of indoor airborne bacteria. *PLoS One*. **7**(4): e34867

Husen, E., (2013). Screening of soil bacteria for plant growth promotion activities in vitro. *Indonesian Journal of Agricultural Science*. **4**(1)

Hutcheson, S.W., (1998). Current concepts of active defense in plants. *Annual Review of Phytopathology*. **36**: 59-90

Ichinose, Y., F. Taguchi, M. Yamamoto, M. Ohnishi-Kameyama, T. Atsumi, M. Iwaki, H. Manabe, M. Kumagai, Q.T. Nguyen & C.L. Nguyen, (2013). Flagellin glycosylation is ubiquitous in a broad range of phytopathogenic bacteria. *Journal of General Plant Pathology*. **79**: 359-365

Ikeda, J.S., C.K. Schmitt, S.C. Darnell, P.R. Watson, J. Bispham, T.S. Wallis, D.L. Weinstein, E.S. Metcalf, P. Adams & C.D. O'Connor, (2001). Flagellar Phase Variation of *Salmonella enterica* Serovar *Typhimurium* Contributes to Virulence in the Murine Typhoid Infection Model but Does Not Influence *Salmonella*-Induced Enteropathogenesis. *Infection and Immunity*. **69**(5): 3021-3030

Irikura, V.M., M. Kihara, S. Yamaguchi, H. Sockett & R. Macnab, (1993). *Salmonella typhimurium* fliG and fliN mutations causing defects in assembly, rotation, and switching of the flagellar motor. *Journal of Bacteriology*. **175**(3): 802-810

Jacobs, J.L. & G.W. Sundin, (2001). Effect of solar UV-B radiation on a phyllosphere bacterial community. *Applied and Environmental Microbiology*. **67**(12): 5488-5496

Jeon, B., K. Itoh, N. Misawa & S. Ryu, (2003). Effects of quorum sensing on flaA transcription and autoagglutination in *Campylobacter jejuni*. *Microbiology and Immunology*. **47**(11): 833-839

Jones, C., M. Homma & R. Macnab, (1989). L-, P-, and M-ring proteins of the flagellar basal body of *Salmonella typhimurium*: gene sequences and deduced protein sequences. *Journal of Bacteriology*. **171**(7): 3890-3900

Jones, J.D. & J.L. Dangl, (2006). The plant immune system. *Nature*. **444**: 323-329

Josenhans, C. & S. Suerbaum, (2002). The role of motility as a virulence factor in bacteria. *International Journal of Medical Microbiology*. **291**: 605-614

Josenhans, C., L. Vossebein, S. Friedrich & S. Suerbaum, (2002). The neuA/flmD gene cluster of *Helicobacter pylori* is involved in flagellar biosynthesis and flagellin glycosylation. *FEMS Microbiology Letters.* **210**: 165-172

Kallmeyer, J., R. Pockalny, R.R. Adhikari, D.C. Smith & S. D'Hondt, (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences*. **109**: 16213-16216

Kim, Y.-K. & L.L. McCarter, (2000). Analysis of the polar flagellar gene system of *Vibrio parahaemolyticus*. *Journal of Bacteriology.* **182**(13): 3693-3704

Kinoshita, M., N. Hara, K. Imada, K. Namba & T. Minamino, (2013). Interactions of bacterial flagellar chaperone–substrate complexes with FlhA contribute to co-ordinating assembly of the flagellar filament. *Molecular Microbiology*. **90**(6): 1249-1261

Kojima, S., (2015). Dynamism and regulation of the stator, the energy conversion complex of the bacterial flagellar motor. *Current Opinion in Microbiology.* **28**: 66-71

Kojima, S. & D.F. Blair, (2004). The bacterial flagellar motor: structure and function of a complex molecular machine. *International Review of Cytology.* **233**: 93-134

Komeda, Y., K. Kutsukake & T. Iino, (1980). Definition of additional flagellar genes in *Escherichia coli* K12. *Genetics.* **94**: 277-290

Komoriya, K., N. Shibano, T. Higano, N. Azuma, S. Yamaguchi & S.I. Aizawa, (1999). Flagellar proteins and type III-exported virulence factors are the predominant proteins secreted into the culture media of *Salmonella typhimurium*. *Molecular Microbiology.* **34**(4): 767-779

Konkel, M.E., J.D. Klena, V. Rivera-Amill, M.R. Monteville, D. Biswas, B. Raphael & J. Mickelson, (2004). Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus. *Journal of Bacteriology.* **186**(11): 3296-3303

Kutsukake, K., Y. Ohya & T. Iino, (1990). Transcriptional analysis of the flagellar regulon of *Salmonella typhimurium*. *Journal of Bacteriology*. **172**(2): 741-747

Kutsukake, K., Y. Ohya, S. Yamaguchi & T. Iino, (1988). Operon structure of flagellar genes in *Salmonella typhimurium*. *Molecular and General Genetics MGG*. **214**: 11-15

Leclerc, G., S.P. Wang & B. Ely, (1998). A New Class of *Caulobacter crescentus* Flagellar Genes. *Journal of Bacteriology*. **180**(19): 5010-5019

Li, Z., F. Dumas, D. Dubreuil & M. Jacques, (1993). A species-specific periplasmic flagellar protein of *Serpulina* (Treponema) *hyodysenteriae*. *Journal of Bacteriology*. **175**(24): 8000-8007

Lindow, S.E., (1991). Determinants of epiphytic fitness in bacteria. In: Microbial ecology of leaves. Springer. pp. 295-314

Lindow, S.E. & M.T. Brandl, (2003). Microbiology of the phyllosphere. *Applied and Environmental Microbiology*. **69**(4): 1875-1883

Liu, B., B. Hu, Z. Zhou, D. Guo, X. Guo, P. Ding, L. Feng & L. Wang, (2012). A novel non-homologous recombination-mediated mechanism for *Escherichia coli* unilateral flagellar phase variation. *Nucleic Acids Research*. gks040

Lloyd, S.A., F.G. Whitby, D.F. Blair & C.P. Hill, (1999). Structure of the C-terminal domain of FliG, a component of the rotor in the bacterial flagellar motor. *Nature*. **400**: 472-475

Logan, S.M., (2006) Flagellar glycosylation–a new component of the motility repertoire? *Microbiology*. **152**: 1249-1262

Logan, S.M., J.F. Kelly, P. Thibault, C.P. Ewing & P. Guerry, (2002). Structural heterogeneity of carbohydrate modifications affects serospecificity of *Campylobacter* flagellins. *Molecular Microbiology*. **46**: 587-597

Lowenthal, A.C., M. Hill, L.K. Sycuro, K. Mehmood, N.R. Salama & K.M. Ottemann, (2009). Functional analysis of the *Helicobacter pylori* flagellar switch proteins. *Journal of Bacteriology*. **191**(23): 7147-7156

Macnab, R.M., (1992). Genetics and biogenesis of bacterial flagella. *Annual Review of Genetics*. **26**: 131-158

Macnab, R.M., (1999). The bacterial flagellum: reversible rotary propellor and type III export apparatus. *Journal of Bacteriology*. **181**(23): 7149-7153

Macnab, R.M., (2003). How bacteria assemble flagella. *Annual Reviews in Microbiology*. **57**: 77-100

Maleki, A., M.A. Jebelli, B. Hayati, H. Daraei & F. Gharibi, (2015). Microbial deactivation properties of generation 2 and 4 poly (amidoamine) dendrimers on common bacteria found in the aqueous environment. *Antimicrobial Resistance and Infection Control*. **4**(1): P35

Maruthamuthu, S., N. Muthukumar, M. Natesan & N. Palaniswamy, (2008). Role of air microbes on atmospheric corrosion. *Current Science*. **94**(3): 359-363

Mathews, M.A., H.L. Tang & D.F. Blair, (1998). Domain Analysis of the FliM Protein of *Escherichia coli*. *Journal of Bacteriology*. **180**(21): 5580-5590

Maurer, L.M., E. Yohannes, S.S. Bondurant, M. Radmacher & J.L. Slonczewski, (2005). pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. *Journal of Bacteriology*. **187**(1): 304-319

McCarter, L.L., (2001). Polar flagellar motility of the*Vibrionaceae*. *Microbiology and Molecular Biology Reviews*. **65**(3): 445-462

McDermott, P.F., F. Ciacci-Woolwine, J.A. Snipes & S.B. Mizel, (2000). High-affinity interaction between gram-negative flagellin and a cell surface polypeptide results in human monocyte activation. *Infection and Immunity.* **68**(10): 5525-5529

McFall-Ngai, M., M.G. Hadfield, T.C. Bosch, H.V. Carey, T. Domazet-Lošo, A.E. Douglas, N. Dubilier, G. Eberl, T. Fukami & S.F. Gilbert, (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences.* **110**(9): 3229-3236

Meindl, T., T. Boller & G. Felix, (2000). The bacterial elicitor flagellin activates its receptor in tomato cells according to the address–message concept. *The Plant Cell.* **12**(9): 1783-1794

Melotto, M., W. Underwood, J. Koczan, K. Nomura & S.Y. He, (2006). Plant stomata function in innate immunity against bacterial invasion. *Cell.* **126**: 969-980

Meng, F., C. Altier & G.B. Martin, (2013). *Salmonella* colonization activates the plant immune system and benefits from association with plant pathogenic bacteria. *Environmental Microbiology.* **15**: 2418-2430

Merritt, P.M., T. Danhorn & C. Fuqua, (2007). Motility and chemotaxis in *Agrobacterium tumefaciens* surface attachment and biofilm formation. *Journal of Bacteriology.* **189**(22): 8005-8014

Minamino, T. & R.M. Macnab, (1999). Components of the *Salmonella* flagellar export apparatus and classification of export substrates. *Journal of Bacteriology.* **181**(5): 1388-1394

Mobley, H., R. Belas, V. Lockatell, G. Chippendale, A.L. Trifillis, D.E. Johnson & J.W. Warren, (1996). Construction of a flagellum-negative mutant of Proteus mirabilis: effect on internalization by human renal epithelial cells and virulence in a mouse model of ascending urinary tract infection. *Infection and Immunity.* **64**(12): 5332-5340

Morohoshi, T., Y. Nakamura, G. Yamazaki, A. Ishida, N. Kato & T. Ikeda, (2007). The plant pathogen *Pantoea ananatis* produces N-acylhomoserine lactone and causes center rot disease of onion by quorum sensing. *Journal of Bacteriology.* **189**(22): 8333-8338

Namba, K. & F. Vonderviszt, (1997). Molecular architecture of bacterial flagellum. *Quarterly Reviews of Biophysics.* **30**(1): 1-65

Neill, S., R. Barros, J. Bright, R. Desikan, J. Hancock, J. Harrison, P. Morris, D. Ribeiro & I. Wilson, (2008). Nitric oxide, stomatal closure, and abiotic stress. *Journal of Experimental Botany.* **59**(2): 165-176

Nimchuk, Z., T. Eulgem, B.F. Holt Iii & J.L. Dangl, (2003). Recognition and response in the plant immune system. *Annual Review of Genetics.* **37**: 579-609

Nothaft, H. & C.M. Szymanski, (2010). Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology.* **8**: 765-778

O'Neil, H.S. & H. Marquis, (2006). *Listeria monocytogenes* flagella are used for motility, not as adhesins, to increase host cell invasion. *Infection and Immunity.* **74**(12): 6675-6681

Ogawa, R., R. Abe-Yoshizumi, T. Kishi, M. Homma & S. Kojima, (2015). Interaction of the C-terminal tail of FliF with FliG from the Na+-driven flagellar motor of *Vibrio alginolyticus*. *Journal of Bacteriology.* **197**(1): 63-72

Oliver, K.M., A.H. Smith & J.A. Russell, (2014). Defensive symbiosis in the real world–advancing ecological studies of heritable, protective bacteria in aphids and beyond. *Functional Ecology.* **28**: 341-355

Olsen, J.E., K.H. Hoegh-Andersen, J. Casadesús, J.T. Rosenkrantz, M.S. Chadfield & L.E. Thomsen, (2013). The role of flagella and chemotaxis genes in host pathogen interaction of the host adapted *Salmonella enterica* serovar *Dublin* compared to the broad host range serovar *S. Typhimurium*. *BMC Microbiology.* **13**: 67

Ormonde, P., P. Hörstedt, R. O'Toole & D.L. Milton, (2000). Role of motility in adherence to and invasion of a fish cell line by *Vibrio anguillarum*. *Journal of Bacteriology.* **182**(8): 2326-2328

Ottemann, K.M. & A.C. Lowenthal, (2002). *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. *Infection and Immunity.* **70**(4): 1984-1990

Ottemann, K.M. & J.F. Miller, (1997). Roles for motility in bacterial–host interactions. *Molecular Microbiology.* **24**(6): 1109-1117

Ozin, A.J., L. Claret, F. Auvray & C. Hughes, (2003). The FliS chaperone selectively binds the disordered flagellin C-terminal D0 domain central to polymerisation. *FEMS Microbiology Letters* **219**: 219-224.

Pandini, A., J. Kleinjung, S. Rasool & S. Khan, (2015). Coevolved Mutations Reveal Distinct Architectures for Two Core Proteins in the Bacterial Flagellar Motor. *PloS One.* **10**: e0142407

Parker, J.L., R.C. Lowry, N.A. Couto, P.C. Wright, G.P. Stafford & J.G. Shaw, (2014). Maf-dependent bacterial flagellin glycosylation occurs before chaperone binding and flagellar T3SS export. *Molecular Microbiology.* **92**(2): 258-272

Power, P. & M. Jennings, (2003). The genetics of glycosylation in Gram-negative bacteria. *FEMS Microbiology Letters.* **218**: 211-222

Ramos, H.C., M. Rumbo & J.-C. Sirard, (2004). Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends in Microbiology.* **12**(11): 509-517

Rashid, M.H. & A. Kornberg, (2000). Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences.* **97**(9): 4885-4890

Redford, A.J., R.M. Bowers, R. Knight, Y. Linhart & N. Fierer, (2010). The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environmental Microbiology.* **12**(11): 2885-2893

Reid, S.W., M.C. Leake, J.H. Chandler, C.-J. Lo, J.P. Armitage & R.M. Berry, (2006). The maximum number of torque-generating units in the flagellar motor of *Escherichia coli* is at least 11. *Proceedings of the National Academy of Sciences.* **103**(21): 8066-8071

Remus-Emsermann, M.N., R. Tecon, G.A. Kowalchuk & J.H. Leveau, (2012). Variation in local carrying capacity and the individual fate of bacterial colonizers in the phyllosphere. *The ISME Journal.* **6**: 756-765

Ridgway, H., M. Silverman & M. Simon, (1977). Localization of proteins controlling motility and chemotaxis in *Escherichia coli*. *Journal of Bacteriology.* **132**(2): 657-665

Rodríguez-Navarro, D.N., M.S. Dardanelli & J.E. Ruíz-Saínz, (2007). Attachment of bacteria to the roots of higher plants. *FEMS Microbiology Letters.* **272**(2): 127-136

Rosenberg, E., G. Sharon, I. Atad & I. Zilber-Rosenberg, (2010). The evolution of animals and plants via symbiosis with microorganisms. *Environmental Microbiology Reports.* **2**(4): 500-506

Ryan, P., R. Ross, G. Fitzgerald, N. Caplice & C. Stanton, (2015). Sugar-coated: exopolysaccharide producing lactic acid bacteria for food and human health applications. *Food & Function.* **6**: 679-693

Saijo-Hamano, Y., N. Uchida, K. Namba & K. Oosawa, (2004). In vitro characterization of FlgB, FlgC, FlgF, FlgG, and FliE, flagellar basal body proteins of *Salmonella*. *Journal of Molecular Biology.* **339**: 423-435

Samatey, F.A., K. Imada, F. Vonderviszt, Y. Shirakihara & K. Namba, (2000). Crystallization of the F41 fragment of flagellin and data collection from extremely thin crystals. *Journal of Structural Biology.* **132**: 106-111

Samatey, F.A., H. Matsunami, K. Imada, S. Nagashima, T.R. Shaikh, D.R. Thomas, J.Z. Chen, D.J. DeRosier, A. Kitao & K. Namba, (2004). Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. *Nature.* **431**: 1062-1068

Scharf, B.E., M.F. Hynes & G.M. Alexandre, (2016). Chemotaxis signaling systems in model beneficial plant–bacteria associations. *Plant Molecular Biology*. **90**(6): 1-11

Schirm, M., S. Arora, A. Verma, E. Vinogradov, P. Thibault, R. Ramphal & S. Logan, (2004). Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *Journal of Bacteriology.* **186**(9): 2523-2531

Schirm, M., I.C. Schoenhofen, S.M. Logan, K.C. Waldron & P. Thibault, (2005). Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins. *Analytical Chemistry.* **77**: 7774-7782

Schirm, M., E. Soo, A. Aubry, J. Austin, P. Thibault & S. Logan, (2003). Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Molecular Microbiology*. **48**(6): 1579-1592

Schmitt, C.K., J.S. Ikeda, S.C. Darnell, P.R. Watson, J. Bispham, T.S. Wallis, D.L. Weinstein, E.S. Metcalf & A.D. O'Brien, (2001). Absence of all components of the flagellar export and synthesis machinery differentially alters virulence of *Salmonella enterica* serovar *Typhimurium* in models of typhoid fever, survival in macrophages, tissue culture invasiveness, and calf enterocolitis. *Infection and Immunity*. **69**(9): 5619-5625

Schuhmacher, J.S., K.M. Thormann & G. Bange, (2015). How bacteria maintain location and number of flagella? *FEMS Microbiology Reviews*. 812-822

Sierro, F., B. Dubois, A. Coste, D. Kaiserlian, J.-P. Kraehenbuhl & J.-C. Sirard, (2001). Flagellin stimulation of intestinal epithelial cells triggers CCL20-mediated migration of dendritic cells. *Proceedings of the National Academy of Sciences*. **98**(24): 13722-13727

Silverman, M. & M. Simon, (1973). Genetic analysis of flagellar mutants in *Escherichia coli*. *Journal of Bacteriology*. **113**(1): 105-113

Silverman, M. & M. Simon, (1980). Phase variation: genetic analysis of switching mutants. *Cell*. **19**(4): 845-854

Sockett, H., S. Yamaguchi, M. Kihara, V.M. Irikura & R.M. Macnab, (1992). Molecular analysis of the flagellar switch protein FliM of *Salmonella typhimurium*. *Journal of Bacteriology*. **174**(3): 793-806

Sourjik, V. & N.S. Wingreen, (2012). Responding to chemical gradients: bacterial chemotaxis. *Current Opinion in Cell Biology*. **24**(2): 262-268

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

Sowa, Y. & R.M. Berry, (2008). Bacterial flagellar motor. *Quarterly Reviews of Biophysics*. **41**(2): 103-132

Stallmeyer, M.B., S.-I. Aizawa, R.M. Macnab & D.J. DeRosier, (1989). Image reconstruction of the flagellar basal body of *Salmonella typhimurium*. *Journal of Molecular Biology*. **205**: 519-528

Steege, D.A., (2000). Emerging features of mRNA decay in bacteria. *Rna*. **6**: 1079-1090

Stocker, B., (1949). Measurements of rate of mutation of flagellar antigenic phase in *Salmonella typhi-murium*. *Journal of Hygiene*. **47**: 398-413

Sun, L., M. Jin, W. Ding, J. Yuan, J. Kelly & H. Gao, (2013a). Posttranslational modification of flagellin FlaB in *Shewanella oneidensis*. *Journal of Bacteriology*. **195**: 2550-2561

Sun, Y., L. Li, A.P. Macho, Z. Han, Z. Hu, C. Zipfel, J.-M. Zhou & J. Chai, (2013b). Structural basis for flg22-induced activation of the *Arabidopsis* FLS2-BAK1 immune complex. *Science.* **342**: 624-628

Surridge, C., (2001). Flagellin shifts gear. *Nature Reviews Molecular Cell Biology.* **2**: 232-232

Taguchi, F., R. Shimizu, Y. Inagaki, K. Toyoda, T. Shiraishi & Y. Ichinose, (2003). Post-translational modification of flagellin determines the specificity of HR induction. *Plant and Cell Physiology.* **44**(3): 342-349

Taguchi, F., M. Yamamoto, M. Ohnishi-Kameyama, M. Iwaki, M. Yoshida, T. Ishii, T. Konishi & Y. Ichinose, (2010). Defects in flagellin glycosylation affect the virulence of *Pseudomonas syringae* pv. *tabaci* 6605. *Microbiology.* **156**: 72-80

Takeuchi, K., H. Ono, M. Yoshida, T. Ishii, E. Katoh, F. Taguchi, R. Miki, K. Murata, H. Kaku & Y. Ichinose, (2007). Flagellin glycans from two pathovars of *Pseudomonas syringae* contain rhamnose in D and L configurations in different ratios and modified 4-amino-4, 6-dideoxyglucose. *Journal of Bacteriology.* **189**(19): 6945-6956

Takeuchi, K., F. Taguchi, Y. Inagaki, K. Toyoda, T. Shiraishi & Y. Ichinose, (2003). Flagellin glycosylation island in *Pseudomonas syringae* pv. *glycinea* and its role in host specificity. *Journal of Bacteriology.* **185**(22): 6658-6665

Tanner, D.E., (2011). Flagellin Translocation and Flagellum Growth. Available from:. http://www.ks.uiuc.edu/Research/flagellum_growth/: Theoretical and Computational Biophysics Group, University of Illinois. [2014, June 18]

Tasteyre, A., M.-C. Barc, A. Collignon, H. Boureau & T. Karjalainen, (2001). Role of FliC and FliD Flagellar Proteins of *Clostridium difficile* in Adherence and Gut Colonization. *Infection and Immunity.* **69**(12): 7937-7940

Thibault, P., S.M. Logan, J.F. Kelly, J.-R. Brisson, C.P. Ewing & P. Guerry, (2001). Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *Journal of Biological Chemistry*. **276**(37): 34862-34870

Thomas, P., S. Kumari, G.K. Swarna & T. Gowda, (2007). Papaya shoot tip associated endophytic bacteria isolated from in vitro cultures and host-endophyte interaction in vitro and in vivo. *Canadian Journal of Microbiology*. **53**(3): 380-390

Tomich, M., C.A. Herfst, J.W. Golden & C.D. Mohr, (2002). Role of flagella in host cell invasion by *Burkholderia cepacia*. *Infection and Immunity*. **70**(4): 1799-1806

Tronick, S.R. & R.J. Martinez, (1971). Methylation of the flagellin of *Salmonella typhimurium*. *Journal of Bacteriology*. **105**(1): 211-219

Tsuda, K., M. Sato, J. Glazebrook, J.D. Cohen & F. Katagiri, (2008). Interplay between MAMP-triggered and SA-mediated defense responses. *The Plant Journal*. **53**: 763-775

Udikovic-Kolic, N., F. Wichmann, N.A. Broderick & J. Handelsman, (2014). Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. *Proceedings of the National Academy of Sciences*. **111**(42): 15202-15207

Ueno, T., K. Oosawa & S.-I. Aizawa, (1994). Domain structures of the MS ring component protein (FliF) of the flagellar basal body of *Salmonella typhimurium*. *Journal of Molecular Biology*. **236**: 546-555

van Asten, F.J., H.G. Hendriks, J.F. Koninkx & J.E. van Dijk, (2004). Flagella-mediated bacterial motility accelerates but is not required for *Salmonella* serotype *Enteritidis* invasion of differentiated Caco-2 cells. *International Journal of Medical Microbiology*. **294**(6): 395-399

Van Der Woude, M.W. & A.J. Bäumler, (2004). Phase and antigenic variation in bacteria. *Clinical Microbiology Reviews*. **17**(3): 581-611

Vatanyoopaisarn, S., A. Nazli, C.E. Dodd, C.E. Rees & W.M. Waites, (2000). Effect of flagella on initial attachment of *Listeria monocytogenes* to stainless steel. *Applied and Environmental Microbiology.* **66**(2): 860-863

Veluchamy, S., S.R. Hind, D.M. Dunham, G.B. Martin & D.R. Panthee, (2014). Natural variation for responsiveness to flg22, flgII-28, and csp22 and *Pseudomonas syringae* pv. *tomato* in heirloom tomatoes. *PloS One.* **9**(9): e106119

Voisin, S., R.S. Houliston, J. Kelly, J.-R. Brisson, D. Watson, S.L. Bardy, K.F. Jarrell & S.M. Logan, (2005). Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. *Journal of Biological Chemistry.* **280**: 16586-16593

Wang, L., K. Tsuda, M. Sato, J.D. Cohen, F. Katagiri & J. Glazebrook, (2009). Arabidopsis CaM binding protein CBP60g contributes to MAMP-induced SA accumulation and is involved in disease resistance against *Pseudomonas syringae*. *PLoS Pathogens.* **5**(2): e1000301

Whipps, J., P. Hand, D. Pink & G.D. Bending, (2008). Phyllosphere microbiology with special reference to diversity and plant genotype. *Journal of Applied Microbiology.* **105**: 1744-1755

Yamamoto, S. & K. Kutsukake, (2006). FljA-mediated posttranscriptional control of phase 1 flagellin expression in flagellar phase variation of *Salmonella enterica* serovar *Typhimurium*. *Journal of Bacteriology.* **188**(3): 958-967

Yao, R., D.H. Burr, P. Doig, T.J. Trust, H. Niu & P. Guerry, (1994). Isolation of motile and non-motile insertional mutants of *Campylobacter jejuni*: the role of motility in adherence and invasion of eukaryotic cells. *Molecular Microbiology.* **14**(5): 883-893

Young, G.M., D.H. Schmiel & V.L. Miller, (1999a). A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. *Proceedings of the National Academy of Sciences*. **96**(11): 6456-6461

Young, G.M., M.J. Smith, S.A. Minnich & V.L. Miller, (1999b). The *Yersinia enterocolitica* motility master regulatory operon, flhDC, is required for flagellin production, swimming motility, and swarming motility. *Journal of Bacteriology*. **181**(9): 2823-2833

Zampronio, C.G., G. Blackwell, C.W. Penn & H.J. Cooper, (2011). Novel glycosylation sites localized in *Campylobacter jejuni* flagellin FlaA by liquid chromatography electron capture dissociation tandem mass spectrometry. *Journal of Proteome Research.* **10**: 1238-1245

Zeng, H., A.Q. Carlson, Y. Guo, Y. Yu, L.S. Collier-Hyams, J.L. Madara, A.T. Gewirtz & A.S. Neish, (2003). Flagellin is the major proinflammatory determinant of enteropathogenic *Salmonella*. *The Journal of Immunology*. **171**: 3668-3674

Zeng, W., M. Melotto & S.Y. He, (2010). Plant stomata: a checkpoint of host immunity and pathogen virulence. *Current Opinion in Biotechnology*. **21**(5): 599-603

Zhou, M., Y. Yang, P. Chen, H. Hu, P.R. Hardwidge & G. Zhu, (2015). More than a locomotive organelle: flagella in *Escherichia coli*. *Applied Microbiology and Biotechnology*. **99**(8): 8883-8890

Zieg, J., M. Hilmen & M. Simon, (1978). Regulation of gene expression by site-specific inversion. *Cell*. **15**(1): 237-244

1.8 Figures and tables

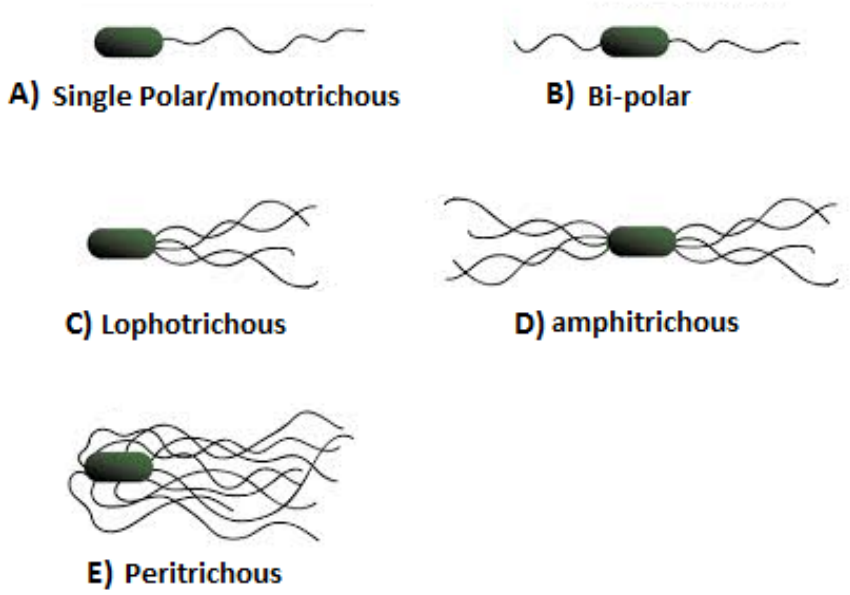


Figure 1. The different flagellations of bacteria. A) Single polar/monotrichous, B) bi-polar, C) lophotricous, D) amphitrichous and E) peritrichous.

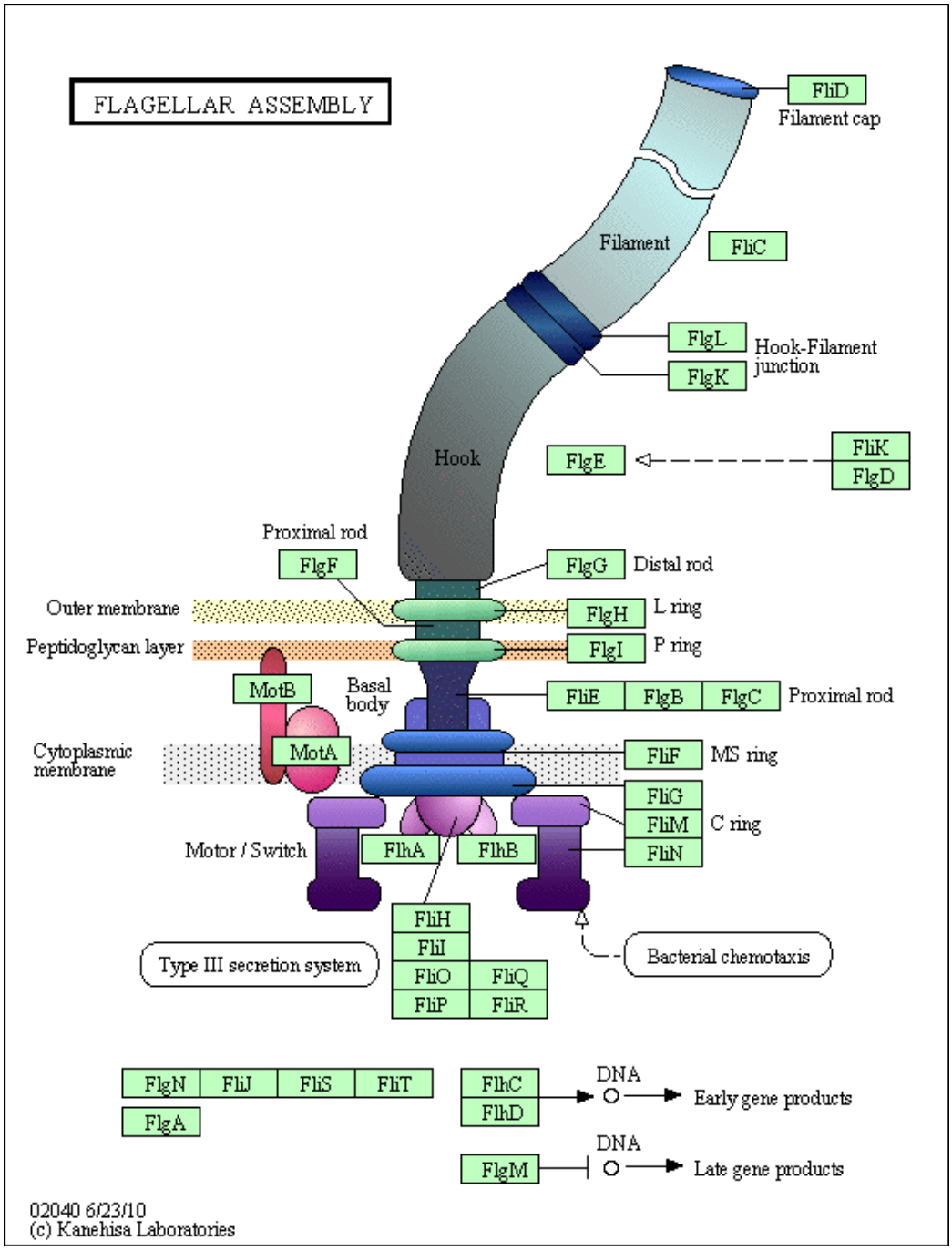


Figure 2. The model structure of the flagellar motor in gram-negative bacteria (Hayakawa and Ishizuka, 2012).

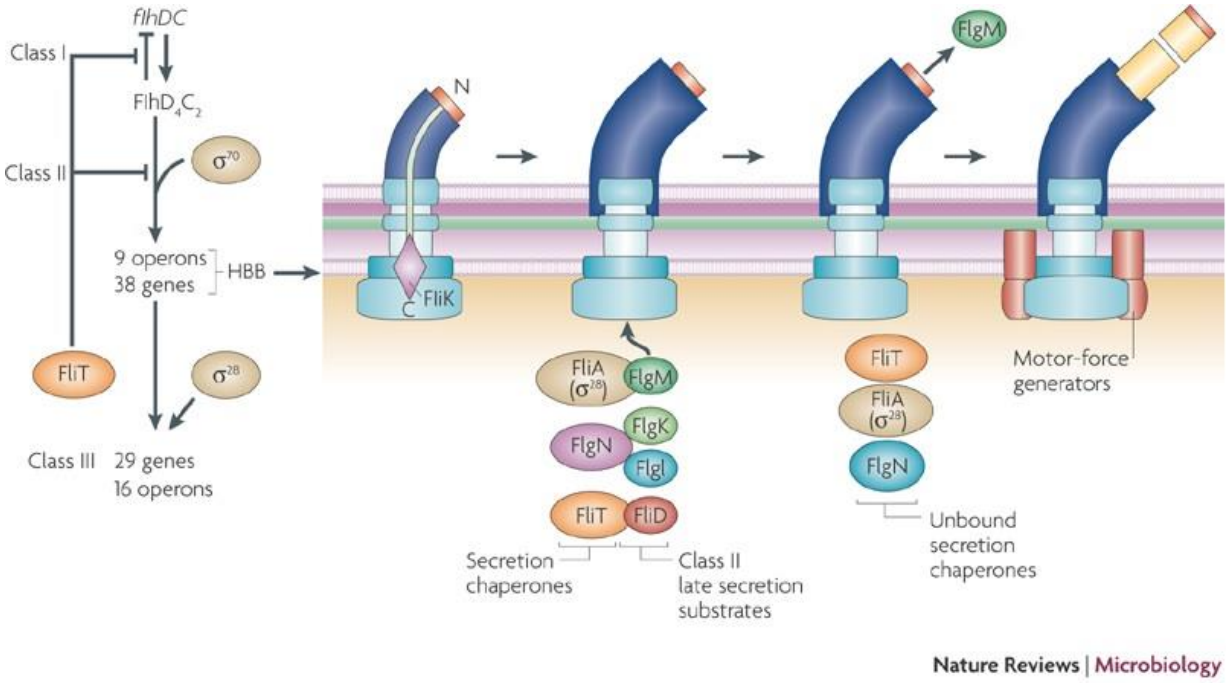
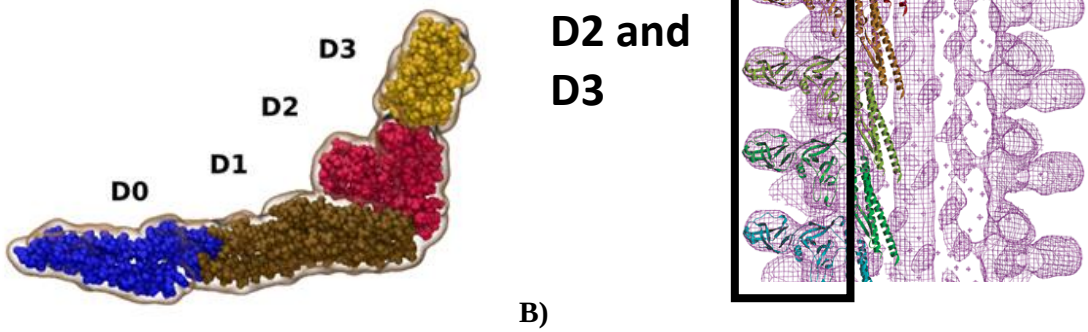


Figure 3. Assembly and regulation of the flagellum (Chevance & Hughes, 2008).



A) **B)**

Figure 4. **(A) A flagellin protein**. Indicating the different domains present; D0, D1, D2 and D3 (Tanner, 2011). **(B) The flagellar filament**. Indicating the orientation of the flagellin protein and indicating that the flagellin glycosylation takes place at the D2 and D3 domain of the flagellin protein. As the D0 and D1 domains are highly conserved and he D2 and D3 domains are hyper-variable (Surridge, 2001).

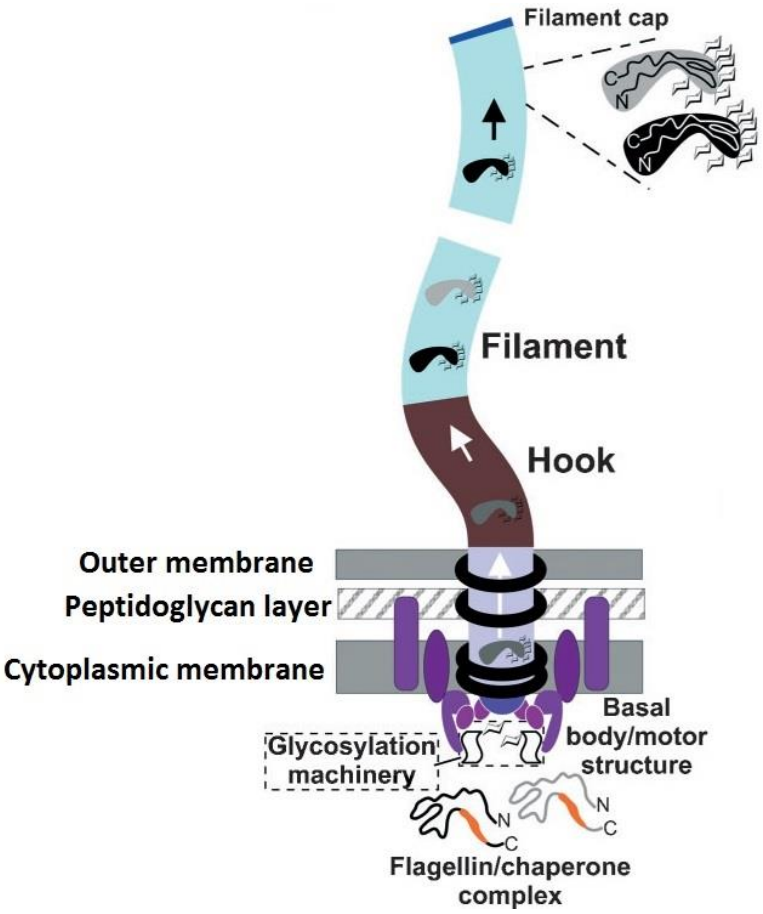


Figure 5. Assembly of the flagellin glycosylation. The O-glycosylation of the flagellin occurs either in the basal body environment, as the proteins and glycans are being exported through the type III secretion system channel, or at the cytoplasmic membrane close to the basal body (Logan, 2006). These glycan chains are linked onto serine and threonine residues of the flagellin surface exposed domain by a glycosyltransferase that is located in the FGI (Hayakawa & Ishizuka, 2012). Glycosylated flagellin are then secreted via the type III secretion system channel, through the basal body and hook till the filament tip, where they are arranged into the growing filament with the assistance of the filament cap (FliD).

CHAPTER TWO

COMPARATIVE GENOMIC ANALYSES OF THE FLAGELLIN GLYCOSYLATION ISLANDS OF SEVENTEEN *PANTOEA ANANATIS* STRAINS

2.1 Abstract

Pantoea ananatis is a plant pathogenic enterobacterium that infects a broad range of agronomic crops. The pathogenicity mechanisms of this bacterium remain largely unknown. However, flagella have recently been shown as essential for motility and consequently play an indispensable role in the infection process. As the main structural protein of the flagellum, flagellin, is highly immunogenic, it can be recognised by the host and result in the launching of concerted host defence responses. Some bacteria have evolved means of avoiding this phenomenon, including glycosylation of the flagellin protein, essentially masking it from host detection. A comparative genomic analysis of seventeen *P. ananatis* strains for which genome sequences are available revealed the presence of flagellin glycosylation islands (FGIs) located adjacent to the flagellin (*fliC*) genes in all seventeen strains. These FGIs can be classified into four distinct types, all with common and unique genes present. The islands encode for proteins involved in glycosylation of the flagellin with distinct sugars, and distinct acetyl, methyl and formyl side branches. The presence of a FGI could explain the broad range of hosts that *P. ananatis* can infect without host detection.

2.2 Introduction

Pantoea ananatis is a Gram-negative, facultatively anaerobic, rod-shaped, peritrichously flagellated member of the family *Enterobacteriaceae* (Walcott *et al.*, 2002, Coutinho & Venter, 2009). It was first discovered on pineapples in the Philippines, where it caused brown rot of the pineapple fruitlets (Serrano, 1928). Since then it has been isolated from a wide range of environments, in particular as part of the epi- and endophytic community on a broad range of plants. There has been particular interest in *P. ananatis* as it can cause disease symptoms on a wide range of plant hosts, including onions (Gitaitis & Gay, 1997, Walcott *et al.*, 2002), melons (Wells *et al.*, 1987), pineapples (De Maayer *et al.*, 2011), maize (Paccola-Meirelles *et al.*, 2001, Bomfeti *et al.*, 2007), rice (Cother *et al.*, 2004), E*ucalyptus* (Coutinho *et al.*, 2002), Sudan-grass (Walcott *et al.*, 2002), and buckwheat (Walcott *et al.*, 2002). Furthermore, a case of clinical disease has been attributed to this species (De Baere *et al.*, 2004). *P. ananatis* has also been found associated with insect hosts, including larvae of mulberry pyralid (*Glyphodes duplicalis*), silkworms (*Bombyx mori*) and brown plant hoppers (*Nilaparyata lugens*)(Watanabe *et al.*, 1996, Watanabe & Sato, 1999).

Phytopathogenic strains of *P. ananatis* cause a variety of symptoms on their host plants including blight (forming lesions, that withers and then kills certain parts of the infected plant), die-back (continual death of the plant from the tips of the shoots), rot and decay (weakening, decomposition and breakdown of plant tissue)(Coutinho *et al.*, 2002). Some of these wide spectrum diseases include centre rot of onions (Bomfeti *et al.*, 2007), brown rot of pineapples (Coutinho & Venter, 2009), stalk rot of rice (Cother *et al.*, 2004), leaf blight in *Eucalyptus* (Coutinho *et al.*, 2002) and necrotic spot and streaking of maize (Paccola-Meirelles *et al.*, 2001) have been observed. While there is some knowledge of the epidemiology, ecology and host range of *P. ananatis*, its pathogenicity determinants remain poorly understood (Coutinho & Venter, 2009, De Maayer *et al.*, 2014). What has been observed is that typical pathogenicity determinants observed in closely related phytopathogens, such as Type II, III and IV secretion systems, phytotoxins and phytohormones, are absent from *P. ananatis* (De Maayer *et al.*, 2011). By studying the genome sequences of *P. ananatis*, several putative pathogenicity determinants have been identified, including lipopolysaccharides, exopolysaccharides, Type VI secretion systems and an ice-nucleation protein (De Maayer *et al.*, 2011). The genes coding for the flagella have been elucidated from the genome of *P. ananatis* and the flagella were also shown to be crucial for symptom development on onion seedlings (Weller-Stuart, 2015).

The flagellum is a filamentous organ located in the bacterial membrane protruding to the exterior of the cell. Its main function among phytopathogenic bacteria include motility, attachment, pathogenicity and host defence avoidance (Power & Jennings, 2003, Takeuchi *et al.*, 2003, Herrera *et al.*, 2008). The main structural protein of the flagellar filament, flagellin (FliC), is highly immunogenic, being recognized by plant and animal hosts and resulting in the initiation of host defence mechanisms in response to pathogen invasion (Felix *et al.*, 1999, Zipfel & Felix, 2005, Cai *et al.*, 2011, Sun *et al.*, 2013a). As a result, pathogens have evolved a number of flagellin-related means to avoid host detection. Some bacteria encode multiple, antigenically distinct copies of FliC, which can be expressed interchangeably, to avoid host detection and defence response (Bonifield & Hughes, 2003, Hayakawa & Ishizuka, 2012). Another mechanism to avoid host flagellin recognition and response is through the post-translational modification of the flagellin protein with a glycan chain (Power & Jennings, 2003). This modification process is known as flagellin glycosylation (Logan, 2006). Aside from host defence avoidance, several distinct roles have been ascribed to

flagellin glycosylation. In some cases flagellin glycosylation is essential for motility and flagellar synthesis and stability. Other reported functions include adhesion, biofilm formation, swarming motility and invasion, expression of the lipopolysaccharide O-antigen, virulence and host specificity (Takeuchi *et al.*, 2003, Hayakawa & Ishizuka, 2012). Flagellin glycosylation has been observed and studied in several phytopathogens, including *Pseudomonas syringae* and *Ralstonia solanacearum* (Logan, 2006, Takeuchi *et al.*, 2007, Hayakawa & Ishizuka, 2012). Glycosylation in *Ps. syringae* pv. *glycinea* and *Ps. syringae* pv. *tabaci* 6605 has been observed to increase the hydrophilicity of the flagella improving its motility (Takeuchi *et al.*, 2003). It has been hypothesized that these glycosylations stabilize the filament structure, improve swarming, swimming speed and provide smooth swimming and virulence (Arora *et al.*, 2005, Taguchi *et al.*, 2006, Taguchi *et al.*, 2008). The lack of these genes would result in the bacterium being non-motile as the flagella will not be assembled due to some of the flagellar expression being incomplete. Knock-out mutagenesis of flagellin glycosylation genes performed in *Ps. syringae* resulted in reduced virulence (Takeuchi *et al.*, 2003, Taguchi *et al.*, 2010). Furthermore in 2003, Takeuchi *et al*., observed that flagellin glycosylation was mainly responsible for the recognition of the plant defence response mechanisms and that this could possibly be the mechanism that determines the compatibility of host specificity (Takeuchi *et al.*, 2003).

To date, flagellin glycosylation has not been observed in the peritrichously flagellate members of the family *Enterobacteriaceae* (Merino Montero & Tomàs Magaña, 2014). Here, by means of genomic and comparative genomic approaches, we have characterized the flagellin glycosylation loci in seventeen strains of *P. ananatis in silico*. The genes responsible for flagellin glycosylation are located on a genomic island within a flagellum biosynthetic locus, termed the flagellin glycosylation island (FGI). The *P. ananatis* FGIs are highly versatile, coding for distinct flagellin glycans and the effects of these distinct flagellin glycans in terms of the biology of *P. ananatis* are discussed.

2.3 Materials and methods

2.3.1 Identification of flagellin glycosylation islands in the *P. ananatis* genomes

The genomes of seventeen *P. ananatis* strains, isolated from distinct ecological niches and geographical locations have been sequenced (Table 1). The gene regions stretching from *fli*T to *fli*Z, encompassing the FGIs, were extracted from the seventeen genomes. The gene regions were structurally annotated using FgenesB (Tyson *et al.*, 2004). Further standardization of the protein coding sequence (CDS) sets was undertaken by means of localized BlastP and BlastN analyses, using Bioedit v. 7.0.9.0 (Hall, 1999). The G+C contents of the FGI regions were also determined using Bioedit v 7.0.9.0 (Hall, 1999).

2.3.2 Ortholog identification between the P. ananatis FGI regions

The CDS sets in the FGI region of each strain were pair-wise compared using localized BlastP analyses with Bioedit v. 7.0.9.0 (Hall, 1999). Orthologs were assumed for the CDSs sharing > 75% amino acid identity over 70% of the alignment length. Average amino acid values were calculated for the orthologous CDS sets by the formula: [SUM (amino acid identity x alignment length for each CDS)/100] / [SUM (alignment lengths for orthologous CDSs)] * 100.

2.3.3 In silico analyses of the FGI CDS sets

The CDS sets were functionally annotated by BlastP comparison of the CDSs against the NCBI non-redundant (nr) protein database and identification of the closest Blast Hits. The proteins were further compared against the Conserved Domain Database (CDD) using the Batch CD-search program (Marchler-Bauer & Bryant, 2004). Glycosyl transferases were classified to the family level by comparison against the Carbohydrate Active enzymes (CAZY) database using the dbCAN web-server (Yin *et al.*, 2012, Lombard *et al.*, 2014).

2.3.4 Phylogenetic analyses and construction of an FGI typing dendrogram

Phylogenies were constructed on the basis of the concatenated gene sequences of four commonly used housekeeping markers, namely AtpD (ATP synthase subunit beta), GyrB (DNA gyrase subunit beta), InfB (translation initiation factor IF-2) and RpoB (RNA polymerase β-subunit). The concatenated sequences were aligned using the MAFFT v. 7 alignment server (Katoh & Standley, 2013) with default parameters. Neighbour-joining trees were constructed with the Molecular Evolutionary Genetics Analysis (MEGA) v. 5.0.3 software package (Kumar *et al.*, 2008), using the Tamura and Nei model, complete gap deletion and bootstrap analysis (n=1000). Using the same approach a neighbour-joining tree was constructed on the basis of the four copies of the flagellin (*fliC*) genes upstream of the FGI regions in the complete genome sequences of *P. ananatis* strains AJ13355, LMG20103 and PA13. A dendrogram was constructed on the basis of the presence/absence of orthologs for each of the FGI proteins among the seventeen strains. Present orthologs were scored with a 1, while absent orthologs were scored with a 0. The presence/absence matrix was used to generate a distance matrix using Bionumerics v. 6.6 (Applied Maths N.V., Belgium) with the parameters: absolute values and Pearssons correlation. The distance matrix was used to generate an Unweighed Pair Group Method with Arithmetic Mean (UPGMA) dendrogram using Phylip v. 3.69 (Felsenstein, 2002). The FGIs were categorized into their distinct types on the basis of the similarity cut-off values of 50%.

2.4 Results and discussion

2.4.1 General properties of the *P. ananatis* strains FGI

The complete and draft genomes of seventeen *P. ananatis* strains (Table 1) have been sequenced, and the region from the *fliT* to the *fliZ* gene was extracted. All seventeen of the compared strains incorporate an FGI region (Table 2, Figure 1). At the 5’ end the FGIs are flanked by genes coding for the flagellar assembly and export proteins FliT and FliS, as well as the filament cap protein FliD and flagellin protein (FliC)(Figure 1). The 3’ flanking region incorporates genes coding for the flagellar biosynthesis protein FliZ and the RNA polymerase sigma factor FliA. The universal presence of FGIs among the *P. ananatis* strains is in contrast to *Pseudomonas aeruginosa*, where this region incorporates an FGI only in some strains

(Dasgupta *et al.*, 2003). The FGI region of the seventeen *P. ananatis* strains range in size from 7.8 to 22.7 kilobases with a G+C content ranging from 41.52 to 48.22% (with an average G+C content of 46.24%)(Table 2). This is typically 7.31% below the average G+C content of the genome, suggesting that this region has been derived through horizontal gene. Between six and twenty proteins are encoded in the *P. ananatis* FGIs (Table 2). The differences in size and number of proteins encoded within the FGI regions indicate extensive variability among the *P. ananatis* strains. This has likewise been observed among strains of *Ps. aeruginosa* as well as species in the genera *Campylobacter*, *Aeromonas*, and *Helicobacter* (Table 4)(Arora *et al.*, 2001, Logan, 2006).

2.4.2 The *P. ananatis* FGIs can be clustered into four distinct types

The CDS sets in the FGI of the seventeen *P. ananatis* strains were pair-wise compared using localized BlastP analysis, where orthologs were identified between the compared strains on the basis of the cut-off values described in section 2.3.2. A total of twenty-seven distinct proteins are encoded in the FGIs of the seventeen *P. ananatis* strains (Table 3). Of these, only three are conserved among all the compared strains, namely RmlD (dTDP-4-dehydrorhamnose reductases), Edn (an HNH endonuclease) and the hypothetical protein Hyp1. These proteins are, however, well-conserved sharing between 69.43 and 100 % amino acid identity among the seventeen compared strains. The presence of a homing type endonuclease (Edn) provides further evidence of horizontal acquisition of the *P. ananatis* FGIs. These homing endonucleases are capable of their own integration and replication in a host genome (Yahara *et al.*, 2009). A further twenty-three proteins are shared by two or more strains, while one protein Glt4, a predicted glycosyltransferase, is unique to one strain, *P. ananatis* SD-1. High levels of sequence identity can be observed between orthologs for each of the twenty-six distinct FGI proteins shared by two or more strains (Table 3). A similarity matrix was constructed based on the presence/absence of the 27 CDSs that were encoded on the FGIs. This matrix was then used to create a UPMGA dendrogram to reflect the similarity values between each of the compared FGI strains. A similarity cut-off value of 50% was used and it was observed that the FGIs could be clustered into four distinct types, Type I-IV (Figure 2). Type I FGIs are most common, with nine *P. ananatis* strains carrying this FGI. Two and five strains, respectively, contain Type II and III FGIs. Only one strain, *P. ananatis* SD-1 has a Type IV FGI. The UPMGA dendrogram was compared to the neighbour-joining

phylogenetic tree of the four concatenated housekeeping markers (Figure 2). In general, poor correlation could be observed between the FGI Type and strain phylogeny. This is likely due to the limited variability among strains of *P. ananatis* in the house-keeping markers. Among fifteen of the compared strains (with the exception of SD-1 and PaMB1), on average only 29 polymorphisms could be observed between the concatenated alignments for each pair of strains (alignment length = 10.516 nucleotides). The remaining two strains consistently grouped separate from the remaining fifteen strains for the four house-keeping markers as well as the concatenated gene/protein sequences. This suggests that *P. ananatis* PaMB1 and SD-1 may represent a distinct subspecies of *P. ananatis* or even a distinct *Pantoea* species. This may account for the distinct FGI (Type IV) observed for *Pantoea* sp. SD-1. On the other hand, *Pantoea* sp. PaMB1 clusters with four other *P. ananatis* strains on the basis of the FGI Type (Type III), suggesting that horizontal gene exchange between the Type III FGI *P. ananatis* strains and PaMB1 may have taken place.

2.4.3 The *P. ananatis* strains FGI carry genes for several distinct sugar biosynthetic pathways

Four distinct glycosyltransferases (Glt1-4) are encoded within the FGIs of the seventeen *P. ananatis* strains (Table 3 and 5). The Glt1 protein is found in FGI Types I, II and IV, Glt2 in FGI Type I, while Glt3 and Glt4 are found in FGI Type III and Type IV, respectively. Using the dbCAN Blast tool the glycosyltransferases were classified according to their Glycosyl Transferase families (Coutinho *et al.*, 2003, Yin *et al.*, 2012). The glycosyltransferases *glt*1, *glt*2, *glt*3 and *glt*4 are all part of the GT2 family which is known to have glycosyltransferase activity with a wide range of sugars (Coutinho *et al.*, 2003). Therefore the type of sugars attached to the *P. ananatis* flagellin proteins cannot solely be determined on the basis of the glycosyltransferases (Henrissat & Davies, 2000).

All four FGI Types code for the rhamnose sugar protein, dTDP-dehydrorhamnose reductase RmlD (Table 3 and 5). The RmlD protein is usually accompanied by the proteins: glucose-1-phosphate thymidylyltransferase RmlA, thymidine diphosphate (dTDP)-glucose 4,6 dehydratase RmlB and dTDP-4-dehydrorhamose 3,5-epimerase RmlC (Giraud & Naismith, 2000, Li & Reeves, 2000). Together these four proteins would form part of the rhamnose

pathway which catalyses the sequential conversion of dTDP-D-glucose to dTDP-L-rhamnose (Giraud *et al.*, 2000, Giraud & Naismith, 2000, Li & Reeves, 2000). The rhamnose pathway is initiated by RmlA, which transfers a thymidyl monophosphate nucleotide to glucose-1-phosphate to form dTDP-glucose (Giraud & Naismith, 2000). RmlB is then responsible for the catalyses of the oxidation of the C4 hydroxyl group of dTDP-glucose, followed by hydration in order to form dTDP-4-keto 6-deoxy-D-glucose (Giraud & Naismith, 2000). RmlC catalyses an infrequent double epimerase reaction occurring at positions C3 and C5 of the dTDP-4-keto 6-deoxy-D-glucose (Giraud & Naismith, 2000). Finally RmlD reduces the C4 keto function in dTDP-4-keto 6-deoxy-D-glucose to produce the end product of the pathway, dTDP-L-rhamnose (Giraud & Naismith, 2000).

RmlB-coding genes are absent from the FGIs of all seventeen *P. ananatis* strains. Intact copies of the *rmlC* and *rmlA* genes are only present in FGI Types I and IV. In the FGI Type II and III *P. ananatis* strains, there is a single 104 amino acid reading frame with the 5'end sharing orthology with the N-terminal 42 amino acids of the RmlA protein, while the 72 C-terminal amino acid residues share orthology with the C-terminal of the RmlC protein. The reading frames for these proteins thus seem to be subject to a deletion event, to generate a single reading frame. It is uncertain if the truncated, combined protein is functional, and this would have to be determined experimentally. It should be noted that in the O-antigen biosynthetic locus of all seventeen compared *P. ananatis* strains, intact copies of *rmlA*, *rmlB* and *rmlC* are present, and it is possible that these complement the defects in rhamnosyl biosynthesis (Chiang & Mekalanos, 1999).

In addition to rhamnosyl biosynthetic enzymes, several other biosynthetic enzymes are encoded within the *P. ananatis* FGIs. These enzymes include 3-oxoacyl-ACP reductases, transketolases, long chain fatty acid CoA ligases, Acyl-protein synthases, Acyl-CoA reductases, and oxidoreductases (Table 3 and 5). Two 3-oxoacyl ACP reductase proteins, FabG1 and FabG2 are encoded in all nine FGI Type I strains. This protein is part of the β-ketoacyl-(ACP) reductase family and catalyzes the first step in the cycle of the fatty acid synthesis (Sheldon *et al.*, 1992, Zaccai *et al.*, 2008). The long chain fatty acid CoA ligase, FadD is also present in all nine FGI Type I strains. FadD is an outer membrane bound fatty acid that assists in transport (DiRusso & Black, 2004). In *E. coli* it is hypothesized that FadD

is responsible for the abstraction of fatty acids through the activation of CoA thioesters (Klein *et al.*, 1971). Furthermore, these strains encode orthologs of the Acyl-protein synthase LuxE and LuxC Acyl-CoA reductase, which are also involved in fatty acid biosynthesis (Soly & Meighen, 1991, Lin *et al.*, 1993). The presence of these fatty acid biosynthetic genes suggests that in addition to a predicted rhamnosyl sugar in the flagellin glycan, this sugar chain may further be substituted with fatty acids and hence in the Type I FGI strains the flagellin may be modified with a liposaccharide (Edwards *et al.*, 2005).

Two transketolase enzymes, TktA1 and TktA2, are also encoded in all FGI Type I strains. These enzymes play a role in the catalysis of the transfer of a two-carbon unit from ketose phosphates to aldose phosphates in the presence of Mg^2 and provides a link between glycolysis and the pentose phosphate pathway (Kochetov *et al.*, 1975, Kovina & Kochetov, 1998, Schneider & Lindqvist, 1998). The second transketolase differs from the first in that it possesses a pyrimidine binding domain of 1-deoxy-D-xylulose-5-phosphate synthase, and requires the cofactor thiamine diphosphate and Ca^2 (Jordan, 2003). The role of the transketolases in flagellin glycan synthesis, transfer or attachment remains to be elucidated. An NAD-binding Rossman fold oxidoreductase (MviM) is present in both FGI Type II strains (Kingston *et al.*, 1996). This enzyme catalyses oxidation-reduction reactions, but its role in flagellin glycosylation remains unclear. The long chain fatty acid CoA ligase, FadD was present in all nine strains of FGI Type I. FadD is an outer membrane bound fatty acid that assists in transport (Klein *et al.*, 1971, DiRusso & Black, 2004).

2.4.4 The *P. ananatis* FGIs show evidence of further glycan modifications

Apart from the sugar and fatty acid biosynthetic and glycan transfer proteins, other proteins involved in modification of the flagellin glycan chain have been observed. These modifications include the addition of amino-, acetyl-, methyl-, and formyl- groups to the sugar backbone (Table 3 and 5)(Nothaft & Szymanski, 2010). To date the functions of all these glycan modifications are largely unknown (Nothaft & Szymanski, 2010). An acetyltransferase in the flagellar region of *Xanthomonas oryzae* has been observed to be essential for flagellin glycosylation as well as playing a role in motility, exopolysaccharide production and biofilm formation (Li et al. 2015). The genes involved in the incorporation of

an acetamidino group in the N-glycan of the flagellum of *Methanococcus maripaludis* have also been identified (Jones et al. 2012).

An aminotransferase (*degT*) is encoded in the FGIs of sixteen *P. ananatis* strains (all except the Type IV FGI strain *P. ananatis* SD-1). Restricted to the nine Type I FGI strains is a gene coding for an acetyltransferase (*maa*). An S-adenosylmethionine-dependent methyltransferase of the FkbM family (*fbk*M1) is present in seven of the nine Type I FGI strains. A distinct copy (FkbM2 38.96 % average amino acid identity to FkbM1) is present in the FGI of both Type II FGI strains. Furthermore a predicted formyltransferase (*vio*F) is present only in the five Type III FGI strains. A second distinct putative formyltransferase (*car*B) is present in both Type II FGI strains. The presence of these enzymes in the *P. ananatis* FGIs suggests that further modifications of the flagellin glycan chain occurs frequently and is highly variable. These glycan modifications are thus expected to add to the complexity and variability of the flagellin glycan in *P. ananatis* strains.

2.4.5 *P. ananatis* strains vary in both the number and type of flagellin genes

Comparative analysis of the seventeen *P. ananatis* strains showed that in six strains the FGI regions were flanked by three flagellin genes (*fli*C), while the other eleven strains were flanked by four *fli*C genes (Figure 1). Two of the *fliC* gene copies, *fliC2* and *fliC4*, were universal to all seventeen compared strains, with the encoded flagellin proteins sharing 81.11 and 70.84 % average amino acid identity among the compared strains. The *fliC1* gene was present in fourteen of the compared *P. ananatis* strains, and the encoded proteins shared 73.48% average amino acid identity. The *fliC3* gene was present in sixteen strains, with an average amino acid identity of 75.67 % for the encoded FliC3 protein. It should be noted that, due to the high level of sequence conservation among the *fliC* gene sequences, these genes frequently result in misassemblies, which are difficult to resolve without PCR amplification and Sanger sequencing approaches. As such, gaps could be observed occurring in the *fliC* region among those *P. ananatis* strains for which draft genomes are available. Four copies of *fliC* could be observed among three strains for which complete genome sequences are available.

A neighbour-joining tree was constructed using the gene sequences of the four copies of *fliC* for three *P. ananatis* strains for which complete genomes are available, namely AJ13355, LMG 20103 and PA13 (Figure 3). The protein sequences for three of the *fliC* genes cluster together (*fliC1*, *fliC2* and *fliC3*), while *fliC4* clusters separately from the other *fliC* copies. Relatively high levels of amino acid identity (ranging between 88 and 100 %) could be observed for the FliC2 and FliC4 proteins, suggesting these are well conserved. Lower homology could be observed between the FliC1 and FliC3 proteins with the other FliC copies, while the FliC1 and FliC3 proteins of the three compared strains share 73.48 % and 75.67 % average amino acid identity, respectively (Figure 3). While the protein sequences of the *fli*C2 and *fli*C4 genes of the three strains cluster together in distinct clades, the FliC3 protein is more variable. As such the *P. ananatis* AJ13355 FliC3 protein clusters with the FliC2 copy of the same strain. Similarly, the *fliC3* gene of *P. ananatis* PA13 clusters with its FliC1 copy, while the *P. ananatis* LMG 20103 FliC3 protein clusters with its FliC2 ortholog. This suggests that some gene duplications may have occurred within the distinct strains after diversification.

Multiple copies of flagellin genes have also been described in other bacteria, including *Escherichia coli*, *Campylobacter* sp., *Vibrio* sp., *Salmonella* sp., and *Pseudomonas* sp. (Déziel *et al.*, 2001, Ikeda *et al.*, 2001, McCarter, 2001, Soutourina & Bertin, 2003, Achouak *et al.*, 2004, Schirm *et al.*, 2004, Logan, 2006, Liu *et al.*, 2012, Zhou *et al.*, 2015). The significance of these distinct *fliC* copies is that they can be expressed at different times and under different conditions in order to generate genetically and antigenically distinct flagellar filaments, known as flagellar phase variation (Ikeda *et al.*, 2001, Bonifield & Hughes, 2003, Aldridge *et al.*, 2006). Through this process, recognition by host receptors of this antigenic protein, can be avoided and subsequently host defence mechanisms are not activated (Bonifield & Hughes, 2003).

The highly antigenic nature of the flagellin protein can be linked to the flg22 epitope (Felix *et al.*, 1999, Jones & Dangl, 2006), a highly conserved domain with a short amino acid sequence (QRLSTGSRINSAKDDAAGLQIA) found at the N-terminal of the protein sequence (Sun *et al.*, 2013b). It forms part of the D0 domain of flagellin in the flagella filament found in the inner core of the filament (Sun *et al.*, 2013b). This epitope is recognised

by plant receptors, such as the FLR2 receptor in *Arabidopsis* (Gómez-Gómez & Boller, 2000), on the cell wall, triggering host defence responses (Felix *et al.*, 1999). A CLUSTALW alignment was performed with the FliC protein amino acid sequences of the four complete *P. ananatis* strains and the flg22 amino acid sequence of *Ps. aeruginosa* (Thompson *et al.*, 1994). The flg22 antigen region of all four FliC copies differed by at least five amino acids from that of the reference (Figure 4). Identical flg22 regions were observed for FliC1, FliC2 and FliC3, while FliC4 had a distinct flg22 amino acid sequence (differs by one amino acid). Hence, it would be expected that *P. ananatis* could phase vary between the FliC1/2/3 and FliC4 copies in response to the host receptors that could detect the particular flg22 antigen. However, it was also found that there were three FliC with multiple changes to the flg22 epitope. The FliC2 of LMG20103 had an insertion of nine amino acids after the fifth amino acid. These differences can lead to limited or no host defence responses, as it has been observed that a single amino acid change can change the epitope functionality (Gómez-Gómez *et al.*, 1999, Meng *et al.*, 2013). Thus, the distinct amino acid observed in the FliC4 flg22 sequences compared to the other three FliCs, as well as the insertion in the LMG20103 FliC2 sequence may result in antigenically distinct flagellins, which may not be detected by the host and subsequently not result in the launching of host defence responses.

2.5 Conclusion

Our comparative genomic analyses revealed the presence of genomic islands within the flagellar biosynthetic locus *fliTSZA-fliCD* of all seventeen strains of *P. ananatis* for which genomes are currently available. These genomic islands encode glycosyltransferases and sugar biosynthetic enzymes indicating that they encode the requirements for posttranslational modification of the flagellin protein in this species. The presence of these flagellin glycosylation islands in all seventeen compared strains suggests that flagellin glycosylation is a common trait of members of this species. The flagellin glycosylation islands are, however, highly variable, both in size and gene content, suggesting that variation exists in the flagellin glycans of different strains of this species. As such, the seventeen strains could be distinguished into four distinct types on the basis of the gene content of their FGI loci.

Rhamnose biosynthetic genes are present within the FGI of all seventeen compared strains, suggesting that the flagellins are glycosylated with a single type of sugar. However,

variability exists in terms of the gene complement, with the Type II and III FGI strains having truncated and fused *rml*C and *rml*A products. What the effects of this fusion are on the functioning of these enzymes remains to be elucidated. The most extensive variability, however, is in terms of the genes coding for proteins involved in other modifications of the glycans. The differential methylation, amination, acetylation and formylation may have profound effects on the structure and functions of the flagellin glycan among the different *P. ananatis* strains. Furthermore, genetic evidence is provided here that the glycans of the Type I FGI strains may further be substituted with a lipoyl group, and hence the flagellin of these strains may be modified with a liposaccharide.

We furthermore characterized the flagellin gene content of three *P. ananatis* strains for which complete genomes are available. Four copies of the *fliC* gene are present in these strains, and this is likely the case for all the compared strains. Variability can be observed in terms of their *flg*22 epitopes, suggesting that *P. ananatis* is capable of flagellin phase variation. The capacity of both flagellin phase variation and flagellin glycosylation would provide *P. ananatis* with a competitive advantage when it comes to avoiding plant host recognition and concomitant defence responses. This may be particularly pertinent considering the broad range of plant hosts that *P. ananatis* is able to colonise and infect.

2.6 References

Achouak, W., S. Conrod, V. Cohen & T. Heulin, (2004). Phenotypic variation of *Pseudomonas brassicacearum* as a plant root-colonization strategy. *Molecular Plant-Microbe Interactions.* **17**(8): 872-879

Aldridge, P.D., C. Wu, J. Gnerer, J.E. Karlinsey, K.T. Hughes & M.S. Sachs, (2006). Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. *Proceedings of the National Academy of Sciences.* **103**(30): 11340-11345

Arora, S.K., M. Bangera, S. Lory & R. Ramphal, (2001). A genomic island in *Pseudomonas aeruginosa* carries the determinants of flagellin glycosylation. *Proceedings of the National Academy of Sciences.* **98**(16): 9342-9347

Arora, S.K., A.N. Neely, B. Blair, S. Lory & R. Ramphal, (2005). Role of motility and flagellin glycosylation in the pathogenesis of *Pseudomonas aeruginosa* burn wound infections. *Infection and Immunity.* **73**(7): 4395-4398

Bomfeti, C.A., W.F. Meirelles, E.A. Souza-Paccola, C.R. Casela, A.d.S. Ferreira, I.E. Marriel & L.D. Paccola-Meirelles, (2007). Evaluation of commercial chemical products, in vitro and in vivo in the control of foliar disease, maize white spot, caused by *Pantoea ananatis*. *Summa Phytopathologica.* **33**(1): 63-67

Bonifield, H.R. & K.T. Hughes, (2003). Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. *Journal of Bacteriology.* **185**(12): 3567-3574

Cai, R., S. Yan, H. Liu & B.A. Vinatzer, (2011). Reconstructing host range evolution of bacterial plant pathogens using *Pseudomonas syringae* pv. *tomato* and its close relatives as a model. *Infection, Genetics and Evolution.* **11**: 1738-1751

Chiang, S.L. & J.J. Mekalanos, (1999). rfb mutations in *Vibrio cholerae* do not affect surface production of toxin-coregulated pili but still inhibit intestinal colonization. *Infection and Immunity*. **67**(2): 976-980

Cother, E., R. Reinke, C. McKenzie, V. Lanoiselet & D. Noble, (2004). An unusual stem necrosis of rice caused by *Pantoea ananas* and the first record of this pathogen on rice in Australia. *Australasian Plant Pathology*. **33**(4): 495-503

Coutinho, P.M., E. Deleury, G.J. Davies & B. Henrissat, (2003) An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology*. **328**(2): 307-317

Coutinho, T., O. Preisig, J. Mergaert, M. Cnockaert, K.-H. Riedel, J. Swings & M. Wingfield, (2002). Bacterial blight and dieback of *Eucalyptus* species, hybrids, and clones in South Africa. *Plant Disease*. **86**(1): 20-25

Coutinho, T.A. & S.N. Venter, (2009). *Pantoea ananatis*: an unconventional plant pathogen. *Molecular Plant Pathology*. **10**(3): 325-335

Dasgupta, N., M.C. Wolfgang, A.L. Goodman, S.K. Arora, J. Jyot, S. Lory & R. Ramphal, (2003). A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Molecular Microbiology*. **50**(3): 809-824

De Baere, T., R. Verhelst, C. Labit, G. Verschraegen, G. Wauters, G. Claeys & M. Vaneechoutte, (2004). Bacteremic infection with *Pantoea ananatis*. *Journal of Clinical Microbiology*. **42**(9): 4393-4395

De Maayer, P., W.Y. Chan, E. Rubagotti, S.N. Venter, I.K. Toth, P.R. Birch & T.A. Coutinho, (2014). Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics*. **15**: 404

De Maayer, P., S.N. Venter, T. Kamber, B. Duffy, T.A. Coutinho & T.H. Smits, (2011). Comparative genomics of the type VI secretion systems of *Pantoea* and *Erwinia* species reveals the presence of putative effector islands that may be translocated by the VgrG and Hcp proteins. *BMC Genomics.* **12**: 576

Déziel, E., Y. Comeau & R. Villemur, (2001). Initiation of biofilm formation by *Pseudomonas aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. *Journal of Bacteriology.* **183**(4): 1195-1204

DiRusso, C.C. & P.N. Black, (2004). Bacterial long chain fatty acid transport: gateway to a fatty acid-responsive signaling system. *Journal of Biological Chemistry.* **279**(48): 49563-49566

Edwards, K.J., S. Allen, B.W. Gibson & A.A. Campagnari, (2005). Characterization of a cluster of three glycosyltransferase enzymes essential for *Moraxella catarrhalis* lipooligosaccharide assembly. *Journal of Bacteriology.* **187**(9): 2939-2947

Felix, G., J.D. Duran, S. Volko & T. Boller, (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *The Plant Journal.* **18**(3): 265-276

Felsenstein, J., (2002). {PHYLIP}(Phylogeny Inference Package) version 3.6 a3.

Giraud, M.-F., G.A. Leonard, R.A. Field, C. Berlind & J.H. Naismith, (2000). RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. *Nature Structural & Molecular Biology.* **7**: 398-402

Giraud, M.-F. & J.H. Naismith, (2000). The rhamnose pathway. *Current Opinion in Structural Biology.* **10**: 687-696

Gitaitis, R. & J. Gay, (1997). First report of a leaf blight, seed stalk rot, and bulb decay of onion by *Pantoea ananas* in Georgia. *Plant Disease.* **81**: 1096-1096

Gómez-Gómez, L. & T. Boller, (2000). FLS2: An LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Molecular Cell.* **5**: 1003-1011

Gómez-Gómez, L., G. Felix & T. Boller, (1999). A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*. *The Plant Journal.* **18**(3): 277-284

Hall, T.A., (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: Nucleic Acids Symposium Series. **41:** 95-98

Hayakawa, J. & M. Ishizuka, (2012). *Flagellar Glycosylation: Current Advances*. INTECH Open Access Publisher.

Henrissat, B. & G.J. Davies, (2000). Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiology.* **124**(4): 1515-1519

Herrera, C.M., M.D. Koutsoudis, X. Wang & S.B. Von Bodman, (2008). *Pantoea stewartii* subsp. *stewartii* exhibits surface motility, which is a critical aspect of Stewart's wilt disease development on maize. *Molecular Plant-Microbe Interactions.* **21**(10): 1359-1370

Ikeda, J.S., C.K. Schmitt, S.C. Darnell, P.R. Watson, J. Bispham, T.S. Wallis, D.L. Weinstein, E.S. Metcalf, P. Adams & C.D. O'Connor, (2001). Flagellar Phase Variation of *Salmonella enterica* Serovar *Typhimurium* Contributes to Virulence in the Murine Typhoid Infection Model but Does Not Influence *Salmonella*-Induced Enteropathogenesis. *Infection and Immunity.* **69**(5): 3021-3030

Jones, J.D. & J.L. Dangl, (2006). The plant immune system. *Nature.* **444**: 323-329

Jordan, F., (2003). Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions. *Natural Product Reports.* **20**: 184-201

Katoh, K. & D.M. Standley, (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution.* **30**(4): 772-780

Kingston, R.L., R.K. Scopes & E.N. Baker, (1996). The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure.* **4**(12): 1413-1428

Klein, K., R. Steinberg, B. Fiethen & P. Overath, (1971). Fatty acid degradation in *Escherichia coli*. *European Journal of Biochemistry.* **19**(3): 442-450

Kochetov, G., N. Tikhomirova & P. Philippov, (1975). The binding of thiamine pyrophosphate with transketolase in equilibrium conditions. *Biochemical and Biophysical Research Communications.* **63**(4): 924-930

Kovina, M.V. & G.A. Kochetov, (1998). Cooperativity and flexibility of active sites in homodimeric transketolase. *FEBS Letters.* **440**(1-2): 81-84

Kumar, S., M. Nei, J. Dudley & K. Tamura, (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics.* **9**(4): 299-306

Li, Q. & P.R. Reeves, (2000). Genetic variation of dTDP-L-rhamnose pathway genes in *Salmonella enterica*. *Microbiology.* **146**: 2291-2307

Lin, J.-W., Y.-F. Chao & S.-F. Weng, (1993). Nucleotide sequence of the *luxC* gene encoding fatty acid reductase of the lux operon from Photobacterium *leiognathi*. *Biochemical and Biophysical Research Communications.* **191**(1): 314-318

Liu, B., B. Hu, Z. Zhou, D. Guo, X. Guo, P. Ding, L. Feng & L. Wang, (2012). A novel non-homologous recombination-mediated mechanism for *Escherichia coli* unilateral flagellar phase variation. *Nucleic Acids Research*. gks040

Logan, S.M., (2006). Flagellar glycosylation–a new component of the motility repertoire? *Microbiology.* **152**: 1249-1262

Lombard, V., H.G. Ramulu, E. Drula, P.M. Coutinho & B. Henrissat, (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research.* **42**: D490-D495

Marchler-Bauer, A. & S.H. Bryant, (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Research.* **32**(2): W327-W331

McCarter, L.L., (2001). Polar flagellar motility of the *Vibrionaceae*. *Microbiology and Molecular Biology Reviews.* **65**(3): 445-462

Meng, F., C. Altier & G.B. Martin, (2013). *Salmonella* colonization activates the plant immune system and benefits from association with plant pathogenic bacteria. *Environmental Microbiology.* **15**: 2418-2430

Merino Montero, S. & J. Tomàs Magaña, (2014). Gram-negative flagella glycosylation. *International Journal of Molecular Sciences.* **15**: 2840-2857

Nothaft, H. & C.M. Szymanski, (2010). Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology.* **8**: 765-778

Paccola-Meirelles, L., A. Ferreira, W. Meirelles, I. Marriel & C. Casela, (2001). Detection of a bacterium associated with a leaf spot disease of maize in Brazil. *Journal of Phytopathology.* **149**: 275-279

Power, P. & M. Jennings, (2003). The genetics of glycosylation in Gram-negative bacteria. *FEMS Microbiology Letters.* **218**: 211-222

Schirm, M., S. Arora, A. Verma, E. Vinogradov, P. Thibault, R. Ramphal & S. Logan, (2004). Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *Journal of Bacteriology.* **186**(9): 2523-2531

Schneider, G. & Y. Lindqvist, (1998). Crystallography and mutagenesis of transketolase: mechanistic implications for enzymatic thiamin catalysis. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*. **1385**(2): 387-398

Serrano, F., (1928). Bacterial fruitlet brown-rot of Pineapple in the Philippines. *Philippine Journal of Science*. **36**(3): 271-305

Sheldon, P.S., R.G. Kekwick, C.G. Smith, C. Sidebottom & A.R. Slabas, (1992). 3-Oxoacyl-[ACP] reductase from oilseed rape (*Brassica napus*). *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*. **1120**: 151-159

Soly, R.R. & E.A. Meighen, (1991). Identification of the acyl transfer site of fatty acyl-protein synthetase from bioluminescent bacteria. *Journal of Molecular Biology*. **219**(1): 69-77

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

Sun, L., M. Jin, W. Ding, J. Yuan, J. Kelly & H. Gao, (2013a). Posttranslational modification of flagellin FlaB in *Shewanella oneidensis*. *Journal of Bacteriology*. **195**: 2550-2561

Sun, Y., L. Li, A.P. Macho, Z. Han, Z. Hu, C. Zipfel, J.-M. Zhou & J. Chai, (2013b). Structural basis for flg22-induced activation of the *Arabidopsis* FLS2-BAK1 immune complex. *Science*. **342**: 624-628

Taguchi, F., Y. Ogawa, K. Takeuchi, T. Suzuki, K. Toyoda, T. Shiraishi & Y. Ichinose, (2006). A homologue of the 3-oxoacyl-(acyl carrier protein) synthase III gene located in the glycosylation island of *Pseudomonas syringae* pv. *tabaci* regulates virulence factors via N-acyl homoserine lactone and fatty acid synthesis. *Journal of Bacteriology*. **188**(24): 8376-8384

Taguchi, F., S. Shibata, T. Suzuki, Y. Ogawa, S.-I. Aizawa, K. Takeuchi & Y. Ichinose, (2008). Effects of glycosylation on swimming ability and flagellar polymorphic transformation in *Pseudomonas syringae* pv. *tabaci* 6605. *Journal of Bacteriology.* **190**(2): 764-768

Taguchi, F., M. Yamamoto, M. Ohnishi-Kameyama, M. Iwaki, M. Yoshida, T. Ishii, T. Konishi & Y. Ichinose, (2010). Defects in flagellin glycosylation affect the virulence of *Pseudomonas syringae* pv. *tabaci* 6605. *Microbiology*. **156**: 72-80

Takeuchi, K., H. Ono, M. Yoshida, T. Ishii, E. Katoh, F. Taguchi, R. Miki, K. Murata, H. Kaku & Y. Ichinose, (2007). Flagellin glycans from two pathovars of *Pseudomonas syringae* contain rhamnose in D and L configurations in different ratios and modified 4-amino-4, 6-dideoxyglucose. *Journal of Bacteriology*. **189**: 6945-6956

Takeuchi, K., F. Taguchi, Y. Inagaki, K. Toyoda, T. Shiraishi & Y. Ichinose, (2003). Flagellin glycosylation island in *Pseudomonas syringae* pv. *glycinea* and its role in host specificity. *Journal of Bacteriology*. **185**(22): 6658-6665

Thompson, J.D., D.G. Higgins & T.J. Gibson, (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.* **22**(22): 4673-4680

Tyson, G.W., J. Chapman, P. Hugenholtz, E.E. Allen, R.J. Ram, P.M. Richardson, V.V. Solovyev, E.M. Rubin, D.S. Rokhsar & J.F. Banfield, (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. **428**: 37-43

Walcott, R., R. Gitaitis, A. Castro, F. Sanders Jr & J. Diaz-Perez, (2002). Natural infestation of onion seed by *Pantoea ananatis*, causal agent of center rot. *Plant Disease.* **86**(2): 106-111

Watanabe, K., H. Kawakita & M. Sato, (1996). Epiphytic bacterium, *Erwinia ananas*, commonly isolated from rice plants and brown planthoppers (Nilaparvata lugens) in hopperburn patches. *Applied Entomology and Zoology*. **31**: 459-462

Watanabe, K. & M. Sato, (1999). Gut Colonization by an Ice Nucleation Active Bacterium, *Erwinia* (*Pantoea*) *ananas*, Reduces the Cold Hardiness of Mulberry Pyralid Larvae. *Cryobiology*. **38**: 281-289

Weller-Stuart, T., (2015). Genomic and functional characterization of motility in *Pantoea ananatis*. In: Department of Microbiology. University of Pretoria. pp. 176

Wells, J., W. Sheng, M. Ceponis & T. Chen, (1987). Isolation and characterization of strains of *Erwinia ananas* from honeydew melons. *Phytopathology*. **77**(3): 511-514

Yahara, K., M. Fukuyo, A. Sasaki & I. Kobayashi, (2009). Evolutionary maintenance of selfish homing endonuclease genes in the absence of horizontal transfer. *Proceedings of The National Academy of Sciences*. **106**(44): 18861-18866

Yin, Y., X. Mao, J. Yang, X. Chen, F. Mao & Y. Xu, (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*. **40**(W1): W445-W451

Zaccai, N.R., L.G. Carter, N.S. Berrow, S. Sainsbury, J.E. Nettleship, T.S. Walter, K. Harlos, R.J. Owens, K.S. Wilson & D.I. Stuart, (2008). Crystal structure of a 3-oxoacyl-(acylcarrier protein) reductase (BA3989) from *Bacillus anthracis* at 2.4-Å resolution. *Proteins: Structure, Function, and Bioinformatics*. **70**: 562-567

Zhou, M., Y. Yang, P. Chen, H. Hu, P.R. Hardwidge & G. Zhu, (2015). More than a locomotive organelle: flagella in *Escherichia coli*. *Applied Microbiology and Biotechnology*. **99**(8): 8883-8890

Zipfel, C. & G. Felix, (2005). Plants and animals: a different taste for microbes? *Current Opinion in Plant Biology*. **8**: 353-360

2.7 Tables and figures

Table 1. *Pantoea ananatis strains* included in this study. The geographic location and host of each of the strains are shown. The bioproject and genome assembly numbers are indicated, as are the genome sizes in base pairs (bp), number of contigs, their sequence status and genomic G+C contents.

Pantoea ananatis strain	Geographic location	Host	Bioproject number	Assembly number	Genome size bp	Contigs	Seq status	G+C Content
AJ13355	Japan	Soil	PRJDA60793	GCA_000270125.1	4877280	2	Complete	53,64
B1-9	Korea	Onion	PRJEA68179	SAMEA2272456	5112318	23	Draft	53,49
LMG 20103	RSA	*Eucalyptus*	PRJNA43085	GCA_000025405.2	4703373	2	Complete	53,69
LMG 2665	USA	Pineapple	PRJNA245803	SAMN02740635	4798552	9	Draft	53,39
BRT 175	Canada	Strawberries	PRJNA202242	SAMN02471932	4832469	12	Draft	53,74
DAR 76143	Australia	Rice	PRJDB1421	GCA_000467085.1	5225989	29	Draft	53,43
S6	Australia	Maize	PRJEB7511	SAMEA3303336	4344781	77	Draft	54,14
S7	Australia	Maize	PRJEB7512	SAMEA3303364	5026033	11	Draft	53,51
S8	Australia	Maize	PRJEB7513	SAMEA3303365	4935197	13	Draft	53,56
PA13	Korea	Rice	PRJNA74285	GCA_000233595.1	4867131	2	Complete	53,58
BD442	RSA	Maize	PRJNA245793	SAMN02739895	4798552	11	Draft	53,59
LMG 5342	USA	Human	PRJEA78277	GCA_000283875.1	4908144	2	Complete	53,32
B40	Japan	Rice	PRJDA69975	GCA_000333515.1	4645034	728	Draft	53,31
PA4	RSA	Onion	PRJNA245793	SAMN02739895	5151438	16	Draft	53,55
CFH71	USA	Cotton	PRJNA287251	SAMN03779113	4311532	4	Draft	53,16
PaMB1	China	Plant	PRJNA259615	SAMN03009826	4756252	177	Draft	53,84
SD-1	China	Rice	PRJNA232443	SAMN02951867	4945802	24	Draft	53,34

Table 2. Characteristics of the *P. ananatis* flagellin glycosylation islands. The sizes of the FGI regions and the number of proteins encoded on the FGIs were calculated. The G+C content (%) and G+C deviation from the whole genome are shown.

***Pantoea ananatis* strain**	**Size (bp)**	**# CDS**	**FGI G+C Content (%)**	**Deviation from genomic G+C content**
AJ13355	22711	20	47,58	-6.06
B1-9	22720	20	47,58	-5.91
LMG 20103	22711	20	47,59	-6.10
LMG 2665	22711	20	47,58	-5.81
BRT 175	21612	19	47,75	-5.99
DAR 76143	22775	19	47,50	-5.93
S6	15803	18	47,98	-6.16
S7	15836	19	47,98	-5.53
S8	15836	18	47,61	-5.95
PA13	16036	13	48,22	-5.36
BD442	16036	12	48,22	-5.37
LMG 5342	8729	8	43,92	-9.40
B40	8227	6	43,86	-9.45
PA4	8730	7	44,04	-9.51
CFH71	8962	8	43,78	-9.38
PaMB1	7894	6	43,33	-10.51
SD-1	10877	8	41,52	-11.82

Table 3. Proteins encoded on the *P. ananatis* FGIs. A short name for each of the proteins is provided. The sizes and putative function (as predicted by BlastP analysis against the NCBI nr protein database and Conserved Domain Database are given. The number of *P. ananatis* strains in which orthologs of each distinct FGI protein are found was determined and the average amino acid identity between orthologs was calculated.

FGIPAN	Protein Name	Size aa	Putative Function	Protein Domain	# strains with orthologs	AVE AA ID %
FGIPAN_01	Glt1	1230	Glycosyltransferase group 2, Glycosyl transferase 1-4	Glycosyltransferase group 1	12	94,32
FGIPAN_02	Glt2	1151	O-antigen biosynthesis protein- glycosyltransferase group 2		11	82,93
FGIPAN_03	DegT	381	Aminotransferase	DegT Aminotransferase	16	79,73
FGIPAN_04	Maa	216	Acetyltransferase		9	99,34
FGIPAN_05	FkbM1	377	S-adenosylmethionine-dependent methyl transferase		7	99,91
FGIPAN_06	AcpP	78	Hypothetical protein		9	99,21
FGIPAN_07	FabG1	247	3-oxoacyl-ACP-reductase		9	99,39
FGIPAN_08	FabG2	258	3-oxoacyl-ACP-reductase		9	98,08
FGIPAN_09	TktA1	237	Transketolase N	Thiamine diphosphate binding domain	9	80,05
FGIPAN_10	TktA2	321	Transketolase pyr	Pyrimidine binding domain, C-terminal domain	9	87,34
FGIPAN_11	FadD	469	Long-chain-fatty-acid--CoA ligase – AMP binding enzyme		9	84,92
FGIPAN_12	LuxE	359	Acyl-protein synthase		9	99,47
FGIPAN_13	LuxC	410	Acyl-CoA reductase		9	99,68
FGIPAN_14	FkbM2	368	Methyltransferase 21		2	100
FGIPAN_15	MviM	369	Oxidoreductase- NAD-binding Rossmann fold		2	99,7
FGIPAN_16	CarB	388	Putative formyltransferase	ATP-grasp domain	2	99,4
FGIPAN_17	Glt3	1097	O-antigen biosynthesis protein		1	100
FGIPAN_18	Hyp1	280	Hypothetical protein		11	91,41
FGIPAN_19	Glt4	1183	family 2 glycosyltransferase	Glycosyl transferase group 1	5	99,11
FGIPAN_20	VioF	260	methionyl-tRNA formyltransferase		5	83,72
FGIPAN_21	RmlD	288	dTDP-4-dehydrorhamnose reductase	Rml substrate binding domain	17	69,43
FGIPAN_22	RmlC	177	dTDP-4-dehydrorhamnose 3,5-epimerase		10	99,88
FGIPAN_23	RmlA	288	Glucose-1-phosphate thymidylyltransferase		10	99,43
FGIPAN_24	ZmpB	106	Hypothetical protein (zinc metalloprotease)		9	98,81
FGIPAN_25	Edn	324	Hypothetical protein (HNH nuclease)		17	100
FGIPAN_26	Hyp2	52	No blast hit		17	97,66

Table 4. The closest match for each FGI protein outside of the species. The organism name, locus tag of the encoded protein, average amino acid identity, bitscore and e-values are given.

FGIPAN	Closest relative	Locus_Tag	AVE AA ID%	Ave Bitscore	Ave E-value
FGIPAN_01	*Pantoea stewartii* subsp. *Indologenes* LMG 2632	HA47_20900	88	211,7333333	7,07E-68
FGIPAN_02	*Pantoea stewartii* subsp. *Indologenes*	HA47_20895	96,3	266,9411765	4,35E-89
FGIPAN_03	*Pantoea stewartii* subsp. *Indologenes*	HA47_20890	94,42	831,9411765	0
FGIPAN_04	*Pantoea stewartii* subsp. *stewartii* DC283	CKS_2661	97,89	548	2,14E-31
FGIPAN_05	*Pantoea stewartii* subsp *stewartii* DC283	CKS_2661	95,74	551,75	1,18E-10
FGIPAN_06	*Pantoea stewartii* subsp *stewartii* DC283	CKS_2661	96,98	563,5	6,25E-117
FGIPAN_07	*Pantoea stewartii*	NL54_12840	96,16	538,5588235	4,12E-19
FGIPAN_08	*Pantoea stewartii*	NL54_12835	75,73	1983,916667	0
FGIPAN_09	*Pantoea stewartii* subsp. *Indologenes*	HA47_20870	59,18	1358,090909	0
FGIPAN_10	*Pantoea stewartii*	NL54_12825	79,27	618,125	5,00E-121
FGIPAN_11	*Dickeya Zeae* EC1	W909_12315	54,48	250	6,33E-80
FGIPAN_12	*Enterobacter cloacae*	OA46_04415	39,98	282,7142857	3,86E-88
FGIPAN_13	*Enterobacter* sp. 638	Ent638_2517	80,19	131,6666667	1,21E-37
FGIPAN_14	*Enterobacter* sp. 638	Ent638_2516	70	359,4444444	4,22E-122
FGIPAN_15	*Enterobacter cloacaea*	OA46_04445	77,51	406,8888889	5,56E-120
FGIPAN_16	*Enterobacter cloacae* S611	EDP2_2424	68,23	348,4444444	1,56E-113
FGIPAN_17	*Enterobacter cloacae* S611	EDP2_2423	70,78	478,8888889	1,19E-155
FGIPAN_18	*Enterobacter cloacae* S611	EDP2_2422	55,27	503,6666667	7,00E-171
FGIPAN_19	*Cronobacter sakazakii* ATCC BAA-894	ESA_01298	70,69	521	0
FGIPAN_20	*Pectobacterium carotovorum* subsp. *Brasiliense*	IW01_01585	51	438,8889	2,33E-148
FGIPAN_21	*Pantoea stewartii* subsp. *Indologenes*	HA47_20860	96,46	734	0
FGIPAN_22	*Pantoea stewartii* subsp. *stewartii* DC283	CKS_3296	97,97	763,5	0
FGIPAN_23	*Pantoea stewartii*	NL54_12810	96,91	783,5	0
FGIPAN_24	*Pantoea vagans*	ID11_09080	34,05	566	1,00E-178
FGIPAN_25	*Pantoea stewartii* subsp. *Indologenes*	HA47_20845	79,19	465,8182	9,15E-154
FGIPAN_26	*Enterobacter cloacaea* S611	EDP2_2416	51,78	1228,6	0
FGIPAN_27	*Pectobacterium carotovorum* subsp. *carotovorum* UGC32	G033_08650	69,57	348,8	1,20E-86
FGIPAN_28	*Pantoea stewartii* subsp. *Indologenes*	HA47_20840	81,16	457,4706	4,12E-80
FGIPAN_29	*Pantoea stewartii* subsp. *Stewartii* DC283	CKS_3304	92,09	343	1,00E-117
FGIPAN_30	*Pantoea stewartii* subsp *stewartii* DC283	CKS_3304	69,38	107	1,38E-26
FGIPAN_31	*Pantoea stewartii*	NL54_12790	94,8	568,2	0,00E+00
FGIPAN_32	-	-	-	-	-
FGIPAN_33	*Pantoea stewartii* subsp. *Indologenes*	HA47_20830	89,72	615,06	0,00E+00
FGIPAN_34	No hits				
FGIPAN_35	*Pantoea* sp. 3.5.1.	EP46_03346	97,82	472,7059	4,59E-166
FGIPAN_36	*Pantoea stewartii*	NL54_12775	96,3	355,06	7,82E-120

Table 5. Common and unique proteins among the different *P. ananatis* FGI Types. The common proteins belong to either two, three or all of the FGI Types and * is for the combination.

Type	Unique Proteins	Common Proteins		
		In 2 FGI Types	In 3 FGI Types	In all FGI Types
I	- Acyltransferase (Maa)	- Hypothetical protein (Hyp1)(FGI Type I, II)	- Glycosyltransferase (Glt1)(FGI Type I, II, IV)	- dTDP-4-dehydrorhamnose reductase (RmlD)
	- Methyltransferase (FkbM1)	- O-antigen biosynthesis Glycosyl transferase (Glt2)(FGI Type I, II)	- Aminotransferase (DegT)(FGI Type I, II, III)	- Zinc metalloprotease (ZmpB)
	- Hypothetical protein (AcpP)	- rmlCA* (FGI Type II, III)		- HNHc nuclease (Edn)
	- 3-oxoacyl-ACP-reductase 1 (FabG1)	- dTDP-4-dehydrorhamnose 3,5-epimerase (RmlC)(FGI Type I, IV)		- Hypothetical protein (Hyp2)
	- 3-oxoacyl-ACP-reductase 2 (FabG2)	- Glucose-1-phosphate thymidyltransferase (RmlA) (FGI Type I, IV)		
	- Transketolase 1 (TktA1)			
	- Transketolase 2 (TktA2)			
	- Long-chain-fatty-acid—CoA ligase (FadD)			
	- Acyl-protein synthase (LuxE)			
	- Acyl-CoA reductase (LuxC)			
II	- Methyltransferase (FkbM2)			
	- Oxidoreductase (MviM)			
	- Carbomyl phosphate synthase (CarB)			
III	- O-antigen biosynthesis glycosyltransferase (Glt3)			
	- Methionyl-tRNA formyltransferase (vioF)			
IV	- Glycosyltransferase (Glt4)			

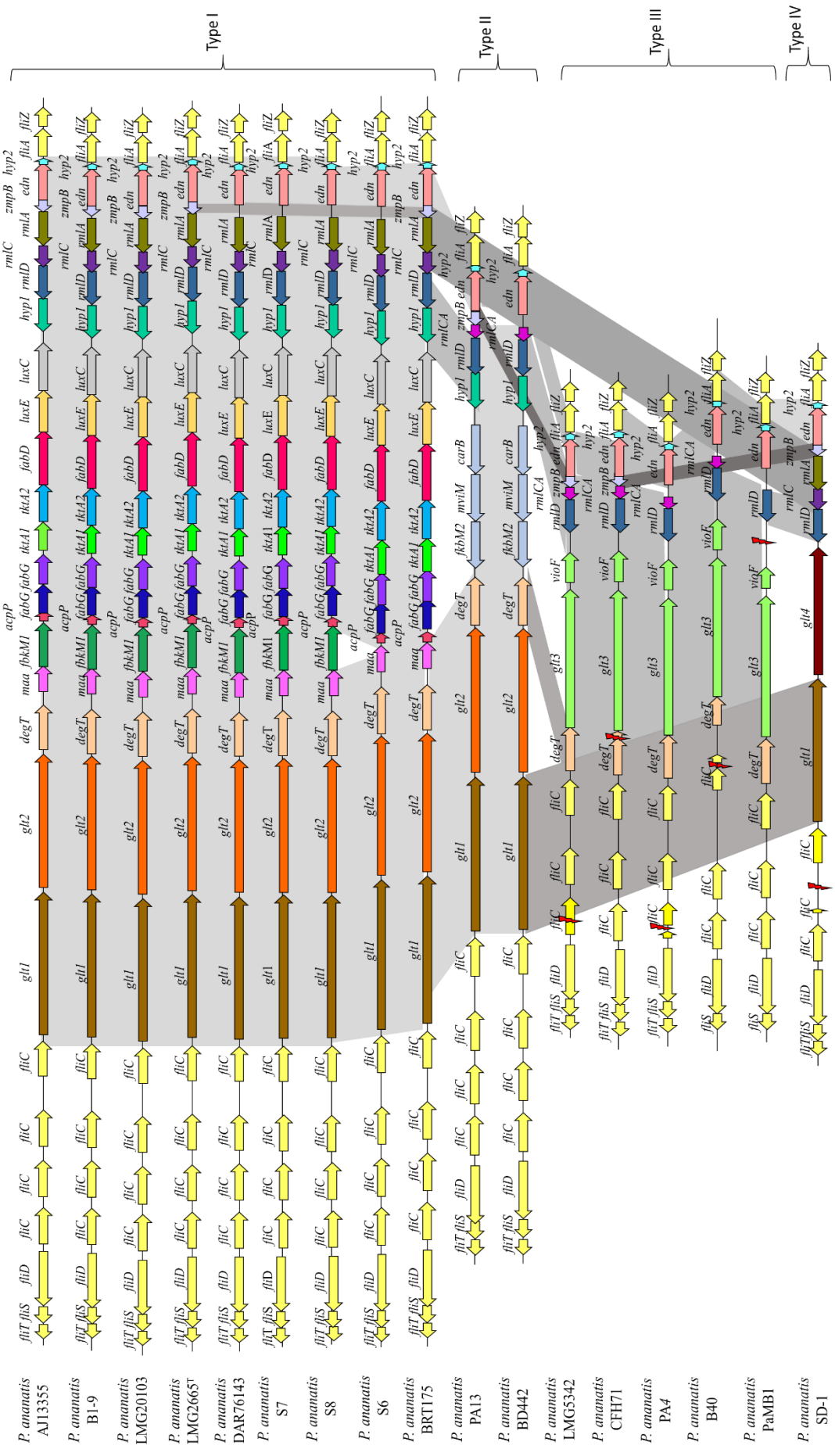


Figure 1: Genomic comparative analysis of the FGI regions in seventeen *P. ananatis* strains. Indicating four distinct FGI Types.

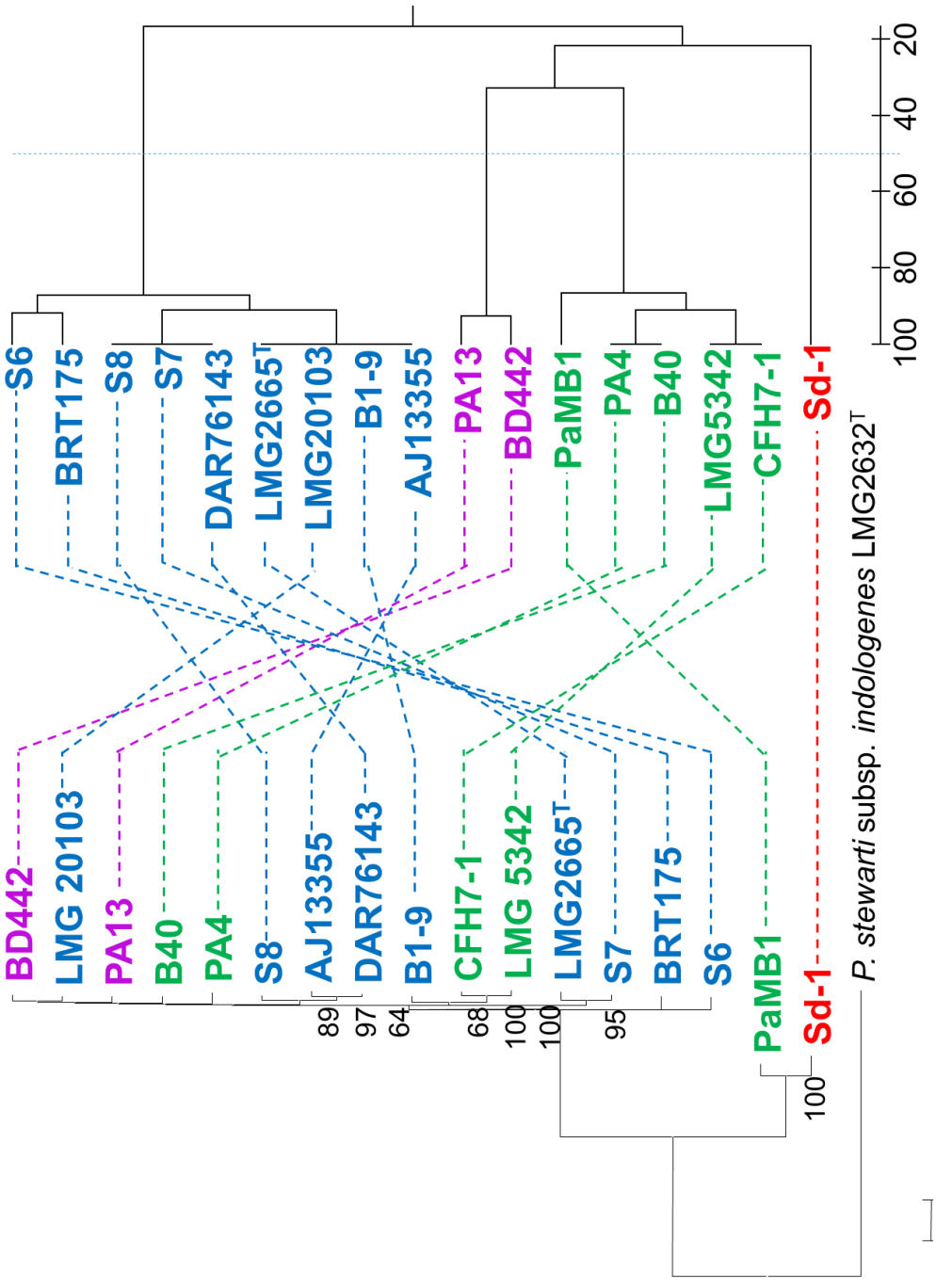


Figure 2: Flagellin glycosylation typing dendrogram versus house-keeping phylogeny. A neighbour-joining phylogeny (left) was constructed on the basis of the concatenated *atpD*, *gyrB*, *infB* and *rpoB* gene sequences. Bootstrap support (n=1,000) is shown. A FGI typing UPGMA dendrogram (right) was constructed on the basis of presence/absence of orthologs of each of the proteins encoded in the FGI loci. The FGI type cut-off values of 50% similarity are shown.

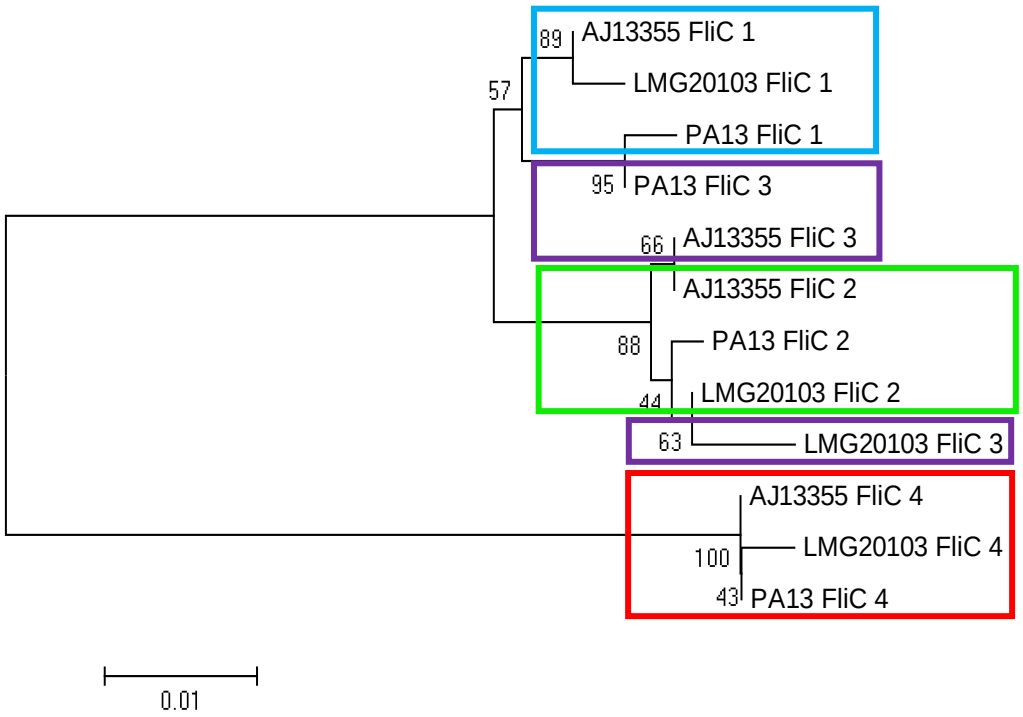


Figure 3. Neighbour-joining phylogeny of the FliC proteins of *P. ananatis* strains AJ13355, LMG 20103and PA13. Bootstrap support (n = 1,000) is shown.

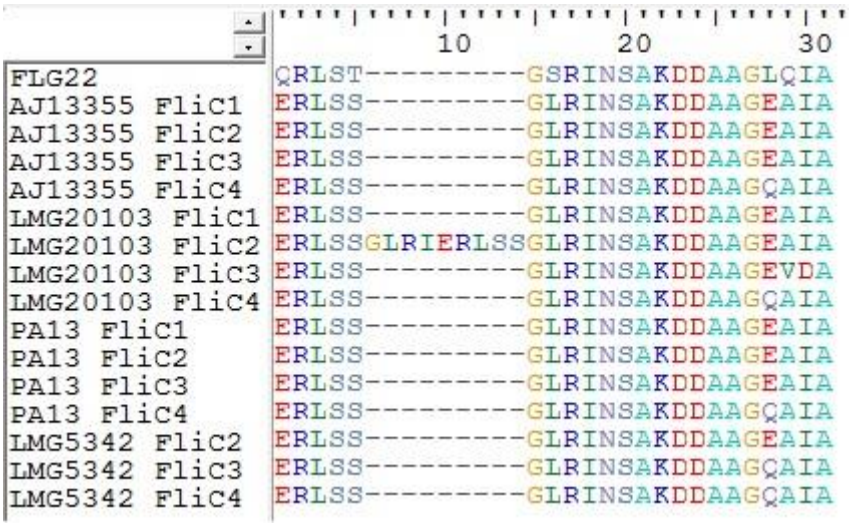


Figure 4. CLUSTALW alignment of the flg22 amino acid sequences of the flagellin copies. For the complete sequenced *P. ananatis* strains LMG20103, PA13, AJ13355 and LMG5342, indicating the difference in the amino acid sequences.

CHAPTER THREE

TYPING OF *P. ANANATIS* STRAINS ON THE BASIS OF VARIABILITY IN THEIR FLAGELLIN GLYCOSYLATION ISLANDS

3.1 Abstract

Pantoea ananatis is a ubiquitous bacterium that is distributed globally, within a wide range of hosts associated with insects, a range of plant diseases, the environment, as well as humans. Our comparative genomic analyses (Chapter 2) revealed the extensive versatility in the flagellin glycosylation islands of seventeen genome sequenced strains, which could be used to distinguish *P. ananatis* strains into four distinct FGI types. Here we wanted to ascertain if flagellin glycosylation type could be linked to the host and/or geographic location from which *P. ananatis* strains were isolated. By means of flagellin glycosylation island (FGI) gene specific PCR assays the presence of FGIs, as well as the FGI type that they belong to were determined for a large collection of *P. ananatis* strains isolated from a wide range of hosts and geographical locations. All of the screened strains showed genetic evidence that FGIs are present and belonged to three distinct FGI types. No direct correlation between FGI type and host/geographical location could be observed. However, the universal presence of FGIs suggest a means by which *P. ananatis* strains can potentially infect a broad range of plant hosts.

3.2 Introduction

The name of the genus *Pantoea* is derived from the Greek adjective *Pantoios* signifying "of all sorts and sources" (Brady *et al.*, 2010). As such, members of this genus are frequently isolated from a wide range of environmental sources including soil (Kageyama *et al.*, 1992, Andersson *et al.*, 1999, Dastager *et al.*, 2009), water and air (Poubol & Izumi, 2005, Morohoshi *et al.*, 2007). Furthermore, they are commonly found in association with vertebrate (De Baere *et al.*, 2004, Delétoile *et al.*, 2009), invertebrate (Watanabe *et al.*, 1996) as well as plant hosts (Bomfeti *et al.*, 2007, Delétoile *et al.*, 2009). *Pantoea ananatis* typifies the genus in this respect, being ubiquitous and isolated from six of the seven (except Antarctica) continents on Earth to date (Serrano, 1928, Paccola-Meirelles *et al.*, 2001, Coutinho *et al.*, 2002, Walcott *et al.*, 2002, Hara *et al.*, 2012). Strains of *P. ananatis* are isolated from a wide range of environmental sources including soil (Andreeva *et al.*, 2011, Hara *et al.*, 2012), water (Morohoshi *et al.*, 2007), food (meat packaging)(Ercolini *et al.*, 2006), sorghum fermentation (Mohammed *et al.*, 1991) and aviation fuels (Rauch *et al.*, 2006). *P. ananatis* strains are also frequently found associated with insects, including ticks

brown grasshoppers, mulberry pyralids (Watanabe *et al.*, 1996, Watanabe & Sato, 1999), ticks and fleas (Murrell *et al.*, 2003). *P. ananatis* has also been isolated from humans and has been associated with a clinical case of septicaemia and bacteraemia (De Baere *et al.*, 2004). Most commonly, however, *P. ananatis* strains are isolated from plants (Coutinho & Venter, 2009). They form part of the endophytic and epiphytic flora on a broad range of plant hosts, including coffee (Nunes & De Melo, 2006), tea (Goto *et al.*, 1988), ginseng (Cho *et al.*, 2007), gypsophila (Cooksey, 1986), honey dew melons (Ceponis *et al.*, 1985, Wells *et al.*, 1987), pineapple (Serrano, 1928), strawberry (Obata *et al.*, 1990), citrus (Wang *et al.*, 2008), papaya (Thomas *et al.*, 2007), pepper (Kang *et al.*, 2007), pearl millet (Frederickson *et al.*, 1997), sorghum (Morales-Valenzuela *et al.*, 2007), sudan-grass (Walcott *et al.*, 2002), buckwheat (Walcott *et al.*, 2002) and cotton lint (Chun & Perkins, 1997).

There has been particular interest in this species as it represents an emerging pathogen on a broad range of plant hosts, including the agronomic crops maize (Paccola-Meirelles *et al.*, 2001, Bomfeti *et al.*, 2007), rice (Cother *et al.*, 2004, Goszczynska *et al.*, 2007), onion (Walcott *et al.*, 2002) and *Eucalyptus* (Coutinho *et al.*, 2002). Little is known about the pathogenicity factors involved in infection and disease symptom formation. Furthermore, little is known about the host specificity of *P. ananatis* strains. The question has arisen whether a strain from one host can infect another host. It has been observed that *Eucalyptus* and maize strains can infect onion seedlings (De Maayer, 2010, Shyntum *et al.*, 2012, Weller-Stuart, 2015). However, Kido *et al.* (2010) reported that strains isolated from rice, foxtail millet, pineapple and river water were able to cause palea browning on rice, while strains isolated from melon were unable to cause disease on rice, but did cause internal fruit rot of melons (Kido *et al.*, 2010). This suggests that some form of host specificity exists in some strains of the species. Research must thus be undertaken to determine whether there are any genetic or phenotypic determinants underlying host specificity in this species.

Flagellin glycosylation has been linked to host specificity in some phytopathogenic bacteria. For example, flagellin glycosylation mutants of *Pseudomonas syringae* pv. *glycinea* were able to grow and cause symptom-like changes in the non-host tobacco, while these mutants did not cause symptoms on the original host soybean (Takeuchi *et al.*, 2003). This suggests that flagellin glycosylation is involved in recognition by the plant host and may drive

compatibility between the phytopathogen and its plant host(s). From the comparative genomic analyses of the *P. ananatis* flagellin glycosylation loci (Chapter 2), it is evident that there is extensive variability in these loci, and the compared strains could be subdivided into four distinct FGI types. This raises the question as to whether strains with a particular type of FGI can be linked to a particular host with which a *P. ananatis* strain is associated, or if strains with the same FGI type can infect a range of different hosts. Furthermore, are particular FGI types associated with distinct geographical origins?

A large collection of *P. ananatis* strains isolated from six continents and from a wide range of different hosts are available at the Forestry and Agricultural Biotechnology Institute. By means of genetic screening of variable gene markers, we undertook to classify these strains according to their FGI types. Furthermore, we attempted to identify if there is any correlation between FGI type and the hosts they infect and the geographic locations from which the strains were isolated.

3.3 Method and Materials

3.3.1 *P. ananatis* strain collection, growth and DNA extraction

Forty strains of *P. ananatis* isolated from diverse plant hosts and environmental niches, from every continent (with the exception of Antarctica) are maintained in the FABI Bacterial Culture Collection (FBCC)(Table 1). Strains were reconstituted from glycerol stocks and maintained on LB (Luria-Bertani) agar and incubated at 28˚C for 48 hours. Genomic DNA was extracted from all forty strains using the Zymo Research Quick-gDNA extraction kit (Zymo Research D3006), as per the manufacturer's instructions.

3.3.2 Primer design for conserved FGI genes and variable genes

Genes for FGI screening (presence/absence and typing) were selected on the basis of our comparative genomic analyses (Chapter 2 Figure 1). The gene sequences selected in this study were extracted from the FGI loci of the seventeen genome sequenced strains and

aligned using the MAFFT server (Katoh & Standley, 2013). Conserved regions at the 5' and 3' ends of the target genes were selected for primer design. Primer sequences were validated for hairpin, homodimer and heterodimer formation using the IDT DNA oligonucleotide analyzer (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/)(Table 2).

Two genes conserved among all seventeen genome sequenced strains (the HNH endonuclease gene *edn* and dTDP-4-dehydrorhamnose reductases *rmlD*) were selected for primer design to ascertain the presence of FGIs in the test strains (Figure 1, Tables 2 and Table 3). Furthermore, a primer set was designed (FGIPF and FGIPR) on the basis of sequences complementary to the flanking genes (*fliC* and *fliA*). Absence of an FGI region would result in PCR amplicons for this primer set. Four genes, *caiC* (Acyl CoA synthase), *mviM* (oxidoreductase), *glt3* (glycosyltransferase) and *glt4* (glycosyltransferase) which are each unique to a particular FGI type (Type I – IV, respectively) were selected for primer design for primary FGI typing (Figure 1, Tables 2 and 3). To account for the possibility of sequence variation in the priming sites which would result in no amplification in a particular strain additional secondary primer sets were designed for two genes which are shared by two or more, but not all strains, for further verification of the primary FGI type screening. These include *glt2* (glycosyltransferase found in FGI Type I and II) and *degT* (aminotransferase found in FGI type I, II and III)(Figure 1, Tables 2 and 3).

3.3.3 Polymerase chain reaction (PCR) screening protocols

The following PCR reagent mixture was used for all primer sets: 6.25 µM dNTPs, 2.5 µl 10x Buffer (Thermo Scientific), 2.0 µl $MgC\ell_2$ (Thermo Scientific), 1.25 U Super-Therm Polymerase (Thermo Scientific), 0.5 µM forward primer, 0.5 µM reverse primer, 2µl DNA and nuclease-free water (Thermo Scientific) to a total volume of 25 µl. PCR amplification was undertaken for all primer sets using an Eppendorf Mastercycler according to the following protocol: initial denaturation 94˚C for 5 minutes (min), 30 cycles [denaturation 94˚C for 30 seconds (sec), primer annealing at 56˚C for 1 min, elongation at 72˚C for 1 min and 30 sec] and a final elongation at 72˚C for 7 min. The amplification of the *edn* gene fragments was performed using this standard protocol. For the amplification of the *caiC*, *glt2*, *glt3* and *fgi* gene fragments the annealing step was amended to 60˚C for 50 sec. For the

amplification of *degT*, *glt4* and *rmlD* gene fragments the annealing step was amended to 62˚C for 55 sec and *mviM* gene fragments the annealing step was amended to 56˚C for 1 min.

3.3.4 Visualisation of PCR amplicons

Following PCR amplification, the amplicons were electrophoresed on a 1% agarose gel, composed of 1 g SeaKem® LE Agarose powder (Lonza) in 100 ml of 1xTAE buffer. Two µl of amplicon was mixed with 1µl of Gel Red™ dye and loaded in each well. The gel was electrophoresed in a BioRad wide mini-sub® Cell GT using a BioRadPowerPac™ Basic, at 100V for 30 minutes. The gel was subsequently visualised under a UV light and the presence/absence of bands and sizes of the bands compared.

3.4 Results and discussion

3.4.1 FGI loci are present in all screened *P. ananatis* strains

Two genes which are universal to all four FGI types (Chapter 2), *edn*, coding for an HNH endonuclease (Figure 2) and *rmlD*, coding for dTDP-4-dehydrorhamnose reductases (Results not shown) were PCR amplified from a collection of forty *P. ananatis* strains. Amplicons could be observed for both genes in all forty strains, suggesting that FGI loci are universally present among *P. ananatis* strains. This was further confirmed by PCR amplification with a forward and reverse primer which are complementary to the *fliC* and *fliA* genes flanking the FGI loci in the genome sequenced strains, at the 5' and 3' ends, respectively. No bands could be observed for any of the strains, indicating that inserts, which are too large to be amplified using conventional Taq polymerase are present in all forty of the tested strains (Figure 6). These two PCR amplification strategies combined provide strong evidence that flagellin glycosylation islands are integrated in the flagellum biosynthetic cluster 3 of all *P. ananatis* strains, and that FGIs form part of the universal genetic complement of this species.

3.4.2 FGI Types I, II and III, but not Type IV, are prevalent among *P. ananatis* strains.

Forty *P. ananatis* strains were genetically screened to determine the FGI Type encoded on their genomes (Table 4 and 5). This data was combined with that obtained in Chapter 2 on the genome sequenced *P. ananatis* strains to determine the prevalence of distinct FGI types. Seven of the genome sequenced strains were included in the genetic screening, to validate the results obtained in the PCR assays. The genetic screening results for all seven strains matched that observed from the comparative genomic analyses (Chapter 2). Isolates or DNA were not physically available for an additional ten genome sequenced strains, but these were incorporated in the overall analysis. Thus fifty strains in total were utilized in the typing scheme. Twenty of the forty screened strains produced PCR amplicons for both *caiC* (Figure 3) and the secondary marker *glt2* (Figure 4), indicating that they belong to FGI type I. Thus the majority of strains (25 out of fifty strains, 50%) belong to FGI type I. The second most common FGI type among the *P. ananatis* strains is Type III (positive for *glt3* (Figure 5), negative for other Type I and II FGI genes i.e. *degT*, *glt2*, *caiC* and *mviM*), with 11 strains (22% of the total) showing genetic/genomic evidence of this locus. Only five strains (four genetically screened and one genomic strain) are positive for the primary marker *mviM* and secondary marker *glt2* and thus belong to FGI type II (10% of the fifty strains).

No amplicons were obtained for the *glt4* (glycosyltransferase) gene fragment in any of the forty screened *P. ananatis* strains. The Type IV FGI therefore appears to be unique to one strain, *P. ananatis* SD-1. As discussed in Chapter 2, this strain may potentially represent a novel subspecies of *P. ananatis*, or even a novel *Pantoea* spp. As such, the Type IV FGI may be a hallmark species to strains closely related to SD-1 (Chapter 2 Figure 2). It should be noted, however, that *P. ananatis* PaMB1 also clusters with SD-1, on the basis of a house-keeping gene phylogeny, but clusters with other *P. ananatis* strains on the basis of its Type III FGI. Screening of large dataset of *P. ananatis* strains thus needs to be undertaken to determine the prevalence of the Type IV FGI in this species.

3.4.3 The *P. ananatis* flagellin glycosylation island may show greater genetic variability than the current scope of FGI types.

While thirty-one of the forty genetically screened strains could be typed according to their flagellin glycosylation island gene content using the PCR assays, nine strains showed conflicting results and therefore could not be accurately placed into their different FGI types (Table 4 and 5). Three strains *P. ananatis* BCC098, BCC135 and BCC637 produced amplicons for the *glt2* gene suggesting they contain either a Type I or II FGI, but were negative for the FGI Type I gene *caiC* and the FGI type II gene *mviM*. No bands were obtained for any of the typing genes for one strain, BCC631. By contrast, a mixture of type I and II genes could be observed in *P. ananatis* BCC0102, while BCC0151 produced amplicons for the Type II (*mviM*) and Type III (*glt3*) genes and BCC626 and BCC641 produced bands for both Type II genes (*mviM* and *glt2*), as well as the Type III gene *glt3*. The reason for amplicon products being produced for distinct FGI Type gene fragments may be two fold. It may be possible that lower sequence conservation among some of the genes may result in non-specific binding of primers, and thus the amplification of false positive bands. Alternatively, it is possible that some strains may contain a mixture of genes from more than one distinct FGI type, as a result of horizontal gene exchange or recombination events. This may result in far greater variability in the flagellin glycosylation loci than that assumed in our comparative genomic analyses and this PCR screening analysis.

3.4.4 No correlation could be observed between FGI type and the host from which a *P. ananatis* strain was isolated.

Twenty-five of the compared strains belong to FGI type I. The majority (22 strains of the twenty-five) were isolated from a range of different plant hosts, including onion, *Eucalyptus*, pineapple, *Cattleya*, rice, maize, melons and strawberries (Figure 7). However, strains were not restricted to plant hosts. For example, the Type I FGI strain *P. ananatis* AJ13355 was isolated from soil, and both BCC030 and BCC053 were isolated from insects. Five of the compared strains belong to FGI Type II. All strains were isolated from a range of different plant hosts, including rice, sugarcane, Cassia and maize. Eleven of the compared strains belong to FGI Type III. The majority were isolated from a range of different plant hosts,

including wheat, rice, onion, cotton ball and banana. However, Type III FGI strain *P. ananatis* LMG5342 was isolated from humans and BCC0024 was isolated from insects.

Particular plant hosts of the compared strains belong to either one or a few FGI Types. The majority of the hosts belong to only one FGI type. Three plant hosts are unique to FGI Type I *P. ananatis* strains, including pineapple, *Cattleya* and strawberries. Two hosts belong to only FGI Type II, namely *Cassia*, and sugarcane. Two plant hosts are restricted to FGI Type III strains, banana and cotton, while PaMB1 was isolated from an unknown plant host. The human isolate also belongs to FGI Type III. It should be noted that there are only single strains isolated from each of these hosts. The one exception is pineapple, where both isolates belong to FGI Type I. For the other hosts for which there are multiple isolates, they belong to different FGI types, including melons (three strains; FGI Type I and untypable), *Eucalyptus* (five strains; FGI Type I and III), onion (six strains; FGI type I and III and untypable), maize (FGI Type I and II) and Rice (FGI Type I, II, III and IV). As such, there does not appear to be direct correlation between the FGI type a *P. ananatis* strain belongs to and the host that it infects.

3.4.5 There is poor correlation between *P. ananatis* FGI Type and geographic source of isolation.

The twenty-five FGI Type I strains were isolated from seven different countries in four continents, including RSA (Africa), USA and Canada (North America), Japan, Thailand and South Korea (Asia) and Australia (Australia) (Figure 7).The majority of the FGI Type I strains were isolated from Australia (28% of strains), RSA (24% of strains) and the USA (24% of strains). The five Type II FGI strains were isolated from four countries, with one from Africa (RSA), two from Asia (South Korea and Japan) and two from South America (Brazil). The eleven Type III FGI strains were isolated from five different continents, including Africa (RSA and Zimbabwe), Asia (China, India and Japan), Europe (Hungary) North America (USA), South America (Uruguay). The single Type IV strain was isolated from China, while the eight untypable strains were from RSA (four strains) and USA (four strains, respectively). When considering single countries, *P. ananatis* strains with distinct FGI types can also be observed, e.g. strains isolated from South Africa belong to FGI Type I,

II and III, while isolates from the USA belong to FGI Type I, III and the untypable FGI. The observation that *P. ananatis* isolates belonging to the same FGI Type have been isolated from different continents, and distinct FGI types being found in the same country indicate that there is no direct correlation between FGI type and the geographic source of isolation.

3.5 Conclusion

By means of genetic screening through the use of PCR assays, we have shown that flagellin glycosylation islands are likely universal to this species and furthermore suggests that flagellin glycosylation is a common genotypic and phenotypic trait for the species. It however, remains to be determined if the genetic presence of an FGI translates into a functional glycosylation system for the flagellum. This would have to be determined using knock-out mutagenesis and phenotypic assays.

Our genetic screening was relatively successful in classifying the *P. ananatis* strains according to their FGI type, with 32 out of 40 screened isolates being assigned to FGI Type I, II and III. The remaining eight strains could not be classified, and this may be due to additional variability in the flagellin glycosylation loci among *P. ananatis* isolates. Most of the strains belonged to FGI Type I, followed by FGI Type III and Type II. Alternatively, sequence variation in the priming site may have hampered classification of these untypable strains. Primers could potentially be designed on several other genetic targets that could provide additional differentiation of the strains into their distinct FGI Types.

The strains incorporated in this study included those isolated from several distinct plant hosts, with their aim of trying to establish a link between FGI Type and host specificity. From our analyses, there does not appear to be a direct correlation between the FGI Type of the *P. ananatis* isolate and the host it infects. As such, flagellin glycosylation may not play a determinative role in host specificity, as observed in other plant associated bacteria. Additionally, the strains incorporated in this study may not be truly host specific.

Unexpectedly, strains belonging to the different FGI types were isolated from multiple different continents, and strains from the same country (e.g. RSA) belonged to different FGI types. This raises questions about the evolution of the distinct FGI types. Are similar evolutionary pressures, or even similar horizontal gene exchange events occurring in these different geographic locales? Do they co-occur with other bacteria with the same genetic FGI make-up so that horizontal exchange can occur? A more likely explanation would that, with the global trade in plant material, *P. ananatis* isolates with distinct FGI types have been translocated across the world, and the subsequent co-mingling of populations have given rise to the global diversity of FGI loci. This presents an exciting avenue for future research into the evolution, structure and functionality of flagellin glycosylation in *P. ananatis*.

3.6 References

Andersson, A., N. Weiss, F. Rainey & M. Salkinoja-Salonen, (1999). Dust-Borne Bacteria in Animal Sheds, Schools and Children's Day Care Centres. *Journal of Applied Microbiology*. **86**(4): 622-634

Andreeva, I.G., L.I. Golubeva, T.M. Kuvaeva, E.R. Gak, J.I. Katashkina & S.V. Mashko, (2011). Identification of Pantoea Ananatis Gene Encoding Membrane Pyrroloquinoline Quinone (Pqq)-Dependent Glucose Dehydrogenase and Pqqabcdef Operon Essential for Pqq Biosynthesis. *FEMS Microbiology Letters*. **318**: 55-60

Bomfeti, C.A., W.F. Meirelles, E.A. Souza-Paccola, C.R. Casela, A.d.S. Ferreira, I.E. Marriel & L.D. Paccola-Meirelles, (2007). Evaluation of Commercial Chemical Products, in Vitro and in Vivo in the Control of Foliar Disease, Maize White Spot, Caused by *Pantoea Ananatis*. *Summa Phytopathologica*. **33**(1): 63-67

Brady, C.L., I. Cleenwerck, S.N. Venter, K. Engelbeen, P. De Vos & T.A. Coutinho, (2010). Emended Description of the Genus *Pantoea*, Description of Four Species from Human Clinical Samples, *Pantoea Septica* Sp. Nov., *Pantoea Eucrina* Sp. *Nov*., *Pantoea Brenneri* Sp. *Nov*. And *Pantoea Conspicua* Sp. *Nov*., and Transfer of *Pectobacterium Cypripedii* (Hori 1911) Brenner Et Al. 1973 Emend. Hauben Et Al. 1998 to the Genus as *Pantoea Cypripedii* Comb. *Nov*. *International Journal of Systematic and Evolutionary Microbiology*. **60**: 2430-2440

Ceponis, M., J. Wells & R. Cappellini, (1985). Bacterial Brown Spot of Honey Dew Melons. *Horticultural Science*. **20**(2): 302-303

Cho, K.M., S.Y. Hong, S.M. Lee, Y.H. Kim, G.G. Kahng, Y.P. Lim, H. Kim & H.D. Yun, (2007). Endophytic Bacterial Communities in Ginseng and Their Antifungal Activity against Pathogens. *Microbial Ecology*. **54**(2): 341-351

Chun, D.T. & H. Perkins, (1997). Profile of Bacterial Genera Associated with Cotton from Low Endotoxin and High Endotoxin Growing Regions. *Annuals of Agricultural and Environmental Medicine*. **4**: 233-242

Cooksey, D., (1986). Galls of *Gypsophila Paniculata* Caused by *Erwinia Herbicola*. *Plant Disease*. **70**(5): 464-468

Cother, E., R. Reinke, C. McKenzie, V. Lanoiselet & D. Noble, (2004). An Unusual Stem Necrosis of Rice Caused by *Pantoea Ananas* and the First Record of This Pathogen on Rice in Australia. *Australasian Plant Pathology*. **33**(4): 495-503

Coutinho, T., O. Preisig, J. Mergaert, M. Cnockaert, K.-H. Riedel, J. Swings & M. Wingfield, (2002). Bacterial Blight and Dieback of *Eucalyptus* Species, Hybrids, and Clones in South Africa. *Plant Disease*. **86**(1): 20-25

Coutinho, T.A. & S.N. Venter, (2009). *Pantoea Ananatis*: An Unconventional Plant Pathogen. *Molecular Plant Pathology*. **10**(3): 325-335

Dastager, S.G., C. Deepa & A. Pandey, (2009). Isolation and Characterization of High-Strength Phenol-Degrading Novel Bacterium of the *Pantoea* Genus. *Bioremediation Journal*. **13**(4): 171-179

De Baere, T., R. Verhelst, C. Labit, G. Verschraegen, G. Wauters, G. Claeys & M. Vaneechoutte, (2004). Bacteremic Infection with *Pantoea Ananatis*. *Journal of Clinical Microbiology*. **42**(9): 4393-4395

De Maayer, P., (2010). Gemone Comparisons to Identify Selected Pathogenicity Factors of a Plant-Associated *Pantoea Ananatis* Strain. In: Department of Microbiology. University of Pretoria. pp. 278

Delétoile, A., D. Decré, S. Courant, V. Passet, J. Audo, P. Grimont, G. Arlet & S. Brisse, (2009). Phylogeny and Identification of *Pantoea* Species and Typing of *Pantoea Agglomerans* Strains by Multilocus Gene Sequencing. *Journal of Clinical Microbiology*. **47**(2): 300-310

Ercolini, D., F. Russo, E. Torrieri, P. Masi & F. Villani, (2006). Changes in the Spoilage-Related Microbiota of Beef During Refrigerated Storage under Different Packaging Conditions. *Applied and Environmental Microbiology.* **72**(7): 4663-4671

Frederickson, D., E. Monyo, S. King & G. Odvody, (1997). A Disease of Pearl Millet in Zimbabwe Caused by *Pantoea Agglomerans*. *Plant Disease.* **81**(8): 959-959

Goszczynska, T., W. Botha, S. Venter & T. Coutinho, (2007). Isolation and Identification of the Causal Agent of Brown Stalk Rot, a New Disease of Maize in South Africa. *Plant Disease.* **91**(6): 711-718

Goto, M., B.L. Huang, T. Makino, T. Goto & T. Inaba, (1988). A Taxonomic Study on Ice Nucleation-Active Bacteria Isolated from Gemmisphere of Tea (Thea Sinensis L.), Phyllosphere of Vegetables and Flowers of *Magnolia Denudata* Desr. *Annals of the Phytopathological Society of Japan.* **54**(2): 189-197

Hara, Y., N. Kadotani, H. Izui, J.I. Katashkina, T.M. Kuvaeva, I.G. Andreeva, L.I. Golubeva, D.B. Malko, V.J. Makeev & S.V. Mashko, (2012). The Complete Genome Sequence of *Pantoea Ananatis* Aj13355, an Organism with Great Biotechnological Potential. *Applied Microbiology and Biotechnology.* **93**(1): 331-341

Kageyama, B., M. Nakae, S. Yagi & T. Sonoyama, (1992). *Pantoea Punctata* Sp. *Nov*., *Pantoea Citrea* Sp. *Nov*., and *Pantoea Terrea* Sp. *Nov*. Isolated from Fruit and Soil Samples. *International Journal of Systematic and Evolutionary Microbiology.* **42**(2): 203-210

Kang, S.H., H. Cho, H. Cheong, C. Ryu, J.F. Kim & S. Park, (2007). Two Bacterial Entophytes Eliciting Both Plant Growth Promotion and Plant Defense on Pepper (Capsicum Annuum L.). *Journal of Microbiology and Biotechnology.* **17**(1): 96

Katoh, K. & D.M. Standley, (2013). Mafft Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution.* **30**(4): 772-780

Kido, K., M. Hasegawa, H. Matsumoto, M. Kobayashi & Y. Takikawa, (2010). *Pantoea Ananatis* Strains Are Differentiated into Three Groups Based on Reactions of Tobacco and Welsh Onion and on Genetic Characteristics. *Journal of General Plant Pathology*. **76**: 208-218

Mohammed, S.I., L.R. Steenson & A.W. Kirleis, (1991). Isolation and Characterization of Microorganisms Associated with the Traditional Sorghum Fermentation for Production of Sudanese Kisra. *Applied and Environmental Microbiology*. **57**(9): 2529-2533

Morales-Valenzuela, G., H. Silva-Rojas, D. Ochoa-Martinez, E. Valadez-Moctezuma, B. Alarcon-Zuniga, L. Zelaya-Molina, L. Cordova-Tellez, L. Mendoza-Onofre, H. Vaquera-Huerta & A. Carballo-Carballo, (2007). First Report of *Pantoea Agglomerans* Causing Leaf Blight and Vascular Wilt in Maize and Sorghum in Mexico. *Plant Disease*. **91**(10): 1365-1365

Morohoshi, T., Y. Nakamura, G. Yamazaki, A. Ishida, N. Kato & T. Ikeda, (2007). The Plant Pathogen *Pantoea Ananatis* Produces N-Acylhomoserine Lactone and Causes Center Rot Disease of Onion by Quorum Sensing. *Journal of Bacteriology*. **189**(22): 8333-8338

Murrell, A., S.J. Dobson, X. Yang, E. Lacey & S.C. Barker, (2003). A Survey of Bacterial Diversity in Ticks, Lice and Fleas from Australia. *Parasitology Research*. **89**: 326-334

Nunes, F. & I. De Melo, (2006). Isolation and Characterization of Endophytic Bacteria of Coffee Plants and Their Potential in Caffeine Degradation. *Environmental Toxicology*. **1**: 293-297

Obata, H., K. Takinami, J.-i. Tanishita, Y. Hasegawa, S. Kawate, T. Tokuyama & T. Ueno, (1990). Identification of a New Ice-Nucleating Bacterium and Its Ice Nucleation Properties. *Agricultural and Biological Chemistry*. **54**(3): 725-730

Paccola-Meirelles, L., A. Ferreira, W. Meirelles, I. Marriel & C. Casela, (2001). Detection of a Bacterium Associated with a Leaf Spot Disease of Maize in Brazil. *Journal of Phytopathology*. **149**: 275-279

Poubol, J. & H. Izumi, (2005). Shelf Life and Microbial Quality of Fresh-Cut Mango Cubes Stored in High Co2 Atmospheres. *Journal of Food Science*. **70**(1): M69-M74

Rauch, M.E., H.W. Graef, S.M. Rozenzhak, S.E. Jones, C.A. Bleckmann, R.L. Kruger, R.R. Naik & M.O. Stone, (2006). Characterization of Microbial Contamination in United States Air Force Aviation Fuel Tanks. *Journal of Industrial Microbiology and Biotechnology*. **33**: 29-36

Serrano, F., (1928). Bacterial Fruitlet Brown-Rot of Pineapple in the Philippines. *Philippine Journal of Science*. **36**(3): 271-305

Shyntum, D., S. Venter, L. Moleleki & T. Coutinho, (2012). The Type Vi Secretion System in *Pantoea Ananatis* Plays a Role in Bacterial Competition. In: Phytopathology. AMER PHYTOPATHOLOGICAL SOC 3340 PILOT KNOB ROAD, ST PAUL, MN 55121 USA. pp. 109-109

Takeuchi, K., F. Taguchi, Y. Inagaki, K. Toyoda, T. Shiraishi & Y. Ichinose, (2003). Flagellin Glycosylation Island in *Pseudomonas Syringae* pv. *Glycinea* and Its Role in Host Specificity. *Journal of Bacteriology*. **185** (22): 6658-6665

Thomas, P., S. Kumari, G.K. Swarna & T. Gowda, (2007). Papaya Shoot Tip Associated Endophytic Bacteria Isolated from in Vitro Cultures and Host-Endophyte Interaction in Vitro and in Vivo. *Canadian Journal of Microbiology*. **53**(3): 380-390

Walcott, R., R. Gitaitis, A. Castro, F. Sanders Jr & J. Diaz-Perez, (2002). Natural Infestation of Onion Seed by *Pantoea Ananatis*, Causal Agent of Center Rot. *Plant Disease*. **86**(2): 106-111

Wang, J., B. So, J. Kim, Y. Park, B. Lee & H. Kang, (2008). Genome-Wide Identification of Pathogenicity Genes in *Xanthomonas Oryzae* Pv. *Oryzae* by Transposon Mutagenesis. *Plant Pathology*. **57**: 1136-1145

Watanabe, K., H. Kawakita & M. Sato, (1996). Epiphytic Bacterium, *Erwinia Ananas*, Commonly Isolated from Rice Plants and Brown Planthoppers (Nilaparvata Lugens) in Hopperburn Patches. *Applied Entomology and Zoology*. **31**: 459-462

Watanabe, K. & M. Sato, (1999). Gut Colonization by an Ice Nucleation Active Bacterium, *Erwinia* (*Pantoea*) Ananasreduces the Cold Hardiness of Mulberry Pyralid Larvae. *Cryobiology*. **38**: 281-289

Weller-Stuart, T., (2015) Genomic and Functional Characterization of Motility in *Pantoea Ananatis*. In: Department of Microbiology. University of Pretoria. pp. 176

Wells, J., W. Sheng, M. Ceponis & T. Chen, (1987). Isolation and Characterization of Strains of *Erwinia Ananas* from Honeydew Melons. *Phytopathology*. **77**(3): 511-514

3.7 Tables and figures

Table 1. The fifty *P. ananatis* strains included in this study. The FABI Bacterial Collection (FBCC) strain codes are indicated, as well as the geographic location and host from which they were isolated. Strains for which genome sequences are available (Chapter 2) are indicated in bold, while a * indicates those strains where the genomic information only was incorporated into this analysis.

#	FABI Culture Collection #	External Culture collection #	Origin	Host	FGI Type
1	BCC0024		RSA	Insect	III
2	BCC0030		RSA	Insect	I
3	BCC0053		RSA	Insect	I
4	BCC0083		USA	Onion	I
5	BCC0084		RSA	*Eucalyptus*	I
6	BCC0087		USA	Onion	I
7	BCC0094		USA	Onion	I
8	BCC0098		USA	Sudangrass	?
9	BCC0102		USA	Sudangrass	?
10	BCC0116		RSA	*Eucalyptus*	I
11	BCC0127	**LMG20103**	RSA	*Eucalyptus*	I
12	BCC0128		USA	Pineapple	I
13	BCC0132	**LMG2665^T**	USA	Pineapple	I
14	BCC0135	LMG2676	USA	Wheat	?
15	BCC0147	CTB1135	Japan	Rice	II
16	BCC0149	LMG2675	Europe	Wheat	III
17	BCC0151	ATCC35400	USA	Melons	?
18	BCC0155	LMG2807	USA	*Cattleya*	I
19	BCC0158	LMG2161	India	Rice	III
20	BCC0192	LMG5342	USA	Human	III
21	BCC0195	LMG2678	Zim	Wheat	III
22	BCC0367	THAILAND 7-2	Thailand	*Eucalyptus*	I
23	BCC0370	DAR76141	Aus	Rice	I
24	BCC0372	**DAR76143**	Aus	Rice	I

Table 1. Continued…					
#	**FABI Culture Collection #**	**External Culture collection #**	**Origin**	**Host**	**FGI Type**
25	BCC0373	DAR76144	Aus	Rice	I
26	BCC0375	RAMI7969	Aus	Rice	I
27	BCC0583	URUGUAY43	*Uruguay*	*Eucalyptus*	III
28	BCC0601	ICMP10132	Brazil	Sugarcane	II
29	BCC0602	ICMP12183	Brazil	Cassia	II
30	BCC0626	BD301	RSA	Onion	?
31	BCC0631	BD377	RSA	Onion	?
32	BCC0633	**PA4**	RSA	Onion	III
33	BCC0635	**BD442**	RSA	Maize	II
34	BCC0637	BD561	RSA	Maize	?
35	BCC0639	BD588	RSA	Maize	I
36	BCC0641	BD622	RSA	Maize	?
37	BCC0644	LMG2628	Japan	Banana	III
38	SUPP1791		Japan	Melons	I
39	SUPP2582		Japan	Melons	I
40	BCC1049	**AJ13355**	Japan	Soil	I
41		**B1-9***	Korea	Onion	I
42		**S7***	Aus	Maize	I
43		**S8***	Aus	Maize	I
44		**S6***	Aus	Maize	I
45		**BRT175***	Canada	Strawberries	I
46		**PA13***	Korea	Rice	II
47		**CFH7.1***	USA	Cotton	III
48		**B40***	Japan	Rice	III
49		**PaMB1***	China	Plant	III
50		**SD1***	China	Rice	IV

Table 2. The FGI Type differential markers. Indicating the presence (+) or absence (-) of the gene fragments amplified used for each FGI type in this study.

		Gene-specific primers								
		caiC	***mviM***	***glt3***	***glt4***	***glt2***	***degT***	***edn* (+ve)**	***rmlD* (+ve)**	***fgi* (-ve)**
FGI Type	I	+	-	-	-	+	+	+	+	-
	II	-	+	-	-	+	+	+	+	-
	III	-	-	+	-	-	+	+	+	-
	IV	-	-	-	+	-	-	+	+	-

Table 3. The primers used in this study. The nucleotide sequences, lengths of the primers, G+C content (%), melting temperature and predicted fragment lengths of the amplicons are shown.

Gene	Name	Primer nt sequence	#Bases	G+C%	TM ˚C	Fragment length ~
glt2	glt2 FP	5'-ACCCTGGATAGCCGTTTATCC-3'	21	52,4	56,5	760
	glt2 RP	5'-TTTGCTGTATGTCACCGCCC-3'	20	55	58,1	
edn	edn FP	5'-GAGTCCTTCAACCAGCTATCGG-3'	21	52,4	57,8	800
	edn RP	5'-ACGGGCTAATGCTTTCACG-3'	19	52,6	55,7	
glt3	glt3 FP	5'-CAGCGTCCTGAAGTAGGTATCG-3'	22	54,5	56,8	970
	glt3 RP	5'-AAGCTGGTAAGATGTGCCCACG-3'	22	54,5	59,8	
mviM	mviM FP	5'-CGCCAATGAGGACTACCGGATTGC-3'	24	54,5	57,8	1020
	mviM RP	5'-GCCACCATCTTGACGCAGTTGCGG-3'	24	57,1	59,1	
rmlD	rmlD FP	5'-AATTCACACCAGCTCACCG-3'	20	55	57,6	588
	rmlD RP	5'-CAATTAGGGCGTTGCTTGCTGG-3'	22	54,5	59,3	
caiC	caiC FP	5'-TCGCTGTGATTGACGATAGCGG-3'	22	54,5	59,3	1020
	caiC RP	5'-GGTGAAGGTCGTCTCCGTCC-3'	20	65	59,8	
degT	degT FP	5'-GCTTGAAGGTGATCTACGATGCT-3'	23	52,2	57,8	915
	degT RP	5'-AAAGCACCTGATGAGACAGGCG-3'	22	54,5	59,7	
fgi	FGI FP	5'-GCGATGTAGAGGGTATTCAGCG-3'	22	54,5	57,4	9860
	FGI RP	5'-AGGAACAATCATCCTCGCCC-3'	20	55	57,2	
glt4	glt4 FP	5'-CTATTGTCACTGCTCCCACAAGC-3'	23	52,2	58	960
	glt4 RP	5'-CCCTTACGATGAAGTCGAAGAGC-3'	23	52,2	57,2	

Table 4. The FGI Types for each gene product. The presence (+) and absence (-) of amplicons for the gene-specific primer pairs in the forty genetically screened *P. ananatis* strains.

				Primary typing genes				Secondary typing genes	
Type	**Nr.**	**Strain**	**Name**	***caiC* (I)**	***mviM* (II)**	***glt3* (III)**	***glt4* (IV)**	***glt2* (I and II)**	***degT* (I, II and III)**
III	S1	BCC0024		-	-	+	-	-	+
I	S2	BCC0030		+	-	-	-	+	+
I	S3	BCC0053		+	-	-	-	+	+
I	S4	BCC0083		+	-	-	-	+	+
I	S5	BCC0084		+	-	-	-	+	+
I	S6	BCC0087		+	-	-	-	+	+
I	S7	BCC0094		+	-	-	-	+	+
?	S8	BCC0098		-	-	-	-	+	+
?	S9	BCC0102		+	+	-	-	+	+
I	S10	BCC0116		+	-	-	-	+	+
I	S11	BCC0127	**LMG20103**	+	-	-	-	+	+
I	S12	BCC0128		+	-	-	-	+	+
I	S13	BCC0132	**LMG2665^T**	+	-	-	-	+	+
?	S14	BCC0135	LMG2676	-	-	-	-	+	+
II	S15	BCC0147	CTB1135	-	+	-	-	+	+
III	S16	BCC0149	LMG2675	-	-	+	-	-	+
?	S17	BCC0151	ATCC35400	-	+	+	-	-	+
I	S18	BCC0155	LMG2807	+	-	-	-	+	+
III	S19	BCC0158	LMG2161	-	-	+	-	-	+
III	S20	BCC0192	**LMG5342**	-	-	+	-	-	+
III	S21	BCC0195	LMG2678	-	-	+	-	-	+
I	S22	BCC0367	THAILAND 7-2	+	-	-	-	+	+
I	S23	BCC0370	DAR76141	+	-	-	-	+	+
I	S24	BCC0372	**DAR76143**	+	-	-	-	+	+
I	S25	BCC0373	DAR76144	+	-	-	-	+	+
I	S26	BCC0375	RAMI7969	+	-	-	-	+	+
III	S27	BCC0583	URUGUAY43	-	-	+	-	-	+
II	S28	BCC0601	ICMP10132	-	+	-	-	+	+
II	S29	BCC0602	ICMP12183	-	+	-	-	+	+
?	S30	BCC0626	DB301	+	-	+	-	+	+
?	S31	BCC0631	BD377	-	-	-	-	-	+
III	S32	BCC0633	**PA4**	-	-	+	-	-	+
II	S33	BCC0635	**BD442**	-	+	-	-	+	+
?	S34	BCC0637	BD561	-	-	-	-	+	+
I	S35	BCC0639	BD588	+	-	-	-	+	+
?	S36	BCC0641	BD622	+	-	+	-	+	+
III	S37	BCC0644	LMG2628	-	-	+	-	-	+
I	S38	SUPP1791		+	-	-	-	+	+
I	S39	SUPP2582		+	-	-	-	+	+
I	S40	BCC1049	**AJ13355**	+	-	-	-	+	+

Table 5. Comparison of the relationships between the FGI Type, geographic origin and *P. ananatis* strain host. The origins of the hosts are abbreviated, USA for United States of America, RSA for Republic of South Africa and Aus for Australia.

#	FABI Culture Collection #	External Culture collection #	Origin	Host	FGI Type
14	BCC0135	LMG2676	USA	Wheat	?
17	BCC0151	ATCC35400	USA	Melons	?
8	BCC0098		USA	Sudangrass	?
9	BCC0102		USA	Sudangrass	?
30	BCC0626	BD301	RSA	Onion	?
31	BCC0631	BD377	RSA	Onion	?
34	BCC0637	BD561	RSA	Maize	?
36	BCC0641	BD622	RSA	Maize	?
2	BCC0030		RSA	Insect	I
3	BCC0053		RSA	Insect	I
4	BCC0083		USA	Onion	I
6	BCC0087		USA	Onion	I
7	BCC0094		USA	Onion	I
5	BCC0084		RSA	*Eucalyptus*	I
10	BCC0116		RSA	*Eucalyptus*	I
11	BCC0127	**LMG20103**	RSA	*Eucalyptus*	I
22	BCC0367	THAILAND 7-2	Thailand	*Eucalyptus*	I
12	BCC0128		USA	Pineapple	I
13	BCC0132	**LMG2665^{T}**	USA	Pineapple	I
18	BCC0155	LMG2807	USA	*Cattleya*	I
23	BCC0370	DAR76141	Aus	Rice	I
24	BCC0372	**DAR76143**	Aus	Rice	I

Table 5. Continued…

#	FABI Culture Collection #	External Culture collection #	Origin	Host	FGI Type
25	BCC0373	DAR76144	Aus	Rice	I
26	BCC0375	RAMI7969	Aus	Rice	I
38	SUPP1791		Japan	Melons	I
39	SUPP2582		Japan	Melons	I
40	BCC1049	**AJ13355**	Japan	Soil	I
41		**B1-9**	Korea	Onion	I
35	BCC0639	BD588	RSA	Maize	I
42		**S7**	Aus	Maize	I
43		**S8**	Aus	Maize	I
44		**S6**	Aus	Maize	I
45		**BRT175**	Canada	Strawberries	I
15	BCC0147	CTB1135	Japan	Rice	II
28	BCC0601	ICMP10132	Brazil	Sugarcane	II
29	BCC0602	ICMP12183	Brazil	Cassia	II
33	BCC0635	**BD442**	RSA	Maize	II
46		**PA13**	Korea	Rice	II
16	BCC0149	LMG2675	Hungary	Wheat	III
21	BCC0195	LMG2678	Zim	Wheat	III
19	BCC0158	LMG2161	India	Rice	III
48		**B40**	Japan	Rice	III
27	BCC0583	URUGUAY43	*Uruguay*	*Eucalyptus*	III
32	BCC0633	**PA4**	RSA	Onion	III
37	BCC0644	LMG2628	Japan	Banana	III
47		**CFH7.1**	USA	Cotton	III
49		**PaMB1**	China	Plant	III
1	BCC0024		RSA	Insect	III
20	BCC0192	LMG5342	USA	Human	III
50		**SD1**	China	Rice	IV

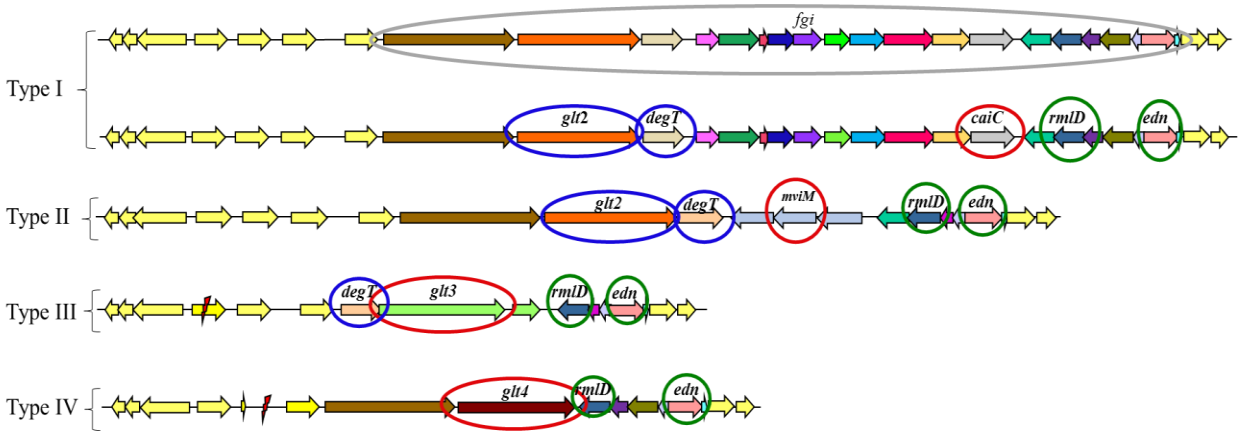


Figure 1. The conserved and variable genes in the different flagellin glycosylation island (FGI) types chosen for genetic screening. Primary type screening gene fragments are indicated in red, while secondary type screening fragments are indicated in blue. The gene fragments utilized to determine the presence of FGI loci are indicated by green circles.

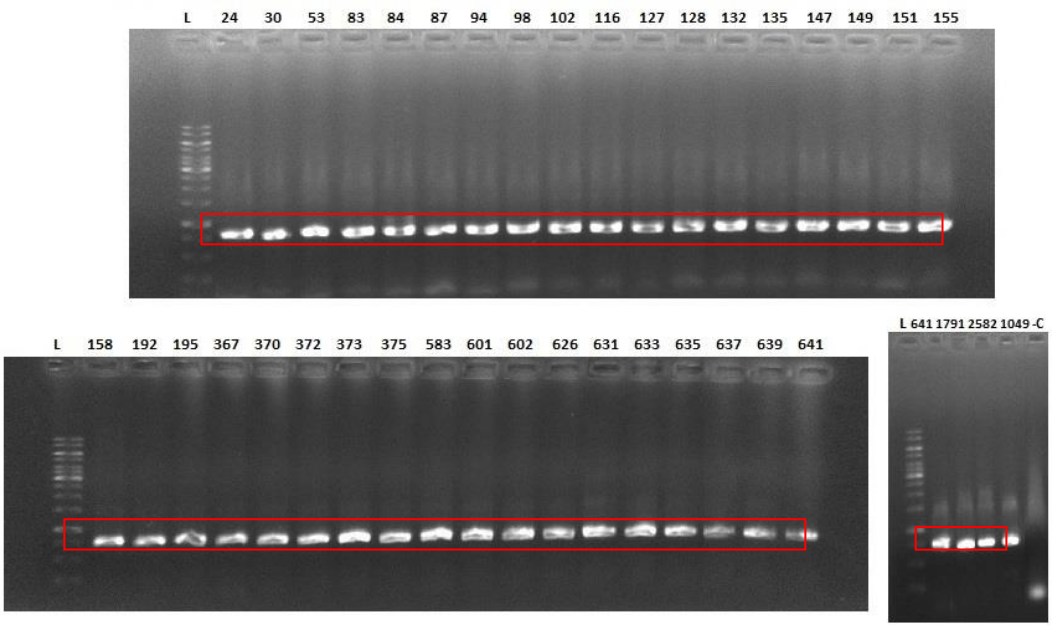


Figure 2. Electrophoresis gel for *edn* amplicons. Indicating the gene-specific Polymerase Chain Reaction (PCR) products of *edn* primer set of forty *P. ananatis* strains, viewed under UV light. The L represents the 1kb DNA ladder (Thermo Fischer) used and –C is the negative control.

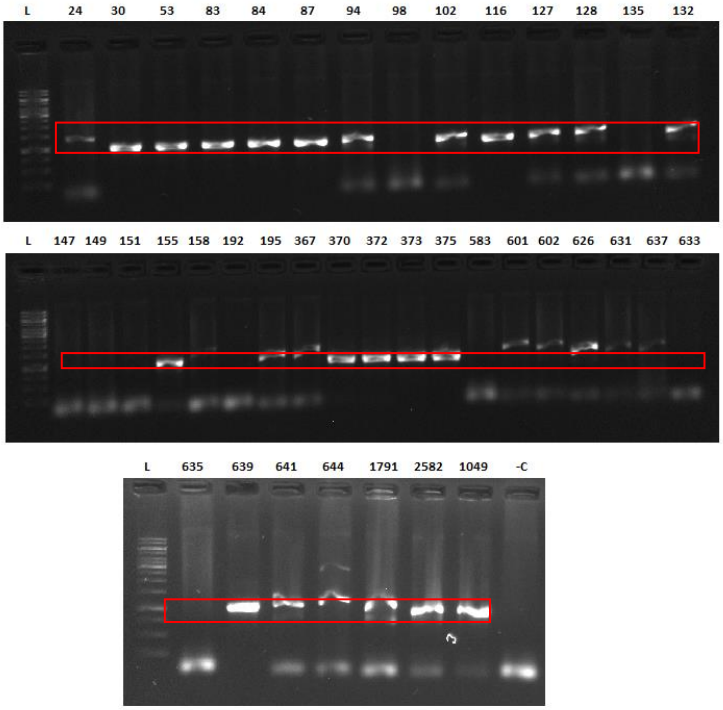


Figure 3. Electrophoresis gel for *caiC* amplicons. Indicating the gene-specific PCR products of *caiC* primer set of forty *P. ananatis* strains, viewed under UV light. The L represents the 1kb DNA ladder (Thermo Fischer) used and –C is the negative control.

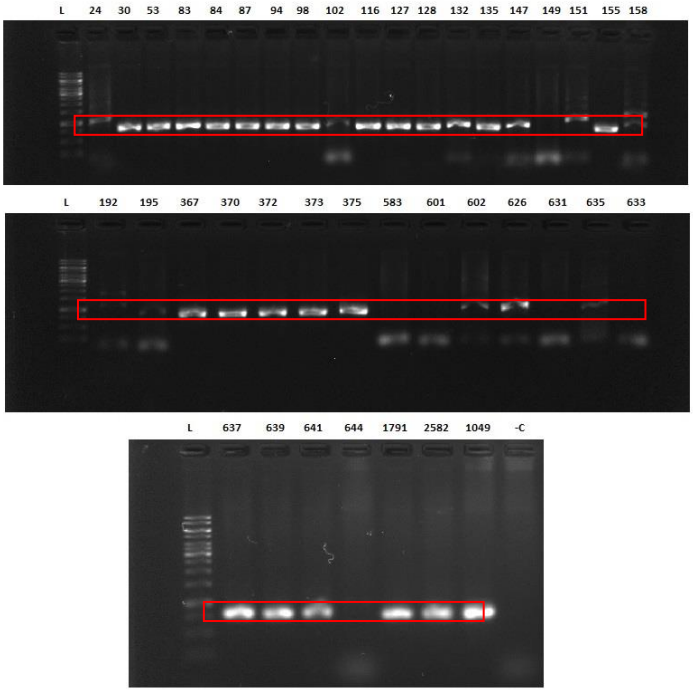


Figure 4. Electrophoresis gel for *glt2* amplicons. Indicating gene-specific PCR products of *glt2* primer set of forty *P. ananatis* strains, viewed under UV light. The L represents the 1kb DNA ladder (Thermo Fischer) used and –C is the negative control.

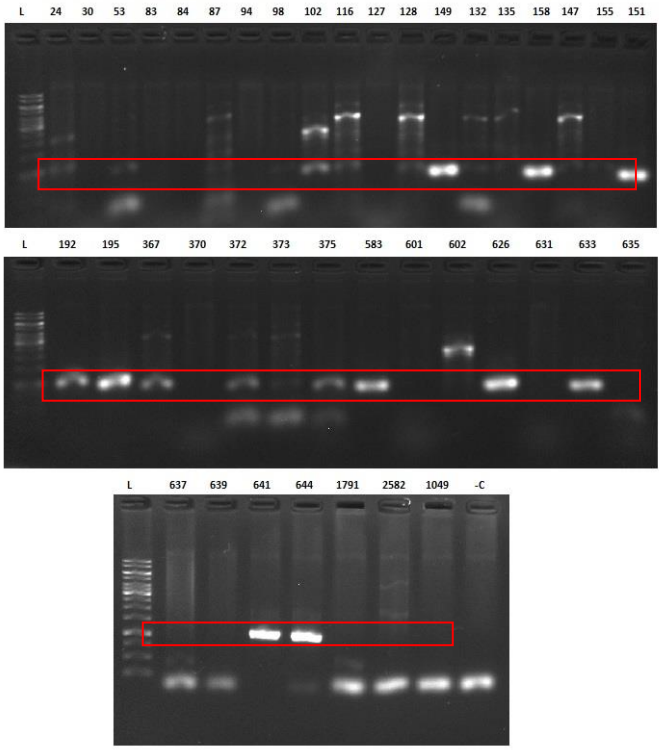


Figure 5. Electrophoresis gel for *glt3* amplicons. Indicating gene-specific PCR products of *glt3* primer set of forty *P. ananatis* strains, viewed under UV light. The L represents the 1kb DNA ladder (Thermo Fischer) used and –C is the negative control.

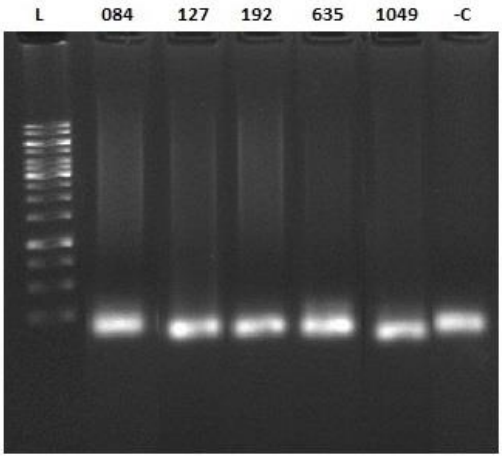


Figure 6. Electrophoresis gel for *fgi* amplicons. Indicating gene-specific PCR products of *fgi* primer set of a small sample of the forty *P. ananatis* strains, viewed under UV light. The L represents the 1kb DNA ladder (Thermo Fischer) used and –C is the negative control. The bright band at the bottom indicates primer dimer.

Figure 7. Geographic location , host and the FGI Type of all the *Pantoea ananatis* strains. *The blue strains are FGI Type I, green strains FGI Type II, purple FGI Type III, red strains FGI Type IV and black strains are untypable.

SUMMARY

Flagellin glycosylation is a common phenomenon among a wide range of bacterial and archaeal lineages. Furthermore, a wide range of functions have been ascribed to this trait, including the assembly, stabilisation and functioning of the flagellum, attachment, host defence avoidance and virulence. In this study, genetic screening and comparative genomic analyses was used to show that flagellin glycosylation islands are universally present among all studied *P. ananatis* strains, regardless of their geographic origin or source of isolation. The genomic analyses undertaken also highlighted the high degree of variability in the genetic architecture of this locus. A typing scheme was developed on the basis of the presence/absence of genes in the islands of the compared strains, and the seventeen genome sequenced strains could be discrimanted into four distinct FGI types. These FGI types differed in terms of the glycosyltransferases encoded on the islands, the sugar biosynthetic genes as well as the gene complement involved in additional methylation, acetylation and formylation of the glycans, which likely translates in highly versatile flagellin glycan structures, containing potentially different sugars with different side chains.

By means of PCR assays designed on the basis of differential gene fragments for each of the FGI types identified from the genomic data, a large number of *P. ananatis* strains isolated from a wide range of environmental sources and geographical locations were screened. The analyses suggested that the diversity of the flagellin glycosylation islands, and thus likely the flagellin glycans themselves, could possibly be capable of numerous functions that stretch beyond the bounds of what was observed here. Furthermore, there was no distinct correlation between the FGI type and the host/geographic locations from which the *P. ananatis* strains were isolated. This may suggest that unlike in some other phytopathogens, flagellin glycosylation does not play a role in host specificity. A role for flagellin glycosylation in virulence is possible, although FGIs were also observed in several isolates from non-plant sources. It should be tested experimentally, however, if these strains are capable of causing disease symptoms when inoculated into a *P. ananatis* host plant. The universal presence of FGIs in *P. ananatis* may also suggest a potential essential role of flagellin glycosylation in the assembly, maintenance or functioning of the flagellum of this species. The global distribution of the different FGI types are very interesting, as this would suggest that the

horizontal gene exchange events from FGIs are purported to have been derived, may have occurred multiple times on several different continents. However, the increased globalization of our planet may also have driven the global distribution of *P. ananatis* strains with the different FGI types. It is thus evident that there are a number of unanswered questions about flagellin glycosylation in *P. ananatis*. At least one of these, is whether the FGI loci are functional and the flagellin glycans are expressed. Future investigations into this fascinating genetic locus must therefore be undertaken.