TABLE OF CONTENTS

CHAPTER 3

COMPARATIVE GENOMICS REVEALS KEY TARGETS FOR ENVIRONMENTAL COLONISATION AND PLANT PATHOGENESIS IN THE WIDE HOST RANGE PATHOGEN *PANTOEA ANANATIS*

CHAPTER 4

IN SILICO IDENTIFICATION AND ANALYSIS OF THE PUTATIVE PATHOGENICITY FACTORS OF *PANTOEA ANANATIS* LMG20103

CHAPTER 5

AN IN DEPTH *IN SILICO* ANALYSIS OF THE TYPE VI SECRETION SYSTEMS OF *PANTOEA ANANATIS* LMG20103

CHAPTER 6

FUNCTIONAL ANALYSIS OF ANANATAN, AN EXOPOLYSACCHARIDE PRODUCED BY *PANTOEA ANANATIS,* WHICH IS HOMOLOGOUS TO STEWARTAN AND AMYLOVORAN AND PLAYS A ROLE IN SYSTEMIC INFECTION OF ONION AND BROWN-ROT DISEASE OF PINEAPPLE

ACKNOWLEDGEMENTS

I would like to extend my heartfelt thanks and appreciation to the following:

- My mom and dad, for your endless love, comfort and understanding. Thank you for giving me the opportunity to do what I have always wanted to do. Dad, I know this thesis comes just too late, but I know you would have been proud of me.

- Joha Grobbelaar, my soon-to-be-wife and my better half. You have been my rock throughout this thesis and I can't wait to share every possible adventure with you.

- My brothers Tim and Maarten and my sister Ivy, as well as my sister in law Gail and my niece Nina.

- My supervisors, Prof. Teresa Coutinho and Prof. Fanus Venter at the University of Pretoria and Dr. Paul Birch at the Scottish Crop Research Institute. We plunged into the deep end of genomics and you have never faltered to give me the support I have needed and gave me the chance to explore and grow into a genomicist.

- Members past and present of Lab 9-35. Thank you for all your help and patience.

- Prof. Emma Steenkamp, Prof. Jaques Theron and Prof. Fourie Joubert. For all their input into my research and my thesis.

- Dr. Ian Toth, Dr. Leighton Pritchard and Dr. Hui Liu. For their help during my research visit to the Scottish Crop Research Institute. To Barbara James, my second mum in Scotland and all my non-Scottish Scotland friends, thank you for making me feel at home in Dundee.

- My best friends Wim, Gregor and Ryan, for their guidance and support.

- Prof. Mike Wingfield and the FABI team, for making me part of the FABI family.

- The National Research Foundation (NRF), the Tree Protection Co-operative Programme (TPCP) and the FABI Centre of Excellence (CoE) for funding this research.

PREFACE

Plant pathogenic bacteria cause devastating diseases on a wide range of agronomic crops and pose a serious threat to food security and the economy. In order to effectively control them and curb the diseases they cause, knowledge is required of the molecular determinants underlying the mechanisms by which they infect, colonise and cause disease symptoms on the host plant. **Chapter 1** discusses the available literature on several phytopathogenic bacteria and their pathogenicity determinants and the techniques used to identify and characterise these pathogenicity factors, with special focus on genome sequencing and post-genomic technologies.

One emerging plant pathogenic bacterium, *Pantoea ananatis*, causes severe diseases on a broad range of agronomically important plants hosts. This pathogen is of particular interest in South Africa as it causes *Eucalyptus* blight and dieback which results in significant economic losses to the forestry industry. Furthermore, *P. ananatis* has been implicated in human disease. Little is known about the pathogenicity determinants it uses to cause disease symptoms. For this reason, the genome of the virulent *Eucalyptus* pathogen *P. ananatis strain* LMG20103 was sequenced. In **Chapter 2**, the sequencing, assembly and annotation of this genome are described. After the full genome was sequenced, it was used for comparative genomic analysis. **Chapter 3** describes the comparison of the LMG20103 genome to those of other plant-pathogenic, as well as animal-pathogenic and non-pathogenic bacteria, in order to answer questions pertaining to the molecular determinants underlying the capacity of this pathogen to survive in almost every environment and cause disease on a wide range of plants as well as humans.

In **Chapter 4**, the whole genome sequence of *P. ananatis* LMG20103 was analysed, in conjunction with the comparative data from the previous chapter, to identify all the genes encoding putative pathogenicity determinants. In this fashion, pathogenicity determinants which have been functionally characterised in other plant pathogenic bacteria could be identified. One interesting pathogenicity determinant identified in Chapter 4, the Type VI secretion system, is likely to play a significant role in pathogenesis. In **Chapter 5**, an in depth *in silico* analysis was performed to study the structures and potential functions of the three Type VI secretion systems in *P. ananatis* LMG20103.

In order to ascertain a pathogenic function for a putative disease determinant, experimental analysis by means of a molecular approach is required. In **Chapter 6**, the function of the exopolysaccharide ananatan, which was identified in Chapter 4, was analysed through knock-out mutation of one of the biosynthetic genes and its function in pathogenesis was determined by infection assays on onion seedlings and pineapple fruits.

Chapter 1:

The pathogenicity factors of plant pathogenic bacteria and their identification and analysis in the genomic era

ABSTRACT

Phytopathogenic bacteria cause diseases on a wide range of agronomic crops, resulting in devastating crop losses and impacting both food security and the economy. With the short-comings of traditional techniques such as cultivar selection and resistance breeding and the use of pesticides in the control of phytopathogenic bacteria, it has been realized that by understanding the mechanisms by which a pathogen causes disease, more directed and effective control measures for plant pathogenic bacteria can be developed. This review focuses on several of the most devastating phytopathogenic bacteria and the pathogenicity determinants they use to infect, colonise and cause symptoms on their host plants. The use of genome sequencing and the latest post-genomic tools for the identification and analysis of these pathogenicity determinants is discussed.

INTRODUCTION

Plant diseases lead to devastating crop losses and have a severe impact on food security, with an estimated 10% global food production loss due to various plant pathogens (Cheatham et al. 2009). The economic effects are also staggering, with plant diseases costing the United States economy approximately US $ 33 billion each year (Pimentel et al. 2000). Plant pathogenic bacteria contribute significantly to the socioeconomic costs of plant diseases, globally affecting innumerable plant species including agronomically important staple food crops, fruit trees, horticultural plants, ornamentals and forestry crops, resulting in severe losses in the field, in greenhouses and during storage of harvested crops (Murillo and Sesma, 2001).

The control and eradication of phytopathogenic bacteria and their diseases is extremely difficult and costly (Murillo and Sesma, 2001). More than US $69 million was spent in the 1900s on the eradication of the citrus canker pathogen *Xanthomonas campestris* pv. *citri* from Florida state, but new outbreaks have since been identified (Shelby, 1998). Control strategies have traditionally relied on preventive measures including husbandry techniques such as crop rotation and limiting the spread of pathogen-infected plant material, the use of agrochemicals and the breeding of resistant crop cultivars (Melchers and Stuiver, 2000). Although having been successfully applied for a number of important plant pathogens, the breeding of resistant cultivars has the drawbacks of being time-consuming, that the selected

8

resistance may be lost and that novel and evolving pathogens can overcome the selected crop resistance (Melchers and Stuiver, 2000). The expediency of agrochemicals is limited due to their high costs, the harmful impacts they may have on the environment and human health and the evolution of tolerant or resistant pathogens (Melcher and Stuiver, 2000).

With the inherent problems of traditional control strategies for phytopathogenic bacteria it has been realised that effective, lasting and novel control strategies can be derived through a better understanding of how a plant pathogen interacts with the plant host and causes disease (Heath, 2001). An understanding of the pathogenicity determinants employed by the pathogen to infect and cause disease on its host will make it possible to develop directed chemical control strategies, biological control with avirulent mutants or non-pathogenic relatives of the pathogen and allow pathogen-specific resistance to be introduced into susceptible host plants through transgenic technologies (Chatterjee and Vidaver, 1986). The characterisation of pathogenicity factors has been greatly enhanced through the advent of molecular techniques. These allow the identification of the genetic basis underlying disease, facilitating the identification of novel pathogenicity determinants and functional analysis of the roles of various pathogenicity factors in disease (Chatterjee and Vidaver, 1986). The recent evolution of the field of genomics has led to the complete genome sequences of a number of important phytopathogens becoming available (Vinatzer and Yan, 2008). These genome sequences provide a complete set of all the genes in the phytopathogen which can be mined, in combination with the latest computational and molecular tools, to identify pathogenicity determinants and determine their functions in plant-pathogen interaction and pathogenesis (Vinatzer and Yan 2008). This information can then be utilised to design and implement control strategies for phytopathogenic bacteria.

PLANT PATHOGENIC BACTERIA

To date more than 100 phytopathogenic bacterial species have been identified, causing a range of diseases including leaf spots, blights, soft rot of fruits and roots, vascular wilts, scabs, tumours and cankers on almost every known cultivated and wild plant host (Murillo and Sesma, 2001). The most economically devastating and thoroughly investigated plant pathogenic bacteria belong to the genera *Erwinia*,

9

Pectobacterium, *Dickeya*, *Pseudomonas* and *Xanthomonas* (Murillo and Sesma, 2001). *Erwinia amylovora* is of great concern to the fruit industry, causing a necrotic disease known as fire blight on Rosaceous hosts including apples, pears and raspberries (Oh and Beer, 2005). *Pectobacterium* and *Dickeya* species cause disease on a wide range of host species. Most significantly they infect potato plants in the field and result in severe post-harvest tuber losses (Toth and Birch, 2005). *Dickeya* strains can reduce potato yields by up to 50% (Laurila et al. 2010). *Pseudomonas syringae* pathovars induce a range of disease symptoms including watersoaking, hypertrophic growth, cankers, chlorosis and leaf necrosis in nearly all cultivated and many wild plant species (Murillo and Sesma, 2001). Similarly, *Xanthomonas* species have been shown to cause disease on 124 monocotyledonous and 268 dicotyledonous plant species (Büttner and Bonas, 2010). Banana wilt caused by *X. campestris* pv. *musacearum* resulted in annual losses of over US $500 million across eastern and central Africa (Tripathi et al. 2010).

PATHOGENICITY FACTORS EMPLOYED BY PLANT PATHOGENIC BACTERIA

Pathogenicity can be defined as the ability of microorganisms to cause disease, thereby allowing them to exploit the nutrients and ecological niche provided by their host (Finlay and Falkow, 1989). Phytopathogenic bacteria have adapted to become proficient, specialised exploiters of the plant host. They do this by means of a complex of pathogenicity factors that allow them to attach and survive on plant surfaces, gain entry into the host, avoid host defenses, invade plant cells, establish themselves, utilise plant nutrients and proliferate within the plant host (Büttner and Bonas, 2003; Finlay and Falkow, 1989). Many phytopathogenic bacteria have adapted to become proficient, specialised exploiters of the plant host.

Plant pathogenic bacteria can be broadly categorised into two groups according to the disease symptoms induced: tumourigenic and necrogenic pathogens. Tumourigenic pathogens provoke abnormal tissue growth in their host by interfering with plant growth signals, thereby forming tumours or galls on the host (Van Sluys et al. 2002). Several important phytopathogens fall into this category, including the crown gall pathogen *Agrobacterium tumefaciens* (Zhu et al. 2000) and *Pantoea agglomerans*.

They do this by the introduction and integration of Ti plasmid DNA into the host genome or by the production of plant hormone-like compounds (Manulis and Barash, 2003). Several plant growth promoting bacteria, including *Sinorhizobium melilotii* and *Mesorhizobium loti* may be considered as tumourigenic as they form tumour-like nodules on the host (Van Sluys et al. 2002). Necrogenic pathogens on the other hand cause plant cell and tissue death. *Pectobacterium atrosepticum*, the causal agent of potato soft rot, is a well-documented example of a pathogen capable of necrogenesis, producing enzymes that break down plant cell wall constituents (Toth et al. 2003). *Pseudomonas syringae* strains produce several toxins which promote necrosis by forming pores in the host cell plasma membrane (Bender et al. 1999).

The mechanisms employed by plant pathogens to infect their host and cause disease symptoms are very diverse. Some factors comprise pathogenicity factors in the broad sense, often indirectly involved in disease and frequently also present in non-pathogenic bacteria. These include motility and chemotaxis, lipopolysaccharides, exopolysaccharides, O-antigens, mechanisms for iron uptake and transport, protection against oxidative stress and regulatory systems involved in the control of pathogenesis (Toth et al. 2006). Other factors are specifically involved in plant disease and are not universally shared. Phytotoxins and cell wall degrading enzymes, specialised exopolysaccharides such as stewartan and amylovoran and the array of secretory systems utilised for the release of specific effector molecules involved in disease are restricted to particular plant pathogens (Toth et al. 2006). Common means of infection have also been identified in animal and plant pathogens, with both making use fimbrial and afimbrial adhesins to attach to host surfaces, using similar secretion systems to inject effectors into host cells and producing homologous lipo- and extracellular polysaccharides (Büttner and Bonas, 2003). Some bacteria have also evolved the ability to cross kingdoms, causing disease in both animal and plant hosts. For example, *A. tumefaciens* has been shown to be able to transfer DNA into the human genome (Lacroix et al. 2006). The necrogenic potato pathogen *Dickeya dadantii* has also been described to be highly virulent on the pea aphid (Grenier et al. 2006). The opportunistic human pathogen *Pseudomonas aeruginosa* infects plants, insect and nematodes (D'Argenio et al. 2001).

Flagellar Motility and Chemotaxis

Most plant pathogens enter their plant hosts through injuries or openings in exposed aerial surfaces. The ability to move towards these sites of infection and away from the inhospitable conditions on exposed plant surfaces is governed by motility through flagella. This flagellar motility is under the control of chemotaxis, where the bacterium is able to perceive and change direction along chemical gradients (Szurmant and Ordal, 2004). In this manner phytopathogenic bacteria can move to their site of infection and ensure their localisation in a safe, nutrient-rich niche.

Motility by means of flagella is an ancient bacterial characteristic. More than 80% of all known bacterial species possess flagella (Soutourina and Bertin 2003). There is enormous diversity in the number, type, location and structure of flagella among bacteria. Most phytopathogenic bacteria express single or multiple polar flagella (Kirov, 2003). The flagellum is a complex, effective organelle. In *Escherichia coli* 49 chromosomally located genes are involved in its biosynthesis, maintenance and functioning (Josenhans and Suerbaum, 2002). It consists of a rotating filament of polymerized subunits of the flagellin protein FliC attached by means of a hook to a basal body consisting of several rings forming an export channel for the export of flagellar elements (Kirov, 2003). The basal body is linked to an inner membrane-embedded MotAB torque-generating motor which in turn is linked to a chemotactic signal transduction cascade (Josenhans and Suerbaum, 2002).

Nutrients such as simple sugars, organic and amino acids are exuded from sites of injury and glandular trichomes and phenols and reactive oxygen species which are released by the plant in response to pathogen attack (Brencic and Winans, 2005). These extracellular attractants and repellents bind to receptors, known as methyl-accepting chemotaxis proteins (MCPs), on the surface of the cell. In the laboratory strain *E. coli* K-12 four genes encoding chemoreceptors responding to distinct chemical signals have been identified. The Tar receptor responds to L-aspartate, dicarboxylic acid and maltose attractants and cobalt and nickel ion repellents, the Trg receptor mediates taxis towards L-serine and away from indole and weak acids, the Tsr receptor mediates taxis towards ribose and galactose and the Tap receptor is involved in taxis towards dipeptides (Taylor et al. 1999). The latter MCP is absent from *Salmonella enterica* strains, which instead carries a Tcp receptor that mediates

taxis towards citrate and away from phenol (Iwama et al. 2006). A further receptor, Aer responds to oxygen concentrations (Szurmant and Ordal, 2004). Changes in attractant/repellent concentrations are transduced from chemoreceptors to the excitation protein CheA which activates the CheY protein. CheY in turn administrates the direction of flagellar rotation. Adaptation to changing concentrations of attractants and repellents occurs through methylation or demethylation of the chemoreceptors by the CheB methylesterase and its antagonist CheR (Taylor et al. 1999). To ensure rapid response to changes in concentrations of nutrients, oxygen or toxic substances the CheZ protein and CheCD complex remove signals allowing prompt perception of new signals (Szurmant and Ordal, 2004).

Motility by means of flagella is an important pathogenicity factor. Its role in plant disease has been described in *A. tumefaciens* and *P. atrosepticum* and mutants of *Pectobacterium carotovorum* subsp. *carotovorum* in genes encoding flagellar components were reduced in their ability to elicit soft rot in Chinese cabbage (Chesnokova et al. 1997; Hossain et al. 2005). The specific role of chemotaxis in pathogenicity has been illustrated in *Ralstonia solanacearum* where pathogenicity of a non-motile flagellin mutant was not different from that of non-mutants, while a chemotactic mutant was less virulent (Yao and Allen, 2006). The role of flagella and chemotaxis in pathogenesis is most significant in the earlier infection phases and is generally not required once infection has been established (Van de Broek et al. 1995). Flagella can also act as adhesive appendages during initial colonization. Once the flagellum becomes tethered it acts as a mechanical signal for upregulation of genes involved in cell invasion, while the motile phenotype is downregulated (Josenhans and Suerbaum, 2002). A role for flagella in biofilm formation has also been described (Kirov, 2003) and the flagellar export apparatus in *Yersinia enterocolitica* acts as a secretory system for the pathogenicity-associated lipoprotein YplA (Young et al. 1999).

Fimbriae, Adhesins and Attachment

In order to colonise a host, bacteria must first adhere to host tissues and surfaces. This is facilitated by adhesins on the outer surface of the bacterium that recognise glycoprotein or glycolipid oligosaccharides on the host cell or tissue and subsequently bind (Proft and Baker, 2008). The effectiveness of adhesins in attachment is however

13

hampered by repulsive forces between the negatively charged bacterium and host cell. This is overcome by fimbriae, filamentous surface appendages consisting of a shaft composed of thousands of pilin subunits with an adhesin located at their tip, that surmount repulsive forces and enable attachment to host surfaces (Proft and Baker, 2008). Fimbriae also function in phage binding, DNA transfer, bacterial aggregation and biofilm formation and twitching and gliding motility (Proft and Baker, 2008). Several outer membrane adhesins not associated with fimbriae, known as non-fimbrial adhesins, also play a major role in attachment to host tissues.

Numerous different types of fimbriae or pili have been identified in bacteria, with many microorganisms hosting more than one type of fimbria. These are broadly classified on the basis of their biosynthetic pathways as chaperone-usher pathway fimbriae, type IV fimbriae and curli pili (Nuccio and Bäumler, 2007). Chaperone-usher assembled fimbriae include the peritrichously distributed type I and type III fimbriae produced by many *Enterobacteriaceae* (Wu and Fives-Taylor, 2001). In uropathogenic *E. coli* type I fimbriae are involved in adhesion to urinary tract cells, cellular aggregation and biofilm formation (Pratt and Kolter, 1998). In the grape and citrus pathogen *Xyllela fastidiosa* they play a role in slowing the rate of movement by twitching motility, a form of movement independent of flagella, when it reaches the site of infection in the xylem (De la Fuente et al. 2007). Flexible polar type IV fimbriae are found in most Gram-negative bacteria and play a role in adhesion, biofilm formation and twitching motility (Craig and Li, 2008). They have been identified in a number of phytopathogens, including *Xanthomonas campestris* and *P. syringae* where they play a role in cell aggregation and protection from environmental stress, rather than adhesion to plant surfaces (Ojanen-Reuhs et al. 1997). Curli pili are coiled surface appendages produced by a number of animal pathogenic *Enterobacteriaceae* via the nucleation/precipitation pathway. They bind to fibronectin, an extracellular protein in mammalian cells (Olsèn et al. 1998). Interestingly, the curli pili of the animal pathogen *Salmonella enterica* are involved in attachment to plant tissues (Barak et al. 2005).

Non-fimbrial adhesins constitute a large group of proteins that are exported via the Type I and Type V secretion systems and become anchored in the outer membrane with an adhesive domain exposed on the cell surface (Gerlach and Hensel, 2007).

These include YadA in *Yersinia* species which binds collagen and fibronectin on the animal host cell membrane, AIDA-I involved in diffuse adherence of uropathogenic *E. coli* to epithelial cells and the filamentous haemagglutinin FhaB in *Bordetella pertussis* (Hoiczyk et al. 2000; Jacob-Dubuisson et al. 2001). Among phytopathogens, non-fimbrial adhesins are restricted to necrogenic pathogens including *P. atrosepticum* and *R. solanacearum* (Bell et al. 2002; Kim et al. 1998). The *Dickeya dadantii* non-fimbrial adhesin HecA plays a role in cell aggregation, which is required for the quorum-sensing dependent production of cell wall degrading enzymes (Rojas et al. 2002).

Iron acquisition, transport and metabolism

Iron is vital for the cellular functioning of all living organisms, playing a role in the tricarboxylic cycle, electron transport, amino acid and nucleotide synthesis, response to oxidative stress and enzyme functioning (Wandersman and Delepelaire, 2004). In the Archaea *Ferroplasma acidiphilum* more than 80% of the proteome consists of iron-containing proteins (Miethke and Mahariel, 2007). In vertebrates and plants iron is maintained in two ionic forms, ferrous (Fe^{2+}) and (Fe^{3+}), frequently complexed with other molecules including citrate and sulphur which act as catalytic centres in enzymatic reactions (Wandersman and Delepelaire, 2004). When Fe^{2+} is reduced in the presence of oxygen, toxic hydroxyl radicals are produced. Vertebrate and plant hosts protect themselves against these compounds by incorporating their iron into storage molecules such as transferrins and ferretins. This iron sequestration, as well as the insoluble nature of Fe^{3+} ions, results in tight iron homeostasis in animals and plants and limited iron availability for pathogenic bacteria (Wandersman and Delepelaire, 2004). Pathogenic bacteria have evolved several intricate mechanisms to solubilise, transport and metabolise iron from their hosts, thereby disturbing the balance of iron within the hosts. This plays a key role in infection. Conversely, withholding iron from the pathogen forms an integral part of host defense in both animals and plants (Ratledge and Dover, 2000). One means of scavenging iron from the host is through the production iron-chelating proteins known as siderophores.

Siderophores are low molecular weight secreted molecules with a high affinity for ferric (Fe^{3+}) iron and ferric salts such as ferric hydroxide and ferric-citrate. More than 500 different types are produced by bacteria, yeasts and fungi, differing in their

specificity, structure, biosynthesis and uptake (Ratledge and Dover, 2000). There is significant diversity in the iron chelating functional groups of siderophores, the most common forms being catechol (e.g. enterobactin and chrysobactin) and hydroxamate (e.g. desferrichrome and desferrioxamine) groups. Other siderophores have a citrate-derived α-hydroxy-carboxylic group (e.g. aerobactin and enterochelin) or aliphatic amine groups (e.g. achromobactin) (Miethke and Mahariel, 2007). The phytopathogen *D. dadantii* produces two siderophores, with achromobactin produced under iron deficiency and in the presence of iron salts, while chrysobactin transports iron attached to weak ligands such as ferric citrate. Achromobactin plays a role at the onset of infection, while chrysobactin is involved in systemic disease (Expert, 1999). Aerobactin provides competitive advantage to *P. carotovorum* in the plant niche (Expert, 1999). *Erwinia amylovora* and *P. agglomerans* use the tripartite siderophore desferrioxamine for iron acquisition and defense against oxidative stress (Dellagi et al. 1998). Siderophore-iron complexes are captured at the bacterial cell surface by high-affinity receptor proteins, including the ferrichrome receptor FhuA, ferrioxamine receptor FoxA, enterochelin receptor FepA and aerobactin receptor Iut (Ratledge and Dover, 2000). Bound iron molecules are then translocated into the periplasm by the Ton system, consisting of ExbB and ExbD and the energy transducer TonB (Larsen et al. 1996). Inner membrane-associated ABC transporters then translocate the iron-siderophore complexes across the inner membrane into the cytoplasm where the iron is dissociated from the siderophores for use by the bacterium (Expert, 1999).

Iron deficiency is detrimental to a pathogenic microorganism. On the other hand, reactive oxygen species which are detrimental to the bacterial cell are produced when iron is present in excessive concentrations (Mey et al. 2005). Precise coordination is thus required for efficient siderophore biosynthesis, siderophore-mediated iron uptake and utilisation. Iron concentrations are perceived by the regulatory protein Fur, a transcriptional repressor that provides fastidious control of iron acquisition and metabolism. Fur also activates the expression of genes silent during iron-sufficient growth including a number of genes encoding pathogenicity determinants (Expert, 1999). In *D. dadantii* Fur negatively regulates genes encoding pectate lyases, which play a key role in its pathogenesis. With a decrease in the intracellular iron pool, Fur repression can upregulate pectate lyase biosynthesis (Expert, 1999).

Quorum-sensing

Pathogenic bacteria do not function as singular cells but make use of an intercellular communication mechanism known as quorum sensing (QS) to coordinate the expression of phenotypes important for pathogenesis by the entire population (de Kievit and Iglewski, 2000). QS functions through the release of small diffusible signalling molecules, autoinducers, into the extracellular medium. When a sufficient number of bacteria produce a threshold quantity of autoinducer, a cognate transcriptional regulator is activated which binds a sequence, the *lux* box, upstream of target genes and induces their expression. In this manner bacterial populations can co-ordinately express pathogenicity proteins (de Kievit and Iglweski, 2000). The main QS molecules produced by Gram-negative bacteria are acyl-homoserine lactones (AHLs), fatty acid derivates with side chains 4 to 15 carbon residues in length (de Kievit and Iglewski, 2000). Differences in side chain length allow bacteria to discriminate their AHLs from those produced by other bacteria (Mohammadi and Geider, 2007). AHLs have the ability to control the expression of various biological functions. This was first noted in *Vibrio fischeri*, a symbiont in the light organ of certain fish and squids (Whitehead et al. 2001). When dense populations of the bacterium are present, high concentrations of the LuxI AHL are produced which bind the transcriptional regulator LuxR and upregulate the expression of the *lux* luciferase genes resulting in bioluminescence (Whitehead et al. 2001). In other organisms, the LuxR protein acts as negative regulator, blocking LuxI from switching on transcriptional activity when at low levels of AHL (von Bodman et al. 1998).

In phytopathogenic bacteria, quorum sensing ensures that copious amounts of pathogenicity determinants are produced at the appropriate time, allowing the pathogen to make a concerted attack and evade host defenses which would be triggered by smaller populations (de Kievit and Iglewski, 2000).Various pathogenicity factors have been shown to be regulated through quorum sensing. In *P. atrosepticum* QS regulates the production of cell wall degrading enzymes (Andersson et al. 2000). In both the fireblight pathogen, *E. amylovora,* and Stewart's wilt pathogen, *Pantoea stewartii* subsp. *stewartii,* production of exopolysaccharides, a major pathogenicity determinant in both pathogens, is QS-regulated (Molina et al. 2005; von Bodman et al. 1998). In the tumourigenic pathogen *P. agglomerans* pv. *gypsophilae* quorum sensing determines the efficiency of gall formation (Chalupowicz et al. 2008).

17

Lipopolysaccharide and Outer membrane components

The outer membrane of Gram negative bacteria is a complex structure that forms an integral, selectively permeable barrier around the bacterium. The major constituents are a phospholipid bilayer situated on the inner surface of the outer membrane and lipopolysaccharides (LPS) which make up the outer layer. Lipoproteins are embedded in the LPS layer (Bos et al. 2007). The outer membrane provides a protective barrier against harmful chemicals, antibiotics and anomalies in pH and osmolarity (Bos et al. 2007). With their location at the outer surface of the bacterium, LPS and lipoproteins come into contact with both plant and animal hosts and thus play important roles in disease (Bos et al. 2007).

LPS is an amphipathic glycolipid that represents the main cell surface component in Gram-negative bacteria (Babinski et al. 2002). It comprises of three structural components: a hydrophilic lipid A moiety which anchors the LPS to the phospholipid, a core oligosaccharide consisting of an inner and outer core domain and an O-side chain (O-antigen) exposed on the cell surface (Babinski et al. 2002). Lipid A is well-conserved among Gram-negative bacteria, including plant pathogens. The core oligosaccharide is more disparate and consists of a complex of variable sugar units (Newman et al. 2000). The O-antigen also consists of highly diverse oligosaccharide repeat units and plays a role in adherence and resistance to complement cell killing (Merino et al. 2000). The immunologically active O-antigen exhibits great variation in the types, arrangement and linkages of sugar subunits, making LPS one of the most variable cell constituents, and has been utilised to classify many pathogens into different serovars (Liu et al. 2008). To date 170 O-antigen groups have been identified in *E. coli* (Scheutz et al. 2004). An additonal antigen, the enterobacterial common antigen is attached to the LPS of Gram-negative enteric bacteria (Kajimura et al. 2005). Several roles for LPS in disease have been described. As pertinent cell constituent, LPS contributes to the selective permeability of the outer membrane and provides a protective barrier to animal pathogenic bacteria against bile salts and enzymatic activity in the intestinal environment (Bos et al. 2007). In the plant context, extrusion of antimicrobial agents of plant origin requires LPS. Increased antibiotic sensitivity has been identified in LPS mutants *in planta* (Dow et al. 2000). The hydrophobic nature of LPS also warrants a role in attachment in plant hosts (Newman et al. 2000) For example, LPS from *R. solanacearum* allows it to bind to tobacco

mesophyll cell walls (Whatley et al. 1980). Conversely, LPS can be recognised by the plant and trigger a defense response (Newman et al. 2000). Some phytopathogenic bacteria have developed means such as the Type III secretion system and its secreted effectors to suppress the defenses induced by LPS (Jakobek and Lindgren, 1993).

Lipoproteins are lipid-modified proteins attached to the cytoplasmic or outer membranes of the bacterial cell. More than 90 lipoproteins are present in *E. coli*, a proportion of which are surface-exposed (Narita et al. 2004). Their main function is ensuring structural integrity of the outer membrane, but they also play a role in uptake of nutrients, attachment to plant and animal cells, binding of bacteriophages and bacteriocins, inducing disease resistance and cytotoxicity (Bos et al. 2007). The most abundant lipoprotein in Gram-negative bacteria, Braun's lipoprotein (Lpp), is important in maintaining cell envelope integrity. In *S. enterica* Lpp contributes to pathogenesis by inducing high levels of pro-inflammatory cytokines, tumour necrosis factor and interleukins, leading to septic shock (Fadl et al. 2005). Other lipoproteins form integral parts of secretion apparatuses involved in export of proteins across the outer membrane, such as VirG in Type III secretion, PulS in Type II secretion and VirB7 in Type IV secretion (Hueck, 1998; Christie et al. 2005). The outer membrane lipoprotein OmlA of *Xanthomonas axonopodis* pv. *citri* has been shown to be involved in plant-microbe interactions (Teixeira-Vanini et al. 2008).

Capsular and extracellular polysaccharides

Bacteria produce an elaborate array of polysaccharides that encapsulate them, creating an advantageous and protective environment for the bacterial cell (Bereswill and Geider, 1997). These occur in two forms. Capsular polysaccharides (CPS) are covalently attached to the phospholipid bilayer or lipid A molecules of LPS (Roberts, 1996). Up to 80 distinct capsular polysaccharides known as K antigens are produced by *E. coli* strains (Schembri et al. 2004). Extracellular polysaccharides (EPS) form a slime layer unattached to the cell, which is nevertheless closely associated with the bacterium (Roberts, 1996). CPS and EPS consist of repeating monosaccharide units of attached to each other through glycosidic linkages. These repeating molecules can consist of the same (homopolysaccharides) or different sugars (heteropolysaccharides) and some of the sugars may be substituted by organic or inorganic molecules (Roberts, 1996). CPS and EPS play a role in protecting bacteria against dessication,

19
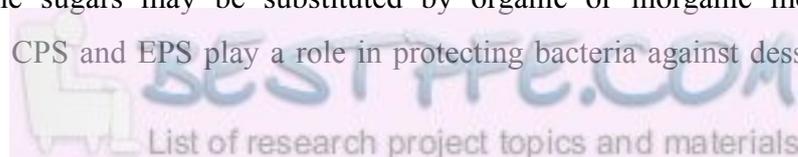
osmotic stress and host defense molecules, predation by other microorganisms and attack by bacteriophages, attachment to surfaces, biofilm formation and mask cell surface components that can elicit host defense responses (Roberts, 1996). CPS and EPS also contribute to the remarkable adaptive capacity exhibited by bacteria that inhabit hostile environments (Politis and Goodman, 1980). EPS and CPS are produced by many phytopathogenic bacteria and represent important pathogenicity determinants, playing many roles in plant-microbe interaction and pathogenesis (Geider, 2000). Well-studied examples include the heteropolysaccharides, amylovoran and stewartan, produced by *E. amylovora* and *P. stewartii,* respectively and the homopolysaccharides cyclic β-1,2 glucan and levan produced by *A. tumefaciens* and *E. amylovora*, respectively.

Stewartan and amylovoran

Stewartan and amylovoran are acidic, heteropolymeric exopolysaccharides that show genetic and structural similarity to *Escherichia coli* group I colonic acid capsules and consist of hepta- and penta-saccharide repeats of galactose, glucuronic acid, glucose and pyruvate residues respectively that form a loosely attached capsule around the cells (Minogue et al. 2002). The principal sugar of both stewartan and amylovoran is galactose, which is not readily available in the plant. Two genes, *galE* and *galF* encode enzymes that convert UDP-glucose into UDP-galactose for use in the EPS backbone (Metzger et al. 1994). A further 12 genes, *amsA-L* and *cpsA-L,* encode enzymes involved in the biosynthesis of amylovoran and stewartan, respectively (Geider, 2000). Roles in survival, colonisation and pathogenesis have been ascribed to stewartan and amylovoran. They bind ions and nutrients, protect against dessication and host defenses and allow the bacteria to elude host recognition (Kim and Geider, 2000; Minogue et al. 2002). Stewartan and amylovoran produced during growth in the intercellular spaces of the leaf results in loss of permeability and bursting of plant cells, leading to accumulation of fluids in the leaf tissues and development of water-soaked lesions and necrosis (Torres-Cabassa et al. 1987). They can also block xylem vessels resulting in wilt of the plant (Bradshaw-Rouse et al. 1981).

Cyclic β-1,2 glucans and levan

Cyclic β-1,2 glucans are produced by a limited number of plant-associated bacteria including *A. tumefaciens*, *Rhizobium* species and *X. campestris* (Breedveld and Miller,

1994). They consist of rings of 17-25 glucose subunits linked by β-1,2-glycosidic bonds and are predominantly localised in the periplasm but are sometimes secreted extracellularly (Breedveld and Miller, 1994). Two proteins are required for cyclic glucan biosynthesis. The ChvB (**ch**romosomal **v**irulence) protein in *A. tumefaciens* and NdvB (**n**odule **d**e**v**elopment) in *Rhizobiaceae* are responsible for initiation, elongation and cyclisation of the glucan. ChvA/NdvA are involved in export of cyclic glucans to the periplasm (Breedveld and Miller, 1994). Several roles for cyclic glucans in both symbiotic and pathogenic plant-microbe interactions have been described. The secreted cyclic glucans of *X. campestris* are transported systemically throughout the plant via the vascular system and suppresses local and systemic host defenses. It acts in a similar fashion to the *P. syringae* Type III effector AvrPto which downregulates expression of *Arabidopsis* genes encoding secreted cell wall components and defense proteins (Rigano et al. 2007). Cyclic glucans also play a role in attachment and nodule formation in *Rhizobium leguminosarum* and tumour development in *A. tumefaciens* (Breedveld and Miller, 1994). *A. tumefaciens chvA/B* mutants could not attach to *Zinnia* leaf mesophyll cells and *S. melilotii* mutants can not colonise alfalfa roots (Breedveld and Miller, 1994). Cyclic glucans also contribute to tolerance to antimicrobial peptides and phytoalexins (Nogales et al. 2006). The homopolysaccharide levan (polyfructosan) is produced by *E. amylovora*, *P. syringae*, *Pantoea agglomerans* and the saprophyte *Pseudomonas fluorescens* (Hettwer et al. 1995). The Levan sucrase (Lsc) enzyme degrades sucrose produced by plants during photosynthesis, releasing the glucose while polymerising fructose residues into levan polysaccharide (Geider, 2000). Mutants in the *lsc* gene exhibit weaker colonisation of pear shoots but are still pathogenic (Geier and Geider, 1993).

Phytotoxins and Phytohormones

Toxins are of particular importance in the pathogenic arsenal of animal and human pathogens. Many phytopathogens also produce secondary metabolites that are toxic to plants (Murillo and Sesma, 2001). These phytotoxins can be glycopeptides, polysaccharides or small metabolites including amino acid analogues (Chatterjee and Vidaver, 1986). Pathovars of *P. syringae* are renowned for the production of a range of phytotoxins. Syringomycin, a nonribosomal peptide, promotes necrosis by pore formation in the host cell membrane (Bender et al. 1999). Coronatine is a type I polyketide synthesised through conjugation of coronafacic acid (CFA) and coronamic

acid (CMA) (Murillo and Sesma, 2001). *Pectobacterium atrosepticum* also encodes a phytotoxic CFA, but lacks CMA genes (Bell et al. 2004). Coronatine is similar in structure to the plant signalling molecule jasmonic acid and by mimicking this molecule allows the pathogen to evade host defense pathways (Toth and Birch, 2005). Coronatine also causes chlorosis, hypertrophy and inhibits root elongation (Uppalapati et al. 2007). Other *P. syringae* toxins such as phaseolotoxin and tabtoxin induce chlorosis, while syringopeptin and syringomycin cause necrosis of plant tissues (Bender et al. 1999). *Xanthomonas albilineans* produces the phytotoxin albicidin, a DNA gyrase inhibitor that induces chlorotic symptoms in sugar cane (Hashimi et al. 2007). 2,5-dihydrophenylalanine produced by *E. amylovora* inhibits the shikimate pathway of plant cells resulting in necrotic symptoms on pear trees (Schwartz et al. 1991).

During plant disease, phytopathogens induce the production of hormones by the plant such as salicylic acid and jasmonate. These are involved in inducing pathogen resistance within the host (Robert-Seilaniantz et al. 2007). Some phytopathogens have evolved the capacity to produce plant hormones themselves, causing imbalances within the plant (Robert-Seilaniantz et al. 2007). This is prevalent among the tumourigenic pathogens *Pseudomonas syringae* pv. savastonoi, *P. agglomerans* and *A. tumefaciens* which produce the phytohormones indole 3-acetic acid (IAA) and cytokinin (Lindow et al. 1998). IAA hormone imbalance results in increased number (hyperplasia) and size (hypertrophy) of plant cells leading to the formation of galls on the plant (Barash and Manulis-Sasson, 2007). IAA production occurs via two pathways, where L-tryptophan is converted into IAA either via the indole acetamide (IAM) or indole-pyruvate (IPyA) pathway (Manulis et al. 1998). The IAM pathway involves the enzymes tryptophan-2-monooxygenase (IaaM) and indole-3-acetimide hydrolase (IaaH) while the IpyA pathway involves an aromatic aminotransferase that converts L-tryptophan to indole pyruvate (IpyA), which is then decarboxylated into IAA (Manulis et al. 1998). *P. agglomerans* pv. *gypsophilae* contains both the IAM and IPyA pathway, where the IAM pathway has been demonstrated to play a role in gall development and the IPyA pathway IAA improves epiphytic fitness by detoxifying surplus tryptophan on plant surfaces, inhibiting the hypersensitive response and down-regulating plant defense genes. IPyA-IAA also loosens the plant cell wall, resulting in leakage of mono- and oligosaccharides (Brandl and Lindow,

1998). Cytokinins are adenine derivatives, synthesised through the isopentenyl transferase Ipt, that induce uncontrolled cell division in the host plant. Mutation of the *P. agglomerans* pv. *gypsophilae ipt* gene resulted in smaller galls being produced (Lichter et al. 1995).

Ice nucleation

Frost-injury results in significant crop losses. Ice crystal formation on plant surfaces can rapidly spread inter- and intracellularly, resulting in mechanical disruption of plant cell membranes and tissues (Lindow et al. 1982). The transition from water to ice can be catalysed by bacterial proteins in a process known as ice nucleation. These ice-nuclei resemble ice crystals and initiate freezing of water at higher temperatures (Edwards et al. 1994). A capacity for ice-nucleation has been identified in a limited number of plant-associated bacteria belonging to the genera *Pseudomonas, Xanthomonas* and *Pantoea* (Edwards et al. 1994). Frost-injury leads to wounds in plant surfaces, creating sites of entry for pathogens into the plant host and making nutrients available (Lindow et al. 1982). Ice nucleating bacteria have also been isolated from clouds, where they increase freezing temperature resulting in rain and snow, thereby facilitating their spread to other host plants (Morris et al. 2008).

Cell wall degrading enzymes

The wall of plant cells consist of a network of three principal polysaccharides: pectin, cellulose and hemicellulose (Van Sluys et al. 2002). Many phytopathogens secrete a range of cell wall degrading enzymes (CWDEs) such as cellulases, xylanases, pectinases and proteases that allow them to break down plant cell wall polymers. Most CWDEs are exported via the Type II secretion system and function extracellularly (Van Sluys et al. 2002). CWDEs are particularly important for pathogenicity in the soft rot pathogens, *P. atrosepticum* and *D. dadantii,* but are produced by many other phytopathogens. *X. campestris* pv. *campestris* produces a replete array of PCWDE including two pectin esterases, two polygalacturonases, five xylanases, four pectate lyases and nine cellulases (Van Sluys et al. 2002).

Pectin is the major polysaccharide present in the primary cell wall and middle lamella between plant cells. It consists of polymers of α-1,4-linked glucuronic acid, L-rhamnose and galacturonic acid which are partially methyl-esterified and acylated (Roy et al. 1999). *Pectobacterium* and *Dickeya* species produce an arsenal of enzymes

23

that systematically degrade pectin, resulting in plant cell bursting and tissue maceration (De Boer, 2003). First, pectin methylesterases and acetylesterases remove methyl and acetyl groups from the pectin backbone (Hugouvieux-Cotte-Pattat et al. 2002). Subsequently, the pectin backbone is degraded by a series of pectate lyases and polygalacturonases, releasing oligogalacturonases that are taken up by the bacterium (Toth et al. 2003). Pectate lyases are the key players in *Dickeya* and *Pectobacterium* soft-rot disease (Roy et al. 1999). The β-1,4-linked glucose homopolymer cellulose provides structural strength to the plant cell wall. Cellulases and β-glucosidases break down the cellulose polymers but are not essential for pathogenesis, rather acting co-ordinately with other CWDEs to assault the host (An et al. 2005; Boccara et al. 1994). Another group of enzymes, the endo-β-1,4-xylanases degrade xylan, the predominant hemicellulose in plant cell walls which is particularly prevalent in monocotyledonous plants (Beliën et al. 2006). Xylanases play a major role in cell wall degradation in rice by *Xanthomonas oryzae* (Beliën et al. 2006). Xylanase A (XynA) produced by *D. dadantii* loosens the cell-wall structure of the host, allowing other secreted enzymes to gain better access to host tissues, rather than participate in the catabolism of hemicellulose (Hurlbert and Preston, 2001). Proteases also partake in the degradation of plant cell wall constituents, breaking down glycoproteins embedded in the plant cell wall. A secreted *Xanthomonas campestris* protease degrades hydroxyproline-rich glycoproteins that provide structural support to the plant cell wall (Dow et al. 1998). Following plant wall polymer degradation, the metabolised intermediates are transported into the bacterial cell and catabolised (An et al. 2005). Periplasmic glucosidases degrade large polymers for translocation into the cytoplasm where cytoplasmic glucosidases convert them into metabolisable sugars (Postma et al. 1993).

Secretion Systems

The interactions between a pathogen and its host are driven by factors present on the surface of the pathogen cell and factors secreted into the extracellular environment or into host cells (Hueck, 1998). In Gram-negative bacteria, six separate secretion systems, Type I-VI, deliver various proteins to the cell surface or into the extracellular milieu. These are found in the majority of plant pathogenic bacteria, although not universally. For example the crown-gall pathogen *A. tumefaciens* does not possess a Type II or III secretion system, while the grape pathogen *Xylella fastidiosa* is lacking a Type III secretion system (Preston et al. 2005). Four out of six secretion systems,

namely the Type I, III, IV and VI systems, transport proteins directly from the cytoplasm to the outside of the cell. The Type II and V secretion systems require preceding systems for translocation of proteins from the cytoplasm into the periplasm and subsequently transfer them across the outer membrane. Translocation across the cytoplasmic membrane is performed by the Sec and Tat pathway (Holland, 2004).

Sec and Tat pathways

The Sec pathway represents the main route for transport of proteins across the cytoplasmic membrane and tranlocates unfolded precursor proteins (Veenendaal et al. 2004). It consists of a membrane-embedded protein-conducting channel made up of the SecY, SecE and SecG proteins and an associated ATPase, SecA. Proteins to be translocated interact with the chaperone SecB and with a complex consisting of SecD, SecF and YajC necessary for efficient protein translocation. YidC then controls the release of proteins from the SecYEG channel once it has been transported across the cytoplasmic membrane (Veenendaal et al. 2004). A second independent pathway runs in parallel with the Sec system. The Twin Arginine Translocation (Tat) pathway transports pre-folded proteins and consists of four proteins, TatA, B, C and E (De Buck et al. 2008). Secreted proteins have a typical twin-arginine motif, S/T-R-R-x-F-L-K in their N-terminal signal peptide and a range of different proteins are transported (De Buck et al. 2008). In *E. coli* ~6% of all secreted proteins are translocated in a Tat-dependent manner (Lee et al. 2006). Substrates include cofactor-binding proteins involved in respiratory and photosynthetic processes, but also many Type II and V secreted pathogenicity proteins and proteins involved in flagellar biosynthesis (De Buck et al. 2008). A role for the Tat pathway in phytopathogenesis has been described for *P. syringae*, with *tat* mutants reduced in siderophores synthesis and motility and attenuated in virulence. In *A. tumefaciens* it was also shown to play a role in chemotaxis in pathogenesis (De Buck et al. 2008).

Type I Secretion System

The Type I secretion system (T1SS) transports proteins of a wide range of functions and ranging in size from 78 to 8,682 amino acids, from the cytoplasm directly to the extracellular environment (Delepelaire, 2004). The translocation system comprises three proteins: an ATP-binding ABC transporter protein positioned in the cytoplasmic membrane, a membrane fusion protein between the inner and outer membrane and a

TolC domain outer membrane protein (Delepelaire, 2004). Secreted proteins include the α-haemolysin of uropathogenic *E. coli,* the iron-acquisition haemophore HasA of *Serratia marcescens* and the SiiE adhesin of *Salmonella enterica* (Gerlach and Hensel, 2007; Delepelaire, 2004). Plant-associated bacteria also make use of the T1SS to secrete proteins involved in plant-microbe interactions, including a *R. solanacearum* adhesin and metalloprotease, the rhizobiocin toxin of *R. leguminosarum* and metalloproteases secreted by *E. amylovora* and *D. dadantii* (Delepelaire, 2004). An *E. amylovora* mutant in the T1SS ABC transporter component *prtD* is disabled in secretion of the protease PrtA and resulted in decreased apple leaf colonisation (Zhang et al. 1999).

Type II Secretion System

The Type II (T2SS) or general secretion pathway exports proteins in two steps, with transport across the cytoplasmic membrane carried out via the Sec and Tat pathways. The second step across the outer membrane is called the terminal branch (Filloux, 2004). The terminal branch is well conserved among Gram negative bacteria and is encoded by 12-16 genes, *gspA-O* and *gspS.* Five of the encoded proteins, GspG-K, share homology with proteins involved in Type IV pilus biosynthesis and have common evolutionary origins (Sauvonnet et al. 2000). These are assembled to form a pilus structure, through which proteins are exported (Filloux, 2004). The secretin GspD forms a pore in the outer membrane through which the pseudopilus extrudes. GspA and B regulate the whole secretion apparatus (Filloux, 2004). A wide array of pathogenicity factors are secreted via the T2SS of animal pathogenic bacteria including pullulanase in *Klebsiella pneumoniae*, cholera toxin in *Vibrio cholerae* and chitinases in *E. coli*. In the phytopathogens *D. dadantii*, *P. atrosepticum* and *X. campestris* cell wall degrading enzymes, proteases and extracellular polysaccharide and the avirulence protein AvrL are Type II secreted (Cianciotto, 2005).

Type III secretion system

Type III secretion systems (T3SS) are found in Gram-negative pathogens of plants and animals and delivers a range of pathogenicity proteins into the cytoplasm of host cells (Preston et al. 2005). In many phytopathogens including *P. syringae, Xanthomonas* spp., soft-rot erwiniae, *E. amylovora* and *P. stewartii* it plays a central role in the interaction with their hosts (Alfano and Collmer, 2004). The T3SS

apparatus of phytopathogenic bacteria consists of a Hrp pilus and is encoded by ~20 *hrp* (**H**ypersensitive **r**esponse and **p**athogenicity) genes organised in 20-40 kb clusters on the genome often together with genes encoding secretion substrates (Preston et al. 2005). Nine of the *hrp* genes, the *hrc* genes, are highly conserved in T3SS biosynthetic loci of both plant and animal pathogens (Alfano and Collmer, 2004). They encode components of the secretion machinery associated with the inner and outer membrane. The extracellular portion of the T3SS differs significantly between animal and plant pathogens. In animal pathogens it consists of a short needle-like protrusion, while phytopathogens have a long flexible Hrp pilus. This is likely due to the fact that they must first penetrate the plant cell wall (Romantschuk et al. 2001).

In animal pathogenic bacteria, the T3SS secretes a number of effectors with different functions, including attachment to host cells, invasion and multiplication in host cells and evasion of host defenses (Galán, 2001). In plant pathogens, Type III secreted effectors mainly function in the manipulation of plant defenses, pathogenesis and influence host specificity (Alfano and Collmer, 2004). The phytopathogens *P. agglomerans* pv. *gypsophilae*, *E. amylovora* and *P. atrosepticum* secrete the T3SS effector DspA, while *P. stewartii* subsp. *stewartii* secretes the homologue WtsE (Ham et al. 2008; Toth et al. 2003). *dspA* mutants are reduced in virulence, while *P. stewartii* WtsE was shown to be required for wilt and watersoaking symptoms in its maize host. DspA/WtsE disrupt host membrane integrity and promote plant cell death as well as suppressing salicylic-acid mediated basal defenses (DebRoy et al. 2004). The heat-stable effector protein HrpN (harpin) also plays a role in pathogenesis. *E. amylovora hrpN* mutants are avirulent, but the harpin is likely to function as a helper for secretion of effectors such as DspA/WtsE and avirulence (Avr) proteins (Alfano and Collmer, 2004). Avr proteins elicit a hypersensitive response (HR) by interacting with a corresponding resistance (R) protein in non-host plants. HR is a rapid defense-associated plant cell death characterised by localised tissue necrosis and phenolic and antimicrobial production at the infection site, preventing further spread of the pathogen (Alfano and Collmer, 2004). In this manner, the host range of Avr-producing pathogens becomes restricted. More than 30 Avr proteins are encoded by *P. syringae* pathovars which cause disease on a restricted range of hosts (Alfano and Collmer, 2004). Some Avr proteins, such as AvrRpt2 of *P. syringae* pv. tomato, can however interfere with host defenses and thus serve as pathogenicity factors (Alfano

and Collmer, 2004). It is hypothesised that Avr proteins first represented pathogenicity factors that suppressed basal host defenses. In response, host plants co-evolved R proteins to recognise Avr proteins, resulting in a HR. The pathogen then evolved novel Avr proteins such as AvrRpt2 which evade R protein recognition and manipulate host defenses. There is thus an intensive molecular arms race between pathogen and host plant (Alfano and Collmer, 2004).

Type IV secretion system

The Type IV secretion system (T4SS) transports infective DNA and proteins into eukaryotic hosts (Christie et al. 2005), playing a role in pathogenicity in animal pathogens such as *Legionella pneumophila* and *Bordetella pertussis* and plant pathogens such as *Xanthomonas axonopodis, A. tumefaciens* and *P. atrosepticum* (Toth et al. 2006). In *A. tumefaciens* the T4SS is encoded by the *virB* operon (*virB1 - B11*) and the *virD* operon (*virD1-D5*). The *virB* cluster encodes an envelope-spanning structure and a transfer pilus (T-pilus) which translocate DNA and proteins. The VirD proteins process the infective DNA to be transferred (Christie et al. 2005). The *A. tumefaciens* T4SS delivers a large single stranded, T-DNA, into the plant where it is integrated into the host genome and results in uncontrolled cell division and formation of crown gall tumours (Christie et al. 2005). Pathogenicity-related proteins secreted by the T4SS in other bacteria include the *Helicobacter pylori* Cag toxin and *L. pneumophila* SidC toxin, while the Ptw Type IV secretion system plays a role in plant disease in the cross-kingdom pathogen *Burkholderia cenocepacia* (Christie et al. 2005). Proteins contributing to nodulation have also been shown to be T4SS delivered by the symbiont *Mesorhizobium loti* (Hubber et al. 2004). Analogous T4SS are involved in the conjugal transfer of plasmid DNA, enabling transfer of antibiotic resistance, fitness and pathogenicity genes among many phytopathogenic bacteria (Christie et al. 2005).

Type V secretion system

The Type V secretion system (T5SS) exports Sec-translocated proteins from the periplasm to the outer bacterial surface in a single ATP-independent step without the requirement of accessory components (Oomen et al. 2004). Secreted proteins, known as autotransporters (AT), consist of an N-terminal signal peptide, an α-domain (passenger) and a C-terminal β-domain (Henderson et al. 2004). The signal peptide

targets the protein to the Sec translocon. Once in the periplasm, the α-domain is transported across the outer membrane via a channel formed by the β-domain (Henderson et al. 2004). Once at the cell surface, the protein is folded and remains attached to the cell surface via the passenger N-terminal. Some passengers contain a protease domain for autoproteolytic cleavage or outer membrane-associated proteases may release the protein extracellularly (Oomen et al. 2004). A mechanical variant of the AT pathway, the two-partner secretion system (TPS) consists of separate α-(TpsA) and β-domain (TpsB) proteins. The TpsB protein forms an outer membrane pore through which the TpsA protein is extruded (Henderson et al. 2004).

Type V secretion proteins range in size from 60-500 kDa and all substrates described to date are involved in pathogenicity. AT pathogenicity determinants include IgA1 proteases in *Neisseria gonorrhoeae* and the AidA-I adhesin in *E. coli* (Henderson et al. 2004). Many phytopathogenic bacteria also encode autotransporters. *P. syringae* has nine AT proteins, including a serine protease, a pertactin-like adhesion, acid phosphatase and lipase, while *P. atrosepticum* encodes two (Preston et al. 2005). TPS proteins include haemolysins such as ShlA in *Serratia marcescens* and adhesins such the filamentous haemagglutinin FhaB of *B. pertussis* (Jacob-Dubuisson et al. 2001). In the soft rot pathogen *D. dadantii*, HecA is TPS secreted and plays a role in adhesion, bacterial aggregation and contact-dependent cell killing (Rojas et al. 2002). A range of TPS proteins have been found in other plant pathogens, including *X. fastidiosa* and *R. solanacearum* (Jacob-Dubuisson et al. 2001).

Type VI secretion system

In 2006, a novel secretion system, termed the Type VI secretion system (T6SS) was identified in *Vibrio cholerae* which induced a cytotoxic effect in the amoeboid model host *Dictyostelium* (Pukatzki et al. 2006). As is the case in the T1SS, T3SS and T4SS, the T6SS delivers effector proteins in a Sec-independent manner (Shrivastava and Mande, 2008). The Type VI secretion apparatus is encoded by 16 conserved genes which are co-localised with effector proteins on the genome (Boyer et al. 2009). Two secreted effector proteins have been identified, VgrG and Hcp. VgrG acts as the puncturing device of the host cell. Three copies of VgrG occur in *V. cholerae* (Pukatzki et al. 2007). The haemolysin co-regulated protein Hcp forms a continuous transportation channel from the cytoplasmic membrane, through the bacterial outer

29

membrane and into the outer membrane of the host cell. As is the case for VgrG, Hcp is also an important cytotoxic effector of the T6SS, suggesting bifunctionality as structural and effector protein (Pukatzki et al. 2006).

T6SSs have been observed in 42 bacterial species belonging to Proteobacteriaceae, although these differ significantly in gene content and structure from the system in *V. cholerae* (Shrivastava and Mande, 2008). It has been shown to play a major role in diseases caused by the animal pathogens *Burkholderia mallei*, *Edwardsiella tarda* and *Pseudomonas aeruginosa* (Schell et al. 2007; Zheng and Leung, 2007; Mougous et al. 2007). Hcp secretion via the T6SS has been described in the phytopathogen *A. tumefaciens* where mutational elimination of Hcp led to reduction in gall formation efficiency. In *P. atrosepticum*, the T6SS did not affect virulence unless Hcp secretion was over-expressed (Wu et al. 2008; Mattinen et al. 2008). In *R. leguminosarum* Type VI secretion blocks infection, colonisation and nodule formation on pea plants and may act as host range determinant (Bladergroen et al. 2003).

Protection against plant defense mechanisms

Plants utilise an extensive arsenal of defense mechanisms, both active and passive, to protect themselves against pathogen attack. The pre-existing passive defenses include structural barriers and antimicrobial phytoanticipins such as catechol and protocatechuic acid which prevent colonization of host tissues by the pathogen (VanEtten et al. 2001). Active or induced defenses include the production of phytoalexin antimicrobials, phenolic acids, ion influxes into host cells, reinforcement of the plant cell wall and callose deposition, oxidative burst and induction of the hypersensitive response (HR), a form of programmed plant cell death which restricts pathogen spread (Abramovitch and Martin, 2004).

In order to successfully invade a plant, phytopathogens must be able to circumvent the phytoanticipins, phytoalexins and other plant antimicrobials. A number of strategies meet this purpose, including enzymatic inactivation of antimicrobials, modification of the target structure, diminished influx of toxic metabolites and utilization of efflux pumps to expel antimicrobial agents (Walsh, 2000). In *E. amylovora*, a phytoalexin-inducible multidrug efflux pump provides resistance to several phytoalexins produced by apple trees in response to pathogen attack (Burse et al. 2004a). The RmrAB efflux pump in the plant symbiont *Rhizobium etli* mediates resistance to phytoalexins

produced by legumes (González-Pasayo and Martínez-Romero, 2000). A cluster of genes in *D. dadantii*, *sapA-F*, provides resistance to several potato-produced antimicrobial peptides, including snakin-1 and α-thionin (López-Solanilla et al. 1998).

The oxidative burst is one of the most rapid plant defense responses initiated following pathogen attack. It involves the production of reactive oxygen species (ROS) such as superoxide (O_2^-) and hydrogen peroxide (H_2O_2) by the plant at the site of invasion (Loprasert et al. 1996). ROS inactivate metabolic enzymes, oxidise proteins and damage DNA in the pathogen (Acuña et al. 2009). H_2O_2 can penetrate the bacterial membrane and react with iron to form highly toxic hydroxyl radicals (Hanyu et al. 2009). Phytopathogenic bacteria use several mechanisms to overcome oxidative stresses invoked by plant defense, including the production of superoxide dismutases and catalases that degrade H_2O_2 (Loprasert et al. 1996). Enzymes such as methionine sulphoxide reductase (MsrA) repair proteins damaged by ROS. *D. dadantii* deletion mutants in *msrA* are impaired in their pathogenicity, as ROS-damaged pathogenicity-related proteins can not be repaired (El Hassouni et al. 1999). The *D. dadantii* pigment indigoidine also plays a role in protection against oxygen radicals, as does the Suf system which is involved in iron acquisition and metabolism (Reverchon et al. 2002; Nachin et al. 2001).

Pathogenicity regulation

Many phytopathogenic bacteria exist as epi- or endophytes on plants or as soil saprophytes until the opportunity for host infection arises. To ensure energy conservation, host defense evasion and effective disease progression, the phytopathogens must express pathogenicity determinants in a coordinated and timely fashion (Mole et al. 2007). Pathogenic bacteria have evolved specific transcriptional regulators to ensure appropriate expression of individual pathogenicity determinants. A broader "global" pathogenicity regulation ensures coordinate expression of multiple pathogenicity factors in response to environmental signals, enabling phenotypic adaptation for effective colonisation and disease induction (Mole et al. 2007).

One of the most effective families of global pathogenicity regulators are the two-component regulatory systems (TCS) which couple a range of different signals, via phosphorylation of their partner-proteins, to an array of pathogenicity effectors, up- or down-regulating them accordingly (Cotter and DiRita, 2000). The best known

example, the PhoP-PhoQ TCS controls almost all identified pathogenicity factors in *Salmonella*. PhoQ represents the sensor and responds to Ca^{2+} and Mg^{2+}, signaling whether *Salmonella* is located intra- or extracellularly to the response regulator PhoP by phosphotransfer, which subsequently regulates the expression of pathogenicity genes (Cotter and DiRita, 2000). Other TCS such as BvgS-BvgA from *Bordetella pertussis* responds to temperature, SO_4^{2-} and nicotinic acid (Cotter and DiRita, 2000). In the phytopathogen *A. tumefaciens* the VirAG TCS regulates *vir* gene expression involved in Type IV secretion (Winans, 1992). In the soft rot pathogens *D. dadantii* and *P. atrosepticum* RsmAB represents the most critical posttranslational regulatory system, controlling the production of CWDEs and Type III effectors (Liu et al. 1998). This same system controls extracellular polysaccharide production in *E. amylovora* and cytokinin biosynthesis in *P. agglomerans* pv. *gypsophilae* (Ma et al. 2001). Transcriptional control is provided by histone-like proteins and alternative sigma factors (Hildebrand et al. 2006). The small nucleoid-associated histone-like protein H-NS is a DNA-binding protein that regulates motility, growth rate and pectate lyase and cellulase production in *P. atrosepticum* (Nasser et al. 2001). Overexpression of *hns* in *E. amylovora* results in reduced virulence, as H-NS acts as transcriptional repressor of EPS and levan production (Hildebrand et al. 2006). The alternative sigma factor *hrpL* is involved in activation of Type III secretion (Mole et al. 2007).

IDENTIFICATION AND ANALYSIS OF PATHOGENICITY FACTORS IN PLANT PATHOGENIC BACTERIA

The identification of pathogenicity determinants and elucidation of their function is imperative to the understanding of a bacterial plant disease and the development of control strategies against the pathogen (Chatterjee and Vidaver, 1986). The development of genetic technologies in the 1970s-80s contributed significantly to the understanding of the phytopathogenic mechanism and plant-microbe interactions (Chatterjee and Vidaver, 1986). The use of random and site-direct mutagenesis enabled the elucidation of the functions of numerous genes utilised by plant pathogens to infect, colonise and induce symptoms on their plant hosts (Chatterjee and Vidaver, 1986). The recent inception of whole genome sequencing has revolutionised the field of phytobacteriology, with novel inexpensive sequencing technologies making this technology available to many research groups (Vinatzer and Yan, 2008). These whole

genomes provide an extensive resource for the identification of the candidate pathogenicity determinants used by a wide range of phytopathogenic bacteria and can be used in combination with novel and existing genetic and genomic technologies to determine their role in plant-microbe interactions and pathogenesis (Toth et al. 2003).

4.1 <u>Genome sequencing</u>

The bacterial genomics era saw its inception with the release of the entire sequence of the pulmonary pathogen *Haemophilus influenzae* (Fleischmann et al. 1995). Since this first genome, there has been a veritable explosion in the field of genomics. To date 1,099 bacterial genomes have been completed and validly published, while 4,626 bacterial genome sequencing projects are currently in progress (Genomes Online Database, July 2010; Kyrpides, 1999). These numbers are increasing at a staggering rate concomitant with the development of novel genome sequencing technologies. The first genomes were sequenced using whole genome shotgun sequencing. This approach involves the generation of a library of small genomic fragments that are cloned and subsequently sequenced using the dideoxynucleotide chain termination method developed by Sanger (Sanger et al. 1977; Mardis, 2008). The Sanger method is the most widely used sequencing technique but is hampered by high cost and labour intensiveness. High-throughput sequencing by this method is thus restricted to a limited number of highly-specialised laboratories (Mardis, 2008). Novel high-throughput sequencing technologies such as 454 Pyrosequencing, Illumina's Solexa technology and the Applied Biosystems SOliD platform circumvent the time-consuming requirement of generating libraries and can be used to sequence whole genomes at a fraction of the cost of shotgun sequencing and in a very short time (von Bubnoff, 2008). Further research and development of novel sequencing techniques, such as nanopore sequencing technology which is currently being explored, will continue to make genome sequencing more rapid, cost-effective and accessible to the scientific community (von Bubnoff, 2008).

Although they account for a limited fraction of the bacterial realm, three-quarters of the completed bacterial genomes belong to medically important representatives of the Kingdom (Doolittle, 2002). This is justified not only by their importance, but also by the huge contributions these genome sequences have made towards the understanding of their pathogenesis and treatment and control of their diseases (Doolittle, 2002). By

33

contrast only 46 phytopathogenic bacteria genomes have been sequenced, accounting for a mere 4.2% of all complete bacterial genomes (Genomes Online Database, July 2010). The first phytobacterial pathogen genome sequenced was that of the citrus variegated chlorosis causal agent, *Xylella fastidiosa* (Simpson et al. 2000). Subsequently, the sequences of the phytopathogens *Agrobacterium tumefaciens* C58, causal agent of crown-gall and the bacterial-wilt pathogen *Ralstonia solanacearum* GMI1000 (Wood et al. 2001; Salanoubat et al. 2002) were published. Other important phytopathogenic bacteria whose genomes have been sequenced include the potato soft rot pathogen *Pectobacterium atrosepticum* SCRI1043, the fire blight pathogen *Erwinia amylovora* CFBP1430, the crucifer black rot pathogen *Xanthomonas campestris* pv. *campestris* ATCC33913 and the tomato and *Arabidopsis* pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (Bell et al. 2004; Smits et al. 2010; da Silva et al. 2002; Buell et al. 2003). Given the impact of emerging and existing phytopathogenic bacteria on food security and the economy and the immeasurable information gained from sequencing their genomes, combined with the reduced time and cost of genome sequencing with novel technologies, many more phytopathogen genomes sequences can be expected in the near future. Currently, the sequencing of a further 76 genomes of phytopathogenic bacteria is being undertaken (Genomes Online Database, July 2010).

4.2 <u>Genome Mining and Postgenomic analysis</u>

With the sequencing of their genomes, the door has been opened for post genomic analysis of many plant pathogens. Postgenomic analysis involves the correlation of aspects encoded on their genome to a phenotype (Vinatzer and Yan, 2008). Once the genome sequence is available, putative genes are identified within the sequence and are annotated in accordance with similar genes in other bacteria. Subsequently, the annotated genome can be mined for genes encoding proteins that underly a particular phenotype and through molecular analyses their function can be confirmed. In this manner numerous aspects encompassed in the plant-pathogen interaction can be analysed, including the functional involvement of various pathogenicity factors in plant disease (Vinatzer and Yan, 2008). A great number of tools have been developed to mine and analyse the genome sequence.

4.2.1 *Comparative genomics*

The true potential of genomics can be seen when the sequences of closely related organisms with different phenotypes are compared. With the availability of the genome sequences of closely related bacteria, differences in their phenotype can be correlated with discrepancies in the genome sequences (Vinatzer and Yan, 2008). Comparison of phytopathogenic bacteria with close non-pathogenic relatives can elucidate the basis of pathogenicity (Vinatzer and Yan, 2008). Comparison of the genomes of the non-pathogenic *Erwinia tasmaniensis* Et1/99 and its virulent relative *E. amylovora* Ea273 revealed the absence of several pathogenicity determinants in *E. tasmaniensis* that are required for *E. amylovora* fire blight (Kube et al. 2008). Similarly, comparison of the genomes of strains of the same species with different host ranges can provide information on host specificity. Differences were observed between the genomes of two *X. campestris* pv. *campestris* strains which could be experimentally linked to their host ranges (Qian et al. 2005). Developments in the field of bioinformatics have led to a large number of software packages and online tools becoming available to perform genome comparisons (Vinatzer and Yan, 2008).

4.2.2 *Postgenomic analysis*

In order to correlate a protein function to a phenotype, experimental analysis is required. A number of tools have been developed and successfully applied for this purpose. One of the first techniques to be developed was mutational analysis, where through the physical or chemical disruption of the gene encoding a particular protein its role in a phenotype could be determined. In this manner, phytopathogenic bacteria could be disrupted in a gene encoding a pathogenicity factor, resulting in an avirulent mutant (Chatterjee and Vidaver, 1986). One of the most utilised means of creating a mutant is through the introduction of transposons, DNA fragments with an inherent capacity to move from one location in a genome to another (Judson and Mekalanos, 2000). These transposons can disrupt the genomic region within which they are inserted (Judson and Mekalanos, 2000). If insertion occurs within a protein coding sequence, the protein is disrupted and its function can be analysed by screening for a change in phenotype (Judson and Mekalanos, 2000). This technique does however have the drawback that in the absence of genetic information of the studied organism, random transposon insertion must be performed and labour-intensive screening is required to identify genes that encode proteins linked to a particular phenotype

35

(Chatterjee and Vidaver, 1986). The advent of gene sequencing technologies, however, allowed the development of more targeted approaches to mutate genes of interest (Judson and Mekalanos, 2000). With the availability of whole genomes sequences, rapid and large-scale screening to identify particular mutants has become possible. For example, using the genome sequence of *P. atrosepticum* SCRI1043 as a backbone, a random transposon library was constructed and was used to screen for insertions in numerous putative pathogenicity genes identified in the genome annotation process. Mutants in these pathogenicity genes were selected and used in plant infection trials to uncover the roles of the encoded proteins in pathogenesis (Bell et al. 2004; Holeva et al. 2004).

Other technologies can also be applied for the postgenomic analysis of pathogenicity factors and plant-pathogen interactions. Microarray technology has been used to rapidly identify genomic differences between related phytopathogens. It can also be used for transcriptomics, to determine those genes that are expressed during plant infection (Lockhart and Winzeler, 2000). By performing microarray analysis with a transposon mutant in a given pathogenicity regulator, its effect on the expression of other pathogenicity factors can also be determined (Lockhart and Winzeler, 2000). Proteomics offers another powerful postgenomic tool and involves direct investigation of the protein content in a cell (Quirino et al. 2010). It has been successfully applied to analyse the roles of various plant pathogen proteins, such as cell surface components or secreted pathogenicity effectors in plant-microbe interaction and pathogenesis (Quirino et al. 2010). With the development of new postgenomic technologies and novel and improved applications of existing postgenomic tools, it can be expected that a more complete picture of how phytobacterial pathogens interact and cause disease in plant hosts will be created in the near future.

CONCLUSIONS

Phytopathogenic bacteria cause severe diseases on an extremely broad range of plants and have a devastating impact on the economy and food security. Nevertheless, the mechanisms by which they cause disease and interact with their plant hosts is poorly understood. The recent sequencing of the whole genomes of several agronomically important plant pathogenic bacteria has highlighted the complex pathogenicity determinants employed by the pathogen to infect, colonise and induce symptoms on their hosts. Whole genome sequencing and postgenomic analysis can be used to identify the various pathogenicity factors utilised by different phytopathogenic bacteria and will create a better understanding of how they cause disease. This information can then be used in turn to develop durable and effective control strategies against these important plant pathogens.

REFERENCES

Abramovitch, R. B. and Martin, G. B. 2004. Strategies used by bacterial pathogens to suppress plant defenses. Curr. Opin. Plant Biol. **7**: 356-364

Acuňa, L. G., García-Calderón, I. L., Elías, A. O., Castro, M. E. and Vásquez, C. C. 2009. Expression of the *yggE* gene protects *Escherichia coli* from potassium tellurite-generated oxidative stress. Arch. Microbiol. **191**: 473-476

Alfano, J. R. and Collmer, A. 2004. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. Annu. Rev. Phytopathol. **42**: 385-414

An, C. L., Lim, W. J., Hong, S. Y., Shin, E. C., Kim, M. K. et al. 2005. Structural and biochemical analysis of the *asc* operon encoding 6-phospho-β-glucosidase in *Pectobacterium carotovorum* subsp. *carotovorum* LY34. Res. Microbiol. **156**: 145-153

Andersson, R. A., Eriksson, A. R. B., Heikinheimo, R., Mäe, A., Pirhonen, M. et al. 2000. Quorum sensing in the plant pathogen *Erwinia carotovora* subsp. *carotovora*: the role of *expR$_{Ecc}$*. Mol. Plant-Microbe Interact. **13**: 384-393

Babinski, K. J., Ribeiro, A. A. and Raetz, C. R. H. 2002. The *Escherichia coli* gene encoding the UDP-2,3-diacylglucosamine pyrophosphatase of Lipid A biosynthesis. J. Biol. Chem. **277**: 25937-25946

Barak, J. D., Gorski, L., Naraghi-Arani, P. and Charkowski, A. O. 2005. *Salmonella enterica* virulence genes are required for bacterial attachment to plant tissues. App. Environ. Microbiol. **71**: 5685-5691

Barash, I. and Manulis-Sasson, S. 2007. Virulence mechanisms and host specificity of gall-forming *Pantoea agglomerans*. Trends Microbiol. **15**: 538-545

Beliën, T., Van Campenhout, S., Robben, J. and Volckaert, G. 2006. Microbial endoxylanases: effective weapons to breach the plant cell-wall barrier or, rather, triggers of plant defense systems? Mol. Plant-Microbe Interact. **19**: 1072-1081

Bell, K., Sebaihia, M., Pritchard, L., Holden, M. T. G., Hyman, L. J. et al. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. Proc. Natl. Acad. Sci. USA **101**: 11105-11110

Bell, K. S., Avrova, A. O., Holeva, M. C., Cardle, L., Morris, W. et al. 2002. Sample sequencing of a selected region of the genome of *Erwinia carotovora* subsp. *atroseptica* reveals candidate phytopathogenicity genes and allows comparison with *Escherichia coli*. Microbiology **148**: 1367-1378

Bender, C. L., Alarcón-Chaidez, F. and Gross, D. C. 1999. *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol. Mol. Biol. Rev. **63**: 266-292

Bereswill, S. and Geider, K. 1997. Characterization of the *rcsB* gene from *Erwinia amylovora* and its influence on exopolysaccharide synthesis and virulence of the fire blight pathogen. J. Bacteriol. **179**: 1354-1361

Bladergroen, M. R., Badelt, K. and Spaink, H. P. 2003. Infection-blocking genes of a symbiotic *Rhizobium leguminosarum* strain that are involved in temperature-dependent protein secretion. Mol. Plant-Microbe Interact. **16**: 53-64

Boccara, M., Aymeric, J.-L. and Camus, C. 1994. Role of endoglucanases in *Erwinia chrysanthemi* 3937 virulence on *Saintpaulia ionantha*. J. Bacteriol. **176**: 1524-1526

Büttner, U. and Bonas, D. 2003. Common infection strategies of plant and animal pathogenic bacteria. Curr. Opin. Plant Biol. **6**: 312-319

Bos, M. P., Robert, V. and Tommassen, J. 2007. Biogenesis of the Gram-negative bacterial outer membrane. Annu. Rev. Microbiol. **61**: 191-214

Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. and Attree, I. 2009. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? BMC Genomics **10**: 104-117

Bradshaw-Rouse, J. J., Whatley, M. H., Coplin, D. L., Woods, A., Sequeira, L. et al. 1981. Agglutination of *Erwinia stewartii* strains with a corn agglutinin: correlation

39

with extracellular polysaccharide production and pathogenicity. Appl. Environ. Microbiol. **42**: 344-350

Brandl, M. T. and Lindow, S. E. 1998. Contribution of indole-3-acetic acid production to the epiphytic fitness of *Erwinia herbicola*. Appl. Environ. Microbiol. **64**: 3256-3263

Breedveld, M. W. and Miller, K. J. 1994. Cyclic β-glucans of members of the family *Rhizobiaceae*. Microbiol. Rev. **58**: 145-161

Brencic, A. and Winans, S. C. 2005. Detection of and Response to Signals Involved in Host-Microbe Interactions by Plant-Associated Bacteria. Microbiol. Mol. Biol. Rev. **69**: 155-194

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T. et al. 2003. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. Proc. Natl. Acad. Sci. USA **100**: 10181-10186

Burse, A., Weingart, H. and Ullrich, M. S. 2004a. The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, *Erwinia amylovora*. Mol. Plant-Microbe Interact. **17**: 43-54

Büttner, D. and Bonas, U. 2010. Regulation and secretion of *Xanthomonas* virulence factors. FEMS Microbiol. Rev. **34**: 107-133

Chalupowicz, L., Manulis-Sasson, S., Itkin, M., Sacher, A., Sessa, G. et al. 2008. Quorum-sensing system affects gall development incited by *Pantoea agglomerans* pv. *gypsophilae*. Mol. Plant-Microbe Interact. **21**: 1094-1105

Chatterjee, A. K. and Vidaver, A. K. 1986. Genetics of pathogenicity factors: application to phytopathogenic bacteria. In <u>Advances in plant pathology volume 4:</u> <u>Application to phytopathogenic bacteria,</u> Orlando, Academic Press**:** 1-218.

Cheatham, M. R., Rouse, M. N., Esker, P. D., Ignacio, S., Pradel, W. et al. 2009. Beyond yield: plant disease in the context of ecosystem services. Phytopathology **99**: 1228-1236

Chesnokova, O., Coutinho, J. B., Khan, I. H., Mikhail, M. S. and Kado, C. I. 1997. Characterization of flagella genes of *Agrobacterium tumefaciens*, and the effect of a bald strain on virulence. Mol. Microbiol. **23**: 579-590

Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S. and Cascales, E. 2005. Biogenesis, architecture, and function of bacterial type IV secretion systems. Annu. Rev. Microbiol. **59**: 451-485

Cianciotto, N. P. 2005. Type II secretion: a protein secretion system for all seasons. Trends Microbiol. **13**: 581-588

Cotter, P. A. and Dirita, V. J. 2000. Bacterial virulence gene regulation: an evolutionary perspective. Annu. Rev. Microbiol. **54**: 519-565

Craig, L. and Li, J. 2008. Type IV pili: paradoxes in form and function. Curr. Opin. Struct. Biol. **18**: 267-277

da Silva, A. C. R., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R. et al. 2002. Comparison of the genomes of two *Xanthomonas pathogens* with differing host specificities. Nature **417**: 459-463

D'Argenio, D. A., Gallagher, L. A., Berg, C. A. and Manoil, C. 2001. *Drosophila* as a model host for *Pseudomonas aeruginosa* infection. J. Bacteriol. **183**: 1466-1471

De Boer, S. H. 2003. Characterization of pectolytic erwinias as highly sophisticated pathogens of plants. Eur. J. Plant Pathol. **109**: 893-899

De Buck, E., Lammertyn, E. and Anne, J. 2008. The importance of the twin-arginine translocation pathway for bacterial virulence. Trends Microbiol. **16**: 442-453

de Kievit, T. R. and Iglewski, B. H. 2000. Bacterial quorum sensing in pathogenic relationships. Infect. Immun. **68**: 4839-4849

De La Fuente, L., Burr, T. J. and Hoch, H. C. 2007. Mutations in Type I and Type IV pilus biosynthetic genes affect twitching motility rates in *Xylella fastidiosa*. J. Bacteriol. **189**: 7507-7510

DebRoy, S., Thilmony, R., Kwack, Y.-b., Nomura, K. and He, S. Y. 2004. A family of conserved bacterial effectors inhibits salicylic acid-mediated basal immunity and promotes disease necrosis in plants. Proc. Natl. Acad. Sci. USA **101**: 9927-9932

Delepelaire, P. 2004. Type I secretion in Gram-negative bacteria. Biochim. Biophys. Acta **1694**: 149-161

Dellagi, A., Paulin, J.-P. and Expert, D. 1998. Dual role of desferrioxamine in *Erwinia amylovora* pathogenicity. Mol. Plant-Microbe Interact. **11**: 734-742

Doolittle, R. F. 2002. Microbial genomes multiply. Nature **416**: 697-700

Dow, J. M., Davies, H. A. and Daniels, M. J. 1998. A metalloprotease from *Xanthomonas campestris* that specifically degrades proline/hydroxyproline-rich glycoproteins of the plant extracellular matrix. Mol. Plant-Microbe Interact. **11**: 1085-1093

Dow, M., Newman, M.-A. and von Roepenack, E. 2000. The induction and modulation of plant defense responses by bacterial lipopolysaccharides. Annu. Rev. Phytopathol. **38**: 241-261

Edwards, A. R., Van Den Bussche, R. A., Wickman, H. A. and Orser, C. S. 1994. Unusual pattern of bacterial ice nucleation gene evolution. Mol. Biol. Evol. **11**: 911-920

El Hassouni, M., Chambost, J. P., Expert, D., Van Gijsegem, F. and Barras, F. 1999. The minimal gene set member *msrA*, encoding peptide methionine sulfoxide reductase, is a virulence determinant of the plant pathogen *Erwinia chrysanthemi*. Proc. Natl. Acad. Sci. USA **96**: 887-892

Expert 1999. Withholding and exchanging iron: interactions between *Erwinia* spp. and their plant hosts. Annu. Rev. Phytopathol. **37**: 307-334

Fadl, A. A., Sha, J., Klimpel, G. R., Olano, J. P., Niesel, D. W. et al. 2005. Murein lipoprotein is a critical outer membrane component involved in *Salmonella enterica* serovar Typhimurium systemic infection. Infect. Immun. **73**: 1081-1096

Filloux, A. 2004. The underlying mechanisms of type II protein secretion. Biochim. Biophys. Acta **1694**: 163-179

Finlay, B. B. and Falkow, S. 1989. Common themes in microbial pathogenicity. Microbiol. Rev. **53**: 210-230

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F. et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269**: 496-512

Galán, J. E. 2001. *Salmonella* interactions with host cells: Type III secretion at work. Annu. Rev. Cell. Dev. Biol. **17**: 53-86

Geider, K. 2000. Exopolysaccharides of *Erwinia amylovora*: structure, biosynthesis, regulation, role in pathogenicity of amylovoran and levan. In <u>Fire blight: the disease and its causative agent</u>. Ed. Vanneste, J.L. Wallingford, CAB International

Geier, G. and Geider, K. 1993. Characterization and influence on virulence of the levansucrase gene from the fireblight pathogen *Erwinia amylovora*. Physiol. Mol. Plant Pathol. **42**: 387-404

Genomes Online Database, 2010. www.genomesonline.org/

Gerlach, R. G. and Hensel, M. 2007. Protein secretion systems and adhesins: The molecular armory of Gram-negative pathogens. Int. J.l Med. Microbiol. **297**: 401-415

González-Pasayo, R. and Martinez-Romero, E. 2000. Multiresistance Genes of *Rhizobium etli* CFN42. Mol. Plant-Microbe Interact. **13**: 572-577

Grenier, A.-M., Duport, G., Pagès, S., Condemine, G. and Rahbé, Y. 2006. The Phytopathogen *Dickeya dadantii* (*Erwinia chrysantemi* 3937) is a pathogen of the pea aphid. Appl. Environ. Microbiol. **72**: 1956-1965

Ham, J. H., Majerczak, D., Ewert, S., Sreerekha, M.-S., Mackey, D. et al. 2008. WtsE, an AvrE-family type III effector protein of *Pantoea stewartii* subsp. *stewartii*, causes cell death in non-host plants. Mol. Plant Pathol. **9**: 633-643

Hanyu, M., Fujimoto, H., Tejima, K. and Saeki, K. 2009. Functional differences of two distinct catalases in *Mesorhizobium loti* MAFF303099 under free-living and symbiotic conditions. J. Bacteriol. **191**: 1463-1471

Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A. and Birch, R. G. 2007. The phytotoxin Albicidin is a novel inhibitor of DNA gyrase. Antimicrob. Agents Chemother. **51**: 181-187

Heath, M. C. 2001. Pathogenicity factors and resistance mechanisms. Physiol. Mol. Plant Pathol. **58**: 147-148

Henderson, I. R., Navarro-garcia, F., Desvaux, M., Fernandez, R. C. and Ala'Aldeen, D. 2004. Type V protein secretion pathway: the autotransporter story. Microbiol. Mol. Biol. Rev. **68**: 692-744

Hettwer, U., Gross, M. and Rudolph, K. 1995. Purification and characterization of an extracellular levansucrase from *Pseudomonas syringae* pv. *phaseolicola*. J. Bacteriol. **177**: 2834-2839

Hildebrand, M., Aldridge, P. and Geider, K. 2006. Characterization of *hns* genes from *Erwinia amylovora*. Mol. Genet. Genom. **275**: 310-319

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A. and Heesemann, J. 2000. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. EMBO J. **19**: 5989-5999

Holeva, M. C., Bell, K. S., Hyman, L. J., Avrora, A. O., Whisson, S. C. et al. 2004. Use of a pooled transposon mutation grid to demonstrate roles in disease development for *Erwinia carotovora* subsp. *atroseptica* putative Type III secreted effector (DspE/A) and helper (HrpN) proteins. Mol. Plant-Microbe Interact. **17**: 943-950

Holland, I. B. 2004. Translocation of bacterial proteins - an overview. Biochim. Biophys. Acta **1694**: 5 - 16

Hossain, M. M., Shibata, S., Aizawa, S.-I. and Tsuyumu, S. 2005. Motility is an important determinant for pathogenesis of *Erwinia carotovora* subsp. *carotovora*. Physiol. Mol. Plant Pathol. **66**: 134-143

Hubber, A., Vergunst, A. C., Sullivan, J. T., Hooykaas, P. J. J. and Ronson, C. W. 2004. Symbiotic phenotypes and translocated effector proteins of the *Mesorhizobium* strain R7A VirB/D4 type IV secretion system. Mol. Microbiol. **54**: 561-574

Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol. Mol. Biol. Rev. **62**: 379-433

Hugouvieux-Cotte-Pattat, N., Shevchik, V. E. and Nasser, W. 2002. PehN, a polygalacturonase homologue with a low hydrolase activity, is coregulated with the other *Erwinia chrysanthemi* polygalacturonases. J. Bacteriol. **184**: 2664-2673

Hurlbert, J. C. and Preston, J. F. 2001. Functional characterization of a novel xylanase from a corn strain of *Erwinia chrysanthemi*. J. Bacteriol. **183**: 2093-2100

Iwama, T., Ito, Y., Aoki, H., Sakamoto, H., Yamagata, S. et al. 2006. Differentia recognition of citrate and a metal-citrate complex by the bacterial chemoreceptor Tcp. J. Biol. Chem. **281**: 17727-17735

Jacob-Dubuisson, F., Locht, C. and Antoine, R. 2001. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. Mol. Microbiol. **40**: 306-313

Jakobek, J. L. and Lindgren, P. B. 1993. Generalized induction of defense responses in bean is not correlated with the induction of the hypersensitive reaction. Plant Cell **5**: 49-56

Josenhans, C. and Suerbaum, S. 2002. The role of motility as a virulence factor in bacteria. Int. J. Med. Microbiol. **291**: 605-614

Judson, N. and Mekalanos, J. J. 2000. Transposon-based approaches to identify essential bacterial genes. Trends. Microbiol. **8**: 521-526

Kajimura, J., Rahman, A. and Rick, P. D. 2005. Assembly of cyclic enterobacterial common antigen in *Escherichia coli* K-12. J. Bacteriol. **187**: 6917-6927

Kim, J. F., Ham, J. H., Bauer, D. W., Collmer, A. and Beer, S. V. 1998. The *hrpC* and *hrpN* operons of *Erwinia chrysanthemi* EC16 are flanked by *plcA* and homologs of

hemolysin/adhesin genes and accompanying activator/transporter genes. Mol. Plant-Microbe Interact. **11**: 563-567

Kim, W.-S. and Geider, K. 2000. Characterization of a viral EPS-depolymerase, a potential tool for control of fire blight. Phytopathology **90**: 1263-1268

Kirov, S. M. 2003. Bacteria that express lateral flagella enable dissection of the multifunctional roles of £agella in pathogenesis. FEMS Microbiol. Lett. **224**: 151-159

Kube, M., Reinhardt, R., Jakovljevic, V., Jock, S. and Geider, K. 2008. The genomic sequence of the fire blight antagonist *Erwinia tasmaniensis* compared with virulence regions of *E. amylovora*. Acta Hort. **793**: 141-144

Kyrpides, N. C. 1999. Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide. Bioinformatics **15**: 773-774

Lacroix, B., Tzfira, T., Vainstein, A. and Citovsky, V. 2006. A case of promiscuity: *Agrobacterium*'s endless hunt for new partners. Trends Genet. **22**: 29-37

Larsen, R. A. Y. A., Myers, P. S., Skare, J. T., Seachord, C. L., Darveau, R. P. et al. 1996. Identification of TonB homologs in the family *Enterobacteriaceae* and evidence for conservation of TonB-dependent energy transduction complexes. J. Bacteriol. **178**: 1363-1373

Laurila, J., Hannukkala, A., Nykyri, J., Pasanen, M., Hélias, V. et al. 2010. Symptoms and yield reduction caused by *Dickeya* spp. strains isolated from potato and river water in Finland. Eur. J. Plant Pathol. **126**: 249-262

Lee, P. A., Tullman-Ercek, D. and Georgiou, G. 2006. The bacterial twin-arginine translocation pathway. Annu. Rev. Microbiol. **60**: 373-395

Lichter, A., Barash, I., Valinsky, L. and Manulis, S. 1995. The genes involved in cytokinin biosynthesis in *Erwinia herbicola* pv. *gypsophilae*: characterization and role in gall formation. J. Bacteriol. **177**: 4457-4465

Lindow, S. E., Arny, D. C. and Upper, C. D. 1982. Bacterial ice nucleation: a factor in frost injury to plants. Plant Physiol. **70**: 1084-1089

Lindow, S. E., Desurmont, C., Elkins, R., McGourty, G., Clark, E. et al. 1998. Occurrence of indole-3-acetic acid–producing bacteria on pear trees and their association with fruit russet. Phytopathology **88**: 1149-1157

Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. et al. 2008. Structure and genetics of *Shigella* O antigens. FEMS Microbiol. Rev. **32**: 627-653

Liu, Y., Cui, Y., Mukherjee, A. and K, C. A. 1998. Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. *carotovora* that controls production of extracellular enzymes and secondary metabolites. Mol. Microbiology **29**: 219-234

Lockhart, D. J. and Winzeler, E. A. 2000. Genomics, gene expression and DNA arrays. Nature **405**: 827-836

López-Solanilla, E., García-Olmedo, F. and Rodríguez-Palenzuela, P. 1998. Inactivation of the *sapA* to *sapF* locus of *Erwinia chrysanthemi* reveals common features in plant and animal bacterial pathogenesis. Plant Cell **10**: 917-924

Loprasert, S., Vattanaviboon, P., Praituan, W., Chamnongpol, S. and Mongkolsuk, S. 1996. Regulation of the oxidative stress protective enzymes, catalase and superoxide dismutase in *Xanthomonas* - a review. Gene **179**: 33-37

Ma, W., Cui, Y., Liu, Y., Dumenyo, K., Mukherjee, A. et al. 2001. Molecular characterization of global regulatory RNA species that control pathogenicity factors in *Erwinia amylovora* and *Erwinia herbicola* pv. *gypsophilae*. J. Bacteriol. **183**: 1870-1880

Manulis, S. and Barash, I. 2003. *Pantoea agglomerans* pvs. *gypsophilae* and *betae*, recently evolved pathogens? Mol. Plant Pathol. **4**: 307-314

Manulis, S., Haviv-Chesner, A., Brandl, M. T., Lindow, S. E. and Barash, I. 1998. Differential involvement of indole-3-acetic acid biosynthetic pathways in pathogenicity and epiphytic fitness of *Erwinia herbicola* pv. *gypsophilae*. Mol. Plant-Microbe Interact. **11**: 634-642

Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. **24**: 133-141

Mattinen, L., Nissinen, R., Riipi, T., Kalkkinen, N. and Pirhonen, M. 2007. Host-extract induced changes in the secretome of the plant pathogenic bacterium *Pectobacterium atrosepticum*. Proteomics **7**: 3527-3537

Melchers, L. S. and Stuiver, M. H. 2000. Novel genes for disease-resistance breeding. Curr. Opin. Plant. Biol. **3**: 147-152

Merino, S., Altarriba, M., Izquierdo, L., Nogueras, M. M., Regué, M. et al. 2000. Cloning and Sequencing of the *Klebsiella pneumoniae* O5 wb gene cluster and its role in pathogenesis. Infect. Immun. **68**: 2435-2440

Metzger, M., Bellemann, P., Bugert, P. and Geider, K. 1994. Genetics of galactose metabolism of *Erwinia amylovora* and its influence on polysaccharide synthesis and virulence of the fire blight pathogen. J. Bacteriol. **176**: 450-459

Mey, A. R., Wyckoff, E. E., Kanukurthy, V., Fisher, C. R. and Payne, S. M. 2005. Iron and Fur regulation in *Vibrio cholerae* and the role of Fur in virulence. Infect. Immun. **73**: 8167-8178

Miethke, M. and Marahiel, M. A. 2007. Siderophore-based iron acquisition and pathogen control. Microbiol. Mol. Biol. Rev. **71**: 413-451

Minogue, T. D., Wehland-von Tebra, M., Bernhard, F. and von Bodman, S. B. 2002. The autoregulatory role of EsaR, a quorum-sensing regulator in *Pantoea stewartii* ssp. *stewartii*: evidence for a repressor function. Mol. Microbiology **44**: 1625-1635

Mohammadi, M. and Geider, K. 2007. Autoinducer-2 of the fire blight pathogen *Erwinia amylovora* and other plant-associated bacteria. FEMS Microbiol. Lett. **266**: 34-41

Mole, B. M., Baltrus, D. A., Dangl, J. L. and Grant, S. R. 2007. Global virulence regulation networks in phytopathogenic bacteria. Trends. Microbiol. **15**

Molina, L., Rezzonico, F., Défago, G. and Duffy, B. 2005. Autoinduction in *Erwinia amylovora*: evidence of an acyl-homoserine lactone signal in the fire blight pathogen. J. Bacteriol. **187**: 3206-3213

Morris, C. E., Sands, D. C., Vinatzer, B. A., Glaux, C., Guilbaud, C. et al. 2008. The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. ISME J. **2**: 321-334

Mougous, J. D., Gifford, C. A., Ramsdell, T. L. and Mekalanos, J. J. 2007. Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*. Nature: Cell Biol. **9**: 797-803

Murillo, J. and Sesma, A. 2001. The biochemistry and molecular genetics of host range definition in *Pseudomonas syringae*. Phytopathol. Mediterr. **40**: 3-26

Nachin, L., El Hassouni, M., Loiseau, L., Expert, D. and Barras, F. 2001. SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. Mol. Microbiology **39**: 960-972

Narita, S.-I., Matsuyama, S.-I. and Tokuda, H. 2004. Lipoprotein trafficking in *Escherichia coli*. Arch. Microbiol. **182**: 1-6

Nasser, W., Faelen, M., Hugouvieux-Cotte-Pattat, N. and Reverchon, S. 2001. Role of the nucleoid-associated protein H-NS in the synthesis of virulence factors in the phytopathogenic bacterium *Erwinia chrysanthemi*. Mol. Plant-Microbe Interact. **14**: 10-20

Newman, M.-A., von Roepenack, E., Daniels, M. and Dow, M. 2000. Lipopolysaccharides and plant responses to phytopathogenic bacteria. Mol. Plant Pathol. **1**: 25-31

Nogales, J., Muños, S., Olivares, J. and Sanjuán, J. 2006. *Sinorhizobium meliloti* genes involved in tolerance to the antimicrobial peptide protamine. FEMS Microbiol. Lett. **264**: 160-167

Nuccio, S.-P. and Baümler, A. J. 2007. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. Microbiol. Mol. Biol. Rev. **71**: 551-575

Oh, C.-S. and Beer, S. V. 2005. Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. FEMS Microbiol. Lett. **253**: 185-192

Ojanen-Reuhs, T., Kalkinnen, N., Westerlund-Wikström, B., van Doorn, J., Haahtela, K. et al. 1997. Characterization of the *fimA* gene encoding bundle-forming fimbriae of the plant pathogen *Xanthomonas campestris* pv. *vesicatoria*. J. Bacteriol. **179**: 1280-1290

Olsén, A., Wick, M. J., Mörgelin, M. and Björk, L. 1998. Curli, fibrous surface proteins of *Escherichia coli*, interact with major histocompatibility complex class I molecules. Infect. Immun. **66**: 944-949

Oomen, C. J., van Ulsen, P., Van Gelder, P., Feijen, M., Tommassen, J. et al. 2004. Structure of the translocator domain of a bacterial autotransporter. EMBO J. **23**: 1257-1266

Pimentel, D., Lach, L., Zuniga, R. and Morrison, D. 2000. Environmental and economic costs of nonindigenous species in the United States. Bioscience **50**: 53-65

Politis, D. J. and Goodman, R. N. 1980. Fine structure of extracellular polysaccharide of *Erwinia amylovora*. Appl. Environ. Microbiol. **40**: 596-607

Postma, P. W., Lengeler, J. W. and Jacobson, G. R. 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. **57**: 543-594

Pratt, L. A. and Kolter, R. 1998. Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. Mol. Microbiol. **30**: 285-293

Preston, G. M., Studholme, D. J. and Caldelari, I. 2005. Profiling the secretomes of plant pathogenic Proteobacteria. FEMS Microbiol. Rev. **29**: 311-360

Proft, T. and Baker, E. N. 2009. Pili in Gram-negative and Gram-positive bacteria – structure, assembly and their role in disease. Cell. Mol. Life Sci. **66**: 613 - 635

Pukatzki, S., Revel, A. T., Sturtevant, D. and Mekalanos, J. J. 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proc. Natl. Acad. Sci. USA **104**: 15508-15513

Pukatzki, S., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W. C. et al. 2006. Identification of a conserved bacterial protein secretion system in Vibrio cholerae

using the Dictyostelium host model system. Proc. Natl. Acad. Sci. USA **103**: 1528-1533

Qian, W., Jia, Y., Ren, S.-X., He, Y.-q., Feng, J.-X. et al. 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. Genome Res. **15**: 757-767

Quirino, B. F., Candido, E. S., Campos, P. F., Franco, O. L. and Krüger, R. H. 2010. Proteomic approaches to study plant–pathogen interactions. Phytochemistry **71**: 351-362

Ratledge, C. and Dover, L. G. 2000. Iron metabolism in pathogenic bacteria. Annu. Rev. Microbiol. **54**: 881-941

Reverchon, S., Rouanet, C., Expert, D. and Nasser, W. 2002. Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. J. Bacteriol. **184**: 654-665

Rigano, L. A., Payette, C., Brouillard, G., Marano, M. R., Abramowicz, L. et al. 2007. Bacterial cyclic β-(1,2)-glucan acts in systemic suppression of plant immune responses. Plant Cell **19**: 2077-2089

Roberts, I. S. 1996. The biochemistry and genetics of capsular polysaccharide production in bacteria. Annu. Rev. Microbiol. **50**: 285-315

Robert-Seilaniantz, A., Navarro, L., Bari, R. and Jones, J. D. G. 2007. Pathological hormone imbalances. Curr. Opin. Plant. Biol. **10**: 372-379

Rojas, C. M., Ham, J. H., Deng, W.-l., Doyle, J. J. and Collmer, A. 2002. HecA, a member of a class of adhesins produced by diverse pathogenic bacteria, contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of *Erwinia chrysanthemi* EC16 on *Nicotiana clevelandii* seedlings. Proc. Natl. Acad. Sci. USA **99**: 13142-13147

Romantschuk, M., Roine, E. and Taira, S. 2001. Hrp pilus – reaching through the plant cell wall. Eur. J. Plant Pathol. **107**: 153-160

Roy, C., Kester, H., Visser, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N. et al. 1999. Modes of action of five different endopectate lyases from *Erwinia chrysanthemi* 3937. J. Bacteriol. **181**: 3705-3709

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S. et al. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature **415**: 497-502

Sanger, F., Nicklen, S. and Coulson, A. R. 1977. DNA sequencing with chain-terminating. Proc. Natl. Acad. Sci. USA **74**: 5463-5467

Sauvonnet, N., Vignon, G., Pugsley, A. P. and Gounon, P. 2000. Pilus formation and protein secretion by the same machinery in *Escherichia coli*. EMBO J. **19**: 2221-2228

Schell, M. A., Ulrich, R. L., Ribot, W. J., Brueggemann, E. E., Hines, H. B. et al. 2007. Type VI secretion is a major virulence determinant in *Burkholderia mallei*. Mol. Microbiology **64**: 1466-1485

Schembri, M. A., Dalsgaard, D. and Klemm, P. 2004. Capsule shields the function of short bacterial adhesins. J. Bacteriol. **186**: 1249-1257

Scheutz, F., Cheasty, T., Woodward, D. and Smith, H. R. 2004. Designation of O174 and O175 to temporary O groups OX3 and OX7, and six new *E. coli* O groups that include verocytotoxin-producing *E. coli* (VTEC):O176, O177, O178, O179, O180 and O181. APMIS **112**: 569-584

Schwartz, T., Bernhard, F., Theiler, R. and Geider, K. 1991. Diversity of the fire blight pathogen in production of dihydrophenylalanine, a virulence factor of some *Erwinia amylovora* strains. Phytopathology **81**: 873-878

Shelby, M. 1998. The importance of home landscapes in the fight against economically important exotic agricultural pests. Proc. Fla. State Hort. Soc. **111**: 209-211

Shrivastava, S. and Mande, S. S. 2008. Identification and functional characterization of gene components of Type VI secretion system in bacterial genomes. PLOS One **3**: 1-11

Simpson, A. J. G., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M. et al. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature **406**: 151-159

Smits, T. H. M., Rezzonico, F., Kamber, T., Blom, J., Goesmann, A. et al. 2010. Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia* spp. Mol. Plant-Microbe Interact. **23**: 384-393

Soutourina, O. A. and Bertin, P. N. 2003. Regulation cascade of £agellar expression in Gram-negative bacteria. FEMS Microbiol. Rev. **27**: 505-523

Szurmant, H. and Ordal, G. W. 2004. Diversity in chemotaxis mechanisms among the Bacteria and Archaea. Microbiol. Mol. Biol. Rev. **68**: 301-319

Taylor, B. L., Zhulin, I. B. and Johnson, M. S. 1999. Aerotaxis and other energy-sensing behavior in bacteria. Annu. Rev. Microbiol. **53**: 103-128

Teixeira-Vanini, M. M., Spisni, A., Sforça, M. L., Pertinhez, T. A. and Benedetti, C. E. 2008. The solution structure of the outer membrane lipoprotein OmlA from *Xanthomonas axonopodis* pv. *citri* reveals a protein fold implicated in protein-protein interaction. Proteins **71**: 2051-2064

Torres-Cabassa, A, Gottesman, S, Frederick, R. D, Dolph, P.J. and Coplin, D. L. 1987. Control of extracellular polysaccharide synthesis in *Erwinia stewartii* and *Escherichia coli* K-12: a common regulatory function. J. Bacteriol. **169**: 4525-4531

Toth, I. K., Bell, K. S., Holeva, M. C. and Birch, P. R. J. 2003. Soft rot erwiniae: from genes to genomes. Mol. Plant Pathol. **4**: 17-30

Toth, I. K. and Birch, P. R. J. 2005. Rotting softly and stealthily. Curr. Opin. Plant. Biol. **8**: 424-429

Toth, I. K., Pritchard, L. and Birch, P. R. J. 2006. Comparative genomics reveals what makes an enterobacterial plant pathogen. Annu. Rev. Phytopathol. **44**: 305-336

Tripathi, L., Mwaka, H., Tripathi, J. N. and Tushemereirwe, W. K. 2010. Expression of sweet pepper Hrap gene in banana enhances resistance to *Xanthomonas campestris* pv. *musacearum*. Mol. Plant Pathol. **11**: 1-11

Uppalapati, S. R., Ishiga, Y., Wangdi, T., Kunkel, B. N., Anand, A. et al. 2007. The phytotoxin Coronatine contributes to pathogen fitness and is required for suppression of salicylic acid accumulation in tomato inoculated with *Pseudomonas syringae* pv. *tomato* DC3000. Mol. Plant-Microbe Interact. **20**: 955-965

Van De Broek, A., Lambrecht, M. and Vanderleyden, J. 1995. Bacterial chemotactic motility is important for the initiation of wheat root colonization by *Azospirillum brasilense*. J. Bacteriol. **144**: 2599-2606

Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E. A., Menck, C. F. M., da Silva, A. C. R. et al. 2002. Comparative genomic analysis of plant-associated bacteria. Annu. Rev. Phytopathol. **40**: 169-189

VanEtten, H., Temporini, E. and Wasmann, C. 2001. Phytoalexin (and phytoanticipin) tolerance as a virulence trait: why is it not required by all pathogens? Physiol. Mol. Plant Pathol. **59**: 83-93

Veenendaal, A. K. J., van Der Does, C. and Driessen, A. J. M. 2004. The protein-conducting channel SecYEG. Biochim. Biophys. Acta **1694**: 81 - 95

Vinatzer, B. A. and Yan, S. 2008. Mining the genomes of plant pathogenic bacteria: how not to drown in gigabases of sequence. Mol. Plant Pathol. **9**: 105-118

von Bodman, S. B., Majerczak, D. R. and Coplin, D. L. 1998. A negative regulator mediates quorum-sensing control of exopolysaccharide production in *Pantoea stewartii* subsp. *stewartii*. Proc. Natl. Acad. Sci. USA **95**: 7687-7692

von Bubnoff, A. 2008. Next-generation sequencing: the race is on. Cell **132**: 721-723

Walsh, C. 2000. Molecular mechanisms that confer antibacterial drug resistance. Nature **406**: 775-781

Wandersman, C. and Delepelaire, P. 2004. Bacterial iron sources: from siderophores to hemophores. Annu. Rev. Microbiol. **58**: 611-647

Whatley, M. H., Hunter, N., Cantrell, M. A., Hendrick, C., Keegstra, K. et al. 1980. Lipopolysaccharide composition of the wilt pathogen, *Pseudomonas solanacearum*. Plant Physiol. **65**: 557-559

Whitehead, N. A., Barnard, A. M. L., Slater, H., Simpson, N. J. L. and Salmond, G. P. C. 2001. Quorum-sensing in gram-negative bacteria. FEMS Microbiol. Rev. **25**: 365-404

Winans, S. C. 1992. Two-way chemical signaling in *Agrobacterium*-plant interactions. Microbiol. Rev. **56**: 12-31

Wood, D. W., Setubal, J. C., Kaul, R., Monks, D. E., Kitajima, J. P. et al. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science **294**: 2317-2323

Wu, H. and Fives-Taylor, P. M. 2001. Molecular strategies for fimbrial expression and assembly. Crit. Rev. Oral Biol. Med. **12**: 101-115

Wu, H.-Y., Chung, P.-C., Shih, H.-W., Wen, S.-R. and Lai, E.-M. 2008. Secretome analysis uncovers an Hcp-family protein secreted via a Type VI secretion system in *Agrobacterium tumefaciens*. J. Bacteriol. **190**: 2841-2850

Yao, J. and Allen, C. 2006. Chemotaxis is required for virulence and competitive fitness of the bacterial witl pathogen *Ralstonia solanacearum*. J. Bacteriol. **188**: 3697-3708

Young, G. M., Schmiel, D. H. and Miller, V. L. 1999. A new pathway for the secretion of virulence factors by bacteria: The flagellar export apparatus functions as a protein-secretion system. Proc. Natl. Acad. Sci. USA **96**: 6456-6461

Zhang, Y., Bak, D. D., Heid, H. and Geider, K. 1999. Molecular characterization of a protease secreted by *Erwinia amylovora*. J. Mol. Biol. **289**: 1239-1251

Zheng, J. and Leung, K. Y. 2007. Dissection of a type VI secretion system in *Edwardsiella tarda*. Mol. Microbiol. **66**: 1192-1206

Zhu, J., Oger, P. M., Schrammeijer, B., Hooykaas, P. J. J., Farrand, S. K. et al. 2000. The bases of crown gall tumorigenesis. J. Bacteriol. **182**: 3885-3895

Chapter 2:

The genome sequence of *Pantoea ananatis* LMG20103, the causative agent of *Eucalyptus* blight and dieback

ABSTRACT

Pantoea ananatis is a broad host range plant pathogen which infects economically important crops such as rice, maize and onion. In South Africa, *P. ananatis* causes blight and dieback of several clones, hybrids and species of the important forestry resource *Eucalyptus*, resulting in devastating losses. In this chapter, the whole genome sequence of a highly virulent *Eucalyptus*-pathogenic *P. ananatis* strain, LMG20103, was sequenced, assembled and annotated. Pertinent genome and protein metrics are discussed. This is the first phytobacterial pathogen genome to be sequenced in Africa and the first member of the genus *Pantoea*, which hosts a number of important plant pathogens, to be completely sequenced and published.

INTRODUCTION

Pantoea ananatis is an emerging plant pathogen that infects a wide range of plant hosts including rice, maize, onion, melons and pineapple (Coutinho and Venter, 2009). In South Africa, it has been implicated in diseases of maize, onion and *Eucalyptus* (Goszczynska et al. 2006; Goszczynska et al. 2007; Coutinho et al. 2002). *Eucalyptus* blight and dieback by *P. ananatis* is of particular concern to the forestry industry as it results in significant losses of seedlings in nurseries. In 2008, blight and dieback led to the loss of 200,000 seedlings in a single nursery (Sean de Haas, Mondi Fountains Nursery, personal communication). Very little is known about the means by which *P. ananatis* infects and causes symptoms on its various plant hosts. Understanding the mechanism of disease could provide solutions to curb crop losses or eradicate the pathogen.

One means of gaining insight into how a pathogen causes disease is by sequencing its genome and mining the genome sequence for candidate genes involved in the pathogen interaction with the plant, the infection process and symptom development (Vinatzer and Yan, 2008). This has been successfully employed to analyse the pathogenesis of a number of important plant pathogens including the tomato pathogen *Pseudomonas syringae*, grapevine-associated *Xylella fastidiosa* and the potato soft rot organism *Pectobacterium atrosepticum* (Vinatzer and Yan, 2008; Buell et al. 2003; Bell et al. 2004). The first genome to be sequenced was that of *Haemophilus influenzae* by means of the classical Sanger sequencing (Fleischmann et al. 1995). Using the Sanger approach to sequence genomes is expensive and time-consuming.

However, novel "next generation" sequencing technologies have recently been developed which are capable of generating genomic data much faster and at a reduced cost. One such technology was developed by 454 Life Sciences (Roche Inc., Switzerland). This sequencer generates large amounts of genome data through an automated system utilising a pyrosequencing approach (Margulies et al. 2005). Briefly, the genome is fragmented through nebulisation and specialised adapters added to each fragment. Fragments are immobilised on beads immersed in a droplet of emulsion and clonally amplified. Following amplification, fragment-coated beads are loaded onto a fibre optic slide and sequence data is generated by detecting light emitted through the chemiluminescent cleavage of an inorganic pyrophosphate when a particular nucleotide is incorporated in the target sequence (Margulies et al. 2005). Sequencing with this technology yields data of similar accuracy to traditional Sanger sequencing, but at $1/10^{th}$ to $1/100^{th}$ of the sequencing cost, significantly reduced sequencing time and without the cloning bias associated with Sanger sequencing (Rothberg and Leamon, 2008). Following sequencing, the whole genome must be assembled and annotated. Annotation is the process by which biological information is attached to the genome sequence (Stein, 2001). The first step of annotation involves the identification all the genes in the genome sequence. Subsequently, these genes can be compared to those for which structural and functional data is available and thereby each gene is annotated. Given the benefits of genome sequences as inexhaustible information resources and the development of novel cost- and time-effective technologies for genome sequencing, this approach to studying the biology, epidemiology and pathology is being utilised increasingly to study mammalian, invertebrate and plant pathogens and has become available to the developing world.

In this chapter the sequencing, by means of 454 pyrosequencing, assembly and annotation of the genome of the *Eucalyptus* pathogen *Pantoea ananatis* LMG20103 are described. Several genome metrics were determined. All the genes on the *P. ananatis* genome were identified and the functions and subcellular localisation of the encoded proteins are discussed. The presence of mobile elements such as plasmids and integrated phage elements were determined. The annotated genome sequence was submitted to the National Center for Biotechnology Information genome database (Genome project: 43085; Genome Accession: CP001875).

MATERIALS AND METHODS

Strain Selection

Pantoea ananatis LMG20103 was selected for sequencing. This strain was isolated from a diseased *Eucalyptus grandis* x *nitens* hybrid in a plantation in Piet Retief, South Africa (Coutinho et al. 2002). It is stored in the Forestry and Agricultural Biotechnology Institute Bacterial Culture Collection under the designation BCC0127 and is maintained at the Laboratorium voor Microbiologie Gent (LMG) culture collection as LMG20103. This strain was selected for sequencing as it was the most virulent in pathogenicity trials on *Eucalyptus grandis* x *nitens* hybrid clones.

DNA extraction

Sequencing by 454 pyrosequencing technology requires 10 μg of purified DNA. DNA was extracted from a single colony of LMG20103 using the Qiagen tissue extraction kit (Qiagen, USA). This yielded 100ng/μl as determined by Nanodrop (Thermo Scientific, USA). Six extracts were pooled and precipitated with 1:100 3M Sodium Acetate and centrifuged at 13,000 rpm for 30 minutes. The supernatant was aspirated and the pellet washed twice with 250 μl 100% ethanol. After aspiration the pellet was airdried and resuspended in 500 μl Tris-EDTA buffer (50 mM TE). The 16S rDNA and *gyrB* sequences were determined from the genomic DNA for strain and purity confirmation.

Genome Sequencing

LMG20103 DNA was submitted to Inqaba Biotec.[TM] (South Africa). Five runs using the Roche GS20 sequencer (Roche, Switzerland) were performed with four and a half plates filled with nebulised LMG20103 DNA. The GS20 sequencer produces reads 80 to 120 nucleotides in size with an expected yield of 20 Megabases (Mb) per plate from 200,000 reads. Thus 4 ½ runs would yield a total of 90 Mb of sequence, which would provide 18x sequencing depth to a genome of 5 Mb in size.

2.4 Genome Assembly

Initial assembly of the 80-120 nt reads was performed at Inqaba Biotec.[TM] using the Roche Newbler Assembler linked to the GS20 sequencer. This yielded 356 contigs ranging in size from 81 to 317,166 nucleotides. Subsequently two different draft assemblies were produced. The first generation assembly was based on the contiguous nature of contigs larger than 500 nucleotides as determined by a scaffolding approach.

59
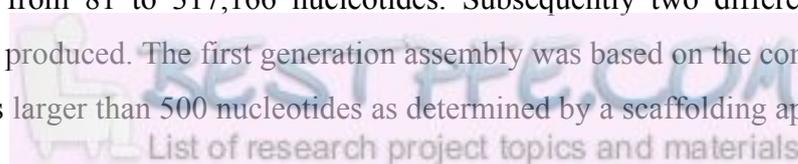
274 contigs smaller than 500 nt were excluded but analysed by performing BlastN analysis against the NCBI nucleotide database. The second generation assembly made use of re-assembly, including <500 nt contigs, as well as a scaffolding approach and gap closure by PCR.

First generation draft assembly

The first generation assembly was performed with 83 contigs larger than 500 nt in size. BLASTN analysis of the last 1000 nucleotides on either end of each contig against the NCBI nucleotide database was performed. The NCBI ORF-finder application (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used to detect open reading frames within these last 1000 nucleotides and the amino acid sequences were subjected to BLASTP against the NCBI database. The percentage nucleotide and amino acid identity between the contig ends and homologues in the complete genome sequences of ten closely related *Enterobacteriaceae* strains was determined. Strains used were *Salmonella enterica* subsp. *enterica* serovar Paratyphi ATCC 9150, *Pectobacterium atrosepticum* SCRI1043, *Escherichia coli* K-12, *E. coli* O157:H7 str. Sakai, *Serratia proteamaculans* 568, *Enterobacter sakazakii* BAA-894, *Enterobacter* sp. 638, *Yersinia pestis* KIM, *Citrobacter koseri* ATCC BAA-895 and *Klebsiella pneumoniae* MGH 78578. Local BLASTN was also performed against contigs available for *Pantoea stewartii* subsp. *stewartii* DC283. When two sequence ends showed homology to contiguous sequence in five or more of the bacterial strains, the contigs were considered as contiguous in the LMG20103 genome and assembled. On the basis of the location where homology ended, an appropriate number of Ns were inserted to signify gaps. In the case of overlap the coinciding end was deleted from one contig to ensure accurate fit of the contigs.

Second Generation Draft Assembly

Newbler Assembler version 2.00.00 (Roche Inc., Switzerland) was used to reassemble contigs, including those smaller than 500 nt, yielding 117 contigs. These were assembled using a scaffolding approach as above. PCR gap closure was performed. Among the small contigs, repeat regions constituting fragments of the 16S, 23S and 5S rDNA genes and the ITS were found and by means of PCR and assembly of the ribosomal DNA fragments six complete copies of the 16S-ITS-23S-5S rRNA operons were assembled and these were incorporated in the second generation assembly.

<u>Prediction of protein coding sequences and protein annotation</u>

Open Reading frames for protein coding sequences (CDS) were predicted using a combination of gene prediction algorithms and systems. Initial genes were predicted by Glimmer v2.1.3 by the BASYS annotation system (Van Domselaar et al. 2005). This method predicts a significant number of false positives, especially in regions where GC content is greater than 60%. A newer version of Glimmer (v3.0.2) (http://www.cbcb.umd.edu/software/glimmer; Delcher et al. 1999; Salzberg et al. 1998), FgenesB (www.softberry.com) and AMIgene were also utilised for CDS prediction. AMIGene is employed in the MaGe annotation system and is reported to improve prediction of small and atypical genes (Bocs et al. 2003).

Amino acid sequences for the open reading frames predicted by the abovementioned methods were compared by local BlastP analysis. Those predicted by three or four of the prediction algorithms were maintained. Open reading frames predicted by only one or two of the methods were analysed by BlastP against the NCBI Protein database and were added to the total open reading frame set if they produced significant hits. The predicted protein coding genes were manually validated by BlastP of the CDS amino acid sequence against the NCBI protein database. Each of the encoded proteins was given a unique numerical locus tag with the prefix PANA_ and subsequently annotated. Initial annotation of CDSs was performed using the BASys automated annotation system (Van Domselaar et al. 2005). This server makes use of over 70 bioinformatic tools to generate a report for each CDS detailing the gene and protein names, nucleotide and amino acid sequences and orientation of genes, synonyms, G+C content, COG function, specific function, metabolic importance, gene ontology, homologues and paralogues, the presence of signal peptides and transmembrane helices and subcellular localisations. The MaGe annotation system provided similar information for each of the CDSs. Both automated annotation systems make use of local nucleotide and protein databases to annotate the CDSs. Annotation of all LMG20103 CDSs was validated by BlastN and BlastP analyses against the nucleotide and protein databases on the NCBI web-server (Http://www.ncbi.nlm.nih.gov).

61

Genome sequence metrics

The genome size was determined from the more complete second generation draft assembly. The G+C content was calculated from the sum of guanine and cytosine residues divided by the overall nucleotide content of the assembly. Coding intensity for the protein coding sequences (CDS) was determined by division of the sum of the lengths of each individual protein coding gene by total genome length. Operons were predicted using the FGenesB software package (www.softberry.com) which determines operons on the basis of the distances between the predicted genes, conservation of neighbouring genes in operons in other genomes and identification of putative promoters and terminators. Predicted operons were compared to the composite set of predicted genes. Operons and transcriptional units for those genes predicted by FGenesB which were not kept in the composite dataset were eliminated. The origin of replication was determined by BlastN analysis of the *E. coli* K-12 *oriC* nucleotide sequence (X02820) against the LMG20103 genome sequence.

RNA Analysis

Some genes encode functional RNAs involved in ribosome synthesis (rRNA), transfer of amino acids during translation (tRNAs) and small non-coding RNAs (ncRNAs) which play various roles, including regulation of mRNA binding and maturation of tRNAs (Griffiths-Jones et al. 2005). The locations of the genes encoding the 16S, 23S and 5S rRNA whose sequences were determined by PCR and Sanger sequencing, were validated by BLASTN analysis of the *E. coli* K-12 16S, 23S and 5S rDNA sequences against the LMG20103 chromosome. The rRNA genes were given the locus tags PANA_r#. Transfer RNAs (tRNAs) were predicted using ARAGORN v1.2 (Laslett and Canback, 2004; Http://130.235.46.10/ARAGORN). The locations were mapped on the LMG20103 chromosome and the proportions of different tRNA species determined. Locus_tags PANA_t# were assigned. Non-coding RNAs (ncRNAs) were detected using Rfam (Griffiths-Jones et al. 2005; http://rfam.janelia.org) and classified as misc_RNA#.

Protein metrics

Functional classification of proteins

The conserved orthologous groups (COG) determined by BASYS and MaGe for each of the predicted LMG20103 CDSs were validated by BLASTN analysis against the

NCBI and COG databases. BLAST links to the Interpro (Hunter et al. 2009; Http://www.ebi.ac.uk/interpro), Conserved Domain Database (CCD; Marchler-Bauer et al. 2005; Http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), Protein Cluster (PRK; Klimke et al. 2009; Http://www.ncbi.nlm.nih.gov/proteinclusters), The Institute For Genomic Research FAMily (TIGRFAM; Haft et al. 2003; Http://www.jcvi.org/cms/research/projects/tigrfams/) and Gene Ontology (GO) (Harris et al. 2004; Http://www.geneontology.org)) databases were followed to determine the COG to which each protein belongs. Those proteins for which no COG could be determined in this manner, were assigned to the COG "Unknown function".

Protein Subcellular Localisation

Bacterial proteins are initially localised in the cytoplasm. A proportion of these proteins are then translocated across the cytoplasmic membrane via secretory pathways to become embedded in the cytoplasmic and outer membrane, localised in the periplasm or transported across the outer membrane to the extracellular medium (Gardy et al. 2003). Determination of the subcellular localization of bacterial proteins can provide information about their potential functions. Several software packages and web-servers were used to predict where each of the LMG20103-encoded proteins are localised. Proteins translocated across the inner membrane often carry an N-terminal signal peptide directing proteins to the secretory machinery. Signal peptides are subsequently cleaved by signal peptidase I (SPase I). SignalP 3.0 (Bendtsen et al. 2004; Http://www.cbs.dtu.dk/services/SignalP) and LipoP 1.0 (Juncker et al. 2003; Http://www.cbs.dtu.dk/services/LipoP) were utilised to detect the presence of signal peptides and signal peptidase cleavage sites. Some proteins are translocated from the cytoplasm via the specialised Twin-arginine translocation (Tat) pathway and can be distinguished by the presence of a specific twin-arginine motif sequence. Tat translocated proteins were analysed using the TatP server (Bendtsen et al. 2005; Http://www.cbs.dtu.dk/services/TatP). Membrane-associated proteins are character-ised by the presence of a transmembrane domain. The occurrence of such a domain in LMG20103 proteins was determined using TMHMM 2.0 (Krogh et al. 2001; http://www.cbs.dtu.dk/services/TMHMM). Proteins with a transmembrane domain but lacking a signal peptide or SPaseI cleave site were classified as inner membrane proteins located on the cytoplasmic side. Proteins with a signal peptide but lacking a SPaseI site were classed as inner membrane proteins exposed at the periplasmic

surface, while those with a transmembrane domain, a signal peptide and SPaseI cleavage site were considered as periplasm-exposed outer membrane proteins. Lipoproteins are processed by signal peptidase II (SPaseII) and the presence of a SPaseII binding site was predicted with LipoP (Juncker et al. 2003; Http://www.cbs.dtu.dk/services/LipoP). LMG20103 proteins with a SPAseII cleavage site and an aspartic acid residue two amino acids upstream of the cleavage site were considered to be lipoproteins localised to the inner membrane while all other lipoproteins may be linked to the outer membrane (Seydel et al. 1999).

The results from these analyses were collated, subdividing proteins into membrane-associated, cytoplasmic, periplasmic and secreted groups. This was validated using PSORTb 3.0 (Http://www.psort.org/psortb/index.html) which takes into consideration presence of signal peptides, transmembrane helices and location-specific motifs, amino acid composition and comparison to homologous proteins for which the location has been experimentally determined (Yu et al. 2010). Further validation was done using the EchoLOCATION tool to analyse those LMG20103 protein sequences for which homologues exist in the model organism *E. coli* K-12 (Misra et al. 2005; Http://www.york.ac.uk/res/thomas/echolocadv.cfm). Assignment of proteins to subcellular locations was based on the cumulative scores of SignalP, TatP, LipoP, Psort and EchoLOCATE. Proteins which span more than one cellular compartment were assigned to their final location based on the region with the highest score.

Analysis of extrachromosomal and phage elements

Plasmid Analysis

BlastP and BlastN analyses with the amino acid and nucleotide sequences of plasmid associated proteins from phylogenetic relatives against the *P. ananatis* LMG20103 genome were performed. Proteins annotated as putative conjugative and non-conjugative plasmid-associated proteins were identified from the total protein set of LMG20103.

Analysis of Phage elements

Integrated phage elements were initially identified using the ACLAME (A CLAssification of Mobile genetic Elements) and Prophinder (Lima-Mendez et al. 2008; Http://aclame.ulb.ac.be/Tools/Prophinder) databases. Subsequently, BlastP of

proteins adjacent to those areas identified as part of phage elements by Prophinder was performed to elucidate the full phage elements. The locations of integrated phages were determined and phage elements compared to the most closely related phages integrated in the genomes of phylogenetic relatives.

Genome submission

The genome sequence and its annotations were converted into Genbank format using Sequin 9.50 (National Center for Biotechnology Information, USA). Annotations added in the Genbank file included all the protein coding sequences, rRNA, tRNA, misc_RNA sequences, protein translations, protein subcellular localisations and COG classifications. The Genbank sequence was submitted to the NCBI server with the genome accession CP001875 and under the genome project 40385.

RESULTS

Genome sequencing

A final concentration of 86.92 μg DNA (174 ng/ μl) was prepared. Purity was confirmed by comparing the OD_{260} to OD_{280} ratios with a high purity value of 1.93 for the DNA sample. This DNA was submitted to Inqaba biotec. [TM] for sequencing. A total of 991,246 reads amounting to 97,152,414 (~97 Mb) nucleotides were sequenced, with a yield per plate of 21.6 Megabases in 220,256 reads. The average read length was 98 nucleotides. The amount of sequence obtained thus exceeded the expected yield of 90 Mb. Given the final genome size of 4,690,298 nt, four and a half runs of 454 sequencing thus gave a genome coverage depth of ~20.7x.

Genome Assembly

Blast analysis of <500 nt contigs

274 contigs smaller than 500 nucleotides were analysed by BLASTN analysis against the NCBI nucleotide database to determine if any should be included in assembly. 238 contigs (86.8%) showed significant homology (88.24% average nucleotide identity) to nucleotide stretches in the large contigs (Table 2.1). These mostly constituted repeat elements which were excluded from the first generation draft assembly, but were included in the second. Five contigs showed homology to sequences in the human genome and may be due to contamination during DNA or sequencing preparation. This accounts for 0.0008% of total DNA sequenced. Sixteen <500 nt contigs showed

65

>98% average nucleotide identity to 5S, 16S and 23S ribosomal DNA (rDNA) as well as internal transcribed spacer (ITS) regions. These could not be integrated into contigs for the first generation assembly but after PCR amplification and Sanger sequencing six copies of the rRNA operon were included in the second generation assembly.

First generation draft assembly

The 83 contigs were assembled into nine large scaffolds, ranging in size from ~7 kb to ~1.52 Mb (Fig. 2.1). BLASTN analysis of 1,000 nucleotides at the proximal and distal ends of the scaffolds showed no homology to contiguous coding regions in the genomes of other organisms, nor did ORFfinder recognise open reading frames within these regions. Thus further assembly by this approach proved impossible and the scaffolds were assembled end to end, from scaffold one to nine to form a single stretch of DNA which was then submitted for prediction of open reading frames and annotation.

Second generation draft assembly

In silico re-assembly, PCR gap closure and scaffolding yielded fourteen scaffolds which were further assembled into three super-contigs using scaffolding with the genome sequences discussed in the materials and methods section (Fig. 2.2). The three super-scaffolds were assembled end to end for the final draft assembly. All further analyses were performed using this more accurate second generation genome assembly.

Prediction of protein coding sequences and protein annotation

The four gene prediction methods employed yielded varying numbers of predicted protein coding genes (Table 2.2). BLASTP comparison of the subset detected by each method showed that a total of 3,614 protein coding sequences were predicted by all four programs. A combined 735 ORFs were predicted uniquely by one of the four algorithms. Following BLASTP analysis, 99 of these were retained in the total set of predicted open reading frames. Another 634 ORFs were predicted by either two or three of the programs utilised and 568 of these were retained. BLASTP analysis revealed that 274 coding sequences were split by some of the methods. These did not contain a recognisable stop codon and were thus incorporated into single reading frames. A total of 4,237 protein coding sequences were identified in the LMG20103 genome. This is similar to *Enterobacter* sp. 638, which encodes 4,240 protein coding

sequences on its 4.69 Megabase genome. A total of 2,085 (49.2%) CDSs are encoded on the forward strand while 2,152 (50.8%) are encoded on the complementary (reverse) strand (Fig. 2.3). The encoded proteins range in size from 27 aa (PANA_2004 – Hypothetical protein) to 4180 aa (PANA_2607 - large repetitive protein) with an average protein size of 324 amino acids.

Genome metrics

Genome size

The second generation assembly yielded a genome 4,690,298 nucleotides in size. A genome size of 4.5-5 Mb is typical among *Enterobacteriaceae* (Table 2.3). The larger genome of the closest phylogenetic relative *P. stewartii* subsp. *stewartii* DC283 (5 Mb) may be due to 25% of its predicted genome content being associated with 10-13 plasmids (Coplin et al. 1981). Similarly the partial assembly of the genome of the leaf cutter ant-associated *Pantoea* sp. At9b is 6,26 Mb in size.

G+C content

The G+C content of the genome is 53.69%, which is typical among the *Enterobacteriaceae* whose G+C base compositions range between 38-60% (Brenner and Farmer, 2006; Table 2.3). The value for LMG20103 is below the 55.1-60.6% range predicted by traditional methods in the original description paper for the genus (Gavini et al. 1989) but falls within the range (53.1-55.3%) determined by the thermal denaturation method for ten strains of *P. ananatis* by Mergaert et al. (1993) and is identical to that predicted for the draft genome sequence of *P. stewartii* subsp. *stewartii* DC283.

Average gene size and coding intensity

The average gene size in LMG20103 was calculated as 973 nt (Table 2.3). This is larger than the average gene size for all bacteria as determined by Xu et al. (2006). An even larger average gene size, 988 nt, was calculated, however, for *Dickeya dadantii* Ech568. A high coding intensity was also observed in the LMG20103 genome, 88%, which is only superseded by *Citrobacter koseri* ATCC BAA-895 (Table 2.3).

Operon prediction

Operon prediction with the 4,300 gene set predicted by FGenesB yielded 2,473 (57.5%) genes occurring as single-gene transcriptional units, while 1,827 (42.5%)

genes form part of 854 polycistronic operons. When the corrected subset of 4,237 predicted genes was analysed, 2,477 (58.5%) genes are present in the form of transcriptional units, while 1,760 (41.5%) are clustered in 817 operons. This is comparable to *Escherichia coli* K-12, where 2,508 (56%) of 4,493 predicted genes occur as transcriptional units, while 1,670 (44%) form part of operons. The largest operon is 14.6 kb in size and encompasses 18 genes (PANA_0548 to PANA_0565) that encode proteins involved in lipid transport and metabolism.

Location of the origin of replication

BlastN analysis with the *oriC* of *E. coli* K-12 showed the origin of replication on the LMG20103 chromosome is located at nucleotides 18,517-18,807 (Fig. 2.4). This 290 nt region shows highest similarity to *Kluyvera ascorbata* DSM4611 *oriC* (DQ227489.1: 76% identity in 277 nt). As is the case in most *Enterobacteriaceae* the *oriC* region is flanked by *mioC* encoding a flavodoxin involved in modulating replication initiation and *gidA* which encodes a tRNA modifying enzyme.

RNA Analysis

Ribosomal RNAs (rRNAs)

Six copies of the 16S-23S-5S rDNA operon are interspersed on the chromosome of LMG20103 (Fig. 2.5). Similarly, between five and seven copies are found in most members of the family *Enterobacteriaceae*. The genes encoding the 16S, 23S and 5S rRNAs are typical in length, while the internal transcribed spacer between the 16S and 23S rDNA vary in size from 379 to 613 nt (Fig. 2.6). Two copies of the 5S rDNA are associated with rRNA operon 5. A duplicated 5S rDNA is found in some rRNA operons on the genomes of many bacteria.

Transfer RNAs (tRNAs)

A total of 67 tRNAs were predicted by ARAGORN v1.2 and are interspersed across the entire genome sequence (Fig 2.7). They encompass tRNAs required for the translocation of all twenty canonical codons during protein synthesis (Fig. 2.8). One copy each of Asparagine-, Cysteine-, Histidine-, Isoleucine-, Phenylalanine- and Tryptophan transfer RNA are encoded on the genome, while eight Leucine-transfer RNA-coding genes (12% of total) are present. This correlates with the high Leucine content (11.08%) in the amino acid sequences of LMG20103 proteins. Fewer tRNAs

are encoded on the chromosome of LMG20103 than on those of *E. tasmaniensis Et*1/99 (81), *P. atrosepticum* SCRI1043 (76) and *E. coli* O157:H7 str. Sakai (105). However, ARAGORN analysis of the incomplete genome sequence of *P. stewartii* subsp. *stewartii* DC283 predicted only 51 transfer RNAs.

Non-coding RNAs

Rfam analysis revealed a further 30 ncRNAs which are not involved in amino acid transfer or ribosome biosynthesis. Six of these, PANA_ncRNA6, -7, 14-16 and -29, encode riboswitches, which form part of mRNA molecules and can bind target molecules (Tucker and Breaker, 2005). These were annotated as misc_binding in the final genome sequence. Three ncRNAs (PANA_ncRNA9, -24 and -28) represent un-translated 5' leader sequences and were annotated as misc_features. PANA_ncRNA11, 26 and 27 are cis-acting elements and were also re-annotated as misc_features. The remaining eighteen ncRNAs show sequence homology to ncRNAs which play a variety of regulatory roles in other microorganisms. Of interest are the PANA_ncRNA22 and -23 which show significant homology to the *E. coli ryhB* (AF480876.1) RNA. *ryhB* is repressed by high levels of iron and down-regulates the expression of iron storage proteins including ferretins and superoxide dismutases (Wandersman and Delepelaire, 2004).

Protein metrics

Functional classification of LMG20103 proteins

The proteins encoded by LMG20103 were classified according to their Conserved Orthologous Groups (COGs). The distribution of proteins among COGs is shown in Table 2.4. A high proportion of proteins are involved in carbohydrate (10.31%) and amino acid (9.28%) transport and metabolism, a typical feature among bacteria. No known function could be assigned to 18.48% of LMG20103 proteins, significantly lower than the 26.30% of proteins of unknown function observed among all bacteria for which COG predictions were made (www.ncbi.nlm.nih.gov). This may be due to fewer proteins of unknown function being present but likely results from the more exhaustive approach applied for COG classification for LMG20103. However, in *Serratia proteamaculans* 568 a total of 18.24% of encoded proteins are of unknown function. A general function prediction was made for 8.24% of the proteins which contain functional domains but whose function has not been experimentally

69
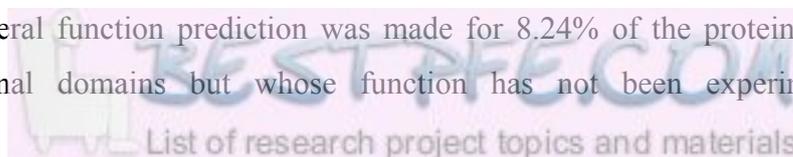
elucidated. The COG proportions in LMG20103 were compared to those in fourteen closely related *Enterobacteriaceae* (Table 2.5). A higher proportion of LMG20103 proteins are involved in carbohydrate and lipid transport and metabolism. A larger proportion of proteins involved in cell wall and envelope biogenesis and defense mechanisms are also present. By contrast, far fewer *P. ananatis* proteins are involved in secondary metabolite biosynthesis and in intracellular trafficking and secretion. The significance of these results is discussed in greater detail in Chapter 3.

Subcellular localisations of LMG20103 proteins

Analysis of signal peptides, transmembrane-helices and secretion motifs allowed prediction of the putative localisation of all the LMG20103 proteins to the different compartments of the bacterial cell. The majority (60.49%) of proteins are located in the cytoplasm. A further 888 proteins (20.96%) contain trans-membrane domains but lack signal peptides and are thus likely localised in the inner membrane on the cytoplasmic side. The other 786 proteins (18.55%) are translocated across the cytoplasmic membrane and become embedded in the inner or outer membrane or are localised in the periplasm or secreted to the extracellular medium (Fig. 2.9 and Table 2.6). Twenty seven of the translocated proteins carry a putative Tat signal peptide, but only 22 of these contain a Tat motif. The remaining 764 proteins are likely secreted by other means, including the Sec pathway. LipoP analysis predicted 117 lipoproteins (2.76% of total proteins) on the basis of signal peptidase activity.

Proteins show a similar distribution to those in *Escherichia coli* K-12 (Table 2.6). A larger subset of proteins is located in the *E. coli* periplasm, while more *P. ananatis* proteins are associated with the inner membrane. A significantly larger number of proteins are secreted to the extracellular environment by LMG20103. Proteins with an extracellular localisation include a polygalacturonase, minor cellulose and extracellular protease, as well as an autotransporter lipase, an adhesin and a filamentous haemagglutinin which may represent putative pathogenicity factors (Chapter 4). A proportion of the extracellular proteins (15 of 55) also contain transmembrane helices, suggesting they may remain attached to the outer membrane.

Analysis of extrachromosomal and phage elements

Analysis of plasmids

BlastN and BlastP analyses against the *P. ananatis* LMG20103 genome were performed with the *P. stewartii* and *P. agglomerans tra* genes and their encoded proteins associated with conjugative plasmids. The PANA_0260 protein shows 95% amino acid identity to *P. stewartii* ACV-0284133, a homologue of outer membrane protein TraF involved in conjugation pilus synthesis (Arutyunov et al. 2010). However, no further conjugal transfer or conjugative plasmid elements could be identified. Screening of the LMG20103 nucleotide sequence and CDS sequences revealed a further five potential proteins associated with non-conjugative plasmids. PANA_2327 and 2328 encode proteins showing 71% and 67% aa identity to PemK (ABD47747) and PemI (AAZ03779) in *Klebsiella pneumoniae,* respectively. These two proteins are involved in stable maintenance of the R100 plasmid. The former stabilises the plasmid by killing the host cells while PemI counteracts PemK (Tsuchimoto et al. 1988). Another CDS, encoded by PANA_4186 shares 64% identity with replication protein A (RepA) in *E. coli* (AAA71884) which forms part of the minimal region required for plasmid replication (Saul et al. 1989). Adjacent to *repA*, PANA_4187 and 4188 encode proteins showing 81% and 57% aa identity, respectively, to *Enterobacter sakazakii* ParA and ParB (ESA_pESAp05553-05554). ParA encodes an ATPase that energises binding of ParB to centromere-like sequences enabling plasmid partitioning. However, orthologues of ParA and ParB have also been identified on bacterial chromsomes and may play a role in effective chromosome partitioning and replication (Gerdes et al. 2000). LMG20103 may thus contain the minimum complement of genes for partitioning, replication and stable maintenance of a non-conjugative plasmid. However, given the limited number of plasmid-associated proteins LMG20103 does not seem to encode a plasmid.

Identification of phage elements in P. ananatis LMG20103

Three phage elements were identified in the genome sequence of LMG20103 using Prophinder. BlastP of CDSs surrounding the predicted prophage elements elucidated the full extent of the integrated phages and identified a further phage element in the genome. The four phages mapped to distinct locations on the chromosome (Fig. 2.10). BlastP of the CDSs contained within the phage elements indicated that phage 2 and 3

belong to the P2 family of temperate lytic phages and showed significant homology to phage Fels-2 found in the genome of *Salmonella enterica* serovar Typhimurium LT2 (Nilsson and Haggård-Ljungquist; 2007). Phage 2 carries 48 protein coding genes, with 28 of these (58%) sharing homologues to the 46 genes carried by the Fels-2 phage (Fig. 2.11 and Table 2.7). Phage 3 carries 18 protein coding genes, with sixteen of these sharing homologues to genes in Phage 2 and the Fels-2 phage. The absence of genes encoding capsid proteins and phage regulatory elements suggest phage element 3 may not encode a full phage complement. Phage elements 1 and 4 belong to the P4 phage family and show highest homology to the *E. fergusonii* ATCC35469 P4 phage. Phage element 1 carries nine protein coding genes with six (66%) sharing homologues in the *E. fergusonii* P4 phage which encodes twelve phage proteins (Fig. 2.12 and Table 2.8). Phage 4 encodes eleven proteins with nine sharing homologues in *E. fergusonii* P4. Phage 1 and 4 lack most phage structural genes and P4 phage family members have been shown to scavenge P2 phage proteins for structural components, DNA packaging and for host cell lysis (Liu et al. 1997).

DISCUSSION

The genome of the *Eucalyptus* blight and dieback causative agent, *Pantoea ananatis* LMG20103, was sequenced using Roche 454 GS20 sequencing technology and subsequently assembled and annotated. This constitutes the first genome of a plant pathogenic bacterium to be sequenced in Africa. Draft sequences of two *Pantoea* species, namely the pathogen responsible for Stewart wilt of maize, *Pantoea stewartii* subsp. *stewartii* DC283 (Http://asap.ahabs.wisc.edu/research-projects/plant-pathogen-genome-projects/pantoea-stewartii-stewartii-dc283-genome-project.html) and the insect-associated *Pantoea* sp. At9b (NCBI Accession number NZ_ACYJ00000000) are available, but their assembly and publication is still pending. Thus the *Pantoea ananatis* LMG20103 genome is also the first in the genus *Pantoea* to be assembled, annotated and published (De Maayer et al. 2010). Four and half sequencing runs with the Roche GS20 sequencer gave a sequencing depth of ~20.7x and allowed near-complete assembly of the genome. Two approaches of assembly were employed with the first yielding nine supercontigs, while the second assembly gave a more accurate assembly, yielding three supercontigs which were assembled end-to-end. With the inherent problems of repeat-sequencing associated with the 454 technology and with assembly of these repeats, full assembly into one chromosome is extremely difficult.

The small gaps between contigs identified during assembly and genome scaffolding suggest that very little genomic data is missing from the LMG20103 sequence.

The LMG20103 genome sequence shows features typical among *Enterobacteriaceae* with respect to genome metrics such as genome size, G+C content, average gene size, coding intensity and number of proteins and structural RNAs. Similarly, a typical number and distribution of rRNAs and tRNAs were observed. Analysis of proteins revealed that a significant proportion of proteins function in the transport and metabolism of amino acids and carbohydrates. This may be linked to the ecological robustness of this species, with strains having been isolated from a wide variety of ecological niches including rivers, soil, aviation fuel tanks, insect guts and a broad range of plant hosts (Coutinho and Venter, 2009; Chapter 3). By contrast, when compared to the genome sequences available for all other *Enterobacteriaceae*, far fewer proteins play a role in intracellular trafficking and secretion. This may be correlated to the absence of several important secretion systems as discussed in Chapter 4. However, more proteins are secreted to the extracellular medium in LMG20103 than *E. coli* K-12, as can be gauged from the subcellular localisations determined for all LMG20103 proteins. This may be due to the pathogenic nature of LMG20103, with a number of the extracellular proteins representing putative pathogenicity factors, while *E. coli* K-12 is a commensal organism.

The presence of mobile genetic elements in the genome sequence of LMG20103 was determined. No plasmids were detected in this strain. This lack of plasmids is unusual among *Pantoea* species as between 10 and 13 plasmids were identified in 31 virulent strains of *P. stewartii* subsp. *stewartii* ranging in size from 2.8 to 210 Megadaltons. Gypsophila and beet-pathogenic *P. agglomerans* pv. *gypsophilae* and pv. *betae* strains also harbour a ~135 kb plasmid which carries a 75 kb pathogenicity island (Barash and Manulis-Sasson, 2009). Four chromosomally integrated phage elements were detected in LMG20103. This is far fewer than in *E. coli* O157:H7 and *S. enterica* CT-18 where 27 and 11 integrated phages elements, respectively, have been detected (http://bicmku.in:8082/prophagedb/index.php). Analysis of the *P. stewartii* DC283 proteins without homologues in *P. ananatis* (Chapter 3) indicates the presence of at least 30 phage elements in the former. The limited number of phage elements in LMG20103 suggests this strain carries a means of resistance to the entry and

integration of lytic phages into the chromosome. The lack of plasmids and limited number of phage elements suggests the genome of LMG20103 may not be as prone to horizontal gene transfer through conjugation and transduction as many of its close phylogenetic relatives.

The complete genome sequence of LMG20103 will enable, by means of genome comparison, mutagenesis and molecular experiments, the study of the biology, ecology and epidemiology of this important emerging plant pathogen. It will provide an information resource for identification of pathogenicity factors and analysis of the role these factors play in *P. ananatis* LMG20103 disease on *Eucalyptus* and other hosts infected with *P. ananatis*. Furthermore, it may provide a greater understanding of the plant and clinical pathogens encompassed in the genus *Pantoea*.

REFERENCES

Arutyunov, D., Arenson, B., Manchak, J. and Frost, L. S. 2010. F-plasmid TraF and TraH are components of an outer membrane complex involved in conjugation. Microbiology **192**(6): 1730-1734

Barash, I. and Manulis-Sasson, S. 2009. Recent evolution of bacterial pathogens: The gall-forming *Pantoea agglomerans* case. Annu. Rev. Phytopathol. **47**: 133-152

Bell, K., Sebaihia, M., Pritchard, L., Holden, M. T. G., Hyman, L. J. et al. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. Proc. Natl. Acad. Sci USA **101**: 11105-11110

Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340**: 783-795

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T. and Brunak, S. 2005. Prediction of twin-arginine signal peptides. BMC Bioinformatics **6**(167): 1-9

Brenner, D.J. and Farmer, J.J. III. 2005. Family I. *Enterobacteriaceae*. In Bergey's manual of systematic bacteriology, 2nd ed. vol. 2. The *Proteobacteria* part B: the *Gammaproteobacteria*. Brenner, D., Krieg. N.R., Staley, J.T. and Garrrity, G.M. Springer, New York, pp 587–607

Bocs, S., Cruveiller, S., Vallenet, D., Nuel, G. and Médigue, C. 2003. AMIGene: Annotation of MIcrobial Genes. Nucleic Acids Res. **31**(13): 3723-3726

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T. et al. 2003. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv tomato DC3000. Proc. Natl. Acad. Sci USA **100**: 10181-10186

Coplin, D. L., Rowan, R. G., Chisholm, D. A. and Whitmoyer, R. E. 1981. Characterization of plasmids in *Erwinia stewartii*. Appl. Environ. Microbiol. **42**(4): 599-604

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

Coutinho, T. A. and Venter, S. N. 2009. *Pantoea ananatis*: an unconventional plant pathogen. Mol. Plant Pathol. **10**: 325-335

De Maayer, P., Chan, W. Y., Venter, S. N., Toth, I. K., Birch, P. R., Joubert, F., et al. 2010. Genome sequence of *Pantoea ananatis* LMG20103, the causative agent of *Eucalyptus* blight and dieback. J. Bacteriol. **192**(11): 2936-2937

Delcher, A. L., Harmon, D., Kasif, S., White, O. and Salzberg, S. L. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27**(23): 4636-4641

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F. et al. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science **269**: 496-512

Gardy, J. L., Spencer, C., Wang, K., Ester, M., Tusnády, G. E. et al. 2003. PSORT-B: improving protein subcellular localization prediction for Gram-negative bacteria. Nucleic Acids Res. **31**(13): 3613-3617

Gavini, F., Mergaert, J., Mielcarek, C., Izard, D., Kersters, K. et al. 1989. Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Erwing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. Int. J. Sys. Bacteriol. **39**(3): 337-345

Gerdes, K. and Jensen, R. B. 2000. Plasmid and chromosome partitioning: surprises from phylogeny. Mol. Microbiol. **37**(3): 455-466

Goszczynska, T., Botha, W. J., Venter, S. N. and Coutinho, T. A. 2007. Isolation and identification of the causal agent of brown stalk rot, a new disease of maize in South Africa. Plant Dis. **91**(6): 711-718

Goszczynska, T., Venter, S. N. and Coutinho, T. A. 2006. PA 20, a semi-selective medium for isolation and enumeration of *Pantoea ananatis*. J. Microbiol. Methods **64**: 225-231

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R. et al. 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. **33**: 121-124

Haft, D. H., Selengut, J. D. and White, O. 2003. The TIGRFAMs database of protein families. Nucleic Acids Res. **31**(1): 371-373

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, R. et al. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. **32**: 258-261

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A. et al. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res. **37**: 211-215

Juncker, A. S., Willenbrock, H., von Heijne, G., Brunak, S., Nielsen, H. et al. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Prot. Sci. **12**: 1652-1662

Klimke, W., Agarwala, R., Badretdin, A., Chetvernin, S., Ciufo, S. et al. 2009. The National Center for Biotechnology Information's Protein Clusters Database. Nucleic Acids Res. **37**: 216-223

Krogh, A., Larsson, B., Heijne, V. and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305**: 567-580

Laslett, D. and Canback, B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32**(1)

Lima-Mendez, G., van Helden, J., Toussaint, A. and Leplae, R. 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. Bioinformatics **24**(6): 863-865

Liu, T. A. O., Renberg, S. K. and Haggård-Ljungquist, E. 1997. Derepression of Prophage P2 by Satellite Phage P4: Cloning of the P4 ε Gene and Identification of Its Product. J. Virol. **71**(6): 4502-4508

Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., Deweese-Scott, C., Geer, L. Y. et al. 2005. CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res.**33**: 192-196

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S. et al. 2005. Genome sequencing in open microfabricated high density picoliter reactors. Nature **437**(7057): 376-380

Mergaert, J., Verdonck, L. and Kersters, K. 1993. Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes*. Int. J. Sys. Bacteriol. **43**(1): 162-173

Misra, R. V., Horler, R. S. P., Reindl, W., Goryanin, I. I. and Thomas, G. H. 2005. EchoBASE: an integrated post-genomic database for *Escherichia coli*. Nucleic Acids Res. **33**: 329-333

Nilsson, A. S. and Haggård-Ljungquist, E. 2007. Evolution of P2-like phages and their impact on bacterial evolution. Res. Microbiol. **158**

Rothberg, J. M. and Leamon, J. H. 2008. The development and impact of 454 sequencing. Nature Biotechnol. **26**: 1117-1125

Salzberg, S. L., Delcher, A. L., Kasif, S. and White, O. 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. **26**(2): 544-548

Saul, D., Spiers, A. J., McAnulty, J., Gibbs, M. G., Bergquist, P. L. et al. 1989. Nucleotide sequence and replication characteristics of RepFIB, a basic replicon of IncF plasmids. J. Bacteriol. **171**(5): 2697-2707

Seydel, A., Gounon, P. and Pugsley, A. P. 1999. Testing the `+2 rule' for lipoprotein sorting in the Escherichia coli cell envelope with a new genetic selection. Mol. Microbiol. **34**(4): 810-821

Stein, L. 2001. Genome annotation: from sequence to biology. Nat. Rev.: Genet. **2**: 493-503

Tsuchimoto, S., Ohtsubo, H. and Ohtsubo, E. 1988. Two genes, *pemK* and *pemI*, responsible for stable maintenance of resistance plasmid R100. J. Bacteriol. **170**(4): 1461-1466

Tucker, B. J. and Breaker, R. R. 2005. Riboswitches as versatile gene control elements. Curr. Opin. Struct. Biol. **15**: 342-348

Van Domselaar, G. H., Stothard, P., Shrivastava, S., Cruz, J. A., Guo, A. et al. 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. **33**: 455-459

Vinatzer, B. A. and Yan, S. 2008. Mining the genomes of plant pathogenic bacteria: how not to drown in gigabases of sequence. Mol. Plant Pathol. **9**(1): 105-118

Wandersman, C. and Delepelaire, P. 2004. Bacterial iron sources: from siderophores to hemophores. Annu. Rev. Microbiol. **58**: 611-647

Xu, L., Chen, H., Hu, X., Zhang, R., Zhang, Z., Luo, Z. W., et al. 2005. Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two Kingdoms. Mol. Biol. Evol. **23**(6), 1107-1108

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Lo, R., Dao, P. et al. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics **26**(13): 1608-1615

FIGURES AND TABLES

Table 2.1: BLAST results of contigs smaller than 500 nt in size.

Number of Contigs	Blast result	Percentage of total 274 contigs
238	High homology to sequences in large contig subset	88%
16	Ribosomal DNA (16S/23S/5S) and ITS	6%
5	No hit	2%
5	Human DNA	2%
5	Enterobacterial DNA	2%

Table 2.2: Open reading frames predicted in the genome sequence of LMG20103 by four different algorithms. ORFs unique to each method are indicated while those retained in the set of LMG20103 ORFs are shown in brackets.

ORF prediction program	Predicted ORFs	ORFs unique to prediction method
Glimmer 2.1.3	5,018	420 (42 retained)
Glimmer 3.0.2	4,792	160 (11 retained)
FGenesB	4,532	73 (16 retained)
AMIGene	4,549	82 (30 retained)

Table 2.3: Genomic attributes of *P. ananatis* LMG20103 and nineteen phylogenetic relatives in the family *Enterobacteriaceae*. A Neighbour-joining phylogeny based on the ClustalW alignment of the Gyrase B (GyrB) amino acid sequences is shown and bootstrap values are indicated (n=1,000). Information for other genomes was obtained from their relative genome project pages (Http://www.ncbi.nlm.nih.gov) and for *P. agglomerans* C9-1 from Http://www.cost783.ch

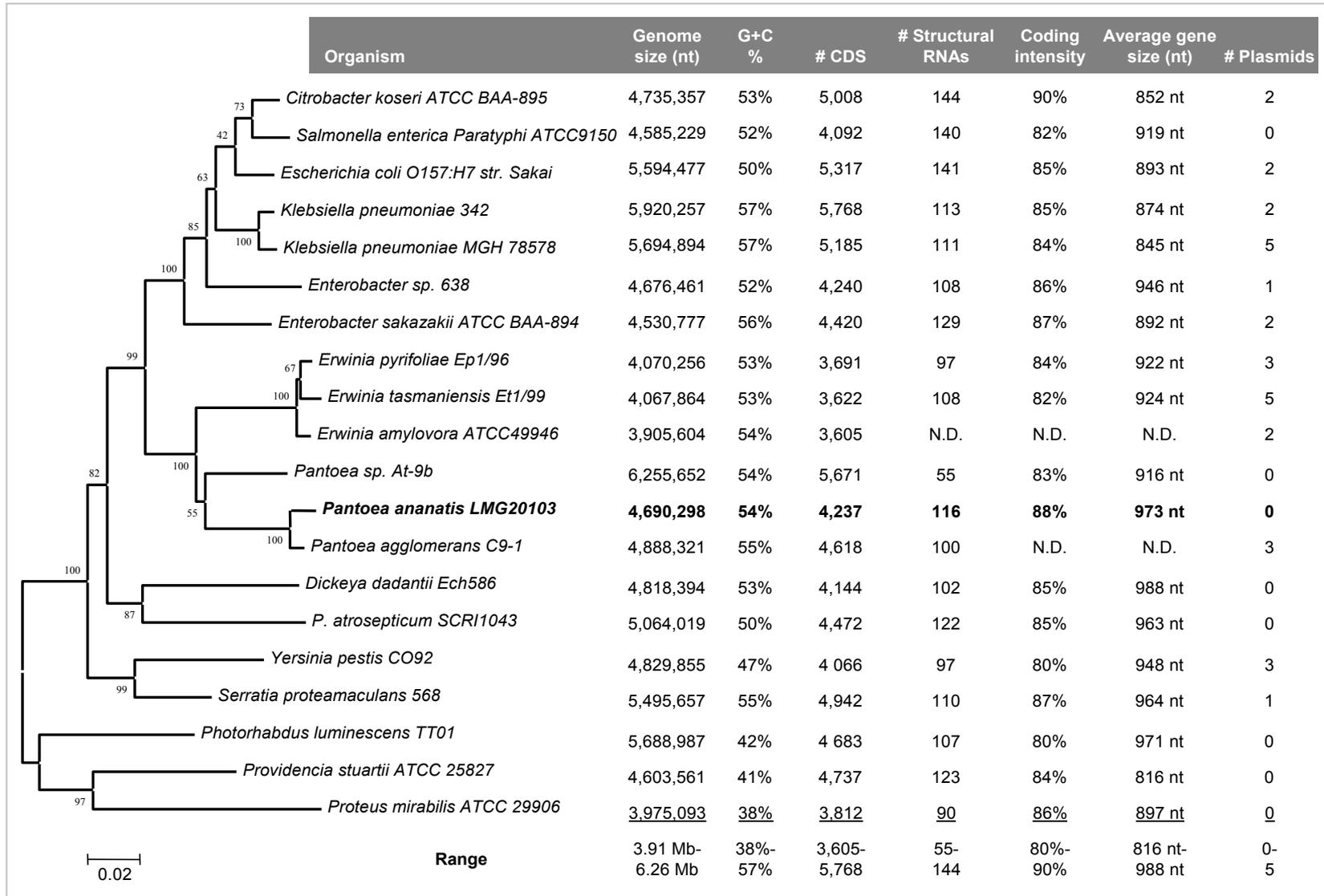| Organism	Genome size (nt)	G+C %	# CDS	# Structural RNAs	Coding intensity	Average gene size (nt)	# Plasmids
Citrobacter koseri ATCC BAA-895	4,735,357	53%	5,008	144	90%	852 nt	2
Salmonella enterica Paratyphi ATCC9150	4,585,229	52%	4,092	140	82%	919 nt	0
Escherichia coli O157:H7 str. Sakai	5,594,477	50%	5,317	141	85%	893 nt	2
Klebsiella pneumoniae 342	5,920,257	57%	5,768	113	85%	874 nt	2
Klebsiella pneumoniae MGH 78578	5,694,894	57%	5,185	111	84%	845 nt	5
Enterobacter sp. 638	4,676,461	52%	4,240	108	86%	946 nt	1
Enterobacter sakazakii ATCC BAA-894	4,530,777	56%	4,420	129	87%	892 nt	2
Erwinia pyrifoliae Ep1/96	4,070,256	53%	3,691	97	84%	922 nt	3
Erwinia tasmaniensis Et1/99	4,067,864	53%	3,622	108	82%	924 nt	5
Erwinia amylovora ATCC49946	3,905,604	54%	3,605	N.D.	N.D.	N.D.	2
Pantoea sp. At-9b	6,255,652	54%	5,671	55	83%	916 nt	0
Pantoea ananatis LMG20103	**4,690,298**	**54%**	**4,237**	**116**	**88%**	**973 nt**	**0**
Pantoea agglomerans C9-1	4,888,321	55%	4,618	100	N.D.	N.D.	3
Dickeya dadantii Ech586	4,818,394	53%	4,144	102	85%	988 nt	0
P. atrosepticum SCRI1043	5,064,019	50%	4,472	122	85%	963 nt	0
Yersinia pestis CO92	4,829,855	47%	4 066	97	80%	948 nt	3
Serratia proteamaculans 568	5,495,657	55%	4,942	110	87%	964 nt	1
Photorhabdus luminescens TT01	5,688,987	42%	4 683	107	80%	971 nt	0
Providencia stuartii ATCC 25827	4,603,561	41%	4,737	123	84%	816 nt	0
Proteus mirabilis ATCC 29906	3,975,093	38%	3,812	90	86%	897 nt	0
Range	3.91 Mb- 6.26 Mb	38%- 57%	3,605- 5,768	55- 144	80%- 90%	816 nt- 988 nt	0- 5

0.02

Table 2.4: COGs identified for the proteins encoded by *P. ananatis* LMG20103. The COG codes, descriptions, number of proteins belonging to each COG and percentages of the total number of proteins total are shown.

Code	Description	COGs	%
J	Translation, ribosomal structure and biogenesis	182	4.30
A	RNA processing and modification	1	0.02
K	Transcription	336	7.93
L	Replication, recombination and repair	163	3.85
B	Chromatin structure and dynamics	0	0.00
D	Cell cycle control, mitosis and meiosis	38	0.90
Y	Nuclear structure	0	0.00
V	Defense Mechanisms	69	1.63
T	Signal Transduction Mechanisms	140	3.30
M	Cell wall/membrane biogenesis	260	6.14
N	Cell motility	103	2.43
Z	Cytoskeleton	0	0.00
W	Extracellular structures	0	0.00
U	Intracellular trafficking and secretion	36	0.85
O	Posttranslational modification, protein turnover, chaperones	142	3.35
C	Energy production and conservion	187	4.41
G	Carbohydrate transport and metabolism	437	10.31
E	Amino acid transport and metabolism	393	9.28
F	Nucleotide transport and metabolism	88	2.10
H	Coenzyme transport and metabolism	141	3.33
I	Lipid transport and metabolism	120	2.83
P	Inorganic ion transport and metabolism	217	5.12
Q	Secondary metabolites biosynthesis, transport and catabolism	51	1.20
R	General function prediction only	349	8.24
S	Function Unknown	784	18.48
Total		**4,237**	**100.0%**

Table 2.6: Proteins localised in distinct subcellular compartments in LMG20103 and *E. coli* K-12.

Localisation	*P. ananatis* LMG20103	%	*E. coli* K-12	%
Cytoplasm	2,563	60.5	2,873	66.1
Inner Membrane	1,258	29.7	973	22.4
Periplasmic	241	5.7	336	7.7
Outer Membrane	120	2.8	149	3.4
Extracellular	55	1.3	14	0.3
Total	**4,237**		**4,345**	

Table 2.5: Percentage of proteins belonging to the different COG functional c... *ananatis* LMG20103 and thirteen closely related *Enterobacteriaceae*. Values which fall below the normal range among other *Enterobacteriaceae* in *P. ananatis* LMG20103 are shown in red, while the COGs above the normal range are shown in green.
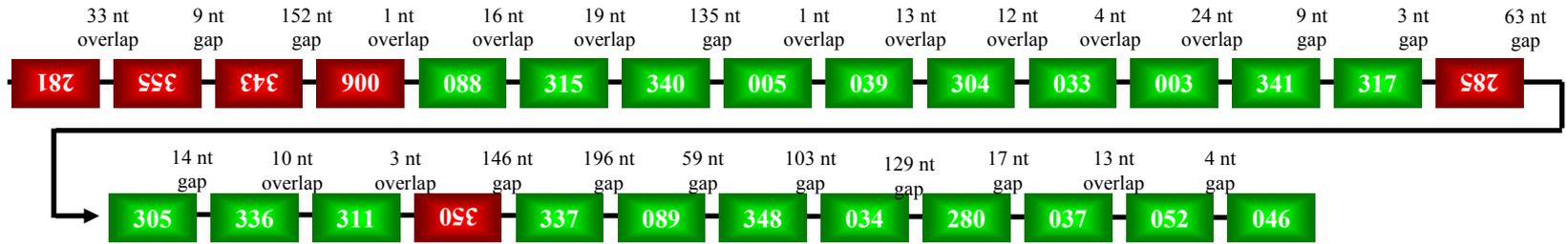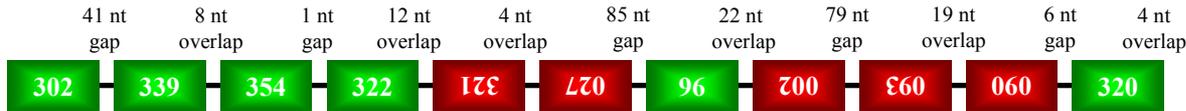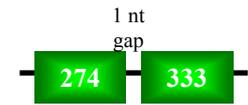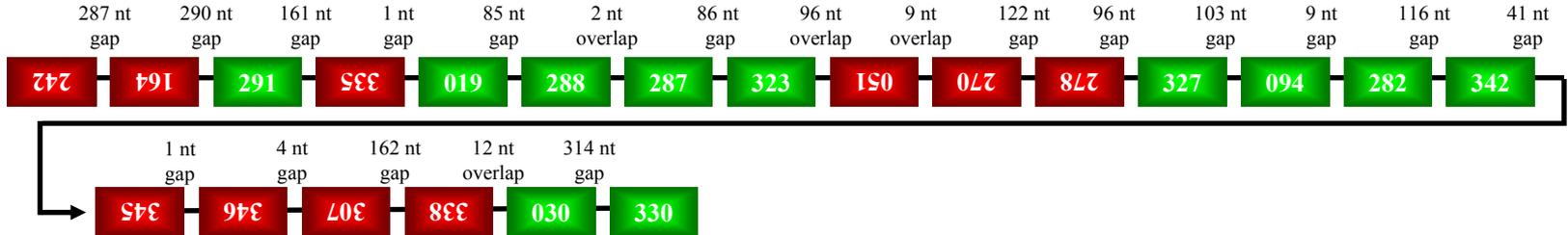
Species	Translation, ribosomal biogenesis	RNA processing and modification	Transcription	Replication, recombination, repair	Chromatin structure and dynamics	Cell cycle control, mitosis/meiosis	Defense mechanisms	Signal transduction mechanisms	Cell wall/membrane biogenesis	Cell motility	Cytoskeleton	Extracellular structures	Intracellular trafficking/secretion	Postranslational modification	Energy production/conversion	Carbohydrate transport/metabolism	Amino acid transport/metabolism	Nucleotide transport/metabolism	Coenzyme transport/metabolism	Lipid transport/metabolism	Ion transport/metabolism	Secondary metabolite biosynthesis	General function prediction only	Unknown function
Citrobacter koseri ATCC BAA-895	3.13	0.02	5.88	2.82	0.00	0.66	0.90	3.55	3.94	2.82	0.00	0.02	2.66	2.65	5.07	7.06	8.01	1.42	3.15	1.90	6.26	1.97	8.56	27.55
S. enterica Paratyphi A ATCC 9150	3.94	0.02	6.74	3.70	0.00	0.72	0.99	3.45	4.88	2.10	0.00	0.00	2.37	3.35	5.94	7.60	8.48	1.75	3.54	1.83	5.07	1.58	10.23	21.72
E. coli O157:H7 str. Sakai	3.19	0.03	5.82	5.39	0.00	0.61	0.90	3.34	4.34	2.00	0.00	0.07	2.66	2.71	4.91	6.41	7.16	1.50	2.63	1.86	5.16	1.45	9.82	28.04
Klebsiella pneumoniae 342	2.96	0.01	8.17	3.11	0.01	0.59	1.02	3.33	3.87	0.84	0.00	0.00	1.68	2.35	5.06	8.74	9.73	1.47	3.03	2.09	6.45	2.02	10.54	22.93
K. pneumoniae MGH 78578	3.29	0.02	8.23	4.16	0.02	0.67	1.20	3.34	4.09	0.98	0.00	0.00	1.83	2.54	5.01	9.30	9.59	1.54	3.13	2.13	6.68	1.98	10.89	19.38
Enterobacter sp. 638	3.99	0.02	7.28	3.76	0.00	0.99	1.05	4.05	4.79	1.85	0.00	0.04	1.91	2.78	4.65	7.77	9.04	1.59	3.12	2.13	6.42	1.75	11.15	19.87
E.sakazakii ATCC BAA-894	3.71	0.02	6.57	3.54	0.02	0.73	0.90	3.78	4.52	2.20	0.00	0.02	2.01	3.07	4.25	7.10	7.65	1.75	3.01	1.93	4.82	1.63	10.18	26.59
Pantoea ananatis LMG20103	4.22	0.02	7.55	3.82	0.00	0.73	**1.98**	3.07	**5.59**	2.93	0.00	0.00	**0.92**	2.76	3.94	**10.29**	8.76	2.01	3.07	2.19	4.32	**1.25**	10.55	20.01
Erwinia tasmaniensis Et1/99	4.56	0.02	6.49	3.82	0.00	0.85	1.00	3.79	5.01	3.12	0.00	0.00	3.62	3.14	3.82	5.46	9.30	2.05	3.29	2.07	5.51	1.45	9.88	21.75
P. atrosepticum SCRI1043	3.59	0.02	7.06	3.65	0.00	0.86	0.90	3.97	4.72	2.07	0.00	0.00	2.42	2.88	4.74	7.06	8.96	1.69	3.17	2.19	6.43	2.05	10.42	21.15
Yersinia pestis CO92	3.88	0.02	5.57	7.93	0.00	0.75	0.94	2.83	4.54	2.89	0.00	0.09	3.41	2.87	4.11	7.24	7.69	1.63	3.11	1.89	5.40	1.65	8.65	22.91
Serratia proteamaculans 568	3.42	0.02	8.52	3.34	0.02	0.63	0.93	3.32	4.29	1.78	0.00	0.00	2.22	2.73	4.83	7.74	10.12	1.80	2.91	2.54	6.44	2.30	11.86	18.24
Proteus mirabilis ATCC 29906	4.50	0.02	6.57	4.46	0.00	0.84	1.10	2.71	4.63	3.00	0.00	0.07	3.31	3.14	5.23	4.63	7.74	2.11	3.21	1.99	5.37	1.77	10.00	23.60
P. luminescens subsp. laumondii TTO1	3.65	0.02	6.62	5.62	0.02	0.97	1.26	2.70	4.14	1.89	0.04	0.02	2.47	2.42	3.45	4.01	6.97	1.61	3.41	2.32	3.75	2.90	9.33	30.41
Average	3.68	0.02	6.89	4.25	0.01	0.76	1.01	3.40	4.44	2.12	0.00	0.03	2.51	2.82	4.70	6.93	8.50	1.69	3.13	2.07	5.67	1.88	10.12	23.40
Range	2.96-4.56	0.01-0.03	5.57-8.52	2.82-7.93	0.00-0.002	0.61-0.99	0.90-1.26	2.70-4.05	3.87-5.01	0.84-3.12	0.00-0.04	0.00-0.09	1.68-3.62	2.35-3.35	3.45-5.94	5.46-9.30	7.16-10.12	1.42-2.11	2.66-3.41	1.83-2.54	3.75-6.68	1.45-2.90	8.56-11.15	18.24-30.41

Tree node support values (left of species names): 68, 46, 66, 85, 100, 99, 97, 92, 100, 100, 98.

Table 2.7: Proteins in P2-like phage elements Phage 2 and 3 integrated into the LMG20103 chromosome. Homologous proteins in *S. enterica* Typhimurium LT2 Fels-2 prophage are shown.

Gene	Product	Phage 2	Phage 3	Fels-2 P2 Phage
int	Phage integrase	PANA_3178	-	STM2739.Fels2
CI	Phage immunity repressor	PANA_3177	PANA_3422	STM2738.Fels2
x1	Phage regulatory protein	-	-	STM2737.Fels2
a1	Phage regulatory protein	PANA_3176	-	-
CII	Phage regulatory protein	PANA_3175	-	STM2736.Fels2
fil	Hypothetical phage protein gp32	PANA_3174	-	STM2735.Fels2
X2	Hypothetical phage protein	-	-	STM2734.Fels2
X3	Hypothetical phage protein	-	-	STM2733.Fels2
b1	*Hypothetical protein*	-	PANA_3421	-
a2	*Hypothetical protein*	PANA_3173	-	-
a3	*Hypothetical protein*	PANA_3172	-	-
a4/b2	Hypothetical phage protein gp33	PANA_3171	PANA_3420	STM2732.Fels2
a5/b3	Hypothetical phage protein gp34	PANA_3170	PANA_3419	STM2731.Fels2
Dam	DNA adenine methyltransferase	PANA_3169	-	STM2730.Fels2
a6	Hypothetical phage protein gp57	PANA_3168	-	-
gpA	Replication protein A	PANA_3167	PANA_3418	STM2729.Fels2
a7	Hypothetical phage protein	PANA_3166	-	STM2728.Fels2
TUM	SOS operon protein	PANA_3165	PANA_3417	STM2727.Fels2
X4	Hypothetical phage protein	-	-	STM2726.Fels2
X5	Hypothetical phage protein	-	-	STM2724.Fels2
hsdR	DNA cytosine methyltransferase	PANA_3164	-	-
a8	Hypothetical phage protein gp96	PANA_3163	-	-
a9	Hypothetical phage protein gp98	PANA_3162	-	-
a10	Phage membrane protein	PANA_3161	-	-
a11	Phage acyl carrier protein	PANA_3160	-	-
gpQ	Capsid portal protein	PANA_3159	-	STM2723.Fels2
gpP	Phage terminase ATPase subunit	PANA_3158	-	STM2722.Fels2
gpO	Phage capsid scaffolding protein	PANA_3157	-	STM2721.Fels2
gpN	Phage capsid protein	PANA_3156	-	STM2720.Fels2
gpM	Phage small terminase subunit	PANA_3155	-	STM2719.Fels2
gpL	Phage capsid completion protein	PANA_3154	-	STM2718.Fels2
gpX	Phage tail protein	PANA_3153	PANA_3416	STM2717.Fels2
nucE	Phage holin	PANA_3152	PANA_3415	STM2716.Fels2
nucD2	Prophage lysozyme	PANA_3151	PANA_3414	STM2715.S.Fels2
x6	Hypothetical phage protein	-	-	STM2714.2.Fels2
x7	Hypothetical phage protein	-	-	STM2714.1.Fels2
lysB	Lysis protein B	PANA_3150	PANA_3413	STM2714.Fels2
a12	*Hypothetical protein*	PANA_3149	-	-
lysC	Lysis protein C	PANA_3148	-	STM2713.Fels2
gpR	Phage tail completion protein	PANA_3147	PANA_3412	STM2712.Fels2
b4	*Hypothetical protein*	-	PANA_3411	-
gpS	Phage tail completion protein	PANA_3146	-	STM2711.Fels2
a13	*Hypothetical protein*	PANA_3145	-	-
gpV	Phage baseplate assembly protein	PANA_3144	PANA_3410	STM2710.Fels2
gpW	Phage baseplate assembly protein	PANA_3143	PANA_3409	STM2709.Fels2
gpJ	Phage baseplate assembly protein	PANA_3142	PANA_3408	STM2708.Fels2
gpI	Phage tail protein	PANA_3141	PANA_3407	STM2707.Fels2
gpH	Phage tail-collar protein	PANA_3140	PANA_3406	STM2706.Fels2
gpG	Phage tail-fibre protein	-	PANA_3405	STM2704.Fels2
gpH	Phage tail-collar protein	-	-	STM2703.Fels2
tfaE	Phage tail assembly chaperone	PANA_3139	-	-
din	DNA invertase	-	-	STM2702.Fels2
b5	*Hypothetical protein*	-	PANA_3404	-
gpFI	Phage tail-sheath protein	PANA_3138	PANA_3403	STM2701.Fels2
gpFII	Phage tail-tube protein	PANA_3137	PANA_3402	STM2700.Fels2
gpE	Phage tail protein	PANA_3136	PANA_3401	STM2699.Fels2
b6	*Hypothetical protein*	-	PANA_3400	-
gpE2	Phage tail protein	PANA_3135	PANA_3399	STM2698.Fels2
gpT	Phage tail tape measure protein	PANA_3134	PANA_3398	STM2697.Fels2
gpU	Phage tail protein	PANA_3133	PANA_3397	STM2696.Fels2
gpD	Phage late-control protein	PANA_3132	PANA_3396	STM2695.Fels2
gpB	Phage Transcriptional activator Ogr	-	PANA_3395	STM2694.Fels2
b7	*Hypothetical protein*		PANA_3394	
b8	*Hypothetical protein*		PANA_3393	
int	Phage integrase	PANA_3131	PANA_3392	-

84

Table 2.8: Proteins encoded in each of the P4 phages (Phage 1 and 4) and homologues identified in the *Escherichia fergusonii* P4 phage.

Gene	Product	Phage 1	Phage 4	*E. fergusonii* P4 phage
int	Phage integrase	PANA_2912	PANA_3913	EFER_0816
c1	*Hypothetical protein*	-	PANA_3914	-
y1	Hypothetical phage protein		-	EFER_0815
alpA	Phage transcriptional regulator	PANA_2916	PANA_3915	EFER_0814
c2	Hypothetical phage protein	-	PANA_3916	EFER_0813
c3	Hypothetical phage protein	-	PANA_3917	EFER_0811
c4	Hypothetical phage protein	-	PANA_3918	EFER_0810
c5/d4	Hypothetical phage protein	PANA_2919	PANA_3919	EFER_0809
Alpha	Phage DNA primase	PANA_2920	PANA_3920	EFER_0808
sid	Capsid head size protein	-	PANA_3921	EFER_0807
y2	Hypothetical phage protein		-	EFER_0806
ogr	Phage transcriptional activator	PANA_2915	PANA_3922	EFER_0805
sidI	Capsid head size protein	-	PANA_3923	
CI	Phage immunity repressor	PANA_2917	-	EFER_0812
d1	*Hypothetical protein*	PANA_2913	-	-
d2	*Hypothetical protein*	PANA_2914	-	-
d3	Hypothetical phage protein	PANA_2918	-	-

Table 2.8: Proteins encoded in each of the P4 phages (Phage 1 and 4) and homologues identified in the *Escherichia fergusonii* P4 phage.
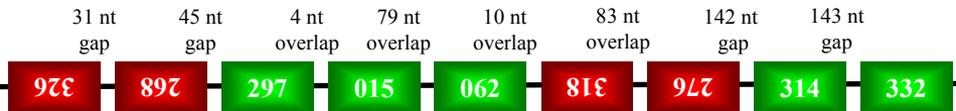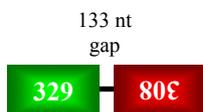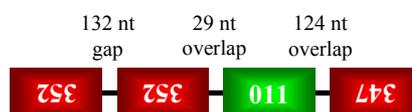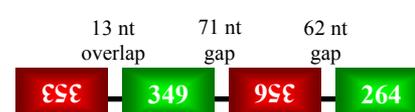
Figure 2.1: First draft assembly of the LMG20103 genome into nine scaffolds. Putative gap sizes were predicted through BLASTN against the NCBI nucleotide database, where the same gap size was found in five or more closely related strains. Contigs in red were inserted in reverse complement, while those indicated in green contigs were assembled in their forward form.
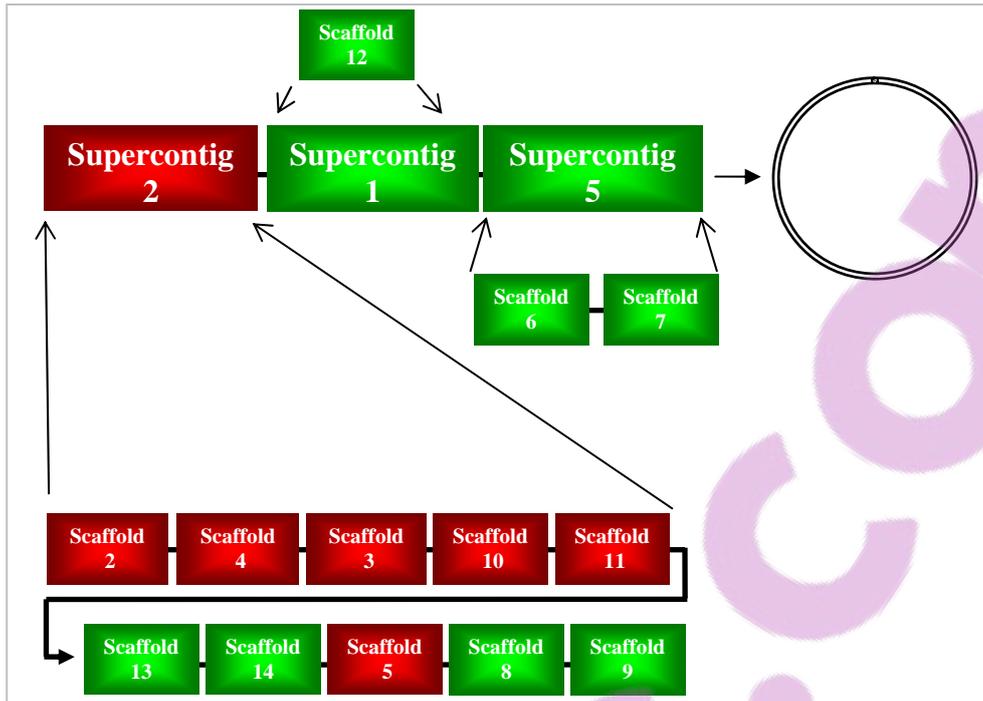
Figure 2.2: Schematic diagram of the second generation draft assembly. The scaffolds were assembled into super-contigs as indicated. Red scaffolds/contigs indicates those assembled in reverse complement while those in green were assembled in original orientation.
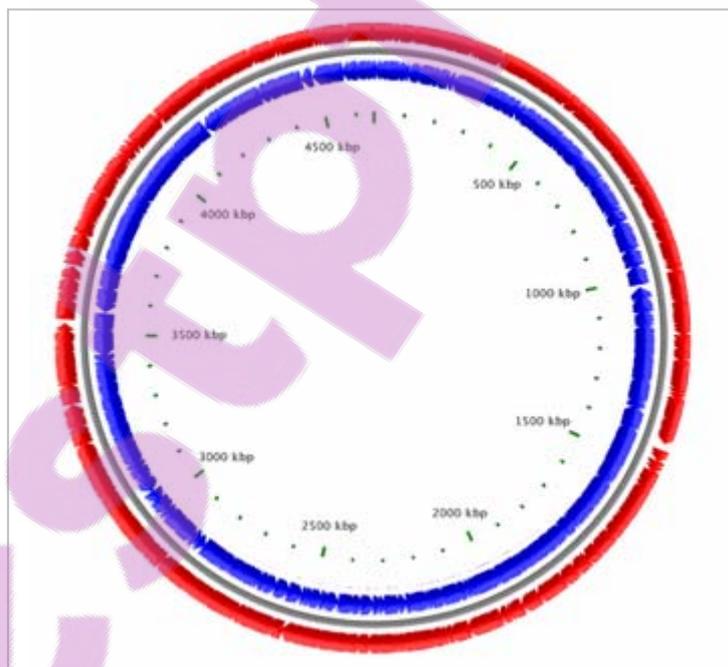
Figure 2.3: BASYS diagram of the protein coding genes on the LMG20103 chromosome. CDSs encoded on the forward strand are coloured in red, while those encoded on the reverse strand are depicted in blue.
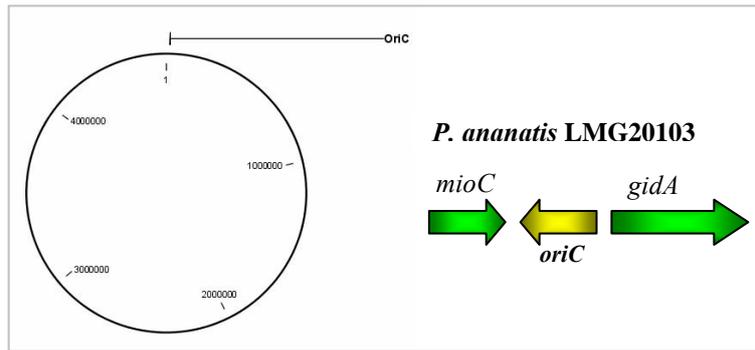
Figure 2.4: location of the origin of replication *oriC* and a schematic diagram of the flanking genes.
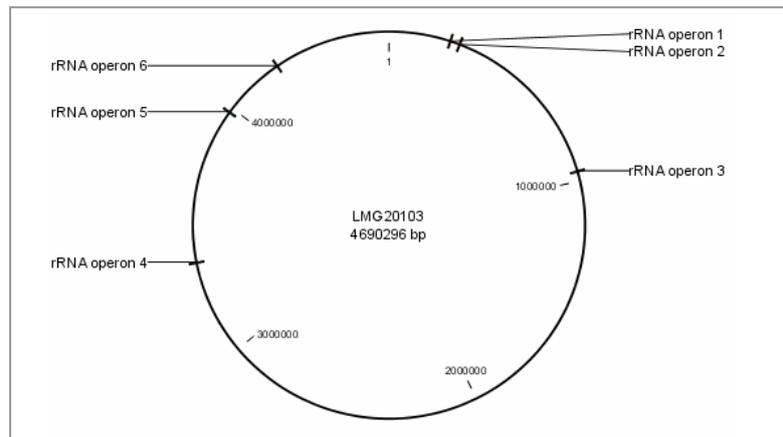

Figure 2.5: chromosomal locations of the rRNA operons in *P. ananatis* LMG20103.
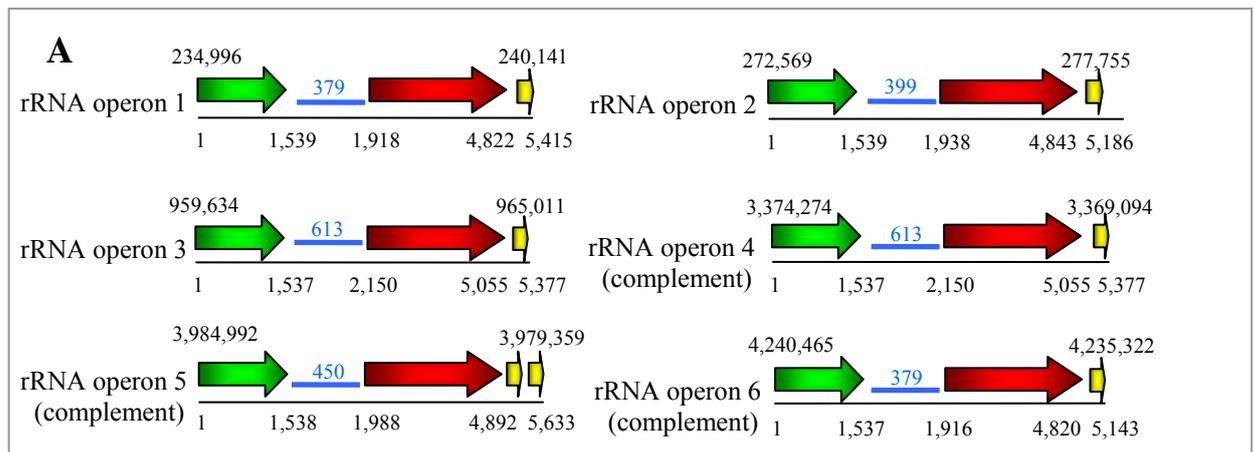

Figure 2.6: Schematic diagram of the rRNA operons in *P. ananatis* LMG20103. 16S rDNA genes are coloured in green, 23S rDNA in red and 5S rDNA in yellow. The ITS region is coloured blue. The size of the ITS region and locations of the operons within the genome are shown.
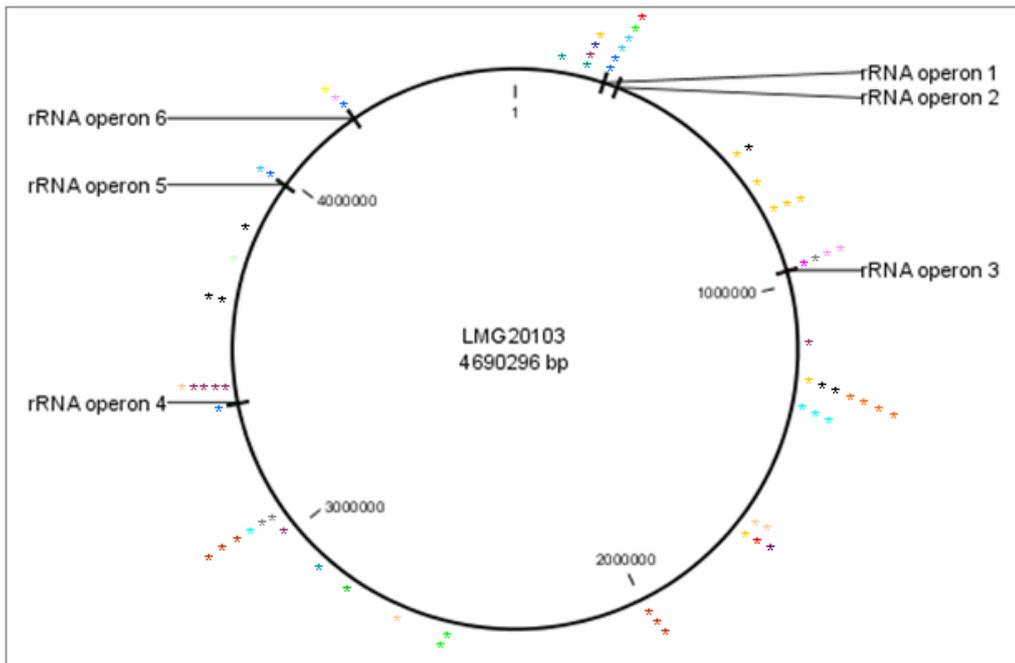
Figure 2.7: Chromosomal distribution of tRNA-coding sequences on the genome of *P. ananatis* LMG20103. The different colour stars indicate the different codon-specific transfer RNAs and follow the colour-scheme used in Figure 2.8.
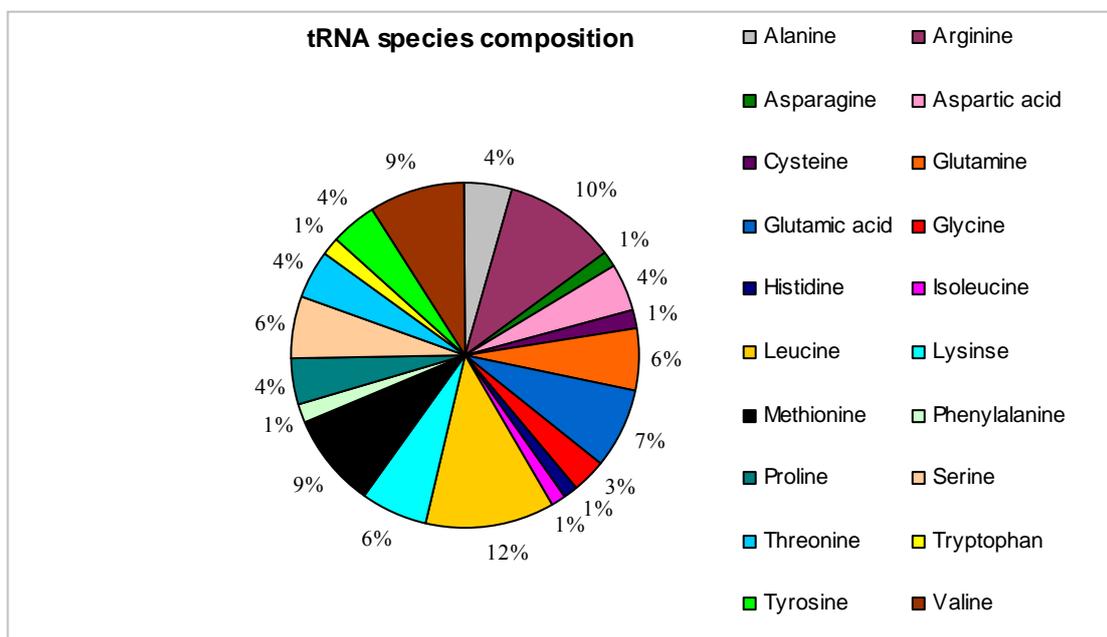

Figure 2.8: Pie chart showing the proportions of the different codon-specific transfer RNAs in *P. ananatis* LMG20103.
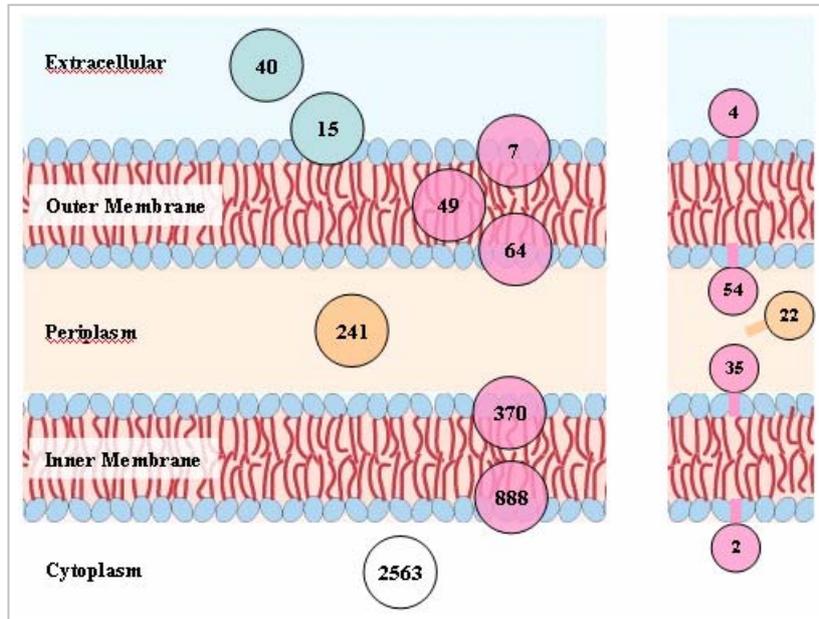
Figure 2.9: Predicted localisations of LMG20103 proteins in the different cellular compartments. The localisations of the predicted lipoproteins are shown on the right.
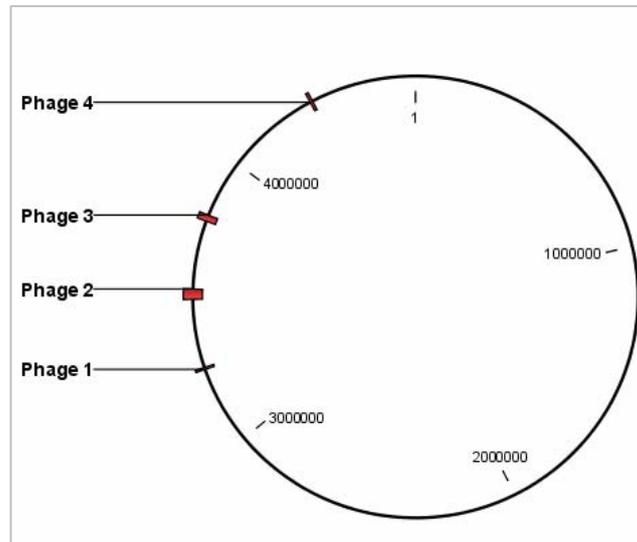

Figure 2.10: Chromosomal integration sites of the two P4 phages (Phage 1 and 4) and the two P2 phages (Phage 2 and 3) in *P. ananatis* LMG20103.
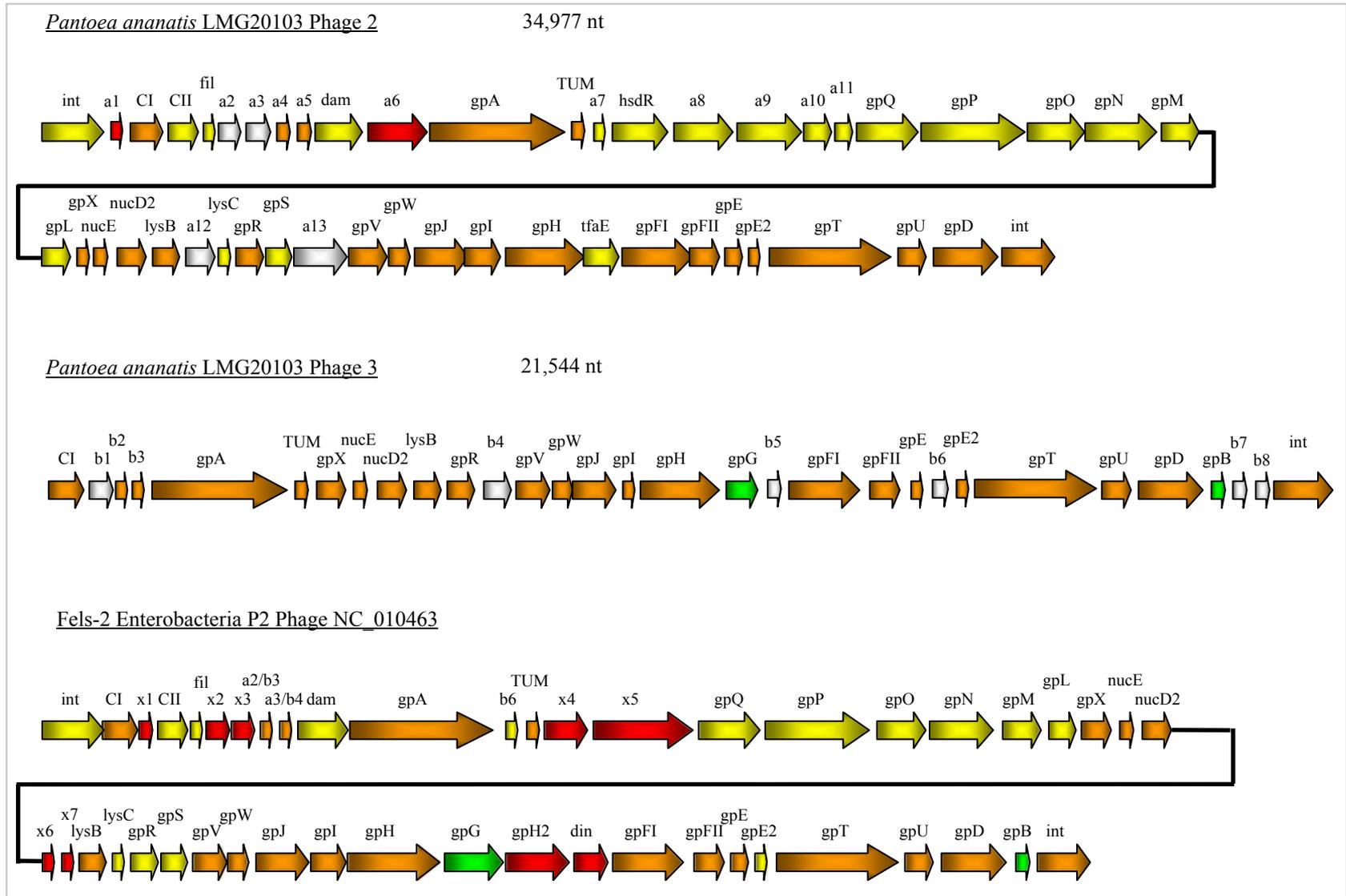
Figure 2.11: Schematic diagram of genes of two integrated P2-type bacteriophages in *P. ananatis* LMG20103, which show similarity to Enterobacteria phage Fels-2. Shaded in yellow are phage genes shared between Phage 2 and Phage Fels2, in green those shared between Phage 2 and Phage Fels2. Orange arrows are genes sharing homologues in all three phages, while those shaded in red represent phage proteins restricted to a particular phage. White arrows indicate those genes for which no evidence of phage origin is available.
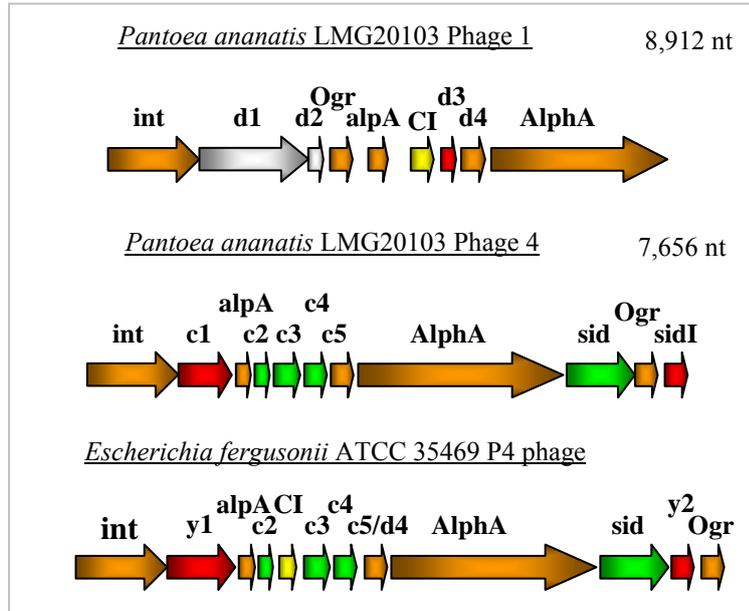
Figure 2.12: Schematic diagram of the genes of two integrated P4-type bacteriophages in *P. ananatis* LMG20103 which show similarity to the *Escherichia fergusonii* ATCC 35469 P4 phage. Shaded in yellow are genes shared between Phage 1 and *E. fergusonii* P4 phage, in green are those shared between Phage 4 and *E. fergusonii* P4 phage. Orange blocks are genes shared between all three phages, while those shaded in red are only found in the individual phages. White genes indicate those for which no evidence of phage origin was found.

CHAPTER 3:

Comparative Genomics reveals key targets for environmental colonisation and plant pathogenesis in the wide host range pathogen *Pantoea ananatis*

ABSTRACT

Pantoea ananatis is a ubiquitous organism that is found in a wide range of environments and has further been implicated in a wide range of plant diseases as well as human pathogenesis. By means of genome comparisons an attempt was made to identify the molecular basis underlying its ecological success and cross-kingdom pathogenesis. Comparison to closely related members of the *Enterobacteriaceae* revealed that the *Eucalyptus* pathogen *P. ananatis* LMG20103 has an extensive flexible genome. This flexible genome was compared against all the genomes that are publicly available which showed that it carries a number of genes which are restricted to bacteria occupying distinct ecological niches, namely plant-associated, animal-associated and insect-associated bacteria. These may encode proteins that play a role in its persistence in these different hosts as well as the environment. A number of these genes reside on genomic islands which indicates that *P. ananatis* has adapted through horizontal acquisition to become adept at successfully colonise a range of environments. Several of these islands also carry pathogenicity determinants which may play a role in pathogenesis of both the plant and vertebrate host.

INTRODUCTION

The ability of bacteria to survive within and exploit almost every environment on earth can be ascribed to their extensive capacity of adaptive evolution. This is exemplified by members of the genus *Pseudomonas* that have evolved to thrive in natural environments including fresh and marine water and the terrestrial environment, as well forming intimate relationships with both plants and animals (Spiers et al. 2000). The exploitation of novel ecological niches by a bacterium can be facilitated by point mutations and gene conversions, genome rearrangements and the acquisition of foreign DNA, frequently in the form of stretches of DNA known as horizontally acquired islands (HAIs). The latter allows rapid adaptation of a bacterium to become adept at survival, nutrient-acquisition, competitiveness and proliferation within novel ecological niches (Schmidt and Hensel, 2004). Some pathogenic bacteria have also adapted to infect animal or plant hosts by acquiring DNA from other bacteria in the form of pathogenicity islands (PAIs) which carry factors capacitating infection and symptom development. Other pathogenic bacteria have evolved the potential for cross-kingdom pathogenicity, such as *Pseudomonas aeruginosa*, which

carries two horizontally acquired islands that carry pathogenicity genes required for disease in both humans and plants (van Baarlen et al. 2007).

Over 1,000 prokaryotic genomes have been sequenced to date (Lagesen et al. 2010). This has seen the concomitant development of genome comparison which has become a useful tool to study the biology of microorganisms inhabiting different ecological niches and the acquisition of genes for adaptation to novel niches (Binnewies et al. 2006; Schmidt and Hensel, 2004). Comparison of the genome sequences of closely related bacteria allows the identification of those genes which are shared between the compared strains and those which are found in only a subset of the bacteria or are unique to a particular bacterium. The set of protein coding sequences which are shared between the compared isolates belonging to a species or genus is known as the "core genome" and includes genes which are indispensable for normal cell functioning and survival (Medini et al. 2005; Vinatzer and Yan, 2008). These genes are transmitted vertically from a common ancestor of the compared strains (Vinatzer and Yan, 2008). The proportion of CDSs of a strain which are not shared with the other compared bacteria is purported to have been horizontally acquired subsequent to divergence of the compared organisms and is referred to as the "flexible" or "accessory genome" (Vinatzer and Yan, 2008). Within this flexible genome are genes which encode proteins involved in the fitness of a bacterium in the environmental niches it inhabits as well as genes which are involved in pathogenesis on the host it infects (Vinatzer and Yan, 2008). For example, the comparison of all available *Pseudomonas* genomes showed the extensive genome plasticity within this genus and identified metabolic differences between strains isolated from plants, insects and animals and from the natural environment allowing them to thrive in their respective ecological niches, which could be linked to their flexible genomes (Perumal et al. 2008). Comparison of genomes of the non-pathogenic pear epiphyte *Erwinia tasmaniensis* and the closely related pear pathogens *E. pyrifoliae* and *E. amylovora* revealed several differences between the non-pathogenic and pathogenic isolates, identifying pathogenicity islands and fitness islands differentially distributed among the compared strains (Smits et al. 2010).

Pantoea ananatis is a member of the family *Enterobacteriaceae* and is characterised by its ubiquity in nature and its ability to cause disease in a wide range of plants. It

95

has been found in a wide array of environments including rivers, soil, meat packaging and aviation fuel tanks (Coutinho and Venter, 2009; Ercolini et al. 2006). It is habitually isolated from plant roots, leaves and stems and exists as part of epiphytic and endophytic flora on a broad range of plants (Coutinho and Venter, 2009). Since its identification as the causal agent of fruitlet rot of pineapple in the Philippines in 1928 (Serrano, 1928) it has been implicated in diseases of a wide range of host crops including maize and onion, *Eucalyptus*, sudangrass and honeydew melons (Coutinho and Venter, 2009). Individual isolates also appear to be capable of causing disease on a wide range of hosts. For example, the *Eucalyptus* isolate *P. ananatis* LMG20103 can induce symptom development on onion and maize (Chapter 6; results unpublished). *P. ananatis* is also associated with insects, including tobacco thrips which act as vectors for onion pathogenic strains, mulberry pyralids, ticks and fleas, demonstrating its ability to persist in invertebrate hosts (Coutinho and Venter, 2009). Its implication in bacteraemic infections further reveals its capacity for proliferation and potential to cause disease in human hosts (De Baere et al. 2004).

The ubiquity of *P. ananatis* suggests it has adapted to proliferate in a wide range of environments. This has likely occurred through horizontal acquisition of genes from microorganisms within the environments it co-inhabits. Similarly, it may have acquired foreign DNA allowing it to persist and cause disease in a range of plant, insect and human hosts. The protein coding sequences (CDSs) encoded on the complete genome sequence of the *Eucalyptus* pathogen *P. ananatis* strain LMG20103 were compared to those of a large set of bacteria. Firstly, the CDSs of LMG20103 were compared to the genomes of closely related animal-pathogenic, plant-associated and plant-pathogenic members of the family *Enterobacteriaceae*, including a draft genome assembly of its nearest phylogenetic relative, *Pantoea stewartii* subsp. *stewartii*, the causal agent of Stewart's wilt of maize. This revealed a large subset of proteins which are shared among the compared *Enterobacteriaceae*, but also a significant number of proteins which are found only in the compared plant-associated and animal-pathogenic bacteria as well as a set of proteins for which no homologues could be found in the compared organisms. The latter two subsets were subsequently compared to all the genomes publicly available on the National Center for Biotechnology Information (NCBI) server to confirm those proteins with homologues only in plant-pathogenic, plant-associated as well as animal-associated bacteria.

Horizontally acquired islands were also predicted to determine if any of the protein coding genes with homologues restricted to bacteria occupying distinct ecological niches have been acquired recently.

MATERIALS AND METHODS

Comparison between *P. ananatis* LMG20103 and closely related *Enterobacteriaceae*

The protein coding sequences (CDSs) of the *Eucalyptus*-pathogenic strain *P. ananatis* LMG20103 were determined in Chapter 2. These CDSs were compared to all the CDSs encoded on the genomes of eight closely related *Enterobacteriaceae*, namely the enterohaemorrhagic human pathogen *E. coli* O157:H7 (Hayashi et al. 2001), the human typhoid pathogen *S. enterica* subsp. *enterica* serovar Paratyphi ATCC9150 (McClelland et al. 2004), the apple epiphyte *E. tasmaniensis* Et1/99 (Kube et al. 2008), the poplar endophytes *Enterobacter* sp. 638 and *Serratia proteamaculans* 568 (Taghavi et al. 2009) and the pear pathogen *Erwinia amylovora* Ea273 (Sebaihia et al. 2010). The draft assemblies of the genome of *Pantoea* sp. At-9b isolated from the fungus garden of leaf-cutter ants (Pinto-Tomás et al. 2009) and the Stewart's wilt agent *Pantoea stewartii* subsp. *stewartii* DC283 (ASAP database; Https://asap.ahabs.wisc.edu/asap/; results unpublished) were also included. Comparisons were performed by local BlastP using the Bioedit 7.0.5.3 software package (Hall, 1999) using a reciprocal best hit (RBH) approach. RBH involves BLAST analysis of one genome against another, followed by a reciprocal BLAST analysis with the second against the first compared genome. Homologues are assumed only when one CDS each in the two separate genomes represents the best hit in a BLAST against the other genome (Moreno-Hagelsieb and Latimer, 2007). Homologues were considered as those CDSs with greater than 30 % amino acid (aa) identity over 80% of the average length of the two proteins. CDSs sharing homology in both of the compared genomes and those unique to one of the genomes were tabulated and the average amino acid identity between shared CDSs was determined.

The conserved orthologous group (COG) functions for *P. ananatis* LMG20103 were determined as described in Chapter 2. By the same means the COG functions for all *P. stewartii* subsp. *stewartii* DC283 CDSs were determined. The COG functions of the shared and unique proteins were determined for five more compared *Enterobacteriaceae* for which COG information is publicly available on the NCBI

97

server, namely *S. proteamaculans* 568, *Enterobacter* sp. 638, *E. tasmaniensis* Et1/99, *E. coli* O157:H7 str. Sakai and *S enterica* subsp. Paratyphi ATCC9150. The relative percentages of COGs missing from each organisms being compared and the average percentages of CDSs belonging to particular COGs among the compared organisms were ascertained.

Comparison between *P. ananatis* LMG20103 and all available genomes

The LMG20103 CDSs with homologues restricted to the animal-pathogenic or plant-associated *Enterobacteriaceae* used in the comparison described above and those *P. ananatis* CDSs without homologues in any of the these enterobacteria were compared by BlastP to 1,099 bacterial genomes which are publicly available on the NCBI server. On the basis of the blast hits, the CDSs were classed into groups with homologues restricted to the animal-associated bacteria (AAB) (including bacteria associated with insects (IAB), humans and animals), the plant-associated bacteria (PAB) including those belonging to the family *Enterobacteriaceae* (PAE) and non-*Enterobacteriaceae* (NEPAB), *Pantoea* species (homologues in the genomes of *P. stewartii* DC283 and/or *Pantoea* sp. At-9b only) and those restricted to *P. ananatis* LMG20103. The COG and specific functions of the proteins with homologues restricted to the different classes were determined based on available literature. A diagram of the comparison between the protein coding sequences of LMG20103 and those of 52 other bacteria was constructed using Genome Diagram (Pritchard et al. 2006). Compared bacterial genomes included those of the two *Pantoea* species sequenced to date, three *Erwinia* species, 22 strains belonging to the family *Enterobacteriaceae* representing plant-pathogenic, plant-associated, animal-pathogenic and insect-associated species, as well as those of twenty five plant-associated and plant-pathogenic bacteria from other families (Fig. 3.3). Comparison was performed as per the criteria outlined above.

Identification of horizontally acquired islands (HAIs)

Horizontally acquired islands (HAIs) in *P. ananatis* LMG20103 were identified by a combination of approaches. Firstly, two techniques were used to identify compositional deviations within the genome sequence of LMG20103. The software package Oligowords 1.1 (Reva and Tümmler, 2005) identifies variance in **o**ligonucleotide **u**sage (OU) within the genome sequence. Nucleotide usage is not

98

random within bacterial genomes, being influenced by the physicochemical properties of the sequence, including codon usage, base stacking energy and equal mutational pressure throughout the genome (Reva and Tümmler, 2005; Vernikos and Parkhill, 2006). Moving along the genome sequence one base at a time, the program looks for 4-mer "words" consisting of all combinations of the four nucleotide bases and calculates their distribution. Regions different from the normal distribution of the 4-mer words are considered to form part of horizontally transferred regions (Reva and Tümmler, 2005). Alienhunter (Vernikos and Parkhill, 2006; Sanger Institute) extends on the OU method by searching for compositional anomalies with 1-8 nucleotides "words". This package subsequently computes the boundaries of putatively horizontally transferred DNA regions by means of a two-state Hidden Markov Model (Vernikos and Parkhill, 2006).

Recently transferred HAIs are frequently characterized by differences in G+C content, as the sequence tends to still resemble that of the donor organism (Schmidt and Hensel, 2004). The LMG20103 genome was divided into 1,000 nt blocks and their G+C content was established. The standard deviation for all blocks was determined. Those blocks whose G+C content fell above and below the range of average G+C content (53.69 %) ± standard deviation were retained. Where adjacent blocks fell outside the normal range they were clustered and analysed. HAIs are also frequently associated with mobile genetics elements such as transposases, insertion elements and integrases. Furthermore tRNAs are frequently associated with HAIs as they are highly conserved among different bacteria and function as anchor points for insertion (Schmidt and Hensel, 2004). The presence of tRNAs, transposases, insertion elements and integrases was determined for the predicted islands. The HAIs were mapped on a GenomeDiagram and the proteins encoded within the HAIs compared to the PAB, AAB, IAB, *Pantoea*-specific CDSs.

RESULTS AND DISCUSSION

Comparison between *P. ananatis* LMG20103 and closely related *Enterobacteriaceae*
The 4,237 *P. ananatis* LMG20103 CDSs were compared to those encoded on the genomes of eight closely related *Enterobacteriaceae*. This comparative analysis showed that between 57.5 and 78.0% of *P. ananatis* CDSs share homologues in the compared organisms, indicating an extensive core set of proteins are encoded on the

99
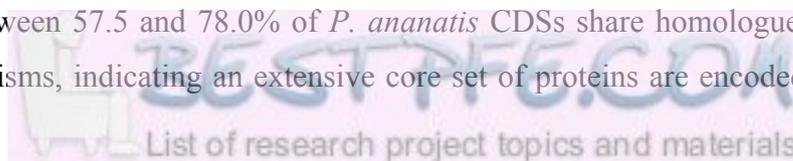
genomes of all the compared strains (Table 3.1). Interestingly, more LMG20103 proteins shared homologues in the fungal-garden associated *Pantoea* sp. At9b than in the more closely related *P. stewartii* DC283. This may be due to a more similar lifestyle and ecology shared between *P. ananatis* LMG20103 and *Pantoea* sp. At9b, but is more likely due to the 6.6 Mb *Pantoea* sp. At9b genome, which encodes 879 more proteins than the 5.5 Mb *P. stewartii* DC283 genome, many of which share homologues in LMG20103. Similarly, when analysing the proportion of CDSs in the other enterobacteria with homologues in *P. ananatis*, a far greater relative proportion of homologues were noted with organisms that have smaller genomes (*E. amylovora* Ea273 and *E. tasmaniensis* Et1/99), while organisms encoding a larger set of CDSs (*Pantoea* sp. At9b and *E. coli* O157:H7) share a lower percentage of homologous CDSs (Table 3.1). The average aa identity between homologues, however, follows the phylogeny inferred by comparison of the DNA gyrase B (GyrB) aa sequences and ranges from 71.9 % (*S. proteamaculans*) to 91.5 % (*P. stewartii*) (Table 3.1).

From the comparison it could be seen that a proportion of the CDSs are encoded on the LMG20103 genome as well as those of all eight compared organisms. This set of protein coding sequences makes up the core genome of the compared *Enterobacteriaceae* which is vertically transmitted from their common ancestor (Vinatzer and Yan, 2008). A proportion of the CDSs encoded on the flexible LMG20103 genome are however not shared with all the compared bacteria. A core genome of 1,728 CDSs was observed from the comparison of the eight enterobacteria, which encompasses 40.7% of the total CDSs on the LMG20103 genome, while its flexible genome carries 59.3% of its CDSs. The core genome of the three sequenced *Pantoea* strains was also determined, with 2,855 of the total LMG20103 CDSs (67.4%) shared in common with both the fungal-garden associated *Pantoea* sp. At9b and the Stewart's wilt pathogen *P. stewartii* subsp. *stewartii* DC283. The flexible genome of LMG20103 among the sequenced *Pantoea* encompasses 1,382 CDSs, of which 450 CDSs (32.6%) are shared in common with *Pantoea* sp. At9b only, 345 CDSs (25.0%) are shared in common with *P. stewartii* DC283 only and 587 CDSs (42.4%) are only found in *P. ananatis* LMG20103 and not in the two compared *Pantoea* strains.

Proteins encoded on the enterobacterial core genome include those involved cellular functions such as ribosome synthesis, DNA replication cell division as well as flagellum biosynthesis. Also shared are proteins encoding components of systems involved in the transport and metabolism of several carbohydrate, amino acid, inorganic ion, lipid and nucleotide substrates, which may be linked to the similar metabolic profiles among the *Enterobacteriaceae* that have been derived from a common ancestor. Other proteins show a more incongruent distribution among the compared bacteria. The COG functions of components encoded on the flexible genomes of *P. ananatis* LMG20103, *P. stewartii* DC283 and the five enterobacterial members for which COG data is available on the NCBI web server, were determined. This showed that CDSs involved in several COG functions are more prevalent in LMG20103, while others are more dominant in the compared *Enterobacteriaceae*.

Functions of *P. ananatis* LMG20103 CDSs without homologues in compared *Enterobacteriaceae*

Comparison of the flexible genomes of LMG20103 and those of the six *Enterobacteriaceae* for which COG data is available revealed a large proportion of CDSs that encode proteins involved in carbohydrate transport and metabolism, defense mechanisms, cell motility and chemotaxis and cell wall/envelope/membrane biogenesis occur in *P. ananatis* but have no homologues in the compared *Enterobacteriaceae* (Fig. 3.1; Table 3.2).

Carbohydrate transport and metabolism

On average, 210 CDSs (12.93 % of the average CDSs missing) encoding proteins involved in the transport and metabolism of carbohydrates by *P. ananatis* LMG20103 are absent from the flexible genomes of the six compared organisms. By contrast, an average 89 CDSs (4.53 % of the average enterobacterial flexible genome missing) are absent from LMG20103. This implies *P. ananatis* LMG20103 has a capacity to make use of a broader spectrum of carbohydrate resources or can do so more effectively than the compared enterobacteria. This is particularly prevalent in the LMG20103-*P. stewartii* DC283 comparison where 102 CDSs involved in the transport and catabolism of carbohydrates such as ribose, xylose, arabinose, galactose, cellobiose, glucose, fructose, lactose, rhamnose, mannose and galacturonate are present in LMG20103 but missing from *P. stewartii* DC283. Likewise, proteins for the transport

and degradation of the sugar alcohol sorbitol, are encoded on the genome of *P. ananatis* (PANA_0833 – PANA_0838) but missing from *P. stewartii* DC283. Sorbitol is the main substrate for carbohydrate transport in Rosaceous plants and serves as a major nutrient source for the pear and apple pathogen *E. amylovora* (Aldridge et al. 1997). *P. ananatis* has been found to exist as epiphyte on apples (Torres et al. 2005). The presence of these carbohydrate and transport mechanisms in *P. ananatis* may be linked to its capacity to infect a broad range of host plants and survive in a range of environments, while *P. stewartii* subsp. *stewartii* has only been isolated from a limited number of maize cultivars and as limited host range pathogen may preclude the requirement for these systems. The more distantly related pine endophyte *S. proteamaculans* 568, on the other hand, encodes 202 CDSs with a role in this function not found in LMG20103 suggesting different carbon substrates may be utilized by these two organisms.

Cell wall/membrane/envelope biogenesis

An average of 64 CDSs (4.21 % of the average CDSs missing in compared *Enterobacteriaceae*) involved in cell wall/membrane/envelope biogenesis are missing from the genomes of the six compared *Enterobacteriaceae*. By contrast an average 47 CDSs belonging to this COG class in the compared *Enterobacteriaceae* are missing from LMG20103 (Fig. 3.1; Table 3.2). The unusual LPS domains encoded by *P. ananatis* LMG20103 (Chapter 4) may in part account for this disparity, with 25 CDSs encoding proteins involved in LPS biosynthesis sharing no homologues in any of the six compared enterobacteria. Fewer discrepancies in this COG could be noted for the flexible genomes of *P. ananatis* and those of the plant associated bacteria, *Enterobacter* sp. 638, *S. proteamaculans* 568, *E. tasmaniensis* Et1/99 and *P. stewartii* DC283. There are 38 CDSs shared between them which are not found in *S. enterica* ATCC9150 and *E. coli* O157:H7 str. Sakai which encode proteins with a role in the production of exopolysaccharides that they likely require for persistence in the plant environment (Chapter 4).

Cell motility and chemotaxis

An average of 51 CDSs (3.29 % of the average CDSs missing in the compared *Enterobacteriaceae*) encoding proteins with a role in cell motility and chemotaxis in *P. ananatis* share no homologues in the compared *Enterobacteriaceae*. By contrast

only 26 CDSs (1.49% of the average CDSs missing from LMG20103) involved in this function are found in the compared enterobacteria but not in *P. ananatis* LMG20103 (Fig. 3.1; Table 3.2). This is once again evident in the *P. ananatis-P. stewartii* comparison, with 27 CDSs involved in this function missing from *P. stewartii* DC283, while it carries six motility and chemotaxis genes with no homologues in LMG20103. This can be attributed to the larger number of *P. ananatis* genes that encode methyl-accepting chemotaxis proteins (MCPs). A total of 21 CDSs encoding MCPs share no homologue in the other *Enterobacteriaceae*, while five are restricted to plant-associated bacteria and these may play a role in the chemotaxis-driven motility towards plant nutrients and away from toxic metabolites (Chapter 4).

Defense mechanisms

An average of 29 CDSs (1.92% of the flexible genome missing from the other *Enterobacteriaceae*) encoding proteins involved in defense share no homologues in the six compared enterobacteria. By contrast, between zero and six proteins involved in this function in the compared enterobacteria share no homologue in *P. ananatis* LMG20103 (Fig. 3.1; Table 3.2). These include CDSs encoding efflux pumps for the extrusion of antimicrobial substances and lactamases to break down antimicrobials (Chapter 4). As the LMG20103 genome was compared to those of both clinically important and plant-associated enterobacteria, this suggests that LMG20103 is well-adapted to protect itself from toxic metabolites and antimicrobials in the clinical environment, as well as those encountered in the plant environment.

Functions of CDSs in compared *Enterobacteriaceae* without homologues *P. ananatis* LMG20103

Comparison of the flexible genomes of *P. ananatis* LMG20103 and those of the six *Enterobacteriaceae* also revealed the presence of a large number of genes encoding proteins involved in intracellular trafficking and secretion as well as inorganic ion transport and metabolism in the compared *Enterobacteriaceae* that are absent from the LMG20103 genome (Fig. 3.1; Table 3.2).

Intracellular trafficking and secretion

An average 79 CDSs (4.51% of the average CDSs missing from LMG20103) that encode proteins involved in secretion are present in the compared enterobacteria but are missing from LMG20103 (Fig. 3.1; Table 3.2). This can be ascribed to the

presence of Type II secretion systems (T2SS) and Type III secretion systems (T3SS) which are notably absent from *P. ananatis* LMG20103 (Chapter 4). T2SSs are encoded on the genomes of all the compared organisms with the exception of *E. coli* O157:H7 str. Sakai. The only organism aside from LMG20103 which is missing a T3SS is the pine-endophyte *Enterobacter* sp. 638. Notably, analysis of the *P. stewartii* DC283 genome showed that it encodes four distinct T3SSs. The Type II and III secretion systems play a major role in the diseases caused by both plant and animal pathogens, which raises questions regarding the means by which *P. ananatis* causes disease in its plant hosts (Hueck et al. 1998; Sauvonnet et al. 2000).

Inorganic ion transport and metabolism

An average of 108 CDSs (5.89% of the average CDSs missing from LMG20103) encoding proteins for inorganic ion transport and metabolism are found in the six compared *Enterobacteriaceae* but are missing from *P. ananatis*. By contrast an average of 54 CDSs in *P. ananatis* are involved in this COG, but are missing from the compared enterobacteria. Transport and catabolic systems for most inorganic ions are shared between *P. ananatis* and its close phylogenetic relatives. However the notable differences are due to alternate means utilized for acquisition and metabolism of iron by *P. ananatis* and the compared enterobacteria. Additional systems for the uptake of iron complexed in enterobactin siderophores are present in *S. enterica* ATCC9150*, E. coli* O157:H7*, Enterobacter* sp. 638 and *S. proteamaculans* with ferrichrome-complexed iron acquisition systems present in *E. coli* O157:H7 and *Enterobacter* sp. 638. *E. tasmaniensis* has homologues to proteins involved in the uptake of iron in dicitrate complexes and biosynthesis of a homologue to the *D. dadantii* achromobactin siderophore. *E. coli* O157:H7, *E. tasmaniensis, S. proteamaculans* and *Enterobacter* sp. 638 also encode hemin uptake and utilization proteins. Hemins consist of oxidized Fe^{3+}-bound tetrapyrroles and this complex makes up a major prosthetic group in many mammalian proteins, representing the main means for iron storage in these hosts. Hemin uptake systems thus provide a means for scavenging iron from mammalian hosts (Tong and Guo, 2009). This capacity is nonexistent in *P. ananatis* LMG20103 which, like *S. enterica* ATCC9150, lacks homologues to hemin uptake and catabolism proteins. By contrast LMG20103 does encode a system for the acquisition of iron in enterochelin siderophores with homology to iron uptake systems in *Yersinia* and *Pectobacterium* species but lacking in the compared enterobacteria.

Comparison to plant-associated, animal-pathogenic and environmental bacteria

The *P. ananatis* CDSs were compared to those in the 1,099 genomes sequenced to date by BlastP against the NCBI protein database. This enabled identification of protein coding sequences which share homology only to those CDSs occurring in microorganisms occupying distinct ecological niches. Those CDSs with homologues only in the other sequenced *Pantoea* species or restricted to *P. ananatis* LMG20103 were also identified. In total, 24.0 % of all CDSs share homology to proteins encoded by organisms belonging to the distinct niche groups, while 76.0 % of LMG20103 CDSs share homologues in all of the groups (Fig. 3.2). Of those restricted to organisms occupying a specific niche, 127 CDSs (3.0 % of the total CDSs) share homologues only in animal associated bacteria (AAB), with a proportion of these found only in insect-associated bacteria (IAB). A number of CDSs were also found in animal-associated bacteria and either or both of the *Pantoea* species and these were also included in the AAB group. A much larger proportion of CDSs (727 CDSs - 17.2 % of total) share homology with proteins found only in plant-associated and plant-pathogenic bacteria. A further 159 CDSs (3.8 % of the total) share homologues only in the genomes of the other sequenced *Pantoea* species, *P. stewartii* DC283 or *Pantoea* sp. At9b or are restricted to the sequenced strain *P. ananatis* LMG20103 and are thus considered as genus and strain specific, respectively. The CDSs with homologues are restricted to plant-, animal-and insect-associated bacteria as well as those restricted to members of the genus *Pantoea* and to the strain LMG20103 were mapped onto a Genome Diagram (Fig. 3.3). This showed that genes with homologues restricted to bacteria occupying distinct ecological niches are scattered around the genome. Horizontally acquired islands (HAIs) were also predicted revealing the presence of thirty five islands on the LMG20103 genome (Table 3.3). These range in size from 2.3 kb (HAI32) to 37.5 kb (HAI21) and are interspersed around the genome (Fig. 3.4). Five of the HAIs are located adjacent to tRNAs, while 13 HAIs incorporate mobile genetic elements including transposases, insertion sequences and phage integrases. The HAIs were overlayed on the Genome Diagram depicting the genes which are restricted to plant-, animal- and insect-associated as well as the genus and strain specific genes, which revealed that there was significant overlap between the islands and the genes specific to bacteria occupying distinct ecological niches (Fig. 3.4). Nineteen of the islands (54.3% of the total HAIs) are only found in plant-associated bacteria, while three and two islands, respectively are only found in

animal-associated and insect-associated bacteria. Five of the HAIs encode phage-associated proteins indicating they may represent integrated phage elements. Another island, HAI34 encodes several putative plasmid-associated proteins which suggests a plasmid may have been integrated into the chromosome.

Pantoea ananatis CDSs unique to strain LMG20103

A total of 103 CDSs (2.43 % of total genome content) are restricted to *P. ananatis* LMG20103. From this total, 84 CDSs had no blast hit when compared to the NCBI protein database and are thus considered to be of unknown function. Predicted CDSs of unknown function are termed ORFans and are a common feature in every sequenced bacterial genome (Daubin and Ochman, 2004). Nineteen CDSs show limited homology to proteins in distantly related environmental bacteria or contain domains associated with known proteins, including PANA_0570-0573 which encode proteins putatively involved in transport and metabolism of an unknown carbohydrate or amino acid substrate.

P. ananatis LMG20103 CDSs unique to the genus Pantoea

56 CDSs (1.27 % of total genome content) share homologues in the genomes of *P. stewartii* and/or *Pantoea* sp. At9b. 25 CDSs (44.6 %) are found in both, while 14 CDSs (25.0 %) are restricted to *P. ananatis* and *Pantoea* sp. At9b and 17 CDSs (30.4 %) are found in *P. ananatis* and *P. stewartii* only. Most genus-specific CDSs are of unknown function (46 CDSs – 82.1 %). However, five CDSs encode proteins with conserved domains for function in cell wall biosynthesis. PANA_0385 and PANA_0387 encode two putative glycosyltransferases, while PANA_0386 encodes a potential polysaccharide flippase. These are located on the predicted 7.6 kb island HAI3 (Fig. 3.4; Table 3.3), which has a G+C content 8.92% below the genome average, suggesting horizontal acquisition from a distantly related organism, for which no genome sequence is available, prior to divergence of the *Pantoea* strains. The proteins encoded on the island may be involved in the biogenesis of a surface or extracellular polysaccharide specific to the *Pantoea* species and their distant relative. Similarly, PANA_1379 and PANA_1380 encode a putative glycosyltransferase and polysaccharide flippase, respectively.

106

P. ananatis LMG20103 CDSs restricted to Animal-associated bacteria (AAB)

A total of 52 CDSs (1.22 % of all *P. ananatis* LMG20103 CDSs) share homologues only in animal-associated bacteria including animal-pathogenic *Salmonella, Escherichia* and *Yersinia* strains. The majority of these, 29 CDSs, encode proteins of unknown function. However, several proteins involved in carbohydrate transport and metabolism share homologues only in AAB, including PANA_0540 which encodes a predicted hexuronate transporter (84% aa identity to the fish pathogen *E. tarda* ExuT – ETAE_1915) and PANA_1084, encoding an isomerase involved in the catabolism of fucose (75% aa identity to FucI in the human-pathogenic *K. pneumoniae* MGH78578 - KPN_00592). PANA_1520-1524 encode five proteins for the biogenesis of a type 1 fimbria with significant homology to the *S. enterica* subsp. *enterica* serovar Typhimurium LT2 Stb fimbria involved in attachment and intestinal persistence in mice (Fig. 3.5; Chapter 4; Weening et al. 2005). These genes are located on HAI8 suggesting recent horizontal acquisition of this fimbrial biogenesis locus from a *Salmonella*, *Proteus* or *Yersinia* species to which this fimbria is restricted (Weening et al. 2005). Furthermore, PANA_1089 encodes a putative non-fimbrial autotransporter adhesin with homology to AidA-I in *E. coli* O157:H7 str. Sakai (ECs0362 – 47 % aa identity) (Fig. 3.5). AidA-I is involved in diffuse adherence in cases of infantile diarrhea caused by enteropathogenic *E. coli* strains (Henderson et al. 2004). The presence of these adherence factors and homologues to carbohydrate transport and metabolism genes specific to animal-associated bacteria in *P. ananatis* LMG20103 suggest they may serve in its potential survival and persistence in the animal host. Furthermore, PANA_1760 shows homology to β-lactamases in clinical strains of *Enterobacter cloacae* (SFO-1 – BAA76882 – 65 % aa identity) and *Citrobacter sedlakii* (Sed-1 – AAK63223 – 64 % aa identity). These provide resistance to a broad spectrum of β-lactam antibiotics frequently used in the clinical environment including penicillins and cephalosporins (Matsumoto and Inoue, 1999; Petrella et al. 2001). This concurs with the occurrence of *P. ananatis* in the clinical environment (De Baere et al. 2004; Rezzonico et al. 2009). The absence of pathogenicity factors such as animal pathogen-specific toxins and haemolysins suggests LMG20103 may not cause human disease or if it does, manifests itself as opportunistic pathogen. Alternatively, novel pathogenicity factors may exist among the proteins of unknown function limited to the AAB group. Two predicted islands, HAI32 AND 33 (Table 3.3), are located within a predicted Type VI secretion system

(Chapter 5). The proteins encoded on these two islands share homology to putative Type VI effector proteins which are restricted in distribution to animal pathogenic bacteria and have been shown to play a role human and animal pathogenesis (Filloux et al. 2008; Chapter 5).

A further 54 *P. ananatis* CDSs (1.27 % of total on the LMG20103 genome) share homologues to proteins restricted to one or both of the *Pantoea* strains for which genome sequences are available and the animal-associated bacteria. These include CDSs encoding homologues to components of a glucosamine/glucuronide transport and catabolism pathway (PANA_3491–3499) and a mannose/fructose transport and catabolism locus (PANA_2342–2345). PANA_1618 shows 55 % aa identity to the transcriptional regulator HosA in the enteropathogenic *E. coli* UMN026 (ECUMN_3063). HosA plays a role in the regulation of swimming motility and cellular aggregation and is restricted to pathogenic *E. coli* strains (Ferrándiz et al. 2005). Although neither *P. stewartii* nor *Pantoea* sp. At9b have been isolated from the clinical environment, other *Pantoea* species generally associated with plants, including *P. dispersa* and *P. agglomerans* have been linked to human and animal disease (Cruz et al. 2007; Schmid et al. 2003). *Pantoea* sp. At9b was also isolated from fungus gardens produced by leaf cutter ants (Pinto-Tomás et al. 2009). Thus the proteins sharing homologues in the compared *Pantoea* species and animal-associated bacteria could play a role in persistence and competence in animal hosts.

Twenty one CDSs (0.5 % of total encoded on the LMG20103 genome) share homologues only in insect-associated bacteria. Ten CDSs share homology with proteins found only in the entomopathogenic *Photorhabdus luminescens* subsp. *laumondii* TT01. Nine of these CDSs (PANA_3280–3291) are housed on the 18.2 kb island HAI23. Interestingly, PANA_3820 and PANA_3284 encode proteins showing homology to nikkomycin biosynthetic proteins encoded by *P. luminescens* plu1441 and plu1874 respectively. This antibiotic is also produced by *Streptomyces* species and has acaricidal, fungicidal and insecticidal activities (Liao et al. 2010). PANA_3289 and PANA_3291 encode homologues to the *Streptomyces rubellomurinus* FbrC protein involved in production of an antimalarial compound (ABB90932 - 58 % aa identity) and the fosmidomycin Fom-1 protein of *Streptomyces fradiae* (ACG70831 – 49 % aa identity), respectively (Eliot et al. 2008; Woodyer et

al. 2006). *P. ananatis* LMG20103 may thus have horizontally acquired biosynthetic genes for the production of an insecticidal peptide. Another putative island, the 4 kb HAI6, carrying five genes, is also restricted in distribution to the insect-associated bacteria, but no known function could be ascribed to any of the encoded proteins (Fig. 3.4; Table 3.3). A 10.1 kb locus, PANA_2715-2726, encodes 10 proteins which show highest homology to the SefA-SefI fimbrial biogenesis locus in *Serratia entomophila* (48 % average aa identity) and the mannose-resistant fimbriae of *Photorhabdus luminescens* (42 % average aa identity). These fimbriae are involved in attachment and colonization of insect hosts (Meslet-Cladiere et al. 2004). Lower homology is also observed between some *P. ananatis* proteins encoded by this locus and other bacteria for which no information of insect-association is available, but homologues to all proteins are only found in the bacteria mentioned above, suggesting a role for the PANA_2717-2726 encoded fimbriae in attachment and colonization of insect hosts. This agrees with the observed association of some phytopathogenic *P. ananatis* strains with vectoring insects (Gitaitis et al. 2003).

P. ananatis LMG20103 CDSs restricted to Plant-associated bacteria (PAB)

A large number of *P. ananatis* CDSs (728 CDSs – 17.2 % of total genome content) share homology with proteins found only in plant-associated bacteria (PAB). Of these, 568 CDSs (13.4 % of total genome content) are found in both plant-associated *Enterobacteriaceae* (PAE) and non-*Enterobacteriaceae*, while a large number of CDSs (160 CDSs – 3.8 % of total) only share homologues in the NEPAB (non-enterobacterial plant-associated bacteria). The high number of PAB-restricted CDSs coincides with the frequent isolation and occurrence of *P. ananatis* in the plant environment. While 32.8% of PAB-specific proteins are of unknown function, a large proportion may play a role in efficient colonisation, nutrient utilization and persistence in or on the plant, while a subset may be involved in phytopathogenicity.

A significant number of the PAB-specific proteins are involved in transport and metabolism of sugars substrates including gluconate (PANA_0300-0303), ribose (PANA_0366-0370), myo-inositol (PANA_4050-4053), galactose (PANA_2419-2421), xylose and arabinose (PANA_0506-0510), maltose (PANA_2710-2714) and ascorbate (PANA_1860-1865). These may enable uptake and catabolism of plant-specific carbohydrates. A pathway for utilization of the rare sugar alcohol xylitol is
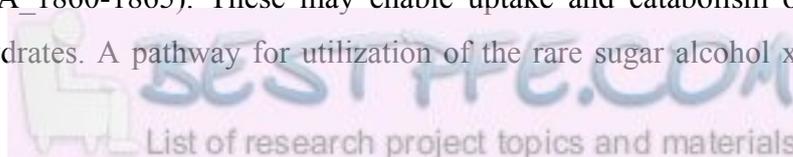
also present with a 7 kb cluster (PANA_1665-1671) encoding proteins with 99 % average aa identity to XytR-Xdh-XytBCDEF in a wheat-associated *P. ananatis* strain (Sakakibara and Saha, 2008). A 12.2 kb island, HAI25 (PANA_3341-3351) encodes homologues to components of the β-ketoadipate pathway for degradation of protocatechuate from phenolic compounds and plant cell wall lignin polymers abundant in the plant environment (Harwood and Parales, 1996; Table 3.3). Similarly, HA11 encompasses ten genes, PANA_1618-1628, which are likely involved in the degradation of phenolic compounds (Matsui et al. 2006). Acquisition of these islands could protect the bacterium against phenolic compounds which make up a major component of the plant defense arsenal against bacterial infection. PANA_4017-4020 encode enzymes involved in degradation of starch and share homologues only in NEPAB including *Agrobacterium* and *Rhizobium* species (76% average aa identity to *A. tumefaciens* Ct58 Atu4833-4836) (Ohnishi et al. 1992). Homologues to components of several amino acid transport and catabolism systems are likewise restricted to the PAB. PANA_3059-3063 and PANA_3319-3322 encode components of polar amino acid transport systems. Opines such as octopine are amino-sugar metabolites that are released from plant tumours induced by *A. tumefaciens* and these are used by this bacterium as nutrient source (Zanker et al. 1992). PANA_1272-1275 encode homologues to the octopine transport proteins, OccM, - P, -Q and –T (54% average aa identity – *A. tumefaciens* AAA50510/4/5/7). By carry octopine catabolic genes, *P. ananatis* may obtain nutrients from plants co-infected by tumour-inducing pathogens.

PAB-restricted inorganic ion uptake and metabolism systems are also encoded on the *P. ananatis* chromosome. A 12 kb cluster, PANA_1658- 1663, encodes homologues to components involved in the assimilation of nitrate/nitrite (Wu and Stewart, 1998) and these are co-located on the predicted island HAI12 with a putative Type VI secretion system. This nitrate and nitrite assimilation system may allow *P. ananatis* to use these substrates as sole nitrogen source during aerobic growth by conversion to ammonia (Wu and Stewart, 1998). PANA_1853-1858 may play a role in the release of phosphorus by degradation of insoluble phosphate in the plant niche, as they encode homologues to enzymes involved in the pyrroloquinoline quinone pathway (*K. pneumoniae* CAA41578-84: PqqABCDEF – 69 % average aa identity; Kim et al. 2003). Furthermore, PANA_0959 and PANA_3446 encode homologues to phytases

in *Raoultella terrigena* (67% aa identity to PhyK – CAE01322) and *Enterobacter* sp. 4 (82% aa identity to PhyA - AAC05186), respectively. Phytases degrade phytate, a common myo-insitol-phosphate storage compound in higher plants, thereby releasing the carbohydrate myo-inositol and phosphate (Kang et al. 2006). Several CDSs encode proteins putatively involved in acquisition of iron which is in short supply in the plant environment and thus represents a nutrient which is actively contested by both the plant host and phytopathogen (Wandersman and Delepelaire, 2004). A 6.8 kb cluster (PANA_2588-2593) encodes homologues to a siderophore receptor, ferric-dictrate sensor components FecI/FecR and a TonB-ExbBD complex for the uptake of siderophore-bound iron (Chapter 4; Fig. 3.6). PANA_3313-3315 encode homologues to a ferric (Fe^{3+})-iron transporter while PANA_1106 encodes a putative Fe^{3+}-hydroxamate transporter. The restricted distribution of homologues these CDSs in plant-associated bacteria suggests they enable *P. ananatis* LMG20103 to compete for iron in the plant.

Several PAB-restricted proteins involved in protection against plant defenses are also present in LMG20103. PANA_0118 and 0119 encode homologues to Ohr and the transcriptional regulator OhrR (*E. tasmaniensis* ETA_34010-34020 - 78% average aa identity) involved in resistance to organic hydroperoxides produced by plants in defense against infection (Mongkolsuk et al. 1998). These are incorporated in the predicted acquired island HAI1. PANA_1871 encodes a catalase restricted to PAB which provides resistance to hydrogen peroxide (ETA17240: 87 % aa identity). Several predicted multidrug efflux pumps are also restricted to the PAB, with PANA_1430-1433, PANA_2736-2738 and PANA_3394-3395 encoding efflux systems which may be involved in the extrusion of plant-produced antimicrobials and toxins. Furthermore, PANA_0239 and PANA_1808 encode β-lactamases that may specifically degrade plant antimicrobials. *P. ananatis* LMG20103 also encodes several enzymes that could enable it to outcompete other plant-associated bacteria and fungi for limited space and nutrients in the plant niche. PANA_1807 encodes a protein with 97 % aa identity to the albicidin detoxifying enzyme AlbD of *Pantoea dispersa* (AAB71813) which degrades the phytotoxin produced by the leaf scald pathogen *Xanthomonas albilineans* (Zhang and Birch, 1997). LMG20103 may also produce antimicrobials, with PANA_3811 encoding a homologue to phenazine antibiotic biosynthesis enzyme PhzF of *Pantoea* sp. At9b (Pat9bDRAFT_5898 - 70 %

aa identity; Blankenfeldt et al. 2004). However, further phenazine biosynthetic genes are absent. PANA_2584 and 2585 encode homologues to the microcin maturation protein TldD and microcin-processing peptidase TldE. Microcins are produced by *Enterobacteriaceae* and are active against a broad range of bacteria (Allali et al. 2002).

While *P. ananatis* LMG20103 CDSs encoding proteins for the biogenesis and functioning of flagella are found in most motile bacteria, homologues to seventeen CDSs encoding methyl accepting chemotaxis proteins (MCPs) are restricted to PAB. These may enable *P. ananatis* movement toward plant-specific attractants or away from toxic plant metabolites (Yao and Allen, 2006). PANA_2279-2281, located on the horizontally acquired island HA14, encode homologues of the enzymes involved in glycosylation of flagellin in the plant pathogen *Pseudomonas syringae* (VioATB: BAH58340-2 - 30% average aa identity), while PANA_2278 encodes a protein with 52 % aa identity to a sugar O-acetyltransferase in *Enterobacter* sp. 638 (Ent638_2518). These proteins attach amino-sugar units to the flagellum and contribute to swarming and swimming motility, biofilm formation and adhesion, with *P. syringae* glycosylation mutants reduced in virulence on tobacco (Nguyen et al. 2009). *P. ananatis* LMG20103 could thus have horizontally acquired the capacity to glycosylate its flagella to facilitate colonization of plant surfaces and tissues (Chapter 4). Colonisation may also be facilitated by the PAB-restricted Csu fimbria encoded on HAI24 (PANA_3330-3334). Survival on plant surfaces may be aided by production of carotenoids, pigments that provide bacterial cells with a means of protection against UV irradiation and damage from phototoxic agents (To et al. 1994). A 5.7 kb cluster (PANA_4158-4163) encodes enzymes involved in the biosynthesis of carotenoid pigment (CrtEXYIBZ – 99 % average aa identity to *P. ananatis* ATCC19321 BAA14124-14129) with homologues restricted to a number of plant-associated bacteria.

A number of *P. ananatis* CDSs encode homologues to PAB-restricted proteins which have been shown experimentally to be involved in plant pathogenesis and symptom development in other phytopathogens. The functions of these putative *P. ananatis* LMG20103 pathogenicity factors are discussed in detail in chapter 4. These include homologues to proteins involved in the degradation of plant cell including an endo-

1,4-β-xylanase (PANA_0376), two polygalacturonases (PANA_0324 and PANA_1095) and two putative pectin acetylesterases (PANA_0136 and PANA_2632) (Chapter 4). PANA_4011 encodes a homologue to the cyclic β1,2-glucan biosynthetic protein NdvB which produce a polysaccharide that may aid in resistance against localized and systemic defenses in the host plant (Rigano et al. 2007). The synthesis of this polysaccharide may circumvent the absence of a Type III secretion system in LMG20103 (Chapter 4). Restricted to plant-pathogenic strains of *P. ananatis* and *P. agglomerans*, *X. campestris* and *P. syringae*, the ice-nucleation protein encoded by PANA_0590 may induce wounds in frost-damaged plants allowing *P. ananatis* to gain access into the host (Lindow et al. 1982). The 21.4 kb horizontally acquired island HAI12, encompassing the CDSs PANA_1650-1657, encodes homologues to components of a Type VI secretion system restricted to plant-associated bacteria (Locus I; Chapter 5). Similarly, six LMG20103 CDSs within T6SS locus II share homologues only in plant-associated bacteria (Chapter 5). These include CDSs encoding components of the predicted regulatory mechanism for the secretion system (PANA_2353, PANA_2360 and PANA_2361) which appears to be specific to plant-associated bacteria, while the 3.1 kb island HAI16 carries three genes which encode putative secretion substrates for the secretion system. Furthermore, HAI17 carries two genes encoding homologues of the known Type VI secretion effector Hcp (Filloux et al. 2008; Chapter 5). This suggests that *P. ananatis* has horizontally acquired several pathogenicity determinants from other phytopathogenic bacteria which allow it to cause disease on its plant hosts.

CONCLUSIONS

Pantoea ananatis is a ubiquitous microorganism with an inherent capacity to survive and proliferate in the natural environment and an ability to form intimate relationships with plants, as well as insects and animals (Coutinho and Venter, 2009. This suggests it has evolved to proliferate in a multitude of ecological niches. By means of comparison of the genome of the *Eucalyptus*-pathogenic strain *Pantoea ananatis* LMG20103 to those of ecologically and phylogenetically related bacteria we attempted to identify the molecular basis behind its ecological success and its ability to cause disease in a broad range of plant hosts. We also looked at how it may further have evolved to persist in both invertebrate and vertebrate hosts and potentially cause disease in these hosts.

An initial comparison was performed between the genomes of *P. ananatis* LMG20103 and eight animal- and plant-associated relatives of the family *Enterobacteriaceae*. This revealed that while these eight compared bacteria share a significant core genome, an extensive flexible genome is evident. These differences may be correlated to discrepancies in the lifestyles of *P. ananatis* and the compared bacteria. LMG20103 possesses many protein coding genes involved in the uptake and metabolism of carbohydrates, cell wall components and defense mechanisms which are absent from the compared bacteria and that may ensure its survival in many different environments. This may also extend to persistence in different plants, with major differences in the above-mentioned functions between the broad host range pathogen *P. ananatis* LMG20103 and the maize-restricted pathogen *Pantoea stewartii* subsp. *stewartii* DC283. Conversely, within the flexible genome of the compared *Enterobacteriaceae* are genes encoding components of both Type II and Type III secretion systems, major pathogenicity determinants in both plant- and animal-pathogenic bacteria, which are absent from *P. ananatis* LMG20103. This suggests LMG20103 is an unusual pathogen which may possess alternative means to cause disease. One such means is the recently described Type VI secretion system which is discussed in detail in Chapter 5. In phytopathogenic bacteria the major function of the T3SS is the secretion of effector proteins which manipulate plant defense responses (Toth et al. 2006). Pathogens evolve by horizontal acquisition of effector proteins allowing them to effectively infect host plants. There is, however, growing evidence that plants co-evolve with the pathogen to recognise effectors and adapt their immune response by what is known as the guard hypothesis. Effectors then become avirulence (avr) proteins which are recognised by cognate resistance (R) proteins in the plant resulting in a Hypersensitive reaction (HR), an immune response restricting spread of the pathogen and disease development (Jones and Dangl, 2006). By this means the host range of the pathogen becomes restricted. By its lack of a T3SS, *P. ananatis* LMG20103 may have chosen to bow out of the molecular arms race with its host plants which may explain its broad host range (Chapter 4).

Comparison of the *P. ananatis* LMG20103 genome to all available genome sequences identified an extensive number of protein coding sequences with homologues in bacteria with restricted distribution in the plant, insect and animal niches. Many of

these reside on the 35 predicted horizontally acquired islands, indicative of horizontal acquisition of genes that would allow LMG20103 to survive and proliferate and potentially cause disease in these hosts. A significant proportion of the CDSs with homology restricted to plant-associated bacteria are involved in carbohydrate and amino acid transport and catabolism, indicative of their involvement in the utilisation of plant-specific nutrients. Other sequences specific to this group of bacteria including genes encoding proteins involved in survival in the plant environment and protection against plant defenses. This highlights the evolution of *P. ananatis* to become adept at survival and proliferation in the plant niche. In one Type VI secretion locus, two putative effector-coding islands with homology to plant-associated bacteria only are present, suggesting recent specialisation of this secretion system for a plausible role in plant pathogenesis. A number of proteins were shared only with insect-associated bacteria, including several encoding a fimbria and a putative insecticidal toxin. This agrees with the frequent isolation of *P. ananatis* from insects and the use by this phytopathogen of insects as vectors (Coutinho and Venter, 2009). Genes with homologues restricted to animal-associated bacteria encode a fimbria and a non-fimbrial adhesin which may enable *P. ananatis* to persist in a vertebrate host. This concurs with evidence of *P. ananatis* in the clinical environment (De Baere et al. 2004). Horizontally acquired islands carrying Type VI effector proteins restricted in distribution to animal-pathogenic bacteria suggesting recent adaptation of *P. ananatis* LMG20103 which may thus represent an emerging vertebrate pathogen.

The extent of horizontal acquisition is however likely underrepresented as prediction was based on parameters of recent acquisition, namely extensive G+C deviation and nucleotide usage deviation. Horizontal acquisition from more closely related organisms will thus not be obvious. A diagrammatic representation of the comparisons between *P. ananatis* LMG20103 and phylogenetically and ecologically related organisms (Fig. 3.4) suggests far more extensive occurrence of horizontal gene transfer within the LMG20103 genome. Regardless, comparative genomics revealed extensive mosaicism in the genome of *P. ananatis* LMG20103. Given the frequent isolation of this species from the natural environment and its intimate association with hosts belonging to both the plant and animal kingdoms, this provides evidence of a bacterium which is well-adapted for its ecological success and with the potential to be a cross-kingdom pathogen.

REFERENCES

Aldridge, P., Metzger, M. and Geider, K. 1997. Genetics of sorbitol metabolism in *Erwinia amylovora* and its influence on bacterial virulence. Mol. Gen. Genet. **256**: 611-619

Allali, N., Afif, H., Couturier, M. and Van Melderen, L. 2002. The highly conserved TldD and TldE proteins of *Escherichia coli* are involved in microcin B17 processing and in CcdA degradation. J. Bacteriol. **184**: 3224-3231

Binnewies, T. T., Motro, Y., Hallin, P. F., Lund, O., Dunn, D. et al. 2006. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. Funct. Integr. Genomics **6**: 165-185

Blankenfeldt, W., Kuzin, A. P., Skarina, T., Korniyenko, Y., Tong, L. et al. 2004. Structure and function of the phenazine biosynthetic protein PhzF from *Pseudomonas fluorescens*. Proc. Natl. Acad. Sci. USA **101**: 16431-16436

Coutinho, T. A. and Venter, S. N. 2009. *Pantoea ananatis*: an unconventional plant pathogen. Mol. Plant Pathol. **10**: 325-335

Cruz, A. T., Cazacu, A. C. and Allen, C. H. 2007. *Pantoea agglomerans*, a plant pathogen causing human disease. J. Clin. Microbiol. **45**(6): 1989-1992

Daubin, V. and Ochman, H. 2004. Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. Genome Res. **14**: 1036-1042

De Baere, T., Verhelst, R., Labit, C., Verschraegen, G., Wauters, G. et al. 2004. Bacteremic infection with *Pantoea ananatis*. J. Clin. Microbiol. **42**: 4393-4395

Eliot, A. C., Griffin, B. M., Thomas, P. M., Johannes, T. W., Kelleher, N. L. et al. 2008. Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of the potent antimalarial compound FR900098. Chem. Biol. **15**: 765-770

Ercolini, D., Russo, F., Torrieri, E., Masi, P. and Villani, F. 2006. Changes in the spoilage-related microbiota of beef during refrigerated storage under different packaging conditions. Appl. Environ. Microbiol. **72**: 4663-4671

Ferrándiz, M.-J., Bishop, K., Williams, P. and Withers, H. 2005. HosA, a member of the SlyA family, regulates motility in enteropathogenic *Escherichia coli*. Infect. Immun. **73**: 1684-1694

Filloux, A., Hachani, A. and Bleves, S. 2008. The bacterial Type VI secretion machine: yet another player for protein transport across membranes. Microbiology **154**: 1570-1583

Gitaitis, R. D., Walcott, R. R., Wells, M. L., Diaz Perez, J. C. and Sanders, F. H. 2003. Transmission of *Pantoea ananatis*, causal agent of center rot of onion by tobacco thrips, *Frankliniella fusca*. Plant Dis. **87**: 675-678

Hall, T. A. 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**: 95-98

Harwood, C. S. and Parales, R. E. 1996. The β-ketoadipate pathway and the biology of self-identity. Annu. Rev. Microbiol. **50**: 553-590

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K. et al. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8**: 11-22

Henderson, I. R., Navarro-garcia, F., Desvaux, M., Fernandez, R. C. and Ala'Aldeen, D. 2004. Type V protein secretion pathway: the autotransporter story. Microbiol. Mol. Biol. Rev. **68**: 692-744

Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol. Mol. Biol. Rev. **62**: 379-433

Jones, J. D. G. and Dangl, J. L. 2006. The plant immune system. Nature **444**: 323-329

Kang, S. H., Cho, K. K., Bok, J. D., Kim, S. C., Cho, J. S. et al. 2006. Cloning, sequencing and characterization of a novel phosphatase gene, *phoI*, from soil bacterium *Enterobacter* sp. 4. Curr. Microbiol. **52**: 243-248

Kim, C. H., Han, S. H., Kim, K. Y., Cho, B. H., Kim, Y. H. et al. 2003. Cloning and expression of pyrroloquinoline quinone (PQQ) genes from a phosphate-solubilizing bacterium *Enterobacter intermedium*. Curr. Microbiol. **47**: 457- 461

Kube, M., Migdoll, A. M., Kuhl, H., Beck, A., Reinhardt, R. et al. 2008. The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. Environ. Microbiol. **10**: 2211-2222

Lagesen, K., Ussery, D. W. and Wassenaar, T. M. 2010. Genome update: the thousandth genome – a cautionary tale. Microbiology **156**: 603-608

Liao, G., Li, J., Li, L., Yang, H., Tian, Y. et al. 2010. Cloning, reassembling and integration of the entire nikkomycin biosynthetic gene cluster into *Streptomyces ansochromogenes* lead to an improved nikkomycin production. Microb. Cell Fact. **9**: 1-7

Lindow, S. E., Arny, D. C. and Upper, C. D. 1982. Bacterial ice nucleation: a factor in frost injury to plants. Plant Physiol.**70**: 1084-1089

Matsui, T., Yoshida, T., Hayashi, T. and Nagasawa, T. 2006. Purification, characterization, and gene cloning of 4-hydroxybenzoate decarboxylase of *Enterobacter cloacae* P240. Arch. Microbiol. **186**: 21-29

Matsumoto, Y. and Inoue, M. 1999. Characterization of SFO-1, a plasmid-mediated inducible class A β-Lactamase from *Enterobacter cloacae*. Antimicrob. Agents Chemother. **43**: 307-313

McClelland, M., Sanderson, K. E., Clifton, S. W., Latreille, P., Porwollik, S. et al. 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. Nature Genet. **36**: 1268-1274

Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R. 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. **15**: 589-594

Meslet-Cladiere, L. M., Pimenta, A., Duchaud, E., Holland, I. B. and Blight, M. A. 2004. In vivo expression of the mannose-resistant fimbriae of *Photorhabdus temperata* K122 during insect infection. J. Bacteriol. **186**: 611-622

Mongkolsuk, S., Praituan, W., Loprasert, S., Fuangthong, M. and Chamnongpol, S. 1998. Identification and characterization of a new organic hydroperoxide resistance (*ohr*) gene with a novel pattern of oxidative stress regulation from *Xanthomonas campestris* pv. phaseoli. J. Bacteriol. **180**: 2636-2643

Moreno-Hagelsieb, G. and Latimer, K. 2007. Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics **24**: 319-324

Nguyen, L. C., Yamamoto, M., Ohnishi-Kageyama, M., Andi, S., Taguchi, F. et al. 2009. Genetic analysis of genes involved in synthesis of modified 4-amino-4,6-dideoxyglucose in flagellin of *Pseudomonas syringae* pv. tabaci. Mol. Genet. Genomics **282**: 595-605

Ohnishi, H., Kitamura, H., Minowa, T., Sakai, H. and Ohta, T. 1992. Molecular cloning of a glucoamylase gene from a thermophilic *Clostridium* and kinetics of the cloned enzyme. Eur. J. Biochem. **207**: 413-418

Perumal, D., Lim, C. S., Chow, V. T. K., Sakharkar, K. R. and Sakharkar, M. K. 2008. A combined computational-experimental analyses of selected metabolic enzymes in *Pseudomonas* species. Int. J. Biol. Sci. **4**: 309-317

Petrella, S., Clermont, D., Casin, I., Jarlier, V. and Sougakoff, W. 2001. Novel class A β-Lactamase Sed-1 from *Citrobacter sedlakii*: genetic diversity of β-lactamases within the *Citrobacter* genus. Antimicrob. Agents Chemother. **45**: 2287-2298

Pinto-Tomás, A., Anderson, M. A., Suen, G., Stevenson, D. M., Chu, F. S. T. et al. 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. Science **326**: 1120-1123

Pritchard, L., White, J. A., Birch, P. R. J. and Toth, I. K. 2006. GenomeDiagram: a python package for the visualization of large-scale genomic data. Bioinformatics **22**: 616-617

Reva, O. N. and Tümmler, B. 2005. Differentiation of regions with atypical oligonucleotide composition in bacterial genomes. BMC Bioinformatics **6**: 1-12

119

Rezzonico, F., Smits, T. H. M., Montesinos, E., Frey, J. E. and Duffy, B. 2009. Genotypic comparison of *Pantoea agglomerans* plant and clinical strains. BMC Microbiol. **9**: 204-223

Rigano, L. A., Payette, C., Brouillard, G., Marano, M. R., Abramowicz, L. et al. 2007. Bacterial cyclic β-(1,2)-glucan acts in systemic suppression of plant immune responses. Plant Cell **19**: 2077-2089

Sakakibara, Y. and Saha, B. C. 2008. Isolation of an operon involved in xylitol metabolism from a xylitol-utilizing *Pantoea ananati*s mutant. J. Biosci. Bioeng. **106**: 337-344

Sauvonnet, N., Vignon, G., Pugsley, A. P. and Gounon, P. 2000. Pilus formation and protein secretion by the same machinery in *Escherichia coli*. EMBO J. **19**: 2221-2228

Schmid, H., Weber, C., Bogner, J. R. and Schubert, S. 2003. Isolation of a *Pantoea dispersa*-like strain from a 71-year-old woman with acute myeloid leukemia and multiple myeloma. Infection **31**(1): 66-67

Schmidt, H. and Hensel, M. 2004. Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol. Rev. **17**: 14-56

Sebaihia, M., Quail, M. A., Perna, N. T., Glasner, J. D., Declerck, G. A. et al. 2010. Complete genome sequence of the plant pathogen *Erwinia amylovora* strain ATCC 49946. J. Bacteriology **192**: 2020-2021

Serrano, F. B. 1928. Bacterial fruitlet brown-rot of pineapple in the Philippines. Philip. J. Sci. **36**: 271-300

Smits, T. H. M., Rezzonico, F., Kamber, T., Blom, J., Goesmann, A. et al. 2010. Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia* spp. Mol. Plant-Microbe Interact. **23**: 384-393

Spiers, A. J., Buckling, A. and Rainey, P. B. 2000. The causes of *Pseudomonas* diversity. Microbiology **146**: 2345-2350

Taghavi, S., Garafola, C., Newman, L., Hoffman, A., Weyens, N. et al. 2009. Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. Appl. Environ. Microbiol. **75**: 748-757

To, K.-Y., Lee, L.-Y., Hung, C.-H., Chen, C.-L., Chang, Y.-S. et al. 1994. Analysis of the gene cluster encoding carotenoid biosynthesis in *Erwinia herbicola* Eho13. Microbiology **140**: 331-339

Tong, Y. and Guo, M. 2009. Bacterial heme-transport proteins and their heme-coordination modes. Arch. Biochem. Biophys. **481**: 1-15

Torres, R., Teixidó, N., Usall, J., Abadias, M. and Viñas, I. 2005. Post-harvest control of *Penicillium expansum* on pome fruits by the bacterium *Pantoea ananatis* CPA-3. J. Hort. Sci. Biotech. **80**: 75-81

Toth, I. K., Pritchard, L. and Birch, P. R. J. 2006. Comparative genomics reveals what makes an enterobacterial plant pathogen. Annu. Rev. Phytopathol. **44**: 305-336

van Baarlen, P., van Belkum, A., Summerbell, R. C., Crous, P. W. and Thomma, B. P. H. J. 2007. Molecular mechanisms of pathogenicity: how do pathogenic microorganisms develop cross-kingdom host jumps? FEMS Microbiol. Rev. **31**: 239-277

Vernikos, G. S. and Parkhill, J. 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics **22**: 2196-2203

Vinatzer, B. A. and Yan, S. 2008. Mining the genomes of plant pathogenic bacteria: how not to drown in gigabases of sequence. Mol. Plant Pathol. **9**: 105-118

Wandersman, C. and Delepelaire, P. 2004. Bacterial iron sources: from siderophores to hemophores. Annu. Rev. Microbiol. **58**: 611-647

Weening, E. H., Barker, J. D., Laarakker, M. C., Humphries, A. D., Tsolis, R. M. et al. 2005. The *Salmonella enterica* Serotype Typhimurium *lpf*, *bcf*, *stb*, *stc*, *std*, and *sth* fimbrial operons are required for intestinal persistence in mice. Infect. Immun. **73**: 3358-3366

Woodyer, R. D., Shao, Z., Thomas, P. M., Kelleher, N. L., Blodgett, J. A. V. et al. 2006. Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. Chem. Biol. **13**: 1171-1182

Wu, Q. and Stewart, V. 1998. NasFED proteins mediate assimilatory nitrate and nitrite transport in *Klebsiella oxytoca* (*pneumoniae*) M5a1. J. Bacteriol. **180**: 1311-1322

Yao, J. and Allen, C. 2006. Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*. J. Bacteriol. **188**: 3697-3708

Zanker, H., Von Lintig, J. and Schröder, J. 1992. Opine transport genes in the octopine (*occ*) and nopaline (*noc*) catabolic regions in Ti plasmids of *Agrobacterium tumefaciens*. J. Bacteriol. **174**: 841-849

Zhang, L. and Birch, R. G. 1997. The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans* to sugarcane. Proc. Natl. Acad. Sci. USA **94**: 9984-9989

Table 3.1: Comparison of the *P. ananatis* LMG20103 genome against eight closely related *Enterobacteriaceae*. A neighbour-joining tree based on the clustalW alignment of the Gyrase B (GyrB) amino acid sequences with bootstrap support (n=1,000) is shown. The genome sizes, number of protein coding sequences (CDSs) on the genome, the proportion of CDSs in the compared enterobacteria with homologues in *P. ananatis* and proportions of *P. ananatis* CDSs with homologues in the compared enterobacteria are presented as well as the percentage amino acid identity for homologues are indicated. The different colours are based on ecology of the organism, with pink representing animal-pathogenic bacteria, orange for non-pathogenic plant-associated *Enterobacteriaceae* and green for the plant-pathogenic *Enterobacteriaceae*.

	Genome size	Number of CDSs	% *P. ananatis* CDSs with homologues in compared organism	% CDSs with homologues in *P. ananatis*	Average % CDS amino acid identity
Pantoea ananatis LMG20103	4.69 Mb	4,237	100%	100%	100%
P. stewartii subsp. stewartii DC283	5.00 Mb	5,090	75.7%	63.0%	91.5%
Pantoea sp. At-9b	6.26 Mb	5,969	78.0%	55.4%	81.8%
Erwinia amylovora Ea273	3.91 Mb	3,488	58.3%	70.8%	76.2%
Erwinia tasmaniensis Et1/99	4.07 Mb	3,622	61.4%	71.9%	76.2%
Escherichia coli O157:H7 str. Sakai	5.59 Mb	5,318	58.9%	46.9%	73.0%
S. enterica Paratyphi ATCC9150	4.59 Mb	4,403	57.5%	55.3%	72.9%
Enterobacter sp. 638	4.68 Mb	4,240	64.2%	64.2%	72.5%
Serratia proteamaculans 568	5.50 Mb	4,942	66.0%	56.6%	71.9%

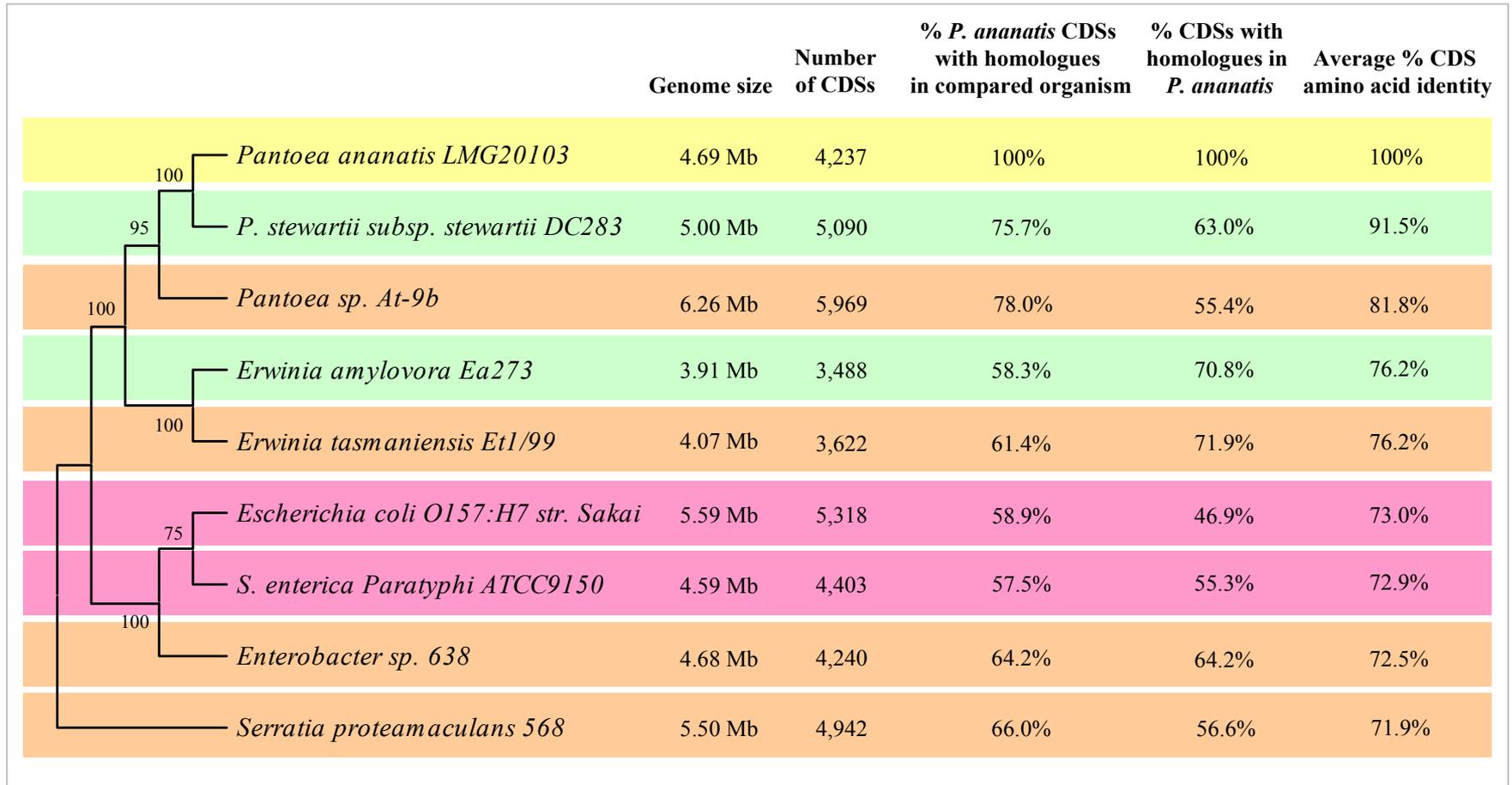Bootstrap values: 100, 95, 100, 100, 75, 100

FIGURES AND TABLES

Table 3.2: Proportions of CDSs belonging to the different COGs which are pr... ompared bacteria (*P. stewartii* subsp. *stewartii* DC283, *S. protemaculans* 568, *Enterobacter* sp. 638, *S. enterica* subsp. *enterica* Paratyphi ATCC9150 and *E. coli* O157:H7 str. Sakai - **In**) but not in *P. ananatis* LMG20103 and vice versa (**Not in**). Average proportions for those missing from the five compared bacteria, and those in the five compared enterobacteria missing from LMG20103 are also given.

	In *P. stewartii* DC283	Not in *P. stewartii* DC283	In *S. protemaculans* 568	Not in *S. protemaculans* 568	In *Enterobacter* sp. 638	Not in *Enterobacter* sp. 638	In *E. tasmaniensis* Et1/99	Not in *E. tasmaniensis* Et1/99	In *S. enterica* ATCC9150	Not in *S. enterica* ATCC9150	In *E. coli* O157:H7 Sakai	Not in *E. coli* O157:H7 Sakai	In other organisms average	Not in other organisms average
Information Processing and Storage	7.71	10.02	13.32	11.87	11.49	11.95	10.08	12.67	10.58	12.38	11.55	11.94	10.79	12.16
Translation, ribosomal structure and biogenesis	0.11	1.85	1.21	1.04	1.38	0.99	1.63	1.16	1.17	1.05	1.00	1.15	1.08	1.21
RNA processing and modification	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
Transcription	1.60	7.39	10.07	8.95	8.31	9.50	7.24	9.91	7.60	9.77	5.52	9.30	6.75	9.14
DNA replication, recombination and repair	6.00	0.78	2.03	1.87	1.80	1.45	1.21	1.59	1.81	1.55	5.00	1.49	2.98	1.46
Cellular Processes	8.83	12.45	12.70	14.64	14.34	14.06	20.01	12.61	15.03	14.49	13.55	14.47	14.08	13.79
Chromatin structure and dynamics	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Defense mechanisms	0.37	2.14	0.00	1.73	0.00	1.72	0.00	1.77	0.09	2.17	0.11	2.01	0.10	1.92
Cell division and chromosome partitioning	0.00	0.10	0.20	0.07	0.52	0.07	0.28	0.06	0.17	0.06	0.20	0.11	0.23	0.08
Signal transduction mechanisms	0.05	1.46	3.41	2.64	5.22	2.77	5.04	3.00	3.93	3.16	3.69	3.21	3.56	2.71
Cell wall/envelope/membrane biogenesis	1.33	4.18	2.79	4.79	2.80	4.03	1.56	3.55	3.11	4.28	2.83	4.42	2.40	4.21
Cell motility and chemotaxis	0.32	2.63	1.21	3.89	1.04	3.56	3.26	2.75	1.60	3.39	1.52	3.50	1.49	3.29
Extracellular structures	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.03	0.00
Intracellular trafficking, secretion, and vesicular transport	6.75	0.58	3.12	0.28	2.71	0.40	7.81	0.24	3.41	0.28	3.29	0.23	4.51	0.33
Posttranslational modification, protein turnover, chaperones	0.00	1.36	1.94	1.25	1.99	1.52	2.06	1.22	2.72	1.17	1.80	0.98	1.75	1.25
Metabolism	1.54	27.92	38.98	29.49	36.04	31.55	25.62	34.15	35.08	34.98	26.62	32.15	27.31	31.71
Energy production and conversion	0.00	1.75	5.15	2.71	4.75	2.57	2.34	2.88	7.30	3.00	5.09	2.64	4.11	2.59
Carbohydrate transport and metabolism	0.74	9.92	6.63	12.56	6.41	12.34	2.20	15.18	6.61	12.66	4.60	11.94	4.53	12.43
Amino acid transport and metabolism	0.37	8.27	10.83	7.22	9.45	8.12	7.88	6.67	8.42	9.05	6.23	8.78	7.20	8.02
Inorganic ion transport and metabolism	0.21	2.82	7.41	2.01	8.59	3.23	7.24	3.92	6.22	4.78	5.66	3.85	5.89	3.43
Lipid transport and metabolism	0.05	2.14	2.79	1.94	1.80	1.65	2.27	2.57	1.34	2.05	1.43	1.72	1.61	2.01
Nucleotide transport and metabolism	0.00	0.19	0.79	0.21	0.24	0.79	0.07	0.24	0.60	0.89	0.51	0.69	0.37	0.50
Coenzyme transport and metabolism	0.00	1.75	2.00	1.39	2.23	1.52	1.42	1.10	2.51	1.11	1.49	1.26	1.61	1.36
Secondary metabolites biosynthesis, transport and catabolism	0.16	1.07	3.38	1.46	2.56	1.32	2.20	1.59	2.07	1.44	1.60	1.26	2.00	1.36
Poorly Characterized proteins	81.92	49.61	35.01	44.00	38.13	42.44	44.29	40.58	39.31	38.15	48.28	41.45	47.82	42.70
General function prediction only	20.73	10.21	12.83	10.62	11.44	9.97	8.16	10.89	8.98	9.38	9.58	10.73	11.95	10.30
Proteins of unknown function	61.19	39.40	22.18	33.38	26.69	32.48	36.12	29.68	30.32	28.76	38.71	30.71	30.80	32.40

Table 3.3: Predicted horizontally acquired islands in *P. ananatis* LMG20103. The start and end locations are given as well as the size of the islands. The G+C content and deviation from the average (53.69 %) are shown. Those islands incorporated into the genome adjacent to tRNAs are indicated as are those islands which encode predicted transposases, insertion elements as well as phage integrases. By BlastP the restricted occurrence of the proteins encoded on the islands in the plant-associated bacteria (PAB), insect-associated bacteria (IAB), animal-associated bacteria (AAB) and those specific to *Pantoea* spp. were determined and the putative functions encoded on the islands are shown.

Island	Origin	Start	End	Size (kb)	G+C %	Difference to average	Number of CDSs	tRNA site	Mobile element	Putative Function
HAI1	PAB	122088	129410	7,3	46.63	-7.06	8	+		Tryptophan catabolism Defense against H_2O_2
HAI2	PAB	337294	340426	3,1	46.15	-7.54	3	-	+	EPS/LPS
HAI3	Pantoea	458783	466352	7,6	44.77	-8.92	9	-	+	Cell wall biosynthesis
HAI4	PAB	602683	624797	22,1	39.41	-14.28	19	+	+	Unknown function
HAI5	-	670474	674671	4,2	39.42	-14.27	6	-	-	Unknown function
HAI6	IAB	1060258	1064283	4,0	42.52	-11.17	5	-	-	Unknown function
HAI7	Phage	1212680	1225151	12,5	43.49	-10.2	14	+	+	O-antigen conversion
HAI8	AAB	1706875	1715428	8,6	47.72	-5.97	8	-	-	Fimbrial biogenesis Metabolic proteins
HAI9	PAB	1768593	1785051	16,5	46.65	-7.04	13	-	-	Siderophore receptor
HAI10	PAB	1791247	1800208	9.0	46.16	-7.53	10	-	-	Lipid metabolism
HAI11	PAB	1813528	1823948	10,4	47.32	-6.37	11	-	-	Hydroxybenzoate and Carbohydrate catabolism T6SS
HAI12	PAB	1845969	1867340	21,4	58.99	5.3	16	-	-	Nitrate/nitrite catabolism
HAI13	-	1999494	2010738	11,2	43.15	-10.54	9	-	-	Fimbrial biogenesis
HAI14	PAB	2526676	2547978	21,3	47.49	-6.2	16	-	-	LPS O-antigen biosynthesis
HAI15	PAB	2600778	2604874	4,1	38.13	-15.56	2	-	-	Unknown function
HAI16	PAB	2625643	2628732	3,1	40.36	-13.33	3	-	-	T6SS effectors
HAI17	PAB	2719987	2722632	2,6	42.86	-10.83	3	-	-	T6SS effectors
HAI18	PAB	2755606	2771508	15,9	38.68	-15.01	12	-	-	O-antigen biosynthesis
HAI19	PAB	3189218	3192808	3,6	38.01	-15.68	4	-	+	Unknown function
HAI20	Phage	3268343	3294216	25,9	46.69	-7	31	-	+	Integrated phage element
HAI21	Phage	3507399	3544855	37,5	54.68	0.99	16	-	+	Integrated phage element
HAI22	PAB	3624534	3628681	4,1	47.49	-6.2	6	-	-	Unknown function
HAI23	IAB	3648158	3666319	18,2	39.33	-14.36	15	+	+	Insecticidal peptide
HAI24	PAB	3709707	3714969	5,3	58.73	5.04	5	-	-	Fimbrial biogenesis
HAI25	PAB	3721666	3733850	12,2	61.04	7.35	13	-	-	Lignin phenol degradation
HAI26	Phage	3779521	3801065	21,5	52.95	-0.74	31	+	-	Integrated phage
HAI27	PAB	4215435	4222462	7,0	47.38	-6.31	5	-	-	Unknown function
HAI28	-	4297808	4308959	11,2	48.39	-5.3	9	-	-	LPS biosynthesis
HAI29	Phage	4323023	4341320	18,3	47.92	-5.77	21	-	+	Integrated phage element
HAI30	PAB	4439015	4443791	4,8	47.58	-6.11	3	-	+	Manganese/Iron transport
HAI31	PAB	4497139	4503460	6,3	41.57	-12.12	5	-	+	Unknown function
HAI32	AAB	4575926	4578154	2,2	40.96	-12.73	2	-	-	T6SS effector - Pyocin S
HAI33	AAB	4583292	4588707	5,4	42.32	-11.37	4	-	-	T6SS effectors
HAI34	Plasmid	4632858	4646227	13,4	45.33	-8.36	14	-	+	Unknown function
HAI35	-	4685773	4690298	4,5	46.4	-7.29	2	-	+	Siderophore receptor

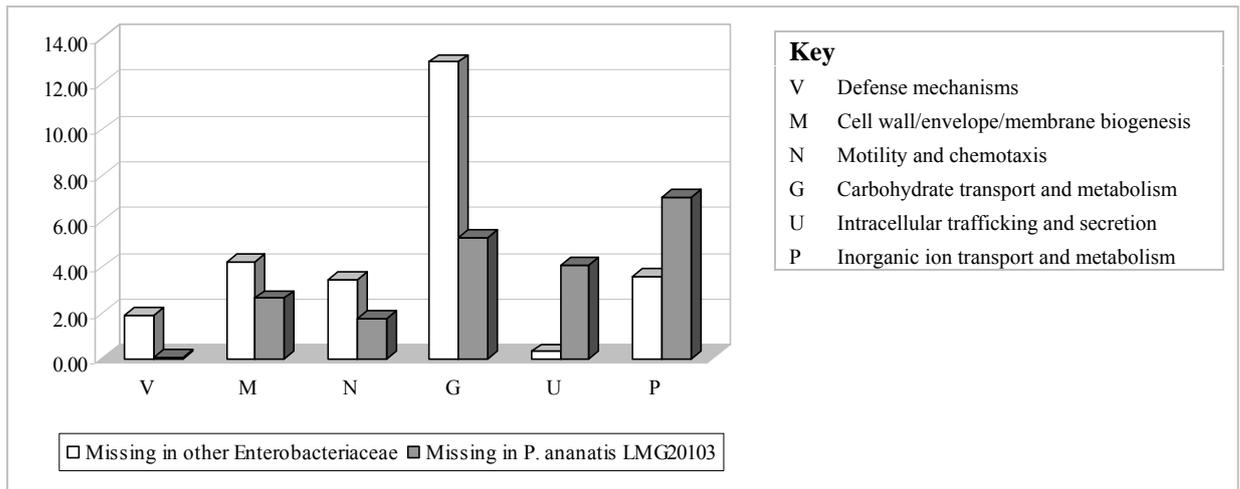| Key	
V	Defense mechanisms
M	Cell wall/envelope/membrane biogenesis
N	Motility and chemotaxis
G	Carbohydrate transport and metabolism
U	Intracellular trafficking and secretion
P	Inorganic ion transport and metabolism

Figure 3.1: Relative percentages of CDSs belonging to several specific COG functional groups present in *P. ananatis* LMG20103 and absent in the other compared *Enterobacteriaceae* and those missing from *P. ananatis* but present in the other *Enterobacteriaceae*.
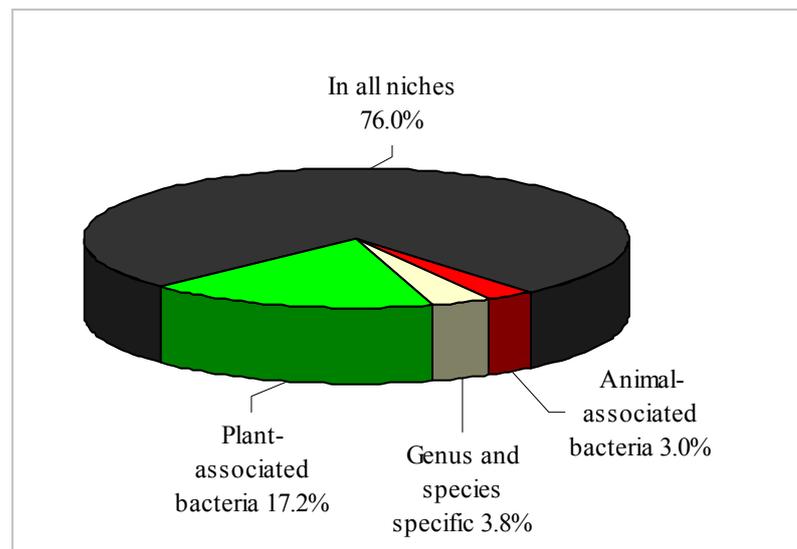
Figure 3.2: Proportions of *P. ananatis* LMG20103 CDSs with homologues only among plant-associated bacteria (PAB), animal-associated bacteria (AAB) and those for which homology is restricted to the genus or strain as well as those which are found in representatives of all groups.

126
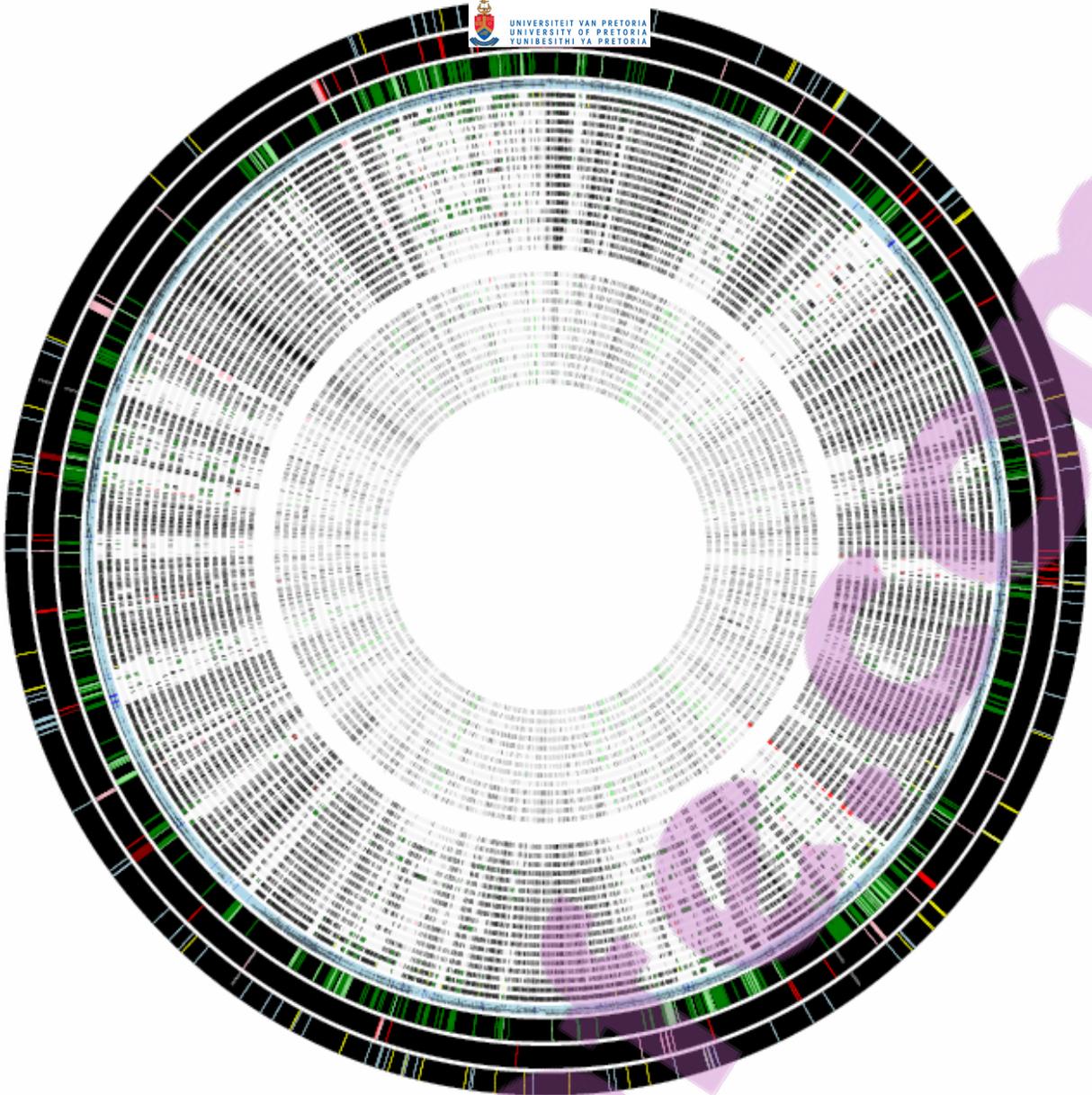
Figure 3.3: Genome comparative diagram constructed with Genome diagram (Pritchard et al. 2006). The outer tracks show those protein coding sequences restricted to the plant-associated bacteria (*Enterobacteriaceae* and non-*Enterobacteriaceae* - dark green; non-*Enterobacteriaceae* only - light green), animal-associated bacteria (Animal-associated bacteria only – red; animal-associated bacteria and *Pantoea* spp. – pink; insect-associated bacteria only – dark red) and those with homologues restricted to *Pantoea* spp. only – yellow, while those restricted to *P. ananatis* LMG20103 are shown in light blue. 31 groups of bacteria are included in the diagram, arranged on the basis of phylogeny from outside to inside: *P. stewartii* subsp. *stewartii* DC283, *Pantoea* sp. At9b, *Erwinia* spp. (*amylovora* Ea273*, tasmaniensis* Et1/99, *pyrifoliae* Ep1/96), *E. sakazakii* ATCC BAA-894, *Enterobacter* sp. 638, *S. enterica* pvs. (Paratyphi ATCC9150, Typhimurium LT2), *Citrobacter* spp. (*koseri* ATCCBAA-895, *rodentium* ICC168, *youngae* ATCC29220), *K. pneumoniae* 342, *K. pneumoniae* MGH78578, *E. coli* K-12, *E. coli* O157:H7 str. Sakai, *Pectobacterium* spp. (*atrosepticum* SCRI1043, *carotovorum* PC1, *wasabiae* WPP163), *Dickeya* spp. (*dadantii* Ech568, *zeae* ECH1519), *Y. pestis* CO92, *S. proteamaculans* 568, *Photorhabdus luminescens* TT01, *P. stuartii* ATCC25827, *P. mirabilis* ATCC 29906, *P. aeruginosa* PA01, *Pseudomonas* spp. (*entomophila* L48, *fluorescens* Pf-5, *putida* W619), *P. syringae* pvs. (tabaci ATCC11528, oryzae str. 1_6, phaseolicola 1448A, syringae B728a, tomato DC3000), *X. fastidiosa* M23, *Xanthomonas* spp. (*axonopodis* pv. *citri* 306, *campestris* ATCC33913, 85-10, *oryzae* KACC10331), *Acidovorax* spp. (*avenae* ATCC19860, *citrulli* AAC00-1), *R. solanacearum* GM1000, *C. taiwanensis* LMG19424, *Burkholderia* spp. (*graminis* C4D1M, *phymatum* STM815), *S. meliloti* 1021, *Agrobacterium* spp. *tumefaciens* C58, *vitis* S4*)* and *Rhizobium* spp. (*etli* CFN42, *leguminosarum* 3841)
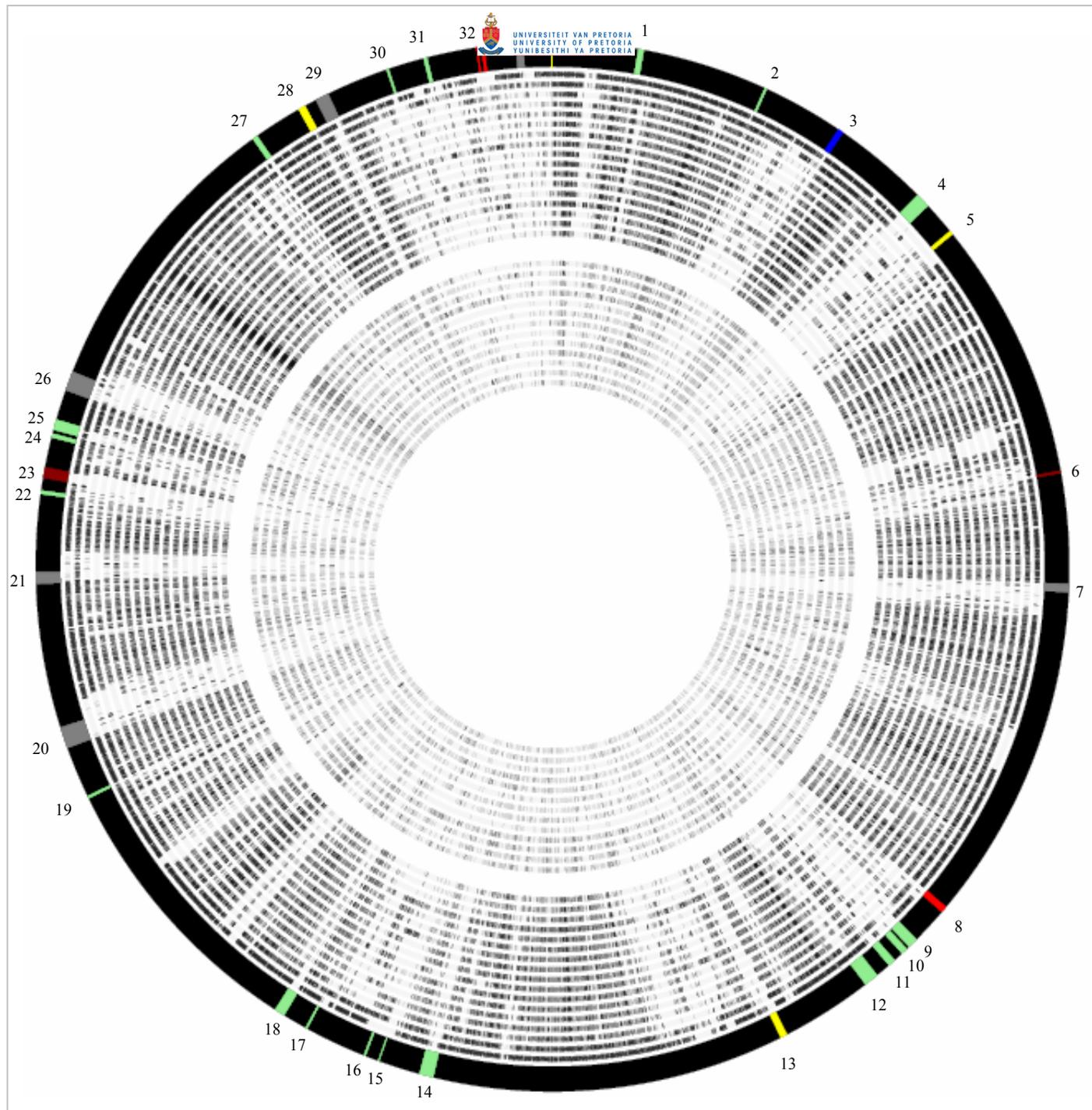
Figure 3.4: Genome diagram showing the locations of predicted horizontally acquired islands in the genome of *P. ananatis* LMG20103. Islands are shown in the outer ring and coloured according the presence of CDSs on each island with homologues restricted to plant-associated bacteria (green), animal-pathogenic bacteria (red), insect-associated bacteria (dark red) and those specific to *Pantoea* species (blue). Islands harbouring integrated phage and plasmid elements are coloured grey, while those carrying CDSs with homologues in a subset of bacteria occupying all ecological niches are coloured in yellow. The inner diagram shows the CDSs with homology in 31 groups of bacteria incorporated into a Genome diagram and arranged as per Fig. 3.3. White regions indicate where no homologues are present to the respective *P. ananatis* CDS are present in compared genomes, while those showing 30-60% aa identity are light grey, those sharing 60-80% identity are dark grey and those with 80-100% aa identity are coloured in black.
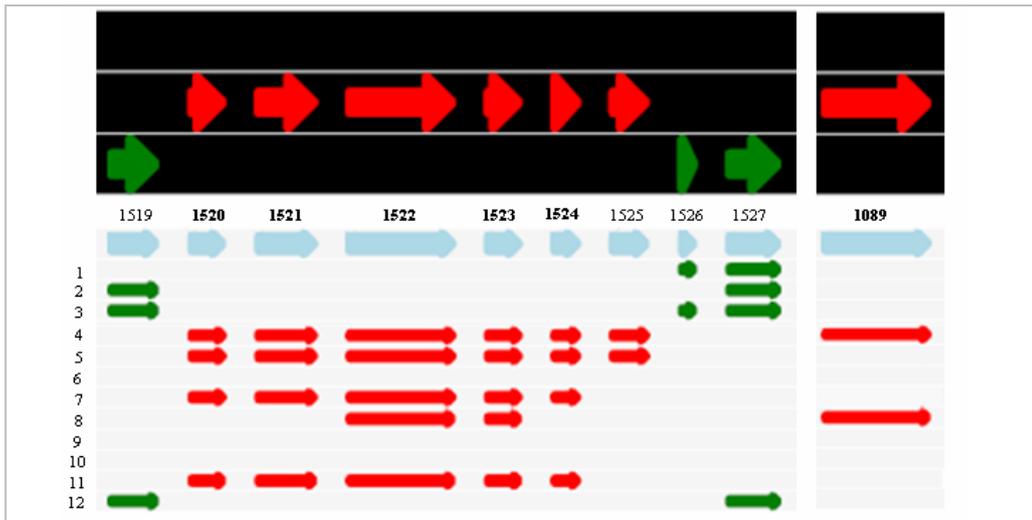
Figure 3.5: Fragments of linear genome diagram showing the *P. ananatis* LMG20103 CDSs (light blue) with homologues restricted to the animal-associated bacteria (in red) while other CDSs are restricted to the plant-associated bacteria (in green). Regions include the putative fimbria (PANA_1520-1524) and the autotransporter adhesion (PANA_1089). The tracks for the compared bacteria are as follows: 1. *P. stewartii* subsp. *stewartii* DC283 2. *Pantoea* sp. At9b 3. *Erwinia* spp. 4. *Salmonella* spp. 5. *Citrobacter* sp. 6. *K. pneumoniae* KPK342 (plant-pathogen) 7. *K. pneumoniae* MGH78578 (human pathogen) 8. *E. coli* O157:H7 9. *Pectobacterium* spp. 10. *Dickeya* spp. 11. *Yersinia pestis* CO92 12. *S. proteamaculans* 568
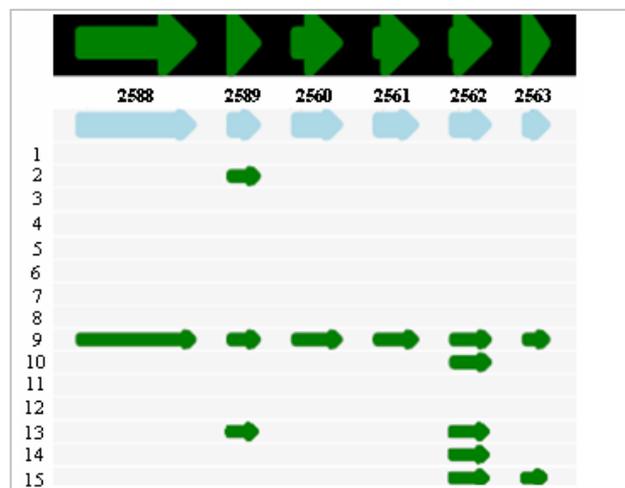
Figure 3.6: Fragment of a linear genome diagram showing the region of *P. ananatis* LMG20103 CDSs (PANA_2558-2563) encoding putative iron transport and metabolism proteins (light blue) with homologues restricted to plant-associated bacteria (green). Tracks for the compared genomes are as follows: 1. *P. stewartii* DC283 2. *Pantoea* sp. At9b 3. *Erwinia* spp. 4. *Salmonella* spp. 5. *Citrobacter* sp. 6. *K. pneumoniae* KPK342 (plant-pathogen) 7. *K. pneumoniae* MGH78578 (human pathogen) 8. *E. coli* O157:H7 9. *Pectobacterium* sp. 10. *Dickeya* sp. 11. *Yersinia pestis* CO92 12. *S. proteamaculans* 568 13. *Pseudomonas syringae* pvs. 14. *Xylella fastidiosa* M23 15. *R. solanacearum* GM1000

Chapter 4:

In silico identification and analysis of the putative pathogenicity factors of *Pantoea ananatis* LMG20103

ABSTRACT

Phytopathogenic bacteria make use of an extensive arsenal of pathogenicity determinants to infect, colonise and induce symptoms on their plant hosts. Little is known about the pathogenicity factors employed by the broad host range pathogen, *Pantoea ananatis*. The genome sequence of the *Eucalyptus*-pathogenic *P. ananatis* strain, LMG20103, was mined to identify putative pathogenicity factors. This revealed that it encodes a vast number of proteins which may be involved in gaining entry into the plant, establishing itself and causing disease symptoms. Interestingly, Type II, III and IV secretion systems, which play a major role in pathogenesis in many related plant pathogens, are absent from LMG20103. However, a number of pathogenicity determinants are produced which may circumvent the need for these secretion systems.

INTRODUCTION

Plant pathogenic bacteria pose a serious threat to food security and the economy, particularly in the current climate of globilisation, growing population numbers and global warming (Gregory et al. 2009). Gaining an understanding of the genetic mechanisms utilized by phytopathogenic bacteria to infect, colonise and cause disease on their hosts is key to understanding their pathology and the development of control strategies (Van Sluys et al. 2002). Phytopathogenic bacteria make use of a replete array of pathogenicity factors to infect their plant hosts, many of which have been functionally characterised (Vinatzer and Yan, 2008).

Traditionally, the identification and functional analysis of pathogenicity determinants has been a hit-and-miss process, with the use of random mutagenesis and gene or locus sequencing to identify individual factors with a role in pathogenesis (Chatterjee and Vidaver, 1986). Recent developments in the field of genomics has led to the whole genome sequences of a number of important phytopathogenic bacteria becoming available. These whole genome sequences provide extensive information resources which can mined to identify homologues to characterized pathogenicity determinants in the genome of a particular plant pathogen (Van Sluys et al. 2002). Whole genome sequencing has shown just how extensive the array of pathogenicity factor employed by many phytopathogenic bacteria is. Analysis of the *Pseudomonas*

131

syringae pv. tomato DC3000 genome sequence revealed that it carries 298 genes that encode factors with a role in causing disease on its plant host, encompassing 5% of the total genome content (Buell et al. 2003). In combination with genome-based functional approaches, 393 protein coding sequences (~9% of the total gene content) involved in pathogenesis were identified in *Pectobacterium atrosepticum* SCRI1043 (Bell et al. 2004). The sequencing of an increasing number of genomes has also facilitated comparative genomics, whereby comparison of a given pathogen to the genomes of other sequenced pathogens can reveal shared disease determinants, many of which have been functionally characterized (Van Sluys et al. 2002). Furthermore, differences in the virulence of two organisms can be correlated to discrepancies in their genomes and in this manner novel candidate pathogenicity determinants can be identified which can then be experimentally characterised.

Pantoea ananatis is an emerging pathogen on a broad range of agronomic crops such as onion, maize, rice and *Eucalyptus* (Coutinho and Venter, 2009). Little is known however about the mechanisms underlying its capacity to cause disease (Coutinho and Venter, 2009). In this chapter the genome sequence of the *Eucalyptus* pathogenic *P. ananatis* LMG20103 was mined for putative pathogenicity determinants. By cross-referencing with available literature on functionally characterised pathogenicity determinants, a large number of putative pathogenicity factors were identified. Through *in silico* analyses these pathogenicity factors were characterised and their potential roles in *P. ananatis* disease on its plant hosts are discussed.

MATERIALS AND METHODS

The amino acid sequences for the 4,237 protein coding genes (CDSs) of *P. ananatis* LMG20103 identified in Chapter 2 were mined for putative pathogenicity determinants by BlastP against the NCBI database (www.ncbi.nlm.nih.gov/Blast/) and by local BlastP against the sequenced genomes of closely related phytopathogenic bacteria. Those *P. ananatis* proteins which are homologous to pathogenicity factors in other bacteria for which functional data is available were classified according to their pathogenic function. Conserved domains in the pathogenicity proteins were identified by comparison against the Conserved domain database (Marchler-Bauer and Bryant, 2004). The subcellular localisations and presence of signal peptides determined in Chapter 2 were analysed and protein features such as molecular weights and

isoelectric points were determined for a number of putative pathogenicity factors. Neighbour-joining trees were constructed for several pathogenicity proteins based on clustalW alignment of their amino acid sequences, using Mega 4.0 (Tamura et al. 2007).

RESULTS AND DISCUSSION

3.1 Overview of pathogenicity factors produced by *P. ananatis* LMG20103

Analysis of the LMG20103 genome revealed the presence of 433 protein coding sequences which share homology with proteins that have been experimentally determined to play roles in pathogenesis in other plant pathogens (Fig. 4.1). Thus ~10% of the total proteins encoded on the genome may play a role in the diseases caused by *P. ananatis* LMG20103 on its *Eucalyptus* host. This is comparable to the ~9% of proteins in the potato pathogen *Pectobacterium atrosepticum* SCRI1043 with a predicted role in soft rot disease (Bell et al. 2004). The putative *P. ananatis* LMG20103 pathogenicity proteins form components of several distinct systems performing different pathogenicity functions, including motility and chemotaxis, iron acquisition from the host, protection against plant host defenses, biosynthesis of extracellular polysaccharides and lipopolysaccharides, adhesion of host surfaces and cells, plant cell wall degradation, quorum sensing, secretion of pathogenicity effectors and regulation of pathogenicity factors (Fig. 4.1). These are discussed in detail below.

Motility and Chemotaxis

Motility through flagella

Flagellar motility enables a pathogen to migrate toward an infection site, traverse host surfaces, colonise tissues and evade host defenses and thus represents a major pathogenicity determinant (Tans-Kersten et al. 2001). Flagella have also been shown to determine the host range of plant pathogens (Takeuchi et al. 2003). Genes encoding flagellum structural and functional components as well as a chemotaxis cascade are present in *P. ananatis* LMG20103 indicating that it is capable of directed motility. Furthermore, *P. ananatis* carries a flagellin glycosylation island which may influence the function of its flagella.

Flagella in P. ananatis LMG20103

Forty-eight genes encoding proteins involved in flagellum biosynthesis and functioning are present in the LMG20103 genome, typical of peritrichously flagellated bacteria of the family *Enterobacteriaceae* (Josenhans and Suerbaum, 2002). These are located in eight clusters encompassing sixteen operons. The encoded proteins show 94% (in 46 CDSs) and 68% (in 41 CDSs) average aa identity, respectively, to orthologues in *P. stewartii* DC283 and *E. coli* K-12. Four copies of the *fliC* which encodes flagellin, the main structural component of the flagellum, are present in *P. ananatis* LMG20103. Similarly, four copies are present in *P. stewartii* DC283, while *E. coli* and *Salmonella* strains encode one or two copies (Bonifield and Hughes, 2003). Flagellin copy number dictates the phenomenon of phase variation. As flagellin represents a potent surface antigen which can induce defense responses in plant hosts, variation of the flagellin expressed on the cell surface enables bacteria to avoid triggering host defenses (Bonifield and Hughes, 2003). Discrepancies in antigenicity result from disparities in the surface exposed amino acids in the central portion of the flagellin protein, while the N- and C-terminal regions are conserved (Bonifield and Hughes, 2003). Three *fliC* copies encode proteins with high homology to each other, while the PANA_3356-encoded FliC protein differs substantially (Table 4.1) suggesting phase variation may occur between the PANA_3356 and PANA_2281/4026/4061-encoded flagellins. This is supported by differences in the central region of the FliC sequences, while the N- and C-terminal regions are more conserved (Fig. 4.2). The four flagellins may thus allow *P. ananatis* to evade host defense initiation.

Flagellin glycosylation in P. ananatis LMG20103

While flagellum biosynthetic and functional components are well-conserved among motile bacteria, some bacteria add *N*- and *O*-linked carbohydrate subunits to their flagella which have marked effects on the flagellum function. Flagellin glycosylation has been detected in a limited subset of plant and animal pathogens including *Pseudomonas* and *Xanthomonas* species (Takeuchi et al. 2003). Several genes, located on a glycosylation island (GI) directly upstream of *fliC* control the synthesis, polymerisation and activation of the carbohydrate subunits attached to the flagellin (Arora et al. 2004). A role for flagellin glycosylation in pathogenicity has been observed in *P. syringae* pv. glycinea where GI mutants fail to invade soybean.

134

Flagellin glycosylation is hypothesized to allow the pathogen to avoid flagellin detection by the host (Takeuchi et al. 2003).

The predicted 21.3 kb *P. ananatis* LMG20103 island HAI14 encodes sixteen proteins (Chapter 3) and is located between the *fliA* gene (PANA_2264) and one of the four flagellin genes (PANA_2281) (Fig. 4.3). Comparison to *Enterobacteriaceae* members shows this island to be absent from its close phylogenetic relatives. Furthermore HAI14 has a G+C content of 47.49%, 6.2% below the genome average, indicative of recent horizontal acquisition. Six and four HAI14 CDSs show homology to flagellin glycosylation proteins in *P. aeruginosa* JJ692 and *P. syringae* pv. tabaci 6605, respectively (Table 4.2, Fig. 4.3). HAI14 thus likely represents a flagellin glycosylation island. Furthermore, orthologous roles can be ascribed to several non-homologous proteins in *P. ananatis* HAI14 and the *P. syringae* pv. tabaci GI, with PANA_2276 and BAH58343 encoding acyl-carrier protein synthases, while PANA_2277 and BAH58341 encode methyltransferases (Table 4.2). PANA_2280 and PANA_2281 encode deoxyhexose glycosyltransferases with homology to *vioT* and *orfN* in *P. syringae* pv. tabaci and *P. aeruginosa*, respectively.

Nine genes at the 5' end of the *P. ananatis* LMG20103 GI share no homologues in the *Pseudomonas* GIs (Fig. 4.3). Four of these genes, PANA_2266-2269 encode homologues to proteins involved in the synthesis of dTDP-rhamnose which is incorporated in the O-antigen of several bacteria, including *Serratia marcescens* and *P. syringae* (Saigí et al. 1999; Takeuchi et al. 2003). The location of rhamnose biosynthetic enzymes in the *P. ananatis* GI suggests that it adds rhamnose sugar subunits to its flagellin proteins and in this way may be able to avoid host detection of the flagellar proteins.

Chemotaxis

The ability of a plant pathogenic bacterium to direct the rotation of its flagellum to move towards infections sites and away from plant defenses is governed by the process of chemotaxis. Chemotaxis functions through surface receptors, known as methyl-accepting chemotaxis proteins (MCPs), which detect chemicals such as simple sugars released at the infection site or plant defense molecules and pass a signal down a chemotaxis cascade which modifies the direction of flagellum rotation (Brencic and Winans, 2005; Szurmant and Ordal, 2004).

Homologues to all proteins that make up the typical chemotaxis cascade are encoded on the *P. ananatis* LMG20103 genome. Analysis of the genome however revealed an unusually large number of MCP-encoding genes, namely 42 distinct genes. By contrast, the two organisms in which chemotaxis has been thoroughly characterised, *E. coli* K-12 and *S. enterica* serovar Typhimurium LT2, encode five MCPs each (Taylor et al. 1999; Iwama et al. 2006). To elucidate the significance of the extensive MCP gene set in *P. ananatis* LMG20103, MCP genes were identified in a number of closely related animal-pathogenic, plant-associated and plant-pathogenic bacteria (Fig. 4.4). The number of receptor genes does not follow a phylogenetic pattern, but rather appears to be linked to the lifestyle and ecology of the organisms. Animal-pathogenic and animal-associated strains carry between 0 (the aflagellate *K. pneumoniae*) and 9 (*P. mirabilis*) MCP genes. Exceptions are *Cronobacter sakazakii* ATCC BAA-894 and *B. cenocepacia* J2315 which encode 24 and 18 receptors respectively. Both are opportunistic human pathogens which are frequently isolated from water and soil, suggesting the extra MCPs may play a role in regulating directed motility in the environment (Holden et al. 2009; Kucerova et al. 2010). Between 13 (*E. tasmaniensis* Et1/99) and 28 (*R. etli* CFN42) chemoreceptor genes are found in non-pathogenic plant-associated bacteria (Fig. 4.4). The greater number of MCPs in these bacteria suggest many of them are involved in taxis in the plant environment. Interestingly, between 23 (*R. solanacearum*) and 53 (*B. glumae*) MCPs are encoded by phytopathogenic representatives (Fig. 4.4). The higher number of MCPs in plant-pathogenic bacteria compared to plant-associated bacteria suggests that a subset of chemoreceptors play a direct role in pathogenesis. Specific roles for chemotaxis in plant disease have been described in several phytopathogens including *D. dadantii* and *R. solanacearum* (Antúnez-Lamas et al. 2009; Yao and Allen, 2007).

Comparison of the amino acid sequences of the *P. ananatis* MCPs to those which have been functionally elucidated in *E. coli* K-12 (Tsr, Trg, Tar and Tap) and *S. enterica* serovar Typhimurium LT2 (Tcp) showed that a subset of *P. ananatis* MCPs clustered with these receptors and likely respond to similar attractants and repellents (Fig. 4.5). PANA_0296 shows homology to *E. coli* K-12 Aer protein and thus regulates taxis towards niches with optimal oxygen concentrations. A role in plant pathogenesis has been described in *R. solanacearum*, where the Aer receptor is required for root and xylem colonisation and virulence (Yao and Allen, 2007).

PANA_0056 shows 41% aa identity to BdlA in *P. aeruginosa* PA1432 (Q9I3S1) which responds to environmental signals and regulates dispersal of biofilms, returning the bacterium to a planktonic state for dispersal to a novel niche (Morgan et al. 2006). Twenty-three *P. ananatis* MCPs (55% of total) share homologues only in plant-associated bacteria, including non-pathogenic inhabitants of the plant niche (Fig. 4.5). This agrees with the observation that more MCPs are found in plant-associated bacteria. Seven MCPs (17% of total) share homologues only in plant-pathogenic bacteria. It can be speculated that these motivate taxis towards plant exudates or away from plant defense molecules released during infection.

Fimbria, non-fimbrial adhesins and attachment

Adherence to host tissues and surfaces is one of the primary steps in host colonisation. This role is fulfilled by fimbrial and non-fimbrial adhesins (Proft and Baker, 2009). Seven putative fimbriae belonging to three fimbrial families and four non-fimbrial adhesins are encoded on the *P. ananatis* LMG20103 genome (Fig. 4.6; Fig. 4.7).

Fimbriae in P. ananatis LMG20103

Type 1 fimbriae

Type 1 fimbriae are widespread among Gram-negative bacteria and are synthesised via the chaperone-usher pathway (Sauer et al. 2000). Prepilin subunits are transported to the periplasm via the Sec machinery. In the periplasm prepilin subunits form a stable complex with the FimC chaperone which folds them for assembly (Sauer et al. 2000). The chaperone-subunit complex is targeted to the outer membrane usher FimD which forms a channel allowing pilin subunits to pass to the cell exterior (Sauer et al. 2000). The major exported pilin FimA forms a pilus rod. A tip fibrillum, consisting of FimF and -G is attached to the rod. FimD attaches the adhesion FimH to the fibrillum tip, which then interacts with target surfaces or host cell receptors (Proft and Baker, 2009). Four type 1 fimbrial biogenesis loci are present in *P. ananatis* LMG20103.

Stb-type fimbria (Cluster III). The *stb* type 1 fimbria of *S. enterica* enhances its intestinal persistence in mice and is encoded by five genes, *stbABCDE* (Weening et al. 2005). An 8.5 kb horizontally acquired region in the LMG20103 genome (HAI8) carries five genes, PANA_1520-1524 which homologues with 59% average aa identity to StbABCDE of *S. enterica* serovar Typhi CT18 (STY0369-STY0373).

PANA_1524 encodes a homologue to major pilin StbA with a chaperone StbB and outer membrane ushers StbCD encoded by PANA_1521-1523. PANA_1520 encodes a homologue to the tip adhesin StbE with lower identity (47%) to StbE in *S. enterica* (STY0369). The absence of homologues in plant-associated bacteria however suggests this fimbria may be involved in *P. ananatis* attachment to tissues in animal hosts.

GltF-YchADEF-type fimbria (Cluster IV). PANA_1796-1798 show homology to genes in the *E. coli* K-12 *gltF-yhcADEF* cluster which encodes a putative fimbria (Nuccio and Bäumler, 2007). The PANA_1797 protein is similar to the outer membrane usher FimD and shares 22% aa identity with *E. coli* K-12 YhcD (NP_417683.1). PANA_1796 encodes a predicted outer membrane protein with 20% aa identity to YhcF (NP_4176686.1). PANA_1798 shows 32% identity to the FimC-like YhcA chaperone (NP_417682.1). No homologues to YhcE or GltF are found in *P. ananatis*. The absence of prepilin or tip adhesin genes in the cluster suggests this chaperone/usher system may not synthesise functional fimbriae, but rather simple fibrillar structures (Nuccio and Bäumler, 2007).

Mannose-resistant Sef/Mrk-like Type I fimbriae (Cluster V). A 10.1 kb cluster (PANA_2715-PANA_2726) encodes twelve proteins with nine of these showing homology to components of mannose-resistant type 1 fimbriae. This constitutes the most extensive type 1 pilus in *P. ananatis* LMG20103. Highest homology was observed with proteins for the characterised fimbriae in the insect-associated *Serratia entomophila* (Sef fimbria - Hurst et al. 2003) and *Photorhabdus temperata* K122 (Mrf fimbria - Meslet-Cladiere et al. 2004; Table 4.3). In our laboratories, *P. ananatis* has been isolated from Mmyrid insects in *Eucalyptus* plantations (results unpublished) and phytopathogenic strains have been isolated from several vectoring insects (Gitaitis et al. 2003). Similarity to the *S. entomophila* and *P. temperata* fimbrial systems suggests the *P. ananatis* LMG20103 cluster V fimbria could be involved in attachment in an insect host.

Csu type 1 fimbria (Cluster VI). Csu fimbria, encoded by the *csuABCDE* locus, are found in the human pathogen *Y. pestis* and the environmental bacterium *Pseudomonas putida* and allows them to attach to surfaces in their natural environment (Tomaras et al. 2003). PANA_3330-3331 encode homologues to the major pilin subunits CsuA

and B. PANA_3332-3333 are similar to chaperone CsuC and usher CsuD and PANA_3334 to the tip adhesin CsuE, with 75% average aa identity to *Pantoea* sp. At9b CsuABCDE (Pat9b_5398-5402). The role of Csu fimbriae in attachment to biotic surfaces suggests the Cluster VI fimbria may be involved in *P. ananatis* attachment to plant surfaces.

Type 3 fimbria (Cluster I)

Type 3 MR/K (**M**annose **r**esistant ***K**lebsiella*-like) fimbriae are produced by several *Enterobacteriaceae* and play a role in human disease, mediating attachment to renal and bronchial cells and abiotic surfaces such as urinary catheters (Ong et al. 2008). Although a role in plant pathogenesis has not been determined, the type 3 fimbriae of nitrogen-fixing *Klebsiella* spp. are involved in attachment to roots (Korhonen et al. 1983). Type 3 fimbriae are encoded by six genes, *mrkEABCDF*, which synthesise fimbriae via a chaperone-usher pathway (Huang et al. 2009). MrkA is the major fimbrial subunit while MrkB and MrkC encode a chaperone and outer membrane usher for fimbrial assembly. MrkD encodes a tip adhesin, MrkE controls fimbrial expression and MrkF regulates fimbrial length and serve as signal for bacterial aggregation and biofilm formation (Huang et al. 2009). PANA_0623-0625 show 31% average aa identity to *K pneumoniae* MrkABC (AAA25093-25095). The PANA_0626 protein shares 35% aa identity with the *E. tasmaniensis* MrkD adhesin (ETA_28550). PANA_0627 encodes a MrkB orthologue (33% aa identity to PANA_0624). PANA_2743, located separately from the other Csu fimbria biogenesis genes, encodes MrkE homologue with 71% aa identity to *K. pneumoniae* AAA25092. No MrkF homologue is present. Given the role of type 3 fimbriae in *Klebsiella* spp. adhesion to roots, the *P. ananatis* fimbria may be involved in attachment to plant surfaces.

Type 4 fimbriae (Cluster IIA-E)

Type 4 fimbriae are long, flexible structures that are assembled via their own dedicated pathway. Unlike type 1 and 3 fimbriae the major structural subunits, are polymerised in the cytoplasm and exported as a whole across the outer membrane (Nudleman and Kaiser, 2004). PilA prepilins are polymerised and transported across the inner membrane by PilC and the PilB ATPase. During extrusion the PilD leader peptidase cleaves the N-terminal peptide and N-methylates mature pilins (Nudleman and Kaiser, 2004). The pilins are translocated across the outer membrane through a

139
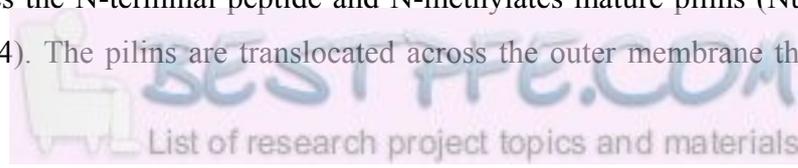
PilQ channel formed by PilQ which is assembled with the aid of the PilF, -N and –P lipoproteins. Pilin translocation through the PilQ secretin is energised by the PilM ATPase (Nudleman and Kaiser, 2004). The PilT ATPase acts in the opposite direction to PilB to retract the fimbria (Nudleman and Kaiser, 2004). The ability to retract the fimbria enables movement through twitching motility and facilitates close contact between pathogen and host cell for pathogenicity factor secretion into the cell (Nudleman and Kaiser, 2004).

In *P. ananatis* LMG20103 nine genes encoding putative type 4 fimbrial components are located in five clusters (cluster IIA-E; Fig. 4.6). Cluster IIA (PANA_0733-0735) encodes homologues to PilABC in *Pantoea* sp. At9b (Pat9bDRAFT_3115-3117: 60% average aa identity). Cluster IIE contains genes encoding PilMNQ homologues with PANA_3669, 3672 and 3673 sharing 60% average aa identity to Pat9bDRAFT_4411 (PilQ), 4414 (PilN) and 4415 (PilM). A homologue to the PilQ assembly lipoprotein PilP is absent from the genome. However, SignalP analysis of PANA_3670 located between *pilQ* and *pilN-pilM* in cluster IIIC shows it contains a signal peptide typical for outer membrane lipoproteins and may thus encode a PilP homologue. The PANA_2860 gene (Cluster IIB) encodes a PilF homologue with 62% aa identity to Pat9bDRAFT_1074. PANA_3625 (Cluster IID) encodes a homologue to leader peptidase PilD (Pat9bDRAFT_3256: 42% aa identity), while PANA_3219 (Cluster IIC) is homologous to PilT (Pat9bDRAFT_3537: 73% identity), indicating that *P. ananatis* can retract its type 4 fimbria. Homologues to all proteins for fimbrial assembly and functioning are thus present in *P. ananatis* (Fig. 4.7). Type 4 fimbriae are involved in diverse functions including DNA transfer, surface dispersal, twitching motility and biofilm formation (Li et al. 2007). The role of the *P. ananatis* type 4 fimbria could thus encompass a range of functions.

Non-fimbrial adhesins in P. ananatis LMG20103

ShlA/HecA/FhaB-like adhesin (PANA_0907)
BlastP with the PANA_0907 aa sequence against the NCBI resolves this protein into the HecA/FhaB/ShlA family of haemagglutinins, adhesins and haemolysins. Members are characterised by a large size, high molecular mass, high glycine content and are secreted via a Type V two-partner secretion system (Jacob-Dubuisson et al. 2001). This family also includes the **C**ontact-**d**ependent growth **i**nhibitor CdiA of

uropathogenic *E. coli* which forms a fibre on the cell surface (Aoki et al. 2009). When this fibre comes into contact with the outer membrane protein BamA found in most bacterial and eukaryotic cells, it down-regulates ATP levels, proton motive force and cell division in these cells. CdiA thus behaves as toxin, inhibiting growth of competing bacteria or host cells (Aoki et al. 2009). It can also induce a CDI phenotype within the producing strain, reducing respiration and membrane potential, providing resistance to oxidative stress and host antimicrobials (Aoki et al.2009).

Comparison to the functionally characterised members of the HecA/ShlA/FhaB family shows that in terms of both sequence similarity and shared protein structural features, the PANA_0907 protein is most similar to CdiA (Fig. 4.8; Table 4.4). This protein could thus protect *P. ananatis* from oxidative stress and host antimicrobials or may serve as phytotoxin. However, conserved domain analysis of the PANA_0907 sequence reveals that it carries two arginine-glycine-aspartate (RGD) motifs, which are found in adhesins such as the *Bordetella pertussis* filamentous haemagglutinin FhaB and play a role in adhesion (Relman et al. 1989), as well as a conserved haemagglutinin active domain (Pfam05860), suggesting the PANA_0907 protein may yet play a role in adhesion.

AidA-I-like adhesin (PANA_1089)

The PANA_1089 protein shows significant homology to the *E. coli* Type V autotransporter adhesin AidA-I which is involved in diffuse adherence to epithelial cells, aggregation, biofilm formation and pathogenesis (Henderson et al. 2004). The C-terminal region of the AidA-I protein is incorporated into the outer membrane and translocates the N-terminal adhesive region to the cell surface where it is displayed (Charbonneau et al. 2006). Alignment of the PANA_1089 and *E. coli* O157:H7 str. Sakai AIDA-I (ECs0362) proteins indicated that homology is limited to their C-terminal regions (47% in 500 aa). This region contains a Pertactin-like subgroup 2 domain (PL2-passenger: CDD01344) and an outer membrane AidA domain (AidAc: TIGR01414) (Fig. 4.9). The 532 aa N-terminal region of the PANA_1089 protein shows no homology to any known proteins. The presence of a PL2-passenger and AidAc domain for outer membrane translocation and two RGD motifs suggests the PANA_1089 protein represents an adhesin with an unusual adhesive domain.

Fibronectin-binding protein (PANA_1495)

The 198 aa PANA_1495 protein shares highest homology (56% aa identity) with the uropathogenic *E. coli* UTI89 fibronectin binding protein B (YP_540246). Although an exact function for this protein has not been determined in the *Enterobacteriaceae*, a similar protein in *Streptococcus* species binds fibronectin and fibrinogen proteins in the eukaryotic extracellular matrix and is important for invasion and colonisation of animal cells (Rocha and Fischetti, 1999). Homologues in phytopathogenic *Pectobacterium* spp. and the endophyte *E. tasmaniensis* suggest the PANA_1495 adhesin may play a role in attachment in plant hosts.

SiiE-like Type I secreted adhesin (PANA_2607)

PANA_2607 encodes the largest protein in *P. ananatis* LMG20103 (4,179 aa in size) with most significant homology to an adhesin in *Yersinia mollaretii* ATCC 43969 (ZP_04640954: 40% aa identity), the large repetitive protein YeeJ in *P. atrosepticum* (ECA1099: 38% aa identity) and the adhesin SiiE in *S. enterica* serovar Typhimurium (CAE11797: 25% aa identity). *siiE* is located on pathogenicity island SPI4 which also carries a type 1 secretion system (Morgan et al. 2007). Similarly, PANA_2607 is located adjacent to genes encoding T1SS components (PANA_2603-2605). T1SS adhesin proteins are typified by a large size required for proper function, allowing projection of the adhesin beyond LPS O-antigen chains and contact with host cell receptors buried beneath cell surface structures (Gerlach et al. 2008). Other common features are the presence of up to 50 glycine-rich immunoglobulin repeat domains and a C-terminal secretion signal (Table 4.5; Delepelaire, 2004). A consensus repeat was identified by BlastP with the *Y. molaretti* adhesin repeats against the PANA_2607 protein. Thirty-one repeats, 76 aa in length with 40.4% average aa identity to the consensus sequence, are interspersed throughout the amino acid sequence with exception of the N- and C-terminal regions (Fig. 4.10) where the adhesive domain is suggested to reside in SiiE (Gerlach et al. 2008). SiiE is involved in attachment of *Salmonella* strains to intestinal epithelial cells and colonisation of calf intestines (Main-Hester et al. 2008). Given the structural and sequence similarities between the PANA_2607 and SiiE proteins (Fig. 4.10; Table 4.5), a similar role in attachment to host cells can be envisaged for the *P. ananatis* protein.

Iron acquisition

Phytopathogenic bacteria scavenge the limited supply of iron in the host and thereby disrupt iron-requiring cellular processes in the plant (Wandersman and Delepelaire, 2004). Siderophores, iron-chelating proteins, are secreted into the extracellular environment. The iron-bound siderophore is recognised by a cognate outer membrane receptor and transported to the periplasm via a channel formed by the receptor. This process is energised by a TonB-ExbBD complex (Létoffé et al. 2004). Subsequently, a specific ABC transport system translocates the iron-siderophore complex to the cytoplasm where it is dissociated to release the bound iron (Bamford et al. 2008). Production and utilisation of multiple siderophores and iron uptake systems increases the capacity of a pathogen to survive in iron-poor and -competitive environments and hosts (Luck et al. 2001). Analysis of the *P. ananatis* genome revealed that it produces two siderophores, eight outer membrane receptors and several inner membrane iron transport systems for uptake of both ferrous and ferric iron (Fig. 4.13). This suggests it possesses means for efficient iron acquisition within the plant host. Given the fine balance of iron within these hosts, scavenging by *P. ananatis* is thus likely detrimental to its plant and potentially its vertebrate hosts.

Siderophores

Two siderophore biosynthetic loci could be identified on the LMG20103 chromosome. A 5.8 kb locus carrying four genes, PANA_0349-353, encodes enzymes for the biosynthesis of the hydroxamate siderophore aerobactin which is produced by many *Enterobacteriaceae*. The encoded proteins share 78% average aa identity with the IucA-D aerobactin biosynthetic enzymes of *Shigella flexneri* M90T (AAD44746-44750). The second locus, PANA_0636-0639, encodes three proteins with homology to the enzymes involved in the biosynthesis of the alcaligin and desferrioxamine (DFO) siderophores produced by *Bordetella pertussis* and *E. amylovora*, respectively. Alcaligin biosynthesis has been extensively studied and involves a cluster of seven genes, *alcABCDERS*. PANA_0636 and PANA_0637 encode homologues to *B. pertussis* biosynthetic enzymes AlcA and AlcBC (BP2456 and BP2457-2458) respectively (Fig. 4.11). No homologues to the protein of unknown function AlcD, the AlcE iron-sulphur protein and AlcR transcriptional regulator are present. PANA_0638 shows homology to AlcS (BP2462), a permease that exports alcaligin from the *B. pertussis* cell (Brickman et al. 2007). No homologue to AlcS is present in *E.*

amylovora. Aside from iron scavenging from the host plant, DFO has been shown to protect *E. amylovora* from oxidative damage and contributes to plant cell wall damage and electrolyte leakage (Dellagi et al. 1998). Given the extensive homology between the *P. ananatis* proteins and those for DFO synthesis in *E. amylovora, P. ananatis* produces a highly similar siderophore which likely functions in phytopathogenesis.

Outer membrane siderophore receptors

Eight putative outer membrane ferri-siderophore receptors are encoded on the *P. ananatis* LMG20103 genome. These show a wide distribution in terms of the iron-loaded siderophores they can bind and transport into the periplasm (Fig. 4.12). PANA_0354 encodes a homologue of IutA in *S. flexneri* M90T (76% aa identity – AAD44750) and is located adjacent to the aerobactin biosynthetic genes, confirming a role as aerobactin receptor (Purdy and Payne, 2001). The PANA_0639 protein shares 86% aa identity with *E. amylovora* desferrioxamine receptor FoxR (CAA11064) and is also located adjacent to DFO biosynthetic genes, indicating a role in uptake of ferric-DFO in *P. ananatis*.

Six further siderophore-specific receptors with no cognate siderophore biosynthetic systems are present in *P. ananatis* LMG20103. These are likely utilised to scavenge iron-loaded siderophores produced by other microorganisms sharing the same niche. Xenosiderophore scavenging is common among bacteria and allows acquisition of iron bound to several structurally related siderophores without the energy cost of siderophore biosynthesis (Brickman et al. 2007). PANA_2945 shows homology to enterobactin receptors FepA, CirA and IroN in *S. enterica and E. coli* and may allow *P. ananatis* to scavenge ferric-enterochelin from other *Enterobacteriaceae*. PANA_0775 shows homology to *P. agglomerans ferrichrome* siderophore receptor FhuA (81% aa identity - CAA74355). A second copy encoded by PANA_1585 shows 54% identity to CAA74355. The PANA_4237 protein shows 57% aa identity to FhuE in *E. coli* K-12 (CAA35616) which detects and binds coprogen, a hydroxamate siderophore produced by the fungal genera *Neurospora* and *Penicillium* (Cui et al. 2006). Two further genes encode homologues to ferri-siderophore receptors of unknown substrate-specificity restricted to plant-associated bacteria (PANA_4210 – 44% aa identity - *P. syringae* pv. tomato DC3000 PSPTO_3574; PANA_2588 – 52%

aa identity - *P. atrosepticum* SCRI1043 ECA1275). As pertinent surface structures, siderophores receptors are highly immunogenic. To avoid host defense systems, receptor proteins can be expressed phase variably (Wandersman and Delepelaire, 2004). By maintaining eight distinct receptors *P. ananatis* LMG20103 may use them alternately and thus avoid host detection and defense response.

TonB and ExbB/D complexes

The TonB complex, consisting of the TonB protein and the auxillary proteins ExbB and ExbD, energises the translocation of iron-siderophore complexes across the outer membrane and plays a role in the release of empty siderophores from the cell surface (Létoffé et al. 2004). In *E. coli* there is only one set of TonB-ExbBD, while it hosts seven potential receptors. The receptors thus compete for the limited TonB transducer. Other bacteria have overcome this limitation by carrying multiple copies of the TonB complex (Wandersman and Delepelaire, 2004). Two TonB complexes are encoded by *P. ananatis* LMG20103. BlastP with the PANA_2591-2593 protein sequences identified homologues in a restricted number of plant-associated bacteria with highest identity to TonB-ExbBD in *P. atrosepticum* SCRI1043 (ECA1270-2: 53% average aa identity). PANA_2591-2593 are located adjacent to genes encoding homologues of the FecI and FecR (PANA_2589-2590) iron sensor proteins and a siderophore receptor (PANA_2588). The latter protein shows highest homology to a receptor in *P. atrosepticum* and this TonB complex is likely involved in energising iron uptake in the plant environment. The second TonB (PANA_2077) shows 71% aa identity to *P. agglomerans* TonB (CAC79956). Located separately on the chromosome PANA_3336-3337 encode proteins with 94% average aa identity to *P. stewartii* DC283 ExbBD (ACV-28149-280150).

Inner membrane transport systems

Transport of iron across the inner membrane occurs via periplasmic binding protein-dependent (PBT) systems consisting of a periplasmic siderophore-binding component, inner membrane components and an energy-coupling ATPase (Wandersman and Delepelaire, 2004). Six putative PBT systems are encoded on the LMG20103 genome. PANA_0776-0778 show 65% average aa identity to FhuBCD in *E. coli* K-12 and are located downstream of the *fhuA* receptor gene PANA_0775. FhuBCD are involved in internalisation of hydroxamate iron-siderophore complexes including

ferrichrome, aerobactin and ferrioxamine (Wandersman and Delepelaire, 2004). PANA_2946-2949 encode four proteins with 64% aa identity to a siderophore transport system in *Y. enterocolitica* 8081 (YE3583-3586) located adjacent to a *cirA* homologue (PANA_2945) and thus likely transport an enterobactin xenosiderophore. Similarly, PANA_2549 and 2550 encode proteins for transport ferric-enterobactin (85% aa identity to *Pantoea* sp. At9b - Pat9bDRAFT_4218-4219). The PANA_3313-3315 proteins show homology to the AfuABC PBT system proteins found in plant-associated bacteria (46% average aa identity to *A. tumefaciens* C58 - Atu2012-2014). PANA_2335-2337 and PANA_2971-2973 encode components of ferric-dicitrate transport systems. However, no FecA ferric-dicitrate receptor homologues are present, suggesting different iron-siderophores may be transported by these two systems.

Under anaerobic and reducing conditions, iron predominantly occurs in the ferrous form (Fe^{2+}). Fe^{2+} diffuses freely across the outer membrane and is subsequently transported by an ABC ferrous ion transporter (Wandersman and Delepelaire, 2004). Three such systems are encoded by *P. ananatis* LMG20103. PANA_0383-0384 encode homologues to components of the FeoAB pump (63% aa identity to *Citrobacter koseri* BAA-895 CKO4830-1) (Wandersman and Delepelaire, 2004). SitABCD in *S. enterica* and avian pathogenic *E. coli* transports ferrous iron as well as manganese. Manganese contributes towards protection against oxidative stress by activation of the manganese-dependent superoxide dismutase SodA and SitABCD thus plays a role in pathogenesis (Sabri et al. 2008). PANA_4200-4203 encode proteins with 78% average aa identity to SitABCD in *E. coli* χ7122 (AAT11264-11267). PANA_2250-2252 encode components homologous to the EfeUOB Fe^{2+} uptake system (Cao et al. 2007).

Iron dissociation from siderophores
Once the siderophore has been taken up, ferric iron is released by degradation of the siderophore or reduction to a ferrous state (Bamford et al. 2008). Iron from ferric-enterobactin is removed by the ferric enterobactin esterase Fes (Bamford et al. 2008). PANA_0567 contains a conserved enterochelin esterase domain (COG2382) and shows 66% aa identity to Fes in *E. coli* K-12 (BAE77569). FhuF releases iron from ferrichrome siderophores (Bamford et al. 2008) but no homologues are present in *P. ananatis* LMG20103. The PANA_0347 shows 56% aa identity to the FAD-binding 9

siderophore-interacting protein of *S. proteamaculans* (Spro_0924). It is located directly downstream of PANA_0349-0354 which encode the predicted aerobactin siderophore and may thus dissociate iron from aerobactin.

<u>Quorum sensing</u>

Bacterial plant pathogens make use of an intercellular communication mechanism known as quorum sensing (QS) for the density dependent regulation of a number of pathogenicity factors, thereby ensuring that they are expressed at effective concentrations at the appropriate phase of infection (de Kievit and Iglweski, 2000). Gram-negative pathogens do this through the release of signalling molecules, Acyl-homoserine lactones (AHLs) which, when present at threshold levels, trigger a cognate transcriptional factor that regulates the expression of genes encoding pathogenicity determinants (de Kievit and Iglweski, 2000).

The genome of *P. ananatis* LMG20103 encodes at least three distinct QS systems. PANA_1955-1956 encode homologues to AHL-biosynthetic proteins EanI and EsaI and the cognate transcriptional regulators EanR and EsaR in the onion-pathogen *P. ananatis* SK-1 (100% aa identity – BAF69064-5) and *P. stewartii* (93% average aa identity - AAA82096-7), respectively. *P. ananatis* LMG20103 thus synthesises the 3-oxohexanoyl AHL identified in *P. ananatis* SK-1 and *P. stewartii* subsp. *stewartii* which ensures the timely production of adequate levels of exopolysaccharide, a major pathogenicity determinant in both of these organisms (Morohoshi et al. 2007; Koutsoudis et al. 2006). In the absence of a QS system, EPS would be produced prematurely when insufficient bacterial cells are presents, which would have drastic effects on early interactions between the pathogen and the plant, dissemination of the pathogen within the host and the development of symptoms and disease (Koutsoudis et al. 2006). The EanI/EsaI AHL activates the transcriptional regulator EanR/EsaR which in turn binds to complementary DNA region, a *lux* box, upstream of the target gene *rcsA* which encodes a regulator of EPS production (Morohoshi et al. 2007; Koutsoudis et al. 2006). A RcsA homologue is encoded by PANA_2308 and a similar *lux* box exists upstream of this gene, indicating that the PANA_1955-1956 QS system regulates EPS production in *P. ananatis* LMG20103, which as is shown in Chapter 6 also represents a major pathogenicity determinant, in a similar fashion as the *P. ananatis* SK-1 and *P. stewartii* QS systems. PANA_1416-1417 show 36% average aa

identity to RhlIR in *P. aeruginosa* PAO1 (EAZ59602-59602). The RhlI autoinducer and RhlR regulator control synthesis of the rhamnolipid biosurfactant which plays a major role in *P. aeruginosa* virulence (Ochsner et al. 1995). As putative rhamnolipid biosynthetic genes are found in *P. ananatis* LMG20103 the RhlIR homologues may regulate expression of a similar biosurfactant. However, *rhlIR* are clustered with rhamnolipid biosynthetic genes *rhlAB* in *P. aeruginosa* while they occur separately and may thus regulate a different phenotype in *P. ananatis*. PANA_2619 encodes a protein with 98% aa identity to the QS signalling molecule LuxS in *P. stewartii* (ACV-0280243). LuxS homologues have been found in several plant-associated bacteria and have been shown to interspecies communication. LuxS production can mislead another bacterium occupying the same ecological niche into assumption of high population density and interfere in their coordinate gene expression (Mohammadi and Geider, 2007). The LMG20103 LuxS may thus give this organism a competitive advantage over other microorganisms co-inhabiting the plant environment.

Lipopolysaccharides

Lipopolysaccharide (LPS) is a pertinent bacterial surface component which plays a major role in plant-microbe interactions and pathogenesis. It consists of three structural domains: a lipid A moiety, a core oligosaccharide, and a surface-exposed O-antigen polysaccharide (Babinski et al. 2002). Analysis of the *P. ananatis* genome indicates that while the lipid A moiety is similar to those in other *Enterobacteriaceae*, unusual core and O-antigen domains are present in the *P. ananatis* LMG20103 LPS. The genetic and structural properties of the *P. ananatis* core and O-antigen components are discussed below as well as their putative functions in pathogenesis.

The LPS core region of P. ananatis LMG20103

The core region is an oligosaccharide polymer consisting of an inner and outer domain (Canals et al. 2006). The enterobacterial inner core generally contains one or more 3-deoxy-α-$_D$-*manno*-octo-2-ulopyranosonic acid (Kdo) residues and two or three heptose ($_L$-$_D$-Hep) residues. The outer core consists of sequentially added hexose molecules that link the inner core and O-antigen (Grizot et al. 2006). The inner core of some enterobacteria contains only Kdo molecules while in others only heptose occurs. Extensive variability has also been observed in the outer core in terms of the

hexose sugars incorporated and the presence of side chains (Coderch et al. 2004). In *P. ananatis* LMG20103 biosynthetic enzymes for both Kdo and heptose are encoded with four enzymes for Kdo biosynthesis sharing 81% average aa identity to *E. coli* K-12 orthologues, while three genes encode proteins with 84% average aa identity to the heptose biosynthetic enzymes in *E. coli* K-12. Genes for inner core assembly (*waa*) are co-localised with those for outer core assembly (*wab*) in the *waa/b* gene cluster (Fig. 4.14). HldD (PANA2381 – 83% aa identity - *E. tasmaniensis* ETA_00850) catalyses transfer of heptose residues to the Kdo residue. WaaA transfers Kdo to the lipid A moiety (PANA_3898 - 83% aa identity - *S. marcescens* AAC44432). Heptose residues are assembled on the backbone by three heptosyltransferases, *waaC* (PANA_3888), *waaF* (PANA_3887) and *waaQ* (PANA_3896) with 84% average aa identity to orthologues in *S. marcescens* N28b. WaaE (PANA_3899 - 70% aa identity - *S. marcescens* AAC44433) transfers a glucose side chain to the first inner core heptose (Regué et al. 2005a).

Analysis of the *waa/wab* cluster indicates the *P. ananatis* LPS inner and outer moieties are most similar to those in *S. marcescens* N28b and the type 1 and 2 core regions in *K. pneumoniae* C3 and 52145 respectively (Fig. 4.14; Coderch et al. 2004; Regué et al. 2005a). The WabG transferase (PANA_3897 – 60% aa identity to *S. marcescens* N28b AAD28801) adds the first outer core galacturonic acid residue to the second inner core heptose subunit. WabH and WabN (PANA_3892-3893 – 60% average aa identity *K. pneumoniae* KPN_03967-03970) catalyse addition of a glucosamine residue to the galacturonic acid (Regué et al. 2005b). WabK (PANA_3889 – 64% identity - *S. marcescens* AAL23756) and WabM (PANA_3891 – 58% aa identity - *S. marcescens* AAL23758) are found in *K. pneumoniae* core type 2 and *S. marcescens* and are responsible for adding a glucose disaccharide to the glucosamine residue. In *K. pneumoniae* core type 1, this is replaced by a phosphorylated Kdo residue through WabI/WabJ, for which no homologues are found in *P. ananatis*. The WabO glycosyltransferase (PANA_3895 - 42% aa identity *K. pneumoniae* AAX20106) encoded in *K. pneumoniae* type 1 and 2 core *waa/wab* clusters but not *S. marcescens* adds a galacturonic acid residue to the third inner core heptose PANA_3894 shows 60% aa identity to Orf12 in *S. marcescens* N28b (AAD28803) which incorporates a heptose residue into the outer core absent from *K. pneumoniae* type 1 and 2 outer cores (Fig. 4.14; Coderch et al. 2004). PANA_3890

149

encodes a homologue of WaaL (46% aa identity – *E amylovora* EAM_0086) which is responsible for ligation of O-antigen to the outer core domain. An *E. amylovora waaL* mutant is sensitive to hydrogen peroxide and antimicrobial peptides. WaaL thus plays a role in protection against plant defenses (Berry et al. 2009).

Based on homology to the *waa/wab* clusters of *S. marcescens* N28b, *K. pneumoniae* C3 and 52145, a schematic model of the *P. ananatis* LMG20103 core domains could be constructed (Fig. 4.15). A role for LPS core in pathogenesis has been demonstrated in these organisms and thus the *P. ananatis* core domain may also function in disease (Coderch et al. 2004; Regué et al. 2005a). The glucose disaccharide in the *K. pneumonia* type 2 outer core has been demonstrated to effect hypervirulence over type 1 core strains (Regué et al. 2005a). As a similar disaccharide occurs in the LMG20103 outer core, it would be of interest to study the core regions in other *P. ananatis* strains to identify potential differences in their LPS core structures.

The P. ananatis LMG20103 LPS O-antigen

As surface-exposed moiety of LPS, the O-antigen plays several roles in plant disease, providing protection against plant antimicrobials and preventing hypersensitive response (Silipo et al. 2005). However, O-antigens can be recognised by plants, along with other surface molecules collectively named Pathogen Associated Molecular Patterns (PAMPs) and elicit plant defense responses (Silipo et al. 2005). O-antigens consist of polymers of repeat sugar units constructed by enzymes encoded in the *wb* gene cluster. One set of genes encodes enzymes for synthesis of O-antigen sugars, a second set encodes transferases that assemble the sugars in O-repeat subunits and a third encodes proteins for processing, export and assembly of O-repeat subunits in the O-antigen chain (Fitzgerald et al. 2003). Diversity has been observed in sugar biosynthetic and transferase genes as use of different sugar subunits and structural organisation determines the extensive O-antigen variation in antigenically distinct bacteria (Drummelsmith and Whitfield, 1999). By contrast, only two pathways for O-antigen assembly/export have been identified. The Wzx/Wzy pathway exports heteropolysaccharide O-antigens, while the less common ABC Wzm-wzt pathway transports mainly homopolysaccharides (Fitzgerald et al. 2003).

A ~13.5 kb *P. ananatis* LMG20103 cluster carries 11 *wb* genes (Fig. 4.16; PANA_2482-2492) and is flanked by the ananatan biosynthesis gene *galE* and

gluconate dehydrogenase gene *gnd*. Similarly the *wb* loci of several *E. coli* and *Salmonella* strains occur between *gnd* and colanic acid biosynthesis genes (Kido et al. 1995). The LMG20103 *wb* locus has an average G+C content of 37.66% (~16% below genome average) and occurs on the predicted horizontally acquired island HAI18 (Chapter 3). Similarly, the O-antigen loci of several *E. coli* and *Shigella* strains have G+C contents ~20% below genome average, indicating that O-antigen diversity has arisen through horizontal acquisition of *wb* genes (Liu et al. 2008).

PANA_2491-2492 encode homologues of the GDP-mannose precursor biosynthetic enzymes ManB and ManC (Jayaratne et al. 1994). Mannose is incorporated in the O-antigens of many bacteria including the mannan homopolymer of *K. pneumoniae* O3 and in the side chains of the *S. enterica* O6 heteropolymer O-antigen (Fitzgerald et al. 2003). Downstream of *manBC* are two genes encoding homologues of the ABC transporter Wzm (PANA_2490) and ATP-binding permease Wzt (PANA_2489) with 54% aa identity to Wzm-Wzt in *K. pneumoniae* O5 (AAF04380-04381). Orthologues are involved in the export of cell-surface glycoconjugates including glycoproteins, capsular polysaccharides and O-antigens (Cuthbertson et al. 2007). BlastP analysis with the *E. coli* K-12 Wzx/Wzy proteins indicate that most close phylogenetic relatives of *P. ananatis* for which whole genome sequences are available, including *Pantoea* and *Erwinia* species, make use of the Wzx-Wzy pathway for O-antigen export, revealing an unusual means for O-antigen export in *P. ananatis* LMG20103.

Downstream of *wzm* and *wzt* two CDSs encode homologues to the D-rhamnose biosynthetic enzymes Rmd (PANA_2484 – 61% identity - *D. dadantii* Dd703_3276) and Gmd (PANA_2485 – 86% aa identity - *D. dadantii* Dd703_3277). This rare sugar is found in a limited subset of bacteria as part of their cell surface glycans, including the O-antigens of *X. campestris* and *P. syringae* (Kneidinger et al. 2001). *P. aeruginosa* produces two types of O-antigens, A-band and B-band. While the B-band heteropolymer differs among *P. aeruginosa* serovars, the A band is antigenically conserved and consists solely of D-rhamnose (de Kievit and Lam, 1997). The GDP-mannose dehydratase Gmd shows a wide distribution and also catalyses the first step in L-fucose, D-talose and D-perosamine synthesis (King et al. 2009). Rmd, which reduces the Gmd product GDP-6-deoxy-D-*lyxo*-hexos-4-ulose to D-rhamnose is however restricted in distribution (King et al. 2009). Downstream of *gmd* and *rmd*,

151

two genes (PANA_2482-2483) encode homologues to the O-antigen mannosyl-transferases of *E. coli* O9 (WbdB and WbdC ACV53841-2 – 56% average aa identity) and the A-band rhamnosyltransferases in *P. aeruginosa* PAO1 (WbpY and WbpZ AAC38771-2 – 47% average aa identity). Similar homology occurs between the *E. coli* O99 mannosyl- and *P. aeruginosa* PAO1 rhamnosyltransferases and a conserved motif has been identified in these glycosyltransferases (Fig. 4.17; Geremia et al. 1996; Rocchetta et al. 1999). The presence of *gmd* and *rmd* in *P. ananatis* LMG20103 suggests a D-rhamnose O-antigen similar to the *P. aeruginosa* A-band. WbdC/WbpZ add a mannose or rhamnose subunit respectively to an *N*-acetylglucosamine attached to a undecaprenyl-phosphate lipid carrier (Und-P). WbdB/WbpY then adds two mannose/rhamnose subunits to the first.

A third enzyme (WbdA/WbpX) adds a final mannose/rhamnose to the polymer (Rocchetta et al. 1999). No WbdA/WbpZ homologue is present in LMG20103. Instead, adjacent to *rmd*, PANA_2846 encodes a glycosyltransferase with 39% aa identity to *Burkholderia glumae* BGR1 Bglu_lg06540. The presence of a conserved motif (Fig. 4.17) suggests it transfers a mannose/rhamnose sugar. Given the similarity between PANA_2482/WbpZ/WbdC and PANA_2483/WbpY/WbdB, it is plausible that the PANA_2486 protein transfers the last sugar subunit in the *P. ananatis* O-antigen. The Wzx-Wzy pathway makes use of the Wzz protein to regulate O polysaccharide chain length. O-antigen length is an important determinant of heterogeneity and is imperative for effective LPS function (Carter et al. 2007). The absence of Wzz in the ABC-dependent pathway may be resolved by introduction of an *O*-methylated sugar as the O-antigen terminal residue, terminating chain elongation (Rocchetta et al. 1999). The PANA_2488 protein contains a conserved *S*-adenosyl-methionine (SAM) domain (cl12011), which is involved in methylation phospholipids, proteins and sugars (Chiang et al. 1996), suggesting PANA_2488 may introduce a methyl group into the last sugar of the *P. ananatis* O-antigen chain.

The *P. ananatis* LMG20103 *wb* cluster suggests synthesis of an unusual O-antigen. The presence of genes for biosynthesis of the rare D-rhamnose and ABC-dependent export indicate an O-antigen unlike any found in other *Enterobacteriaceae*. Rather, the *P. ananatis* LMG20103 O-antigen appears to be similar to those found in the phytopathogens *P. syringae* and *X. campestris* and the opportunistic human pathogen

P. aeruginosa (Rocchetta et al. 1999). During chronic *P. aeruginosa* infections the B-band O-antigen expressed in early stages of infection, is selectively lost and the A-band antigen becomes dominant. This adaptive alteration provides a means to evade host defenses and survive within the host (Dasgupta and Lam, 1995). The potential D-rhamnose homopolymer may allow *P. ananatis* LMG20103 to evade plant defenses. Unlike *P. aeruginosa* however, no B-band O-antigen has been observed. Conversely, three genes, PANA_2266-2269 encode homologues to the RmlACD enzymes (*Pectobacterium wasabiae* Pecwa_3016-3018 – 71% average aa identity) which encode the L-isomer rhamnose found in heteropolymers of *S. flexneri* 2a and *S. marcescens* O4 (Saigí et al. 1999). *P. ananatis* LMG20103 may thus alternate between an L- and D-isomer rhamnose O-antigen.

Capsular and Extracellular polysaccharides

Capsular or exopolysaccharides are common structural features that encapsulate bacterial cells, create an advantageous environment and protect the bacterium against dessication, osmotic stress and host defenses (Bereswill and Geider, 1997). A direct role in pathogenesis has also been ascribed to several exopolysaccharides (Kim and Geider, 2000). Analysis of the *P. ananatis* LMG20103 genome revealed the presence of genes for the biosynthesis of four distinct exopolysaccharides. The genetic and structural homologies and potential pathogenic functions of these polysaccharides are discussed below. Another set of genes shows homology to those involved in linear β-glucan synthesis. While this is not a true exopolysaccharide it is also discussed.

Ananatan exopolysaccharide

P. stewartii and *E. amylovora* produce the exopolysaccharides stewartan and amylovoran, respectively. These show genetic and structural similarity to *Escherichia* group I colonic acid capsules and consist of sugar repeats of galactose, glucuronic acid, glucose and pyruvate residues that form a loosely attached capsule around the cells (Minogue et al. 2002). The stewartan and amylovoran biosynthetic pathways have been extensively characterised and involve twelve proteins, CpsA-L and AmsA-L respectively. Twelve protein-coding genes with homology to stewartan/amylovoran genes were identified in *P. ananatis* LMG20103 and are likely involved in the biosynthesis and export of a similar exopolysaccharide which we have termed ananatan. In this section the sequence homology and a putative model for ananatan

biosynthesis are discussed. As stewartan and amylovoran constitute major pathogenicity determinants in *E. amylovora* and *P. stewartii*, structural and functional aspects of the ananatan exopolysaccharide were analysed and are discussed in detail in Chapter 6.

The twelve ananatan biosynthetic genes PANA_2495-2506 were termed *anaA-L* in accordance with the stewartan biosynthetic genes (*cpsA-L*). AnaA-L show 95% (AAO05910-21 - 12 of 12 CDSs) and 67% (CAA54879-90 - 11 of 12 CDSs) average aa identity to stewartan (Cps) and amylovoran (Ams) biosynthetic proteins, respectively. Eight Ana proteins also show 49% average aa identity to *E. coli* K-12 colanic acid enzymes (Table 4.6). The gene order is conserved between the ananatan and stewartan loci, but limited synteny is observed between the ananatan and colonic acid loci (Fig. 4.18). This suggests ananatan genes synthesise an exopolysaccharide highly similar to stewartan in a similar fashion to *P. stewartii* (Fig. 4.19). Adjacent to the ananatan cluster, PANA_2493 and 2494 encode GalE and GalF homologues with 97% and 82% aa identity to GalEF in *P. stewartii* DC283 (ACV0284082-0284083) and *E. amylovora* ATCC49946 (EAM_2161-2162). GalE and GalF convert glucose to galactose, the principle amylovoran and stewartan sugar (Metzger et al. 1994). The proximal position and homology in GalEF indicate similar galactose predominance in ananatan. PANA_2506 encodes a homologue to the CpsA/AmsG glycosyltransferases which add a galactose molecule to a phosphorylated undecaprenyl lipid carrier. Subsequent sugars are sequentially added to stewartan and amylovoran by CpsG, -D, -K, -E, –F and -J and AmsD, -C, -K, -B, -E and –J, respectively (Geider, 2000). PANA_2496 (AnaK), PANA_2497 (AnaJ), PANA_2499 (AnaG), PANA_2500 (AnaF), PANA_2501 (AnaE) and PANA_2502 (AnaD) show significant homology to the *P. stewartii* and *E. amylovora* glycosyltransferases. In keeping with the stewartan and amylovoran biosynthetic pathways, AnaA, -G and -D likely add sugars to the backbone, while AnaK, -E, -F and -J add side chain sugars (Fig. 4.19; Geider, 2000). Subsequently the sugar repeats are transferred across the inner and outer membrane by a Wzabc complex in a similar fashion as the LPS O-antigen (Drummelsmith and Whitfield, 1999). The inner membrane protein Wzc transfers the repeat units to the outer membrane protein Wza which transports them across the outer membrane. Wzb, an acid phosphatase, recycles the lipid carrier for further EPS production (Drummelsmith and Whitfield, 1999). PANA_2503-2505 encode Wzabc homologues.

PANA_2495 shows homology to AmsL/CspL which are also involved in sugar repeat unit export, but their exact function is not known. At the outer surface, the sugar repeats are polymerised into the growing EPS by the CpsH/AmsF polymerase (Geider, 2000). PANA_2498 (AnaH) encodes a homologous polymerase.

Group 4 capsule

Group 4 capsule (G4C) is produced by several enteric bacteria including *E. coli, Shigella* and *Salmonella* species.G4C, also known as O-antigen capsule, is identical in structure to the LPS O-antigen and utilises the same Wzy-Wzy-dependent biosynthetic pathway (Peleg et al. 2005). G4C biosynthetic mutants exhibit enhanced O-antigen production due to shunting of repeat units to the O-antigen synthesis pathway (Shifrin et al. 2008). G4C biosynthesis in *E. coli* is regulated by the Rcs phosphorelay system (Ferrières and Clarke, 2003). G4C capsule extends 0.1-1 μm from the cell surface (Shifrin et al. 2008). As shorter antigens such as intimins, antigen 34 and AIDA-I do not extend as far from the bacterial surface, they are masked by G4C. The FimH adhesin of type 1 fimbriae may become coated with capsular material rendering them non-functional (Schembri et al. 2004). The Type III secretion system is also masked by G4C (Shifrin et al. 2008). It may thus play a role in early infection, where cell surface components that can be recognised by the host and trigger defense responses are masked to avoid detection (Shifrin et al. 2008).

E. coli G4C consists of O-antigen polymers secreted and assembled through an operon encoding 7 proteins, GfcABCDE-Etk-Etp (Shifrin et al. 2008). PANA_0242-0245 show 61% average identity aa identity to the capsule assembly proteins GfcABCD in *E. coli* K-12 respectively (Fig. 4.20). The *E. coli* Etk, Etp and GfcE proteins show homology to colonic acid capsule export proteins Wzc, Wzb and Wza respectively (Peleg et al. 2005). BlastP analysis with the *E. coli* K-12 Etk, Etk and GfcE revealed high homology to the ananatan export proteins AnaCIB (Fig. 4.20; PANA_2503-2505 – 59% average aa identity). No further homologues could be found in the *P. ananatis* LMG20103 genome suggesting AnaCIB may be involved in the biosynthesis of both ananatan and group 4 capsule. A plausible hypothesis is that *P. ananatis* responds to an environmental signal early in infection and via the Rcs phosphorelay triggers deployment of AnaCIB (Wzabc) for G4C synthesis, allowing the bacterium to escape recognition by the host and its associated defense responses.

Once established in the host the Rcs phosphorelay could shunt AnaCIB towards ananatan biosynthesis to allow attachment, host cell disruption and vascular occlusion (Fig 4.21).

Rhamnolipid

P. aeruginosa produces extracellular glycolipids consisting of L-rhamnose and 3-hydroxyalkanoic acid (HAA). These rhamnolipids are biosurfactants that enable the bacterium to overcome surface tension surrounding the cells and swarm for efficient surface colonisation (Déziel et al. 2003). Rhamnolipid synthesis involved three proteins. RhlA synthesises HAA and RhlB transfers TDP-L-rhamnose to HAA. The RhlC rhamnosyltransferase adds a second rhamnose, converting mono- to di-rhamnolipids (Soberón-Chávez et al. 2005). PANA_0754 encodes a protein with 30% and 36% aa identity to the *P. aeruginosa* (PA3479) and *D. dadantii* (Dd703_2296) RhlA proteins, respectively. PANA_0753 shows 31% and 30% aa identity to RhlB in *D. dadantii* (Dd703_2480) and *P. aeruginosa* (PA3478), respectively. No RhlC homologue is present in *P. ananatis* or *D. dadantii* suggesting they only produce monorhamnolipids. The presence of homologues to the quorum sensing components RhlIR which coordinate expression of *rhlAB*, supports the argument for rhamnolipid production by *P. ananatis* (Ochsner et al. 1995). Rhamnolipids may provide a means for *P. ananatis* LMG20103 to overcome surface tension at plant interfaces and colonise host tissues.

Cyclic and linear β-glucans

Cyclic glucans are glucose homopolymers produced by a limited number of plant-associated bacteria and play a role in both symbiotic and pathogenic plant-microbe interactions (Breedveld and Miller, 1994). In *A. tumefaciens*, two proteins are required for cyclic glucan biosynthesis. The ChvB (**ch**romosomal **v**irulence) protein is responsible for initiation, elongation and cyclisation of the glucan, while ChvA is involved in export of cyclic glucans to the periplasm (Breedveld and Miller, 1994).

PANA_4011 encodes a 2,867aa, 321 kDa protein which shows 47% and 37% identity respectively to NdvB of *X. campestris* pv. *campestris* ATCC 33913 (XCC2055) and ChvB of *A. tumefaciens* C58 (Atu2730). No homologous proteins are present in *P. stewartii* or *Erwinia* spp. A neighbour-joining tree based on a clustalW alignment of available ChvB/NdvB aa sequences (Fig. 4.22) indicates two branches, with NdvB

homologues in tumourigenic plant pathogens and plant-nodulating bacteria forming one branch while PANA0875 clusters with the NdvB proteins of the plant-associated *Enterobacteriaceae*, *X. campestris* and *P. putida*. This may be correlated to the roles of the cyclic glucans in one branch in tumour development and nodulation, while PANA_4011 may have a similar role in suppressing local and systemic plant host defenses as the cyclic glucan of *X. campestris*. The *P. ananatis* β-1,2-glucan may thus compensate for its lack of a Type III secretion system by serving a similar purpose. No NdvA/ChvA homologue was found in *P. ananatis*, nor could homologues be found in other NdvB-positive bacteria in the PANA_4011 clade (Fig. 4.22), suggesting an alternative means for secretion into the periplasm or that they secrete cyclic glucans into the extracellular environment by an as yet undefined mechanism.

Linear β-1,2-glucans or membrane-derived oligosaccharides (MDOs) are envelope constituents found in most Proteobacteria that consist of branched β-1,2-linked homopolymers of 5-12 D-glucose residues (Bohin, 2000). Two main enzymes are involved in synthesis of MDOs. The inner membrane glucosyltransferase MdoH forms linear polyglucose chains, while the periplasmic MdoG polymerises and adds branches to the linear glucan backbone (Bohin, 2000). PANA_1439-1440 encode homologues with 76% average aa identity to *P. stewartii* DC283 (ACV-0281905-0281906) and *E. coli* K-12 (CAA45521-45522) MdoGH. Other proteins add side chains and substitutions. MdoC transfers succinyl groups from the cytoplasm to the glucan backbone in the periplasm, while MdoB transfers phosphoglycerol (Bohin, 2000). PANA_1438 encodes a protein with 83% aa identity to MdoC in *P. stewartii* (ACV-0281904) indicating the *P. ananatis* linear glucan is succinylated. The PANA_0602 protein shares 62% identity with *E. coli* K-12 MdoB (P93401). The *P. ananatis* MDO thus contains a phosphoglycerol substitution while no MdoB homologue is present in *P. stewartii* DC283. The periplasmic protein MdoD controls glucan backbone size (Lequette et al. 2004). PANA_2621 shows 97% aa identity to MdoD in *P. stewartii* DC283 (ACV-0281354). PANA_3472 encode an MdoD paralogue with 42% aa identity to PANA_2621. MDOs play a role in phytopathogenicity. *P. syringae mdo* mutants are reduced in virulence. *D. dadantii mdo* mutants are reduced in exoenzyme production and avirulent on potato tubers, as MDOs modulate exoenzyme gene expression (Bouchart et al. 2007). The *P. ananatis* MDOs may thus play a role in phytopathogenesis.

Phytotoxins and Phytohormones

Toxins represent an integral part of the pathogenic arsenal of some phytopathogens (Bender et al. 1999). Analysis of the *P. ananatis* LMG20103 genome showed no homologues to genes encoding any *P. syringae* phytotoxins, *P. atrosepticum* coronofacic acid or *E. amylovora* DHP are present. However, PANA_0531-0532 show homology to the rhizobitoxin biosynthetic genes *rtxE and -F* in the legume symbiont *Bradyrhizobium elkanii* (33% average aa identity - BAB55903-55904). Rhizobitoxin inhibits ethylene defenses in host plants for efficient *B. elkanii* nodulation (Yasuta et al. 2001). Homologues to the main Rhizobitoxin structural component RtxA and desaturase RtxC are however absent from *P. ananatis*. PANA_0531-0532 are located on HAI4 with encoded proteins in this island sharing homologues in plant-associated bacteria only. *P. ananatis* LMG20103 may thus have horizontally acquired biosynthetic genes for a phytotoxin, but its structure, genetics and function would need to be elucidated.

Although no role for haemolysins in plant pathogenesis has been described, four *P. ananatis* CDSs show significant homology to enzymes in vertebrate pathogens with established haemolytic and pathogenic effects. PANA_4024 encodes a protein with significant homology to *Providencia rettgeri* haemolysin Hlx (PROVRETT_05838 – 45% aa identity). PANA_1278 shows homology to the HlyF haemolysin found in human pathogenic *E. coli* (AAG43374 – 52% aa identity; Morales et al. 2004), while the PANA_3183 protein is similar to HlyIII in *Vibrio vulnificus* (AAP50516 – 57% aa identity; Chen et al. 2004). The PANA_3513 protein shows 61% aa identity to the HlyA haemolysin produced by the opportunistic human pathogen *Aeromonas hydrophila* (ABD59012). Both HlyA and the PANA_3183 protein contain two cystathionine B synthase (CBS) domains which are involved pore formation in host cell membranes resulting in cell lysis (Erova et al. 2007). BlastP analysis with the four putative *P. ananatis* haemolysins indicates homologues are also found in plant-associated bacteria, suggesting they may have cytolytic effects on plant cells. The lack of genes encoding established phytotoxins however suggests phytotoxins do not make a significant contribution to the plant diseases caused by *P. ananatis* LMG20103.

Some tumourigenic pathogens induce the development of cankers by causing a hormonal imbalance in host plants through the production of the plant hormones

indole acetic acid (IAA) and cytokinin (Barash and Manulis-Sasson, 2007). BlastP with the *P. agglomerans* pv. *gypsophilae* IAA biosynthetic proteins IaaM and IaaH revealed the absence of homologues to these two enzymes in *P. ananatis* suggesting that, unlike its close phylogenetic relative, it does not produce gall-inducing IAA. However, a homologue to indole pyruvate decarboxylase IpdC (PANA_2745: 76% aa identity to *P. agglomerans* 299R AAB06571), an enzyme involved in the IpyA (indole pyruvate) pathway utilised for the production of IAA by many epiphytic bacteria, is present. No homologue is present for the aminotransferase required in the initial step of the IpyA pathway (Fig. 4.23) The absence of this enzyme could be overcome by two tryptophanases, encoded by PANA_0116 and PANA_3984 (43% aa identity to each other), which may convert L-tryptophan to indole and pyruvate that could then be shunted to IpdC for degradation to IAA. BlastP with the cytokinin biosynthetic enzymes Etz of *P. agglomerans* and Ipt of *R. solanacearum* showed no homologues are present in *P. ananatis* LMG20103. It can be deduced that as necrogenic pathogen, *P. ananatis* LMG20103 does not produce the IAA or cytokinins required to induce tumour formation in its *Eucalyptus* host. Homologues to enzymes involved in IpyA pathway suggest that it does produce IAA which may potentially improve its epiphytic fitness on plant surfaces.

Ice nucleation

A limited number of plant-associated bacteria belonging to the genera *Pseudomonas, Xanthomonas* and *Pantoea* produce ice nucleation proteins (InaA) which may result in frost injury and wounding, providing a site of entry for ice-nucleating phytopathogens (Lindow et al. 1982). PANA_0591 encodes a 1,306 aa ice-nucleation protein, showing 94% aa identity to *P. ananatis* KUIN-3 InaA (Q47879). Comparison of available InaA sequences (Fig. 4.24) shows PANA_0591 clustering with other *Pantoea* and *P. syringae* InaA homologues. *P. ananatis* LMG20103 may thus utilise the InaA protein to induce damage on aerial surfaces of *Eucalyptus* and hence gain entry into the plant. The PANA_0591-encoded InaA protein contains conserved domains typical of substrate proteins secreted by the Type I secretion system, suggesting InaA is secreted in this manner.

Plant cell wall degradation

Many phytopathogenic bacteria secrete cell wall degrading enzymes (CWDEs) such as cellulases, xylanases, pectinases and proteases that allow them to break down plant cell wall polymers. This enables the pathogen to penetrate and progress through plant tissues and releases nutrients (Toth et al. 2003). *P. ananatis* LMG20103 carries ten CDSs encoding putative CWDEs that may degrade plant cell wall cellulose, hemicellulose, pectin and glycoproteins (Table 4.7). Genes encoding components for degradation product transport and metabolic pathways are also present. This suggests *P. ananatis* can utilise CWDEs to attack its hosts and use the degraded products as nutrient source. However, genes encoding key enzymes in plant cell wall degradation, including pectin lyases produced by many soft rot pathogens, are absent, suggesting a subsidiary role for *P. ananatis* CWDEs.

Pectin degrading enzymes

Pectobacterium and Dickeya spp. produce an arsenal of enzymes that systematically degrade pectin polymers in the plant cell wall, catalysing plant cell bursting and tissue maceration (De Boer, 2003). The first enzyme, methylesterase, removes methyl groups (Hugouvieux-Cotte-Pattat et al. 2002). No methylesterase homologues are present in *P. ananatis*. Acetyl groups are removed by pectin acetylesterases PaeX and PaeY to liberate polygalacturonate (Shevchik and Hugouvieux-Cotte-Pattat, 2003). No PaeY homologue is present, but the PANA_0136 protein shows 55% aa identity to *D. dadantii* PaeX (CAD45188.1). The *Dickeya* PaeX enzyme has a signal peptide for extracellular secretion via the T2SS. SignalP analysis indicates the PANA_0136 protein lacks a signal peptide and absence of a T2SS suggests a different means of secretion or intracellular function.

Subsequently polygalacturonate is degraded by a series of pectate lyases and polygalacturonases, releasing oligogalacturonates that are taken up by the bacterium (Roy et al. 1999). No pectate lyase homologues exist in the genome, correlating with the absence of soft rot symptoms in plants infected with *P. ananatis*. However, PANA_1095 and PANA_0324 encode two putative polygalacturonases. The former protein shows 49% aa identity to the Type II secreted endopolygalacturonase Peh1 of *P. atrosepticum* (M83222.1). The presence of a signal peptide suggests a periplasmic localisation. *Agrobacterium vitis* also produces a PehA polygalacturonase but lacks a

T2SS and significant periplasmic PehA activity was measured (Herlache et al. 1997). PANA_0324 shows 86% identity to a *P. atrosepticum* polygalacturonase (ECA3552). No signal peptide is present, indicating cytoplasmic localisation. As homologues to PANA_0324 and PANA_1095 are restricted to plant-associated bacteria they likely play a role in degradation of plant-derived pectic polymers. However, the lack of homologues to pectate lyases suggests pectin degradation is not a major pathogenicity determinant in *P. ananatis*. Rather, it may utilise oligogalacturonates released by pectin lyases of other phytopathogens inhabiting the same plant niche.

Cellulases and cellulose degradation

D. dadantii produces two cellulases, Cel8Y and Cel5Z which act in synergy to degrade cellulose. Cel8Y catalyses degradation of long cellulose molecules before Cel5Z converts them into dimers and trimers that are transported into the bacterial cell by a phosphoenolpyruvate-dependent pathway and hydrolysed by β-glucosidases (Zhou and Ingram, 2000). PANA_0132 encodes a putative cellulase showing 57% aa identity to *D. dadantii* to Cel8Y (AAA24818). It does not contain a signal peptide and Psort analysis indicates a periplasmic localisation. No homologue to *cel5Z* is present in the LMG20103 genome. Cel5Z mutation in *D. dadantii* had no effect on virulence, while Cel8Y is necessary early in infection (Boccara et al. 1994). The PANA_2614 CelY homologue may thus degrade long cellulose polymers in host plant cell walls.

Xylanases and hemicellulose degradation

Phytopathogens produce endoxylanases that break β-1,4-xylosidic bonds in the hemicellulose xylans, releasing xylo-oligosaccharides that are transported into the cell where β-xylosidase hydrolysis releases free sugars. Removal of substitutions requires arabinofuranosidase and acetylesterase enzymes (Tsujibo et al. 2001). PANA_0376 encodes a putative endoxylanase with 43% identity to *Acidobacterium capsulatum* XynA (BAB40957). A Tat signal peptide is present, indicating translocation to the periplasm. PANA_3781 and PANA_2623 also encode proteins with homology to endoxylanases. Similarity to the PANA_0136-encoded pectin acetylesterase (PANA_3781: 32% aa identity; PANA_2623: 31% aa identity) suggests however they may remove acetyl side chains from xylan to facilitate efficient backbone hydrolysis by the PANA_0376 endoxylanase. A 26 aa signal peptide in PANA_3781, absent in PANA_2623 suggests periplasmic or extracellular function, while the PANA_2623

protein is cytoplasmic. PANA_2594 encodes a protein with 47% aa identity to the *Streptomyces lividans* AbfA (AAA61708) which removes L-arabinofuranosyl side chains from the xylan backbone and converts them to L-arabinose (Margolles and de los Reyes-Gavilán, 2003). Lack of a signal peptide indicates cytoplasmic localisation for the PANA_2594 protein. Another homologue is encoded by PANA_0739, with 53% aa identity to *Cellvibrio japonicus* AbfA (CJA_3012). Psort analysis suggests the PANA_0739 protein is secreted by an unknown secretory mechanism.

Proteases

Phytopathogen proteases also partake in degradation of plant cell wall constituents. PANA_2851 encodes a homologue of *Pectobacterium carotovorum* metalloprotease Prt1 (AAA28458 – 66% aa identity), which induces degradation of plant tissues and releases amino acids and peptides which serve as nutrients (Kyöstiö et al. 1991).

Transport and catabolism of plant cell polymer products

Following plant wall polymer degradation, intermediates are transported into the cell by carbohydrate transport systems and degraded into metabolisable sugars by means of periplasmic and cytoplasmic glucosidases (Postma et al. 1993). Six putative cytoplasmic and two periplasmic β-glucosidases, and five cytoplasmic α-glucosidases are encoded by *P. ananatis* LMG20103. The periplasmic glucosidases encoded by PANA_1246 and PANA_1953 show 77 and 44% aa identity to *E. coli* BglX (ABY91285) and *Streptomyces thermoviolaceus* β-xylosidase BxlA (BAD02389), respectively. BxlA degrades xylan intermediates to release xylose sugars (Tsujibo et al. 2001). The PANA_3309-encoded α-glucosidase shares 72% aa identity with *P. carotovorum* YicI (ZP_03825495) which also degrades xylan polymers (Lovering et al. 2005).

PANA_1567-1569, PANA_4222-4225 and PANA_0061-0064 show homology to components of three carbohydrate transport systems in *P. carotovorum* (Hong et al. 2006). PANA_0929-PANA_0933 are homologous to the CelABCDF transporter involved in cellulose intermediate transport (Yang et al. 1996). Two specialised transport systems are involved in import of pectin intermediates. The ExuT pathway imports galacturonates and glucuronates and the KdgT pathway transports pectin catabolic intermediates 2-keto-3-deoxygluconate (KDG), 5-keto-4-deoxyuronate (DKI) and 2,5-diketo-3-deoxygluconate (DKII) (San Francisco et al. 1996).

162

PANA_0750 and PANA_3457 show 44% and 88% aa identity to *D. dadantii* KdgT (ABF-0019248) and ExuT (ABF-0018695) transporters respectively. PANA_0540 and PANA_3310 encode homologues to PANA_0750 (45 and 49% aa identity respectively) indicating more than one ExuT transporter is present. Inside the cell galacturonate delivered by ExuT is converted into KDG via the UxaABC galacturonate catabolism pathway. KduI and KduD also convert DKI to DKII and DKII to KDG respectively. KDG is converted into utilisable pyruvate and 3-phosphoglyceraldehyde by KdgK and KdgA (San Francisco et al. 1996). PANA_3454-3456 encode proteins with 71% average aa identity to UxaABC in *P. atrosepticum* (ECA0645-0647). PANA_0582 encodes a KduI enzyme and the adjacent gene PANA_0581 encodes a KduD homologue. KdgK homologues are encoded by PANA_0075 and PANA_0138 while a KdgA is encoded by PANA_2190.

Secretion systems

Gram-negative plant and animal pathogens use specialised protein secretion systems to deliver pathogenicity factors across the inner and outer membranes of the cell (Lavander et al. 2006). Six distinct secretion systems (Type I to VI) have been described. Type I, III, IV and VI transport pathogenicity proteins directly across both inner and outer membranes while Type II and V pathways require an intermediate translocation across the inner membrane. Inner membrane translocation can occur by two means, the general secretory pathway (Sec) or via the Twin Arginine Translocation (Tat) pathway (Lavander et al. 2006). Analysis of *P. ananatis* LMG20103 genome revealed the presence of Type I, V and VI secretory pathways, as well as a Sec and Tat system. No functional Type II, III and IV secretory machineries are present. This is significant as these play vital roles in pathogenesis in many closely related plant and animal pathogens.

The general secretory (Sec) pathway

The Sec pathway transports unfolded precursor proteins with an amino terminal signal peptide from the cytoplasm to the periplasm through an ATP-driven translocase complex (Fig. 4.25; Mori and Ito, 2001). Ten proteins are involved in Sec translocation and homologues to each of these proteins are encoded on the *P. ananatis* LMG20103 genome, sharing 97 and 83% average aa identity to the Sec components in *P. stewartii* DC283 and *E. coli* K-12, respectively. Three signal peptidases remove

the signal peptides from substrate proteins prior to Sec translocation. Signal peptidase I (SpI), encoded by *lepB* (PANA_2894) removes the leader peptide from most proteins, while signal peptidase II (SpII) cleaves the leader peptide from lipoproteins (*lspA* − PANA_0670). Prepilin peptidases (SpIII) (PANA_3625) remove the signal sequence from type 4 fimbrial proteins (Stathopoulos et al. 2000). SignalP analysis (Chapter 2) indicated 575 *P. ananatis* LMG20103 CDSs contain signal peptidase. This constitutes 13.6% of the total proteins encoded on the genome and includes many putative *P. ananatis* pathogenicity determinants, including proteins involved in fimbrial and flagellar biogenesis, iron acquisition, cell wall degradation, protection against plant defenses, Type V secreted proteins and Type VI secretion machinery proteins as well as the ananatan exopolysaccharide biosynthetic protein AnaB. This suggests an important role in pathogenesis for the Sec translocation pathway in *P. ananatis* LMG20103.

The Twin-arginine translocation pathway

The Tat pathway transports a specific set of folded, co-factor bound proteins (Mori and Ito, 2001). The Tat system consists of TatA, B and C which are encoded in a single operon. PANA_0193-0195 show 88% average aa identity to TatABC *P. stewartii* DC283 (ACV-0283443-0283445) (Fig. 4.26). A role for the Tat pathway in pathogenesis has been described for several phytopathogenic bacteria, including *P. syringae* and *A. tumefaciens* (Bronstein et al. 2005; Ding and Christie, 2003). Twenty seven Tat substrate proteins were identified in *P. ananatis* LMG20103 using TatP 3.0 and the Tatfind 1.4 server (Chapter 2; Http://signalfind.org/tatfind), encompassing 0.6% of the total protein content and 3.3% of all proteins translocated across the inner membrane in *P. ananatis*. This is comparable to the twenty six *E. coli* K-12 Tat substrates, with ten of these common to both. Several putative *P. ananatis* pathogenicity factors appear to be secreted via the Tat pathway. PANA_3781 encodes a putative endo-glucanase involved in cell wall degradation. Two MdoD orthologues involved in periplasmic glucan biosynthesis carry Tat signal peptides as does a type 1 fimbria minor pilin subunit protein (PANA_2716). FhuD is Tat secreted in *E. coli* K-12 (Palmer et al. 2005). A Tat signal peptide is also present in the *P. ananatis* homologue (PANA_1168) and in the AnaH protein (PANA_2498) involved in ananatan biosynthesis (Chapter 6). Given the importance of EPS, attachment through

fimbriae and iron acquisition in pathogenesis, the Tat pathway is likely a significant pathogenicity determinant in *P. ananatis* LMG20103.

Type I secretion system

The Type 1 secretion system (T1SS) translocates a number of important pathogenicity determinants in phytopathogenic bacteria (Delepelaire, 2004). The secretory apparatus comprises of three proteins, a membrane fusion protein, an outer membrane protein and an ATP-binding protein. A single T1SS is encoded by *P. ananatis* LMG20103 with PANA_2603-2605 sharing 94% and 69% average aa identity to the T1SS proteins in *P. stewartii* DC283 (ACV-0281370-2) and *P. atrosepticum* SCRI1043 (ECA1096-9), respectively (Fig. 4.27). Type 1 secreted proteins belong to the RTX (**R**epeats in **ToX**in) family (COG2931). They are typically large and acidic, with an isoelectric point below 4.0, have a low cysteine content and contain up to 50 glycine-rich repeats (Delepelaire, 2004). BlastP analysis identified two putative T1SS substrates encoded by PANA_2607 and PANA_0591, which show similarity to RTX family members (Table 4.8). These proteins are the largest and third largest encoded on the genome, respectively. PANA_2607 is located adjacent to the PANA_2603-2605 T1SS genes. On the basis of sequence homology it encodes a non-fimbrial adhesin, while PANA_0591 encodes an ice-nucleation protein whose means of secretion has not previously been identified.

Type II secretion system

The Type II secretion system (T2SS) secretes a number of pathogenicity factors including plant cell wall degrading enzymes (Hazes and Frost, 2008). T2SSs have been identified in the genomes of all phytopathogenic bacteria sequenced to date with the exception of *A. tumefaciens* C58. BlastP analysis with the *K. pneumoniae* Pul and *P. stewartii* Out T2SS proteins shows that seven proteins share homologues in *P. ananatis* LMG20103. However, BlastP with the *P. ananatis* homologues against the *K. pneumoniae* and *P. stewartii* genomes indicate closer homology to predicted Type IV fimbria biosynthetic proteins, suggesting a T2SS is missing. This raises questions on how the putative secreted CWDEs in *P. ananatis* LMG20103 are exported. The Type IV fimbria and flagellar biosynthetic system have been shown to function as secretory apparatus (Sauvonnet et al. 2000; Young et al. 1999) and may perform similar functions in *P. ananatis*.

Type III Secretion System

The Type III secretion system (T3SS) secretes a number of effector proteins which allow phytopathogenic bacteria to manipulate basal plant defenses and plays a role in determining the host range of a plant pathogen (Toth et al. 2006). Given its conserved nature among *Enterobacteriaceae* and phytopathogenic bacteria, a Type III secretion system was expected in *P. ananatis* LMG20103. BlastP against the *P. ananatis* LMG20103 genome with the T3SS apparatus and effector protein arrays of *P. stewartii* DC283, *P. atrosepticum* SCRI1043, *P. syringae* pv. *tomato* DC3000, *Y. pestis* CO92 and *S. enterica* LT2 revealed no homologues are present, demonstrating the atypical absence of a T3SS from *P. ananatis* LMG20103. T3SS loci show characteristics of pathogenicity islands and may have arisen in many phytopathogenic bacteria through horizontally acquired from other organisms (Alfano and Collmer, 2004). *P. ananatis* LMG20103 may not have come into contact with co-occurring bacteria carrying T3SS islands, or may have lost the capacity or requirement of Type III secretion. Two genes PANA_0751 and PANA_2458 encode homologues to HopAN1 in *P. syringae* pv. tomato T1 (EEB56950 - 26 and 45% aa identity respectively). Hop (**H**rp **o**uter **p**rotein) is the generic name given to those proteins secreted by the T3SS of phytopathogens (Lindeberg et al. 2005). The presence of these genes suggests a T3SS system was present in LMG20103 but may subsequently have been lost. Alternatively, the *hop* genes may have been acquired separately. The implications of not having a T3SS are that *P. ananatis* may have chosen to extract itself from the arms race with host plants and can therefore infect a wide range of hosts without basal defenses restricting their spread to novel hosts. It may further facilitate the existence of *P. ananatis* as endophyte and epiphyte on a wide range of hosts. The role of the T3SS in suppressing host defenses does however raise questions on how *P. ananatis* LMG20103 can circumvent the lack of a T3SS and fulfil this purpose. Production of cyclic glucans may preclude requirement of a T3SS, as they can suppress both basal and systemic plant defenses. This glucan is also produced by *A. tumefaciens* which is missing a T3SS (Rigano et al. 2007). *P. ananatis* may also be able to protect itself against host defenses. Several protective measures against plant defenses are encoded on the *P. ananatis* genome.

Type IV secretion system

The Type IV secretion system functions in conjugal transfer of DNA between bacteria, but also in the transport of DNA and proteins from the bacterial cytoplasm into the cytoplasm eukaryotic cells (Ding et al. 2003). The archetypal T4SS is the VirB system in the phytopathogen *A. tumefaciens*, which exports a large single-stranded DNA into plant cells, where it is integrated into the plant genome and expresses oncogenes, resulting in uncontrolled cell division and formation of crown gall tumours. No homologues to any VirB proteins of *A. tumefaciens* C58, *P. stewartii* DC283 or *P. atrosepticum* SCRI1043 could be detected in *P. ananatis* LMG20103 indicating that it lacks a T4SS.

Type V secretion system

Type V secretion systems (T5SS) encompass autotransporters (AT) and two-partner secretion (TPS) pathways which secrete large pathogenicity proteins (Fig. 4.28; Henderson et al. 2004). Three putative AT and one TPS proteins are encoded by *P. ananatis* LMG20103 (Table 4.9; Fig. 4.29). One of AT proteins (PANA_1089) and the TPS protein (PANA_0907) carry two RGD motifs in their passenger domains which are associated with attachment to eukaryotic cells and thus likely function as non-fimbrial adhesins (Henderson et al. 2004). The PANA_906 protein shows 55% and 51% aa identity to the secretion partner CdiB in *E. coli* (AAZ57197) and HecB in *D. dadantii* (AAC31980), respectively. The homology and chromosomal co-localisation of PANA_0906 indicate a role as β-domain for Type V secretion of the PANA_0907 protein and may be involved in its activation as described for HecB and CdiB (Aoki et al. 2005).

The PANA_4212 protein contains two GDSL lipase/esterase family domains (Pfam00657) and an AT autotransporter domain (COG5571) (Fig. 4.29). BlastP analysis against the NCBI protein database indicates that homologues are largely restricted to plant-associated bacteria, but are also found in some *Salmonella* strains. Members of the GDSL lipase/esterase family break long-chain acylglycerols (lipases) and short-chain fatty acids esters (esterases) (Henderson et al. 2004). The PANA_4212 protein shows 53% and 31% aa identity respectively to the lipase/esterase EstA in the root associated *Serratia liquefaciens* (AAO38760) and Lip-1 in the entomopathogen *Photorhabdus luminescens* (CAA47020). EstA remains

167

associated with the outer membrane while Lip-1 is secreted (Riedel et al. 2003; Wang and Dowds, 1993). Lack of an autoproteolytic domain suggests the PANA_4212 protein remains associated with the bacterial surface. Although the function of many esterases and lipases remains unknown, some are associated with plant pathogenesis. Acetyl and cinnamoyl esterases participate in hemicellulose degradation, while the *B. subtilis* esterase has been shown to hydrolyse phytotoxins (Riedel et al. 2003). EstA in *S. liquefaciens* catabolises fatty acid esters and has been plays a role in synthesis of acyl homoserine lactones for quorum sensing (Riedel et al. 2003).

The 661 aa, 74 kDa protein encoded by PANA_2402 contains a phosphatidic acid phosphatase domain (Cl00474) (Fig. 4.29) and thus likely represents an enzyme capable of catalysing the hydrolysis of organic phosphomonoesters at an acidic pH. Substrates such as nucleotides and sugar phosphates are converted into inorganic phosphates and dephosphorylated product. This provides a means for the uptake of essential phosphate (Hoopman et al. 2008). PANA_2402 shows highest identity to predicted acid phosphatases in the plant-associated *B. cenocepacia* MCO-3 (Bcenmc03-6306 - 67% aa identity) and the phytopathogen *R. solanacearum* GM1000 (Rsc0131 - 64% aa identity). No homologues are found in the closely related phytopathogens *E. amylovora* or *P. stewartii* subsp. *stewartii* DC283. Acid phosphatases in intracellular pathogens have been demonstrated to affect host signaling pathways controlling the production of reactive oxygen species (Hoopman et al. 2008). The PANA_2402 acid phosphatase may thus have a role in interfering with oxidative burst defenses in host plants.

Type VI secretion

The recently identified Type VI secretion system (T6SS) is found in a number of plant and animal pathogens. This secretion system injects cytotoxins into host cells and plays a major role in pathogenesis (Yahr, 2006). Analysis of the *P. ananatis* LMG20103 genome revealed the presence of three gene loci with high homology to T6SSs in phylogenetic relatives. Given the absence of Type II, III and IV secretion systems in *P. ananatis* LMG20103, these Type VI secretion systems are expected to play a significant role in secretion of pathogenicity factors and in pathogenesis. For this reason an in depth *in silico* analysis of the three T6SS loci was performed which is discussed in Chapter 5.

Protection against plant defenses

Plants produce toxic substances such as antimicrobial peptides, phenolic acids, reactive oxygen species and proteases to protect themselves against infection by phytopathogenic bacteria. (Abramovitch and Martin, 2004). In order to successfully infect a host plant, a phytopathogen must be capable of overcoming these defenses. The *P. ananatis* LMG20103 genome encodes several proteins with homology to those involved in protection against antimicrobial agents, resistance to reactive oxygen species and phenolics acids in other phytopathogens.

Protection against plant-produced antimicrobial agents

Bacteria have developed a number of means to combat the detrimental effects of antimicrobial agents. β-lactamases can inactivate antimicrobials and multidrug efflux pumps can shuttle them out of the cell before they can take effect (Matsuo et al. 2008). Plants produce flavanoid, isoprenoid and alkaloid secondary metabolites known as phytoalexins which result in pathogen growth arrest and sequestration of additional defense mechanisms (Burse et al. 2004a). Means of opposing phytoalexins and other plant antimicrobials are thus necessary for host colonisation and infection. Several β-lactamases and efflux pumps with potential roles in inactivation and extrusion of plant antimicrobials are encoded by *P. ananatis*.

β-lactamases

Several predicted *P. ananatis* LMG20103 proteins show homology to β-lactamases which degrade β-lactam antimicrobial peptides. The PANA_1958 protein shows 79% aa identity to the *P. stewartii* DC283 AmpC cephalosporinase (ACV-0283455) which degrades a range of β-lactam peptides (Coudron et al. 2000). PANA_0502 shares 76% identity with the *Enterobacter* sp. 638 Vat protein (Ent638_0391). Vat inactivates antibiotics produced by *Streptomyces* spp. (Rende-Fournier et al. 1993) and may provide *P. ananatis* with a selective advantage over plant-associated *Streptomyces*. PANA_0239, PANA_4169 and PANA_4682 encode three putative β-lactamases with homologues restricted to plant-associated bacteria which may be involved in inactivation and degradation of plant-produced antimicrobials.

Multidrug efflux pumps

Multidrug efflux pumps belong to five transport families, the ATP-binding cassette (ABC) superfamily, resistance/nodulation/division (RND) family, multidrug and toxin
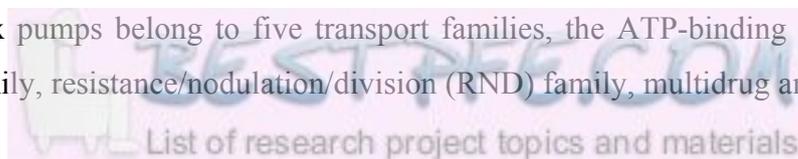
extrusion (MATE) family, major facilitator superfamily (MFS) and the small multidrug resistance (SMR) family (Saier et al. 1998). Thirty-four efflux pumps, encompassing all five families, are present in *P. ananatis* LMG20103 (Table 4.10).

Several predicted *P. ananatis* efflux pumps share homology to efflux systems employed by plant-associated bacteria. One ABC efflux pump, encoded by the Sap antimicrobial resistance locus (SapABCDF), is involved in resistance of *D. dadantii* to the potato antimicrobials snakin-1 defensin Pth-1 and α-thionin (López-Solanilla et al. 1998). PANA_2030-2034 show 73% average aa identity to the SapABCDF proteins of *D. dadantii* (CAA10909-10913) indicating that *P. ananatis* LMG20103 could employ this system for resistance against plant-produced antimicrobials. A role in phytopathogenesis has also been identified for several RND efflux systems. PANA_1021-1023 encode proteins with 81% average aa identity to the AcrABR phytoalexin efflux pump in *E. amylovora* (AAQ21214-21216). *E. amylovora acrB* mutants showed reduced tolerance to apple phytoalexins and diminished fire blight symptoms (Burse et al. 2004a). PANA_1741-1743 encode homologues to components of the YdhIJK fusaric acid RND efflux pump which exports the phytotoxin fusaric acid produced by *Fusarium* spp. (Utsumi et al. 1991). *Fusarium* spp. are associated with stem cankers and rot of *Eucalyptus* spp. (Roux et al. 2001) potentiating a role for *P. ananatis* LMG20103 as biological control agent. The majority of RND efflux pumps (eight of thirteen) share homologues only in plant-associated bacteria, including those encoded by PANA_1288-1290, PANA_1431-1433 and PANA_2736-2738 suggesting significant roles for RND efflux systems in *Pantoea* resistance to plant-produced antimicrobials. The MATE efflux protein NorM protein of *E. amylovora* mediates protection against antibiotics produced by *P. agglomerans*, a common co-inhabitant of pear and apple blossoms, thereby nullifying the competitive advantage of latter organism (Burse et al. 2004b). The PANA_1723 protein shows significant homology to *E. amylovora* NorM (76% aa identity - AAQ21212) while PANA_2374 is homologous to a second NorM efflux system in the *E. amylovora* genome (73% aa identity - EAM2133). These efflux pumps may play a similar role in *P. ananatis* LMG20103 resistance to antibiotics produced by co-inhabitants of the *Eucalyptus* phyllosphere.

Defense against plant oxidative burst and reactive oxygen species

Plants produce toxic reactive oxygen species (ROS) including superoxides (O_2^-), hydrogen peroxide (H_2O_2) and organic hydroperoxides as defense response to pathogen attack (Mongkolsuk et al. 1998). ROS are also produced during oxidative respiration by bacteria themselves (Sabri et al. 2008). Bacteria have evolved mechanisms to detoxify and repair damage associated with reactive oxygen species. H_2O_2 can be converted to O_2 and H_2O by catalases and peroxidases (Hanyu et al. 2009). Three putative catalases are encoded by *P. ananatis*. The PANA_0868 protein shows 78% aa identity to the catalase-peroxidase KatG (ECK3934), while PANA_1753 shares 77% identity with catalase KatE (ECK1730) in *E. coli* K-12. In *Mesorhizobium loti* KatG provides protection against H_2O_2 produced through metabolism, while KatE protects against plant-produced H_2O_2 (Hanyu et al. 2009). Homologues to PANA_1871 are restricted to a subset of plant-associated bacteria including *Erwinia* spp. and *P. stewartii* DC283. This suggests involvement in degradation of plant generated H_2O_2. Five peroxidases are also encoded on the genome. The *X. campestris* Ohr protein (AAC38562) also provides protection against plant-produced hydroperoxides and the PANA_0119 protein shows 64% aa identity to this protein (Mongkolsuk et al. 1998). Superoxides (O_2^-) can be removed by superoxide dismutases (Sod) linked to metal ion co-factors. The copper/zinc co-factored SodC eradicates superoxides produced by oxidative burst in plant hosts (Forest et al. 2000). The PANA_1739 protein shares 87% aa identity to *P. stewartii* SodC (ACV-0281027). The SufABCDSE locus also provides protection against plant-induced oxidative stress and *suf* locus mutants of *D. dadantii* were reduced in virulence (Nachin et al. 2001). PANA_1714-1719 encode proteins with 71% average aa identity to *D. dadantii* SufABCDSE (CAC17124-9).

Resistance to plant phenolics acids

Another defense mechanism deployed by plants is the release of phenolic acids which are toxic to phytopathogens (Uchiyama et al. 2008). Several plant-associated bacteria produce a phenolic acid decarboxylase (PadC) which converts phenolic acids into non-toxic derivatives (Uchiyama et al. 2008). *P. ananatis* LMG20103 encodes a PadC homologue (PANA_0344) showing 77% identity to the phylloplane epiphyte *Klebsiella oxytoca* PadC protein (BAF65031). *P. ananatis* LMG20103 may thus use this decarboxylase to degrade plant defense phenols.

Protection against plant proteases

Plant-produced proteases can breach the bacterial outer membrane. Some Gram-negative bacteria encode the efficient broad-spectrum protease inhibitor α-macroglobulin (Budd et al. 2004). α-macroglobulins are typified by the 1,636 aa *E. coli* periplasmic protein YfhM. The *yfhM* gene is located adjacent to *pbpC* which encodes a peptidoglycan repair enzyme (Doan and Gettins, 2008). When host proteases disrupt the pathogen LPS and peptidoglycan layers, YfhM and PBP1C form a complex, with YfhM trapping the proteases while PBP1C repairs cell membrane breaches (Budd et al. 2004). PANA_2864 encodes a 1,632 aa protein with 57% identity to YfhM in *E. coli* K-12 (ECK2516) and PANA_2863 shows 65% aa identity to PBP1C (ECK2515). These may aid *P. ananatis* in combating host proteases and repairing protease-damaged cell membranes.

Pathogenicity regulation

The effective colonisation and infection of a host plant requires tight regulation of its pathogenicity factors by a phytopathogenic bacterium (Mole et al. 2007). Pathogens have evolved specific transcriptional regulators to ensure appropriate expression of individual pathogenicity determinants. A broader "global" pathogenicity regulation ensures coordinate expression of multiple pathogenicity factors in response to environmental signals, enabling phenotypic adaptation for effective colonisation and disease induction (Mole et al. 2007). Several *P. ananatis* LMG20103 proteins show homology to known global pathogenicity regulators and are discussed below, with focus on the PhoPQ and the Rcs phosphorelay regulators which play major roles in pathogenesis in closely related animal- and phytopathogens.

Global pathogenicity regulation in P. ananatis LMG20103

Most global pathogenicity regulatory systems exist as two-component systems (TCS) consisting of an inner membrane histidine kinase and a cognate cytoplasmic response regulator that acts as positive or negative regulator for gene expression and consequent behavioural adaptation (Eguchi et al. 2004). TCSs interact with each other to produce a complex regulatory network for precise regulation of pathogenicity (Eguchi et al. 2004). Twenty-two distinct TCSs could be identified in *P. ananatis* with several showing homology to TCSs regulating a range of pathogenicity phenotypes in plant and animal pathogens. The PmrAB TCS regulates structural changes in LPS

resulting in resistance to cationic antimicrobials (Roland et al. 1993). PANA_2237-2238 show 67% average aa identity to PmrAB in *S. enterica* serovar Arizonae (SARI_03365-03366). The BvgAS TCS of *B. pertussis* controls expression of fimbriae, adhesins and biofilm biosynthetic genes (Irie et al. 2004). PANA_1196-1197 show homology to BvgAS (29% average aa identity - AAA22969-22970) and may allow *P. ananatis* to regulate adhesin biosynthesis. RssAB modulates swarming motility for rapid colonisation of host surfaces in *S. marcescens* (Wei et al. 2005). PANA_2095-2096 encode RssAB homologues (AAN28325-28326: 69% average aa identity) and could regulate swarming-associated colonisation of plant surfaces. PANA_2519-2520 and PANA_2968-2969 show homology to the BaeSR TCS that regulates several multidrug efflux pumps for resistance to antimicrobials (Nishino et al. 2007). PANA_3996-3997 show homology to a BaeS/R TCS limited to phytopathogens in the genera *Dickeya*, *Xanthomonas* and *Pseudomonas*. This TCS is found upstream of a multidrug efflux pump also restricted to plant-associated bacteria, suggesting it regulates this efflux pump for defense against plant antimicrobials.

The Rcs phosphorelay

The Rcs phosphorelay system regulates production of capsular and extracellular polysaccharides (EPS) including colanic acid in *E. coli*, amylovoran in *E. amylovora* and stewartan in *P. stewartii* subsp. *stewartii* (Castanié-Cornet et al. 2006). It consists of three proteins: the inner membrane hybrid sensor kinase RcsC, the histidine-phosphotransferase domain protein RcsD and the cytoplasmic response regulator RcsB (Castanié-Cornet et al. 2006). PANA_2609-2611 encode homologues with 70% and 93% average aa identity to RcsCBD in *E. amylovora* Ea273 (EAM_2261-2263) and *P. stewartii* DC283 (ACV-0281364-0281366), respectively. RcsC is activated by two independent proteins which respond to environmental signals, the inner membrane lipoprotein DjlA and outer membrane lipoprotein RcsF (Castanié-Cornet et al. 2006). The PANA_0687 shows 78% aa identity to DjlA in *E. amylovora* (EAM_0670) while PANA_0819 encodes a protein with 98% aa identity to *P. stewartii* RcsF (ACV-0284431). RcsF responds to cues including exposure to low temperatures, osmotic strength, dessication and changes in surface molecules such as LPS, peptidoglycan and periplasmic membrane-derived oligosaccharides (Mdo) (Castanié-Cornet et al. 2006).

For regulation of exopolysaccharide biosynthesis, RcsB requires the auxillary transcriptional activator RcsA, forming a complex that binds an RcsAB box promoter region upstream of the exopolysaccharide biosynthetic operon. PANA_2308 shows 99% aa identity to *P. stewartii* RcsA (ACV-0282841). In *P. stewartii* RcsAB boxes are located upstream of *cpsA* and *rcsA* (Carlier and von Bodman, 2006). BlastN analysis with the *P. stewartii* RcsAB box sequences revealed identical RcsAB binding sites in front of *P. ananatis rcsA* (PANA_2308) and *anaA* (PANA_2506) (Fig 4.30). Thus *P. ananatis* LMG20103 likely regulates ananatan EPS synthesis in a similar fashion to *P. stewartii*. Also upstream of *P. stewartii rcsA* is a binding site for the quorum-sensing transcriptional regulator EsaR (von Bodman et al. 2003). BlastN analysis with the EsaR binding sequence of *P. stewartii*, showed a sequence identical in 21/23 nucleotides in front of *rcsA* (Fig. 4.30) which indicates that quorum-sensing control of ananatan synthesis occurs in *P. ananatis* LMG20103.

PhoP-PhoQ global regulator

The PhoPQ TCS consists of sensor kinase PhoQ and receptor PhoP. A role in regulation of *Salmonella* pathogenesis has been demonstrated, as it modulates resistance to antimicrobial peptides and acidic pH allowing the pathogen to survive in macrophages. PhoQ responds to Mg^{2+} ion concentration, sensing transition from the extracellular environment to an intra-macrophage localisation (Fass et al. 2009). In the phytopathogen *D. dadantii* PhoP/Q responds to plant extracts and also regulates expression of genes for survival in the acidic apopolast, resistance to antimicrobials and CWDE production (Llama-Palacios et al. 2003). Upregulated genes in *Salmonella* include *pagP* and the PrmA/B TCS which modify the LPS lipid A moiety for resistance to acidic pH and antimicrobials, the Rcs phosphorelay pathway which modulates EPS biosynthesis and regulators for response to oxidative stress (García-Calderón et al. 2007). PANA_1511-1512 show 68% and 71% average aa identity to PhoPQ in *S. enterica* Typhimurium LT2 (STM1230-1231) and *D. dadantii* (CAD332737-332738), respectively. String 8.2 analysis of the interactions of *S. enterica phoP* and *phoQ* with regulated genes reveals a complex network of interaction (Jensen et al. 2009; Http://string.embl.de). Given the sequence homology to *S. enterica* LT2 PhoPQ and homologues for all of its regulated genes, a similar global regulatory network for pathogenicity may occur in *P. ananatis* (Fig 4.31).

CONCLUSIONS

Analysis of the *P. ananatis* LMG20103 genome revealed the presence of a large number of genes encoding putative pathogenicity factors, encompassing more than 10% of the total genome content. This is comparable to the arsenal of pathogenicity determinants found in the tomato pathogen *P. syringae* pv. tomato DC3000 and potato soft rot pathogen *P. atrosepticum* SCRI1043 (Bell et al. 2004; Buell et al. 2003). Many of the putative pathogenicity factors share similarity to the functionally characterised determinants in other phytopathogens and thus give vital clues about how *P. ananatis* LMG20103 causes blight and dieback of *Eucalyptus*. The roles of several putative *P. ananatis* pathogenicity proteins in disease must, however, be viewed with caution as they may also play roles in normal cellular functioning and environmental survival (Toth et al. 2006). The presence of genes encoding flagellum structural and functional components indicates *P. ananatis* LMG20103 is capable of flagellar motility similar to a wide range of bacteria (Soutourina and Bertin, 2003). Motility, however, also plays a vital role in pathogenesis, allowing the bacterium to move towards colonization and infection sites (Tans-Kersten et al. 2001). The presence of a flagellin glycosylation island in *P. ananatis* hints towards a pathogenic role as do the extensive set of methyl-accepting chemotaxis proteins, many of which share homologues only in plant-pathogenic bacteria. Lipopolysaccharides also form a common feature in Gram-negative bacteria, but the presence of proteins for the synthesis of unusual core and O-antigen domains in the LPS of *P. ananatis* LMG20103 suggest a role in pathogenesis, allowing it to evade plant host defenses (Silipo et al. 2005). Several iron scavenging systems are also present which may function in iron acquisition both in the environment and from plant hosts. A large number of fimbrial and non-fimbrial adhesins are encoded on the genome, many of which share homology to those in environmental bacteria, while one fimbria is found only in insect-associated bacteria and may play a role in attachment to insects that serve as vectors that introduce *P. ananatis* into its plant hosts.

P. ananatis exists as epiphyte and endophyte on plants and as soil saprophyte (Coutinho and Venter 2009). Thus some of the above-mentioned factors may play significant roles in the survival of *P. ananatis* as non-pathogenic inhabitant in the plant environment. However, given the right conditions, *P. ananatis* may adapt to

cause disease in its plant hosts. Quorum sensing mechanism are encoded on the *P. ananatis* LMG20103 genome which may allow individual cells to communicate and adapt its communal phenotype in response to environmental signals and coordinately produce pathogenicity factors involved in disease induction and symptom development (de Kievit and Iglewski, 2000). Several global and specific pathogenicity regulatory systems are also encoded on the genome which may allow coordinate expression of its disease determinants (Mole et al. 2007).

Tumourigenic pathogens induce gall formation on their plant hosts (Barash and Manulis-Sasson, 2007). Analysis of the *P. ananatis* LMG20103 genome showed the absence of pathogenicity determinants for gall formation, indicating it is not a tumourigenic pathogen. Some necrotrophic pathogens make use of phytotoxins to lyse host tissues (Bender et al. 1999). No homologues to major phytotoxins are encoded by *P. ananatis* LMG20103. Other phytopathogenic bacteria make use of cell wall degrading enzymes to breach the plant cell wall and macerate host tissues (Toth et al. 2006). The absence of a Type II secretory apparatus and pectin lyases however suggest CWDEs do not play a major role in *P. ananatis* disease. Instead, an extracellular polysaccharide, which we have named ananatan, is structurally and genetically similar to *P. stewartii* stewartan and *E. amylovora* amylovoran and is likely involved in cell bursting and tissue necrosis (Minogue et al. 2002), indicating a necrotrophic lifestyle for *P. ananatis* LMG20103.

In order to deploy phytotoxins, cell wall degrading enzymes or exopolysaccharides at the site of infection, necrotrophic pathogens must be able to evade or disable host defenses. The majority of phytopathogenic bacteria make use of a Type III secretion system (T3SS) which exports effectors into the host that manipulate plant defenses. The absence of a T3SS in the genome of *P. ananatis* LMG20103 raises question on how this pathogen performs the same function. The presence of genes encoding components of a group 4 capsule suggests *P. ananatis* LMG20103 can alternate between G4C and ananatan. The former can mask immunogenic surface structures and thus prevent host detection during early infection (Shifrin et al. 2008), while the latter would be produced at the appropriate time to induce cell bursting and tissue necrosis (Minogue et al. 2002). Production of cyclic glucans may also allow *P. ananatis* LMG20103 to manipulate plant host defenses (Breedveld and Miller, 1994).

Phytopathogenic bacteria also make use of secretion systems to inject effectors with various pathogenic functions into the plant cell. The absence of Type II, III and IV secretion systems indicates the pathogenicity mechanisms of *P. ananatis* LMG20103 are discrete from those of many closely related phytopathogens. On the other hand, Type I and V secretion systems are present and these appear to secrete several putative pathogenicity determinants. The role of the Type VI secretion system (T6SS) in both animal and plant disease has recently received a lot of attention. As there are three T6SSs encoded on the *P. ananatis* LMG20103 genome, these are likely to play an important part in its pathogenesis and are described in detail in Chapter 5.

Analysis of the *P. ananatis* LMG20103 genome sequence enabled the identification of a large number of putative pathogenicity determinants which suggest this pathogen is well-equipped to cause disease in its host plant *Eucalyptus*. To determine the true function of these factors in disease however, will require functional analyses of the various factors. The availability of the genome sequence and comparative identification of pathogenicity targets provide an excellent springboard for functional analysis of the *P. ananatis* LMG20103 pathogenicity factors. *P. ananatis* strains have been associated with diseases on a wide range of plant hosts. Furthermore, it is frequently isolated as plant epi- and endophyte, as soil saprophyte and even as human pathogen (Coutinho and Venter, 2009). The sequencing of more genomes and comparison of their CDSs to the *P. ananatis* LMG20103 genome can be used to identify differences in disease and pathogenicity factors between the various strains infecting different hosts and contribute to understanding the various factors involved in the pathogenesis and lifestyles of various *P. ananatis* strains.

REFERENCES

Abramovitch, R. B. and Martin, G. B. 2004. Strategies used by bacterial pathogens to suppress plant defenses. Curr. Opin, in Plant Biol. **7**: 356-364

Alfano, J. R. and Collmer, A. 2004. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. Annu. Rev. Phytopathol. **42**: 385-414

Antúnez-Lamas, M., Cabrera-Ordóñez, E., López-Solanilla, E., Raposo, R., Trelles-Salazar, O. et al. 2009. Role of motility and chemotaxis in the pathogenesis of *Dickeya dadantii* 3937 (ex *Erwinia chrysanthemi* 3937). Microbiology **155**: 434-442

Aoki, S. K., Webb, J. S., Braaten, B. A. and Low, D. A. 2009. Contact-dependent growth inhibition causes reversible metabolic downregulation in *Escherichia coli*. J. Bacteriol. **191**: 1777-1786

Arora, S. K., Wolfgang, M. C., Lory, S. and Ramphal, R. 2004. Sequence polymorphism in the glycosylation island and flagellins of *Pseudomonas aeruginosa*. J. Bacteriol. **186**: 2115-2122

Babinski, K. J., Ribeiro, A. A. and Raetz, C. R. H. 2002. The *Escherichia coli* gene encoding the UDP-2,3-diacylglucosamine pyrophosphatase of Lipid A biosynthesis. J. Biol. Chem. **277**: 25937-25946

Bamford, V. A., Armour, M., Mitchell, S. A., Cartron, M., Andrews, S. C. et al. 2008. Preliminary X-ray diffraction analysis of YqjH from *Escherichia coli*: a putative cytoplasmic ferri-siderophore reductase. Acta Crystallogr. **F64**: 792-796

Barash, I. and Manulis-Sasson, S. 2007. Virulence mechanisms and host specificity of gall-forming *Pantoea agglomerans*. Trends Microbiol. **15**: 538-545

Bell, K., Sebaihia, M., Pritchard, L., Holden, M., T. et al. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. Proc. Natl Acad. Sci. USA **101**: 11105-11110

Bender, C. L., Alarcón-Chaidez, F. and Gross, D. C. 1999. *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol. Mol. Biol. Rev. **63**: 266-292

Bereswill, S. and Geider, K. 1997. Characterization of the *rcsB* gene from *Erwinia amylovora* and its influence on exopolysaccharide synthesis and virulence of the fire blight pathogen. J. Bacteriol. **179**: 1354-1361

Berry, M. C., McGhee, G. C., Zhao, Y. and Sundin, G. W. 2009. Effect of a *waaL* mutation on lipopolysaccharide composition, oxidative stress survival, and virulence in *Erwinia amylovora*. FEMS Microbiol. Lett. **291**: 80-87

Boccara, M., Aymeric, J. and Camus, C. 1994. Role of endoglucanases in *Erwinia chrysanthemi* 3937 virulence on *Saintpaulia ionantha*. J. Bacteriol. **176**(5): 1524-1526

Bohin, J.-P. 2000. Osmoregulated periplasmic glucans in Proteobacteria. FEMS Microbiol. Lett. **186**: 11-19

Bonifield, H. R. and Hughes, K. T. 2003. Flagellar phase variation in *Salmonella enterica* is mediated by a posttranslational control mechanism. J. Bacteriol. **185**: 3567-3574

Bouchart, F., Delangle, A., Lemoine, J., Bohin, J.-P. and Lacroix, J.-M. 2007. Proteomic analysis of a non-virulent mutant of the phytopathogenic bacterium *Erwinia chrysanthemi* deficient in osmoregulated periplasmic glucans: change in protein expression is not restricted to the envelope, but affects general metabolism. Microbiology **153**: 760-767

Breedveld, M. W. and Miller, K. J. 1994. Cyclic β-glucans of members of the family *Rhizobiaceae*. Microbiol. Rev. **58**: 145-161

Brencic, A. and Winans, S. C. 2005. Detection of and response to signals involved in host-microbe interactions by plant-associated bacteria. Microbiol. Mol. Biol. Rev. **69**: 155-194

Brickman, T. J., Anderson, M. T. and Armstrong, S. K. 2007. *Bordetella* iron transport and virulence. Biometals **20**: 303-322
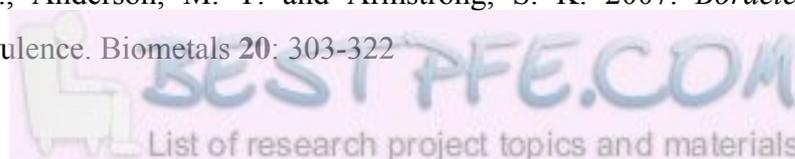
Bronstein, P. A., Marrichi, M., Cartinhour, S., Schneider, D. J. and DeLisa, M. P. 2005. Identification of a twin-arginine translocation system in *Pseudomonas syringae* pv. tomato DC3000 and its contribution to pathogenicity and fitness. J. Bacteriol. **187**: 8450-8461

Budd, A., Blandin, S., Levashina, E. A. and Gibson, T. J. 2004. α-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? Genome Biol. **5**: 1-13

Burse, A., Weingart, H. and Ullrich, M. S. 2004a. The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, *Erwinia amylovora*. Mol. Plant-Microbe Interact. **17**: 43-54

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L. et al. 2003. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv tomato DC3000. Proc. Natl. Acad. Sci. USA **100**(18): 10181-10186.

Burse, A., Weingart, H. and Ullrich, M. S. 2004b. NorM, an *Erwinia amylovora* multidrug efflux pump involved in *in vitro* competition with other epiphytic bacteria. Appl. Environ. Microbiol. **70**: 693-703

Canals, R., Jiménez, N., Vilches, S., Regué, M., Merino, S. et al. 2006. The UDP N-acetylgalactosamine 4-epimerase gene is essential for mesophilic *Aeromonas hydrophila* serotype O34 virulence. Infect. Immun. **74**: 537-548

Cao, J., Woodhall, M. R., Cartron, M. L., Alvarez, J. and Andrews, S. C. 2007. EfeUOB (YcdNOB) is a tripartite, acid-induced and CpxAR-regulated, low-pH Fe^{2+} transporter that is cryptic in *Escherichia coli* K-12 but functional in *E. coli* O157:H7. Mol. Microbiol. **65**: 857-875

Carlier, A. L. and von Bodman, S. B. 2006. The *rcsA* promoter of *Pantoea stewartii* subsp. *stewartii* features a low-level constitutive promoter and an EsaR quorum-sensing-regulated promoter. J. Bacteriol. **188**: 4581-4584

Carter, J. A., Blondel, C. J., Zaldívar, M., Álvarez, S. A., Marolda, C. L. et al. 2007. O-antigen modal chain length in *Shigella flexneri* 2a is growth-regulated through RfaH-mediated transcriptional control of the *wzy* gene. Microbiology **153**: 3499-3507

Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M. et al. 2008. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. **36**: 623-631

Castanié-Cornet, M.-P., Cam, K. and Jacq, A. 2006. RcsF is an outer membrane lipoprotein involved in the RcsCDB phosphorelay signaling pathway in *Escherichia coli*. J. Bacteriol. **188**: 4264-4270

Charbonneau, M.-E., Berthiaume, F. and Mourez, M. 2006. Proteolytic processing is not essential for multiple functions of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence (AIDA-I). J. Bacteriol. **188**: 8504-8512

Chatterjee, A. K. and Vidaver, A. K. 1986. Genetics of pathogenicity factors: application to phytopathogenic bacteria. In Advances in plant pathology volume 4: Application to phytopathogenic bacteria, Orlando, Academic Press**: 1-218.

Chen, Y.-C., Chang, M.-C., Chuang, Y.-C. and Jeang, C.-L. 2004. Characterization and virulence of hemolysin III from *Vibrio vulnificus*. Infect. Immun. **49**: 175-179

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P. et al. 1996. S-adenosylmethionine and methylation. FASEB J. **10**: 471-480

Coderch, N., Piqué, N., Lindner, B., Abitiu, N., Merino, S. et al. 2004. Genetic and structural characterization of the core region of the lipopolysaccharide from *Serratia marcescens* N28b (serovar O4). J. Bacteriol. **186**: 978-988

Coudron, P. E., Moland, E. S. and Thomson, K. S. 2000. Occurrence and detection of AmpC beta-lactamases among *Escherichia coli*, *Klebsiella pneumoniae*, and *Proteus mirabilis* isolates at a veterans medical center. J. Clin. Microbiol. **38**: 1791-1796

Coutinho, T. A. and Venter, S. N. 2009. *Pantoea ananatis*: an unconventional plant pathogen. Mol. Plant Pathol. **10**: 325-335

Cui, Y., Tu, R., Guan, Y. and Chen, S. 2006. Cloning, sequencing and characterization of the *Azospirillum brasilense fhuE* gene. Curr. Microbiol. **52**: 169-177

Cuthbertson, L., Kimber, M. S. and Whitfield, C. 2007. Substrate binding by a bacterial ABC transporter involved in polysaccharide export. Proc. Natl. Acad. Sci. USA **104**: 19529-19534

Dasgupta, T. and Lam, J. S. 1995. Identification of *rfbA*, involved in B-band lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa* serotype O5. Infect. Immun. **63**: 1674-1680

De Boer, S. H. 2003. Characterization of pectolytic erwinias as highly sophisticated pathogens of plants. Eur. J. Plant Pathol. **109**: 893-899

de Kievit, T. R. and Iglewski, B. H. 2000. Bacterial quorum sensing in pathogenic relationships. Infect. Immun. **68**: 4839-4849

de Kievit, T. R. and Lam, J. S. 1997. Isolation and characterization of two genes, *waaC* (*rfaC*) and *waaF* (*rfaF*), involved in *Pseudomonas aeruginosa* serotype O5 inner-core biosynthesis. J. Bacteriol. **179**: 3451-3457

Delepelaire, P. 2004. Type I secretion in Gram-negative bacteria. Biochim. Biophys. Acta **1694**: 149-161

Dellagi, A., Brisset, M.-N., Paulin, J.-P. and Expert, D. 1998. Dual role of desferrioxamine in *Erwinia amylovora* pathogenicity. Mol. Plant-Microbe Interact. **11**: 734-742

Déziel, E., Lépine, F., Milot, S. and Villemur, R. 2003. *rhlA* is required for the production of a novel biosurfactant promoting swarming motility in *Pseudomonas aeruginosa*: 3-(3-hydroxyalkanoyloxy)-alkanoic acids (HAAs), the precursors of rhamnolipids. Microbiology **149**: 2005-2013

Ding, Z., Atmakuri, K. and Christie, P. J. 2003. The outs and ins of bacterial Type IV secretion substrates. Trends Microbiol. **11**: 527-535

Ding, Z. and Christie, P. J. 2003. *Agrobacterium tumefaciens* twin-arginine-dependent translocation is important for virulence, flagellation, and chemotaxis but not Type IV secretion. J. Bacteriol. **185**: 760-771

Doan, N. and Gettins, P. G. 2008. α-Macroglobulins are present in some Gram-negative bacteria. J. Biol. Chem. **283**(42): 28747-28756

Drummelsmith, J. and Whitfield, C. 1999. Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). Mol. Microbiol. **31**: 1321-1332

Eguchi, Y., Okada, T., Minagawa, S., Oshima, T., Mori, H. et al. 2004. Signal transduction cascade between EvgA/EvgS and PhoP/PhoQ two-component systems of *Escherichia coli*. J. Bacteriol. **186**: 3006-3014

Erova, T. E., Sha, J., Horneman, A. J., Borchardt, M. A., Khajanchi, B. K. et al. 2007. Identification of a new hemolysin from diarrheal isolate SSU of *Aeromonas hydrophila*. FEMS Microbiol. Lett. **275**: 301-311

Fass, E., Groisman, E. A., Dorman, C. and Jenal, U. 2009. Controls of *Salmonella* pathogenicity island-2 gene expression. Curr.Opin. Microbiol. **12**: 199-204

Ferrières, L. and Clarke, D. J. 2003. The RcsC sensor kinase is required for normal biofilm formation in *Escherichia coli* K-12 and controls the expression of a regulon in response to growth on a solid surface. Mol. Microbiol. **50**: 1665-1682

Fitzgerald, C., Sherwood, R., Gheesling, L. L., Brenner, F. W. and Fields, P. I. 2003. Molecular analysis of the *rfb* O antigen gene cluster of *Salmonella enterica* serogroup O:6,14 and development of a serogroup-specific PCR assay. Appl. Environ. Microbiol. **69**: 6099-6105

Forest, K. T., Langford, P. R., Kroll, J. S. and Getzoff, E. D. 2000. Cu,Zn superoxide dismutase structure from a microbial pathogen establishes a class with a conserved dimer interface. J. Mol. Biol. **296**: 145-153

García-Calderón, C. B., Casadesús, J. and Ramos-Morales, F. 2007. Rcs and PhoPQ regulatory overlap in the control of *Salmonella* virulence. J. Bacteriol. **189**: 6635-6644

Geider, K. 2000. Exopolysaccharides of *Erwinia amylovora*: structure, biosynthesis, regulation, role in pathogenicity of amylovoran and levan. In Fire blight: the disease and its causative agent. Ed. Vanneste, J.L. Wallingford, CAB International

Geremia, R. A., Petroni, E. A., Ielpi, L. and Henrissat, B. 1996. Towards a classification of glycosyltransferases based on amino acid sequence similarities: prokaryotic α-mannosyltransferases. J. Biol Chem. **318**: 133-138

Gerlach, R. G., Cláudio, N., Rohde, M., Wagner, C. and Hensel, M. 2008. Cooperation of *Salmonella* pathogenicity islands 1 and 4 is required to breach epithelial barriers. Cell. Microbiol. **10**: 2364-2376

Gitaitis, R. D., Walcott, R. R., Wells, M. L., Diaz Perez, J. C. and Sanders, F. H. 2003. Transmission of *Pantoea ananatis*, causal agent of center rot of onion by tobacco thrips, *Frankliniella fusca*. Plant Dis. **87**: 675-678

Gregory, P. J., Johnson, S. N., Newton, A. C. and Ingram, J. S. I. 2009. Integrating pests and pathogens into the climate change/ food security debate. J. Exp. Bot. **60**: 2827-2838

Grizot, S., Salem, M., Vongsouthi, V., Durand, L., Moreau, F. et al. 2006. Structure of the *Escherichia coli* heptosyltransferase WaaC: binary complexes with ADP and ADP-2-deoxy-2-fluoro heptose. J. Mol. Biol. **363**: 383-394

Hanyu, M., Fujimoto, H., Tejima, K. and Saeki, K. 2009. Functional differences of two distinct catalases in *Mesorhizobium loti* MAFF303099 under free-living and symbiotic conditions. J. Bacteriol. **191**: 1463-1471

Hazes, B. and Frost, L. 2008. Towards a system biology approach to study Type II/IV secretion systems. Biochim. Biophys. Acta **1778**: 1839-1850

Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C. and Ala'Aldeen, D. 2004. Type V protein secretion pathway: the autotransporter story. Microbiol. Mol. Biol. Rev. **68**: 692-744

Herlache, T. C., Hotchkiss, A. T., Burr, T. J. and Collmer, A. 2008. Characterization of the *Agrobacterium vitis pehA* gene and comparison of the encoded polygalacturonase with the homologous enzymes from *Erwinia carotovora* and *Ralstonia solanacearum*. Appl. Environ. Microbiol. **63**: 338-346

Holden, M. T. G., Seth-Smith, H. M. B., Crossman, L. C., Sebaihia, M., Bentley, S. D. et al. 2009. The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients. J. Bacteriol. **191**: 261-277

Hong, S. Y., An, C. L., Cho, K. M., Lee, S. M., Kim, Y. H., Kim, M. K., et al. 2006. Cloning and comparison of third β-glucoside utilization (*bglEFIA*) operon with two operons of *Pectobacterium carotovorum* subsp. *carotovorum* LY34. Biosci. Biotechnol. Biochem. **70**(4): 798-807

Hoopman, T. C., Wang, W., Brautigam, C. A., Sedillo, J. L., Reilly, T. J. et al. 2008. *Moraxella catarrhalis* synthesises an autotransporter that is an acid phosphatase. J. Bacteriol. **190**: 1459-1472

Hugouvieux-Cotte-Pattat, N., Shevchik, V. E. and Nasser, W. 2002. PehN, a polygalacturonase homologue with a low hydrolase activity, is coregulated with the other *Erwinia chrysanthemi* polygalacturonases. J. Bacteriol. **184**(10): 2664-2673

Huang, Y.-J., Liao, H.-W., Wu, C.-C. and Peng, H.-L. 2009. MrkF is a component of Type 3 fimbriae in *Klebsiella pneumoniae*. Res. Microbiol. **160**: 71-79

Hurst, M. R., O'callaghan, M. and Glare, T. R. 2003. Peripheral sequences of the *Serratia entomophila* pADAP virulence-associated region. Plasmid **50**: 213-229

Irie, Y., Mattoo, S. and Yuk, M. H. 2004. The Bvg virulence control system regulates biofilm formation in *Bordetella bronchiseptica*. J. Bacteriol **186**: 5692-5698

Iwama, T., Ito, Y., Aoki, H., Sakamoto, H., Yamagata, S. et al. 2006. Differential recognition of citrate and a metal-citrate complex by the bacterial chemoreceptor Tcp. J. Biol. Chem. **281**: 17727-17735

Jacob-Dubuisson, F., Locht, C. and Antoine, R. 2001. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. Mol. Microbiol. **40**: 306-313

Jayaratne, P., Bronner, D., MacLachlan, P. R., Dodgson, C., Kido, N. et al. 1994. Cloning and analysis of duplication *rfbM* and *rfbK* genes involved in the formation of GDP-mannose in *Eschenrchia coli* O9:K30 and participation of *rfb* genes in the synthesis of the group I K30 capsular antigen. J. Bacteriol. **176**: 3126-3139

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C. et al. 2009. STRING 8 - a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. **37**: D412-D416

Josenhans, C. and Suerbaum, S. 2002. The role of motility as a virulence factor in bacteria. Int. J. Med. Microbiol. **291**: 605-614

Kido, N., Torgov, V. I., Sugiyama, T., Uchiya, K., Sugihara, H. et al. 1995. Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. J. Bacteriol. **177**: 2178-2187

Kim, W.-S. and Geider, K. 2000. Characterization of a viral EPS-depolymerase, a potential tool for control of fire blight. Phytopathology **90**: 1263-1268

King, J. D., Poon, K. K. H., Webb, N. A., Anderson, E. M., McNally, D. J. et al. 2009. The structural basis for catalytic function of GMD and RMD, two closely related enzymes from the GDP-D-rhamnose biosynthesis pathway. FEBS J. **276**: 2686-2700

Kneidinger, B., Graninger, M., Adam, G., Puchberger, M., Kosma, P. et al. 2001. Identification of two GDP-6-deoxy-D-lyxo-4-hexulose reductases synthesizing GDP-D-rhamnose in *Aneurinibacillus thermoaerophilus* L420-91T. J. Biol. Chem. **276**: 5577-5583

Korhonen, T. K., Tarkka, E., Ranta, H. and Haahtela, K. 1983. Type 3 fimbriae of *Klebsiella* sp.: molecular characterization and role in bacterial adhesion to plant roots. J. Bacteriol. **155**: 860-865

Koutsoudis, M. D., Tsaltas, D., Minogue, T. D. and von Bodman, S. B. 2006. Quorum-sensing regulation governs bacterial adhesion, biofilm development, and host colonization in *Pantoea stewartii* subspecies stewartii. Proc. Natl. Acad. Sci USA **103**: 5983-5988

Kucerova, E., Clifton, S. W., Xia, X.-q., Long, F., Porwollik, S. et al. 2010. Genome sequence of *Cronobacter sakazakii* BAA-894 and comparative genomic hybridization analysis with other *Cronobacter* species. PLOS One **5**: 1-10

Kyöstiö, S. R. M., Cramer, C. L. and Lacy, G. H. 1991. *Erwinia carotovora* subsp. *carotovora* extracellular protease: characterization and nucleotide sequence of the gene. J. Bacteriol. **173**: 6537-6546

Lavander, M., Ericsson, S. K., Bröms, J. E. and Forsberg, A. 2006. The twin arginine translocation system is essential for virulence of *Yersinia pseudotuberculosis*. Infect. Immun. **74**: 1768-1776

Lequette, Y., Ōdberg-Ferragut, C., Bohin, J. and Lacroix, J. 2004. Identification of *mdoD*, an *mdoG* paralog which encodes a twin-arginine-dependent periplasmic protein that controls osmoregulated periplasmic glucan backbone structures. J. Bacteriol. **186**(12): 3695-3702

Létoffé, S., Delepelaire, P. and Wandersman, C. 2004. Free and hemophore-bound heme acquisitions through the outer membrane receptor HasR have different requirements for the TonB-ExbB-ExbD complex. J. Bacteriol. **186**: 4067-4074

Li, Y., Hao, G., Galvani, C. D., Meng, Y., De La Fuente, L. et al. 2007. Type I and Type IV pili of *Xylella fastidiosa* affect twitching motility, biofilm formation and cell–cell aggregation. Microbiology **153**: 719-726

Lindeberg, M., Stavranides, J., Chang, J. H., Alfano, J. R., Collmer, A. et al. 2005. Proposed guidelines for a unified nomenclature and phylogenetic analysis of Type III Hop effector proteins in the plant pathogen *Pseudomonas syringae*. Mol. Plant-Microbe Interac. **18**: 275-282

Lindow, S. E., Arny, D. C. and Upper, C. D. 1982. Bacterial ice nucleation: a factor in frost injury to plants. Plant Physiol. **70**: 1084-1089

Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. et al. 2008. Structure and genetics of *Shigella* O antigens. FEMS Microbiol. Rev. **32**: 627-653

Llama-Palacios, A., López-Solanilla, E., Poza-Carrión, C., García-Olmedo, F. and Rodríguez-Palenzuela, P. 2003. *Erwinia chrysanthemi phoP-phoQ* operon plays an important role in growth at low pH, virulence and bacterial survival in plant tissue. Mol. Microbiol. **49**: 347-357

López-Solanilla, E., García-Olmedo, F. and Rodríguez-Palenzuela, P. 1998. Inactivation of the *sapA* to *sapF* locus of *Erwinia chrysanthemi* reveals common features in plant and animal bacterial pathogenesis. Plant Cell **10**: 917-924

Lovering, A. L., Lee, S. S., Kim, Y., Withers, S. G. and Strydnaka, N. C. 2005. Mechanistic and structural analysis of a family 31 α-glycosidase and its glycosyl-enzyme intermediate. J. Biol. Chem. **280**(3): 2105-2115

Luck, S. N., Turner, S. A., Rajakumar, K., Sakellaris, H. and Adler, B. E. N. 2001. Ferric dicitrate transport system (Fec) of *Shigella flexneri* 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes. Infect. Immun. **69**: 6012-6021

Main-Hester, K. L., Colpitts, K. M., Thomas, G. A., Fang, F. C. and Libby, S. J. 2008. Coordinate regulation of *Salmonella* pathogenicity island 1 (SPI1) and SPI4 in *Salmonella enterica* serovar Typhimurium. Infect. Immun. **76**(3): 1024-1035

Marchler-Bauer, A. and Bryant, S. H. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. **32**: 327-331

Margolles, A. and de Los Reyes-Gavilán, C. G. 2003. Purification and functional characterization of a novel α-L-arabinofuranosidase from *Bifidobacterium longum* B667. Appl. Environ. Microbiol. **69**: 5096-5103

Matsuo, T., Chen, J., Minato, Y., Ogawa, W., Mizushima, T. et al. 2008. SmdAB, a heterodimeric ABC-type multidrug efflux pump, in *Serratia marcescens*. J. Bacteriol. **190**: 648-654

189

Meslet-Cladiere, L. M., Pimenta, A., Duchaud, E., Holland, I. B. and Blight, M. A. 2004. *In vivo* expression of the mannose-resistant fimbriae of *Photorhabdus temperata* K122 during insect infection. J. Bacteriol. **186**: 611-622

Metzger, M., Bellemann, P., Bugert, P. and Geider, K. 1994. Genetics of galactose metabolism of *Erwinia amylovora* and its influence on polysaccharide synthesis and virulence of the fire blight pathogen. J. Bacteriol **176**(2): 450-459.

Minogue, T. D., Wehland-von Tebra, M., Bernhard, F. and von Bodman, S. B. 2002. The autoregulatory role of EsaR, a quorum-sensing regulator in *Pantoea stewartii* ssp. *stewartii*: evidence for a repressor function. Mol. Microbiol. **44**: 1625-1635

Mohammadi, M. and Geider, K. 2007. Autoinducer-2 of the fire blight pathogen *Erwinia amylovora* and other plant-associated bacteria. FEMS Microbiol. Lett. **266**: 34-41

Mole, B. M., Baltrus, D. A., Dangl, J. L. and Grant, S. R. 2007. Global virulence regulation networks in phytopathogenic bacteria. Trends Microbiol. **15**(8): 363-371

Mongkolsuk, S., Praituan, W., Loprasert, S., Fuangthong, M. and Chamnongpol, S. 1998. Identification and characterization of a new organic hydroperoxide resistance (*ohr*) gene with a novel pattern of oxidative stress regulation from *Xanthomonas campestris* pv. phaseoli. J. Bacteriol. **180**: 2636-2643

Morales, C., Lee, M. D., Hofacre, C. and Maurer, J. J. 2004. Detection of a novel virulence gene and a *Salmonella* virulence homologue among *Escherichia coli* isolated from broiler chickens. Foodborne Pathog. Dis. **1**: 160-165

Morgan, R., Kohn, S., Hwang, S.-H., Hassett, D. J. and Sauer, K. 2006. BdlA, a chemotaxis regulator essential for biofilm dispersion in *Pseudomonas aeruginosa*. J. Bacteriol. **188**: 7335-7343

Morgan, E., Bowen, A. J., Carnell, S. C., Wallis, T. S. and Stevens, M. P. 2007. SiiE is secreted by the *Salmonella enterica* serovar Typhimurium pathogenicity islands 4-encoded secretion system and contributes to intestinal colonization in cattle. Infect. Immun. **75**(3): 1524-1533

190
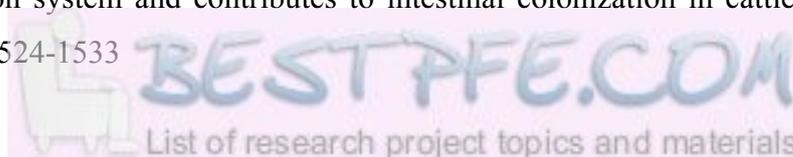
Mori, H. and Ito, K. 2001. The Sec protein-translocation pathway. Trends Microbiol. **9**: 494-500

Morohoshi, T., Nakamura, Y., Yamazaki, G., Ishida, A., Kato, N. et al. 2007. The plant pathogen *Pantoea ananatis* produces N-acylhomoserine lactone and causes center rot disease of onion by quorum sensing. J. Bacteriol. **189**: 8333-8338

Nachin, L., El Hassouni, M., Loiseau, L., Expert, D. and Barras, F. 2001. SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. Mol. Microbiol. **39**: 960-972

Nishino, K., Nikaido, E. and Yamaguchi, A. 2007. Regulation of multidrug efflux systems involved in multidrug and metal resistance of *Salmonella enterica* serovar Typhimurium. J. Bacteriol. **189**: 9066-9075

Nuccio, S.-P. and Baümler, A. J. 2007. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. Microbiol. Mol. Biol. Rev. **71**: 551-575

Nudleman, E. and Kaiser, D. 2004. Pulling together with Type IV pili. J. Mol. Microbiol. Biotechnol. **7**: 52-62

Ochsner, U. R., Reiser, J., Fiechter, A., and Witholt, B. 1995. Production of *Pseudomonas aeruginosa* rhamnolipid biosurfactants in heterologous hosts. Appl. Environ. Microbiol. **61**(9): 3503-3506

Ong, C.-l. Y., Ulett, G. C., Mabbett, A. N., Beatson, S. A., Webb, R. I. et al. 2008. Identification of Type 3 fimbriae in uropathogenic *Escherichia coli* reveals a role in biofilm formation. J. Bacteriol. **190**: 1054-1063

Palmer, T., Sargent, F. and Berks, B. C. 2005. Export of complex cofactor-containing proteins by the bacterial Tat pathway. Trends Microbiol. **13**: 175-180

Peleg, A., Shifrin, Y., Ilan, O., Nadler-Yona, C., Nov, S. et al. 2005. Identification of an *Escherichia coli* operon required for formation of the O-antigen capsule. J. Bacteriol. **187**: 5259-5266

Postma, P. W., Lengeler, J. W. and Jacobson, G. R. 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. **57**: 543-594

Proft, T. and Baker, E. N. 2009. Pili in Gram-negative and Gram-positive bacteria – structure, assembly and their role in disease. Cell. Mol. Life Sci. **66**: 613-635

Purdy, G. E. and Payne, S. M. 2001. The SHI-3 iron transport island of *Shigella boydii* 0-1392 carries the genes for aerobactin synthesis and transport. J. Bacteriol. **183**: 4176-4182

Regué, M., Izquierdo, L., Fresno, S., Jimenez, N., Piqué, N. et al. 2005a. The incorporation of glucosamine into enterobacterial core lipopolysaccharide. J. Biol. Chem. **280**: 36648-36656

Regué, M., Izquierdo, L., Fresno, S., Piqué, N., Corsaro, M. M. et al. 2005b. A second outer-core region in *Klebsiella pneumoniae* lipopolysaccharide. J. Bacteriol. **187**: 4198-4206

Relman, D. A., Domenighini, M., Tuomanen, E., Rappuoli, R. and Falkow, S. 1989. Filamentous hemagglutinin of *Bordetella pertussis*: nucleotide sequence and crucial role in adherence. Proc. Natl. Acad. Sci. USA **86**: 2637-2641

Rende-Fournier, R., Leclercq, R., Galimand, M., Duval, J. and Courvalin, P. 1993. Identification of the *satA* gene encoding a streptogramin A acetyltransferase in *Enterococcus faecium* BM4145. Antimicrob. Agents Chemother. **37**: 2119-2125

Riedel, K., Talker-Huiber, D., Givskov, M., Schwab, H. and Eberl, L. 2003. Identification and characterization of a GDSL esterase gene located proximal to the *swr* quorum-sensing system of *Serratia liquefaciens* MG1. Appl. Environ. Microbiol. **69**(7): 3901-3910

Rigano, L. A., Payette, C., Brouillard, G., Marano, M. R., Abramowicz, L. et al. 2007. Bacterial cyclic β-(1,2)-glucan acts in systemic suppression of plant immune responses. Plant Cell **19**: 2077-2089

Rocchetta, H. L., Burrows, L. L. and Lam, J. S. 1999. Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*. Microbiol. Mol. Biol. Rev. **63**: 523-553

Rocha, C. L. and Fischetti, V. A. 1999. Identification and characterization of a novel fibronectin-binding protein on the surface of group A Streptococci. Infect. Immun. **67**: 2720-2728

Roland, K. L., Martin, L. E., Esther, C. R. and Spitznagel, J. K. 1993. Spontaneous *pmrA* mutants of *Salmonella typhimurium* LT2 define a new two-component regulatory system with a possible role in virulence. J. Bacteriol. **175**: 4154-4164

Roux, J., Steenkamp, E. T., Marasas, W. F. O., Wingfield, M. J. and Wingfield, B. D. 2001. Characterization of *Fusarium graminearum* from *Acacia* and *Eucalyptus* using β-tubulin and histone gene sequences. Mycologia **93**: 704-711

Roy, C., Kester, H., Visser, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N., Robert-Baudouy, J. et al. 1999. Modes of action of five different endopectate lyases from *Erwinia chrysanthemi* 3937. J. Bacteriol. **181**(12): 3705-3709

Sabri, M., Caza, M., Proulx, J., Lymberopoulos, M. H., Breé, A. et al. 2008. Contribution of the SitABCD, MntH, and FeoB metal transporters to the virulence of avian pathogenic *Escherichia coli* O78 χ7122. Infect. Immun. **76**: 601-611

Saidijam, M., Benedetti, G., Ren, Q., Xu, Z., Hoyle, C. J. et al. 2006. Microbial drug efflux proteins of the Major Facilitator Superfamily. Curr. Drug Targets **7**: 793-811

Saier, M. H., Paulsen, I. A. N. T., Sliwinski, M. K., Pao, S. S., Skurray, R. A. et al. 1998. Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. FASEB J. **12**: 265-274

Saigí, F., Climent, N., Piqué, N., Sanchez, C., Merino, S. et al. 1999. Genetic analysis of the *Serratia marcescens* N28b O4 antigen gene cluster. J. Bacteriol. **181**: 1883-1891

San Francisco, M. J., Xiang, Z. and Keenan, R. W. 1996. Digalacturonic acid uptake in *Erwinia chrysanthemi*. Mol. Plant-Microbe Interact. **9**(2): 144-147

Sauer, F. G., Mulvey, M. A., Schilling, J. D., Martinez, J. J. and Hultgren, S. J. 2000. Bacterial pili: molecular mechanisms of pathogenesis. Curr. Opin. Microbiol. **3**: 65-72

Sauvonnet, N., Vignon, G., Pugsley, A. P. and Gounon, P. 2000. Pilus formation and protein secretion by the same machinery in *Escherichia coli*. EMBO J. **19**: 2221-2228

Schembri, M. A., Dalsgaard, D. and Klemm, P. 2004. Capsule shields the function of short bacterial adhesins. J. Bacteriol. **186**: 1249-1257

Shevchik, V. E. and Hugouvieux-Cotte-Pattat, N. 2003. PaeX, a second pectin acetylesterase of *Erwinia chrysanthemi* 3937. J. Bacteriol. **185**(10): 3091-3100

Shifrin, Y., Peleg, A., Ilan, O., Nadler, C., Kobi, S. et al. 2008. Transient shielding of intimin and the Type III secretion system of enterohemorrhagic and enteropathogenic *Escherichia coli* by a group 4 capsule. J. Bacteriol. **190**: 5063-5074

Silipo, A., Molinaro, A., Sturiale, L., Dow, J. M., Erbs, G. et al. 2005. The elicitation of plant innate immunity by lipooligosaccharide of *Xanthomonas campestris*. J. Biol.l Chem.**280**: 33660-33668

Soberón-Chávez, G., Aguirre-Ramírez, M. and R., S. 2005. The *Pseudomonas aeruginosa* RhlA enzyme is involved in rhamnolipid and polyhydroxyalkanoate production. J. Indust. Microbiol. and Biotechnol. **32**: 675-677

Soutourina, O. A. and Bertin, P. N. 2003. Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol. Rev. **27**: 505-523

Stathopoulos, C., Hendrixson, D. R., Thanassi, D. G., Hultgren, S. J., St. Geme III, J. W. et al. 2000. Secretion of virulence determinants by the general secretory pathway in gram-negative pathogens: an evolving story. Microb. Infect. **2**: 1061-1072

Szurmant, H. and Ordal, G. W. 2004. Diversity in chemotaxis mechanisms among the Bacteria and Archaea. Microbiol. Mol. Biol. Rev. **68**: 301-319

Takeuchi, K., Taguchi, F., Inagaki, Y., Toyoda, K., Shiraishi, T. et al. 2003. Flagellin glycosylation island in *Pseudomonas syringae* pv. glycinea and its role in host specificity. J. Bacteriol. **185**: 6658-6665

Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**: 1596-1599

Tans-Kersten, J., Huang, H. and Allen, C. 2001. *Ralstonia solanacearum* needs motility for invasive virulence on tomato. J. Bacteriol. **183**: 3597-3605

Taylor, B. L., Zhulin, I. B. and Johnson, M. S. 1999. Aerotaxis and other energy-sensing behavior in bacteria. Annu. Rev. Microbiol. **53**: 103-128

Tomaras, A. P., Dorsey, C. W., Edelmann, R. E. and Actis, L. A. 2003. Attachment to and biofilm formation on abiotic surfaces by *Acinetobacter baumannii*: involvement of a novel chaperone-usher pili assembly system. Microbiology **149**: 3473-3484

Toth, I. K., Bell, K. S., Holeva, M. C. and Birch, P. R. J. 2003. Soft rot erwiniae: from genes to genomes. Mol. Plant Pathol. **4**: 17-30

Toth, I. K., Pritchard, L. and Birch, P. R. J. 2006. Comparative genomics reveals what makes an enterobacterial plant pathogen. Annu. Rev. Phytopathol. **44**: 305-336

Tsujibo, H., Takada, C., Tsuji, A., Kosaka, M., Miyamoto, K., Inamore, Y. et al. 2001. Cloning, sequencing, and expression of the gene encoding an intracellular β-D-xylosidase from *Streptomyces thermoviolaceus* OPC-520. Biosci. Biotech. Biochem. **65**(8): 1824-1831

Uchiyama, H., Hashidoko, Y., Kuriyama, Y. and Tahara, S. 2008. Identification of the 4-hydroxycinnamate decarboxylase (PAD) gene of *Klebsiella oxytoca*. Biosci. Biotech. Biochem. **72**(1): 116-123

Utsumi, R., Yagi, T., Katayama, S., Katsuragi, K., Tachibana, K. et al. 1991. Molecular cloning and characterization of the fusaric acid-resistance gene from *Pseudomonas cepacia*. Agric. Biol. Chem. **55**: 1913-1918

Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E., Menck, C. F., da Silva, A. C., Ferro, J. A. et al. 2002. Comparative genomic analysis of plant-associated bacteria. Annu. Rev. Phytopathol. **40**: 169-189

Vinatzer, B. A. and Yan, S. 2008. Mining the genomes of plant pathogenic bacteria: how not to drown in gigabases of sequence. Mol. Plant Pathol. **9**(1): 105-118

von Bodman, Ball, J. K., Faini, M. A., Herrera, C. M., Minogue, T. D., Urbanowski, M. L. et al. 2003. The quorum sensing negative regulators EsaR and ExpR$_{Ecc}$, homologues within the LuxR family, retain the ability to function as activators of transcription. J. Bacteriol. **185**(23): 7001-7007

Wandersman, C. and Delepelaire, P. 2004. Bacterial iron sources: from siderophores to hemophores. Ann. Rev. Microbiol. **58**: 611-647

Wang, H. and Dowds, B. C. 1993. Phase variation in *Xenorhabdus luminescens*: cloning and sequencing of the lipase gene and analysis of its expression in primary and secondary phases of the bacterium. J. Bacteriol. **175**(6): 1665-1673

Weening, E. H., Barker, J. D., Laarakker, M. C., Humphries, A. D., Tsolis, R. M. et al. 2005. The *Salmonella enterica* serotype Typhimurium *lpf*, *bcf*, *stb*, *stc*, *std*, and *sth* fimbrial operons are required for intestinal persistence in mice. Infect. Immun. **73**: 3358-3366

Wei, J.-R., Tsai, Y.-H., Soo, P.-C., Horng, Y.-T., Hsieh, S.-C. et al. 2005. Biochemical characterization of RssA-RssB, a two-component signal transduction system regulating swarming behavior in *Serratia marcescens*. J. Bacteriol. **187**: 5683-5690

Yahr, T. L. 2006. A critical new pathway for toxin secretion? New Engl. J. Med. **355**(11): 1171-1172

Yang, M., Luoh, S., Goddard, A., Reilly, D., Henzel, W., Bass, S. et al. 1996. The *bglX* gene located at 47.8 min on the *Escherichia coli* chromosome encodes a periplasmic β-glucosidase. Microbiology **142**: 1659-1665

Yao, J. and Allen, C. 2007. The plant pathogen *Ralstonia solanacearum* needs aerotaxis for normal biofilm formation and interactions with its tomato host. J. Bacteriol. **189**: 6415-6424

Yasuta, T., Okazaki, S., Mitsul, H., Yuhashi, K.-I., Ezura, H. et al. 2001. DNA sequence and mutational analysis of rhizobitoxine biosynthesis genes in *Bradyrhizobium elkanii*. Appl. Environ. Microbio. **67**: 4999-5009

Young, G. M., Schmiel, D. H. and Miller, V. L. 1999. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. Proc. Natl. Acad. Sci. USA **96**: 6456-6461

Zhou, S. and Ingram, L. O. 2000. Synergistic hydrolysis of carboxymethyl cellulose and acid-swollen cellulose by two endoglucanases (CelZ and CelY) from *Erwinia chrysanthemi*. J. Bacteriol. **182**(20): 5676-5682

FIGURES AND TABLES

Table 4.1: Percentage amino acid identity between the four copies of the flagellin coding gene *fliC*

fliC gene	Size (aa)	PANA_2281	PANA_3356	PANA_4026	PANA_4061
PANA_2281	307 aa	100			
PANA_3356	309 aa	72%	100		
PANA_4026	311 aa	92%	92%	100	
PANA_4061	311 aa	97%	71%	72%	100

Table 4.2: Homologies of the CDSs in the *P. ananatis* LMG20103 glycosylation island with GI proteins in *P. syringae* pv. tabaci 6605 and *P. aeruginosa* JJ692. Putative functions of the encoded proteins are shown. Hypothetical proteins in *P. syringae* pv. tabaci are named Hp1 and Hp2.

P. ananatis LMG20103	Protein	*P. syringae* tabaci 6605	Protein	% aa ID	*P. aeruginosa* JJ692	Protein	% aa ID	Predicted Function
PANA_2266	RmlA	-	-	-	-	-	-	Glucose-1-P thymidylyltransferase
PANA_2267	RmlC	-	-	-	-	-	-	dTDP-4-dehydrorhamnose 3,5-epimerase
PANA_2268	RmlC	-	-	-	-	-	-	dTDP-4-dehydrorhamnose 3,5-epimerase
PANA_2269	RmlD	-	-	-	-	-	-	dTDP-4-dehydrorhamnose reductase
PANA_2270	-	-	-	-	-	-	-	Hypothetical protein
PANA_2271	-	-	-	-	-	-	-	Acyl-CoA reductase
PANA_2272	-	-	-	-	-	-	-	Acyl-protein synthetase
PANA_2273	-	-	-	-	-	-	-	AMP-dependent CoA synthetase/ligase
PANA_2274	-	-	-	-	-	-	-	Transketolase
PANA_2275	AcpR	BAH58345	AcpR	30%	AAP35716	OrfD	33%	3-oxoacyl-[acyl-carrier-protein] reductase
PANA_2276	AcpS	-	-	-	AAP35714	OrfB	24%	Acyl carrier protein
PANA_2277	-	-	-	-	AAP35719	OrfH	30%	Methyltransferase
PANA_2278	VioB	BAH58344	VioB	31%	AAP35718	OrfG	22%	O-acetyltransferase
PANA_2279	VioA	BAH58340	VioA	64%	AAP35713	OrfA	63%	Nucleotide sugar aminotransferase
PANA_2280	VioT	BAH58342	VioT	30%	AAP35720	OrfN	25%	Glycosyltransferase
PANA_2281	VioT	BAH58342	VioT	25%	AAP35720	OrfN	35%	Glycosyltransferase
-	-	BAH58337	Hp1	-	-	-	-	Hypothetical protein
-	-	BAH58338	Tp	-	-	-	-	Major facilitator superfamily transporter
-	-	BAH58339	Gs	-	-	-	-	Glutamine synthase
-	-	BAH58341	Mt	-	-	-	-	Methyltransferase
		BAH58343	acpS	-	-	-	-	3-oxoacyl-[acyl-carrier-protein) synthase
-	-	BAH58346	Hp2	-	-	-	-	Hypothetical protein
-	-	-	-	-	AAP35715	OrfC	-	3-oxoacyl-[acyl-carrier-protein) synthase
-	-	-	-	-	AAP35717	OrfF	-	Hydroxylating dioxygenase

Table 4.3: Homologies between the *P. ananatis* and *S. entomophila* Sef and *P. temperata* Mrf fimbrial biosynthetic loci. The percentage amino acid identities and functions of the proteins encoded by the loci are shown.

Protein	Function	Homologue in *S. entomophila*	Homologue in *P. temperata*
PANA_2715	Tip adhesin	SefI: 22%	MrfJ: 26%
PANA_2716	Fibrillum/minor subunit	SefH: 37%	MrfE/F/G: 33%
PANA_2717	Fimbrial chaperone	SefG: 49%	-
PANA_2718	Hypothetical protein	-	-
PANA_2719	Hypothetical protein	-	-
PANA_2720	Hypothetical regulator	-	-
PANA_2721	Fibrillum/minor subunit	SefF: 37%	-
PANA_2722	Activator/Initiator	SefE: 36%	-
PANA_2723	Fimbrial chaperone	SefD: 63%	MrfD: 48%
PANA_2724	Outer membrane usher	SefC: 55%	MrfC: 50%
PANA_2725	Fimbrial anchor	SefB: 30%	MrfB: 30%
PANA_2726	Major pilin subunit	SefA: 55%	MrfA: 40%
Average aa identity		**48%**	**41%**

Table 4.4: Protein properties of the *D. dadantii* invasion protein HecA (AAN38708), the *S. marcescens* haemolysin ShlA (AAA50323), the *B. pertussis* filamentous haemagglutinin FhaB (AAA22974), the *E. coli* contact-dependent growth inhibitor CdiA (AAZ57198) and the PANA_0907 protein.

	HecA	**ShlA**	**PANA_0907**	**FhaB**	**CdiA**
Size	3,848 aa	1,608 aa	3,667 aa	3,590 aa	3,132 aa
Molecular mass	387 kDa	165 kDa	375 kDa	367 kDa	319 kDa
pI	5.44	8.77	5.63	8.92	5.77
Aliphatic index	85.1	73.6	84.5	86.0	83.5
Glycine content	634 (16%)	213 (13%)	472 (13%)	448 (12%)	417 (13%)
Gene GC content	68.8 %	58.7 %	54.1 %	66.7 %	56.0 %
RGD motifs	4	-	2	2	1

Table 4.5: Features of the adhesin proteins in *P. ananatis* and *Yersinia mollaretti* ATCC 43969

Adhesin	*P. ananatis* PANA_2607	*Yersinia mollaretti* ZP_04640954
Size (nt)	12,537 nt (4,179 aa)	12,705 nt (4,235 aa)
Molecular Weight (kDa)	416	425
Ig repeats	31	29

Table 4.7: Predicted cell wall degrading enzymes in *P. ananatis* LMG20103. The molecular weights (Mw) and sizes of the encoded enzymes are shown as is their putative target

CDS	Protein	Enzyme	Putative target	Size (aa)	Mw (kDa)
PANA_0132	CelY	Cellulase	Cellulose	334	38.1
PANA_0136	PaeX	Pectin acetyltransferase	Pectin	310	34.3
PANA_0324	Pgl	Polygalacturonase	Pectin	443	48.4
PANA_0376	XynA	Xylanase	Hemicellulose	411	45.0
PANA_0739	AbfA	Arabinofuranosidase	Hemicellulose	319	37.0
PANA_1095	Peh	Polygalacturonase	Pectin	417	44.5
PANA_2594	AbfA	Arabinofuranosidase	Hemicellulose	511	58.3
PANA_2623	XynB	Xylanase	Hemicellulose	331	35.8
PANA_2851	Prt1	Protease	Glycoproteins	337	37.5
PANA_3781	XynB	Xylanase	Hemicellulose	311	34.1

199

Table 4.8: Protein properties for the two putative *P. ananatis* LMG20103 Type I secretion substrates.

	PANA_2607	PANA_0591
Size (aa)	4,179	1,027
Mol. Wt (kDa)	416	102
Isoelectric point pI	4.12	4.9
Glycine content	473 (11.3%)	167 (16.3%)
Cysteine content	2 (0.004%)	7 (0.7%)
Repeat length (aa)	76	16
Number of repeats	31	16

Table 4.9: Protein features of the Type V secreted proteins in *P. ananatis* LMG20103. The predicted size, length of the signal peptide and passenger domains, the predicted function of the protein and the number of eukaryotic cell-binding RGD domains are shown.

	PANA_4212	PANA_1089	PANA_2402	PANA_0907
Means of secretion	AT	AT	AT	TPS
Predicted function	Lipase/ Esterase	Adhesin	Acid phosphatase	Adhesin
Size (aa)	659	985	661	3,667
Mol. Wt (kDa)	70	104	65	375
Signal peptide size	30 aa	26 aa	44 aa	32 aa
Passenger domain size	382 aa	308 aa	139 aa	3,626 aa
RGB domains in passenger	1	2	1	2

Table 4.10: Multidrug efflux systems in *P. ananatis* LMG20103 and those identified in several related phytopathogens (Saidijam et al. 2006).

Organism	ABC	MFS	RND	MATE	SMR	Total
P. ananatis **LMG20103**	**4**	**12**	**13**	**3**	**2**	**34**
A. tumefaciens C58	7	23	8	3	1	42
R. solanacearum GM1000	6	36	10	4	1	56
P. syringae pv. *tomato* DC3000	4	34	10	5	1	54

200
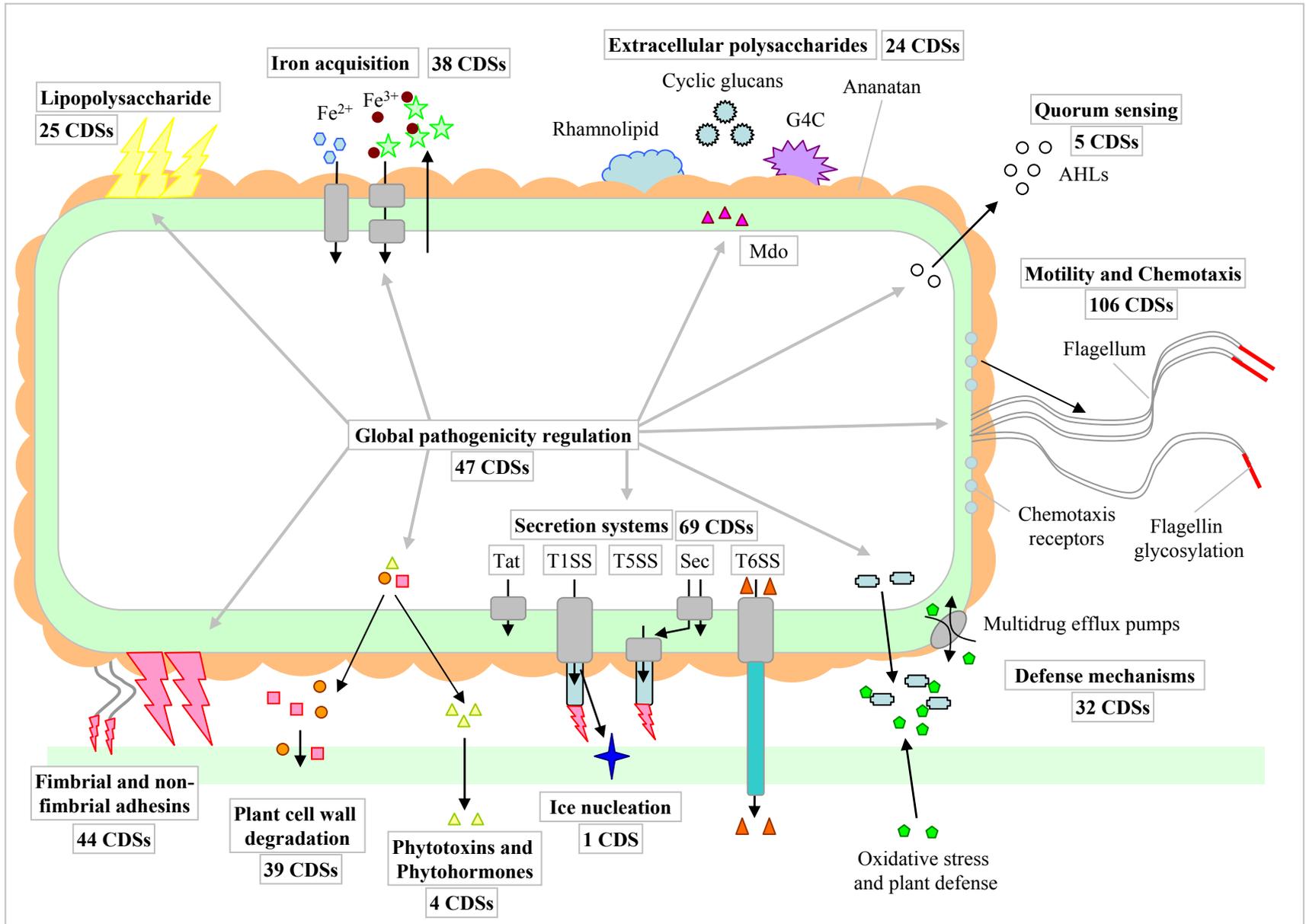
Figure 4.1: Diagram showing the *P. ananatis* LMG20103 protein coding sequences with putative roles in pathogenesis. The various pathogenic functions and number of *P. ananatis* protein coding sequences (CDSs) involved in each function are shown.
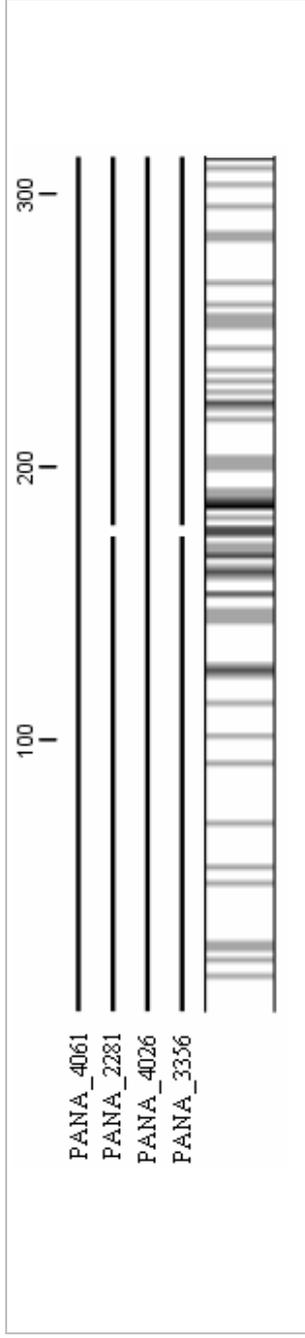
Figure 4.2: Diagram of a ClustalX alignment of the three flagellin homologues in *P. ananatis* LMG20103. The conservation plot indicates aa residues shared in all four copies in white, those shared among two/three flagellin homologues are in grey, while those unique to one FliC are in black.
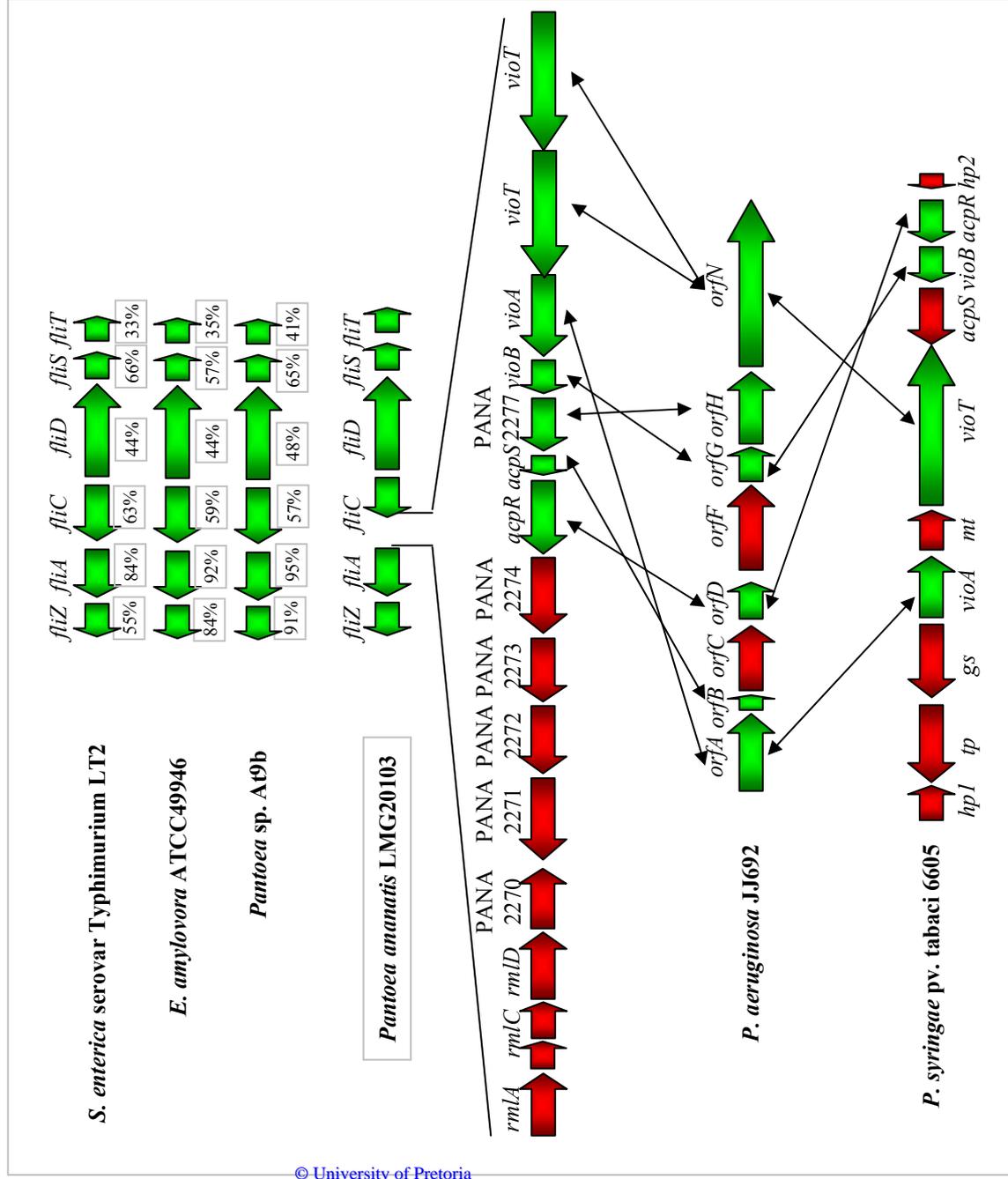
Figure 4.3: Schematic diagram showing the location of the *P. ananatis* LMG20103 flagellin glycosylation island (GI) which is absent from the compared *Enterobacteriaceae*. *P. ananatis* GI genes sharing homologues in the glycosylation islands of *Pseudomonas aeruginosa* JJ692 and *P. syringae* pv. *tabaci* 6605 are shown in green, while those unique to each GI locus are coloured in red.
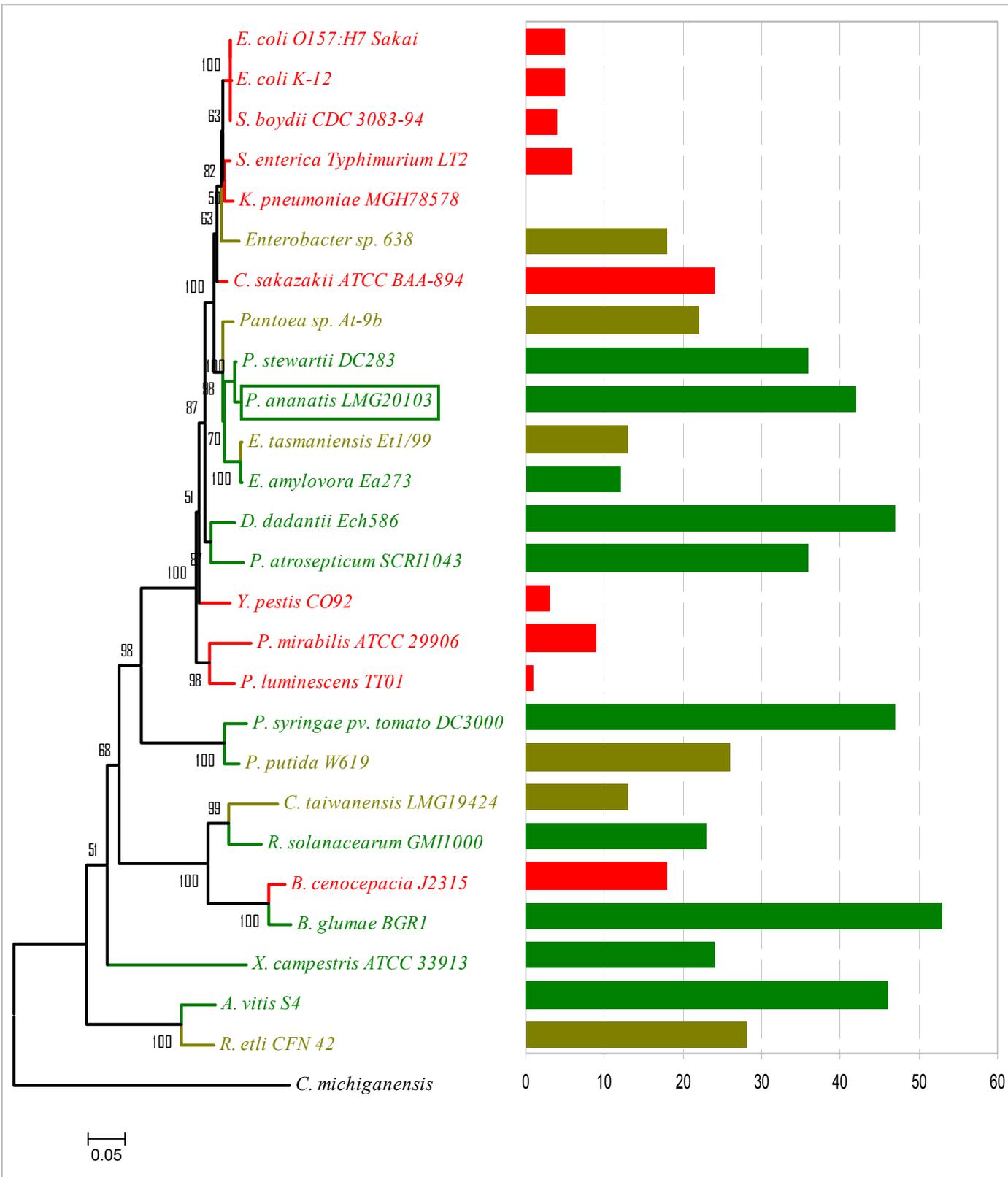
Figure 4.4: Number of methyl-accepting chemotaxis proteins (MCPs) in *P. ananatis* LMG20103 compared to those found in phylogenetic relatives. A neighbour-joining phylogeny was constructed on the basis of a ClustalW alignment of the Gyrase B amino acid sequences with bootstrap support (n = 1,000). Organisms associated with animal hosts are coloured in red, while non-pathogenic plant-associated bacteria are coloured olive and plant-pathogenic bacteria are coloured in green.
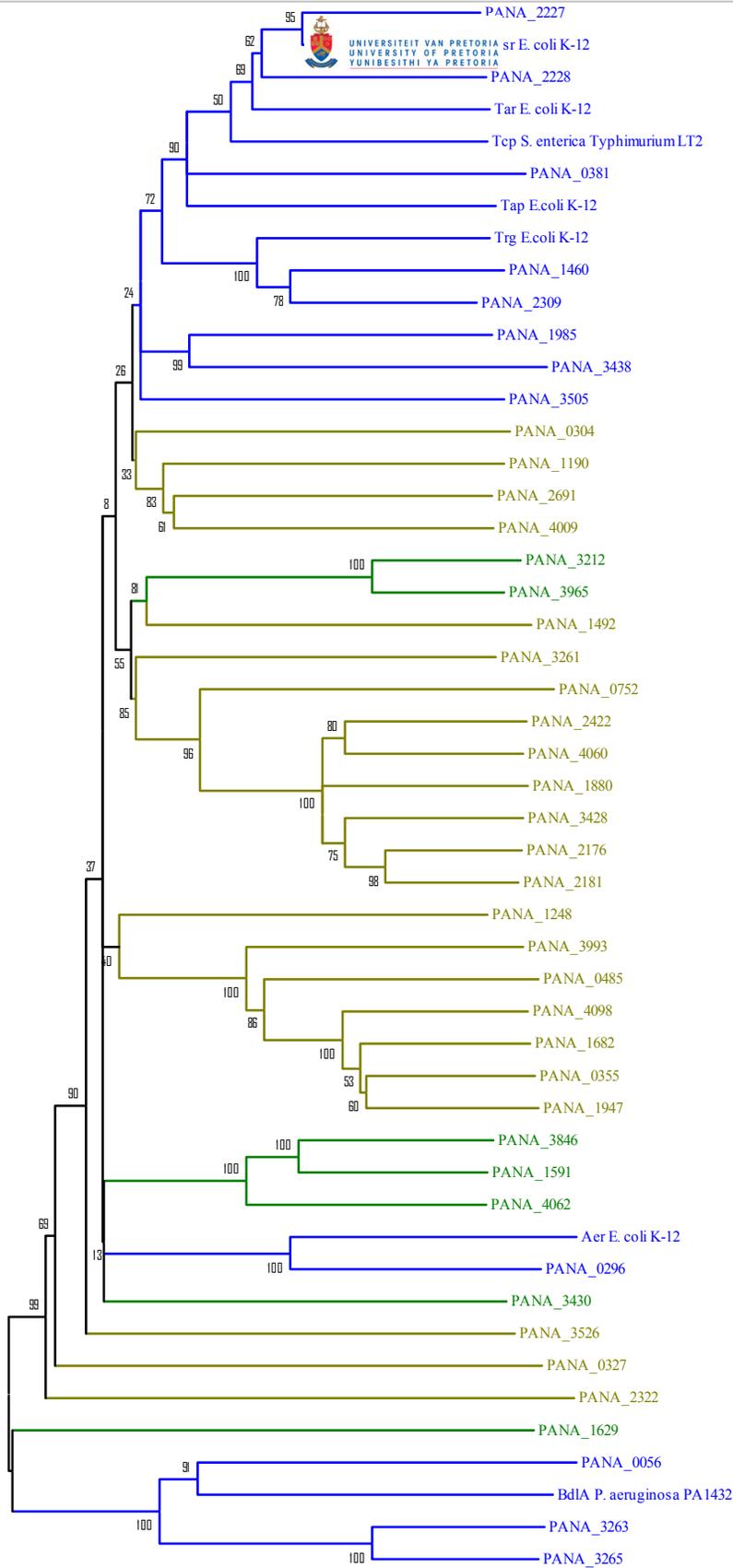
Figure 4.5: Neighbour-joining tree of a clustalW alignment of the amino acid sequences for the methyl-accepting chemotaxis proteins (MCPs) in *P. ananatis* LMG20103, as well as the Tsr (P02942), Tar (P07017), Trg (P05704), Tap (P07018) and Aer (P50466) receptors of *E. coli* K-12, the Tcp (Q02755) receptor of *S. enterica* serovar Typhimurium LT2 and the BdlA (Q9I3S1) receptor of *P. aeruginosa* PA1432. Sequences showing homology among plant-associated bacteria only are shown in olive, while those restricted to plant pathogens only are shown in green. Those receptors with homologues in plant- and animal-associated and environmental bacteria are coloured in blue. Bootstrap (n = 1,000) values are shown.
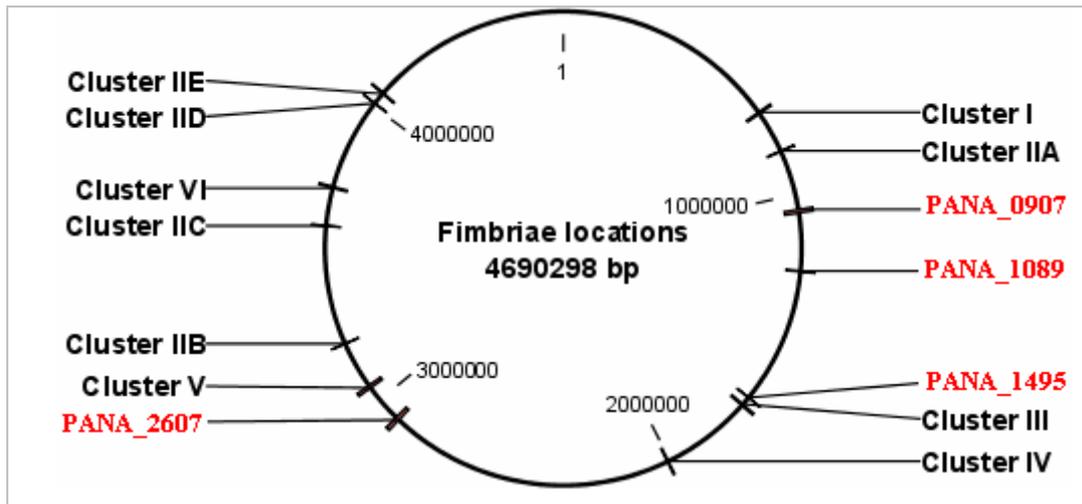
Figure 4.6: Chromosomal locations of locations of fimbrial clusters I-VI and non-fimbrial adhesins encoded by PANA_0907, PANA_1089, PANA_1495 and PANA_2607 (in red) on the *P. ananatis* LMG20103 genome.
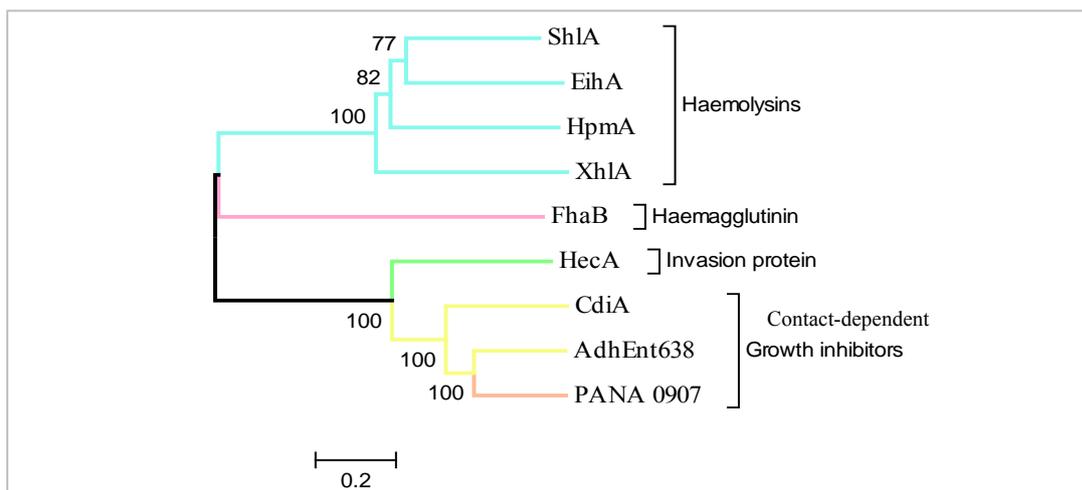
Figure 4.8: Neighbour-joining tree based on clustalW alignment of the PANA_0907 aa sequence and proteins of known function in the ShlA/FhaB/Hec family: FhaB (*B. pertussis* AAA22974), HecA (*D. dadantii* AAN38708), CdiA (*E. coli* AAZ57198) and the haemolysins ShlA (*S. marcescens* AA50323), EihA (*Edwardsiella tarda* AAQ16-190), XhlA (*Xenorhabditis nematophila* AAV33651) and HmpA (*Proteus mirabilis* AAA25657). A putative adhesin from *Enterobacter* sp. 638 (Ent638_ 0053) with highest homology to PANA_0907, was also included.
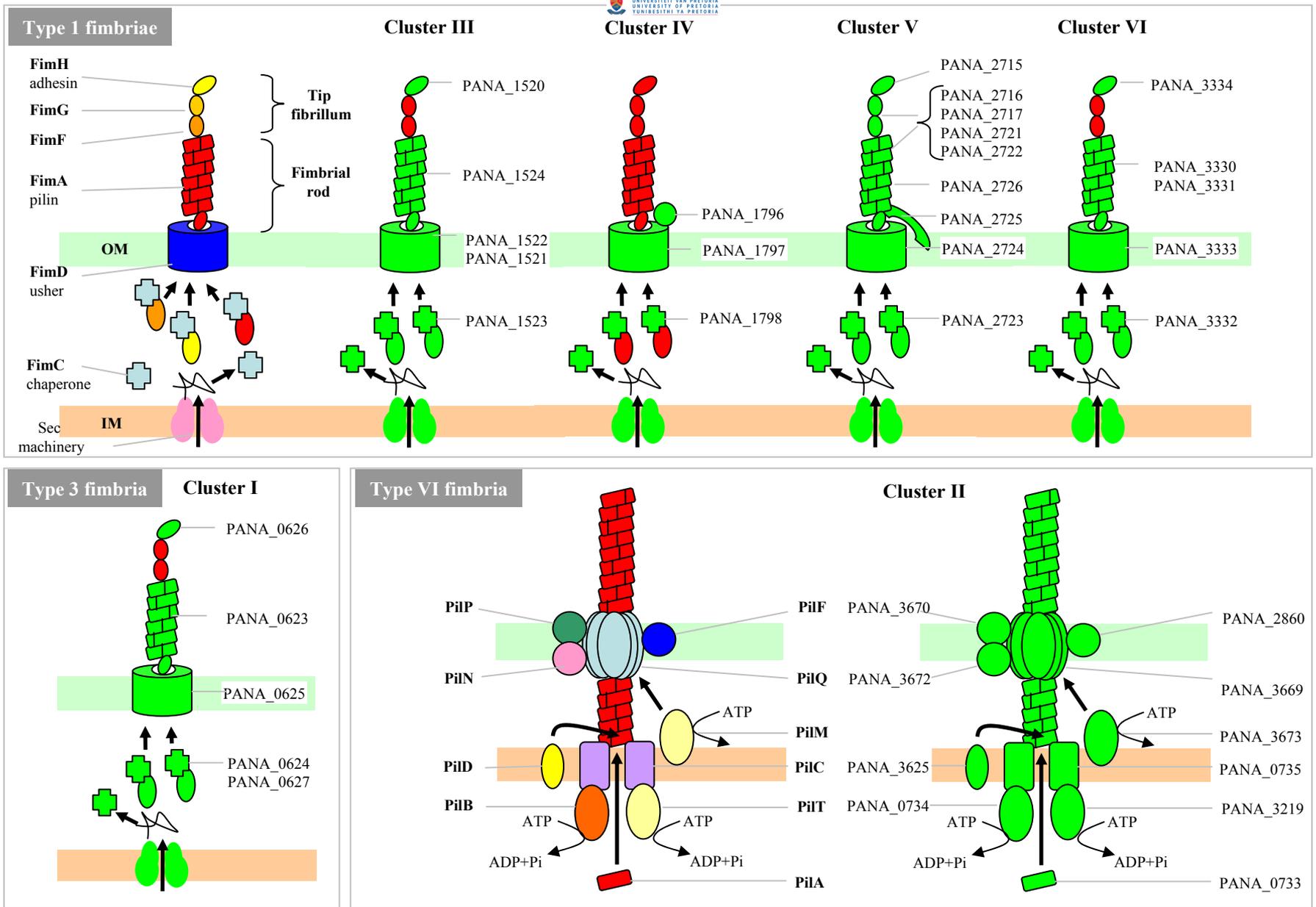
Figure 4.7: Schematic diagrams of the fimbriae encoded by *P. ananatis* LMG20103. Diagrams of assembly components for the type 1, 3 and 4 fimbriae are shown. Proteins with homologues in *P. ananatis* are coloured in green, while proteins lacking homologues in *P. ananatis* are coloured in red.
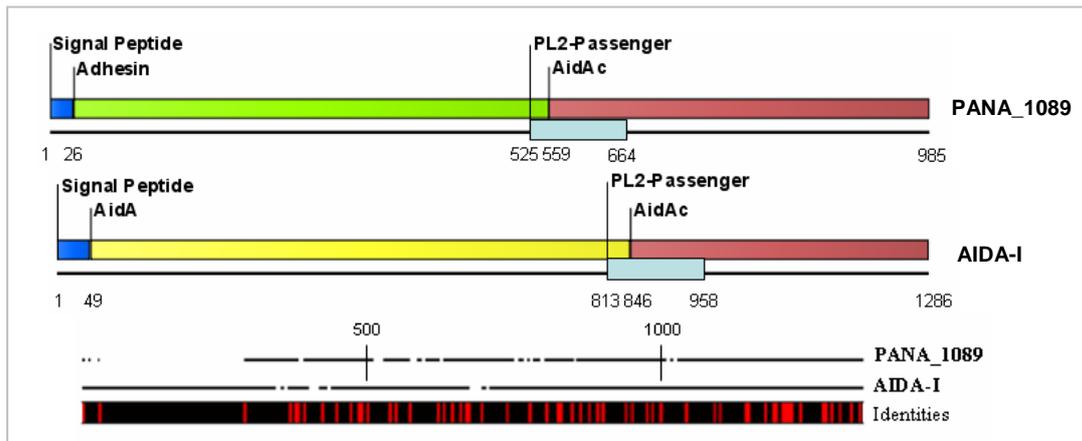
Figure 4.9: Conserved domains in the PANA_1089 protein and *E. coli* O157:H7 str. Sakai AidA-I adhesin (ECs0362). The PL2-passenger and AidAc domains and an N-terminal signal peptide are shared in common, while different domains are noted at the N-terminal end.
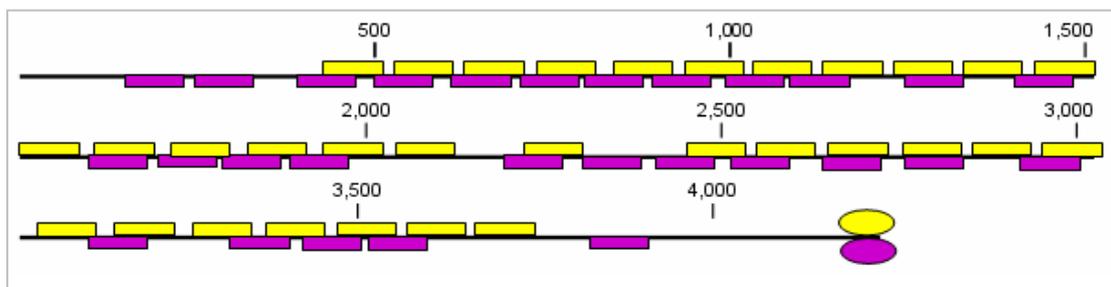
Figure 4.10: Schematic diagram showing the locations of Ig repeat domains in the adhesin proteins of *P. ananatis* (yellow) and *Yersinia mollaretti* ATCC 43969 (pink). Putative adhesive domains are shown at the C-terminal.
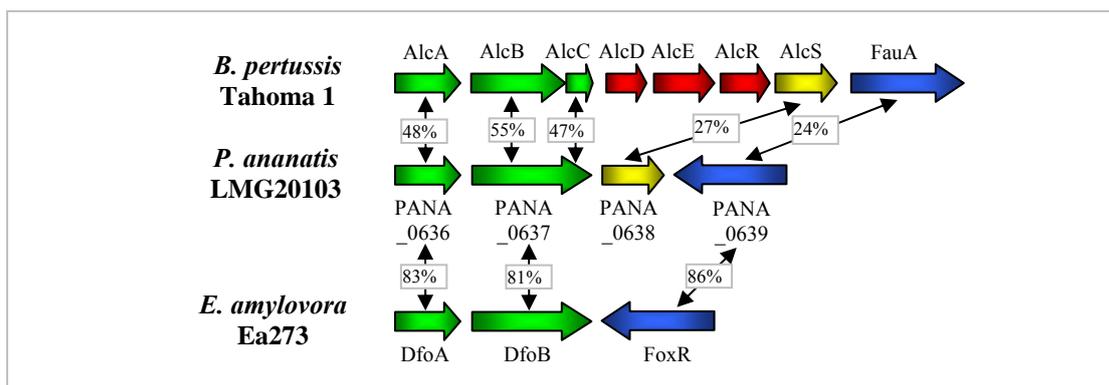
Figure 4.11: The ferrioxamine biosynthetic clusters of *B. pertussis* Tahoma 1, *P. ananatis* LMG20103 and *E. amylovora* ATCC49946. Genes conserved among all three are shown in green, those unique to *B. pertussis* in red and those in *P. ananatis* and *B. pertussis* only in yellow. Blue arrows indicate the cognate receptors. Percentage amino acid identities are shown.

207

Figure 4.12: Neighbour-joining tree showing the phylogeny of the amino acid sequences for the siderophore receptors encoded in *P. ananatis* LMG20103 and homologues in *S. enterica* (Sent), *E. coli*, *P. agglomerans* (Pag), *Erwinia amylovora* (Eamyl), *P. atrosepticum* (Patro), *S. flexneri* (Sflex) and *P. syringae* (Psyr). Included are representatives of coprogen (FhuE), ferrichrome (FhuA), ferrioxamine (Fox), enterobactin (IroN, FepA, CirA) and ferric dicitrate (FecA) outer membrane receptors. The amino acid sequence for the haemophore receptor HasR of *S. marcescens* was used to root the tree.

Figure 4.13: Schematic diagram of the predicted iron transport systems in *P. ananatis* LMG20103.

Figure 4.14: Comparison between the LPS core loci of *S. marcescens* N28b (Coderch et al. 2004), *K. pneumoniae* C3 [core type 1] (Regué et al. 2005b) and *P. ananatis* LMG20103 with Artemis Comparison Tool (T-BlastX). Genes encoding proteins involved in inner core synthesis are coloured in yellow, while those involved in outer core domain synthesis are coloured in orange. *P. ananatis* genes with homologues in *S. marcescens* only are shown in blue, while the CDS with homology in *K. pneumoniae* only is shown in red. The percentage amino acid identity between the respective CDSs is shown.
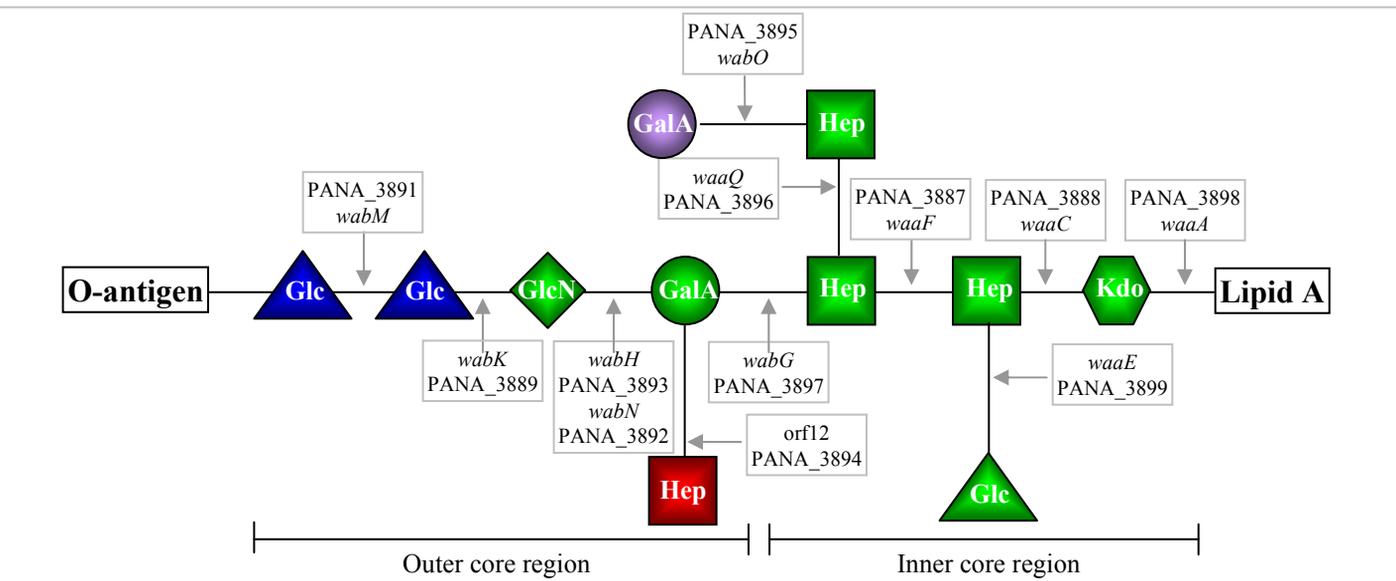
Figure 4.15: Schematic diagram of the *P. ananatis* LMG20103 LPS core domain as inferred from *K. pneumoniae* and *S. marcescens* (Coderch et al. 2004; Regué et al. 2005b). The sugar residues Kdo (3-deoxy-α-D-*manno*-octo-2-ulopyranosonic acid), Hep (Heptose), Glc (glucose), GlcN (glucosamine), GalA (galacturonic acid) incorporated in *P. ananatis* as well as *S. marcescens* and *K. pneumoniae* core type 1 and type 2 are coloured in green, while blue indicates those sugar subunits incorporated in *P. ananatis*, *S. marcescens* and *K. pneumoniae* core type 2 (but not core type 1) and the residue in purple is found in *P. ananatis* and *K. pneumoniae* core type 1 and 2 only and the red residue is incorporated in the outer core of *S. marcescens* and *P. ananatis* only.
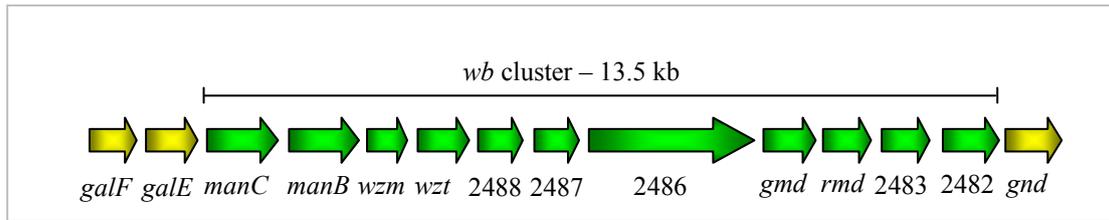
Figure 4.16: The *wb* gene cluster in *P. ananatis* LMG20103 with flanking genes coloured in yellow. PANA_2482-2483 and PANA_2486 encode glycosyltransferases, while PANA_2488 encodes a putative methyltransferase and PANA_2487 encodes a hypothetical protein of unknown function.
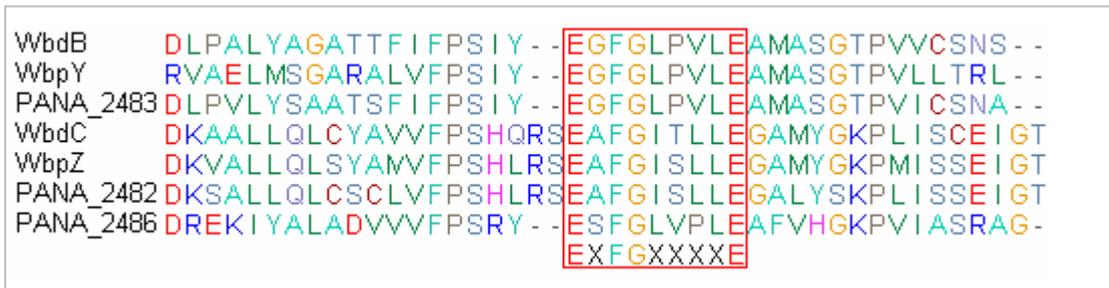

Figure 4.17: ClustalW alignment of a conserved region in the mannosyltransferases of *E. coli* O9 (Wbd), *P. aeruginosa* A-band polysaccharide rhamnosyltransferases (Wbp) and the rhamnosyl-transferases in *P. ananatis* LMG20103. The red block denotes a conserved amino acid region found in both mannosyl- and rhamnosyltransferases (Geremia et al. 1996).

210
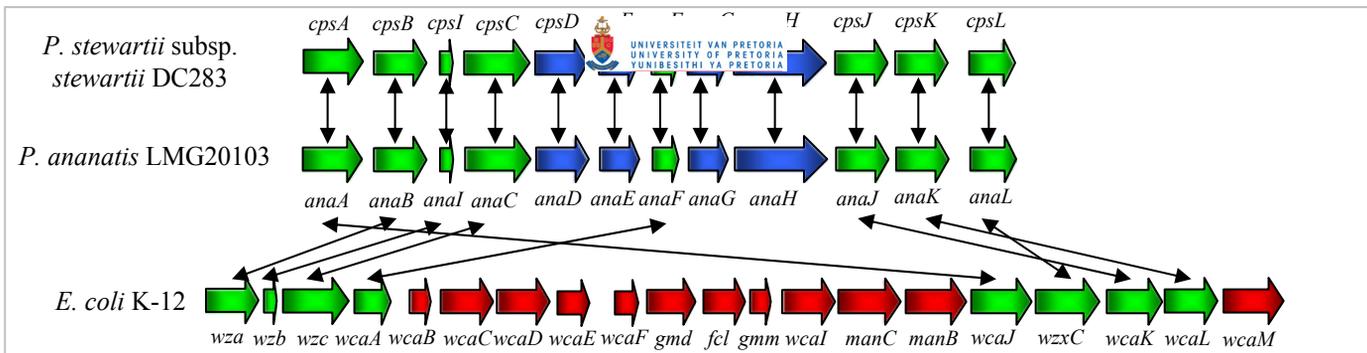
Figure 4.18: Schematic diagram of the *P. stewartii* DC283 stewartan, *P. ananatis* LMG20103 ananatan and *E. coli* colanic acid loci. Those genes with homologues shared in all three clusters are coloured in green, while those with homologues in the stewartan and ananatan loci are shown in blue and those unique to the colanic acid cluster in red.

Table 4.6: Amino acid identities of the ananatan biosynthetic proteins in *P. ananatis* LMG20103 compared to those involved in biosynthesis of stewartan, amylovoran and colanic acid.

| *P. ananatis* LMG20103 | | *P. stewartii* DC283 | | | E.amylovora Ea273 | | | *E. coli* K-12 | | |
Locus-tag	Protein	Locus-tag	Protein	% aa ID	Locus-tag	Protein	% aa ID	Locus-tag	Protein	% aa ID
PANA_2506	AnaA	AAO05910	CpsA	93%	CAA54879	AmsG	77%	AAC77848	WcaJ	29%
PANA_2505	AnaB	AAO05911	CpsB	97%	CAA54880	AmsH	81%	AAC77833	Wza	70%
PANA_2504	AnaI	AAO05912	CpsI	94%	CAA54881	AmsI	69%	AAC77834	Wzb	62%
PANA_2503	AnaC	AAO05913	CpsC	96%	CAA54882	AmsA	74%	AAC77835	Wzc	55%
PANA_2502	AnaD	AAO05914	CpsD	95%	CAA54884	AmsC	35%	-	-	-
PANA_2501	AnaE	AAO05915	CpsE	96%	CAA54883	AmsB	74%	-	-	-
PANA_2500	AnaF	AAO05916	CpsF	93%	-	-	-	AAC77836	WcaA	36%
PANA_2499	AnaG	AAO05917	CpsG	94%	CAA54885	AmsD	32%	-	-	-
PANA_2498	AnaH	AAO05918	CpsH	95%	CAA54887	AmsF	60%	-	-	-
PANA_2497	AnaJ	AAO05919	CpsJ	95%	CAA54888	AmsJ	70%	AAC77850	WcaK	55%
PANA_2496	AnaK	AAO05920	CpsK	95%	CAA54889	AmsK	73%	AAC77851	WcaL	51%
PANA_2495	AnaL	AAO05921	CpsL	95%	CAA54890	AmsL	76%	AAC77849	WzxC	27%
Average identity				95%			67%			49%
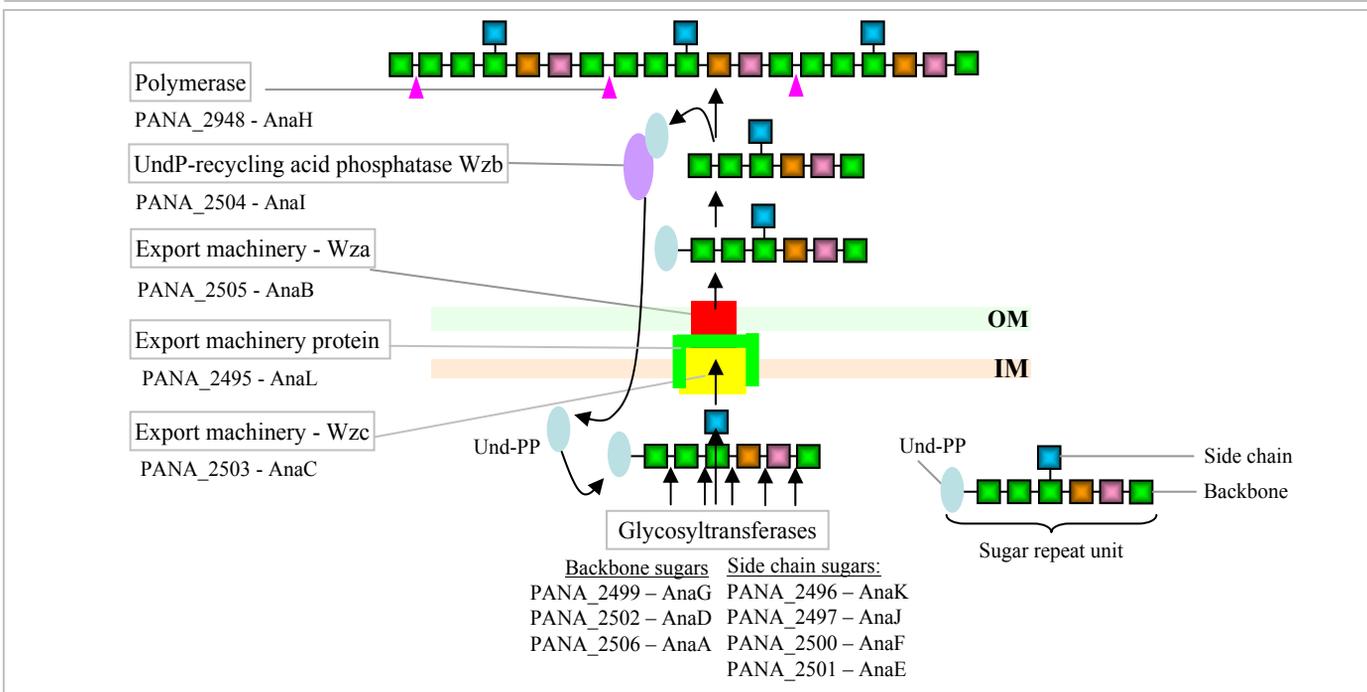
Figure 4.19: Schematic model of the putative biosynthetic pathway of ananatan by *P. ananatis* LMG20103 predicted on the basis of presence of conserved domains and homology to characterised stewartan and amylovoran biosynthetic proteins.
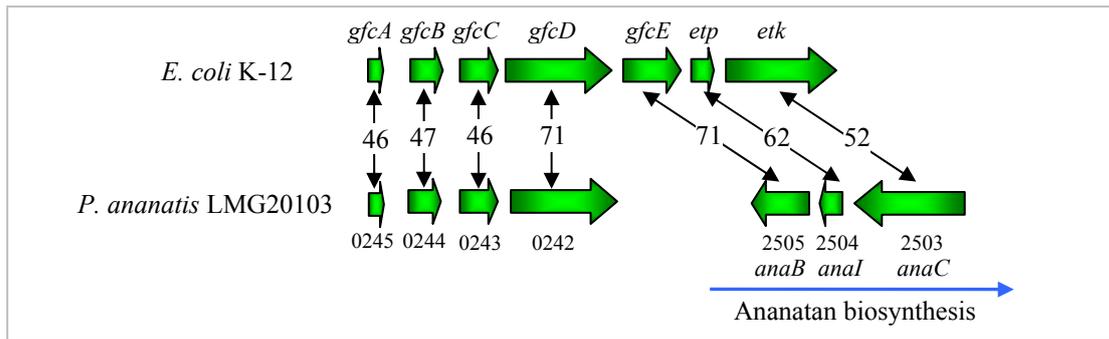
Figure 4.20: The group four capsule biosynthetic locus in *E. coli* K-12 and its homologues in *P. ananatis* LMG20103. Percentage amino acid identities of the encoded proteins are shown.
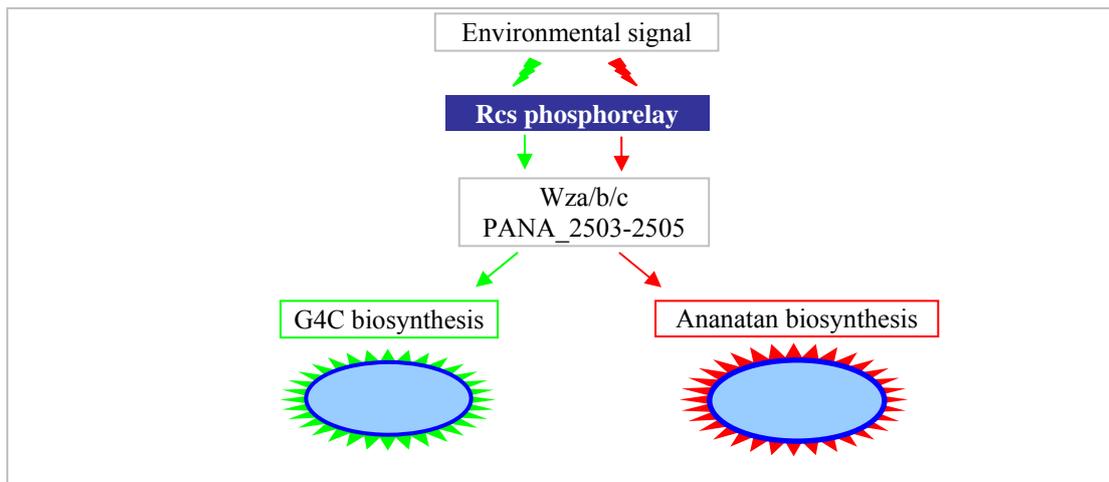

Figure 4.21: Schematic diagram of the putative shunt of the polymerase and export machinery for biosynthesis of alternate group 4 capsule and ananatan via the Rcs phosphorelay pathway.
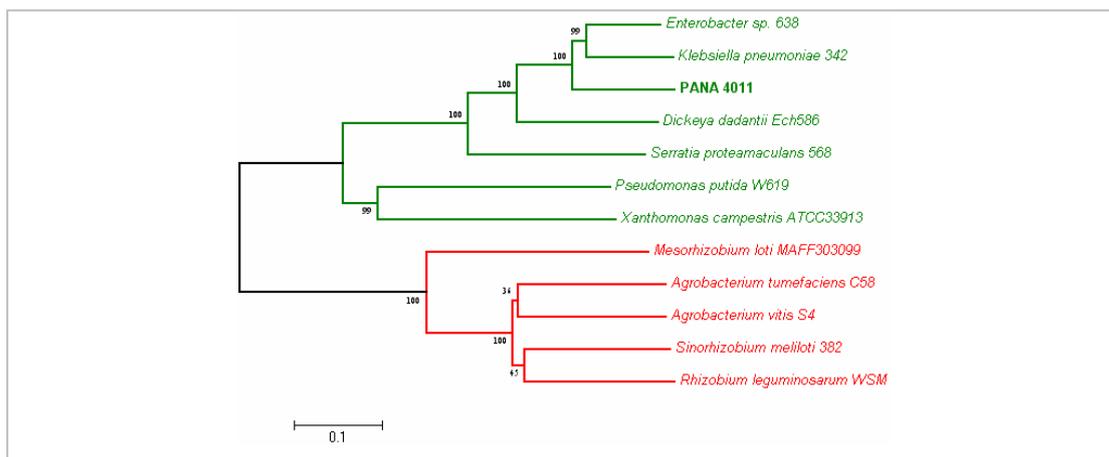

Figure 4.22: Neighbour-joining tree of ClustalW aligned amino acid sequences of the β-1,2-glucan synthase NdvB/ChvB. Branches in red indicate those bacteria which encode a ChvA/NdvA ATP-binding protein involved in β-1,2-glucan export. Branches coloured in green indicate those bacteria where no ChvA/NdvA homologue is present. Bootstrap values (n = 1,000) are displayed.
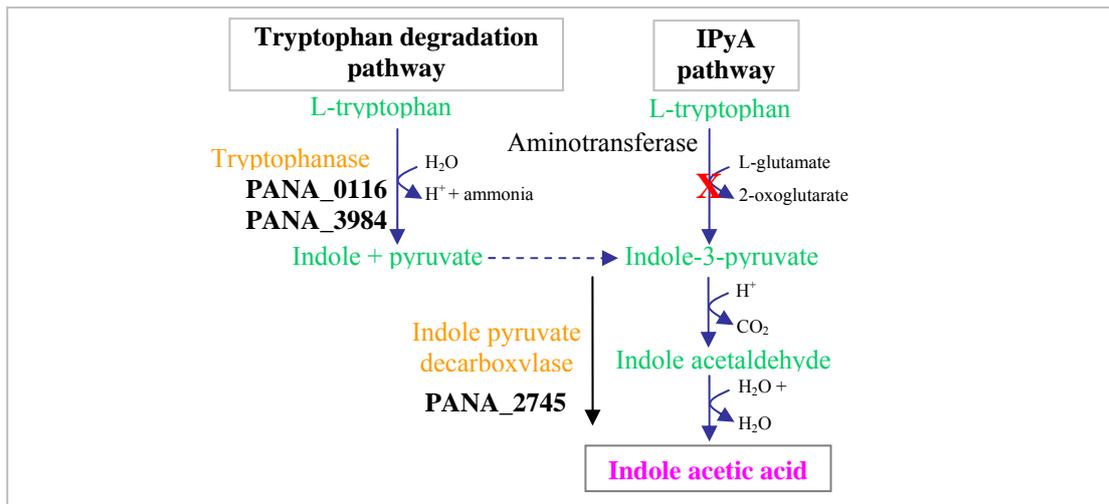
Figure 4.23: Schematic diagram of IAA production adapted from the Metacyc metabolic pathway database (Caspi et al. 2008). Those enzymes for which homologues are present in *P. ananatis* LMG20103 are shown in orange, while those reactions for which homologous enzymes are absent are indicated with a red cross.
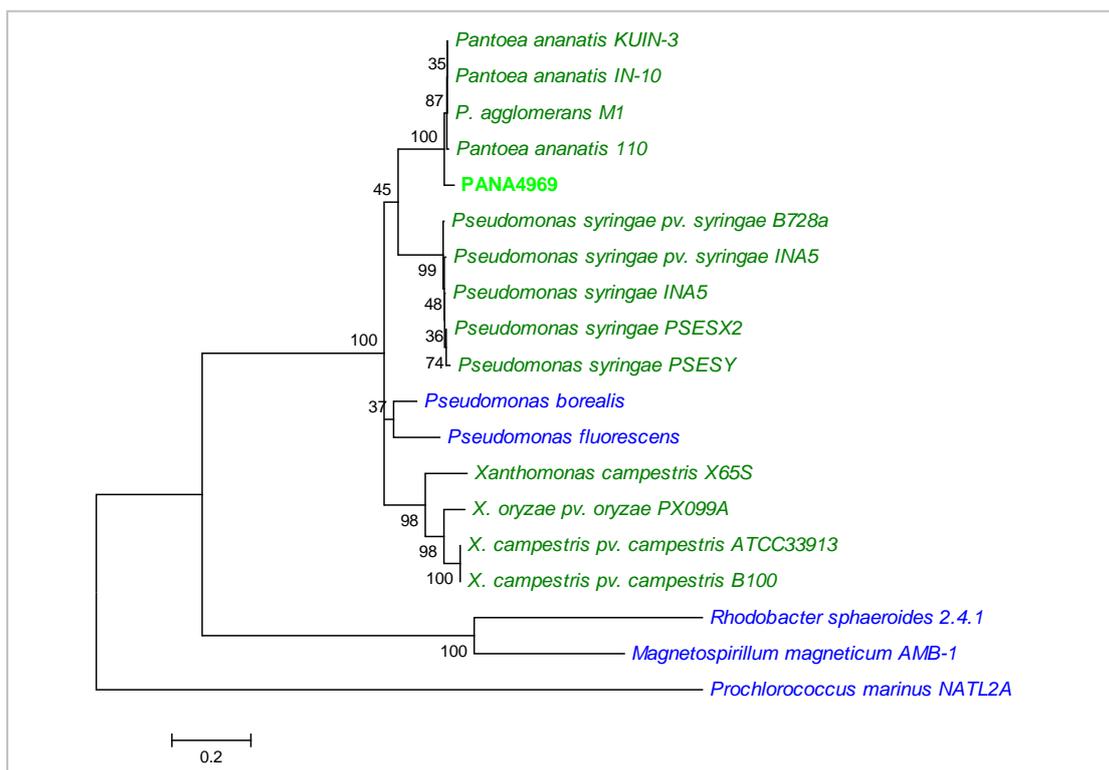

Figure 4.24: Neighbour-joining tree based on a ClustalW alignment of the ice-nucleation protein (InaA) amino acid sequences. Bootstrap support (n = 1,000) is shown. Potential plant pathogens are drawn in green while plant-associated and environmental bacteria are shown in blue.
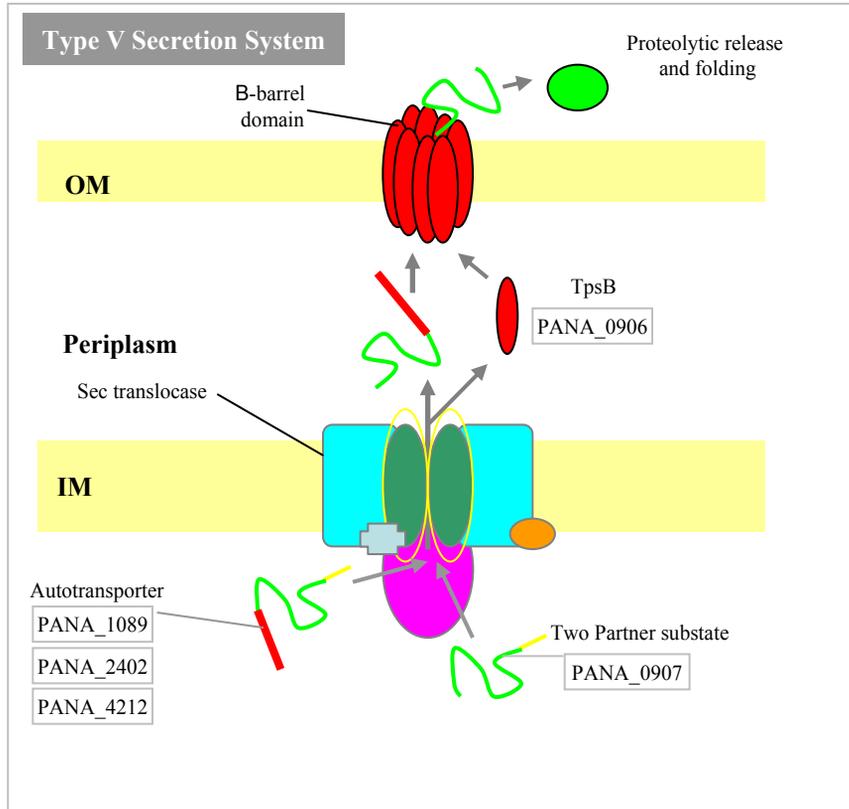
213

Figure 4.25: Schematic diagram of the Sec translocation pathway for protein transport across the inner membrane (IM) in *P. ananatis* LMG20103.

Figure 4.26: Schematic diagram of the *P. ananatis* LMG20103 components for the Twin-arginine translocation pathway involved in protein transport across the inner membrane (IM).

Figure 4.27: Schematic diagram of the Type I secretion pathway in *P. ananatis* LMG0103 as predicted on the basis of sequence homology and domain conservation.

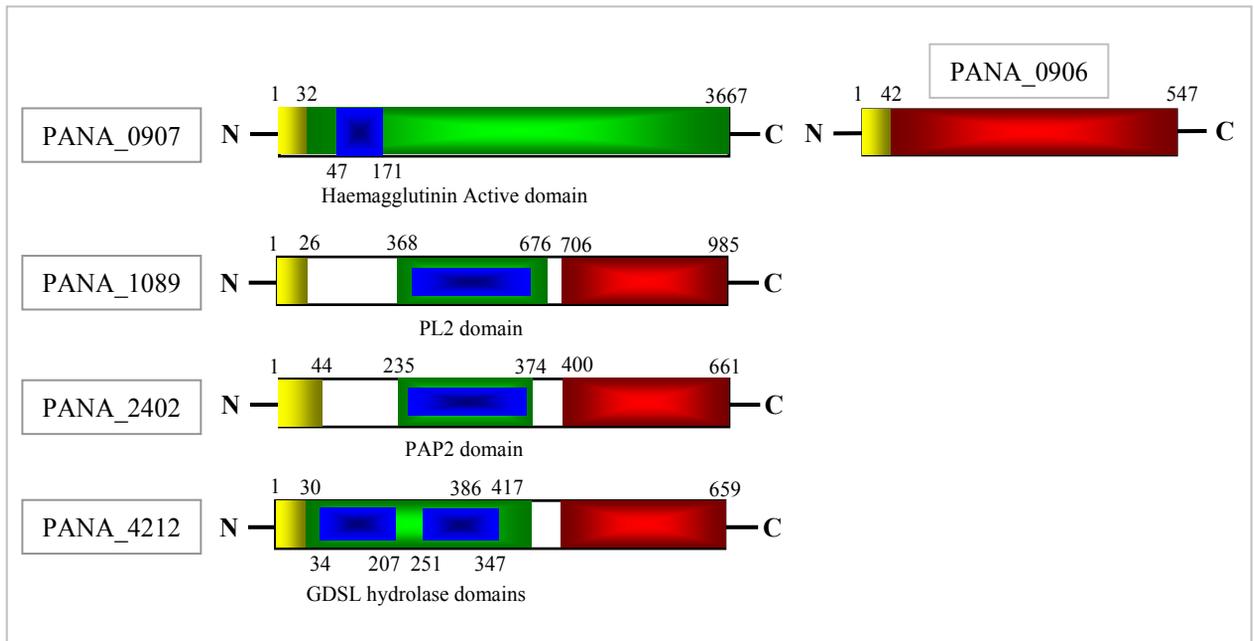Figure 4.28: Schematic diagram of the Type V secretion systems in *P. ananatis* LMG20103.

Figure 4.29: Schematic diagram of the Type V secreted substrates in *P. ananatis* LMG20103. The N-terminal signal peptide is shown in yellow, the passenger domain in green and the C-terminal β-domain in red. The active domains required for the function of the passenger proteins is shown are depicted in blue. The TPS transporter protein encoded by PANA_0906 is shown separately.
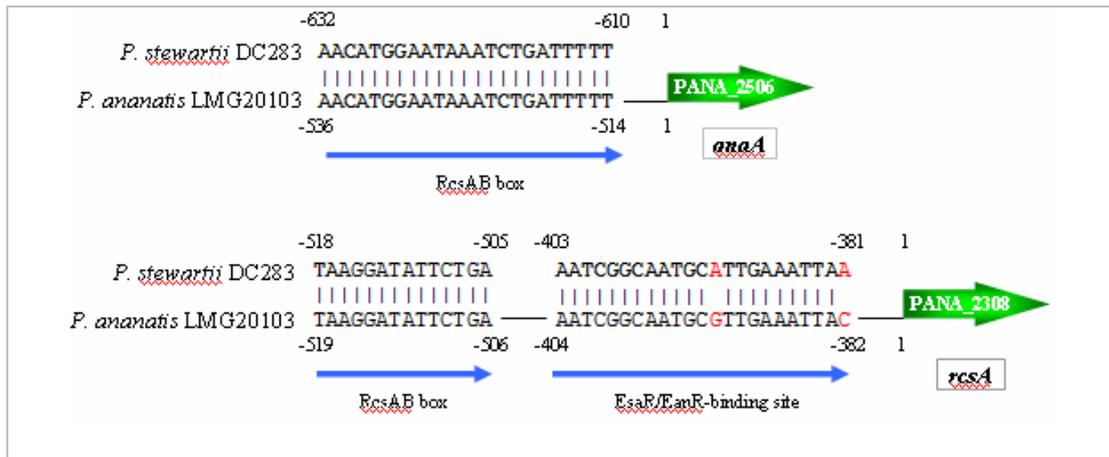
Figure 4.30: Alignment of the DNA fragments upstream of the EPS biosynthetic gene *anaA/cpsA* and EPS regulator *rcsA* in *P. stewartii* subsp. *stewartii* DC283 and *P. ananatis* LMG20103. The RcsAB and EsaR/EanR box alignments are shown, with red bases indicating sequence differences.

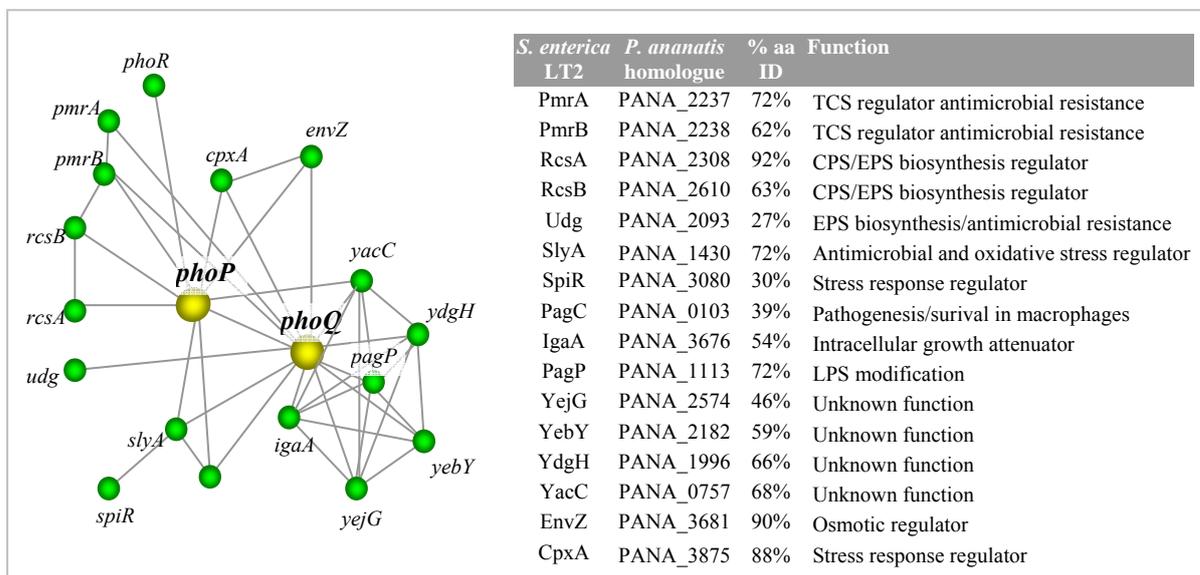| S. enterica LT2	P. ananatis homologue	% aa ID	Function
PmrA	PANA_2237	72%	TCS regulator antimicrobial resistance
PmrB	PANA_2238	62%	TCS regulator antimicrobial resistance
RcsA	PANA_2308	92%	CPS/EPS biosynthesis regulator
RcsB	PANA_2610	63%	CPS/EPS biosynthesis regulator
Udg	PANA_2093	27%	EPS biosynthesis/antimicrobial resistance
SlyA	PANA_1430	72%	Antimicrobial and oxidative stress regulator
SpiR	PANA_3080	30%	Stress response regulator
PagC	PANA_0103	39%	Pathogenesis/surival in macrophages
IgaA	PANA_3676	54%	Intracellular growth attenuator
PagP	PANA_1113	72%	LPS modification
YejG	PANA_2574	46%	Unknown function
YebY	PANA_2182	59%	Unknown function
YdgH	PANA_1996	66%	Unknown function
YacC	PANA_0757	68%	Unknown function
EnvZ	PANA_3681	90%	Osmotic regulator
CpxA	PANA_3875	88%	Stress response regulator

Figure 4.31: Functional association diagram of PhoPQ for *S. enterica* LT2 constructed with String 8.2. The accompanying table shows homologues for the amino acid sequences for all regulated genes present in *P. ananatis* LMG20103 suggesting a similar regulatory network exists in this organism.

Chapter 5:

An in depth *in silico* analysis of the Type VI secretion systems of *Pantoea ananatis* LMG20103

ABSTRACT

The Type VI secretion system (T6SS) is a recently identified pathogenicity determinant utilised by many plant- and animal-pathogenic bacteria. Analysis of the *P. ananatis* LMG20103 genome revealed that it carries three loci encoding this secretion system. In the absence of a Type II, III and IV secretion systems these are likely to play a major role in its pathogenesis. An *in silico* analysis of the three loci indicated that they encode three distinct functional T6SSs. The first system likely plays a role in persistence in the plant environment. The locus II and III T6SSs have likely evolved, through the horizontal acquisition of pathogenicity effectors, to become major pathogenicity determinants in plant and vertebrate hosts respectively.

INTRODUCTION

In 2006 a novel secretion system was identified in two mammalian pathogens, *Vibrio cholerae* and *Pseudomonas aeruginosa* (Pukatzki et al. 2006; Mougous et al. 2006) and was shown to inject cytotoxin into their host cells in a contact-dependent manner (Yahr, 2006). In keeping with the numerical classification scheme for secretion systems it was termed the Type VI secretion system (T6SS). T6SS loci have since been identified in more than 80 bacterial genomes largely restricted to the Proteobacteria, mainly those of animal pathogens, including *Salmonella enterica* and *Yersinia pestis* and plant pathogens such as *Pectobacterium atrosepticum* and *Agrobacterium tumefaciens*, but also in plant symbionts and marine microorganisms (Boyer et al. 2009). No links between the presence of T6SS and phylogeny or ecology can be observed which indicates many microorganisms may have obtained it through horizontal acquisition (Boyer et al. 2009). Some bacteria have obtained multiple loci encoding T6SS. Six T6SS loci are present in *Burkholderia pseudomallei*, with 2.3% of its genome involved in Type VI secretion (Shalom et al. 2007). Four and six loci have also been identified in *Y. pestis* and the insect-pathogen *Photorhabdus luminescens*, respectively (Schell et al. 2007). These differ significantly in the genetics and structure of the secretion apparatus as well as in the substrates proteins that are secreted (Boyer et al. 2009).

Various functions have been linked to T6SSs. The T6SS of *V. cholerae, P. aeruginosa* and *Aeromonas hydrophila* secrete haemolysin co-regulated proteins

218

(Hcp) which have cytolytic effects on host macrophages and it is also involved in virulence in the fish pathogen *Edwardsiella tarda* and the human glanders pathogen *B. mallei* (Suarez et al. 2008; Zheng and Leung, 2007; Schell et al. 2007). Rather than a cytotoxic role, the *S. enterica* and *Y. pestis* T6SSs are involved in persistence in host cells in late stages of infection (Robinson et al. 2009; Parsons and Heffron, 2005). In enteroaggregative *E. coli* it has been implicated in biofilm formation and adherence to abiotic surfaces (Aschtgen et al. 2008). In the phytopathogen *P. atrosepticum* the T6SS is involved in potato stem and tuber infection (Liu et al. 2008) and in *A. tumefaciens* it plays a role in tumourigenesis on host plants (Wu et al. 2008). Two of the three Type VI loci of the opportunistic human pathogen *P. aeruginosa* were also shown to play a role in plant pathogenesis, while a third is involved in chronic lung infection (Lesic et al. 2009). The occurrence of T6SSs in non-pathogenic bacteria indicates they may be involved in non-pathogenic functions including host-symbiont communication, chemotaxis and cell-cell communication (Boyer et al. 2009). In *Rhizobium leguminosarum*, expression of T6SS is involved in symbiosis, blocking nodulation and nitrogen fixation on legume species that are not natural hosts, thus determining host specificity (Bladergroen et al. 2003).

Genes encoding components of the Type VI secretion apparatus and secreted effectors are co-localised in large loci which carry 12-25 protein coding sequences (CDSs) (Filloux et al. 2008). Boyer et al. (2009) identified 16 CDSs conserved among most T6SS loci, belonging to 16 distinct conserved orthologous groups (COGs). These are loosely grouped according to their cellular localisation. A further 13 conserved accessory proteins are distributed among T6SS loci and include specialised transcriptional and post-transcriptional regulators. Three cytoplasmic components have been functionally elucidated. ClpV (COG0542), a hexameric ATPase, is purported to drive secretion (Filloux, 2009). It interacts with the cytoplasmic proteins IglA (COG3516) and IglB (COG3517) which form a tubular complex showing structural homology to the tail sheath of bacteriophage T4 and further act as chaperones that unfold effector proteins for efficient secretion (Blondel et al. 2009; Boyer et al. 2009). Six proteins constitute inner membrane and periplasmic components of the secretion apparatus (Boyer et al. 2009). The IcmF (COG 3523) and DotU (COG3455) proteins share homology with T6SS components and may play roles in membrane stabilisation of the secretion apparatus and improve host cell

© University of Pretoria

adherence and translocation efficiency (Blondel et al. 2009). Two periplasmic lipoproteins, COG3521 and COG3522, are predicted to interact with DotU and IcmF to form a continuous channel connecting the inner and outer membrane (Boyer et al. 2009). Two further proteins, COG3515 and COG3520, represent inner membrane components of unknown function.

Three proteins constitute the portion of the secretory apparatus which passes through the outer membrane into the extracellular milieu and pierces the host cell (Filloux et al. 2008). These proteins structurally resemble phage proteins and thus a phage origin for the T6SS has been hypothesised. The COG3518 protein shows homology to the T4 phage baseplate protein gp25. It also carries a putative lysozyme domain which may hydrolyse *N*-acetyl glucosamine and *N*-acetyl muramic acid of the bacterial peptidoglycan layer and could facilitate passage of the secretory apparatus through the outer membrane and potentially the host cell membrane (Shrivastava and Mande, 2008). Hcp (COG3517) proteins have a hexameric ring structure and shows structural similarity to the tail tube proteins gp19 in the T4 bacteriophage (Leiman et al. 2009). Hcp proteins arrange end to end to form continuous nanotubules which act as a channel across the cytoplasmic and outer membranes for effector transport (Pukatzki et al. 2009). The third protein, VgrG (COG3501), carries an N-terminal gp27-like Rhs_VgrG domain (TIGR03361), a baseplate gpV domain (cl11432) and a gp5 domain (PHA02596). The T4 bacteriophage gp27 protein forms a spike at the end of the bacteriophage tail, while the gp5 protein serves as lysozyme, locally dissolving the bacterial host cell wall (Haggård-Liungquist et al. 1995; Arisaka et al. 2003). These proteins enable the phage tail to puncture the bacterial cell wall to allow bacteriophages to inject DNA into their bacterial hosts (Filloux, 2009). The genetic and structural homology to bacteriophages T4 tail spike gp27 and gp5 proteins suggest VgrG forms a similar tail spike complex on the Hcp tubules that allows the T6SS to puncture both the bacterial and host cell membranes (Filloux, 2009).

Type VI effectors do not carry canonical signal peptides and are secreted in a Sec- or Tat-independent manner, crossing the inner and outer bacterial membranes in a single step (Lesic et al. 2009). The two main secreted effectors identified to date are the Hcp and VgrG proteins which also serve as structural components of the secretion apparatus (Pukatzki et al. 2009). The Hcp protein acts as pore-forming toxin, resulting

220
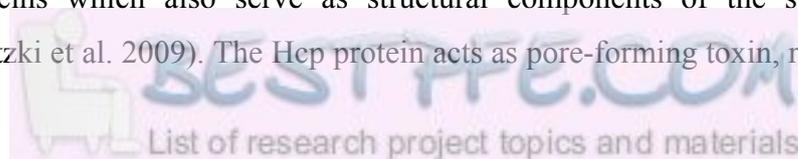
in ion leakage and apoptosis of host cells (Pukatzki et al. 2009). The overexpression of Hcp in the phytopathogen *P. atrosepticum* also resulted in increased virulence (Mattinen et al. 2007). Some VgrG proteins, known as "evolved" VgrGs, contain C-terminal extensions with domains that have been linked to pathogenic functions, including cross-linking of actin in host cells, peptidoglycan degradation and ADP-ribosylation (Blondel et al. 2009). The *S. enterica* VgrG extension contains an S-pyocin domain (COG5529) with a potential role in bacterial cell killing (Blondel et al. 2009). Short tandem repeats and COG4253 functional domains in some VgrGs show structural similarity to the adhesins fibronectin and YadA suggesting a role in host cell adhesion (Pukatzki et al. 2007). It has been proposed that, following completion of the secretory apparatus, the VgrG/Hcp complex passes through the outer membrane and penetrates the host cell membrane. Subsequently the evolved VgrG C-terminal domains will be unfolded and exposed to the target cell cytoplasm where, together with Hcp, it can carry out its pathogenic function (Leiman et al. 2009). The C-terminal domain may also be cleaved off to interact with distant host cell targets (Pukatzki et al. 2009). Several other T6SS substrates have been identified. The *R. leguminosarum* T6SS secretes an RbsB-like protein that has a role in host-range determination (Pukatzki et al. 2009). *B. mallei* secretes a deubiquitinase, TssM, which interferes with various ubiquitin-requiring host cell functions (Schell et al. 2007).

Expression of the T6SS is tightly regulated, at the transcriptional or post-translational level to ensure accurate timing in host-pathogen interactions (Schell et al. 2007). The *B. mallei* T6SS is under the control of an AraC-type transcriptional regulator and the VirAG two-component system (Boyer et al. 2009). The SsrB transcriptional regulator in *S. enterica* upregulates cytotoxicity-related genes including those for Type III secretion and exoenzyme production but downregulates genes for persistence and Type VI secretion (Parsons and Heffron, 2005). Post-translational regulation occurs through the phosphorylation and dephosphorylation activities of serine/threonine protein kinase PpkA and its antagonist serine/threonine phosphatase PppA. These de/phosphorylate a FHA-domain (Fork Head Associated) protein (COG3456) in the T6SS locus. In resting state PppA maintains a dephosphorylated Fha protein. Appropriate environmental signals trigger phosphorylation of the Fha protein by PpkA (Mougous et al. 2007). The activated Fha protein then recruits the secretion apparatus components ClpV, DotU and IcmF to the inner membrane and initiates

T6SS assembly (Boyer et al. 2009). This form of post-translation regulation has been identified in *P. aeruginosa* and *B. cenocepacia* (Boyer et al. 2009).

Pantoea ananatis is a phytopathogen that causes disease on a brought range of agronomic crops including maize, rice, onion, pineapples, melons and *Eucalyptus* (Coutinho and Venter, 2009). It is also frequently isolated as endo- or epiphyte from a wide range of plants (Coutinho and Venter, 2009). Furthermore, it has been isolated from a number of insects which act as vectors and its potential for use as biological control agent against insect pests has been demonstrated (Gitaitis et al. 2003; Watanabe and Sato, 1999). Added to this diverse lifestyle, it has been found to cause human disease (De Baere et al. 2004). The genome of the *Eucalyptus*-pathogenic strain *P. ananatis* LMG20103 was sequenced (Chapter 2; De Maayer et al. 2010). This strain was also shown to infect a number of other plant species including onion and pineapples (Chapter 6). Analysis of the genome sequence of *Pantoea ananatis* LMG20103 allowed the identification of numerous putative pathogenicity determinants (Chapter 4) but highlighted the absence of Type II, III and IV secretion systems, which form important components of the pathogenic arsenal in closely related animal- and plant-pathogens. However, three Type VI secretion loci were identified and in the absence of the above-mentioned secretory systems, Type VI secretion is expected to play a significant role in *P. ananatis* pathogenesis on its host plants and potentially on invertebrate and vertebrate animal hosts. In this chapter an in depth *in silico* analysis of the three T6SS loci was performed. Genetic, as well as structural aspects are discussed, as are the putative evolutionary origins and functions of the Type VI secretion systems of *P. ananatis* LMG20103.

MATERIALS AND METHODS

Identification of Type VI secretion loci in *P. ananatis* LMG20103

BlastN and BlastP analysis with the nucleotide and amino acid sequences for the 16 conserved proteins identified by Boyer and co-workers (2009) was used to identify T6SS loci on the genome sequence of *P. ananatis* LMG20103. By further BlastP analysis against the NCBI protein database with the amino acid sequences for proteins encoded upstream and downstream of the conserved CDSs, the full extent of the T6SS loci was determined. Three loci were identified in *P. ananatis* LMG20103 which were termed T6SS locus I-III. Once fully elucidated the amino acid sequences for the

proteins encoded in the three loci were compared by BLASTP against the NCBI protein database to find homologous loci. The presence of homologous loci in the draft genome sequences for *Pantoea ananatis* LMG2665^T and *Pantoea stewartii* subsp. *stewartii* DC283 was also determined.

Genetic and structural analyses of the *P. ananatis* LMG20103 T6SSs

The protein coding sequences in each of the three T6SS loci in *P. ananatis* LMG20103 were analysed. The three loci were aligned to homologous loci in other organisms and the gene order conservation and percentage identity between the encoded proteins determined. The presence and absence of homologues to the 16 conserved proteins identified by Boyer et al. (2009) was observed. The conserved domains in all the proteins encoded by the three loci were resolved with CDSearch (Marchler-Bauer and Bryant, 2004; Http://www.ncbi.nlm.nih.gov/Structure/cdd/). Their subcellular localisation was determined by identification of signal peptides with SignalP 3.0 (Bendtsen et al. 2004; Http://www.cbs.dtu.dk/services/SignalP/) and the presence of transmembrane helices with TMHMM 2.0 (Krogh et al. 2001; Http://www.cbs.dtu.dk/services/TMHMM/). Lipoproteins were identified by the detection of signal peptidase cleavage sites with the LipoP 1.0 webserver (Juncker et al. 2003; Http://www.cbs.dtu.dk/services/LipoP/). Subcellular localisation was validated with Psortb 3.0 (Yu et al. 2010; Http://psort.org). This data was combined with a pre-existing model from literature to construct a model of the three T6SS apparatuses in *P. ananatis* LMG20103.

Analysis of the evolutionary origins of the T6SS loci in *P. ananatis* LMG20103

The evolutionary origins of the three loci were determined as per Bingle et al. (2008) on the basis of the conserved chaperones IglA and IglB. This entailed a ClustalW alignment with the concatenated IglA and IglB amino acid sequences and a distance tree was calculated by the neighbour-joining method, excluding gaps and with a Poisson correction using Mega 4.1 (Tamura et al. 2007). Protein sequences for homologues in organisms included in the tree by Bingle et al. (2008) were also aligned and incorporated into the tree. As no homologues to these two proteins were found in T6SS locus I, a second tree was constructed using the same methodology, based of the amino acid sequences of the inner membrane protein IcmF, which is

conserved in all three *P. ananatis* loci. The IcmF homologues in the IglA/IglB-containing T6SS loci identified by Bingle et al. (2008) were also included.

<u>Analysis of putative effectors and functions of the *P. ananatis* LMG20103 T6SSs</u>
The T6SS loci and their adjacent genes were analysed for homologues to known Type VI secretion substrates. Evidence for an extracellular localisation for putative effectors was gathered by the *in silico* methods described in section 2.2. Conserved domains in the putative effectors were also identified using CDSearch (Marchler-Bauer and Bryant, 2004). Evidence for horizontal acquisition of the T6SSs and their putative effectors was gauged on the basis oligonucleotide usage and G+C deviation as described in Chapter 3. Based on the evidence from literature putative functions were ascribed to the T6SSs in *P. ananatis* LMG20103.

RESULTS

<u>Type VI secretion loci in *P. ananatis* LMG20103</u>
Using the 16 conserved proteins identified by Boyer et al. (2009) as reference, BlastP identified three putative Type VI secretion loci in *P. ananatis* LMG20103. BlastP of up- and downstream predicted ORFs elucidated the extent of the three T6SS clusters named locus I, II and III respectively which are interspersed across the chromosome (Fig. 5.1). Locus I is ~9 kb in size and encompasses 8 CDSs. It has a G+C content of 56.80%, 3.11% above the genome average and on the basis of G+C deviation and anomalous oligonucleotide usage was assigned to the ~21.4 kb horizontally acquired island HAI12 along with the adjacent nitrate/nitrite assimilation genes. Locus II is ~31 kb in size, encompasses 25 CDSs and has a G+C content 54.24%, 0.55% above the genome average. The 29.9 kb T6SS locus III encompasses 22 CDSs and has a G+C content of 53.62%, 0.08% below the genome average. Two further genes, PANA_2446 and 2447 encode homologues to haemolysin co-regulated proteins (Hcp) typically associated with T6SS loci, but are located separate from the *P. ananatis* T6SS loci (Fig. 5.1). A total of 57 CDSs, ~1.3% of the proteins encoded on the genome, are thus likely involved in this secretory function in *P. ananatis* LMG20103.

The distribution of homologous loci was determined by BlastP against the available genome sequences on the NCBI database and the draft genome sequences of *P. ananatis* LMG2665[T] and *P. stewartii* DC283. Homologues to *P. ananatis* LMG20103

locus I are restricted to all of the closely related *Erwinia* and *Pantoea* strains sequenced to date (Fig. 5.2). Similarly, homologous loci to *P. ananatis* LMG20103 locus II are prevalent in the sequenced *Erwinia* and *Pantoea* strains, with the exception of *Pantoea stewartii* DC283 (Fig. 5.2). Homologous loci are also present in a number of other bacteria, most of which belong to the family *Enterobacteriaceae*. Most of these are plant-associated bacteria, including *Serratia* species and *Pectobacterium wasabiae*, but also several opportunistic human pathogens including *Enterobacter cloacae* and *Cronobacter* species. T6SS locus III shows a more restricted distribution among the sequenced *Erwinia* and *Pantoea* species, being present in *P. ananatis* LMG20103 and the more distantly related *Pantoea* sp. At9b and *E. amylovora* strains, while this locus is not present in the more closely related *P. stewartii* DC283, *P. ananatis* LMG2665^T or in the sequenced strains of *E. pyrifoliae* and *E. tasmaniensis* (Fig. 5.2). Homologous loci are mostly found in more distantly related animal-pathogenic bacteria, including *Salmonella enterica, Yersinia pestis* and *Klebsiella pneumoniae* strains. However, representatives are also present in the plant-associated *Serratia proteamaculans* and the above-mentioned erwinias.

Genetics and structure of the *P. ananatis* LMG20103 T6SSs

Type VI secretion locus I

T6SS locus I carries eight CDSs, PANA_1650-1657. Homologues to all eight encoded proteins are only found in *Pantoea* sp. At9b showing 75% average aa identity, while seven share 88% average aa identity to proteins encoded in a *P. stewartii* subsp. *stewartii* DC283 (Fig. 5.3). The *E. tasmaniensis* Et1/99, *E. amylovora* CFBP1430 and *E. pyrifoliae* Ep1/96 loci on the other hand only share four CDSs in common with T6SS locus I, with 59% average aa identity (Fig. 5.3).

Only three locus I-encoded proteins share homology with conserved T6SS proteins identified by Boyer et al. (2009) (Fig. 5.3; Table 5.1). PANA_1655 and 1656 encode an IcmF (COG3523) and DotU (COG3455) homologue, respectively. The third conserved gene, PANA_1653, encodes a FHA-domain protein, while PANA_1650 and PANA_1652 encode PpkA and PppA homologues respectively, indicating that the T6SS locus I is post-translationally regulated. PANA_1654 encodes a homologue of the ImpM protein in *R. leguminosarum* (AAL17785 – 34% aa identity) which has been suggested to function in regulation of the PpkA/PppA and FHA-domain proteins

(Boyer et al. 2009). As homologues to these four proteins are restricted to plant-associated bacteria, it is likely that locus I T6SS expression is regulated by plant-derived signals. No ClpV1 homologue is encoded in locus I, suggesting this T6SS may assemble and function in an ATP-independent manner. Similarly, no homologues to ClpV1 were found in the *R. leguminosarum imp* T6SS locus, the *Francisella tularensis* T6SS-like cluster and one of the six predicted T6S loci in *Y. pestis* CO92 (Filloux et al. 2008). Alternatively, a ClpV homologue from locus II or III may be sequestered for assembly and functioning of the locus I T6SS when required. Two further putative structural proteins are encoded in locus I. PANA_1651 shows homology to a protein of unknown function in the T6SS loci of *P. stewartii* DC283 (ACV0283603 – 74% aa identity) and *S. proteamaculans* (Spro_3008 – 40% aa identity). SignalP and Psort analysis indicates the presence of a signal peptide and an inner-membrane localisation. LipoP analysis predicts that PANA_1657 encodes a putative outer membrane lipoprotein. It shows 75% aa identity to *Pantoea* sp. At9b Pat9b_2015, with no homologue found in *P. stewartii* DC283. It may serve a similar role to the COG3521 protein, forming a membrane channel for translocation of extracellular components of the secretion apparatus (Boyer et al. 2009).

No homologues to the Hcp and VgrG proteins common to T6SS loci in most organisms, are associated with locus I. Two Hcp homologues, are encoded by PANA_2446 and 2447, located on the chromosome separately from the three T6SS loci. These show high identity to the locus II-associated Hcp protein (PANA_2364) (Table 5.2). As there is extensive homology between locus I and II, these two Hcp homologues may be associated with locus I. The similarities in size, G+C content and homology (Table 5.2) suggest however they could also be linked to locus II. The locus I secretion apparatus thus consists of mainly inner membrane components and one outer membrane lipoprotein. The absence of VgrG, Hcp and ClpV homologues would suggest this locus encodes a non-functional T6SS with missing structural and functional components (Fig. 5.4). However, no homologues to these proteins are found in the *R. leguminosarum imp* T6SS locus either, which is nonetheless functional and secretes an RbsB-like protein (Bladergroen et al. 2003). In the absence of the Hcp/VgrG nanotubules, it is likely the locus I secretion apparatus does not puncture host cells, but it may rather secrete substrate proteins into the extracellular environment. This will need to be elucidated experimentally.

226

Type VI secretion locus II

T6SS locus II encompasses 25 genes, PANA_2349-2373. BlastP of the 25 encoded proteins, showed that the *P. ananatis* T6SS locus II is most homologous to T6SS loci in *Pantoea* sp. At9b (Pat9bDRAFT_2275-2300) and *E. amylovora* CFBP1430 (EAMY3000-3028), with 82% and 71% average aa identity in 21 proteins respectively (Fig. 5.5). Upstream of *P. ananatis* locus II, PANA_2348 encodes a 1,367aa protein sharing 72% aa identity with an Rhs element in *Cronobacter turicensis* z3032 (Ctu_00940), while downstream of locus II PANA_2374 encodes a MATE family multidrug efflux protein. Rhs elements are common features of regions of genomic rearrangement, lending strength to the argument that this locus arose through horizontal acquisition (Das and Chauduri, 2003).

Homologues to all sixteen conserved proteins identified by Boyer et al. (2009) are present, suggesting structural similarity to the predicted T6SS apparatuses in other bacteria (Fig. 5.4; Table 5.1). Additional proteins in locus II include a PANA_2353-encoded serine/threonine protein kinase (PpkA) and PANA_2360-encoded serine/threonine phosphatase (PppA) which together with the FHA-domain protein (PANA_2361) will likely exercise post-translational control over the locus II T6SS. No transcriptional regulator-encoding genes are present within locus II, implying that the *P. ananatis* locus II T6SS is solely under post-translational control. As is the case for locus I, the PpkA, PppA and FHA-domain protein sequences show highest homology to those found in plant-associated bacteria, suggesting regulation of the locus II T6SS occurs in response to plant-derived signals. Two additional locus II genes, PANA_2359 and PANA_2373 encode putative structural proteins of unknown function, with homologues found in the T6SS loci of a limited number of bacteria. The PANA_2359 protein carries a signal peptide as predicted by SignalP and is thus most likely localised in the periplasm or outer membrane, while no signal peptide is present in PANA_2373, suggesting a cytoplasmic localisation (Fig. 5.4).

Type VI secretion locus III

T6SS locus III is ~29.9 kb in size and encodes nineteen predicted proteins, PANA_4130-4151. It has G+C content of 53.62% which is typical for the chromosome. This suggests that, as is the case for loci I and II, locus III may have been derived from a near phylogenetic relative or may represent an ancient

227

characteristic, having been derived prior to speciation. BlastP analysis however indicates that this locus is most homologous to those found in more distantly related *Enterobacteriaceae*. Loci with similar gene content and gene order were found in eight other *Enterobacteriaceae*. A neighbour-joining tree based on a ClustalW alignment of the concatenated amino acid sequences of nine conserved proteins shows that the *P. ananatis* locus III is most homologous to those in the human pathogens *Enterobacter cloaceae* ATCC13047, *K. pneumoniae* MGH78578 and the pine endophyte *S. proteamaculans* 568 (Fig. 5.6). Locus III CDSs showed 80.7% and 80.3% and 78.8% average aa identity to 15, 15 and 13 proteins encoded in the homologous clusters in *E. cloacae* ATCC13047, *S. proteamaculans* 568 and *K. pneumoniae* MGH78578, respectively (Fig. 5.7). By contrast only 67.5% and 56.0% average aa identity in 14 and 16 CDSs were observed in the more closely related *Pantoea* sp. At9b and *E. amylovora* CFBP1430. The most protein coding sequences are shared between *P. ananatis* locus III and a T6SS locus in *Y. pestis* KIM10, with 19 CDSs in common.

Fourteen of the predicted proteins encoded in locus III show homology to 13 of the conserved proteins identified by Boyer et al. (2009) (Fig. 5.4; Table 5.1), suggesting a structurally similar T6SS apparatus is encoded by this locus as found in other bacteria. Absent from locus III are homologues of the Fha domain (COG3456) and COG3913 proteins. As these are involved in post-translational regulation and no homologues to PpkA and PppA are encoded in the locus, this T6SS does not appear to be regulated in this manner. No transcriptional regulator is encoded in the locus either, raising questions as to how the locus III T6SS is regulated. It can be hypothesised that the prominent PpkA-PppA and Fha domain proteins in the structurally incomplete locus I may be involved in regulating the locus III T6SS. However their extensive divergence suggests this is not the case. Also absent is a homologue of the COG4455 protein of unknown function. Given only those proteins involved in post-translational regulation are absent from locus III, the COG4455 protein may also be involved in this function. PANA_4130 and PANA_4137 both encode copies of the membrane-associated ImpA-related protein (COG3515) which has been suggested to play a role in protein export (Mintz and Fives-Taylor, 2000). The significance of multiple copies of this protein in Type VI secretion is not known.

Evolutionary origins of the T6SS loci in *P. ananatis* LMG20103

The evolutionary origins of the *P. ananatis* LMG20103 T6SS were analysed by incorporating the conserved amino acid sequences for the locus II and III IglA and IglB chaperones into the phylogeny constructed by Bingle et al. (2008) (Fig. 5.8A). The locus II and III sequences are situated in separate clusters within the tree, indicating distinct evolutionary origins for these two loci, likely having arisen through two separate horizontal acquisition events. This is supported by BlastP analyses which indicate extensive rearrangement and low homology, with 31% average aa identity in twelve proteins, exist between locus II and III, with the homology likely a function of the structural components required for operation of the two secretion apparatuses. As no IglA and IglB homologues are present in locus I, a tree was constructed based on the amino acid sequences of IcmF (Fig. 5.8B), the only protein conserved in locus I, II and III and the T6SS loci included by Bingle et al. (2008). Although the branches and clusters in the IcmF tree differ somewhat to those in the IglAB tree, locus II and III are again situated in distant clusters (Fig. 5.8B). On the other hand, the locus I IcmF protein groups together with the locus II protein which suggests that these loci have been derived from a common ancestor and one may have arisen through gene duplication of the other. This is supported by the fact that all eight proteins encoded by locus I share homology to proteins encoded in locus II, with 51% average aa identity between them. Homologues to locus II and III are found in many other bacteria including members of the family *Enterobacteriaceae*, while locus I homologues are restricted to *Pantoea* and *Erwinia* species. This suggests locus II was acquired in an ancient horizontal transfer event, while locus I was derived from locus II just prior to divergence of *Pantoea* and *Erwinia*. Locus III, being more restricted in distribution, is likely to have been horizontally acquired subsequent to divergence.

Putative functions and effectors of the *P. ananatis* LMG20103 T6SSs

Type VI secretion locus I

No homologues to Hcp and VrgG proteins, two known effectors of the T6SS, are encoded in locus I. As genes encoding secretion substrates are frequently located adjacent to the genes for assembly of the secretion apparatus, the flanking regions of locus I were analysed for putative effectors. Adjacent to the 5' end of T6SS locus I three genes, PANA_1647-1649, encode ChaABC homologues with a role in cation

transport. A similar transporter is found adjacent to the T6SS loci in the other *Pantoea* and *Erwinia* spp. Downstream of PANA_1657 and co-located with T6SS locus I on horizontally acquired island HAI12 are seven genes encoding enzymes involved in assimilation and metabolism of nitrate and nitrite (Fig. 5.8). Similar genes are found adjacent to the *P. stewartii* DC283 T6SS locus, but homologues are absent from the genomes of *Pantoea* At9b and *Erwinia* species. Approximately 12.4 kb downstream of PANA_1657 an operon consisting of eight genes, PANA_1665-1672, encodes proteins involved in metabolism of the plant hemicellulose component xylitol, a rare capacity identified in *P. ananatis* ATCC19321 (Sakakibara and Saha, 2008). PANA_1665-1672 show 99% average aa identity to the *P. ananatis* ATCC19321 proteins (BAG32259-32266). Within this operon PANA_1667 encodes a protein sharing 48% aa identity with the periplasmic binding protein RbsB in *R. leguminosarum* (AAL14902). In the latter organism, RbsB was demonstrated to be secreted by the T6SS (Bladergroen et al. 2003). RbsB homologues are involved in nitrogen fixation, uptake of ribose and chemotaxis. A role in chemotaxis has also been identified in *Actinobacillus actinomycetemcomitans*, where RbsB can bind AI-2 autoinducer molecules (Pukatzki et al. 2009). PANA_1669 and PANA_1670 show 51% average aa identity to *R. leguminosarum* RbsA (Rleg_4757) and RbsC (Rleg_4756). Similarly, RbsABC are encoded adjacent to a T6SS locus in *Vibrio cholerae* (Bladergroen et al. 2003). The *P. ananatis* LMG20103 RbsB homologue appears to serve a purpose as periplasmic binding protein for xylitol transport and metabolism, but it could also act as ligand for binding of AI-2 signalling molecules and quorum sensing. Secretion of RbsB via the locus I T6SS would need to be experimentally elucidated.

Type VI secretion locus II

Homologues to VgrG (PANA_2352) and Hcp (PANA_2364) are present in locus II and may be secreted as effectors as has been demonstrated in analogous T6SSs in other bacteria. The PANA_2352 protein shows homology over the entire length of the protein sequence only to the VgrG protein of *Cronobacter turicensis* z3032 (Ctu_00910 – 66% aa identity). CDSearch analysis of the conserved domains in the PANA_2352 amino acid sequence indicates this protein represents an evolved VgrG (Fig. 5.7). The 856 aa sequence contains a VgrG domain (amino acid 6-524) at the N-terminal end. A stretch of 95 aa (amino acid 370-464) contains a conserved domain

230
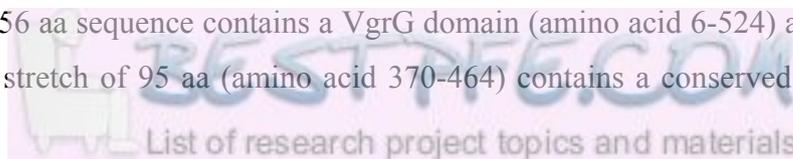
(Cl11432) found in the T4 bacteriophage V protein which forms a spike at the end of the bacteriophage tail (Haggård-Liungquist et al. 1995). Another domain (PHA2596 - amino acid 504-619) is conserved in the T4 bacteriophage gp5 protein which serves as a lysozyme locally dissolving the host cell wall (Arisaka et al. 2003). It can thus be envisioned that these three domain act as host cell wall puncturing device for the Type VI secretion apparatus. A VgrG C-terminal extension stretches from amino acid 620-856. BlastP analysis of this region shows homologues exist in the *C. turiscens* VgrG (Ctu_00910) and is further restricted to plant pathogens belonging to the genus *Xanthomonas* and *Ralstonia*. No conserved domains could be identified within the C-terminal extension. Given the restricted distribution largely among phytopathogenic bacteria, this VgrG protein may play a role in plant disease.

Located between *rhs* (PANA_2348) and *vgrG* (PANA_2352) are three genes, PANA_2349-2351, which are not shared in common with the *Pantoea* sp. At9b and *E. amylovora* CFBP1430 T6SS loci (Fig. 5.5). PANA_2349 encodes a protein 146 aa in size which shows 59% aa identity to the T6SS-associated Ctu_00930 in *C. turicensis* z3032. It also shows 27% identity to ORF5 on a genomic island in the insect pathogen *Photorhabdus luminescens* W14 (AAN64197) which carries a macrophage toxin-like protein and a VgrG protein (Waterfield et al. 2002). The PANA_2350 protein shows 38% aa identity to a T6SS locus-associated hypothetical protein in *Erwinia pyrifoliae* Ep1/96 (EpC_06470). The third protein encoded by PANA_2351 shows limited homology to a hypothetical protein in *C. turicensis* z3032 (Ctu_00920 - 25% aa identity) and homologues are further restricted to plant-associated bacteria. Psort analysis indicates that this protein is localised extracellularly. Its location adjacent to the VgrG-encoding PANA_2352 suggests it may serve as secreted effector, but its function will need to be elucidated. The PANA_2351 gene has a G+C content of 42.25%, 11,44% below the genome average, and may thus have been recently horizontally acquired. In a similar location in the homologous *Pantoea* sp. At9b T6SS locus, a single gene encodes a hypothetical protein (Pat9bDRAFT_2274), while three genes in *E. amylovora* CFBP encode a hypothetical protein (EAMY3001), a flagellar L-ring-like protein (EAMY3002) and a second VgrG protein (EAMY3003). This indicates significant diversification in this region of the T6SS loci of these organisms, with recent horizontal acquisition of putative effector proteins upstream of *vgrG*.

The Hcp-encoding PANA_2364 gene resides on a putative horizontally acquired island HAI16 (Chapter 3; Fig. 5.8). This 3.1 kb island has a G+C content of 40.36%, 13.33% below the genome average, and encompasses 3 genes, PANA_2362-2364. Within this region in *Pantoea* sp. At9b there are three genes Pat9bDRAFT_2288-2290 which encode hypothetical proteins, while two hypothetical proteins are encoded by EAMY_3017-3018 in *E. amylovora* CFBP1430, indicating that is also a region of diversification within T6SS locus II and its homologues. PANA_2362 encodes a 133 aa protein with highest homology to a putative exported protein in *P. atrosepticum* SCRI1043 (ECA2175). Homologues are restricted to plant-associated bacteria belonging to the genera *Pseudomonas*, *Erwinia*, *Pectobacterium* and members of the family *Rhizobiaceae*, indicating it may represent a Type VI-secreted effector in the plant environment. A similar distribution is observed for the Hcp protein encoded by PANA_2364, which shows highest homology to the *E. tasmaniensis* Et1/99 Hcp (ETA_06200 – 96% aa identity). BlastP analysis of the PANA_2363 protein showed that it shares 40% aa identity with the C-terminal extension of the evolved VgrG3 of *Vibrio cholerae* TMA21 (EEF013564). PANA_2363 carries a peptidoglycan-binding (PG-binding) domain (COG3409) at its N-terminal, with 40% aa identity between amino acids 1-89 of PANA_2363 and a region of a PG-binding domain protein of *Dickeya zeae* Ech1591 (Dd1591_4051), while the C-terminal portion of PANA_2363 (amino acid 97- 291) shows 45% aa identity to the EF hand domain protein of *Pseudomonas syringae* pv. *tomato* DC3000 (PSPTO_5206). The PG-binding domain has been suggested to play a role in anchoring the T6SS to the bacterial envelope or it may bind cell membranes of other bacteria and have an antibiotic role (Pukatzki et al. 2007). The homology between the PANA_2363 protein and the C-terminal extension of *V. cholerae* TMA21 VgrG and the conserved peptidoglycan-binding domain suggests the PANA_2363 may represent a further effector secreted by the locus II secretion apparatus. The restricted homology of the EF hand and PG-binding domains in plant-associated bacteria suggest this protein plays a role in phytopathogenesis.

Within the locus II T6SS locus there are thus five putative effector proteins. Experimental validation of their functions is however necessary. The three putative effector proteins encoded by PANA_2362-2364 occur on a horizontally acquired

232

island (HAI16), while PANA_2351 also shows evidence of recent horizontal acquisition. Homologues to these putative effector-encoding genes are restricted in distribution to plant-associated bacteria, including phytopathogens belonging to the genera *Erwinia*, *Pectobacterium* and *Pseudomonas*. This suggests that the more ancient and integrated *P. ananatis* LMG20103 locus II T6SS may have recently evolved through the uptake effectors and become specialised to function in the plant environment, with a potential role in plant disease.

Type VI secretion locus III

Genes encoding Hcp and VgrG homologues are present in locus III. The Hcp protein encoded by PANA_4146 shows significant disparity to the other *P. ananatis* LMG20103 Hcp homologues in its amino acid sequence, length and G+C content of the encoding gene (Table 5.2). This agrees with the distinct evolutionary origin of the locus III T6SS (Fig. 5.2). PANA_4144 encodes a 783 aa protein with most extensive homology to the VgrGs of *K. pneumoniae* MGH78578 (KPN_01329 - 83% aa identity) and *Enterobacter cloacae* ATCC13047 (ECL_01807 – 81% aa identity) but shares only 26% aa identity with the PANA_2352 VgrG encoded in locus II. As is the case for the PANA_2352 VgrG, a C-terminal extension exists in the PANA_4144 protein which thus represents an evolved VgrG. Unlike PANA_2352 however, this extension is common to the VgrG proteins of a large number of plant and animal pathogens including *Klebsiella* spp., *Yersinia* spp. and *Burkholderia* spp. Two overlapping conserved domains are present, DUF2345 and COG4253 (Fig. 5.7). The domain of unknown function DUF2345 has also been identified in the C-terminal extension of the *Pseudomonas entomophila* VgrG (YP_606302) as part of a fibronectin-like adhesion domain (Pukatzki et al. 2007) and COG4253 has been suggested to play a role in attachment to host cells (Boyer et al. 2009). Thus the PANA_4144 VgrG C-terminal extension may be involved in attachment of VgrG to host cells. An alternative function for the COG4253 domain has been suggested, where it acts as linker between the core VgrG protein and its C-terminal extension and modulates the VgrG function, between secretion and virulence (Boyer et al. 2009).

BlastP analysis with the locus II proteins revealed two clusters of protein coding sequences which are not conserved among homologous T6SS loci. These clusters correspond to two horizontally acquired islands predicted on the basis of deviations in

G+C content and oligonucleotide usage (Chapter 3; Fig. 5.8). The first island, HAI32 is ~2.2 kb in size with a G+C content of 40.96% (12.73% below the genome average) and carries two genes, PANA_4135 and PANA_4136. Both encode proteins with homology to proteins found only in animal-associated bacteria. PANA_4135 encodes a 144 aa protein which shows 44% aa identity to a hypothetical protein in *Y. pseudotuberculosis* YPIII (YPK_0624). PANA_4136 encodes a 432 aa protein with an N-terminal region which shows 53% aa identity to the S-pyocin domain protein in *Yersinia pseudotuberculosis* (YPK_0952). No conserved domains could be detected in the C-terminal portion of the protein. S-pyocins are a form of bacteriocin produced by many bacteria that play a role in bacterial cell killing (Michel-Briand and Baysse, 2002). An S-pyocin domain has also been described in the C-terminal extension of a evolved VgrG protein of *Salmonella enterica* (Blondel et al. 2009).

HAI33 is 5.4 kb in size, with a G+C content of 42.32% (11.37% below genome average) and carries five genes, PANA_4139-4143 (Fig. 5.8). As is the case for HAI16 in locus II and HAI32 in locus III several encode putative secreted effector proteins. Four of the five encoded proteins share homology to proteins restricted to animal-associated bacteria. The exception is the 162 aa protein encoded by PANA_4140 which is unique to *P. ananatis* LMG20103. This protein carries a signal peptide and TMMH predicts an internal helix, which suggests this protein is localised in the inner membrane/ periplasm and rather than being an effector, may represent a structural component of the locus III T6SS. Likewise, the PANA_4139 protein has two internal helices and is predicted to be localised in the cytoplasm/inner membrane and is also likely a structural protein. PANA_4142 encodes a 234 aa protein which shows highest homology to a hypothetical protein in *Salmonella enterica* serovar Typhi CT18 (STY3289 – 30% aa identity) with further homologues restricted to animal-pathogenic *Yersinia* and *Salmonella* strains. Similarly, homologues to the 93 aa PANA_4141 protein are restricted to several *Salmonella enterica* strains only and is predicted to encode a secreted protein. As no signal peptide is present in the PANA_4141 protein it may be secreted via the locus III T6SS. PANA_4143 encodes a 743 aa protein which shows highest identity to a zinc-metallopeptidase protein in *Yersinia pestis* KIM10 (y2694 – 53% aa identity). Furthermore, the first 124 N-terminal amino acids share 30% identity to the N-terminal region of the *Vibrio* sp. AND4 VgrG protein (AND4_08927). A zinc-metallopeptidase domain has also been

identified as part of the C-terminal extension of a *Pseudomonas aeruginosa* VgrG protein. It can thus be hypothesised that the PANA_4143 may act as effector protein of the locus III T6SS with a role in pathogenesis in animal hosts.

DISCUSSION

Analysis of the whole genome sequence revealed the presence of three Type VI secretion systems in a strain of the environmentally prolific, broad host range plant pathogen *Pantoea ananatis*. 1.3% of the total proteins encoded on the genome encode structural, functional and secreted components of these T6SS implying that they perform secretory functions important to the typical lifestyle of this organism. The T6SS are encoded in three distinct loci, locus I-III, on the chromosome. Evolutionary analysis revealed that the *P. ananatis* LMG20103 T6SS loci have arisen through at least two separate horizontal acquisition events, with extensive differences in both gene order and sequence conservation between the locus II and III T6SS. On the other hand, all eight proteins encoded by locus I share some homology to those in locus II and this locus may have arisen through partial duplication of the locus II prior to divergence of the *Erwinia* and *Pantoea* species. Evidence of horizontal gene transfer in T6SS locus I however suggests a more complex evolutionary history for this locus. Homologous T6SS loci are well-represented in the *Erwinia* and *Pantoea* species for which whole genomes are available. However, in the nearest phylogenetic relative of *P. ananatis* LMG20103, namely *Pantoea stewartii* subsp. *stewartii* DC283, only a homologue to locus I can be found, while no locus II or III homologues are present. The aggressive, maize-specific pathogen *P. stewartii* DC283 has four Type III secretion systems, which may preclude its need for these T6SSs. This may be correlated to differences in their lifestyles, with the host-specific *P. stewartii* subsp. *stewartii* using its T3SS to cause disease and survive on its maize host, while the T6SSs of *P. ananatis* may allow it to survive in various environmental niches and cause disease in a broad range of hosts.

Analysis of known and predicted structural proteins showed that all major components for functional locus II and III T6SSs are present. By contrast, homologues to the major structural proteins ClpV, Hcp and VgrG are not encoded by T6SS locus I, suggesting it is non-functional. However, several T6SS in other organisms are known to function without these components (Bladergroen et al. 2003).

Two Hcp proteins occur separately from the three T6SS loci on the chromosome. These may be involved in locus I T6SS functioning, but as Hcp and VgrG proteins in genomic locations distinct from the T6SS locus are a common feature to many bacteria, their association with a particular *P. ananatis* T6SS locus seems unlikely. Both locus I and II contain CDSs encoding proteins involved in post-translational regulation of T6SS loci. As these share homology only to proteins in plant-associated bacteria, and given that these regulatory proteins respond to environmental signals, it is likely they are involved in regulating the locus I and II T6SS expression in response to signals in the plant. No transcriptional or post-translational regulatory components are present in T6SS locus III, which raises questions as to how this system is regulated. The T6SS locus in the phytopathogen *Pseudomonas syringae* pv. *syringae* B728a was recently shown to lack both post-translational and transcriptional regulators, but is regulated via two two-component master regulators, LadS and RetS, which are not chromosomally linked to the T6SS locus (Records and Gros, 2010). Twenty-two two component systems are encoded on the *P. ananatis* LMG20103 genome and these may likewise play a role in regulation of the locus III T6SS.

Two structural components, VgrG and Hcp have also been demonstrated to serve as secreted effectors with pathogenic functions in both animal and plant pathogens. Hcp proteins assemble end to end to form a channel through which effectors are secreted, but also acts as pore-forming toxin (Pukatzki et al. 2009). VgrG proteins form membrane puncturing devices, allowing the T6SS to extend through the bacterial outer membrane and access the host cell. Some VgrG proteins, known as evolved VgrGs, have C-terminal extensions with a range of different pathogenic functions (Pukatzki et al. 2009). As no VgrG and Hcp proteins are present, and non-pathogenic functions have been assigned to the T6SSs in several commensal and symbiotic bacteria, it is likely that the locus I T6SS does not play a role in disease expression by LMG20103. The T6SS locus I proteins and the adjacent nitrate/nitrite assimilation operon occur on a horizontally acquired island only found in plant-associated bacteria and locus I homologues are restricted in distribution to *Pantoea* and *Erwinia* species which indicates a locus I T6SS function in the plant environment. A homologue to the *R. leguminosarum* Type VI secreted RbsB protein, with a putative function in quorum sensing, is located close to locus I. This protein may also be secreted by *P. ananatis* locus I, but its secretion, along with the actual function of the locus I system will need

to be experimentally elucidated. As a homologous T6SS locus is found in the pear epiphyte *Erwinia tasmaniensis* and in *E. amylovora* which also has an epiphytic phase in its infection cycle, it is feasible that the LMG20103 locus I T6SS plays a role in the epi- or endophytic survival of this broad host range pathogen on its host plants.

T6SS locus II encodes both VgrG and Hcp proteins. The VgrG protein has C-terminal extension, but the evolved region does not contain a conserved domain and its function can thus not be predicted on the basis of its sequence alone. The locus II Hcp protein is encoded on a predicted horizontally acquired island, along with two genes which encoded putative secreted effectors. Homologues to all three encoded proteins are restricted to plant-associated bacteria. Adjacent to the *vgrG* gene, a further gene, PANA_2351 also shows evidence of recent horizontal acquisition and encodes a putative effector protein with homology restricted to plant-associated bacteria. Alignment of the locus II sequences with those in the highly homologous loci showed significant conservation in both gene order and amino acid identity of the encoded structural proteins. There is however no overlap in the two regions of recent horizontal acquisition, with non-homologous proteins encoded in the overlapping regions in the *Erwinia amylovora* CFBP1430 and *Pantoea* sp. At9b loci. This suggests that while the well-conserved Type VI secretion apparatus is a common trait among *P. ananatis* and its close phylogenetic relatives, their T6SSs have become specialised, through the recent horizontal acquisition of effector proteins to perform specific functions to suit their different lifestyles. As the four recently acquired genes encode putative effector proteins restricted to plant-associated bacteria it is likely that the *P. ananatis* locus II T6SS has become specialised to secrete effectors that play a role in the plant environment and are likely involved in phytopathogenesis.

P. ananatis LMG20103 T6SS locus III also encodes a VgrG and Hcp protein. The PANA_4144-encoded evolved VgrG carries a C-terminal extension with a conserved domain COG4253 for which two putative functions have been assigned, attachment to host cells and linking of C-terminal extensions to the core VgrG protein. Alignment of *P. ananatis* locus III to homologous loci showed that, as is the case for locus II, there are two regions of divergence. These overlap with two potential horizontally acquired islands as determined by their unusual oligonucleotide usage and deviation in G+C content. Homologues to the proteins encoded on both HAIs are restricted to animal-

associated bacteria, mainly in the genera *Yersinia* and *Salmonella*. The first HAI carries a gene that encodes a protein with an S-pyocin domain found in cell killing bacteriocins. This type of domain is also found in an evolved VgrG protein in *Salmonella enterica* (Blondel et al. 2009). It can thus be envisaged that PANA_4136 encodes an effector protein which is secreted via the locus III T6SS that may kill or inhibit bacteria competing for the same nutrient sources or habitat. The second HAI also carries two genes, PANA_4141 and PANA_4143 which encode putative effector proteins. The latter shows homology to a zinc-metallopeptidase. A zinc-metallopeptidase domain is also found in the C-terminal extension of the evolved VgrG in *Pseudomonas aeruginosa* (Pukatzki et al. 2007). As homologues to this protein are restricted to animal pathogens including *Yersinia pestis*, *Salmonella enterica* and *Klebsiella pneumoniae* it is likely to represent a pathogenicity factor secreted via the locus III T6SS. As both a S-pyocin domain and metallopeptidase domain are found as C-terminal extensions in VgrG proteins, and the *P. ananatis* locus III VgrG protein has a linker domain, it is plausible that it can alternately associate with the different C-terminal extension-like proteins encoded by PANA_4143 and 4136. In this manner, *P. ananatis* may modulate the type of evolved VgrG protein secreted via the locus III T6SS when they are required for different functions. It can be envisaged that the core VgrG could attach to and breach the eukaryotic cell membranes, after which the PANA_4143 protein becomes attached to the VgrG and exercises its pathogenic function in the mammalian host. Alternatively, after breaching the bacterial cell wall, the S-pyocin domain protein may become attached to the VgrG protein and this evolved VgrG can kill bacteria competing for limited nutrient sources. This hypothesis will need to be tested experimentally.

In conclusion, *Pantoea ananatis* LMG20103 produces three different Type VI secretion systems which perform diverse functions. One is likely involved in the epiphytic or endophytic survival of this pathogen on plants, while the other two may have evolved through the acquisition of secretion effectors to play a role in disease on plant and animal hosts respectively. As *P. ananatis* has been found in almost every environment, surviving on a wide range of plants, infecting both a wide range of host plants as well as humans, its Type VI secretion systems are likely to play significant roles in its versatile lifestyle.

REFERENCES

Arisaka, F., Kanamaru, S., Leiman, P. and Rossmann, M. G. 2003. The tail lysozyme complex of bacteriophage T4. Int. J. Biochem. Cell. Biol. **35**: 16-21

Aschtgen, M.-S., Bernard, C. S., De Bentzmann, S., Lloubès, R. and Cascales, E. 2008. SciN is an outer membrane lipoprotein required for Type VI secretion in enteroaggregative *Escherichia coli*. J. Bacteriol. **190**: 7523-7531

Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. J.Mol. Biol. **340**: 783-795

Bingle, L. E. H., Bailey, C. M. and Pallen, M. J. 2008. Type VI secretion: a beginner's guide. Curr. Opin. Microbiol. **11**: 3-8

Bladergroen, M. R., Badelt, K. and Spaink, H. P. 2003. Infection-blocking genes of a symbiotic *Rhizobium leguminosarum* strain that are involved in temperature-dependent protein secretion. Mol. Plant-Microbe Interact. **16**: 53-64

Blondel, C. J., Jiménez, J. C., Contreras, I. and Santiviago, C. A. 2009. Comparative genomic analysis uncovers 3 novel loci encoding Type Six secretion systems differentially distributed in *Salmonella* serotypes. BMC Genomics **10**: 1-17

Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. and Attree, I. 2009. Dissecting the bacterial type VI secretion system by a genome wide *in silico* analysis: what can be learned from available microbial genomic resources? BMC Genomics **10**: 104-117

Coutinho, T. A. and Venter, S. N. 2009. *Pantoea ananatis*: an unconventional plant pathogen. Mol. Plant Pathol. **10**: 325-335

Das, S. and Chaudhuri, K. 2003. Identification of a unique IAHP (IcmF associated homologous proteins) cluster *Vibrio cholerae* and other Proteobacteria through *in silico* analysis. In Silico Biol. **3**: 287-300

De Baere, T., Verhelst, R., Labit, C., Verschraegen, G., Wauters, G. et al. 2004. Bacteremic infection with *Pantoea ananatis*. J. Clin. Microbiol. **42**: 4393-4395

De Maayer, P., Chan, W. Y., Venter, S. N., Toth, I. K., Birch, P. R. J. et al. 2010. Genome sequence of Pantoea ananatis LMG20103, the causative agent of Eucalyptus blight and dieback. J. Bacteriol. **192**(11): 2936-2937

Filloux, A. 2009. The Type VI secretion system: a tubular story. EMBO J. **28**: 309-310

Filloux, A., Hachani, A. and Bleves, S. 2008. The bacterial Type VI secretion machine: yet another player for protein transport across membranes. Microbiology **154**: 1570-1583

Gitaitis, R. D., Walcott, R. R., Wells, M. L., Diaz Perez, J. C. and Sanders, F. H. 2003. Transmission of *Pantoea ananatis*, causal agent of center rot of onion by tobacco thrips, *Frankliniella fusca*. Plant Dis. **87**: 675-678

Haggård-Liungquist, E., Jacobsen, E., Rishovd, S., Six, E. W., Nilssen, Ø. et al. 1995. Bacteriophage P2: genes involved in baseplate assembly. Virology **213**: 109-121

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. et al. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. **12**: 1652-1662

Krogh, A., Larsson, B., Heijne, V., G and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305**: 567-580

Leiman, P. G., Basler, M., Ramagopal, U. A., Bonanno, J. B., Sauder, J. M. et al. 2009. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. Proc. Natl. Acad. Sci. USA **106**: 4154-4159

Lesic, B., Starkey, M., He, J., Hazan, R. and Rahme, L. G. 2009. Quorum sensing differentially regulates *Pseudomonas aeruginosa* Type VI secretion locus I and homologous loci II and III, which are required for pathogenesis. Microbiology **155**: 2845-2855

Liu, H., Coulthurst, S. J., Pritchard, L., Hedley, P. E., Ravensdale, M. et al. 2008. Quorum sensing coordinates brute force and stealth modes of infection in the plant pathogen *Pectobacterium atrosepticum*. PLoS Pathog. **4**: 1-11

Marchler-Bauer, A. and Bryant, S. H. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. **32**: 327-331

Mattinen, L., Nissinen, R., Riipi, T., Kalkkinen, N. and Pirhonen, M. 2007. Host-extract induced changes in the secretome of the plant pathogenic bacterium Pectobacterium atrosepticum. Proteomics **7**: 3527-3537

Michel-Briand, Y. and Baysse, C. 2002. The pyocins of *Pseudomonas aeruginosa*. Biochimie **84**: 499-510

Mintz, K. P. and Fives-Taylor, P. M. 2000. *impA*, a gene coding for an inner membrane protein, influences colonial morphology of *Actinobacillus actinomycetemcomitans*. Infect. Immun. **68**: 6580-6586

Mougous, J. D., Cuff, M. E., Raunser, S., Shen, A., Zhou, M. et al. 2006. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. Science **312**: 1526-1530

Mougous, J. D., Gifford, C. A., Ramsdell, T. L. and Mekalanos, J. J. 2007. Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*. Nature Cell Biol. **9**: 797-803

Parsons, D. A. and Heffron, F. 2005. *sciS*, an *icmF* homolog in *Salmonella enterica* serovar Typhimurium limits intracellular replication and decreases virulence. Infect. Immun. **73**: 4338-4345

Pukatzki, S., Mcauley, S. B. and Miyata, S. T. 2009. The Type VI secretion system: translocation of effectors and effector-domains. Curr. Opin. Microbiol.: 11-17

Pukatzki, S., Revel, A. T., Sturtevant, D. and Mekalanos, J. J. 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proc. Natl. Acad. Sci. USA **104**(39): 15508-15513

Pukatzki, S., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W. C. et al. 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. Proc. Natl. Acad. Sci. USA**103**: 1528-1533

Records, A. R. and Gross, D. C. 2010. Sensor kinases RetS and LadS regulate *Pseudomonas syringae* Type VI secretion and virulence factors. J. Bacteriol. **192**: 3584-3596

Robinson, J. B., Telepnev, M. V., Zudina, I. V., Bouyer, D., Montenieri, J. A. et al. 2009. Evaluation of a *Yersinia pestis* mutants impaired in a thermoregulated Type VI-like secretion system in flea, macrophage and murine models. Microb. Pathog. **47**: 243-251

Sakakibara, Y. and Saha, B. C. 2008. Isolation of an operon involved in xylitol metabolism from a xylitol-utilizing *Pantoea ananatis* mutant. J. Biosci. Bioeng. **106**: 337-344

Schell, M. A., Ulrich, R. L., Ribot, W. J., Brueggemann, E. E., Hines, H. B. et al. 2007. Type VI secretion is a major virulence determinant in *Burkholderia mallei*. Mol. Microbiol. **64**: 1466-1485

Shalom, G., Shaw, J. G. and Thomas, M. S. 2007. *In vivo* expression technology identifies a Type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. Microbiology **153**: 2689-2699

Shrivastava, S. and Mande, S. S. 2008. Identification and functional characterization of gene components of Type VI secretion system in bacterial genomes. PLOS One **3**: 1-11

Suarez, G., Sierra, J. C., Sha, J., Wang, S., Erova, T. E. et al. 2008. Molecular characterization of a funtional Type VI secretion system from a clinical isolate of *Aeromonas hydrophila*. Microb. Pathog. **44**: 344-361

Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software Version 4.0. Mol. Biol. Evol. **24**: 1596-1599

Watanabe, K. and Sato, M. 1999. Gut colonization by an ice nucleation active bacterium, *Erwinia* (*Pantoea*) *ananas* reduces the cold hardiness of mulberry Pyralid larvae. Cryobiol. **38**: 281-289

Waterfield, N. R., Daborn, P. J. and Ffrench-Constant, R. H. 2002. Genomic islands in *Photorhabdus*. Trends Microbiol. **10**: 541-545

Wu, H.-Y., Chung, P.-C., Shih, H.-W., Wen, S.-R. and Lai, E.-M. 2008. Secretome analysis uncovers an Hcp-family protein secreted via a Type VI secretion system in *Agrobacterium tumefaciens*. J. Bacteriol. **190**: 2841-2850

Yahr, T. L. 2006. A critical new pathway for toxin secretion? N. Engl. J. Med. **355**: 1171-1172

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Lo, R., Dao, P. et al. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics **26**(13): 1608-1615

Zheng, J. and Leung, K. Y. 2007. Dissection of a Type VI secretion system in *Edwardsiella tarda*. Mol. Microbiol. **66**: 1192-1206

Table 5.1: The sixteen conserved proteins identified by Boyer et al. (2009) in *Pseudomonas aeruginosa* PAO1*, Vibrio cholerae* O1 biovar El Tor, *Rhizobium leguminosarum* biovar *trifolii* RBL5523, *Salmonella enterica* subsp. *enterica* serovar Typhimurium SR-11, *Edwardsiella tarda* PPD130/91, *Burkholderia pseudomallei* K96243 and *Aeromonas hydrophila* ATCC7966. The conserved orthologous group, gene name and predicted protein localisations are shown. Homologues in *P. ananatis* LMG20103 locus I, II and III are shown.

COG ID	Gene name	Predicted Localisation	*P. aeruginosa*	*V. cholerae*	*R. legumi-nosarum*	*S. enterica*	*E. tarda*	*B. pseud-omallei*	*A. hydro-phila*	LMG20103 Locus I	LMG20103 Locus II	LMG20103 Locus III
COG3913	-	Cytoplasm	PA0076		*impM*	*sciT*					PANA_2368	
COG3523	*icmF*	Inner membrane	PA0077	VC_A0120	*impL*	*sciS*	*evpO*	BPS1511	AHA_1845	PANA_1655	PANA_2369	PANA_4138
COG3455	*dotU*	Inner membrane	PA0078	VC_A0115	*impK*	*sciP*	*evpN*	BPS1510	AHA_1840	PANA_1656	PANA_2370	PANA_4148
COG3522	-	Cytoplasm	PA0079	VC_A0114	*impJ*	*sciO*	*evpM*	BPS1509	AHA_1839		PANA_2371	PANA_4149
COG3521	-	Cytoplasm	PA0080	VC_A0113		*sciN*	*evpL*	BPS1508	AHA_1838		PANA_2372	PANA_4132
COG3456	*fha*	Inner membrane	PA0081	VC_A0112	*impI*			AHA_1837	PANA_1653	PANA_2361		
COG3515	-	Outer membrane	PA0082	VC_A0119 VC_A0121	*impA*	*sciA*	*evpK*	BPS1493	AHA_1844 AHA_1846		PANA_2367	PANA_4130 PANA_4137
COG3516	*iglA*	Cytoplasm	PA0083	VC_A0107	*impB*	*sciH*	*evpA*	BPS1496	AHA_1832		PANA_2366	PANA_4151
COG3517	*iglB*	Cytoplasm	PA0084	VC_A0108	*impC/D*	*sciI*	*evpB*	BPS1497	AHA_1833		PANA_2365	PANA_4150
COG3157	*hcp*	Secreted	PA0085			*sciK/M*	*evpC*	BPS1498	AHA_1826		PANA_2364	PANA_4146
COG4455	-	Cytoplasm	PA0086		*impE*	*sciE*					PANA_2358	
COG3518	-	Cytoplasm	PA0087	VC_A0109	*impF*	*sciD*	*evpE*	BPS1499	AHA_1834		PANA_2357	PANA_4131
COG3519	-	Cytoplasm	PA0088	VC_A0110	*impG*	*sciC*	*evpF*	BPS1500	AHA_1835		PANA_2356	PANA_4134
COG3520	-	Cytoplasm	PA0089	VC_A0111	*impH*	*sciB*	*evpG*	BPS1501	AHA_1836		PANA_2355	PANA_4133
COG0542	*clpV*	Inner membrane	PA0090	VC_A0116		*sciG*	*evpH*	BPS1502	AHA_1841		PANA_2354	PANA_4145
COG3501	*vgrG*	Secreted	PA0091 PA0095	VC_A0123		*vgrS*	*evpI*	BPS1503	AHA_1827 AHA_1848		PANA_2352	PANA_4144

FIGURES AND TABLES

Table 5.2: The predicted amino acid sizes and G+C contents of the genes encoding the four Hcp homologues in *P. ananatis* LMG20103. Amino acid identities between the respective Hcp proteins are shown.

	Size (aa)	G+C (%)	PANA _2364	PANA _2446	PANA _2447	PANA _4146
PANA_2364	160	49.90	100			
PANA_2446	160	47.70	64.3	100		
PANA_2447	160	47.62	70.0	67.5	100	
PANA_4146	175	50.95	18.3	17.7	175	100
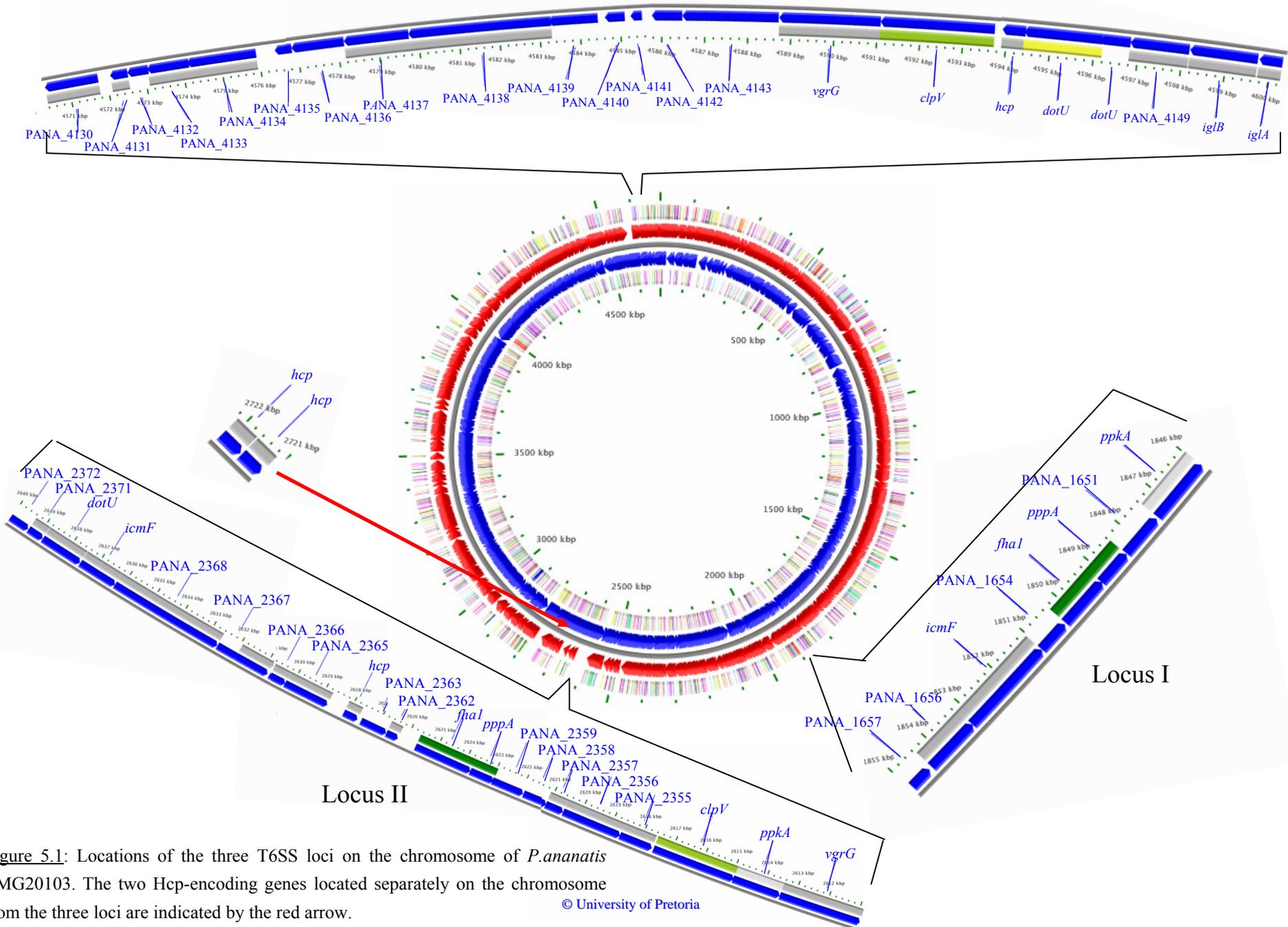
Figure 5.1: Locations of the three T6SS loci on the chromosome of *P.ananatis* LMG20103. The two Hcp-encoding genes located separately on the chromosome from the three loci are indicated by the red arrow.

	Locus I	Locus II	Locus III
Pantoea ananatis LMG20103	+	+	+
Pantoea ananatis LMG2665T	+	+	-
Pantoea stewartii subsp. stewartii DC283	+	-	-
Pantoea sp. At9b	+	+	+
Erwinia amylovora CFBP1430	+	+	+
Erwinia amylovora ATCC49946	+	+	+
Erwinia tasmaniensis Et1/99	+	+	-
Erwinia pyrifoliae DSM12163	+	+	-
Erwinia pyrifoliae Ep1/96	+	+	-

0.01
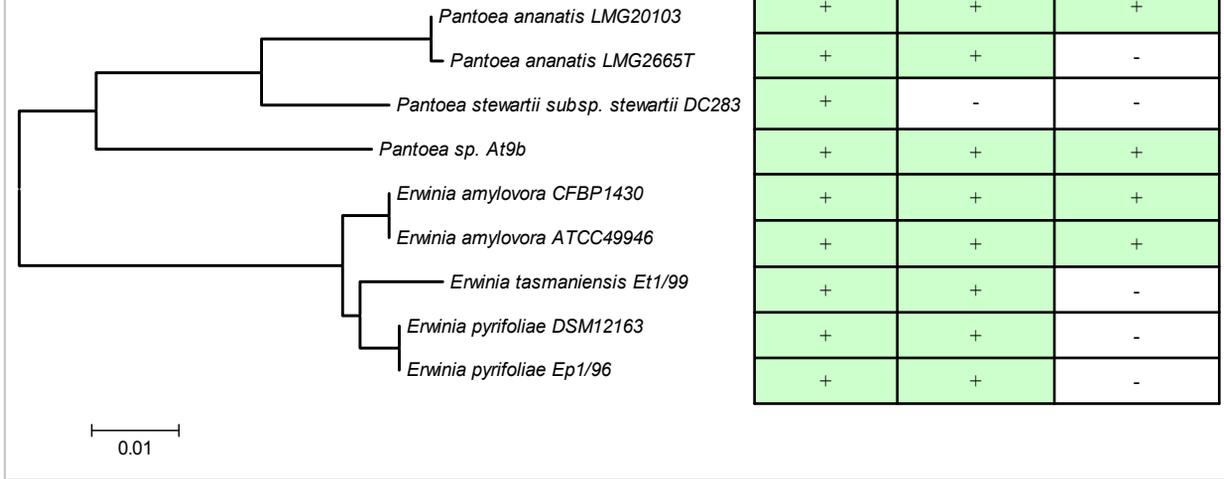
Figure 5.2: The presence/absence in close phylogenetic relatives of T6SS loci homologous to those of *P. ananatis* LMG20103. A neighbour-joining tree of an alignment GyrB amino acid sequences shows the phylogenetic relationships among the included strains.

Protein	COG	Size (aa)	*P. ananatis* LMG20103	*P. stewartii* subsp. *stewartii* DC283	*Pantoea* sp. At9b	*E. amylovora* CFBP1430
-	-	166	PANA_1657	-	Pat9b2015 (72%)	-
DotU	COG3455	405	PANA_1656	ACV283597 (95%)	Pat9b2016 (79%)	EAMY1623 (57%)
IcmF	COG3523	702	PANA_1655	ACV0283599 (93%)	Pat9b2017 (77%)	EAMY1622 (61%)
-	-	234	PANA_1654	ACV0283600 (88%)	Pat9b2018 (71%)	EAMY1621 (53%)
Fha	COG3456	375	PANA_1653	ACV0283601 (87%)	Pat9b2019 (75%)	EAMY1620 (61%)
PppA	-	258	PANA_1652	ACV0283602 (84%)	Pat9b2020 (74%)	-
-	-	315	PANA_1651	ACV0283603 (74%)	Pat9b2021 (63%)	-
PpkA	-	470	PANA_1650	ACV0283604 (88%)	Pat9b2022 (75%)	-
Avg ID				**88%**	**75%**	**59%**

Figure 5.3: Comparison of *P. ananatis* LMG20103 T6SS locus I against homologous loci in *Pantoea stewartii* DC283, *Pantoea* sp. At9b and *Erwinia amylovora* CFBP1430. Homologous CDSs are coloured in green, while yellow denotes those CDSs outside of the T6SS loci shared among bacteria while red CDSs are not shared. Red crosses denote those proteins missing from T6SS clusters. The accompanying table shows the amino acid identities between proteins in the T6SS loci and the average amino acid identity for the loci.
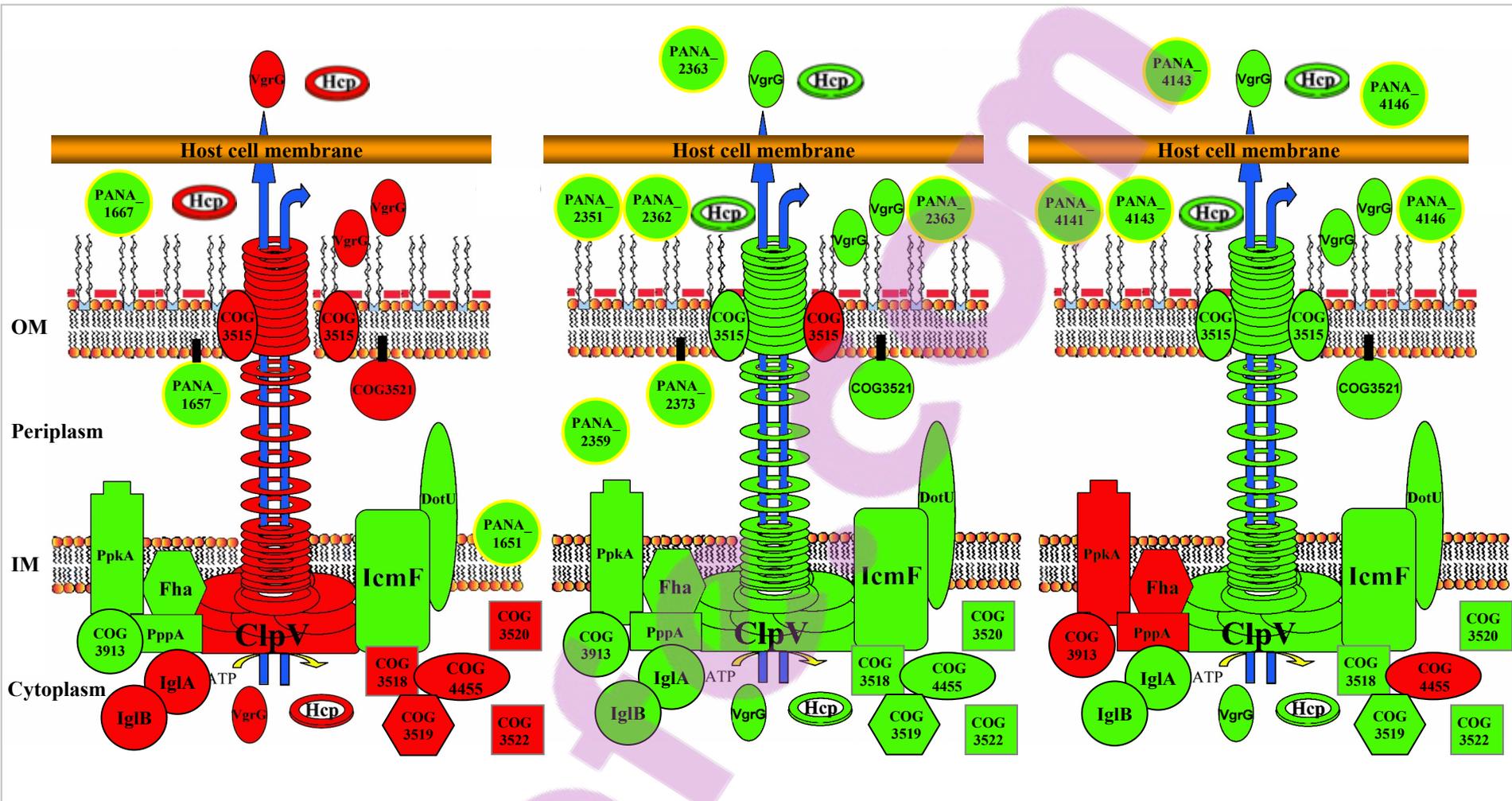
Figure 5.4: Schematic model of the secretion apparatuses encoded by the three T6SS loci of *Pantoea ananatis* LMG20103 based on the models of Lei et al. (2009) and Filloux et al. (2008) and on the predicted subcellular localisations of the additional protein encoded in locus I-III. Those components coloured in red are not encoded in the given locus, while those in green are present. Additional non-conserved proteins are encircled in yellow and their location predicted based on Psort, LipoP and SignalP analysis.
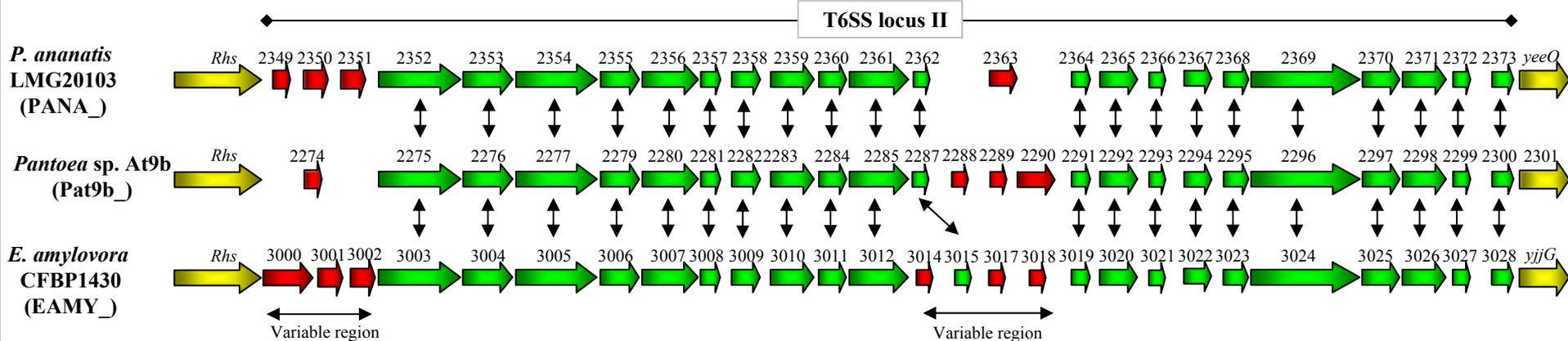
Protein	COG	Size (aa)	*P. ananatis* LMG20103	*Pantoea* At9b	*E. amylovora* CFBP1430
-	-	146	PANA_2349	-	-
-	-	211	PANA_2350	-	-
-	-	197	PANA_2351	-	-
VgrG	COG3501	856	PANA_2352	PAt9b_2275 (74%)	EAMY_3003 (68%)
PpkA	-	503	PANA_2353	PAt9b_2276 (80%)	EAMY_3004 (66%)
ClpV	COG0542	869	PANA_2354	PAt9b_2277 (91%)	EAMY_3005 (83%)
-	COG3520	349	PANA_2355	PAt9b_2279 (82%)	EAMY_3006 (63%)
-	COG3519	625	PANA_2356	PAt9b_2280 (87%)	EAMY_3007 (76%)
-	COG3518	191	PANA_2357	PAt9b_2281 (87%)	EAMY_3008 (72%)
-	COG4455	274	PANA_2358	PAt9b_2282 (72%)	EAMY_3009 (60%)
-	-	351	PANA_2359	PAt9b_2283 (71%)	EAMY_3010 (51%)
PppA	-	263	PANA_2360	PAt9b_2284 (90%)	EAMY_3011 (78%)
Fha	COG3456	632	PANA_2361	PAt9b_2285 (74%)	EAMY_3012 (62%)
-	-	133	PANA_2362	PAt9b_2287 (40%)	EAMY_3015 (39%)
-	-	291	PANA_2363	-	-
Hcp	COG3157	160	PANA_2364	PAt9b_2291 (79%)	EAMY_3019 (95%)
-	COG3517	499	PANA_2365	PAt9b_2292 (94%)	EAMY_3020 (91%)
-	COG3516	180	PANA_2366	PAt9b_2293 (89%)	EAMY_3021 (80%)
-	COG3515	365	PANA_2367	PAt9b_2294 (70%)	EAMY_3022 (67%)
-	COG3913	237	PANA_2368	PAt9b_2295 (81%)	EAMY_3023 (58%)
IcmF	COG3523	1,208	PANA_2369	PAt9b_2296 (87%)	EAMY_3024 (75%)
DotU	COG3455	413	PANA_2370	PAt9b_2297 (90%)	EAMY_3025 (74%)
-	COG3522	448	PANA_2371	PAt9b_2298 (89%)	EAMY_3026 (70%)
-	COG3521	167	PANA_2372	PAt9b_2299 (75%)	EAMY_3027 (51%)
-	-	212	PANA_2373	PAt9b_2300 (71%)	EAMY_3028 (55%)
Avg ID				**82%**	**71%**

Figure 5.5: Comparison of *P. ananatis* LMG20103 T6SS locus II against homologous loci in *Pantoea* sp. At9b and *Erwinia amylovora* CFBP1430. Homologous CDSs are coloured in green, while yellow denotes the CDSs adjacent to the T6SS loci Red arrows indicate CDSs unique to the individual loci. The accompanying table shows the amino acid identities between proteins in the T6SS loci and the average amino acid identity for the loci.
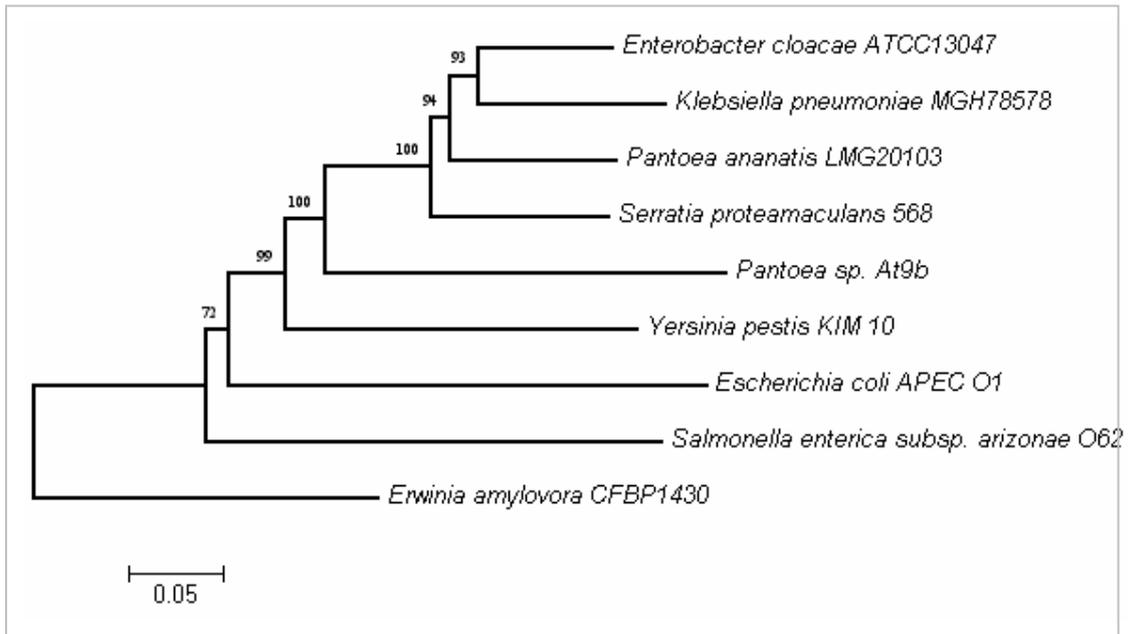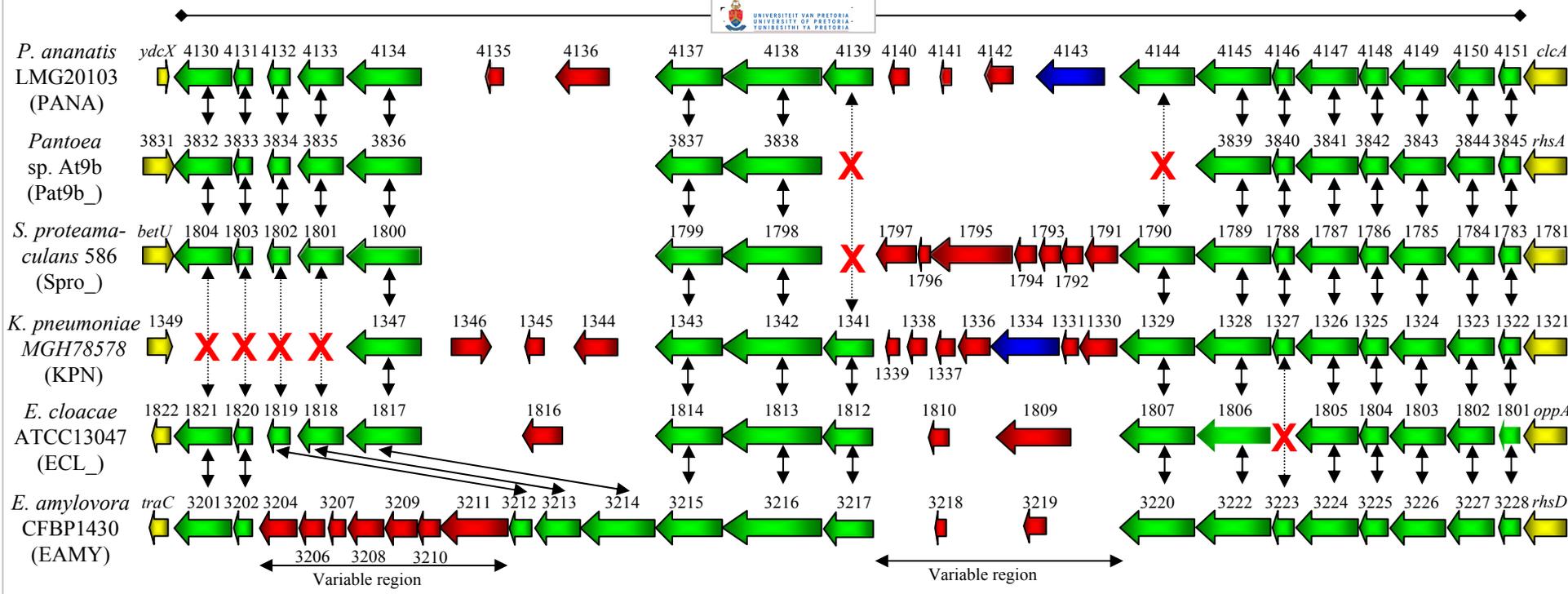
Figure 5.6: Neighbour-joining tree of the concatenated amino acid sequences of nine conserved T6SS proteins conserved in *P. ananatis* LMG20103 locus III and homologous loci found in several closely related *Enterobacteriaceae*. The tree was drawn based on a ClustalW alignment using MEGA 4.1 (Tamura et al. 2007) as described in section 2.3. Bootstrap values (n = 1,000) are shown.

Figure 5.7: Comparison of *P. ananatis* LMG20103 T6SS locus III against homologous loci in *Pantoea* sp. At9b, *Erwinia amylovora* CFBP1430, *Klebsiella pneumoniae* MGH78578 *Serratia proteamaculans* 586, and E*nterobacter cloacae* ATCC13047. Homologous CDSs are coloured in green, while yellow denotes those CDSs adjacent to the T6SS loci. Red arrows denotes CDSs unique to the individual loci, while red crosses indicate absence of proteins in a given region. The accompanying table shows the amino acid identities between proteins in the T6SS loci and the average amino acid identity for the loci.

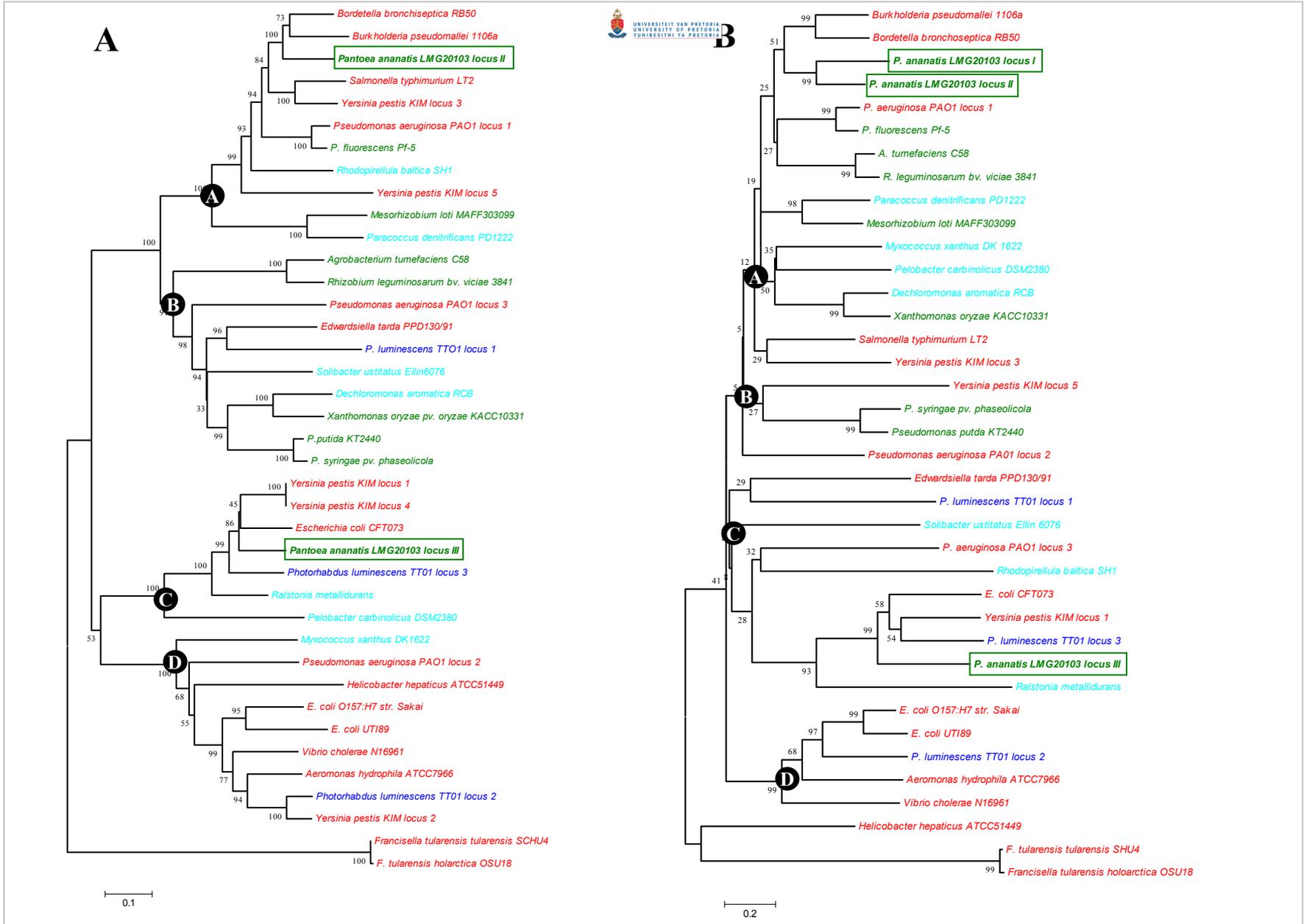| Protein	COG	Size (aa)	*P. ananatis* LMG20103	*E. amylovora* CFBP1430	*Pantoea* sp. At9b	*S. proteamac-ulans* 586	*E. cloacae* ATCC13047	*K. pneumoniae* MGH78578
-	COG3515	463	PANA_4130	EAMY_3201 (65%)	Pat9b3832 (49%)	Spro_1804 (77%)	ECL_01821 (56%)	
-	COG3518	151	PANA_4131	EAMY_3202 (78%)	Pat9b3833 (72%)	Spro_1803 (85%)	ECL_01820 (87%)	
-	COG3521	181	PANA_4132	EAMY_3212 (40%)	Pat9b3834 (52%)	Spro_1802 (66%)	ECL_01819 (62%)	
-	COG3520	361	PANA_4133	EAMY_3213 (54%)	Pat9b3835 (77%)	Spro_1801 (87%)	ECL_01818 (89%)	KPN_1347 (86%)
-	COG3519	599	PANA_4134	EAMY_3214 (62%)	Pat9b3836 (66%)	Spro_1800 (91%)	ECL_01817 (91%)	-
-	-	144	PANA_4135	-	-	-	-	-
-	-	432	PANA_4136	-	-	-	-	-
-	COG3515	533	PANA_4137	EAMY_3215 (50%)	Pat9b3837 (54%)	Spro_1799 (59%)	ECL_01814 (84%)	KPN_1343 (79%)
IcmF	COG3523	1,177	PANA_4138	EAMY_3216 (45%)	Pat9b3838 (58%)	Spro_1798 (78%)	ECL_01813 (84%)	KPN_1342 (76%)
-	-	402	PANA_4139	EAMY_3217 (22%)	-	-	ECL_01812 (67%)	KPN_1341 (67%)
-	-	162	PANA_4140	-	-	-	-	-
-	-	93	PANA_4141	-	-	-	-	-
-	-	234	PANA_4142	-	-	-	-	-
-	-	743	PANA_4143	-	-	-	-	KPN_1334 (77%)
VgrG	COG3501	783	PANA_4144	EAMY_3220 (47%)	-	Spro_1790 (71%)	ECL_01807 (81%)	KPN_1329 (82%)
ClpV	COG0542	884	PANA_4145	EAMY_3222 (74%)	Pat9b3839 (71%)	Spro_1789 (83%)	ECL_01806 (82%)	KPN_1328 (81%)
Hcp	COG3157	175	PANA_4146	EAMY_3223 (81%)	Pat9b3840 (75%)	Spro_1788 (95%)	-	KPN_1327 (95%)
-	-	595	PANA_4147	EAMY_3224 (37%)	Pat9b3841 (44%)	Spro_1787 (74%)	ECL_01805 (70%)	KPN_1326 (64%)
DotU	COG3455	229	PANA_4148	EAMY_3225 (39%)	Pat9b3842 (53%)	Spro_1786 (85%)	ECL_01804 (86%)	KPN_1325 (79%)
-	COG3522	446	PANA_4149	EAMY_3226 (49%)	Pat9b3843 (74%)	Spro_1785 (87%)	ECL_01803 (91%)	KPN_1324 (83%)
IglB	COG3517	515	PANA_4150	EAMY_3227 (90%)	Pat9b3844 (84%)	Spro_1784 (94%)	ECL_01802 (84%)	KPN_1323 (82%)
IglA	COG3516	175	PANA_4151	EAMY_3228 (73%)	Pat9b3845 (64%)	Spro_1783 (79%)	ECL_01801 (74%)	KPN_1322 (71%)
Avg ID				**56.0%**	**67.5%**	**80.3%**	**80.7%**	**78.8%**

Figure 5.8: Phylogenies based on ClustalW alignment with the concatenated IglA and IglB (A) and the IcmF amino acid sequences (B) as per Bingle et al. (2008). A neighbour-joining tree was constructed based on the alignment excluding gaps and with Poisson correction. Bootstrap analysis (n=1,000) was performed. Organisms are coloured according to their ecology, animal-associated bacteria in red, plant-associated bacteria in green, environmental bacteria in cyan and insect-associated bacteria in blue.
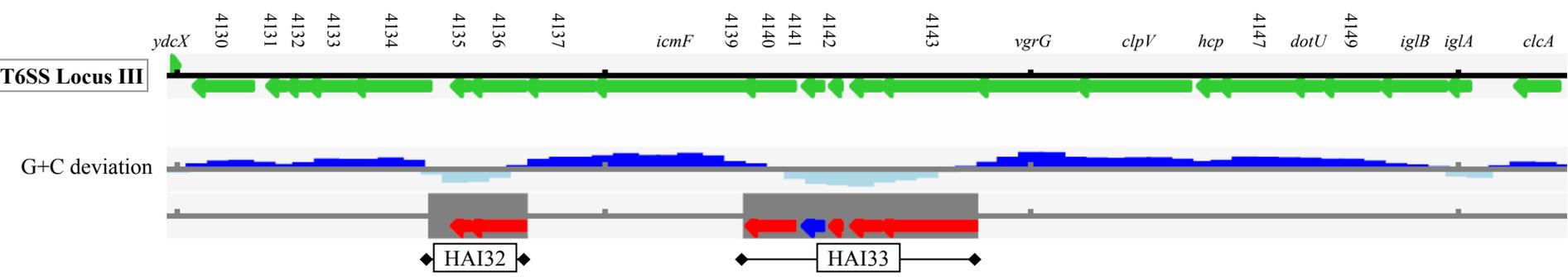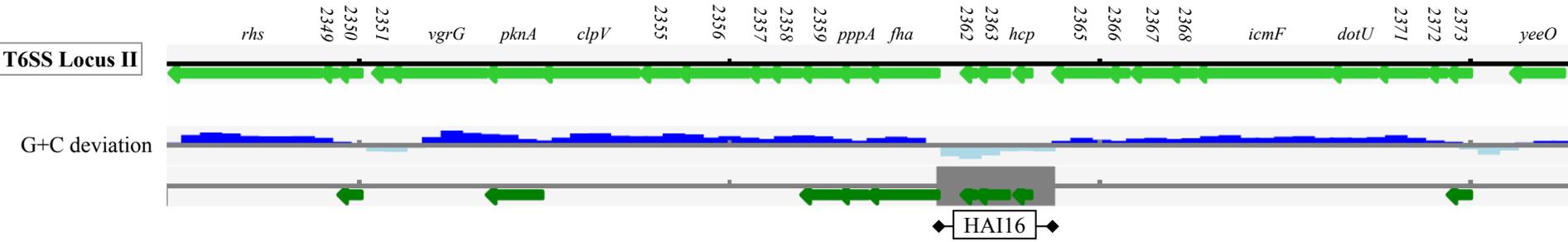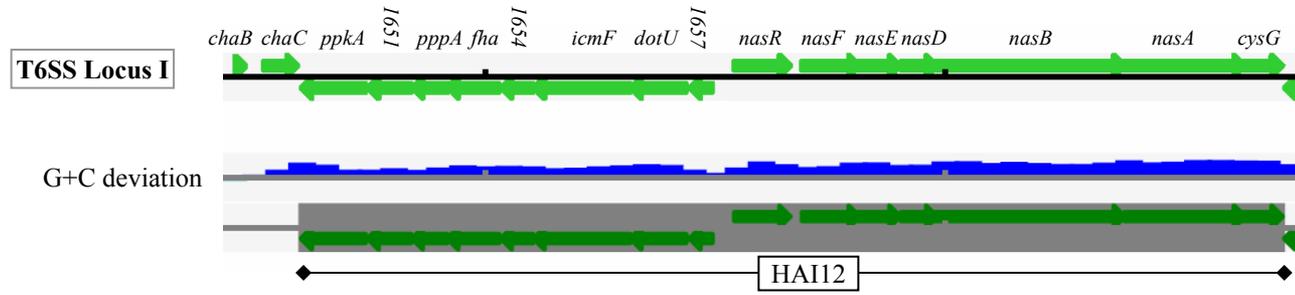
Figure 5.9: The horizontally acquired islands (HAIs) within the *P. ananatis* LMG20103 T6SS loci calculated on the basis of oligonucleotide usage and G+C deviation. Regions with G+C% above the genome average are shown in dark blue, while those below are shown in light blue. Those genes that encode proteins with homologues restricted to plant-associated bacteria are coloured in dark green, those restricted to animal-associated bacteria in red, while the protein unique to LMG20103 is coloured in blue.
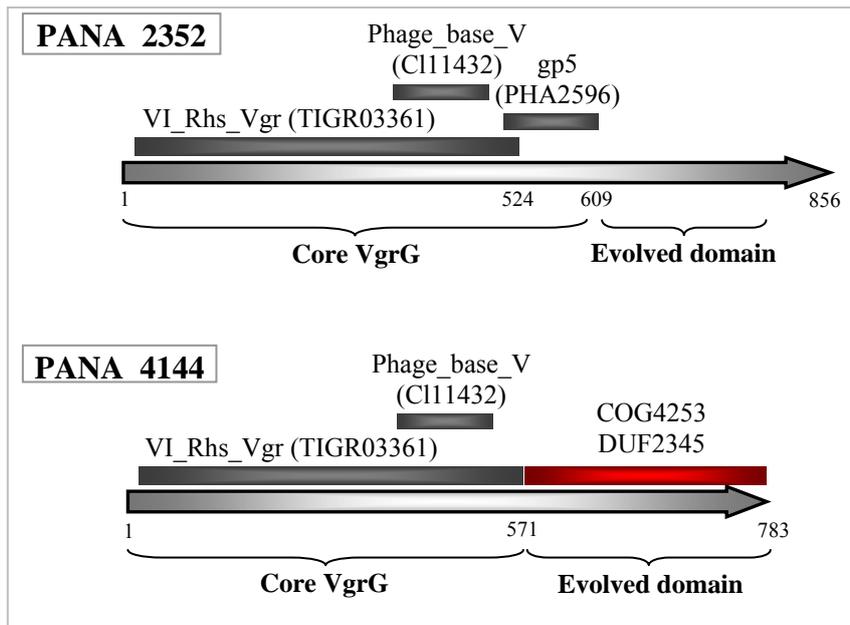
Figure 5.10: Schematic diagrams of the VgrG proteins encoded in *P. ananatis* T6SS locus II (PANA_2352) and locus III (PANA_4144) showing their conserved domains.

Chapter 6:

Functional analysis of Ananatan, an exopolysaccharide produced by *Pantoea ananatis,* which is homologous to stewartan and amylovoran and plays a role in systemic infection of onion and brown-rot disease of pineapple

ABSTRACT

Pantoea ananatis is a broad host range pathogen which causes centre rot of onion and brown rot of pineapple fruit. Little is known about the pathogenicity mechanisms underlying its ability to infect, colonise and cause disease symptoms in its hosts. The whole genome of a *Eucalyptus*-pathogenic strain, *P. ananatis* LMG20103, has recently been sequenced, enabling identification of a large number of putative pathogenicity determinants. One such determinant shows significant homology to the exopolysacch-arides stewartan and amylovoran produced by the closely related phytopathogens *Pantoea stewartii* subsp. *stewartii* and *Erwinia amylovora*, respectively. Structural and functional characterisation of the *P. ananatis* exopolysaccharide, termed ananatan, was undertaken which showed that it plays a major role in the diseases caused by this pathogen on onion and pineapple.

INTRODUCTION

Pantoea ananatis is an emerging plant pathogen with an extensive host range including melons, pineapples, maize, rice, onions and *Eucalyptus*. Symptoms are very diverse and include blight, die-back, rot and decay, depending on the host infected (Coutinho and Venter, 2009). Centre rot of onion caused by *P. ananatis* has resulted in 100% crop losses in fields in Georgia, USA (Gitaitis and Gay, 1997). The disease appears as white streaks with water-soaked margins along the length of the onion leaf. Severe infections result in wilting and bleaching of all leaves with spread down the neck to the bulb which becomes discoloured and decayed (Gitaitis et al. 2002). Blight and dieback have been observed in *Eucalyptus grandis* and *E. grandis* x *nitens* (GN) hybrid seedlings. Water-soaked spots develop on the leaves and coalesce to form larger lesions, with leaf petioles becoming necrotic resulting in abscission of leaves. With severe infections, seedlings appear scorched and repeat infections result in stunting (Coutinho et al. 2002). *P. ananatis* brown-rot disease of pineapple is characterised by brown spots on fruitlets within the fruit, which in severe infections spread to adjacent fruitlets and result in total loss of the fruit (Serrano, 1928). In our laboratory, we have shown that pathogenic *P. ananatis* strains isolated from *Eucalyptus*, onion and pineapple are capable of inducing symptoms on all three host plants, indicating that they are not host-specific (data unpublished).

Although there is considerable information concerning the epidemiology and symptomology of the various plant diseases caused by *P. ananatis*, little is known about the molecular mechanisms underlying pathogenicity. In Chapter 4 all the putative pathogenicity determinants were identified from the genome sequence of the virulent *Eucalyptus* pathogenic *P. ananatis* strain LMG20103. However, the pathogenicity factors were identified based on homology to known pathogenicity factors in other phytopathogenic bacteria and were not functionally characterised. To date, only the quorum-sensing autoinducer, *N-acyl*-homoserine lactone encoded by *eanI* in the onion pathogen *P. ananatis* SK-1 has been experimentally demonstrated to play a role in disease expression (Morohoshi et al. 2007). A mutation in *eanI* resulted in reduced production of extracellular polysaccharide (EPS) and biofilm formation as well as limited necrosis on onion leaves. Although Morohoshi and co-workers (2007) postulated that reduced EPS was responsible for the diminished virulence of the strain this was not conclusively proven. EPS is a major pathogenicity determinant in two close relatives of *P. ananatis*. In the maize pathogen *Pantoea stewartii* subsp. s*tewartii*, copious amounts of stewartan are produced that block the xylem vessels of host plants resulting in occlusion, vascular collapse and wilt (Herrera et al. 2008). Similarly, the biosynthesis of the exopolysaccharide amylovoran by *Erwinia amylovora*, the causative agent of fire blight on several rosaceous plant hosts, results in necrotic lesions, vascular wilt and production of ooze (Bereswill and Geider, 1997). Further roles of EPS in host-microbe interactions include prevention of desiccation, protection against plant host defense recognition and retention of nutrients released from damaged plant cells (Bernhard et al. 1996).

The chemical structures of stewartan and amylovoran have been determined (Nimtz et al. 1996a; Nimtz et al. 1996b). Both are heteropolysaccharides consisting of repeat units of a heptasaccharide and pentasaccharide backbone of glucose, galactose and glucuronic acid subunits, respectively, with two side chains branching from the same galactose. The minor side chain consists of glucose, while the major side chain comprises of glucuronic acid or galactose. In stewartan and amylovoran the major side chain is further decorated with glucose and pyruvate, respectively (Coplin et al. 1996). The genetic determinants for biosynthesis of both types of EPS have been determined. Twelve genes are involved in amylovoran and stewartan synthesis, termed *amsA-L* and *cpsA-L,* respectively (Bernhard et al. 1993; Coplin and

Majerczak, 1990). The repeating saccharide units of stewartan and amylovoran are sequentially assembled on an undecaprenyl phosphate lipid carrier (UndP), transported across the cell membrane and added to a growing stewartan or amylovoran polysaccharide (Langlotz et al. 1999). *amsB, -C, -D, -E, -G, -J and -K* and *cpsA, -D, -E, -F, -G, -J and K* encode glycosyltransferases which catalyse the sequential incorporation of sugars onto the lipid carrier. Final processing, phosphorylation, polymerisation and translocation occur through the activity of the *ams*A, *-F, -H, -I,* and *–L* and *cpsC, -H-, -I* and *–L* enzymes. Downstream of the biosynthetic clusters are *galE* and *galF*, whose products convert glucose to galactose, the principle sugar of amylovoran and stewartan, which is not readily available in the plant host (Metzger et al. 1994). The chemical structure of the EPS has also been determined for a *Eucalyptus* canker-associated bacterium *Erwinia futululu* (Yang et al. 2002) which we have recently correctly identified as *P. ananatis* BCC114 (results unpublished). In this strain, two variants of EPS are produced. One is structurally identical to stewartan, while the other contains a terminal glucose residue replaced by pyruvic acid in the large side chain (Yang et al. 2002).

Genes encoding enzymes for the biosynthesis of an exopolysaccharide showing extensive homology to those involved in stewartan and amylovoran biosynthesis were identified in the genome sequence of *P. ananatis* LMG20103. We have termed this exopolysaccharide ananatan. Given the importance of these EPSs in *P. stewartii* and *E. amylovora* pathogenesis, we undertook to characterise the genetics and structure of ananatan and its function in pathogenesis.

METHODS AND MATERIALS

Strains and culturing conditions

Three *P. ananatis* strains maintained in the Bacterial Culture Collection (BCC) at the University of Pretoria, South Africa, namely the pineapple pathogenic strain LMG2665[T] (Isolated by C. Robbs – 1965), *Eucalyptus*-pathogenic strain LMG20103 (Coutinho et al. 2002) and *Eucalyptus Coniothyrium* canker-associated strain BCC114 (Yang et al. 2002) were maintained on Luria Bertani (LB) agar at 28 °C. *Escherichia coli* S17 carrying the p21 λpir suicide vector which contains a Tn*5* mini-transposon cassette was maintained on LB agar emended with 50 µg/ml kanamycin at 37 °C (Xi et al. 1999). Similarly, *P. ananatis* LMG2665[T] mutants were maintained on

kanamycin LB agar. *P. ananatis*-selective media (PA20) was used for mutant selection (Goszczynska et al. 2006). For EPS quantification and pathogenicity assays strains were grown on EPS-inducing Casamino acid-Peptone-Glucose (CPG) agar emended with 50 μg/ml kanamycin for mutant strains (Bradshaw-Rouse et al. 1981).

Sequence analysis

The genome of the *Eucalyptus* pathogen *P. ananatis* LMG20103 was sequenced (Chapter 2; De Maayer et al. 2010), but this strain proved not to be amenable to genetic manipulation. Therefore a draft genome sequence of the pineapple-pathogenic, genetically alterable strain *P. ananatis* LMG2665[T] was determined using Solexa sequencing (results unpublished). The EPS biosynthetic genes in LMG20103 and LMG2665[T] were identified from the genome sequences by BlastN analysis with *P. stewartii* subsp. *stewartii* stewartan homologues. The amino acid sequences of the EPS genes were determined and compared to those of stewartan and amylovoran. Primers (Inqaba Biotec., South Africa) were designed on the upstream and downstream sequence of the *anaA, -B, -C, -D, -G* and *-H* (Table 6.1) genes from *P. ananatis* LMG20103 and LMG2665[T] to encompass the entire gene sequences. PCR amplification of these biosynthetic genes in *P. ananatis* BCC114 was undertaken with the program: 94 °C for 5 min; 35 cycles [94 °C for 45 sec; 63 °C for 1 min; 72 °C for 2 min]; 72°C for 7 min. Sanger sequencing was performed as per protocol to determine the sequences. A neighbour joining tree was constructed based on the ClustalW alignment of the concatenated amino acid sequences for AnaABCDGH and their homologues in *P. stewartii* and *E. amylovora* using the Mega 4.0 software package (Tamura et al. 2007).

Random transposon knock-out mutagenesis

A transposon mutation grid was constructed based on the protocol by Holeva et al. (2004) for the enterobacterial phytopathogen *Pectobacterium atrosepticum* with modifications specific for *P. ananatis* as described below:

Transposon mutagenesis. Random mutagenesis was performed by mobilisation of the *E. coli* S17-1 λpir suicide plasmid by conjugation into *P. ananatis* LMG2665[T] (de Lorenzo et al. 1990). Two ml of an overnight culture of LMG2665[T] were mixed with 1 ml *E. coli* S17 and 5 ml of 10 mM MgCl$_2$. The mixture was centrifuged and 100 μl of the suspension was plated on LB agar and incubated at 28 ºC overnight. 2 ml of 10

mM MgCl$_2$ was added to the plate and the cells were scraped off and centrifuged. The pellet was resuspended in 200 μl of 10 mM MgCl$_2$, spread on PA20 selective agar amended with 50 μg/ml kanamycin and incubated for 72 h at 28 ºC. Mutants were picked off using a pipette tip and stored in 384-well plates containing 100 μl LB broth with 15% glycerol and 50 μg/ml kanamycin and stored at – 70 ºC. Twenty four 384-well plates amounting to 9,216 individual mutants were selected. Given that this technique introduces transposons within the genome at random, and that 4,237 protein coding sequences are encoded on the *P. ananatis* LMG20103 chromosome (Chapter 2), ~2.2 transposon insertions per protein coding sequence can be expected. Incorporation of the transposon cassette was determined by PCR of DNA extracted from several putative mutants with primers designed on the kanamycin resistance gene (*kan*) in the λpir suicide vector (Table 6.1). Nine mutants selected at random among the twenty-four 384-well microtitre plate and the wild-type LMG2665^T strain were included.

DNA extraction. DNA was extracted from pooled cells. Twenty four pools were made by growing all colonies from each individual 384-well plate on LB agar plates amended with 50 μg/ml kanamycin overnight. Five ml of resuspension buffer (T$_{50}$E$_{50}$) was added and the cells scraped off the plate. DNA was extracted by boiling 50 μl of the pooled cell suspensions for 10 minutes. The boiled suspensions were centrifuged at 13,000 rpm for 1 min and the supernatants transferred to fresh eppendorf tubes.

Screening for mutants. Screening PCRs were performed on the 24 pools with the Tn5-specific primer Tn5R designed on the Tn5 cassette sequence (Jan Michiels, Personal communication) and the gene-specific forward and reverse primers for the six ananatan biosynthetic genes (Table 6.1). PCR amplification was performed as per the program: 95 °C for 5 min; 30x [94 °C for 1 min; 62 °C for 1 min; 72 °C for 2 min]; 72 °C for 1 min. Agarose gel electrophoresis was done with the amplicons, using 1% agarose gels, run in 0.5x TBE buffer at 100 V for 45 minutes. Products unique to a particular 384-well plate for a specific primer combination were extracted from the gel using a Zymoclean Gel DNA recovery kit (Zymo Research, USA) and Sanger sequenced using the Tn5R and *ana* gene-specific primer. The plate containing the correct gene mutation was further sub-divided into four 96 colony pools (sectors), and re-screened by PCR. Positive 96-colony sectors were again divided into twelve pools

260
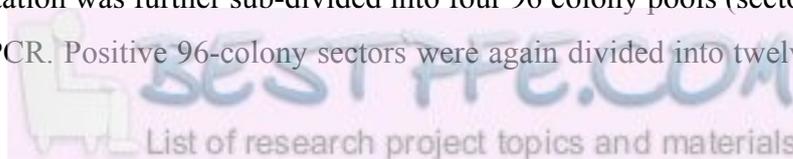
containing eight colonies, re-screened and thereafter the positive pool subdivided into eight single colonies. The amplicons obtained for putative individual mutants were blunt-end ligated into the pJET cloning vector (Fermentas Life Sciences, USA) and amplified and sequenced using the pJET1.2 forward and reverse primers as per protocol (CloneJET manual, Fermentas Life Sciences, USA) to validate the insertion and location of the Tn5 cassette within the target gene.

EPS Quantification Assay

The exopolysaccharides (EPS) produced by the mutant and wild-type strains of *P. ananatis* LMG2665[T] and LMG20103 were quantified by a modified version of the ethanol-precipitation anthrone method described by Coplin et al. (1986). Wild-type and mutant strains were grown to an optical density of 0.6 (measured at O.D. = 600 nm) in 50 ml of CPG broth incubated at 28 ºC and shaken at 150 rpm. Cultures were boiled for 15 minutes and centrifuged at 13,000 rpm for 5 minutes. The EPS-containing supernatant was collected and precipitated in two volumes of 100% ethanol overnight at -20 ºC. After centrifugation, the resulting pellet was reconstituted in 5 ml hot distilled water and 0.6 ml added to a chilled test-tube. 3ml of anthrone reagent (0.2 g anthrone resuspended in 5 ml ethanol added to 45 ml 75% H_2SO_4) was added and the contents of the test-tubes boiled for 10 minutes. The $O.D._{625}$ (measured at 625 nm) was determined and compared to a glucose standard curve, where 0-500 µg/ml $_D$-glucose was added to distilled water and concurrently treated like the tested strains. Six replicates of the two wild-type and mutant strains were performed.

Plant infection assays

Onion seedlings and pineapple fruits were employed in pathogenicity assays as symptom development is rapid and substantial. An initial screen was performed with both the *Eucalyptus*-pathogenic strain LMG20103 and pineapple-pathogenic strain LMG2665[T] to determine their capacity to cause disease on onion and pineapple. Susceptible onion seedlings were stab-inoculated with colonies grown on CPG agar for 48 h at 28 ºC as described by Morohoshi et al. (2007). Subsequently, duplicate inoculations per leaf were performed with the wild-type and mutant strains on three plants each, with three plants inoculated with sterile distilled water as control. The trial was repeated in triplicate. Plants were monitored 2, 3 and 5 days post-inoculation and again after 1-3 weeks. Pineapple fruits were likewise stab-inoculated with 1 ml of

cultures (O.D.$_{625}$ = 0.4) of the wild-type and mutant strains. The inoculation sites were closed with autoclaved paraffin to avoid contamination with secondary pathogens. Four pineapple fruits each were inoculated with the tested wild-type and mutant strains and with sterile distilled water as control. The pineapple fruits were cut open five days post-inoculation and the presence of brown rot lesions was determined.

Statistical analyses

Statistical analyses were performed on the onion lesion measurements using the STATISTICA® (Statsoft, 2001) package. An analysis of variance (ANOVA) and post hoc Tukey test were conducted at 95% confidence intervals.

RESULTS AND DISCUSSION

Genetics of *Pantoea ananatis* ananatan exopolysaccharide biosynthesis

Based on the extensive DNA sequence homology to stewartan biosynthetic genes, twelve genes were identified in the ananatan gene locus of *P. ananatis* LMG20103 and LMG2665^{T}. We termed these genes *anaA* to *anaL* in keeping with the name of the exopolysaccharide **ana**natan while the gene epithets matched those for the homologues in the *P. stewartii* stewartan locus. The *Eucalyptus* strain LMG20103 and pineapple strain LMG2665 ananatan loci are both 18,333 nucleotides in length and share 99% nucleotide identity. Comparison of the amino acid sequences for the ananatan, stewartan and amylovoran biosynthetic genes showed significant sequence identity exists between them (Fig. 6.1). The predicted amino acid sequences of LMG20103 and LMG2665^{T} show 99.9% average aa identity, with 5 amino acid substitutions between them, and show 95% and 67% average amino acid identity to the stewartan and amylovoran homologues, respectively (Fig. 6.1). Lower homology is observed between the putative ananatan/amylovoran glycosyltransferases AnaD/AmsC, AnaF/AmsE and AnaG/AmsD, with 35%, 11% and 31% amino acid identity, respectively. The gene order is conserved between the stewartan and ananatan loci, but an alternate gene order can be observed between the biosynthetic genes for ananatan and amylovoran, where *amsB/amsC* and *anaD/anaE* as well as *amsD/amsE* and *anaF/anaG* occur in an inverse order (Fig. 6.1). Similar differences in gene order and sequence homology between stewartan and amylovoran loci have been correlated to structural differences observed between these exopolysaccharides (Geider, 2000). Extensive genetic homology thus exists between the ananatan,

262

stewartan and amylovoran biosynthetic loci. It is therefore likely that AnaA, -D, -E – F, -G, -K and AnaJ represent glycosyltransferases, while AnaB, -C, -H, -I and –L are involved in the final processing, phosphorylation, and polymerisation of the repeat heptasaccharide units in the ananatan polymer.

Structural analysis of _P. ananatis_ ananatan exopolysaccharide

Specific primers were used to amplify and sequence the _anaA, -B, -C, -D, -G_ and _-H_ genes of _P. ananatis_ BCC114 for which the EPS chemical structure has previously been determined (Fig. 6.2; Yang et al. 2002.). Products for each of the _ana_ genes were obtained. The amino acid sequences for these six genes show 99.3 and 99.5% average identity to those in LMG20103 and LMG2665[T], respectively. A neighbour-joining tree of a PAM62 ClustalW alignment shows that the concatenated amino acid sequences of the _P. ananatis_ BCC114 AnaA, -B, -C, -D, G- and -H homologues clustered with those of LMG20103 and LMG2665[T] (Fig. 6.3). The identical structure observed for the _P. ananatis_ BCC114 EPS and stewartan and the high sequence homology between the ananatan and stewartan biosynthetic enzymes (Yang et al. 2002) thus suggest highly similar structures for stewartan and the EPSs of LMG20103, LMG2665[T] and BCC114. Yang et al. (2002) describe a second EPS produced in _P. ananatis_ BCC114 with a pyruvic acid modification in the major side chain, a feature shared with amylovoran, but not stewartan (Coplin et al. 1996). A terminal pyruvic acid moiety has also been observed in _Dickeya dadantii_ Ech6 EPS, succinoglycan in _Sinorhizobium melilotii_ and in xanthan exopolysaccharides produced by _Xanthomonas_ species (Casas et al. 2000; Becker et al.1993; Yang et al. 2001) and has been suggested to protect the EPS from degradation by α-galactosidases in the plant host and increase EPS viscosity (Casas et al. 2000).

On the basis of homology between the ananatan, stewartan and amylovoran glycosyltransferases, the enzymes involved in the sequential addition of sugars to the ananatan repeat units could be predicted (Fig. 6.4). AnaA thus likely represents the galactosyl transferase that adds the first galactose residue to the UndP lipid carrier, while AnaD adds the other backbone galactose residue. AnaK and -E add glucuronic acid and galactose residues to the major side chain and AnaJ will add the minor glucose side chain. The _cpsF_ and _cpsG_ genes of _P. stewartii_ have been shown to complement an _E. amylovora amsD_ mutant, resulting in amylovoran repeat units with

the backbone galactose replaced by glucose (Geider, 2000). CpsF and –G homologues in the ananatan loci thus likely represent glucosyltransferases that could add a glucose residue in the backbone or the major side chain. BLASTP analysis against the Pfam database with the AnaJ sequences of LMG20103 and LMG2665[T] indicated that these proteins contain conserved domains found in polysaccharide pyruvyltransferases (Pfam04230). AnaJ may thus be bifunctional, introducing either a glucose or pyruvate residue into the ananatan major side chain (Fig. 6.4) and may explain the findings of Yang et al. (2002) of two EPS structures. The presence of a homologue in *P. stewartii* subsp. *stewartii* (CpsJ – 95% aa identity to AnaJ) implies that although not documented, stewartan side chains substituted with pyruvic acid may also exist.

Creation of an *anaD* knock-out mutant

Random transposon insertion mutants (9,216) of *P. ananatis* LMG2665[T] were created. Transposon insertion into the mutant chromosomes was confirmed by PCR with the *kmr-F* and *kmr-R* primers for the kanamycin resistance gene. Subsequently, the twenty-four 384-well microtitre plate pools were screened for mutants with the transposon inserted in the six *ana* genes using *ana* gene-specific forward/reverse and Tn5R primers (Table 6.1). Agarose gel electrophoresis of the amplicons obtained with each of the 12 primer combinations showed numerous bands for most of the 24 pools. This indicates extensive non-specific binding of the primers. Numerous bands unique to one of 24 four pools were excised from the gel for each of the 12 primer combinations. Using Sanger sequencing, the sequences of the *ana* gene-specific-Tn5R products were determined. One 384-well microtitre plate (Plate 24) was found to contain a mutant with a transposon insertion within the *anaD* gene which was amplified with the *anaD-F* and Tn5R combination. This plate was further subdivided into four sectors (A-D) of 96 colonies and re-screened and sector D was found to contain the mutant. Sector D was divided into 16 pools (1-16) of eight colonies. The correct mutant product was amplified in pool 10. The eight (A-H) colonies within this pool were amplified using the *anaD*-F and Tn5R primers and agarose gel electrophoresis showed colony F contains the product of expected size (Fig. 6.5). The product was extracted from the gel and ligated into the pJET cloning vector and subsequently amplified and Sanger sequenced with the pJET specific primers. The location of the transposon cassette within the *anaD* gene was confirmed (Fig. 6.6). The *anaD* mutant was named PANA24D10F.

EPS production in the *anaD* mutant PANA24D10F

An EPS quantitative assay was performed as described for the wild-type strains *P. ananatis* LMG20103 and LMG2665[T] and the *anaD* mutant PANA24D10F. Comparison of the two wild-type strains showed no significant difference in exopolysaccharide production occurred between them (Fig. 6.7), with the *P. ananatis* LMG2665[T] and LMG20103 measurements of 272±31.4 and 253±28.5 μg glucose/ml, respectively (Mean ± standard deviation). In contrast, the *anaD* mutant PAN24DJ1 produced only 158±15.5 μg glucose/ml (Fig. 6.7), showing a significant reduction in EPS biosynthesis compared to the wild-type strains, 42% lower than the wild-type LMG2665[T]. AnaD is a putative galactosyltransferase that is predicted to add a galactose residue to the backbone. Thus disruption of *anaD* would be expected to completely halt EPS production. Mutagenesis of the *E. amylovora amsC* and *P. stewartii cpsD* has also been shown to reduce amylovoran and stewartan production, respectively (Menggad and Laurent, 1998; Dolph et al. 1988). While evidently reduced in glucose content, there are still significant levels of glucose/ml in the PANA24D10F strain. This may be due to the *anaD* mutant compensating by producing other capsular polysaccharides, such as the group 4 capsule (Chapter 4) or membrane-derived oligosaccharides (MDOs) (Chapter 4), to ensure survival, preventing dessication and providing protection against osmotic stress. Furthermore, the EPS extraction and quantification methods employed can be considered to be relatively crude and the presence of carry-over glucose from the growth medium cannot be discounted. However, it can be concluded that *anaD* mutation resulted in significant reduction of EPS production.

Functional analysis of the *P. ananatis* LMG2665[T] *anaD* mutant

As part of this study, pathogenicity assays were undertaken on onion seedlings. Following incubation, localised lesions were observed on seedlings inoculated with the wild-type strains *P. ananatis* LMG20103 and LMG2665[T] and the *anaD* mutant PANA24D10F. Two days post-infection (Dpi) lesions of similar length were observed in plants infected with both wild-type and mutant strains (Fig. 6.8). However, five dpi lesions had become significantly larger in wild type inoculated plants (average lesion length of 74.75 mm and 65.86 mm for LMG20103 and LMG2665[T], respectively) than in those inoculated with mutant PANA24D10F (average lesion length of 17.12 mm) (Fig. 6.8; Fig. 6.9). By seven dpi with the wild-type cultures, leaves became wilted

and dried from their neck to their tip, with leaves that were not directly inoculated also showing similar signs of wilting (Fig. 6.9). By contrast, lesions on the leaves of mutant-inoculated onion seedlings, even 2 to 3 weeks post-inoculation, remained the same size as those observed five dpi and plants showed no obvious systemic infection (Fig. 6.9). Taken together, these results indicate that ananatan may induce systemic infection through vascular occlusion and wilt. A similar pattern has been observed in maize infected with *P. stewartii* subsp. *stewartii*, where stewartan-deficient mutants were unable to induce wilting, but could still elicit restricted necrotic lesions (Dolph et al. 1988; Coplin and Majerczak, 1990). Other pathogenicity determinants could be responsible for formation of necrotic lesions. For example, two large repetitive RTX toxins were recently identified in *P. stewartii* as playing a role in the formation of watersoaked lesions on maize (Roper and von Bodman, 2008).

Pineapple fruits inoculated with the wild-type *Eucalyptus* pathogen *P. ananatis* LMG20103 and pineapple pathogen *P. ananatis* LMG2665^T showed no external symptoms. However, when the pineapple fruits were sliced in half, typical symptoms of brown-rot disease could be observed (Fig. 6.10). Large brown to dark-brown spots extended from the site of inoculation throughout the entire fruitlet towards the fruit core. The tissue around the site of infection appeared macerated and occurred in combination with an alcoholic rotting smell. By contrast, pineapple fruits inoculated with the *anaD* mutant PANA24D10F had small zones of maceration around the site of inoculation, but no brown spots could be observed (Fig. 6.10). No maceration or brown rot symptoms could be observed in the control pineapple fruits inoculated with sterile distilled water. Ananatan can thus be concluded to play a role in brown spot disease on pineapple fruit. The lack of brown spots in pineapple fruits inoculated with the mutant suggests a role for ananatan in water-soaking and tissue necrosis. As only minor localised maceration occurred in the mutant, compared to the lesion extending towards the fruit core in the wild-type strains, ananatan may also be involved in spread of the pathogen throughout the infected fruitlets.

CONCLUSIONS

In this study, genetic loci for the production of an exopolysaccharide, which has been named ananatan, were identified in the pineapple pathogen *P. ananatis* LMG2665^T and *Eucalyptus*-pathogenic *P. ananatis* LMG20103. These show significant genetic homology to the *P. stewartii* and *E. amylovora* loci involved in the production of the structurally similar stewartan and amylovoran exopolysaccharides. Through mutational analysis it was shown that ananatan represent a major pathogenicity determinant in both systemic disease of onion and brown-rot disease of pineapple. Localised lesions in onion seedlings and tissue maceration in pineapple fruits observed when plants were inoculated with the ananatan mutant, however, suggest other pathogenicity factors are also involved in disease. The rapid symptom development on both onions and pineapples indicate that they represent good model systems to test the functions of other pathogenicity determinants in *P. ananatis* disease. The availability of the genome sequences for *P. ananatis* LMG20103 and LMG2665^T and the expansive range of putative pathogenicity targets identified from the genome sequence of *P. ananatis* LMG20103 (Chapter 4) may allow us to identify other pathogenicity genes responsible for lesion formation. Furthermore an extensive set of random mutants was generated in this experiment. Following optimisation of the screening technique, mutational analysis could thus be combined with the genomic data to identify and gain an understanding of the pathogenicity determinants employed by this broad host range, economically important plant pathogen.

REFERENCES

Becker, A., Kleickmann, A., Küster, H., Keller, M., Arnold, W. et al. 1993. Analysis of the *Rhizobium meliloti* genes *exoU*, *exoV*, *exoW*, *exoT*, and *exoI* involved in exopolysaccharide biosynthesis and nodule invasion: *exoU* and *exoW* probably encode glucosyltransferases. Mol. Plant-Microbe Interact. **6**: 735-744

Bereswill, S. and Geider, K. 1997. Characterization of the *rcsB* gene from *Erwinia amylovora* and its influence on exopolysaccharide synthesis and virulence of the fire blight pathogen. J. Bacteriol. **179**: 1354-1361

Bernhard, F., Bellemann, P., Bugert, P., Coplin, D. L. and Geider, K. 1993. Heterologous expression of EPS-genes in *Erwinia amylovora* and *Erwinia stewartii*. Acta Hort. **338**: 265-27

Bernhard, F., Schullerus, D., Bellemann, P., Geider, K., Majerczak, D. R. et al. 1996. Genetics and complementation of DNA regions involved in amylovoran synthesis of *Erwinia amylovora* and stewartan synthesis of *Erwinia stewartii*. Acta Hort. **411**: 269-274

Bradshaw-Rouse, J. J., Whatley, M. A., Coplin, D. L., Woods, A., Sequeira, L. et al. 1981. Agglutination of strains of *Erwinia stewartii* with a corn agglutinin: correlation with extracellular polysaccharide production and pathogenicity. Appl. Environ. Microbiol. **42**: 344-350

Casas, J. A., Santos, V. E. and García-Ochoa, F. 2000. Xanthan gum production under several operational conditions: molecular structure and rheological properties. Enzyme and Microb. Technol. **26**: 282-291

Coplin, D. L. and Majerczak, D. R. 1986. Molecular Cloning of Virulence Genes from *Erwinia stewartii*. J. Bacteriol. **168**: 619-623

Coplin, D. L. and Majerczak, D. R. 1990. Extracellular polysaccharide genes in *Erwinia stewartii*: directed mutagenesis and complementation analysis. Mol. Plant-Microbe Interact. **3**: 286-292

Coplin, D. L., Majerczak, D. R., Bugert, P. and Geider, K. 1996. Nucleotide sequence analysis of the *Erwinia stewartii cps* gene cluster for synthesis of stewartan and comparison to the *Erwinia amylovora ams* cluster for synthesis of amylovoran. Acta Hort. **411**: 251-257

Coutinho, T. A., Preisig, O., Mergaert, J., Cnockaert, M. C., Riedel, K.-H. et al. 2002. Bacterial blight and dieback of *Eucalyptus* species, hybrids, and clones in South Africa. Plant Dis. **86**: 20-25

Coutinho, T. A. and Venter, S. N. 2009. *Pantoea ananatis*: an unconventional plant pathogen. Mol. Plant Pathol. **10**: 325-335

de Lorenzo, V., Herrero, M., Jakubzik, U. and Timmis, K. N. 1990. Mini-Tn*5* transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in Gram-Negative Eubacteria. J. Bacteriol. **172**: 6568-6572

De Maayer, P., Chan, W. Y., Venter, S. N., Toth, I. K., Birch, P. R. J. et al. 2010. Genome sequence of *Pantoea ananatis* LMG20103, the causative agent of *Eucalyptus* blight and dieback. J. Bacteriol. **192**: 2936-2937

Dolph, P. J., Majerczak, D. R. and Coplin, D. L. 1988. Characterization of a gene cluster for exopolysaccharide biosynthesis and virulence in *Erwinia stewartii*. J. Bacteriol. **170**: 865-871

Geider, K. 2000. Exopolysaccharides of *Erwinia amylovora*: structure, biosynthesis, regulation, role in pathogenicity of amylovoran and levan. In <u>Fire blight: the disease and its causative agent</u>. Ed. Vanneste, J.L. Wallingford, CAB International

Gitaitis, R., Walcott, R., Culpepper, S., Sanders, H., Zolobowska, L. et al. 2002. Recovery of *Pantoea ananatis*, causal agent of center rot of onion, from weeds and crops in Georgia, USA. Crop Prot. **21**: 983-989

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

Goszczynska, T., Venter, S. N. and Coutinho, T. A. 2006. PA 20, a semi-selective medium for isolation and enumeration of *Pantoea ananatis*. J. Microbiol. Methods **64**: 225 - 231

Herrera, C. M., Koutsoudis, M. D., Wang, X. and von Bodman, S. B. 2008. *Pantoea stewartii* subsp. *stewartii* exhibits surface motility, which is a critical aspect of Stewart's wilt disease development on maize. Mol. Plant-Microbe Interact. **21**: 1359-1370

Holeva, M. C., Bell, K. S., Hyman, L. J., Avrora, A. O., Whisson, S. C. et al. 2004. Use of a pooled transposon mutation grid to demonstrate roles in disease development for *Erwinia carotovora* subsp. *atroseptica* putative Type III secreted effector (DspE/A) and helper (HrpN) proteins. Mol. Plant-Microbe Interact. **17**: 943-950

Langlotz, C., Ullrich, H., Geider, K., Huber, A., Nimtz, M. et al. 1999. Biosynthesis and characterization of capsular exopolysaccharides from *Erwinia amylovora* and *Erwinia stewartii*. Acta Hort. **489**: 307-313

Menggad, M. and Laurent, J. 1998. Mutations in *ams* genes of *Erwinia amylovora* affect the interactions with host plants. Eur. J. Plant Pathol. **104**: 313-322

Metzger, M., Bellemann, P., Bugert, P. and Geider, K. 1994. Genetics of galactose metabolism of *Erwinia amylovora* and its influence on polysaccharide synthesis and virulence of the fire blight pathogen. J. Bacteriol. **176**: 450-459

Morohoshi, T., Nakamura, Y., Yamazaki, G., Ishida, A., Kato, N. et al. 2007. The plant pathogen *Pantoea ananatis* produces N-acylhomoserine lactone and causes center rot disease of onion by quorum sensing. J. Bacteriol. **189**: 8333-8338

Nimtz, M., Mort, A., Domke, T., Wray, V., Zhang, Y. et al. 1996a. Structure of amylovoran, the capsular exopolysaccharide from the fire blight pathogen *Erwinia amylovora*. Carbohydr. Res. **287**: 59-76

Nimtz, M., Mort, A., Wray, V., Domke, T., Zhang, Y. et al. 1996a. Structure of stewartan, the capsular exopolysaccharide from the corn pathogen *Erwinia stewartii*. Carbohydr. Res. **288**: 189-201

Roper, C. and von Bodman, S. B. (2008). B-151. The Role of RTX Toxins in Virulence of Pantoea stewartii subsp. stewartii, the causal agent of Stewart's wilt of corn. American Society for Microbiology 108th general meeting.

270

Serrano, F. 1928. Bacterial fruitlet brown-rot of pineapple in the Phillipines. The Philippine J. Sci. **36**: 271-300

Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. **24**: 1596-1599

Xi, C., Lambrecht, M., Vanderleyden, J. and Michiels, J. 1999. Bi-functional *gfp*- and *gusA*-containing mini-Tn*5* transposon derivatives for combined gene expression and bacterial localization studies. J. Microbiol. Methods **35**: 85-92

Yang, B. Y., Brand, J. and Montgomery, R. 2001. Pyruvated galactose and oligosaccharides from *Erwinia chrysanthemi* Ech6 extracellular polysaccharide. Carbohydr. Res. **331**: 59-67

Yang, B. Y., Ding, Q. and Montgomery, R. 2002. Extracellular polysaccharides of *Erwinia futululu*, a bacterium associated with a fungal canker disease of *Eucalyptus* spp. Carbohydr. Res. **337**: 2469-2480

FIGURES AND TABLES

Table 6.1: Forward (-F) and reverse (-R) primer sequences for the target genes *anaA, -B, -C, -D, -G* and *-H*, the kanamycin gene (*kan*) and Tn5 transposon on the λpir suicide plasmid (Dr. Jan Michiels, personal communication). The sizes of amplicons and primers, primer G+C% and melting temperatures are shown.

| Target gene	Size (nt)	Amplicon size (nt)	Primer size (nt)	G+C content (%)	Tm (°C)
	Primer sequence				
anaA	1,433	1,583			
anaA-F	5'-GGTTACTTAAGCGTCTGGTGTG-		22	50.0	55.7
anaA-R	5'-GGTTTATCAATGCCGTACCGC-3'		21	52.4	56.6
anaB	1,139	1,288			
anaB-F	5'-CATGTTCAGGGCCTGAGTCAG-3'		21	57.1	57.8
anaB-R	5'-CAGATGTTGCCTACGCAGACG-3'		21	57.1	58.3
anaC	2,189	2,296			
anaC-F	5'-ACGTGTCGTTATCTTGAGCCAG-		21	52.4	57.0
anaC-R	5'-GAACACGCCACTCAGAAATGG-3'		21	52.4	56.5
anaD	1,157	1,278			
anaD-F	5'-GCTGAGAGATCAACCCGGAAG-3'		21	57.1	57.5
anaD-R	5'-TGTGTTGTTACAGGCGACCG-3'		20	55.0	57.8
anaG	1,034	1,138			
anaG-F	5'-GCCTGAAACACCGTTAAGCCG-3'		21	57.1	59.1
anaG-R	5'-AGGTGTCCCATGAGGCACATC-3'		21	57.1	59.4
anaH	2,210	2,281			
anaH-F	5'-ACGGCCTTCATGACGACGGG-3'		20	65.0	62.1
anaH-R	5'-TGCTGTTGGTGTTGGCTGCCG-3'		21	61.9	63.5
kan	792	766			
kmr-F	5'-GAAGAACTCGTCAAGAAGGCG-3'		21	52.4	55.8
kmr-R	5'-TGAACAAGATGGATTGCACGC-3'		21	47.6	56.3
Tn5R	5'-CCGGGGATCCTCTAGAAAGC-3'		20	60.0	57.0
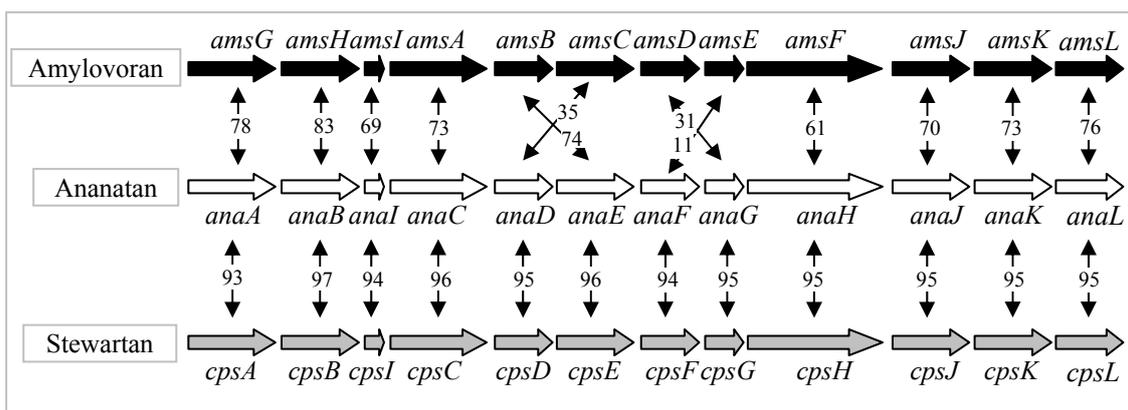
Figure 6.1: Comparison of the *E. amylovora* amylovoran, *P. ananatis* LMG2665[T] ananatan and *P. stewartii* DC283 stewartan biosynthetic loci. An inversion can be observed between *amsB/amsC* and *anaD/E* and the *amsD/E* and *anaF/G* genes. The percentage amino acid identities between the amino acid sequences for each gene are shown.
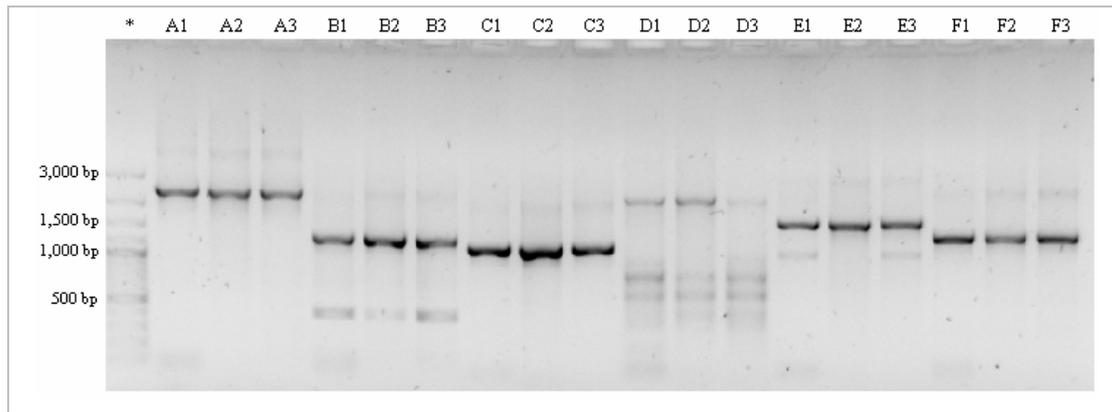
272

Figure 6.2: Agarose gel of the *anaC* (A), *anaD* (B), *anaG* (C), *anaH* (D), *anaA* (E) and *anaB* (F) amplicons for *P. ananatis* LMG20103 (1), *P. ananatis* LMG2665[T] (2) and *P. ananatis* BCC114 (3). A 100 bp DNA ladder was included (*) to estimate product sizes.
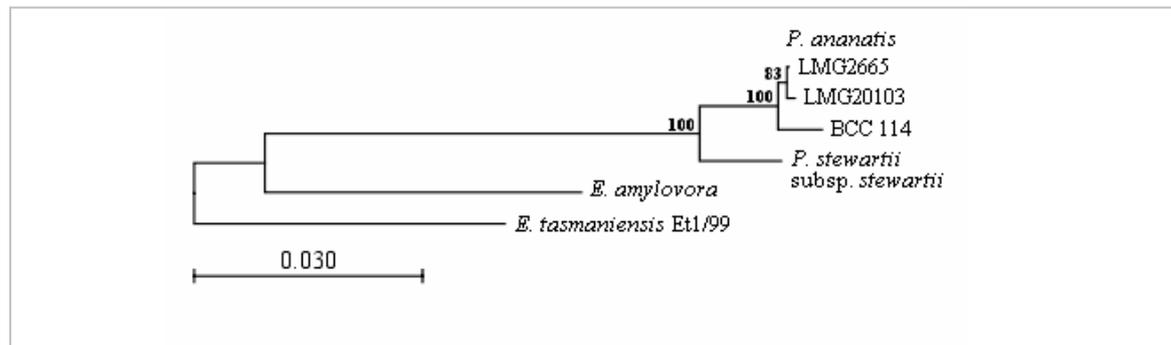

Figure 6.3: Neighbour-joining tree of a PAM62 ClustalW alignment of the concatenated amino acid sequences of AnaABCDGH of *P. ananatis* LMG2665, LMG20103 and BCC114, as well as the concatenated sequences for *P. stewartii* subsp. *stewartii* CpsABCDGH and *E. amylovora* AmsACDFGH. The tree was rooted with the AnaABCDGH homologues in *Erwinia tasmaniensis* Et1/99. Bootstrap values (n=1000) are shown.
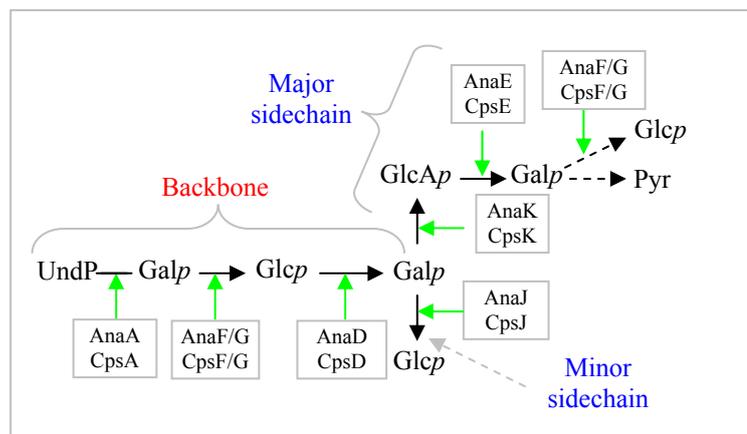
Figure 6.4: Predicted structure of the ananatan repeat polymer. UndP is the undecaprenyl lipid carrier, bound to the sugars Galactose (Gal*p*), Glucose (Glc*p*), Glucuronic acid (GlcA*p*) and Pyruvate (pyr). Enzymes involved in each sugar incorporation step for the exopolysaccharides produced by *P. stewartii* subsp. *stewartii* (Cps) and *P. ananatis* LMG2665[T] are shown in grey blocks.
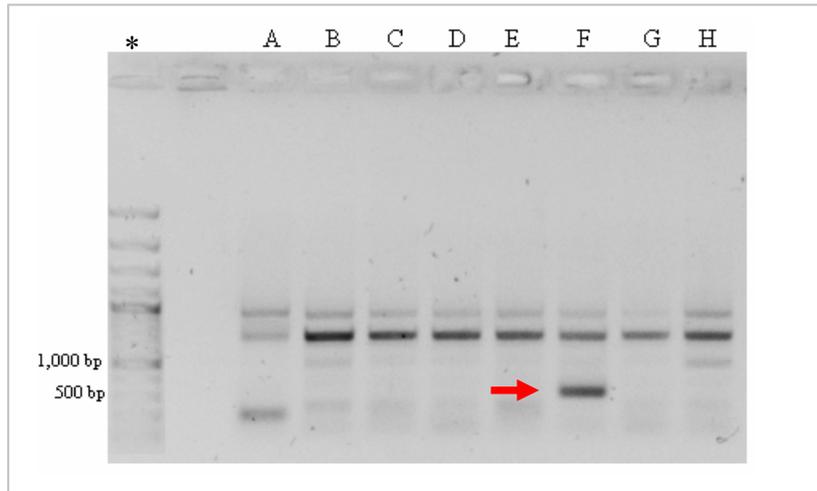
273

Figure 6.5: Agarose gel showing the *anaD-F* – Tn5R amplicons for the eight individual colonies (A-H) on plate 24, sector D, pool 10. The red arrow shows the product of expected size in the *anaD* mutant PANA24D10F. A 100 bp DNA ladder was run (*) to estimate product sizes.
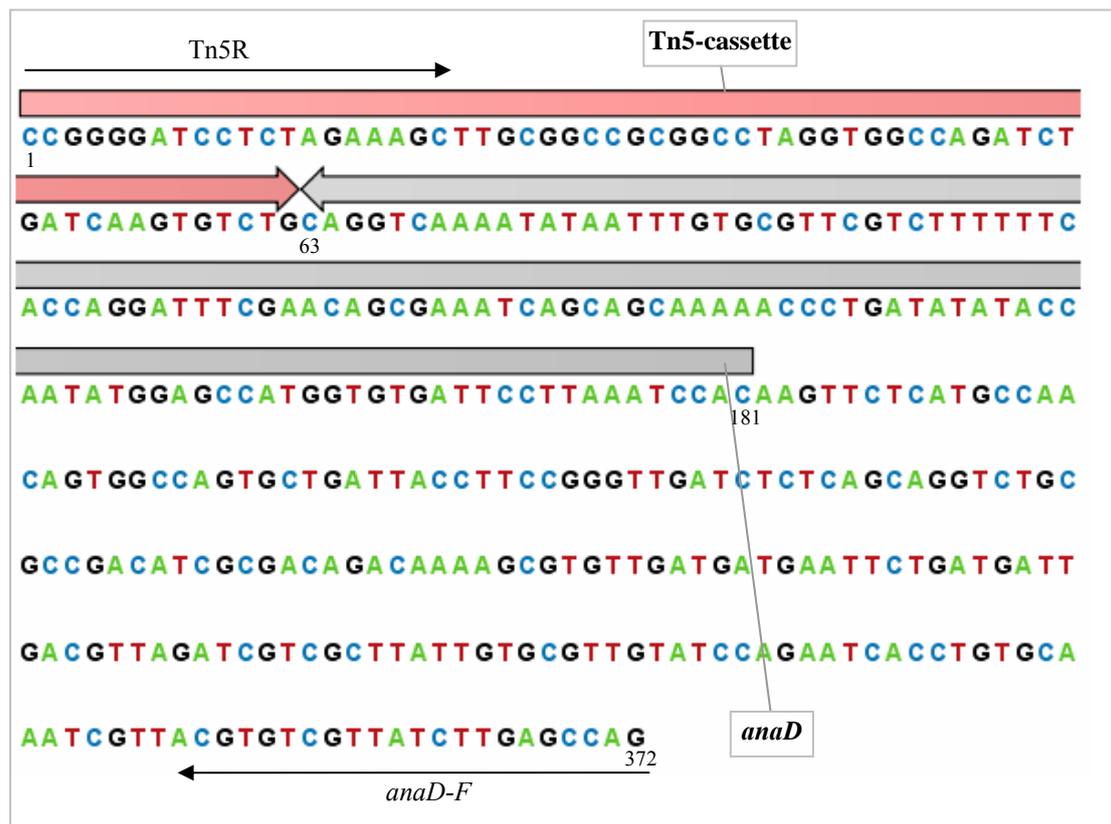

Figure 6.6: Chromosomal integration site of the Tn5 cassette in the ananatan biosynthetic gene *anaD*. The binding sites of the Tn5R and *anaD-F* primers are shown. The *anaD-F*-Tn5R product is 372 nt in size, with the transposon integration falling 118 nt within the *anaD* gene.
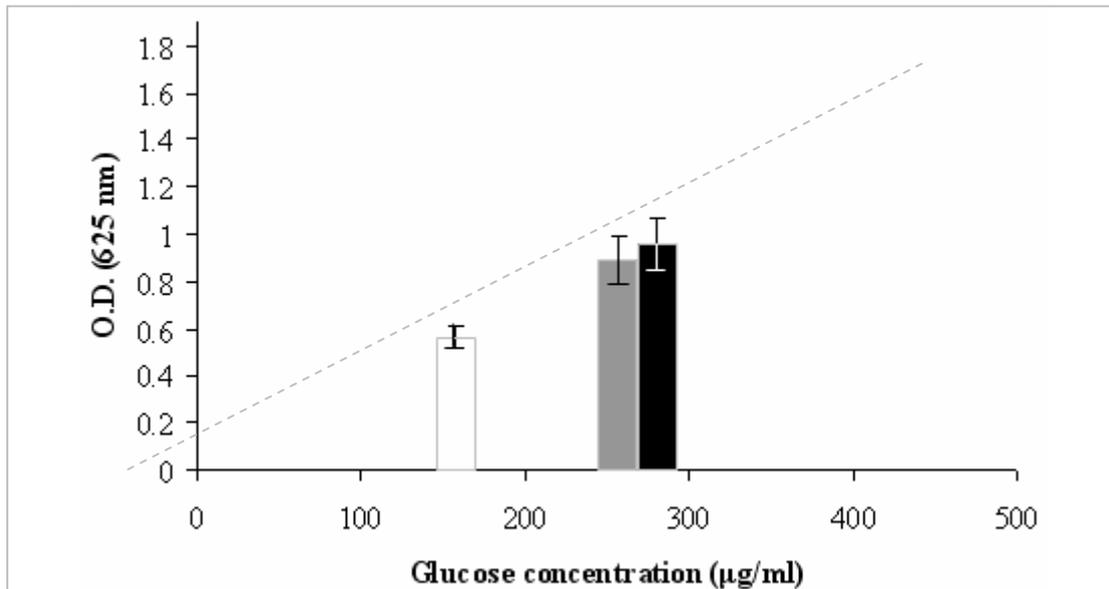
Figure 6.7: Optical densities (O.D.) measured at 625 nm plotted against glucose concentration. The trend line shows the glucose standard measurements. The amount of EPS sugars released during chemical degradation of the PANA24D10F mutant (white bar), and wild-type strains *P. ananatis* LMG20103 (grey bar) and LMG2665[T] (black bar) are shown. Vertical bars for each measurement indicate 95% confidence intervals.
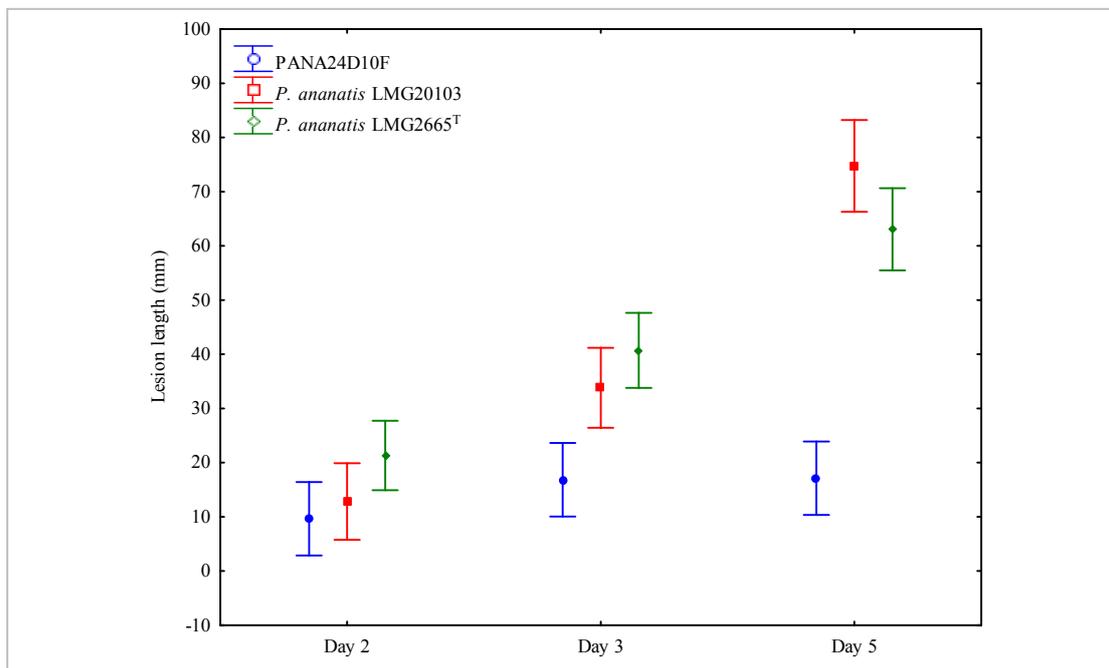

Figure 6.8: Graph of lesion length (mm) over time for onion seedlings inoculated with the wild-type strains *P. ananatis* LMG20103 and LMG2665[T] and *anaD* mutant PANA24D10F. Vertical bars indicate 95% confidence intervals.

275

Figure 6.9: Onion seedlings inoculated with the *anaD* mutant PANA24D10F (C) and the wild-type pathogens *P. ananatis* LMG2665[T] (D) and LMG20103 (E). Plants inoculated with sterile distilled water were included as control (B). Differences in systemic disease could be observed between the wild-type and mutant strains (A), with inoculated plants corresponding to the lesions depicted directly below them.
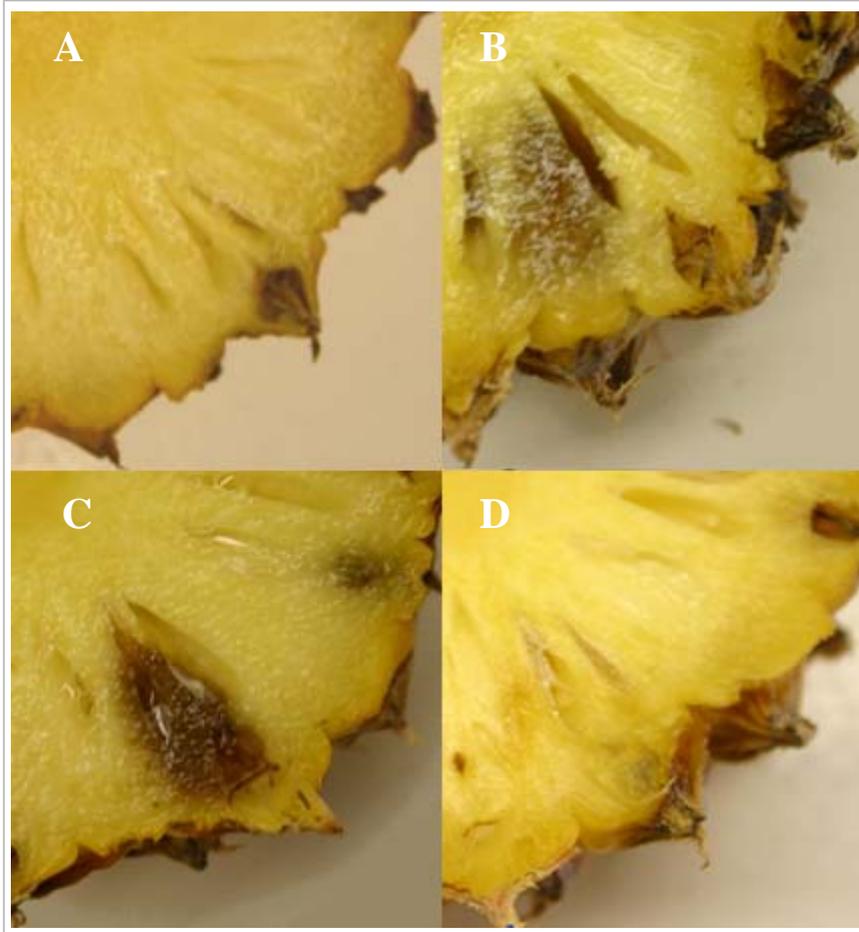
Figure 6.10: Pineapple fruits inoculated with the wild-type *Eucalyptus* pathogen *P. ananatis* LMG20103 (B), pineapple-pathogen LMG2665[T] (C) and *anaD* mutant PANA24D10F (D). Pineapples inoculated with sterile distilled water were included as control (A)

277

SUMMARY

Pantoea ananatis is a ubiquitous organism found in almost every environment on earth. It has been implicated in diseases of a wide range of agronomic crops worldwide, including onion, maize, rice and pineapple, as well as a human disease. In South Africa, *P. ananatis* causes blight and dieback of *Eucalyptus*, resulting in severe losses of this important forestry resource. Nevertheless, little is known about the pathogenicity mechanisms utilised by this pathogen to cause disease in this host.

The whole genome of a highly virulent *Eucalyptus*-pathogenic *P. ananatis* strain, LMG20103, was sequenced. This genome sequence was subsequently mined to identify a vast array of genes encoding putative pathogenicity determinants. Comparative genomics revealed that it has evolved to be able to thrive in a wide range of environments and that this strain carries pathogenicity determinants that may allow it to infect hosts in both the animal and plant Kingdom. Interestingly, no Type II and III secretion systems, which form a major part of the pathogenicity arsenal of many plant pathogenic bacteria are present in *P. ananatis*. However, three loci on the genome encode three distinct copies of the Type VI secretion system, which has recently been demonstrated to play an important role in diseases caused by many plant- and animal-pathogenic bacteria. *In silico* analysis of these secretion systems showed that they likely secrete several pathogenicity effectors which may have a role in *P. ananatis* infection of both plant and animal hosts. Another putative pathogenicity determinant identified from the genome, the exopolysaccharide ananatan, was experimentally demonstrated to play a role in disease expression on both onion seedlings and pineapple fruit. This was done through the production of a library of mutants which encompasses all the genes on the *P. ananatis* genome.

Genome sequencing enabled the identification of all the putative pathogenicity factors of *P. ananatis* LMG20103 and the use of the mutant library and post-genomic techniques has and will allow the functional characterization of many of these pathogenicity determinants. By this means, the mechanisms underlying the disease caused by *P. ananatis* on *Eucalyptus* and other hosts can be better understood. With this information, more directed and effective strategies for the control of this pathogen and its diseases can be developed.