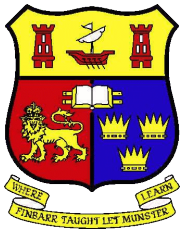


University College Cork

Coláiste na hOllscoile Corcaigh

An analysis of selected secretion systems of *Pseudomonas* species

A Thesis presented to the National University of Ireland

For the Degree of Doctor of Philosophy by

Michael Frank Egan BSc.

School of Microbiology

National University of Ireland,

Cork

Supervisor: Professor Fergal O'Gara

Head of Department: Professor Gerald Fitzgerald

September 2014

Dedicated to my Parents

“When people thought the Earth was flat, they were wrong. When people thought the Earth was spherical, they were wrong. But if you think that thinking the Earth is spherical is just as wrong as thinking the Earth is flat, then your view is wronger than both of them put together.”

-Isaac Amisov

Declaration

I, the undersigned, Frank Egan, hereby declare that this work has not been submitted for another degree, either in University College Cork or elsewhere and that the results presented in this thesis are derived from experiments carried out by myself in University College Cork, with the exception of the following:

Chapter 2

Promoter fusion assays with *Pseudomonas aeruginosa* T6SS genes were undertaken by Dr. Matthieu Barret using plasmid constructs created by Dr. Matthieu Barret.

P. aeruginosa internalisation assays were performed with the assistance of Emma Hennessy.

Chapter 4

Caenorhabditis elegans synchronisation and virulence assays were undertaken by Dr. Olivier Lesouthier University of Rouen, France.
Transcriptome analysis was undertaken by Dr. Matthieu Barret.

Frank Egan

September 2014

Publications:

Barret, M., **Egan, F**., Fargier, E., Morrissey, J.P., and O'Gara, F. (2011) Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: novel clusters and putative effectors uncovered. *Microbiology* **157**: 1726–1739.

Barret, M., **Egan, F**., Moynihan, J., Morrissey, J.P., Lesouhaitier, O., and O'Gara, F. (2013) Characterization of the SPI-1 and Rsp type three secretion systems in *Pseudomonas fluorescens* F113. *Environ. Microbiol. Rep.* **5**: 377–386.

Egan, F., Barret, M., and O'Gara, F. (2014) The SPI-1-like Type III secretion system: more roles than you think. *Front. Plant Sci.* **5**:

Egan, F., Reen, J.F, and O'Gara, F. (2013) Tle Distribution and Diversity in Metagenomic Datasets Reveals Niche Specialisation. *Environ. Microbiol. Rep.* (in press).

Presentations and Conference Contributions:

Egan, F. *et al*. Genomic analysis of the T6SS in *Pseudomonas* spp. Poster presentation, Society for General Microbiology Irish Division Conference IT, Tallaght, 2011.

Egan, F. *et al*. Genomic analysis of the T6SS in *Pseudomonas* spp. Poster presentation, Young Microbiologists Symposium UCC, 2012.

Egan, F. *et al*. Characterisation of the SPI-1 and Rsp type three secretion systems in *Pseudomonas fluorescens* F113. Oral presentation, Young Microbiologists Symposium UCC, 2012.

Egan, F. *et al*. Characterisation of the SPI-1 and Rsp type three secretion systems in *Pseudomonas fluorescens* F113. Poster presentation, Society for General Microbiology Annual Spring Conference, Manchester, 2013

Egan, F. An analysis of selected secretion systems of *Pseudomonas* species. Oral presentation, Postgraduate Student Seminar Series, UCC, 2013.

Table of Contents

Abstract

Direct secretion systems which deliver molecules from one cell to another have huge significance in shaping bacterial communities or in determining the outcome of bacterial associations with eukaryotic organisms. This work examines the roles of the Type III Secretion System (T3SS) and the Type VI Secretion System (T6SS) systems of *Pseudomonas,* a widespread genus including clinical pathogens and biocontrol strains.

Bioinformatic analysis of T6SS phylogeny and associated gene content within *Pseudomonas* identified several T6SS phylogenetic groups, and linked T6SS components VgrG and Hcp encoded outside of T6SS gene loci with their cognate T6SS phylogenetic groups. Remarkably, such "orphan" *vgrG* and *hcp* genes were found to occur in diverse, horizontally transferred, operons often containing putative T6SS accessory components and effectors.

The prevalence of a widespread superfamily of T6SS lipase effectors (Tle) was assessed in metagenomes from various environments. The abundance of the Tle superfamily and individual families varied between niches, suggesting there is niche specific selection and specialisation of Tle.

Experimental work also discovered that *P. fluorescens* F113 uses the SPI-1 T3SS to avoid amoeboid grazing in mixed populations. This finding may represent a significant aspect of F113 rhizocompetence, and the rhizocompetence of other Rhizobacteria.

Abbreviations

ABC	ATP-binding cassette transporters
APEC	Avian pathogenic *Escherichia coli*
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
CaCl2	Calcium chloride
cAMP	Cyclic adenosine monophosphate
COG	Clusters of orthologous genes
DGC	Diguanylate cyclase
DNA	Deoxyribonucleic acid
ETI	Effector-triggered immunity
FPI	*Francisella* Pathogenicity Island
GFP	Green Fluorescent Protein
GTP	Guanosine triphosphate
H2O	Water
HR	Hypersensitive response
Hrp	Hypersensitive response and pathogenicity
HSI	Hcp Secretion Island
IMG	Integrated Microbial Genomes
KH2PO4	Potassium phosphate
LB	Luria Bertani
MgSO4	Magnesium sulphate
NaCl	Sodium chloride
NaHPO4	Sodium phosphate
NGM	Nematode growth medium
NTP	Nucleoside triphosphate
ORF	Open read frame
PAS	Page's Amoeba Saline
PBS	Phosphate Buffered Saline
PDE	Phosphodiesterase
Pfam	Protein Family

PGPR	Plant Growth Promoting Rhizobacteria
PYG	Peptone Yeast Glucose
qRT-PCR	Quantitative real-time polymerase chain reaction
RLU	Relative light unites
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
Rsp	Rhizosphere secreted protein
SPI	*Salmonella* Pathogenicity Island
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
T3SS	Type III Secretion System
T4SS	Type IV Secretion System
T6SS	Type VI Secretion System
Tae	Type six amidase effectors
Tde	Type six DNase effector
Tge	Type six glycosidase hydrolase effector
Tle	Type six lipase effector
Tli	Type six lipase immunity
tRNA	Transfer RNA
Ysc	*Yersinia* secretion component

Units

aa:	Amino acid
bp:	Base pairs
cm:	Centimetre
M:	Molar
ml:	Millilitre
mm:	Millimetre
mM:	Millimolar
nm:	Nanometre
rpm:	Revolutions per minute
µl:	Microlitre
µm:	Micrometre
w/v:	Weight per volume

Figures

Chapter 1

General Introduction

1.1. Direct Secretion Systems

Life is dominated by prokaryotes, single celled organisms lacking membrane-enclosed organelles. Prokaryotes comprise two of the three domains of life: Archaea and Bacteria (Woese, 1987). Bacteria are fundamental part of the ecosystem, and are the most numerous of all biological domains, extending to all niches on Earth which can support life (Whitman *et al*., 1998). It is therefore unsurprising that they display a great degree of phylogenetic and metabolic diversity (Lozupone and Knight, 2007). Despite their microscopic size, they also represent one of the major categories of life by weight (Whitman *et al*., 1998). Needless to say, they have considerable influence over life on Earth. This influence even extends towards abiotic factors, as they have prominent roles in global biogeochemical cycles (Falkowski *et al*., 2008).

In contrast to prokaryotes, eukaryotes have are larger cells with membrane-enclosed organelles. Eukaryotes can be single celled or multicellular organisms. However, even among higher eukaryotes, the associated microbiome contains a staggering amount of bacteria. There are considerably more bacterial cells within the human gut than human cells (Sears, 2005). These are no passive partners. Indeed, in their varied metabolic capabilities underpin a tremendous amount of mutualistic relationships. Due to their metabolic diversity, bacteria can catabolise recalcitrant reactions, such as the degradation of lignocellulose in the termite gut and the conversion of minerals into bioavailable forms which can be used by host plants (Rodríguez and Fraga, 1999). Still other bacteria are pathogens, killing or causing disease in other higher eukaryotes.

Bacteria themselves are no unified group, and bacterial communities contain various phyla, which compete and co-operate with other members of that community, as well as eukaryote hosts. Unravelling the nature of these interactions and elucidating their mechanistic basis is of vital importance. Understanding how bacteria mediate these interacting networks are exerting their considerable impacts on our world is a prerequisite for harnessing beneficial effects and managing negative effects. Biological relationships, between and within species, are a complex web of signalling, of direct and indirect effects, of attack and defence, of mutualism and competition. Nowhere do the interactions become more direct and intimate than in

the case of those specialised secretion systems which can transport molecules directly from a synthesising cell into a target cell.

Such secretion systems are particularly frequent and best characterised in Proteobacteria, a widespread and influential bacterial phylum. This is best exemplified by the genus *Pseudomonas*, which colonises a variety of niches with and employs a variety of different lifestyle strategies. Moreover, direct secretion systems are highly abundant within many of its diverse species. Therefore, this genus is an ideal model to explore the mechanisms and functions of such secretions systems, and the impacts they may have in different niches.

1.2 *Pseudomonas*

Pseudomonas is a widespread genus of gram negative Proteobacteria. In total, nine species or species complexes have been described for this genus*: P. aeruginosa*, *P. stutzeri* and *P. oleovorans, P. fluorescens, P. syringae*, *P. lutea, P. putida*, *P. anguilliseptica* and *P. straminea* **(Figure 1.1)**. The genus displays a high degree of genetic and metabolic diversity. For example, the species *aeruginosa* within this genus has a core genome of approximately 4,000 genes, with 10,000 additional genes commonly found within the pan genome, and a further 100,000 genes which are rare or unique (Tummler *et al*., 2014). Given the conservation within the *P. aeruginosa* species, the pan genome of other species might be much larger (Silby *et al*., 2011) In addition to possessing several different species with may have large pan genomes, the genomes of many individual strains are large, allowing a high degree of plasticity which enables some species to thrive in a variety of niches (Silby *et al*., 2011). Accordingly, Pseudomonads can be found in many environments and impact on society in several ways **(Figure 1.2)**.

They can act as opportunistic human pathogens, especially the *P. aeruginosa* species, where they are most commonly associated with burn wounds the lungs of cystic fibrous and on solid surfaces such as contact lens and catheters (Butrus *et al*., 1987; Pier *et al*., 1997; Estahbanati *et al*., 2002). Less commonly, species such as *P. stutzeri*, *P. putida* and *P. fluorescens* have also been implicated in human pathogenesis (Noble and Overman, 1994; Yang *et al*., 1996; Hsueh *et al*., 1998).

Figure 1.1. Phylogeny of the *Pseudomonas* genus. Modified from (Duan *et al*., 2013)

Pseudomonads can also act as phytopathogens, with different pathovars of the *P. syringae* species being responsible for the destruction of billions of euro worth of various agricultural crops annually (Hirano and Upper, 1990). In contrast, members of the *P. fluorescens* species complex can have beneficial effects on plant health and crop yield, underscoring the diversity of roles this genus can play. These benefits include production of hormones to promote growth, solubilisation of minerals and biocontrol activities i.e. the antagonism of phytopathogenic organisms (Browne *et al*., 2009; Couillerot *et al*., 2009; Kochar *et al*., 2011). In addition, *Pseudomonas*, especially *P. putida* species, are of interest as agents of bioremediation (Nwachukwu, 2001; Nelson *et al*., 2002). As well as being involved in a plethora of disparate processes, *Pseudomonas* are also easy to culture and commonly found in nature, making them good model organisms for bacteriologists in many different fields.

As mentioned, certain Pseudomonads can make positive contributions to plant health, and are termed Plant Growth Promoting Rhizobacteria (PGPR). PGPR strains exert their beneficial influences in several ways, including direct or indirect protection against phytopathogens. Indirect protection occurs through niche exclusion, where pathogens cannot access the plant surface due to the PGPR colonisation, or from manipulation of the plant immune response in such a way that it is better able to resist pathogens (Chithrashree *et al*., 2011). Direct protection can result from the lethal or otherwise inhibitory effects of the biomolecules produced by the PGPR. Phytopathogens which have been reported to be antagonised by members of the *P. fluorescens* species complex include Chromista, insects and fungi (Radja Commare *et al*., 2002; Rezzonico *et al*., 2005; Daval *et al*., 2011).

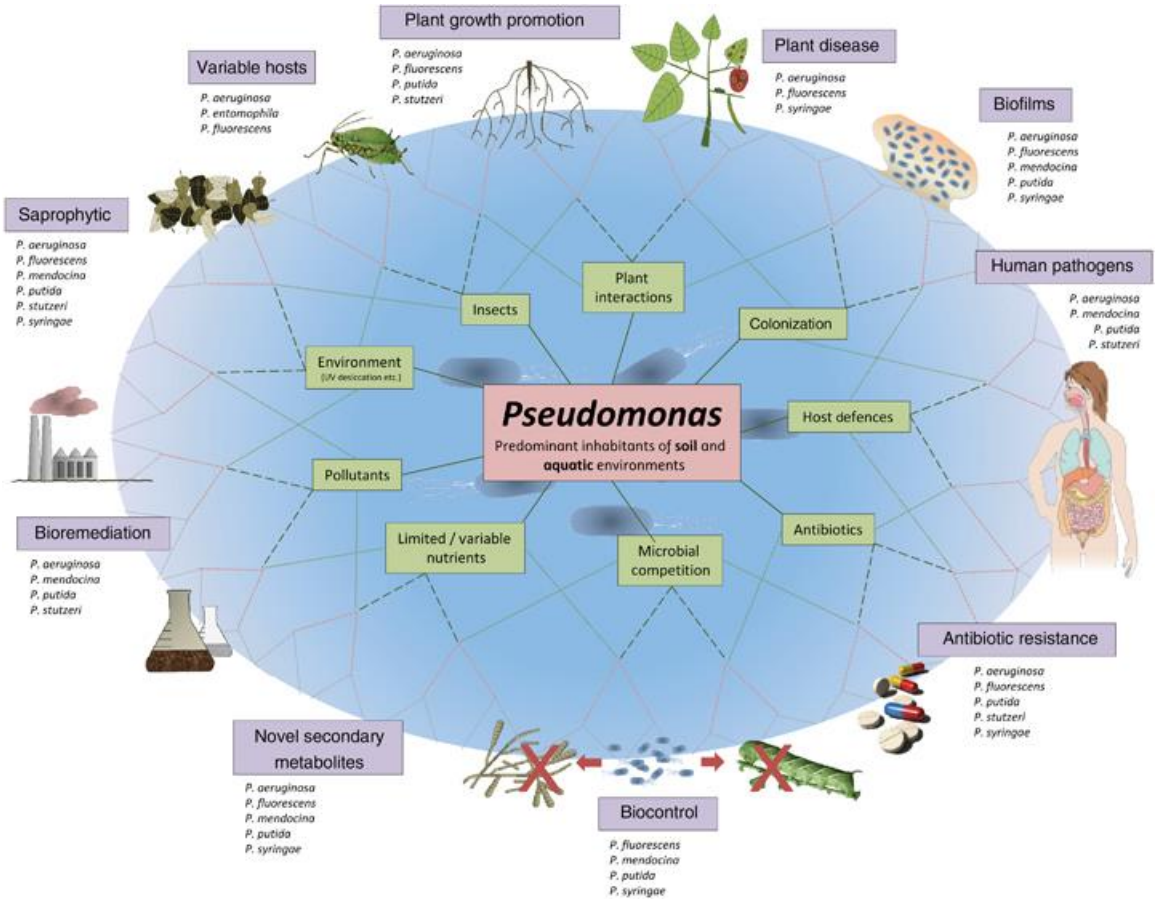


Figure 1.2. As a widespread genus occupying several different niches, *Pseudomonas* has myriad impact on human society. Figure reproduced from (Silby *et al*., 2011).

1.2.1 Model Organisms

One model organism studied in this thesis is the PGPR strain *P. fluorescens* F113, a bacterial isolate from sugarbeet (Shanahan *et al*., 1992). In addition to its ability to synthesise phytohormones and mobilise phosphate, it can antagonise several diverse pathogens, including oomycetes, nematodes, fungi and bacteria (Fenton *et al*., 1992; D. Cronin *et al*., 1997; D. Cronin *et al*., 1997; Barahona *et al*., 2011).

A second model organism in this work is *P. aeruginosa* PA14. *P. aeruginosa* is a species associated with opportunistic and nosocomial infection in humans, but also displays virulence towards a range of organisms including amoeba, insects, nematodes and plants (Lee *et al*., 2006). Much of these virulence mechanisms are known to be conserved between various hosts (Pukatzki *et al*., 2002). For example, biomolecules secreted by the T3SS are necessary for killing several hosts (Brannon

et al., 2009). For this reason various model hosts are considered useful means to identify virulence mechanisms in *P. aeruginosa* infection of humans. PA14 is a commonly studied and particularly virulent strain of *P. aeruginosa*, which possesses pathogenicity islands lacking in other *P. aeruginosa* strains and a non-functional copy of the *ladS* gene, which is known to be involved in virulence regulation (Lee *et al*., 2006; Mikkelsen *et al*., 2011). Like *P. fluorescens* F113, PA14 has several secretion systems encoded in its genome, and these have been shown to be important in various inter-kingdom interactions.

1.2.2 Secretion systems of *Pseudomonas*

Bacterial biomolecules which are used in interactions with other organisms are synthesised in the interior of the bacterial cell, but act outside it. While many molecules are freely diffusible, others require active transport. Some molecules, including virulence factors, are transported via membrane vesicles released from the cell or are released into the general environment by cell lysis (Kadurugamuwa and Beveridge, 1995; Hassett *et al*., 2000). Two secretion pathways exist for the transport of molecules across the cellular membrane that are common to all domains of life. These are the general secretion pathway (Sec) and the Twin Arginine Translocation pathway (Tat) (Natale *et al*., 2008). However, the dual membranes of gram negative bacteria mean that such substrates would only be exported into the periplasm. To facilitate secretion into the extracellular environment, gram negative bacteria have evolved specialised secretion systems (Type I-VI). Though all secretion systems can secrete molecules into the environmental milieu, only the Type III Secretion System (T3SS), Type IV Secretion System (T4SS) and the Type VI Secretion System (T6SS) can be used to secrete directly into target cells **(Figure 1.3)** (Backert and Meyer, 2006; Bleves *et al*., 2010). Such secretion systems are the focus of this thesis. The T4SS is absent from the *Pseudomonas* genomes sequenced to date. However, some do encode conjugation machinery known as the Type IV Pilus which is related to the T4SS (Carter *et al*., 2010; Filloux, 2011). Therefore, this work focuses only on the T3SS and T6SS.

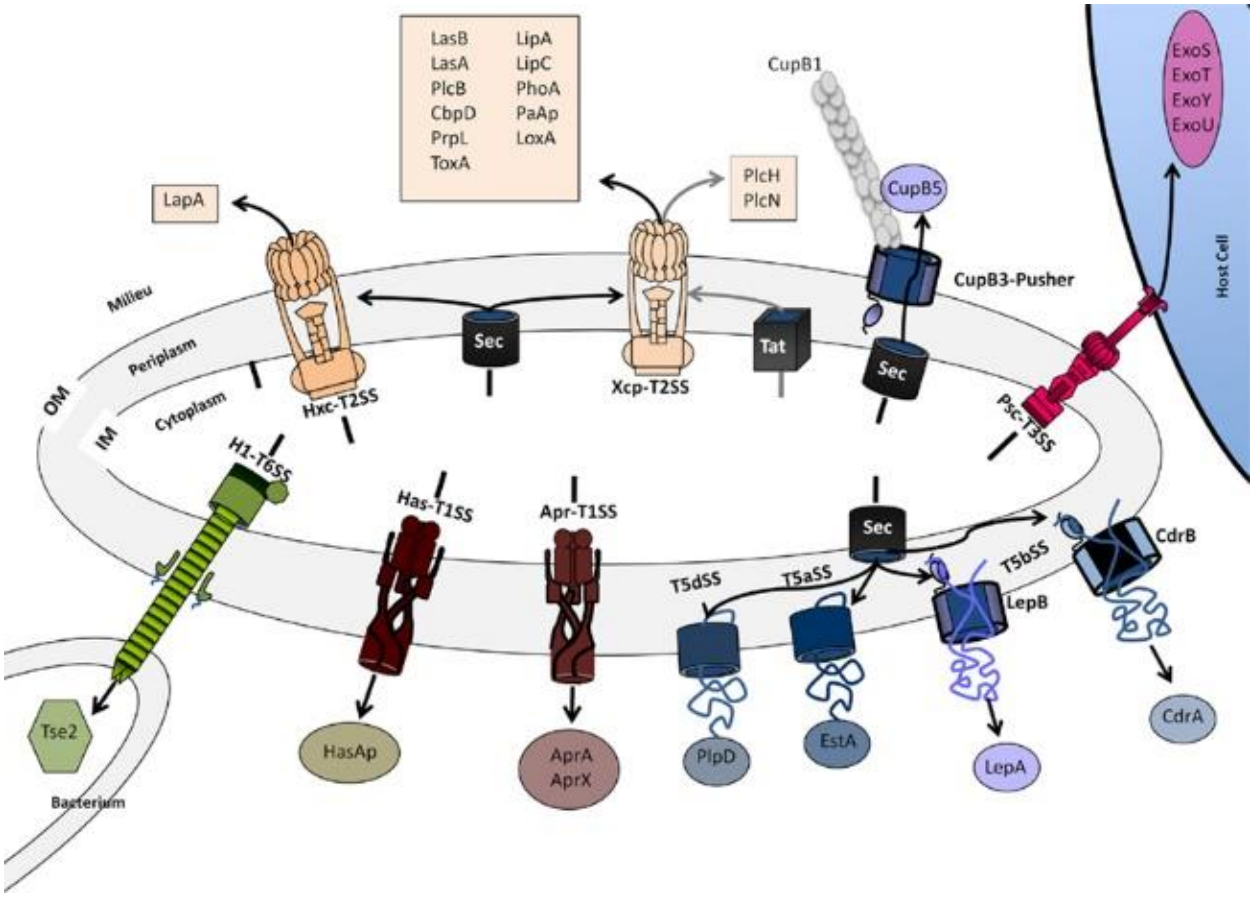


Figure 1.3. *P. aeruginosa* possesses all specialised secretion systems Type1-Type VI except for Type IV. Of these only T3SS and T6SS can facilitate secretion directly into a target cell. Figure reproduced from (Filloux, 2011)**.**

2.1 The T3SS

The involvement of a genetic locus encoded in *Yersinia* species in the secretion of Yop proteins (*Yersinia* outer proteins) in a Sec-independent manner heralded the discovery of a new protein secretion system, which was later coined the T3SS (Michiels *et al*., 1991; Salmond and Reeves, 1993).

The term T3SS is often used as an umbrella term to describe both the non-flagellar T3SS and the flagella as both seem to share an evolutionary history, with the non-flagella T3SS probably being a product of flagella expatiation (Abby and Rocha, 2012). In this text the term T3SS refers exclusively to non-flagellar T3SS, unless otherwise stated. Non-flagellar type III secretion systems are multimeric protein complexes for the secretion of effector molecules into eukaryotic host cells.

2.2 T3SS Structure

T3SS assembly and function requires approximately 20 proteins, though the structure itself is composed of hundreds of multiple copies of these proteins which often assemble as homooligomers **(Figure 1.4)**. The T3SS resembles a macromolecular syringe which can be divided schematically into several parts: a cytoplasmic protein sub-complex which facilitates recognition of T3SS substrates, a basal structure spanning the dual membranes of the bacteria which is necessary for export and anchors the secretion apparatus, and a needle or pilus protruding outwards to contact target cells. The needle or pilus is thought to serve as a conduit for secreted effectors and translocator proteins which form pores in the host membrane and are necessary for effector delivery. These translocators are often considered part of the T3SS apparatus.

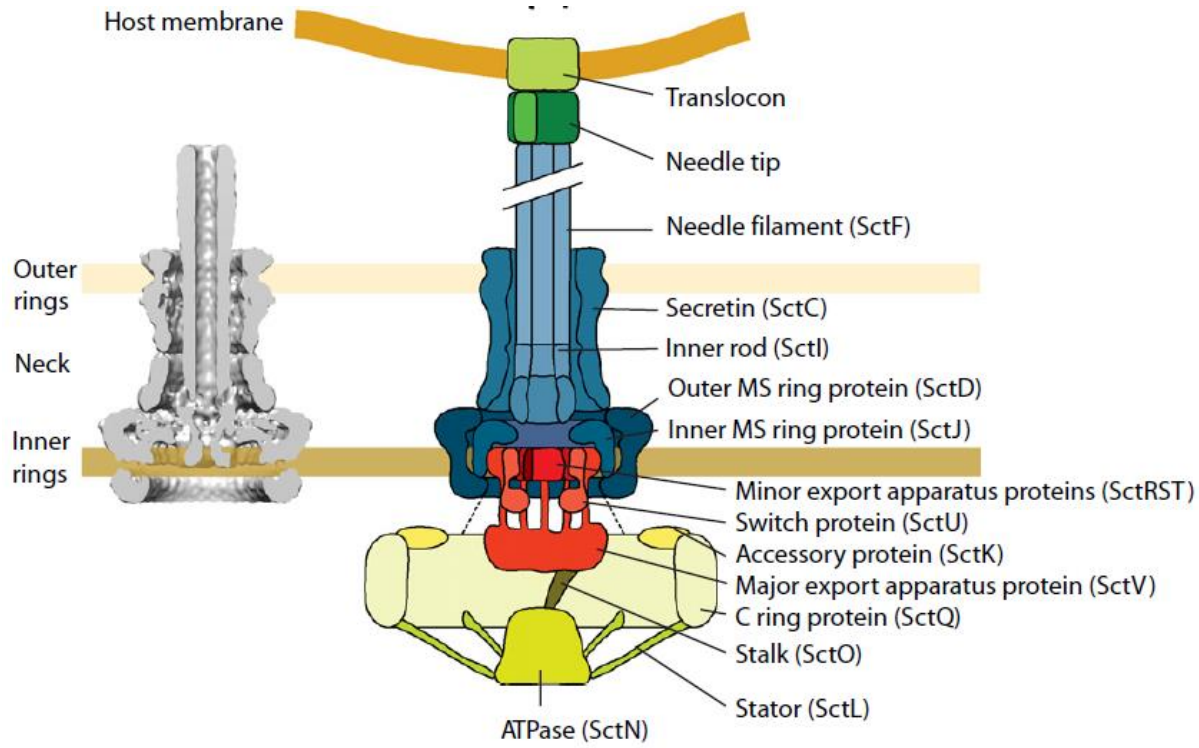


Figure 1.4. Structure of the T3SS. A structural representation of the needle complex from 3D modelling is shown on the left. A labelled diagram of the T3SS and its constituent components is shown on the right. Figure modified from (Diepold and Wagner, 2014).

When understanding the T3SS structure it is important to be aware that certain differences exist between different T3SS, some of which are due to the existence of several distinct phylogenetic families. Multiple phylogenetic analyses based on proteins involved in the assembly of the needle complex have split the T3SSs into 7 distinct families: SPI-1 (also known as the Inv-Mxi-Spa family), SPI-2, Hrp1, Hrp2, Ysc, Rhizobiales (Rhc) and Chlamydiales (Pallen *et al*., 2005; Troisfontaines and Cornelis, 2005; Abby and Rocha, 2012).

From initial characterisation and genomic distribution, the Ysc, Chlamydiales, SPI-1 and SPI-2 families were associated with animal-bacterial interactions while the Rhizobiales, Hrp1 and Hrp2 families were associated with plant-bacterial interactions (Abby and Rocha, 2012).

Homologues of T3SS proteins from different bacteria and belonging to different families are named inconsistently. Hence a unifying nomenclature has been proposed (Hueck, 1998). In this nomenclature proteins are given the prefix Sct (secretion and cellular translocation) and the suffix corresponds to that of the Ysc proteins of the

T3SS in *Yersinia* e.g. SpaS of the SPI-1 T3SS is homologous to YscU of the Ysc T3SS and is therefore described as SctU.

As explained below there are some differences between T3SS belonging to these different families, in terms of structure, assembly and roles. However, nine of the core components are highly conserved across all T3SS: SctC, SctJ, SctN, and SctQ-V. It is these proteins which are used in phylogenetic analyses of the T3SS (Troisfontaines and Cornelis, 2005; Abby and Rocha, 2012). In addition to having these highly conserved proteins, several components which do not appear to be homologous or exhibit low sequence identity appear to fulfil similar functions (Diepold and Wagner, 2014). The overall commonality of T3SS families is probably best demonstrated by the phenomenon of promiscuous secretion; that is, there are multiple reports of effector secretion via non-cognate T3SS families, including effectors which are used to target animals being heterologously expressed and secreted via phytopathogenic T3SSs and vice versa (Anderson *et al*., 1999; Subtil *et al*., 2001).

In the inner membrane an outer ring of SctD encloses an inner ring composed of the lipoprotein SctJ which is anchored in the membrane by a single transmembrane domain (Schraidt *et al*., 2010; Schraidt and Marlovits, 2011). Each of these inner rings have a 24-fold symmetry (Schraidt and Marlovits, 2011). 12-15 copies of SctC form a homo-oligomeric ring of diameter 20 nm which is embedded in the outer membrane (Koster *et al*., 1997; Kubori *et al*., 1998; Spreter *et al*., 2009; Schraidt and Marlovits, 2011; Bergeron *et al*., 2013). The *sctC* gene is split into two ORFs in rhizobial T3SS, seemingly due to acquisition of a second copy of this gene followed by loss of some of the coding region of the original gene (Abby and Rocha, 2012). In limited cases the second ORF itself has apparently been fragmented into two separate ORFs (Gazi *et al*., 2012).

The three separate modular rings form due to the presence of a ring binding motif on each of the subunits (Bergeron *et al*., 2013). The inner and outer membrane rings are proposed to join by the binding of SctC by the C-terminal of SctD, as SctC co-precipitates with this protein fragment (Ross and Plano, 2011). The distance to which the SctD and SctC rings extend into the periplasm can vary, which is proposed

to facilitate temporal variation in the distances between the inner and outer membranes (Kudryashev *et al*., 2013).

In addition to the membrane-bound rings, the basal body includes an export apparatus composed of three minor export proteins (SctR-T) and two major export proteins (SctU and SctV) (Diepold and Wagner, 2014). This export apparatus is enclosed by the inner membrane rings (Wagner *et al*., 2010). The minor proteins are mainly located in the membrane and periplasm, while the major proteins are found in the membrane but their C-terminal domains extend into the cytoplasm (Allaoui *et al*., 1994; Berger *et al*., 2010).

The two major proteins of the export apparatus are best characterised. The cytoplasmic portion of SctV assembles into a ring *in vivo*, through which secreted substrates are expected to pass (Abrusci *et al*., 2013). The SctU protein is termed a switch protein and is involved in the switching between secretion of needle components and translocators (Zarivach *et al*., 2008).

At the cytoplasmic surface of the basal body five T3SS cytoplasmic proteins are recruited: SctN (ATPase), SctL (stator) SctO (stalk) SctQ (C ring) and SctK (accessory protein). T3SS substrates are delivered to the T3SS by chaperones. Chaperones are divided into three classes based on their substrates: Class 1 chaperones bind to effectors, Class 2 chaperones bind translocators and Class3 chaperones bind needle subunits (Burkinshaw and Strynadka, 2014). Class 1 chaperones are further sub-divided based on whether they associate with just one specific cognate effector (Class 1A), or several different effectors (Class 1B). Class 1 chaperones bind via a conserved chaperone binding domain located in the N terminal of the effector proteins (Lilic *et al*., 2006). Class 2 chaperones recognise the translocators through a helical tetratricopeptide domain (Job *et al*., 2010). This domain is also used for binding between needle subunits and one of the chaperones in the Class 3 chaperone heterodimer, with other chaperone in the heterodimer being hypothesised to stabilise the needle-binding chaperone itself. (Quinaud *et al*., 2007; Chatterjee *et al*., 2011).

SctN dissociates these proteins from their cognate chaperones and may unfold them or maintain them in an unfolded state prior to secretion (Akeda and Galán, 2005).

ATPase activity is repressed by SctL (Blaylock *et al*., 2006). SctO is involved in the recognition of the chaperones of translocators (Evans and Hughes, 2009).

The C ring is formed by the SctQ protein (Morita-Ishihara *et al*., 2006). Two proteins make up the C ring in both flagella and T3SS, but in the latter these two proteins are both encoded by a single *sctQ* gene which has an internal promoter (Bzymek *et al*., 2012).

A complex of SctC, SctL and SctK detected in *Salmonella* is termed a sorting complex as it associates with chaperones and enforces a secretion hierarchy (Lara-Tejero *et al*., 2011). Preferential binding of chaperones of early secretion substrates by the sorting complex is hypothesised to be the mechanism by which this hierarchy is enforced. It is not clear whether this finding is transposable to other T3SS.

Once the basal body and cytoplasmic components are in place, the inner rod and needle can be assembled. This needle is composed of helically arranged SctF proteins, with the conserved C-terminal of the SctF subunits pointing inwards (Loquet *et al*., 2012; Demers *et al*., 2013).

The inner rod composed of SctI subunits is enclosed by the basal body (Schraidt and Marlovits, 2011). SctI interacts with SctF and possibly acts to stabilise the needle, as well as being involved in the assembly of the apparatus as discussed below (Yang *et al*., 2013). The association between the needle and the basal body is also likely mediated by SctD and SctJ which have been shown to interact with SctF (Ogino *et al*., 2006).

The needle has an internal diameter of 3 nm which is too small to accommodate most folded T3SS effectors (Marlovits *et al*., 2004). The evidence is that effectors are secreted in an unfolded state (Radics *et al*., 2014). Fusing a stable GFP domain, which resists unfolding, to the C-terminal of three different effectors from *S. enterica* serovar Typhimurium abrogated secretion of all of these substrates. While fusion of ubiquitin to the T3SS-targetting N terminal domain of SptP also abrogated secretion, a similar construct where two residues crucial for ubiquitin stability were mutated was secreted.

While the different T3SS phylogenetic families have differences in T3SS structural proteins, they mainly differ in their translocators and extracellular appendages. This

is presumably because of the greater challenge of delivering effectors across the additional barrier of the cell wall, which can be reinforced through callose deposition at sites of infection (Luna *et al*., 2010) of the short (~60 nm), apparently ancestral needle, the Hrp and Rhc T3SS have long (~1-2 um) flexible pili presumed to be an adaptation to the thickness of the plant cell wall (Cornelis, 2006; Krishnan *et al*., 2011; Abby and Rocha, 2012). The *m*ajor subunits of the pilus are called pilins, and encoded by the *hrpA* gene in the Hrp T3SS *(*Roine *et al., 1997).* Rhc pili are composed of NopA, NopB and a translocator-related NopX. NopA is the major component by protein levels, but the function NopB is unknown, though it is necessary for secretion and co-precipitates with NopA *(*Saad *et al., 2008).* This pilus certainly is capable of crossing the cell wall (Brown *et al*., 2001).

At the tip of the needle in animal-associated T3SS families is a pentameric complex of LcrV subunits, though pentamers containing 4 LrcV subunits and a translocator have been reported (Mueller *et al*., 2005; Broz *et al*., 2007; Veenendaal *et al*., 2007). LcrV binds to the needle via its C terminus (Espina *et al*., 2006). LcrV oligomers form a ring with a 3-4nm internal diameter (Caroline *et al*., 2008). Two hydrophilic translocators are later recruited to the tip, possibly in a stepwise fashion with YopB recruitment preceding YopD recruitment as has been reported for *Shigella* (Olive *et al*., 2007; Epler *et al*., 2009).

These translocators are necessary for effector translocation into host cells, and act to form a membrane-embedded complex resulting in a pore (Montagner *et al*., 2011). Indeed, YopB homologues have structural homology to pore forming toxins such as colicin (Barta *et al*., 2012). LrcV is also required for pore formation (Mota, 2006).

Translocation may occur mainly at lipid rafts (Schoehn *et al*., 2003) This is possibly due to the elevated levels of cholesterol found in lipid rafts as translocators have been shown to bind cholesterol, and depleting membrane cholesterol can render target cells resistant to T3SS-mediated invasion (Hayward *et al*., 2005; Bridge *et al*., 2010)

The phytopathogenic translocator repertoire is also more complex and variable than that of animal pathogens. Proteins with weak homology to the known translocators can be found in plant-associated T3SS (Meyer *et al*., 2006). These proteins probably fulfil the same role, as they have been shown to be required for virulence but not

secretion of effectors into the medium. HrpF can form pores in an artificial membrane (Büttner *et al*., 2002). However, the other proteins appear to play a role in the translocation process. Hpa2 is a lytic transgylcolase interacts with HrpF and the host membrane (Li *et al*., 2011). Single *hrpF* or *hpa2*, but not a *hrpF-hpa2* double mutant, could elicit HR in Rice, suggesting a level of redundancy.

In addition, small, heat stable proteins known as harpins, which are exclusive to and possibly ubiquitous in, bacteria possessing the Hrp T3SS, have been implicated in the translocation process (Kvitko *et al*., 2007; Choi *et al*., 2013). Harpin HrpZ from *P. syringae* forms a pore in artificial membranes, possibly dependent on its interaction with membrane component phosphatidic acid (Lee *et al*., 2001; Haapalainen *et al*., 2011). Harpin HrpW binds to calcium pectate which is found in the cell wall of plants (Charkowski *et al*., 1998) and the HrpW homologue in *Lonsdalea quercina* contains an extension with pectate lyase activity (Yang *et al*., 2014).

2.3 Assembly of the T3SS

Assembly of the T3SS is an ordered process which proceeds in several stages **(Figure 1.5)**. The basal body and export apparatus are formed first and these play a crucial role in assembly of the inner rod and needle, which are exported via the nascent T3SS structure. This is followed by secretion of translocators and finally, the effectors themselves.

The formation of the basal body appears to vary in *Salmonella* and *Yersinia*. In *Salmonella*, assembly is said to be "inside out", as the minor export apparatus proteins localise to the inner membrane prior to the formation of the membrane rings, which then appear to grow outwards (Wagner *et al*., 2010). Conversely, in *Yersinia* the process is termed "outside in" as the outer membrane ring of SctC is assembled prior to the inner membrane rings, while the formation of the export apparatus and membrane rings are independent (Diepold *et al*., 2010, 2011). The membrane rings are composed of homeric rings of SctD, SctJ and SctC and all of these proteins are needed for full stability of the rings complex and basal structure (Sukhan *et al*., 2001). However, low amounts of these rings can form in certain mutants. In *Yersinia* SctC, but not SctJ, is strictly required for the formation of the

SctD ring (Diepold *et al*., 2010). In contrast, SctJ, but not SctC, is strictly required for the formation of the SctD ring in *Salmonella* (Sukhan *et al*., 2001). Therefore, it seems likely that assembly of the membrane rings in *Salmonella* proceeds in a SctJ, SctD, SctC fashion, while the order is reversed in *Yersinia*.

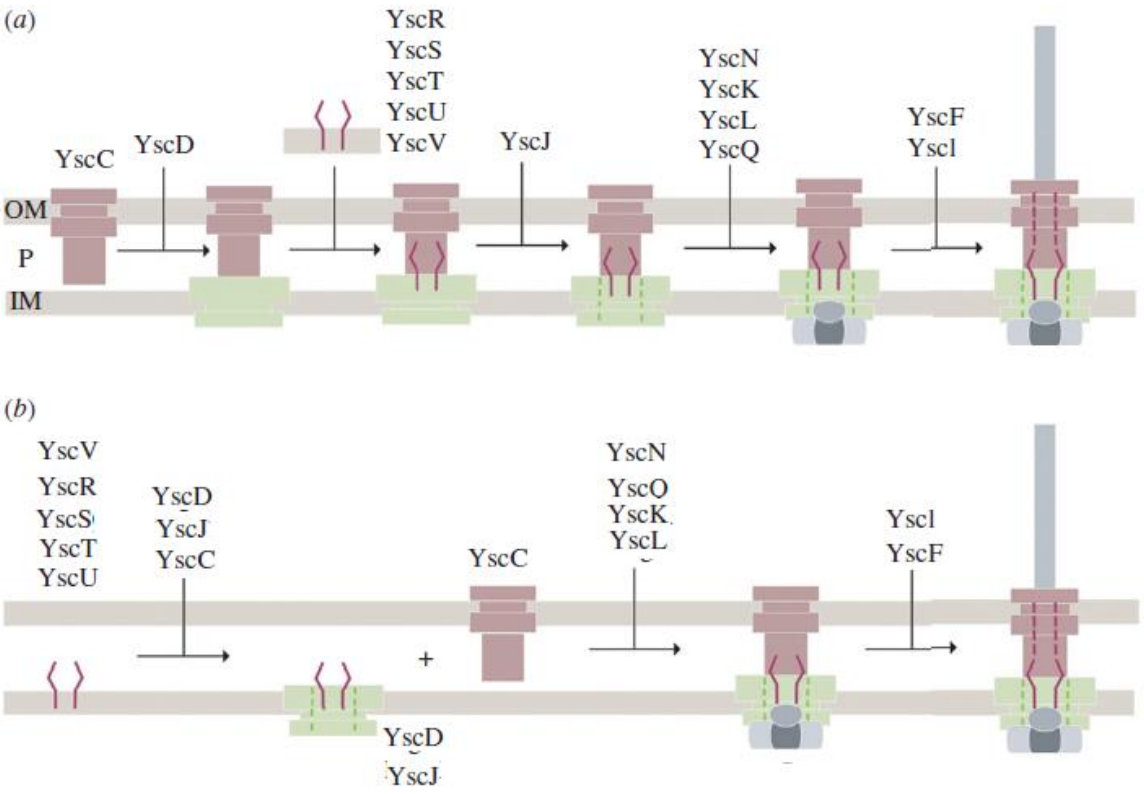


Figure 1.5. T3SS assembly proceeds in an outside-in manner in *Yersinia* **(a)** and an inside out manner in *Salmonella* **(b).** Figure modified from (Kosarewicz *et al*., 2012)

The secretin SctC, which forms the outer membrane ring, is transported through the inner membrane by the Lol transport system. This requires the use of another T3SS protein called pilotin which binds to SctC and is the substrate for the Lol transport system (Koo *et al*., 2012). SctC is a homologue of a Type II Secretion System protein PulD, which also requires a dedicated pilotin protein to access the outer membrane (Collin *et al*., 2011). However, several T3SS lack pilotins and may have alternative mechanisms for crossing the inner membrane. In particular, pilotins have not been reported in plant pathogenic T3SS (Büttner, 2012). The variation in secretins or the transport process is presumably responsible for the varying effects on secretin levels when pilotin is absent. In *Salmonella*, decreased amounts of secretin were present in the cell when the pilotin protein was absent but the opposite was true for *Yersinia* (Crago and Koronakis, 1998; Burghout *et al*., 2004).

Lytic enzymes are often required for local degradation of the peptidoglycan to facilitate insertion of large supramolecular structures including other secretion systems (Dijkstra and Keck, 1996; Koraimann, 2003). Indeed, such lytic genes are often found in the T3SS-encoding genetic loci (Zahrl *et al*., 2005). Deletion of such a gene from a T3SS in *Citrobacter rodentium* resulted in decreased T3SS assembly (Deng *et al*., 2004). The lytic transglycosylase from a *Salmonella* T3SS operon was shown to degrade peptidoglycan when not inhibited by an associated chaperonin, suggesting this process is localised and controlled by other accessory components (Zahrl *et al*., 2005).

Once the basal body is formed, cytosolic components are recruited. Though protein-protein interactions have been reported between various cytoplasmic T3SS components, SctQ and SctK are probably most essential for cytoplasmic complex formation and localisation (Diepold and Wagner, 2014). SctK is necessary for the localisation of the C ring to the basal body, though the N-terminus of SctD has a cytoplasmic FHA domain which binds phosphorylated SctQ (Diepold *et al*., 2010; Barison *et al*., 2012). Possibly this interaction with SctD is not necessary for localisation but for secretion. The C ring also interacts with the stator and ATPase (Jouihri *et al*., 2003; Morita-Ishihara *et al*., 2006). Once the basal body and cytoplasmic proteins are in place, the inner rod and needle subunits can be exported.

When the needle reaches an appropriate length, the SctP ruler/tape measure protein interacts with SctU, ultimately halting the export of the needle subunits and beginning the secretion of the tip and translocator proteins (Journet *et al*., 2003; Marlovits *et al*., 2006; Botteaux *et al*., 2008). SctU, and its homologues from flagella, have a conserved NPTH motif (Ferris *et al*., 2005). Auto-catalytic cleavage of SctU at the asparagine residue of this motif results in a fragment of the SctU cytoplasmic domain being secreted by the T3SS (Zarivach *et al*., 2008; Frost *et al*., 2012). Mutants expressing SctU protein which is not cleaved cannot secrete translocators or the tip protein, but are still technically capable of secreting genuine T3SS effectors (Sorg *et al*., 2007). This may be because the cleaved domain had been blocking the access of the later substrates to the secretion channel (Frost *et al*., 2012). SctO interacts with SctV and cleaved SctU, suggesting it also functions in the substrate specificity switching process (Cherradi *et al*., 2014).

The needle tip and translocators are next to be secreted. In *Shigella* this has been shown to be have its own hierarchy. The needle tip protein is secreted, and appears to sense bile salts, resulting in the recruitment of the translocator YopB (Olive *et al*., 2007). YopB recruits the final translocator YopD upon detection of host cell lipids (Epler *et al*., 2009). Once the translocators form a pore in the host membrane, effectors can enter the target cell.

The signal targeting the effectors for secretion is an ongoing area of research as there appears to be a high degree of diversity and nuance. It has been shown in several cases that the N-terminal amino acids contain the secretion signal (Michiels *et al*., 1991; Sory *et al*., 1995). However, C-terminal signals have also been recorded (Allen-Vercoe *et al*., 2005). Though secretion of effectors may involve recognition by chaperones, not all secreted effectors appear to have chaperones (Anderson *et al*., 2002). While SipB has an essential N-terminal signal sequence, the chaperone binding site is several residues downstream of it, and a C-terminal region is also required for secretion (Kim *et al*., 2007). In other cases it appears that the secretion signal is contained in the untranslated region of mRNA (Niemann *et al*., 2013).

While no consensus T3SS secretion signal has been detected within the N-terminal regions of effectors certain properties are common in these regions such as increased serine prevalence (Wang *et al*., 2011). It has been suggested that disordered regions

may be important, possibly because the structural flexibility allows for interactions with a range of T3SS structural proteins (Buchko *et al*., 2010). A more comprehensive recent analysis suggested that the amino acid composition, water accessibility state and secondary structure are all involved in targeting substrates for export in a co-dependent manner (Wang *et al*., 2013).

2.4 Roles and Effectors of the T3SS

Though restricted to interactions with eukaryotes, interactions mediated by T3SS run the gamut from virulence to symbiosis (Preston, 2007). Therefore, it is sufficient to explore these roles in *Pseudomonas*, as the interactions mediated by the T3SS in this genus are commensurate with the diversity among its myriad constituents. The best studied are of *P. aeruginosa* which uses its Ysc-like T3SS in the pathogenesis of several hosts, including amoeba, mammals, wax moth and zebrafish (Miyata *et al*., 2003; Matz *et al*., 2008; Brannon *et al*., 2009; Galle *et al*., 2012). The T3SS dependent activity against phagocytic cells is likely a key mechanism underpinning these various virulence phenotypes because T3SS is not involved in hosts that lack such cells such as plants and worms (Brannon *et al*., 2009).

Although extremely diverse, described type III effectors are composed of less than 40 motifs or domains which can often interfere with conserved eukaryotic cellular processes such as regulation via G proteins **(Figure 1.6)** (Dean, 2011). Many effectors seem to be the products of gene fusion events resulting in bi-functional proteins with unrelated domains (Stavrinides *et al*., 2006).

Four effectors are found in *P. aeruginosa*: ExoS, ExoT, ExoU and ExoY. ExoU is an acutely toxic phospholipase which disrupts cellular membranes (Sato and Frank, 2004). ExoY is an adenylate cyclase which are enzymes that convert ATP to cAMP (Yahr *et al*., 1998). The adenylate cyclase activity is essential for cell-rounding mediated by ExoY, but not for some level of disruption of the cytoskeleton (Vallis *et al*., 1999; Cowell *et al*., 2005).

ExoS and ExoT each have ADP-ribosylation and GTPase activating domains, and both deter invasion (Cowell *et al*., 2003). Their GTPase activity is targeted against the proteins Rho, Rac and Cdc42 which are Rho GTPases that regulate the actin

cytoskeleton (Goehring *et al*., 1999; Kazmierczak and Engel, 2002). However, the targets of the ADP-ribosylation domains of ExoS and ExoT are different. Rab proteins regulate vesicle traffic, and it is this process which is disrupted through the ADP-ribosylation of Rab5 by ExoS (van der Bliek, 2005; Deng and Barbieri, 2008). ExoT ADP-ribosylates CT10 regulator of kinase proteins which are involved in regulation of phagocytosis (Sun and Barbieri, 2003).

The *P. syringae* species contains many phytopathogens which use T3SS for infection of plants. Unlike the *P. aeruginosa* type III effector repertoire, the amount of effectors in *P. syringae* is high, with 57 families of effectors reported in the pan genome of this species in 2012 (Lindeberg *et al*., 2012). The reason for this high variety of effectors is that the relationship between plant and phytopathogen is effectively an arms race between the plant evolving to recognise effectors and trigger ETI or HR, and the pathogen evolving effectors which can avoid recognition or inhibit the subsequent steps in the plant's response (Espinosa *et al*., 2003; Göhre and Robatzek, 2008; Cunnac *et al*., 2009; Guo *et al*., 2009; Lindeberg *et al*., 2012).

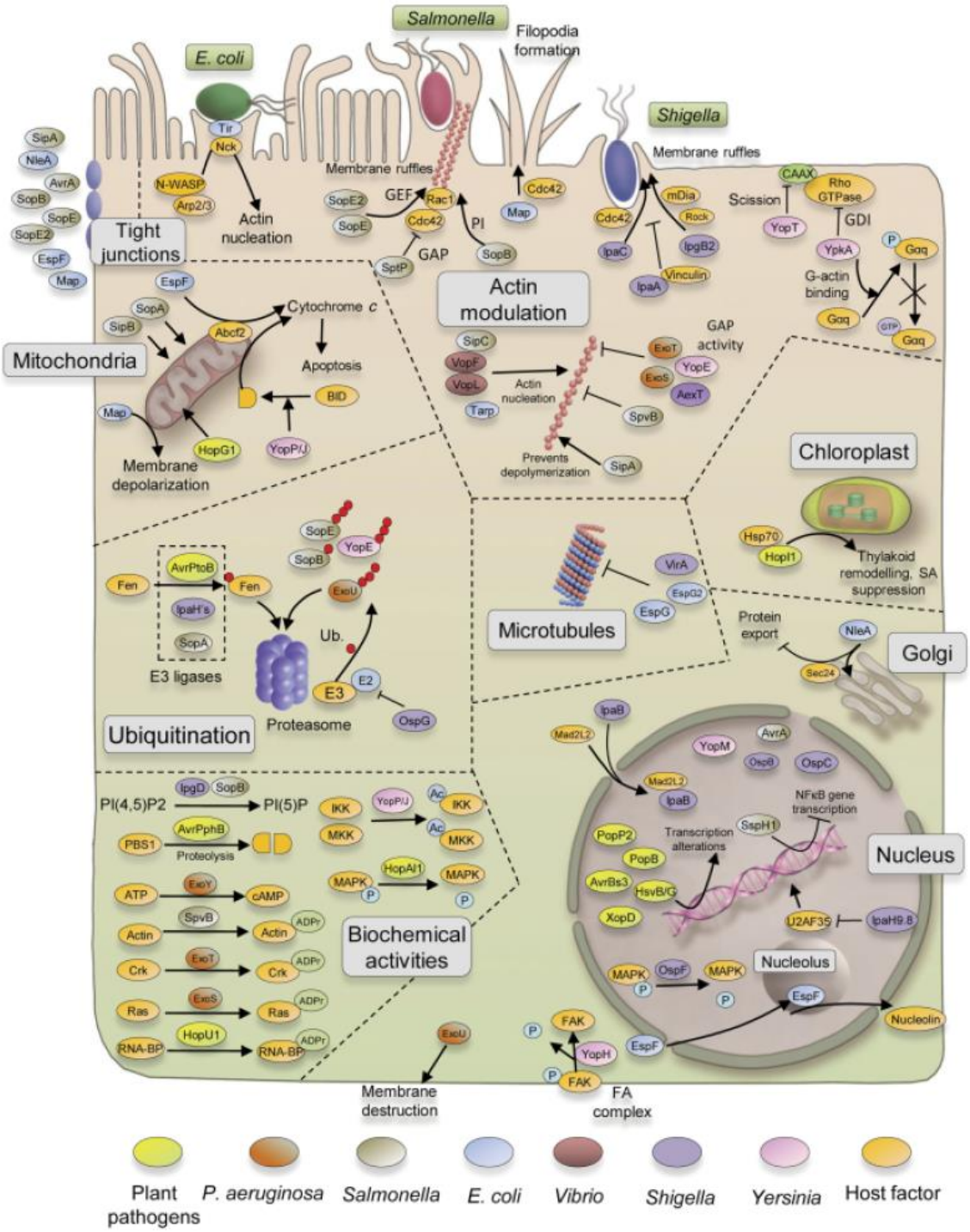


Figure 1.6. Diverse T3SS effectors have conserved targets and mechanisms of actions. These can include GTPase activity to interfere with modulation of host processes by *P. aeruginosa* effectors. Phytopathogenic T3SS effectors are often targeted to the nucleus where they influence gene regulation. Figure reproduced from (Dean, 2011).

The *P. fluorescens* are often associated with plants roots and biocontrol. The ectomycorrizal activity of *P. fluorescens* BBc6R8 is dependent on its T3SS (Cusano *et al*., 2011). *P. fluorescens* KD has a Hrc1-type T3SS which does not inhibit the growth of *Phytium ultimum* but does decrease the levels of the virulence related pectinase polygalacturonase produced by this pathogen (Rezzonico *et al*., 2005). *P. fluorescens* MFM1032 uses a T3SS which probably belongs to the Hrp family to lyse macrophages and interfere with the growth of amoeba (Sperandio *et al*., 2012). As both the KD and MFM1032 strains encode T3SS belonging to Hrp1, these are interesting examples of a role for this T3SS beyond plant-microbial interactions. The fungal wall may present similar challenges to the plant cell wall, so it is perhaps not so surprising that Hrp T3SS can be involved in such interactions. Indeed *Burkholderia rhizoxinica* interacts with a fungus using this system (Lackner *et al*., 2011). However, the interaction between macrophage and MFM1032 mediated by this T3SS is a challenge to the paradigm that certain families are plant or animal associated (Büttner, 2012).

2.5 Regulation of T3SS

Naturally, T3SS are expressed in response to environmental signals transduced by the cell. For example, in *P. aeruginosa*, the T3SS is regulated through the action of both the RetS/GacS signalling cascade and the concentrations of the second messenger cyclic di-GMP (Moscoso *et al*., 2011). A common feature of T3SS regulation is the activation of expression by a protein encoded within the T3SS locus itself.

Two transcriptional regulator are encoded within the SPI-1 T3SS locus: HilA and InvF (Khajanchi *et al*., 2009). Both these proteins positively affect transcription of T3SS. HilA up-regulates the expression of T3SS structural genes and *invF*, and InvF up-regulates expression of effectors, allowing staggered expression of genes according to when the proteins are needed (Darwin and Miller, 1999). Expression of some effectors is directly regulated by both HilA and InvF, suggesting fine tuning of effector expression may be achieved by cooperative efforts of both transcriptional regulators (Thijs *et al*., 2007)

The transcriptional regulator of T3SS in *P. aeruginosa*, ExsA, upregulates the transcription of T3SS genes when free to do so (Frank and Iglewski, 1991). ExsA availability is controlled by a series of protein-protein interactions. ExsD acts an anti-activator of T3SS gene expression by binding ExsA (McCaw *et al*., 2002), ExsC binds to ExsD to prevent it binding to ExsA (Dasgupta *et al*., 2004), and ExsE binds to ExsC to prevent it binding to ExsD (Rietsch *et al*., 2005). Transcription is induced when ExsE is secreted by the T3SS, ultimately relieving the repression of ExsA activity (Rietsch *et al*., 2005).

A similar paradigm exists in *P. syringae* where expression of the T3SS genes is upregulated by the activity of the alternate sigma factor HrpL encoded within the T3SS locus (Xiao *et al*., 1994). This gene is upregulated by the action of transcriptional activators HrpR and HrpS, the GacA/S Two component system (Xiao *et al*., 1994; Chatterjee *et al*., 2003).

Though the T3SS regulators encoded within T3SS loci are strongly associated with their cognate T3SS genes, several non-T3SS genes are also regulated by the action of these regulators (Thijs *et al*., 2007; Mucyn *et al*., 2014). As T3SS is often involved in an invasion process which results in a transfer to a dramatically different environment, this is perhaps unsurprising.

Regulation of T3SS genes by proteins encoded within T3SS loci allows a quick increase in expression. The time taken to achieve maximal expression can be further decreased by positive feedback mechanisms such as the pilus component HrpA upregulating T3SS gene expression (Ortiz-Martín *et al*., 2010). Another feedback mechanism is seen in *Yersinia* where the translocator YopD acts post-transcriptionally to prevent translation of T3SS transcripts until this repression is relieved by secretion of YopD (Kopaskie *et al*., 2013).

In many cases the T3SS mediates process which are vital for survival in or infection of hosts, making it advantageous to have the T3SS poised and ready. For example, low level of T3SS expression occurs even in non-inducing conditions (Rietsch and Mekalanos, 2006; Ortiz-Martín *et al*., 2010). Once fully assembled, the T3SS can be held in a stand-by state. The gatekeeper protein SctW associates with the T3SS via its interactions with SctI and acts to plug the secretion channel (Cherradi *et al*., 2013). It is hypothesised that host cell sensing by the needle causes the dissociation

of the SctI-SctW complex and leaves the secretion channel open. In *Yersinia* another protein, LcrG appears to block the secretion channel of the fully formed T3SS, and its repression is relieved by titration with increased levels of tip protein LcrV (Matson and Nilles, 2001).

The translocation of effectors YopE, YopT or ExoS into host cells limits the subsequent translocation of more effectors by other bacterial cells (Viboud and Bliska, 2001; Aili *et al*., 2008; Cisz *et al*., 2008). The precise mechanism involved remains to be elucidated but appears to involve the enzyme activities of these effectors within the host cells preventing the formation of additional translocation pores (Viboud and Bliska, 2001; Cisz *et al*., 2008).

Seemingly the only role of YopK is to limit translocation of effectors into the target cell. This is possibly dependent on its ability to interact with pore component YopD (Dewoody *et al*., 2013).

Halting additional translocation may be advantageous for preventing sister cells bearing the needless metabolic cost of secreting more proteins into an infected cell, but may also prevent premature death of host cells. The former hypothesis alone does not explain why a *yopK* mutant displays greater cytotoxicity, but is unable to resist killing by neutrophils (Thorslund *et al*., 2013).

3.1 The T6SS

The T6SS is another multimeric protein complex which can secrete molecules lacking an N-terminal signal sequence into target cells or the environmental milieu **(Figure 1.7)**. Genetic loci encoding T6SSs typically have 11-15 genes, though loci containing more are not uncommon and loci containing up to 39 genes have been observed (Murdoch *et al*., 2011). The similarity between some of the T6SS proteins and phage proteins suggests that the T6SS evolved from bacteriophage. As a bacteriophage delivers its payload into the bacterial cell and the T6SS is used to inject substrate molecules out of the bacterial cell, it is often referred to as an "inverted" bacteriophage. As befits an important and widespread secretion system, genes involved in T6SS were implicated in virulence and mutualistic processes before the T6SS was formally identified as a secretion system.

The first such discovery was a hemolysin co-regulated protein, named Hcp, which was shown to be secreted from *Vibrio cholerae* despite lacking a signal peptide (Williams *et al*., 1996). A strain of *Rhizobium leguinosarum* which was unable to form nodules with certain pea plants was subjected to transposon mutagenesis, resulting in mutants that were more effective at nodule formation (Roest *et al*., 1997). The site of insertion was later sequenced fully and found to be in a 14-gene operon called the *imp* (impaired in nodulation) operon. It is now clear that this operon encoded a T6SS. The fact the locus only impaired nodule formation on certain plants suggested this locus was an avirulence locus, which prompted an investigation into any secretion deficiency of the transposon mutant (Bladergroen *et al*., 2003). Protein profiles of the mutant lacked four proteins which are present in that of the wild type, one of which was identified as a ribose binding protein. The authors noted that similar genetic loci were to be found in other gram negative genomes. Concomitant with this work, a *icmF* homologue was identified as being induced in vivo in a rabbit model of infection (Das *et al*., 2002). Deletion of *sci* island of *Salmonella enterica*, which also contains fimbrial genes, resulted in decreased ability of bacteria to enter eukaryotic cells (Folkesson *et al*., 2002). In addition a *Salmonella enterica* strain carrying a transposon in its *icmF* gene was found to be impaired in intracellular growth during the later stages of infection (Parsons and Heffron, 2005). Several *Erwinia tarda* mutants were reported to be impaired in virulence towards fish (Srinivasa Rao *et al*., 2003). Later characterisation

of these mutants revealed that these strains did not secrete three T3SS proteins or two proteins, EvpA and EvpC, which were both encoded in a widespread locus of unknown function (Srinivasa Rao *et al*., 2004). Two of three single deletion mutants in this locus had lower replication rates in host macrophages and one of these was essential for the secretion of EvpC. In addition to these data, it was noted that two of the genes found at genetic loci encoding what are now known to be T6SS included the homologues to the *icmF* and *dotU* genes that code for essential structural components of the T4SS, which suggested that this locus could encode genes involved in secretion.

The secretion system was properly described as such in two papers from 2006. The first identified the secretion system in *Vibrio cholera* after using transposon mutagenesis to identify virulence determinants using the *Dictyostelium discodeum* model (Pukatzki *et al*., 2006). Disruption of the subsequently named *vas* (virulence-associated secretion) genes resulted in the loss of virulence towards both amoeba and macrophages, and abrogated the secretion of four proteins which lacked a secretion signal: three VgrG proteins and Hcp. The widespread nature of similar loci in other bacteria suggested a new and widespread secretion system had been identified -the T6SS.

Three distinct genetic loci encoding these T6SS are present in *P. aeruginosa*, and were named HSI-1, HSI-2 and HSI-3, for Hcp Secretion Island, in a paper which showed that secretion of Hcp1 was dependent on at least two of the genes encoded in the HSI-1 locus: *icmF1* and *clpV1* (Mougous *et al*., 2006). GFP-tagging of one of these genes, the ATPase ClpV1, showed localisation of this protein to distinct locations in the cell periphery. Concentrated foci of the ClpV1-GFP chimera were not observed when either the *icmF1* or *hcp1* genes were disrupted, consistent with the idea of a multimeric complex of interacting proteins. Structural evidence that the Hcp1 protein formed hexameraic rings with an internal diameter of 4 nm suggested that these proteins from a conduit for the passage of secreted substrates.

Virulence gene regulator AggR was found to control expression of a T6SS in *Escherichia. coli*, which was necessary for the secretion of the protein AaiC, which has no known function or homologues outside *E. coli* (Dudley *et al*., 2006).

3.2 Structure and Assembly of T6SS

Knockout mutation studies indicated that 13 proteins are needed for a functional secretion system (Zheng and Leung, 2007). Accordingly, the corresponding 13 genes are the most commonly found genes encoded within T6SS loci (Boyer *et al*., 2009). However, recent results indicate that small proteins containing a PAAR (proline-alanine-alanine-arginine) motif form the end of the needle and are important for complete T6SS activity, though it seems that they may not be absolutely required for a residual level of activity (Shneider *et al*., 2013). Similarly, recent work on one of the core components, ClpV, suggests that it too is not absolutely required for a basal level of T6SS activity (Basler *et al*., 2012). These 14 proteins seem to be essential or at least contribute very strongly to T6SS functioning in almost all cases. In contrast, additional proteins may be required for any or complete functioning of certain T6SS, though whether accessory proteins are required at all, or which accessory proteins are required or what roles they play varies markedly between different T6SS. The naming of these T6SS proteins from different species has been inconsistent through the literature. For example, homologues of one protein are known variously as VasK, EvpO, ImpL and IcmF, and TssM, the latter being in no way related to the T6SS effector protein called TssM that is found in *B. pseudomallei* (Tan *et al*., 2010). To address this issue a common nomenclature for Type VI proteins has been established (Shalom *et al*., 2007). With the exception of VgrG (TssI), Hcp (TssD) and ClpV (TssH) which have been consistently described using their original names, this nomenclature will be followed herein. Under this system, proteins involved in the T6SS are designated Tss (Type Six Secretion) in the case of core genes or Tag (Tss-associated gene) for accessory components e.g. the VasK/EvpO/ImpL/IcmF homologues are designated TssM with this nomenclature.

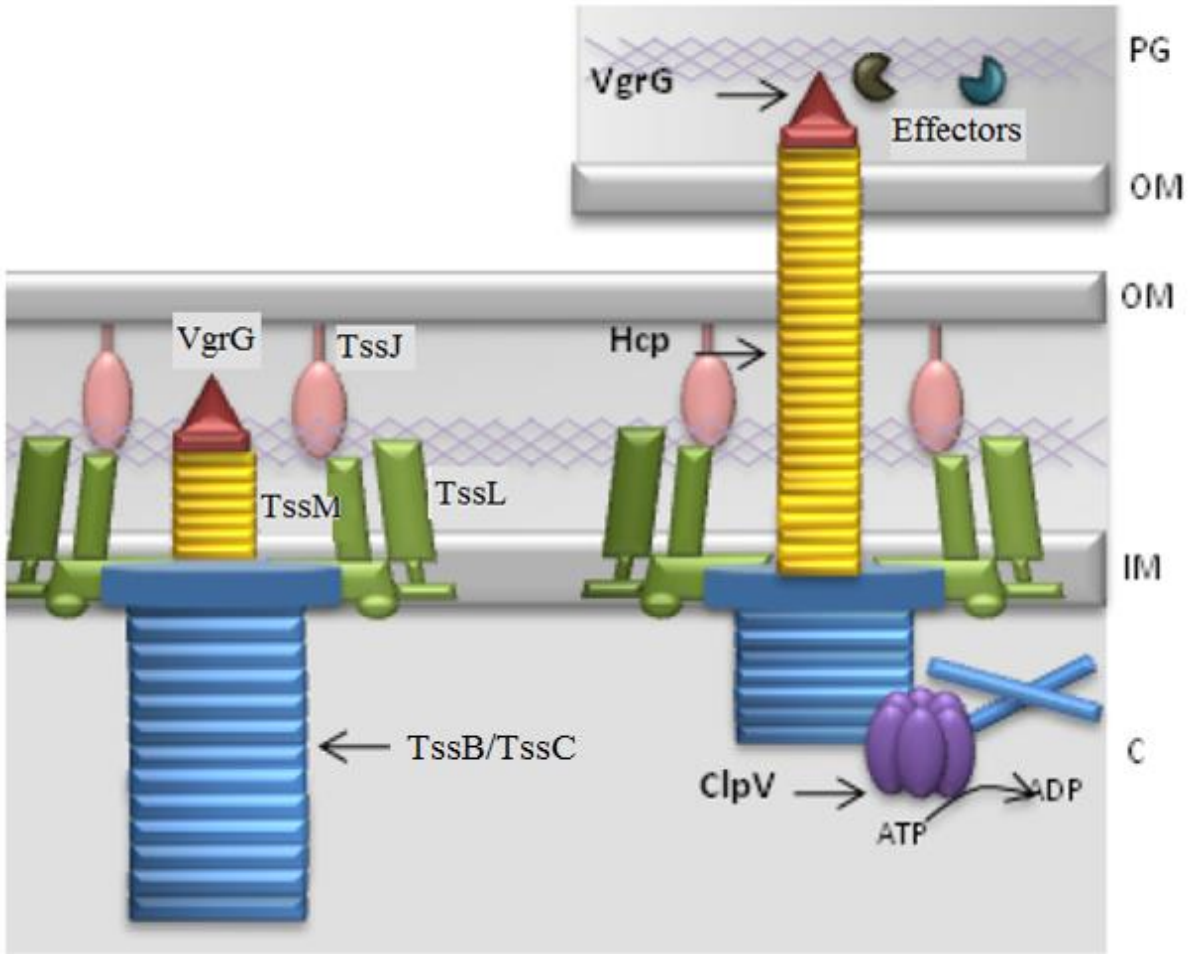


Figure 1.7. The structure of the T6SS in includes extended (left) and contracted (right) forms. The structure is anchored in the membrane by TssJ, TssL and TssM. The baseplate consists of at least TssF, TssG, TssK and possibly TssA and TssE. These components are hypothesised to remain in place between cycles of contraction, destruction and re-assembly. Figure modified from (Filloux, 2013)

Even though they display low degrees of sequence homology, several of the proteins of the T6SS display high levels of structural homology with components of the bacteriophage machinery (Mougous *et al*., 2006; Leiman *et al*., 2009; Pell *et al*., 2009; Veesler and Cambillau, 2011; Brunet *et al*., 2014). Therefore, the T6SS appears to be an adaption of the bacteriophage apparatus by gram negative bacteria for their own ends. Conceptually the T6SS structure can be divided into the bacteriophage analogues which act to puncture cells and deliver effectors and bacterial components which function to anchor the bacteriophage structure in the cell wall and membrane. The contractile bacteriophage delivers its nucleic acid payload through a molecular syringe which enters the bacterial cell after contraction of a

contractile sheath complex forces the syringe outwards. A baseplate which facilitates binding to the target cell and a needle complex which punctures the bacterial membrane are other necessary components of this system. All these structural elements are found in T6SS, though the baseplate appears to be located on the inner membrane and instead of allowing contact with the target cell presumably allows interaction with the membrane-embedded proteins and, by analogy with bacteriophage assembly, may also act to initiate formation of the T6SS syringe (Kostyuchenko *et al*., 2003).

The T6SS syringe and needle are formed by stacked hexameric rings of Hcp and a trimer of VgrG proteins, respectively (Mougous *et al*., 2006; Pukatzki *et al*., 2007). Hcp is structurally similar to the phage lambda tail protein gpV (Mougous *et al*., 2006; Pell *et al*., 2009) while VgrG is similar to the gp27/gp5 complex which forms the bacteriophage spike (Leiman *et al*., 2009). The gp27-like domain of VgrG, which contains a pore which mirrors that of the Hcp hexamer tube, is encoded at the N-terminal and located farther from the target cell envelope. Though Hcp and VgrG are essential structural components, they are also secreted by the T6SS and the hallmark of a functional T6SS is the presence of these proteins in the supernatant. These proteins also are translocated into the target cell and sometimes have toxic C-terminal domains, in which case they are referred to as "evolved" VgrG or Hcp (Pukatzki *et al*., 2007; Blondel *et al*., 2009).

As is the case for contractile bacteriophage, a contractile sheath surrounds the Hcp tubules and is composed of a complex of TssB and TssC proteins which form a cogwheel structure with an internal diameter of 10 nm and an external diameter of 30 nm (Bönemann *et al*., 2009; Lossi *et al*., 2013). This is sufficient to accommodate the Hcp hexamer tubule which has an external diameter of 8-9 nm and an internal diameter of 4 nm (Mougous *et al*., 2006). Both TssB and TssC have shorter half-lives in the absence of the other, suggesting the interaction is necessary for protein stability or resistance to endogenous proteases (Bönemann *et al*., 2009). Cell imaging shows that the sheath is up to 1 μm long in the extended state and approximately half that size in the contracted state (Basler *et al*., 2012).

Another set of proteins act to anchor the T6SS apparatus in the cell membrane, including the core proteins TssJ, TssL and TssM. TssM and TssL are homologous to

the IcmF and DotU proteins which stabilise the Type IV Secretion system (Sexton *et al*., 2004), and these proteins also directly interact to stabilise the T6SS. TssM contains one or three TM regions and the majority of the protein in located in the periplasm. TssM and TssL bind via their respective N-terminal domains (L.-S. Ma *et al*., 2009). TssL forms dimers which is important for its correct functioning (Durand *et al*., 2012). TssL is bound to the cell wall either directly through the action of a C-terminal peptidoglycan binding domain or, in cases where the C-terminal domain is absent from TssL, an accessory protein which interacts with TssL and binds to the cell wall via its own peptidoglycan domain (Aschtgen, Gavioli, *et al*., 2010; Aschtgen, Thomas, *et al*., 2010). Interestingly, which accessory protein carries the peptidoglycan binding domain varies between different phylogenetic clusters of T6SS (Aschtgen, Thomas, *et al*., 2010). A loop between two Beta strands of the outer membrane lipoprotein TssJ are exposed in the periplasm and interact with the C-terminal region of TssM (Felisberto-Rodrigues *et al*., 2011). TssM also has the ability to bind and hydrolyse NTP through its Walker A and Walker B motifs, which is important for its functioning (Ma *et al*., 2012).

The membrane bound proteins and the bacteriophage-related proteins are proposed to be linked by a baseplate-like structure composed of at least three proteins: TssK, TssF and TssG. These proteins form a multiunit complex which is localised to the inner membrane and is analogous to the bacteriophage baseplate (Basler *et al*., 2012; English *et al*., 2014). The best characterised component of the baseplate, TssK, is able to bind to several proteins including Hcp, TssL, TssF TssG and itself (Zoued *et al*., 2013; English *et al*., 2014). TssA and TssE are the only two components with no known roles in the T6SS. As TssE shares similarity with phage baseplate protein, it is likely to also form part of the baseplate. Strains lacking TssE do not form TssB/TssC tubules, suggesting that this baseplate has a role in the assembly of the sheath (Basler *et al*., 2012). TssA is a cytoplasmic protein capable of binding to TssK and is assumed to also form part of the baseplate (Zoued *et al*., 2013).

The T6SS is proposed to fire when a conformational change forces contraction of the TssB/TssC sheath and ejection of the T6SS spike and Hcp tubule **(Figure 1.7)**. Two core components of the T6SS, ClpV and TssM, contain domains with ATPase activity, and therefore had been suggested to energise the secretion system (L.-S. Ma *et al*., 2009), but as other roles have been identified for each of these proteins, the

current thinking is that both these hypotheses are false and the energy required for contraction is contained in the extended formation of the TssB/TssC complex (Ho *et al*., 2014).

ClpV is an ATPase protein related to chaperone ClpB (Mougous *et al*., 2006). ClpV was shown to be necessary for Hcp secretion by *E. tarda*, while deletion of *clpV* in *V. cholerae* abrogated secretion of Hcp in the supernatant, and abolished the T6SS-dependent killing of *D. discodeum* by this bacterium (Zheng and Leung, 2007; Zheng *et al*., 2011). From these studies it was concluded that ClpV was a core component and necessary for secretion. However, subsequent work suggested that some T6SS activity occurred in the absence of ClpV. T6SS-dependent killing of *E. coli* by *V. cholerae* was markedly reduced, but not abolished by deletion of this gene (Zheng *et al*., 2011). Further studies on this protein showed attachment of the ClpV N-terminal sequence to another protein, ClpP, lead to the degradation of TssB, suggesting the N-terminal sequence of ClpV targets TssB (Bönemann *et al*., 2009). A slight reduction in TssC levels was also recorded, presumably due to the lack of stabilising by TssB. TssB only interacts with the N-terminal of ClpV when bound to TssC, which is consistent with the hypothesis that ClpV targets the complex of TssB/TssC after formation of the tail sheath. Remodelling of TssB/TssC complex is dependent on ATP hydrolysis, and occurs through substrate threading through a pore in the ClpV (Bönemann *et al*., 2009). It was later shown that sheath contraction still occurs in *clpV* mutants, but in these mutants the contracted sheaths do not disassemble, indicating that the function of ClpV is to recycle the sheath proteins in an ATP-dependent manner after T6SS firing instead of providing energy for the firing process (Basler *et al*., 2012). This explains why some ClpV mutants display a basal level of T6SS-dependent killing. Similarly, while using the presence of Hcp in supernatant as a proxy for T6SS activity suggested that ClpV was essential to the T6SS process this may be due to Hcp levels simply being below the detection limit in these case, instead of being absent.

The other T6SS component with ATPase activity is TssM. While TssM is an essential component of the secretion system, the secretion defect of an *tssM* mutant was successfully complemented with both a TssM protein carrying an alanine substitution in place of a crucial lysine in the Walker A motif and a TssM protein carrying a double residue substitution in the Walker A motif, indicating that the

NTPase activity of this protein was not necessary for secretion in *E. tarda* (Zheng and Leung, 2007). In contrast, the alanine substitution in the same conserved lysine in TssM of *Agrobacterium tumefaciens* significantly decreased the amount of Hcp in the supernatant, while an additional substitution in the adjacent Glycine abolished Hcp secretion completely (L.-S. Ma *et al*., 2009). As this double residue substitution was not used in the experiments by Zheng and colleagues, these results are not directly comparable. These data may indicate that the requirement for NTPase activity of TssM varies between different T6SS in different strains or that residue substitutions used do not completely abrogate the NTPase activity. Higher degradation of TssM by protease at lower ATP concentrations indicated that NTP binding by TssM resulted in a conformational change, but only after the bound NTP is hydrolysed does Hcp bind to the C-terminal of TssL (Ma *et al*., 2012). The authors proposed a model whereby TssM underwent a conformation change after NTP binding and hydrolysis, resulting in a subsequent change in TssL which made the Hcp binding site accessible. This same study found that the *A. tumefaciens* TssM protein containing substitutions in the Walker A motif retained significant ATPase activity, but this was further decreased in a TssM protein containing additional substitutions in this motif (Ma *et al*., 2012). Thus it likely that the Walker A motif is necessary for T6SS secretion, and that some residual ATPase activity was present in the *E. tarda tssM* mutant, allowing for sufficient Hcp secretion by this bacterium (Zheng and Leung, 2007).

All the core components of the T6SS have known roles, apart from TssE and TssA, and there is evidence that these are baseplate proteins. In contrast the roles of the most accessory proteins are much less well understood. Not only have several accessory proteins got no known function, the list of accessory proteins is possibly inaccurate as it is based on what genes were encoded in T6SS loci in sequenced genomes in 2007 (Shalom *et al*., 2007). While the list has seen some additions, possibly others remain to be added, while some of the current proteins on the list may represent effectors, or proteins with very specific roles in a limited number of T6SS, and should not be considered as true accessory components.

3.3 Roles of T6SS

Given that the earliest studies involving T6SS involved interactions with amoeba, macrophages, plants, and the intestine by known animal pathogens such as *V. cholerae* and *P. aeruginosa*, it is perhaps not surprising that the T6SS was initially seen as a virulence determinant in bacterial-eukaryotic process. This line of enquiry was followed in several subsequent studies, and indeed many interactions between eukaryotes and gram negative bacteria were found to involve the T6SS.

3.3.1 T6SS in bacteria-eukaryotic interactions.

As mentioned, the T6SS was identified by screening transposon mutants defective in killing of *D. discodeum* (Pukatzki *et al*., 2006). Intact T6SS genes were necessary for *B. mallei* virulence in the hamster model (Schell *et al*., 2007). Furthermore, sera from mice, horse and human which were infected with *B. mallei* contained antibodies to Hcp, suggesting the T6SS is also involved in virulence processes in these hosts. Similarly, mice infected with *A. hydrophila* produced antibodies which bound to the Hcp protein from this bacteria and the T6SS-deficient strains were defective in virulence towards both murine macrophages and HeLa cells (Suarez *et al*., 2008). Most interestingly, while T6SS mutants were impaired in virulence in a mouse septicaemia model in this study, it was shown that prior challenge with Hcp prevented *A. hydrophila* infection, suggesting Hcp could be the basis of a potential vaccine.

V. cholerae was shown to cross link actin in J774 murine macrophage cells by the action of an evolved VgrG protein with a C-terminal actin cross linking domain (Pukatzki *et al*., 2007). This activity was later shown to only occur after translocation of the bacteria into the cells and possibly serves to impair the subsequent phagocytosis of cells yet to be phagocytised (A. T. Ma *et al*., 2009)

Expression of T6SS is increased at 26 °C compared to 37 °C in *Y. pestis*, suggesting T6SS is involved in the flea-borne stage of its lifecycle (Pieper *et al*., 2009). However, the relevance of this thermoregulation is unclear as T6SS mutants were found not impaired in a flea model, and instead T6SS was shown to promote uptake

and limit intracellular replication in murine macrophage J774 cells (Robinson *et al*., 2009).

Knocking out either HSI-2 or HSI-3 lead to a reduced *P. aeruginosa* PA14 bacterial load three days after inoculation into *Arabidopsis thaliana* leaves. In contrast, both HSI-2 and HSI-3 had to be knocked out before virulence was impaired in the murine models, suggesting compensation occurs between the two loci (Lesic *et al*., 2009). Whether the basis for this compensation is the secretion of the same effectors by both loci or secretion of effectors with overlapping roles is unknown.

Like other secretion systems, the T6SS is also able to mediate mutualistic interactions as well as pathogenesis. In *H. hepaticus* the T6SS has a role in limiting colonisation and inflammation in a murine model (Chow and Mazmanian, 2010). In addition, the T6SS genes of the plant pathogenic *Pectobacterium astrosepticum* were induced by potato extract, but a T6SS mutant displayed increased virulence towards this host, suggesting that T6SS limits pathogenesis in this particular interaction (Mattinen *et al*., 2007).

In addition, the FPI secretion system of *Francisella tularensis*, which bears similarity with the T6SS, is required for virulence in mice, as it is needed for escaping from the phagosome and reproducing within macrophages (Barker *et al*., 2009).

3.3.2 T6SS in bacterial-prokaryotic interactions

Despite the initial and continuing findings which implicate T6SS in virulent or mutualistic interactions with eukaryotic hosts, many data suggested that T6SS could also be used to target other prokaryotes. The similarity between T6SS and the bacteriophage which targets bacteria suggested this system was capable to delivering molecules into other bacteria (Schwarz *et al*., 2007). Moreover, the identification of a peptidoglycan-binding C-terminal domain on an evolved VgrG and C-terminal extensions of both VgrG and Hcp containing S-type pyocin domains raised the possibility that the T6SS could have an anti-bacterial function (Pukatzki *et al*., 2007; Blondel *et al*., 2009). The fact that this secretion system was present in the genomes of many bacteria which had no known virulence phenotypes suggested that T6SS

could be involved in more general ecological roles, including bacterial-bacterial interactions (Schwarz, Hood, *et al*., 2010). One of the first such roles identified for the T6SS was in mediating contact dependent inhibition in *P. mirabilis,* but as this was dependent on a genetic locus which contained *hcp* and *vgrG* and several other genes which were not known to be involved in T6SS, it was not conclusive proof of the T6SS's ability to target bacterial cells (Gibbs *et al*., 2008). The first concrete evidence of such an ability came from studies on *P. aeruginosa* where three T6SS effectors (Tse1, Tse2 and Tse3) were discovered by analysing the secretome of a *P. aeruginosa* strain over-expressing the HSI-1 system by mass spectrometry (Hood *et al*., 2010). Competition assays between *P. aeruginosa* mutants showed that Tse2 was capable of killing in a T6SS-dependent manner, and that the adjacently-encoded immunity protein Tsi2 was able to rescue strains from Tse2-associated death. Reports of T6SS-dependent bacterial antagonism by *V. cholerae* V52 against *Escherichia, Citrobacter* and *Salmonella* and by *Serratia marcescens* Db10 against *P. fluorescens, E. coli, Enterobacter cloacae* and another strain of *S. marcescens*, soon followed (MacIntyre *et al*., 2010; Murdoch *et al*., 2011). Interestingly, both these strains constitutively express T6SS under laboratory conditions, indicating that this antibacterial functioning may be a key trait.

The bacteria targeted by the T6SS vary. Despite the antibacterial action of *V. cholerae* V52 towards several Gammaproteobacterial species, *P. aeruginosa* was not killed, possibly indicating that it has the appropriate antitoxin gene(s) to neutralise the T6SS effector(s) (MacIntyre *et al*., 2010). In separate co-culture experiments with 31 different bacterial competitors, a T6SS mutant strain of *B. thailandensis* was significantly attenuated compared to the wild type in three cases (Schwarz, West, *et al*., 2010). The strains which outcompeted the mutant, namely *S. proteamaculans, P. fluorescens and P. putida*, were all Gammaproteobacteria. The T6SS had no effect on competition experiments with other proteobacteria or Firmicutes, suggesting a high degree of specificity. This may be due to differences in the exterior of such cells as it has been shown that the capsule of *Campylobacter jejuni* impairs T6SS functioning in this organism (Bleumink-Pluym *et al*., 2013). Moreover, the T6SS mutant *of B. thailandensis* was displaced from biofilms by *P. putida* while the wild type strain was able to persist (Schwarz, West, *et al*., 2010). Interestingly, in competition experiments it was only the proliferation of *B. thailandensis* which was

affected by mutation of T6SS; the other organisms grew to similar extents regardless of the *Burkholderia* T6SS activity. Possibly the secreted effectors impaired the production or delivery of an anti-*Burkholderia* factor, without impairing the growth of other bacteria.

In addition to experimental evidence, it is now clear from the widespread nature of bactericidal T6SS effectors that a range of phylogenetically diverse bacteria employ the T6SS for inter-bacterial killing (Russell *et al*., 2014). Though some T6SS appear to be exclusively antagonistic towards eukaryotes or prokaryotes, anti-bacterial and anti-eukaryotic roles of the T6SS should not be seen as mutually exclusive. Indeed, some T6SS and T6SS effectors which can target both eukaryotes and prokaryotes have been identified.

3.4 Effectors of T6SS

Except in certain instances where T6SS may be playing a physiological role, the function of T6SS ultimately depends on the effectors it delivers. Much progress has been made in this area concurrent with this dissertation. Many of the effectors characterised to date target other bacteria, and these are almost always encoded with an adjacent immunity protein which protects the cell from self-intoxication or "friendly fire" attacks from nearby sister cells. This is also the case for evolved Hcp and VgrG when they encode domains which are toxic for the producing cell (Pukatzki *et al*., 2007; Blondel *et al*., 2009). The first standalone T6SS effectors i.e. not Hcp or VgrG, to be identified were a ribose binding protein of unknown role and EvpP, which is important for virulence towards fish, but whose mechanism is still unknown (Bladergroen *et al*., 2003; Zheng and Leung, 2007; Wang *et al*., 2009). In contrast, the next set of effectors to be identified, the Tse effectors from *P. aeruginosa*, have been characterised (Hood *et al*., 2010).

While the mechanism of Tse2 is not fully understood, structural modelling predicts that it resembles enzymes which bind nucleic acid, and this is in keeping with its ability to induce quiescence, but not death, in target cells (Li *et al*., 2012). Furthermore, nuclease activity is common in polymorphic toxins (Zhang *et al*., 2012). Tse1 and Tse3 were characterised as bactericidal effectors which target the cell wall, and like Tse2, their activity is antagonised by adjacently-encoded

immunity proteins which served to protect the attacking cell or nearby sister cells (Russell *et al*., 2011). Killing mediated through these effectors can be abrogated by expression of immunity genes, increasing the osmolarity of the growth media, or impairing T6SS function. The addition of effector Tse1 exogenously to bacterial cells did not have any effect unless the cells were permeabilised, confirming their delivery into the cell by the T6SS. Neither Tse1 nor Tse3 are able to access the periplasm from the cytosol, suggesting that these must be injected directly into the periplasm, though it is not known if this holds true for all T6SS effectors (Russell *et al*., 2011). HSI-1 dependent delivery of Tse2 between *P. aeruginosa* sister cells was dependent on a mutation resulting in HSI-1 overexpression (Hood *et al*., 2010). However, HSI-1-dependent delivery of Tse1 and Tse3 was detected when the wild type *P. aeruginosa* cells were co-incubated with *P. putida* cells, showing that T6SS killing was stimulated by the presence of non-self bacteria (Russell *et al*., 2011). Similar toxin-antitoxin T6SS effectors, Ssp1 and Ssp2, were identified in *S. marcescens* and were shown to cause the T6SS dependent killing of *E. coli* (English *et al*., 2012). In fact, both these effectors and Tse1 and Tse3 were later shown to belong to two different superfamilies of T6SS effectors (Russell *et al*., 2012). The cell-wall degrading abilities of Tse1 and Tse3 are due to their amidase and glyosidase hydrolase activity, respectively (Russell *et al*., 2011). For this reason these superfamilies are known as Tae (T6SS amidase effectors) and Tge (T6SS glycoside hydrolase effectors) (Whitney *et al*., 2013; Russell *et al*., 2014) The Tae superfamily consists of 4 families while the Tge superfamily consists of 3 families. Importantly, these families are not confined to the *Pseudomonas* genus where they were first identified, making them a widespread effector potentially mediating many interbacterial interactions.

Though interbacterial interactions are clearly very important activities of the T6SS, eukaryotic-targeting effectors have also been characterised. However, most of these target conserved elements in prokaryotes and eukaryotes, with possible exceptions being the aforementioned EvpP and the actin-crosslinking Vgr3. While Tse2 was shown to be toxin when heterologously expressed in yeast cells, targeting of this effector to any eukaryote remains to be demonstrated (Hood *et al*., 2010). In contrast, the effector VasX contributes towards *V. cholerae* virulence against both eukaryotes and prokaryotes. Initially this effector was characterised as being

involved in virulence toward amoeba, and the N-terminal of this effector protein which contains a putative pleckstrin homology domain was shown to bind to phosphatidylinositol phosphates, which are more commonly found in eukaryotes (Miyata *et al*., 2011). However, subsequent modelling showed it possessed a low level of structural homology to the bacterocin colicin (Zheng *et al*., 2011). Moreover, it was later shown that VasX was toxic to bacterial cells and that the presence of VasX in the periplasm disrupts membrane-potential, consistent with a role as a pore forming toxin (Miyata *et al*., 2013). As is the case with other anti-bacterial effectors, a specific immunity protein is encoded adjacently to VasX. Interestingly, the operon containing VasX was shown to have an internal promoter allowing for expression of VasI independent of VasX (Miyata *et al*., 2013).

Another superfamily of T6SS effectors identified concurrent with the work of this thesis were the Tle (Type Six Lipase Effectors). The initial Tle effector, which is now known as Tle2, was identified in *V. cholerae* as a lipase protein which was lethal to both bacteria and amoeba and whose action was antagonised by a specific, adjacently encoded, immunity protein. (Dong *et al*., 2013). Like Tse1 and Tse3, this protein was also shown to be a representative of a superfamily composed of five different families of lipase effectors: Tle1-5 (Russell *et al*., 2013). By targeting a conserved element of bacterial and eukaryote membranes, these T6SS effectors can potentially antagonise a wide range of other organisms.

3.5 Regulation of T6SS

As would be expected of a widespread secretion system with diverse roles, T6SS expression has been optimised for specific bacterial lifestyles by being integrated in various existing general regulatory pathways, many of which control a range of other virulence genes. This is certainly true of T6SS regulation in *P. aeruginosa* which has been an active are of study **(Figure 1.8)** (Bernard *et al*., 2010).

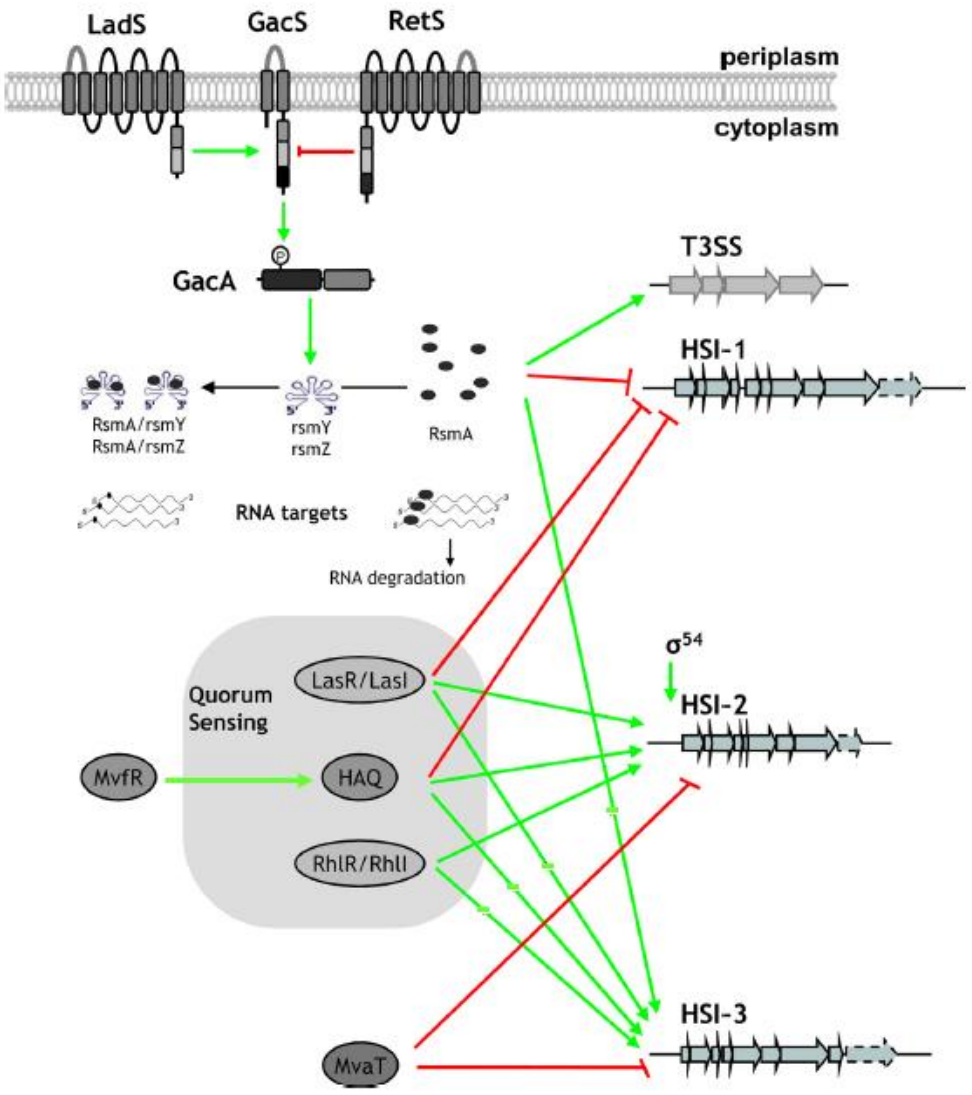


Figure 1.8. The regulation of the three HSI loci of *P. aeruginosa* is dependent on a variety of different pathways, including quorum sensing, the GacS/A signal transduction system and internally encoded enhancer binding proteins. Figure modified from (Bernard *et al*., 2010).

The expression of many virulence factors is controlled in a density-dependent fashion through the levels of quorum sensing molecules (Rutherford and Basler, 2014). Indeed, two T6SS loci in *P. aeruginosa*, HSI-2 and HSI-3, are induced in this manner, though only partially in the case of HSI-3 (Lesic *et al*., 2009). Other examples of quorum sensing dependent regulation of T6SS include T6SS expression in *Pectobacterium astrosepticum*, *Aeromonas hydrophila* and positive and negative regulation of T6SS in *V. cholerae*, as measured by Hcp production, by quorum sensing regulators HapR and LuxO (Liu *et al*., 2008; Ishikawa *et al*., 2009; Khajanchi *et al*., 2009). Regulation by quorum sensing is sometimes indicative of

growth phase dependent regulation, as these molecules accumulate as the bacterial population grows.

For example, expression of HSI-I in *P. aeruginosa* and HSI-III in *P. syringae* is controlled by the sensor kinases RetS/LadS, where RetS activity ultimately represses T6SS expression (Goodman *et al*., 2004; Brencic and Lory, 2009; Records and Gross, 2010). As these sensor kinases control the amount of RsmA which is available to bind to, and thereby repress translation of various transcripts, this pathway regulates T6SS at a post-transcriptional level. One interesting feature of regulation by this pathway is that T3SS and T6SS are regulated by RetS and cyclic di-GMP concentrations in a reciprocal manner in *P. aeruginosa* **(Figure 1.9)** (Moscoso *et al*., 2011).

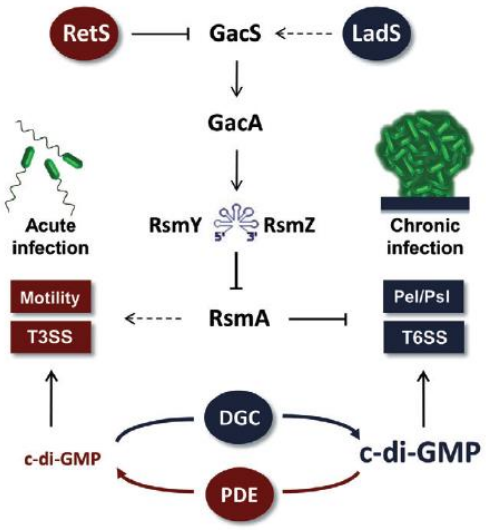


Figure 1.9. In *P. aeruginosa* the switch between the HSI-I T6SS and T3SS is regulated by the post-transcriptional regulator protein RsmA and the concentration of cyclic-di-GMP. DGC and PDE stand for Diguanylate cyclase and Phosphodiesterase which are enzymes involved in the synthesis and degradation of cyclic di-GMP. Figure reproduced from (Moscoso *et al*., 2011).

An interesting example of T6SS regulation has been described in *E. coli* (Brunet *et al*., 2011). In iron replete conditions repression of T6SS expression in enteroaggregative *E. coli* is mediated by the ferric uptake regulator (Fur) protein which binds to the promoter sequence to prevent transcription. When iron is limiting, such as is the case in host animals, Fur repression ceases, allowing expression of the T6SS genes. Furthermore, the GATC motifs in the DNA become available for methylation by DNA adenine methylase. Methylation decreases Fur affinity for the promoter and this methylation is inherited by one of the daughter cells after DNA replication and binary fission, making it an epigenetic mechanism of T6SS control.

In common with many horizontally acquired genes, T6SS genes can be controlled by histone-like nucleoid structuring proteins, as is the case for HSI-II and HSI-III in *P. aeruginosa* which are silenced by MvaT (Castang *et al*., 2008).

In addition to being integrated into more general regulatory networks, T6SS expression can also be mediated by specific regulators, including by an enhancer binding protein which is encoded within some loci (Bernard *et al*., 2011). This enhancer binding protein, encoded by the *sfa2* gene in HSI-II of *P. aeruginosa*, increases the affinity of the alternative sigma factor 54 for the T6SS promoter. This is reminiscent of the activation of T3SS which is often controlled by the activity of regulatory proteins encoded within a locus, such as transcriptional activator HilA or the T3SS enhancer binding protein HrpL (Lindeberg *et al*., 2006; Ellermeier and Slauch, 2007). However, this is more common in T3SS than in T6SS (Boyer *et al*., 2009). Specific activators encoded within T6SS loci reflect the need to increase T6SS rapidly and dramatically, which is consistent with a role of T6SS in antagonistic interactions with either eukaryotic or prokaryotic cells where the death of either party is a possible outcome. In fact, in some cases the importance of having an active T6SS is reflected by constitutive T6SS expression (Murdoch *et al*., 2011). This may be because a T6SS is so frequently used in the life cycle of this bacterium, but alternatively may indicate the importance of having a T6SS ready to fire is great enough to warrant the metabolic costs of transcribing and translating the T6SS genes even if it is not always used.

Firing of a T6SS secretion system is a destructive process, requiring at least partial reformation of the structural apparatus as well as possibly damaging the cell wall and

membrane and using energy in both the processes of assembly, firing and disassembly. Moreover, it is a risky activity as T6SS firing can prompt a counter attack from the target cell (LeRoux *et al*., 2012; Basler *et al*., 2013). Therefore, firing often or indiscriminately could be an undesirable strategy in many cases. However, should a bacterial cell encounter a potential target cell, the lag time between a transcriptional response and T6SS firing may be undesirable as the target cell may only be within attacking range for a limited period of time, and perhaps more importantly the target cell may also be able to kill through its own T6SS or by another method. For this reason, many T6SS loci have accessory genes which regulate the T6SS firing post-translationally.

3.5.1 Post-translational T6SS regulation

Post-translation regulation has been best characterised in *P. aeruginosa* where T6SS activity, and possibly assembly, is dependent on the accessory component TagH which has a Fha (forkhead-associated) domain for binding to phosphorylated threonine residues (Mougous 2007). Where present, Fha is an essential component whose absence abrogates detectable T6SS activity, and in *P. aeruginosa* at least it is necessary for recruitment of ClpV to the T6SS foci (Mougous *et al*., 2007). While ClpV activity is a necessary step in the complete firing and recycling of the T6SS, ClpV recruitment is not sufficient for the T6SS activity so Fha must play other roles (Mougous *et al*., 2007). Indeed, as outlined below, another role for Fha has been described in *A, tumefaciens* where Fha binding to TssL is a necessary step for TssL-Hcp binding (Lin *et al*., 2014). In *P. aeruginosa* two separate regulatory pathways converge on the Fha protein, one involving a repressor protein TagF and the other involving a threonine phosphorylation pathway (Silverman *et al*., 2011).

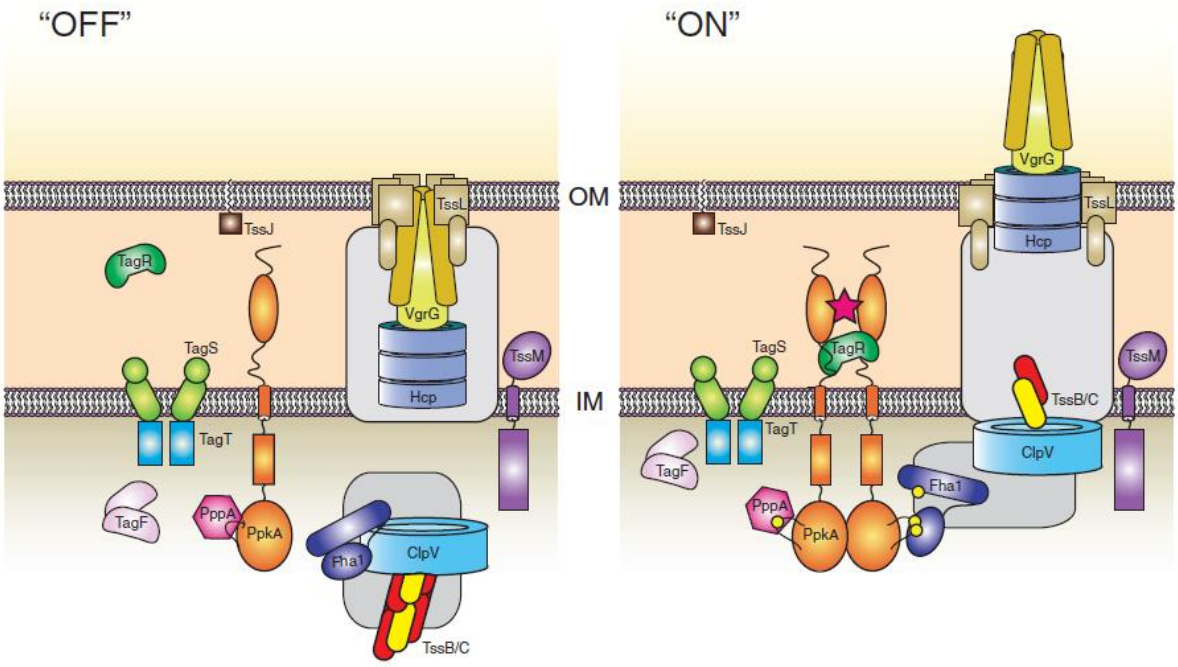


Figure 1.10. Post-translational regulation of the T6SS by the phosphothreonine pathway. Activating conditions, most likely to be related to membrane perturbation, are sensed by the cell ultimately resulting in PpkA dimerisation. After dimerisation and auto-phosphorylation of PpkA, Fha is phosphorylated. Dephosphorylation by the phosphatase PppA acts to reset the post-translational regulatory system. Some T6SS regulated in this manner have additional accessory proteins which act upstream of PpkA. Shown here is periplasmic accessory protein TagR which is necessary for the dimerisation of PpkA from HSI-I of *P. aeruginosa*. Figure reproduced from (Hsu *et al*., 2009).

In the threonine phosphorylation pathway, a membrane-embedded serine-threonine kinase PpkA and a serine threonine phosphatase PppA act antagonistically to determine the phosphorylation state of the Fha protein, and ultimately the activity of the T6SS **(Figure 1.10)**. Dimerisation of periplasmic domains of PpkA prompts auto-phosphorylation of PpkA and subsequent phosphorylation of a threonine residue of Fha, prompting secretion (Mougous *et al*., 2007; Hsu *et al*., 2009). PpkA is necessary for recruitment of ClpV by Fha to the T6SS foci, though this is independent of its own phosphorylation status and capability, suggesting it may be additional structural role (Mougous *et al*., 2007). Overexpression of *ppkA* or mutation of *pppA* does increase ClpV recruitment, but this is likely due to increased levels of T6SS assembly, firing and reassembly instead of phosphorylation of Fha directly leading to increased ClpV recruitment. The phosphatase PppA dephosphorylates both PpkA and Fha. In *P. aeruginosa*, a *pppA* mutant has elevated secretion, but it consistently fires its T6SS from the same arbitrary location while the

wild type can assemble and fire at various locations (Basler *et al*., 2013). Therefore, this post-translation regulatory system can regulate the T6SS spatio-temporally. A majority of T6SS loci do not possess any of the post-translational genes, while just over half of those which possess Fha also encode PpkA and PppA (Ho *et al*., 2014). This implies the existence of other post-translation accessory proteins which may act on Fha.

The second post-translational regulatory mechanism of HSI-I in *P. aeruginosa* is mediated through the TagF repressor protein which acts on Fha (Silverman *et al*., 2011). A *tagF* mutant displays ClpV recruitment to T6SS foci to a similar degree as a *pppA* mutant. Interestingly, a *tagF* mutant does not require PpkA for ClpV recruitment, despite its proposed structural role, which may indicate that this structural role is only required in the presence of the TagF protein. Just as PpkA is not needed for ClpV recruitment in the absence of TagF, T6SS activation only requires phosphorylation of Fha in the absence of TagF.

T6SS activation can occur either by de-repression of TagF or phosphorylation by PpkA. While either pathway acting individually is sufficient for T6SS activity, induction of both methods simultaneously in vitro increases secretion of the effector Tse1, but not Hcp. Though it is not entirely clear from the Western blot shown in this study (Silverman *et al*., 2011), this does not appear to be due to increased levels of cellular Tse1, possibly indicating the existence of a method of preferential substrate secretion which is responsive to environmental signal cues. In any case, this shows that these pathways are not simply two independent pathways with the same targets or a simple binary (on/off) outcome.

While it is not known what activates TagF depression, in *P. aeruginosa* the threonine phosphorylation pathway is activated by surface growth or membrane disruptions and depends on several additional accessory proteins (TagQRST) which act upstream of PpkA in the pathway (Casabona *et al*., 2013). TagQ and TagR are localised to the outer membrane with TagR localisation being dependent on TagQ. TagS and TagT reside as a membrane bound complex. As this complex has ATPase activity it may function as an ABC transporter, but its transported substrates are unknown and neither TagQ nor TagR are dependent on TagST for their localisation. TagR is similar to the periplasmic domains of other transmembrane kinases,

suggesting it may as a receptor or co-receptor for PpkA activation via the C-terminal Von Willebrand Factor A domain which is located in the periplasm (Hsu *et al*., 2009). While the exact signal the TagQRST proteins respond to is unknown, it is associated with membrane disruption. The TagT protein, and presumably the TagQRS, proteins are necessary a T6SS counter-attack in response to T6SS or T4SS activity directed against *P. aeruginosa* by an adjacent cell, and this response also occurs when the membrane is artificially disrupted by Polymixin B (Basler *et al*., 2013; Ho *et al*., 2013).

Even in loci encoding the *ppkA* and *pppA* genes, the additional *tagQRST* gens are often absent, so activation of PpkA is clearly dependent on other factors in these organisms (Ho *et al*., 2014). One such example is in *S. marcescens* which lacks the *tagQRST* genes and the C-terminal Von Willebrand Factor A domain in its PpkA, but still relies on phosphorelay from PpkA to Fha to regulate T6SS activity (Fritsch *et al*., 2013). While in *S. marcescens* Fha was the only protein identified by mass spectrometry to be phosphorylated by PpkA, in *A. tumefaciens* TssL is phosphorylated by PpkA and it is seems that Fha remains unphosphorylated (Lin *et al*., 2014). However, Fha binding to p-TssL is necessary for formation the TssM/TssL/Hcp complex and secretion. The model proposed by Lin and colleagues is that phosphorylation of TssL triggers an ATP dependent dimerisation of TssM, a subsequent change in TssL allows access for Fha binding, and that this allows TssL to bind to Hcp (Lin *et al*., 2014).

3.6 Physiological and other roles of T6SS

While much research has focused on the direct effects of T6SS on other cells, the T6SS has been shown to play physiological roles in several instances. Some of these may be secondary effects, as it is possible that disruption of a large membrane embedded protein complex has downstream effects on membrane fluidity, or attachment. Loss of TssJ from *E. coli* results in diminished biofilm formation (Aschtgen *et al*., 2008), which is potentially an effect of membrane disruption. Auto-agglutination is *Y. pestis* is caused by Hcp protein, presumably through its ability to polymerise (Podladchikova *et al*., 2011). While the *C. jejuni* T6SS is required for virulence towards mice, an apparent side effect of having the T6SS is decreased

survival in bile (Lertpiriyapong *et al*., 2012). The hypothesis given was that bile salts can enter through a secretion channel, and indeed, elevated bile salt levels were detected in the wild type strain compared to the mutant, using expression of a bile-regulated gene as a proxy for cellular bile salt levels. Similarly, increased sensitivity to bile in a *E. coli* strain expressing a T4SS has been reported (Bidlack and Silverman, 2004). In contrast, *Yersinia pseudotuberculosis* survival after exposure to osmotic stress, or bile salt deoxycholate was decreased in a T6SS mutant, a finding given more weight by the fact that expression of the specific T6SS involved was induced in response to osmotic stress through the response regulator OmpR (Gueguen *et al*., 2013). This is probably due to the surprising ability of this T6SS to pump protons out of the cell to maintain intercellular pH (Zhang *et al*., 2013). T6SS in *V. anguillarumis* is also in the bacterial stress response, specifically in regulating the stress response of the bacterium (Weber *et al*., 2009). The hypothesis forwarded for this phenotype is that signals which activate the regulator RpoS are sequestered by T6SS substrates. When the substrates are secreted by the T6SS, they are no longer present to sequester the solutes, which are then free to activate RpoS.

In addition it is possible that reported T6SS-dependent phenotypes do not involve the activity of the T6SS, but instead reflect the fact that some T6SS related genes may have additional roles. Deletion of *ppkA* and *pppA* resulted in a range of changes at the level of gene expression, particularly in oxidative stress response genes and genes regulated by the stationary phase sigma factor RpoS and either of the LasR/RhlR quorum sensing regulators (Goldová *et al*., 2011). No such phenotype has been reported for other mutants lacking other genes in this T6SS locus.

The lack of an appropriate oxidative stress response in particular is in keeping with the role of PpkA and PppA as responders to environmental stress such as a T6SS attack. In *V. alginolyticus* transcriptome analysis revealed that a range of genes were differentially regulated in a *pppA* mutant, including the flagellar genes and the gene encoding the quorum sensing regulator LuxR (Sheng *et al*., 2013) Transcription of *hcp* of HSI-III was reduced when HSI-II, which encodes the PpkA and PppA homologues, was removed. It is reasonable to assume that other genes of the HSI-III operon were also down-regulated by removal of HSI-II. This suggests that knocking out a T6SS can affect gene expression elsewhere, though it is not known if this was another example of regulation by PppA. There is no theoretical reason an indirect

transcriptional response arising from mutation of T6SS genes has to be limited to other T6SS genes.

Two studies involving phenotypes resulting from mutation of the T6SS gene *hcp clpV* and *tssM* in APEC highlight the inconsistencies of phenotypes in different T6SS mutants (de Pace *et al*., 2010, 2011). All of the mutants were unable to adhere to HeLa cells as effectively as the wild type. Type 1 Fimbriae expression was reduced in the *clpV* and *hcp* mutants, and the superior ability of the wild type to adhere to HeLa cells was not evident in the presence of D-mannose, which impairs Type 1 Fimbrial binding. Therefore the decreased ability of the *hcp* and *clpV* mutants seems to be due to the decreased expression of the fimbrial genes. Despite being impaired in HeLa cell binding compared to the wild type, fimbriae expression remained the same in the *tssM* mutant. The *hcp* and *tssM* mutants were impaired in HeLa cell invasion, but the *clpV* mutant was unaffected. Replication in macrophages was impaired in the *tssM* mutant but not the *hcp* or *clpV* mutants. Mutation of the *tssM* gene resulted in reduced expression of cAMP regulator protein, and ultimately the flagellar genes, resulting in a loss of motility. However, no motility defect was seen in the *hcp* and *clpV* mutants.

Discrepancies in experiments aren't constricted to different phenotypes in different T6SS mutants. Under very similar experimental conditions, the same strain of *V. cholerae* was reported to secrete VgrG2, or not secrete VgrG2 (Pukatzki *et al*., 2007; Miyata *et al*., 2013)

Another factor to consider when assessing reports of T6SS-dependent phenotypes is that previous reports of virulence impairment of T6SS mutants in *in vivo* models might actually reflect an inability of the pathogen to successfully kill, or survive among, other bacteria instead of an inability to directly interact with a host in a T6SS-dependent manner.

Paradoxically, the finding that the T6SS has anti-bacterial potential has important implications for preventing and treating infections. If the anti-bacterial activity of the T6SS is a key determinant in the establishment or progression of infection, it remains a target for developing anti-infection strategies. Indeed small molecule inhibitors of the T6SS of *V. cholerae* have been identified (Sun *et al*., 2014). As targeting a bacterial process that in not essential in replication may exert less

selective pressure than bactericidal agents, this may be an attractive method of dealing with the rise of antibiotic resistance. Furthermore, the potential for killing infectious bacteria with the T6SS may represent an alternative to phage therapy for developing new treatments.

4.1 Aims of this study

T6SS is a recently discovered secretion system and many questions in the field remain unanswered. This study aimed to identify the roles of the T6SS in *Pseudomonas* and beyond using both experimental and bioinformatic approaches. As the activities of the T6SS effectors ultimately determine the phenotype of the T6SS, a priority was to identify new effectors. In Chapter 2 T6SS of *Pseudomonas* were examined bioinformatically and characterised experimentally. In addition, new putative effectors were identified and these were linked with certain T6SS based on phylogenetic links. Bioinformatic analysis in Chapter 3 identified possible protein-protein interactions of T6SS accessory proteins, which may aid in discovery of new effectors and the mechanism(s) by which they associate with the T6SS. This chapter also shows that the while the T6SS occurs with similar frequencies across different environments, the Tle effector superfamily is more abundant in certain niches. Furthermore certain niches appear to be selecting for particular Tle families.

In contrast to the T6SS, the T3SS has been well studied, but mostly in the context of disease models. Moreover, each T3SS family has been studied in a limited range of strains. New genomic data is revealing that T3SS phylogenetic families are found in bacteria with different lifestyles to the model bacteria where these T3SS families are best described. Therefore, another aim of this study was to investigate the role of a T3SS belonging to the SPI-1 family in an environmental strain of *Pseudomonas*. In Chapter 4 this T3SS in *P. fluorescens* F113 was shown to be involved in *P. fluorescens* avoidance of amoeboid grazing.

References

Abby, S.S. and Rocha, E.P.C. (2012) The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems. *PLoS Genet* **8**: e1002983.

Abrusci, P., Vergara-Irigaray, M., Johnson, S., Beeby, M.D., Hendrixson, D.R., Roversi, P., *et al*. (2013) Architecture of the major component of the type III secretion system export apparatus. *Nat. Struct. Mol. Biol.* **20**: 99–104.

Aili, M., Isaksson, E.L., Carlsson, S.E., Wolf-Watz, H., Rosqvist, R., and Francis, M.S. (2008) Regulation of *Yersinia* Yop-effector delivery by translocated YopE. *Int. J. Med. Microbiol.* **298**: 183–192.

Akeda, Y. and Galán, J.E. (2005) Chaperone release and unfolding of substrates in type III secretion. *Nature* **437**: 911–915.

Allaoui, A., Woestyn, S., Sluiters, C., and Cornelis, G.R. (1994) YscU, a *Yersinia enterocolitica* inner membrane protein involved in Yop secretion. *J. Bacteriol.* **176**: 4534–4542.

Allen-Vercoe, E., Toh, M.C.W., Waddell, B., Ho, H., and DeVinney, R. (2005) A carboxy-terminal domain of Tir from enterohemorrhagic *Escherichia coli* O157:H7 (EHEC O157:H7) required for efficient type III secretion. *FEMS Microbiol. Lett.* **243**: 355–364.

Anderson, D.M., Fouts, D.E., Collmer, A., and Schneewind, O. (1999) Reciprocal secretion of proteins by the bacterial type III machines of plant and animal pathogens suggests universal recognition of mRNA targeting signals. *Proc. Natl. Acad. Sci.* **96**: 12839–12843.

Anderson, D.M., Ramamurthi, K.S., Tam, C., and Schneewind, O. (2002) YopD and LcrH Regulate Expression of Y*ersinia enterocolitica* YopQ by a Posttranscriptional Mechanism and Bind to *yopQ* RNA. *J. Bacteriol.* **184**: 1287–1295.

Aschtgen, M.-S., Bernard, C.S., Bentzmann, S.D., 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.

Aschtgen, M.-S., Gavioli, M., Dessen, A., Lloubès, R., and Cascales, E. (2010) The SciZ protein anchors the enteroaggregative *Escherichia coli* Type VI secretion system to the cell wall. *Mol. Microbiol.* **75**: 886–899.

Aschtgen, M.-S., Thomas, M.S., and Cascales, E. (2010) Anchoring the type VI secretion system to the peptidoglycan: TssL, TagL, TagP... what else? *Virulence* **1**: 535–540.

Backert, S. and Meyer, T.F. (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr. Opin. Microbiol.* **9**: 207–217.

Barahona, E., Navazo, A., Martínez-Granero, F., Zea-Bonilla, T., Pérez-Jiménez, R.M., Martín, M., and Rivilla, R. (2011) *Pseudomonas fluorescens* F113 Mutant with Enhanced Competitive Colonization Ability and Improved Biocontrol Activity against Fungal Root Pathogens. *Appl. Environ. Microbiol.* **77**: 5412–5419.

Barison, N., Lambers, J., Hurwitz, R., and Kolbe, M. (2012) Interaction of MxiG with the cytosolic complex of the type III secretion system controls *Shigella* virulence. *FASEB J.* **26**: 1717–1726.

Barker, J.R., Chong, A., Wehrly, T.D., Yu, J.-J., Rodriguez, S.A., Liu, J., *et al*. (2009) The *Francisella tularensis* pathogenicity island encodes a secretion system that is required for phagosome escape and virulence. *Mol. Microbiol.* **74**: 1459–1470.

Barta, M.L., Dickenson, N.E., Patil, M., Keightley, A., Wyckoff, G.J., Picking, W.D., *et al*. (2012) The Structures of Coiled-Coil Domains from Type III Secretion System Translocators Reveal Homology to Pore-Forming Toxins. *J. Mol. Biol.* **417**: 395–405.

Basler, M., Ho, B.T., and Mekalanos, J.J. (2013) Tit-for-Tat: Type VI Secretion System Counterattack during Bacterial Cell-Cell Interactions. *Cell* **152**: 884–894.

Basler, M., Pilhofer, M., Henderson, G.P., Jensen, G.J., and Mekalanos, J.J. (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**: 182–186.

Berger, C., Robin, G.P., Bonas, U., and Koebnik, R. (2010) Membrane topology of conserved components of the type III secretion system from the plant pathogen *Xanthomonas campestris* pv. vesicatoria. *Microbiology* **156**: 1963–1974.

Bergeron, J.R.C., Worrall, L.J., Sgourakis, N.G., DiMaio, F., Pfuetzner, R.A., Felise, H.B., *et al*. (2013) A Refined Model of the Prototypical *Salmonella* SPI-1 T3SS Basal Body Reveals the Molecular Basis for Its Assembly. *PLoS Pathog* **9**: e1003307.

Bernard, C.S., Brunet, Y.R., Gavioli, M., Lloubès, R., and Cascales, E. (2011) Regulation of Type VI Secretion Gene Clusters by σ54 and Cognate Enhancer Binding Proteins. *J. Bacteriol.* **193**: 2158–2167.

Bernard, C.S., Brunet, Y.R., Gueguen, E., and Cascales, E. (2010) Nooks and Crannies in Type VI Secretion Regulation. *J. Bacteriol.* **192**: 3850–3860.

Bidlack, J.E. and Silverman, P.M. (2004) An Active Type IV Secretion System Encoded by the F Plasmid Sensitizes *Escherichia coli* to Bile Salts. *J. Bacteriol.* **186**: 5202–5209.

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.

Blaylock, B., Riordan, K.E., Missiakas, D.M., and Schneewind, O. (2006) Characterization of the *Yersinia enterocolitica* Type III Secretion ATPase YscN and Its Regulator, YscL. *J. Bacteriol.* **188**: 3525–3534.

Bleumink-Pluym, N.M.C., van Alphen, L.B., Bouwman, L.I., Wösten, M.M.S.M., and van Putten, J.P.M. (2013) Identification of a Functional Type VI Secretion System in *Campylobacter jejuni* Conferring Capsule Polysaccharide Sensitive Cytotoxicity. *PLoS Pathog* **9**: e1003393.

Bleves, S., Viarre, V., Salacha, R., Michel, G.P.F., Filloux, A., and Voulhoux, R. (2010) Protein secretion systems in *Pseudomonas aeruginosa*: A wealth of pathogenic weapons. *Int. J. Med. Microbiol.* **300**: 534–543.

Van der Bliek, A.M. (2005) A sixth sense for Rab5. *Nat. Cell Biol.* **7**: 548–550.

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**: 354.

Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H., and Mogk, A. (2009) Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J.* **28**: 315–325.

Botteaux, A., Sani, M., Kayath, C.A., Boekema, E.J., and Allaoui, A. (2008) Spa32 interaction with the inner-membrane Spa40 component of the type III secretion system of *Shigella flexneri* is required for the control of the needle length by a molecular tape measure mechanism. *Mol. Microbiol.* **70**: 1515–1528.

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.

Brannon, M.K., Davis, J.M., Mathias, J.R., Hall, C.J., Emerson, J.C., Crosier, P.S., *et al*. (2009) *Pseudomonas aeruginosa* Type III secretion system interacts with phagocytes to modulate systemic infection of zebrafish embryos. *Cell. Microbiol.* **11**: 755–768.

Brencic, A. and Lory, S. (2009) Determination of the regulon and identification of novel mRNA targets of *Pseudomonas aeruginosa* RsmA. *Mol. Microbiol.* **72**: 612–632.

Bridge, D.R., Novotny, M.J., Moore, E.R., and Olson, J.C. (2010) Role of host cell polarity and leading edge properties in *Pseudomonas* type III secretion. *Microbiology* **156**: 356–373.

Browne, P., Rice, O., Miller, S.H., Burke, J., Dowling, D.N., Morrissey, J.P., and O'Gara, F. (2009) Superior inorganic phosphate solubilization is linked to phylogeny within the *Pseudomonas fluorescens* complex. *Appl. Soil Ecol.* **43**: 131–138.

Brown, I.R., Mansfield, J.W., Taira, S., Roine, E., and Romantschuk, M. (2001) Immunocytochemical Localization of HrpA and HrpZ Supports a Role for the Hrp

Pilus in the Transfer of Effector Proteins from *Pseudomonas syringae* pv. tomato Across the Host Plant Cell Wall. *Mol. Plant. Microbe Interact.* **14**: 394–404.

Broz, P., Mueller, C.A., Müller, S.A., Philippsen, A., Sorg, I., Engel, A., and Cornelis, G.R. (2007) Function and molecular architecture of the *Yersinia* injectisome tip complex. *Mol. Microbiol.* **65**: 1311–1320.

Brunet, Y.R., Bernard, C.S., Gavioli, M., Lloubès, R., and Cascales, E. (2011) An Epigenetic Switch Involving Overlapping Fur and DNA Methylation Optimizes Expression of a Type VI Secretion Gene Cluster. *PLoS Genet* **7**: e1002205.

Brunet, Y.R., Hénin, J., Celia, H., and Cascales, E. (2014) Type VI secretion and bacteriophage tail tubes share a common assembly pathway. *EMBO Rep.* **15**: 315–321.

Buchko, G.W., Niemann, G., Baker, E.S., Belov, M.E., Smith, R.D., Heffron, F., *et al*. (2010) A multi-pronged search for a common structural motif in the secretion signal of *Salmonella enterica* serovar Typhimurium type III effector proteins. *Mol. Biosyst.* **6**: 2448–2458.

Burghout, P., Beckers, F., Wit, E. de, Boxtel, R. van, Cornelis, G.R., Tommassen, J., and Koster, M. (2004) Role of the Pilot Protein YscW in the Biogenesis of the YscC Secretin in *Yersinia enterocolitica*. *J. Bacteriol.* **186**: 5366–5375.

Burkinshaw, B.J. and Strynadka, N.C.J. (2014) Assembly and structure of the T3SS. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1843**: 1649–1663.

Butrus, S.I., Klotz, S.A., and Misra, R.P. (1987) The Adherence of *Pseudomonas aeruginosa* to Soft Contact Lenses. *Ophthalmology* **94**: 1310–1314.

Büttner, D. (2012) Protein Export According to Schedule: Architecture, Assembly, and Regulation of Type III Secretion Systems from Plant- and Animal-Pathogenic Bacteria. *Microbiol. Mol. Biol. Rev.* **76**: 262–310.

Büttner, D., Nennstiel, D., Klüsener, B., and Bonas, U. (2002) Functional Analysis of HrpF, a Putative Type III Translocon Protein from *Xanthomonas campestris* pv. vesicatoria. *J. Bacteriol.* **184**: 2389–2398.

Bzymek, K.P., Hamaoka, B.Y., and Ghosh, P. (2012) Two Translation Products of *Yersinia yscQ* Assemble To Form a Complex Essential to Type III Secretion. *Biochemistry (Mosc.)* **51**: 1669–1677.

Caroline, G., Eric, F., Bohn, Y.-S.T., Sylvie, E., and Attree, I. (2008) Oligomerization of PcrV and LcrV, Protective Antigens of *Pseudomonas aeruginosa* and *Yersinia pestis*. *J. Biol. Chem.* **283**: 23940–23949.

Carter, M.Q., Chen, J., and Lory, S. (2010) The *Pseudomonas aeruginosa* Pathogenicity Island PAPI-1 Is Transferred via a Novel Type IV Pilus. *J. Bacteriol.* **192**: 3249–3258.

Casabona, M.G., Silverman, J.M., Sall, K.M., Boyer, F., Couté, Y., Poirel, J., *et al*. (2013) An ABC transporter and an outer membrane lipoprotein participate in

posttranslational activation of type VI secretion in *Pseudomonas aeruginosa*. *Environ. Microbiol.* **15**: 471–486.

Castang, S., McManus, H.R., Turner, K.H., and Dove, S.L. (2008) H-NS family members function coordinately in an opportunistic pathogen. *Proc. Natl. Acad. Sci.* **105**: 18947–18952.

Charkowski, A.O., Alfano, J.R., Preston, G., Yuan, J., He, S.Y., and Collmer, A. (1998) The *Pseudomonas syringae* pv. tomato HrpW Protein Has Domains Similar to Harpins and Pectate Lyases and Can Elicit the Plant Hypersensitive Response and Bind to Pectate. *J. Bacteriol.* **180**: 5211–5217.

Chatterjee, A., Cui, Y., Yang, H., Collmer, A., Alfano, J.R., and Chatterjee, A.K. (2003) GacA, the Response Regulator of a Two-Component System, Acts as a Master Regulator in *Pseudomonas syringae* pv. tomato DC3000 by Controlling Regulatory RNA, Transcriptional Activators, and Alternate Sigma Factors. *Mol. Plant. Microbe Interact.* **16**: 1106–1117.

Chatterjee, C., Kumar, S., Chakraborty, S., Tan, Y.W., Leung, K.Y., Sivaraman, J., and Mok, Y.-K. (2011) Crystal Structure of the Heteromolecular Chaperone, AscE-AscG, from the Type III Secretion System in *Aeromonas hydrophila*. *PLoS ONE* **6**: e19208.

Cherradi, Y., Hachani, A., and Allaoui, A. (2014) Spa13 of *Shigella flexneri* has a dual role: chaperone escort and export gate-activator switch of the type III secretion system. *Microbiology* **160**: 130–141.

Cherradi, Y., Schiavolin, L., Moussa, S., Meghraoui, A., Meksem, A., Biskri, L., *et al*. (2013) Interplay between predicted inner-rod and gatekeeper in controlling substrate specificity of the type III secretion system. *Mol. Microbiol.* **87**: 1183–1199.

Chithrashree, Udayashankar, A.C., Chandra Nayaka, S., Reddy, M.S., and Srinivas, C. (2011) Plant growth-promoting rhizobacteria mediate induced systemic resistance in rice against bacterial leaf blight caused by *Xanthomonas oryzae* pv. oryzae. *Biol. Control* **59**: 114–122.

Choi, M.-S., Kim, W., Lee, C., and Oh, C.-S. (2013) Harpins, Multifunctional Proteins Secreted by Gram-Negative Plant-Pathogenic Bacteria. *Mol. Plant. Microbe Interact.* **26**: 1115–1122.

Chow, J. and Mazmanian, S.K. (2010) A Pathobiont of the Microbiota Balances Host Colonization and Intestinal Inflammation. *Cell Host Microbe* **7**: 265–276.

Cisz, M., Lee, P.-C., and Rietsch, A. (2008) ExoS Controls the Cell Contact-Mediated Switch to Effector Secretion in *Pseudomonas aeruginosa*. *J. Bacteriol.* **190**: 2726–2738.

Collin, S., Guilvout, I., Nickerson, N.N., and Pugsley, A.P. (2011) Sorting of an integral outer membrane protein via the lipoprotein-specific Lol pathway and a dedicated lipoprotein pilotin. *Mol. Microbiol.* **80**: 655–665.

Cornelis, G.R. (2006) The type III secretion injectisome. *Nat. Rev. Microbiol.* **4**: 811–825.

Couillerot, O., Prigent-Combaret, C., Caballero-Mellado, J., and Moënne-Loccoz, Y. (2009) *Pseudomonas fluorescens* and closely-related fluorescent pseudomonads as biocontrol agents of soil-borne phytopathogens. *Lett. Appl. Microbiol.* **48**: 505–512.

Cowell, B.A., Evans, D.J., and Fleiszig, S.M.J. (2005) Actin cytoskeleton disruption by ExoY and its effects on *Pseudomonas aeruginosa* invasion. *FEMS Microbiol. Lett.* **250**: 71–76.

Cowell, B.A., Twining, S.S., Hobden, J.A., Kwong, M.S.F., and Fleiszig, S.M.J. (2003) Mutation of *lasA* and *lasB* reduces *Pseudomonas aeruginosa* invasion of epithelial cells. *Microbiology* **149**: 2291–2299.

Crago, A.M. and Koronakis, V. (1998) *Salmonella* InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. *Mol. Microbiol.* **30**: 47–56.

Cronin, D., Moënne-Loccoz, Y., Fenton, A., Dunne, C., Dowling, D.N., and O'Gara, F. (1997) Ecological interaction of a biocontrol *Pseudomonas fluorescens* strain producing 2,4-diacetylphloroglucinol with the soft rot potato pathogen *Erwinia carotovora* subsp. atroseptica. *FEMS Microbiol. Ecol.* **23**: 95–106.

Cronin, D., Moenne-Loccoz, Y., Fenton, A., Dunne, C., Dowling, D.N., and O'Gara, F. (1997) Role of 2,4-Diacetylphloroglucinol in the Interactions of the Biocontrol Pseudomonad Strain F113 with the Potato Cyst Nematode *Globodera rostochiensis*. *Appl. Environ. Microbiol.* **63**: 1357–1361.

Cunnac, S., Lindeberg, M., and Collmer, A. (2009) *Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions. *Curr. Opin. Microbiol.* **12**: 53–60.

Cusano, A.M., Burlinson, P., Deveau, A., Vion, P., Uroz, S., Preston, G.M., and Frey-Klett, P. (2011) *Pseudomonas fluorescens* BBc6R8 type III secretion mutants no longer promote ectomycorrhizal symbiosis. *Environ. Microbiol. Rep.* **3**: 203–210.

Darwin, K.H. and Miller, V.L. (1999) InvF Is Required for Expression of Genes Encoding Proteins Secreted by the SPI1 Type III Secretion Apparatus in *Salmonella typhimurium*. *J. Bacteriol.* **181**: 4949–4954.

Dasgupta, N., Lykken, G.L., Wolfgang, M.C., and Yahr, T.L. (2004) A novel anti-anti-activator mechanism regulates expression of the *Pseudomonas aeruginosa* type III secretion system. *Mol. Microbiol.* **53**: 297–308.

Das, S., Chakrabortty, A., Banerjee, R., and Chaudhuri, K. (2002) Involvement of in vivo induced *icmF* gene of *Vibrio cholerae* in motility, adherence to epithelial cells, and conjugation frequency. *Biochem. Biophys. Res. Commun.* **295**: 922–928.

Daval, S., Lebreton, L., Gazengel, K., Boutin, M., Guillerm-Erckelboudt, A.-Y., and Sarniguet, A. (2011) The biocontrol bacterium *Pseudomonas fluorescens* Pf29Arp strain affects the pathogenesis-related gene expression of the take-all fungus

Gaeumannomyces graminis var. tritici on wheat roots. *Mol. Plant Pathol.* **12**: 839–854.

Dean, P. (2011) Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol. Rev.* **35**: 1100–1125.

Demers, J.-P., Sgourakis, N.G., Gupta, R., Loquet, A., Giller, K., Riedel, D., *et al*. (2013) The Common Structural Architecture of *Shigella flexneri* and *Salmonella typhimurium* Type Three Secretion Needles. *PLoS Pathog* **9**: e1003245.

Deng, Q. and Barbieri, J.T. (2008) Modulation of Host Cell Endocytosis by the Type III Cytotoxin, *Pseudomonas* ExoS. *Traffic* **9**: 1948–1957.

Deng, W., Puente, J.L., Gruenheid, S., Li, Y., Vallance, B.A., Vázquez, A., *et al*. (2004) Dissecting virulence: Systematic and functional analyses of a pathogenicity island. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 3597–3602.

Dewoody, R., Merritt, P.M., and Marketon, M.M. (2013) YopK controls both rate and fidelity of Yop translocation. *Mol. Microbiol.* **87**: 301–317.

Diepold, A., Amstutz, M., Abel, S., Sorg, I., Jenal, U., and Cornelis, G.R. (2010) Deciphering the assembly of the *Yersinia* type III secretion injectisome. *EMBO J.* **29**: 1928–1940.

Diepold, A. and Wagner, S. (2014) Assembly of the bacterial type III secretion machinery. *FEMS Microbiol. Rev.* n/a–n/a.

Diepold, A., Wiesand, U., and Cornelis, G.R. (2011) The assembly of the export apparatus (YscR,S,T,U,V) of the *Yersinia* type III secretion apparatus occurs independently of other structural components and involves the formation of an YscV oligomer. *Mol. Microbiol.* **82**: 502–514.

Dijkstra, A.J. and Keck, W. (1996) Peptidoglycan as a barrier to transenvelope transport. *J. Bacteriol.* **178**: 5555–5562.

Dong, T.G., Ho, B.T., Yoder-Himes, D.R., and Mekalanos, J.J. (2013) Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. *Proc. Natl. Acad. Sci.* **110**: 2623–2628.

Duan, J., Jiang, W., Cheng, Z., Heikkila, J.J., and Glick, B.R. (2013) The Complete Genome Sequence of the Plant Growth-Promoting Bacterium *Pseudomonas* sp. UW4. *PLoS ONE* **8**: e58640.

Dudley, E.G., Thomson, N.R., Parkhill, J., Morin, N.P., and Nataro, J.P. (2006) Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative *Escherichia coli*. *Mol. Microbiol.* **61**: 1267–1282.

Durand, E., Zoued, A., Spinelli, S., Watson, P.J.H., Aschtgen, M.-S., Journet, L., *et al*. (2012) Structural Characterization and Oligomerization of the TssL Protein, a Component Shared by Bacterial Type VI and Type IVb Secretion Systems. *J. Biol. Chem.* **287**: 14157–14168.

Ellermeier, J.R. and Slauch, J.M. (2007) Adaptation to the host environment: regulation of the SPI1 type III secretion system in *Salmonella enterica* serovar Typhimurium. *Curr. Opin. Microbiol.* **10**: 24–29.

English, G., Byron, O., Cianfanelli, F., Prescott, A., and Coulthurst, S. (2014) Biochemical analysis of TssK, a core component of the bacterial Type VI secretion system, reveals distinct oligomeric states of TssK and identifies a TssK-TssFG sub-complex. *Biochem. J.*

English, G., Trunk, K., Rao, V.A., Srikannathasan, V., Hunter, W.N., and Coulthurst, S.J. (2012) New secreted toxins and immunity proteins encoded within the Type VI secretion system gene cluster of *Serratia marcescens*. *Mol. Microbiol.* **86**: 921–936.

Epler, C.R., Dickenson, N.E., Olive, A.J., Picking, W.L., and Picking, W.D. (2009) Liposomes Recruit IpaC to the *Shigella flexneri* Type III Secretion Apparatus Needle as a Final Step in Secretion Induction. *Infect. Immun.* **77**: 2754–2761.

Espina, M., Olive, A.J., Kenjale, R., Moore, D.S., Ausar, S.F., Kaminski, R.W., *et al*. (2006) IpaD Localizes to the Tip of the Type III Secretion System Needle of *Shigella flexneri*. *Infect. Immun.* **74**: 4391–4400.

Espinosa, A., Guo, M., Tam, V.C., Fu, Z.Q., and Alfano, J.R. (2003) The *Pseudomonas syringae* type III-secreted protein HopPtoD2 possesses protein tyrosine phosphatase activity and suppresses programmed cell death in plants. *Mol. Microbiol.* **49**: 377–387.

Estahbanati, H.K., Kashani, P.P., and Ghanaatpisheh, F. (2002) Frequency of *Pseudomonas aeruginosa* serotypes in burn wound infections and their resistance to antibiotics. *Burns* **28**: 340–348.

Evans, L.D.B. and Hughes, C. (2009) Selective binding of virulence type III export chaperones by FliJ escort orthologues InvI and YscO. *FEMS Microbiol. Lett.* **293**: 292–297.

Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008) The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320**: 1034–1039.

Felisberto-Rodrigues, C., Durand, E., Aschtgen, M.-S., Blangy, S., Ortiz-Lombardia, M., Douzi, B., *et al*. (2011) Towards a Structural Comprehension of Bacterial Type VI Secretion Systems: Characterization of the TssJ-TssM Complex of an *Escherichia coli* Pathovar. *PLoS Pathog* **7**: e1002386.

Fenton, A.M., Stephens, P.M., Crowley, J., O'Callaghan, M., and O'Gara, F. (1992) Exploitation of gene(s) involved in 2,4-diacetylphloroglucinol biosynthesis to confer a new biocontrol capability to a *Pseudomonas* strain. *Appl. Environ. Microbiol.* **58**: 3873–3878.

Ferris, H.U., Furukawa, Y., Minamino, T., Kroetz, M.B., Kihara, M., Namba, K., and Macnab, R.M. (2005) FlhB Regulates Ordered Export of Flagellar Components via Autocleavage Mechanism. *J. Biol. Chem.* **280**: 41236–41242.

Filloux, A. (2011) Protein secretion systems in *Pseudomonas aeruginosa*: an essay on diversity, evolution, and function. *Cell. Infect. Microbiol. - Closed Sect.* **2**: 155.

Filloux, A. (2013) The rise of the Type VI secretion system. *F1000Prime Rep.* **5**:

Folkesson, A., Löfdahl, S., and Normark, S. (2002) The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. *Res. Microbiol.* **153**: 537–545.

Frank, D.W. and Iglewski, B.H. (1991) Cloning and sequence analysis of a trans-regulatory locus required for exoenzyme S synthesis in *Pseudomonas aeruginosa*. *J. Bacteriol.* **173**: 6460–6468.

Fritsch, M.J., Trunk, K., Diniz, J.A., Guo, M., Trost, M., and Coulthurst, S.J. (2013) Proteomic Identification of Novel Secreted Antibacterial Toxins of the *Serratia marcescens* Type VI Secretion System. *Mol. Cell. Proteomics* **12**: 2735–2749.

Frost, S., Ho, O., Login, F.H., Weise, C.F., Wolf-Watz, H., and Wolf-Watz, M. (2012) Autoproteolysis and Intramolecular Dissociation of *Yersinia* YscU Precedes Secretion of Its C-Terminal Polypeptide YscUCC. *PLoS ONE* **7**: e49349.

Galle, M., Jin, S., Bogaert, P., Haegman, M., Vandenabeele, P., and Beyaert, R. (2012) The *Pseudomonas aeruginosa* Type III Secretion System Has an Exotoxin S/T/Y Independent Pathogenic Role during Acute Lung Infection. *PLoS ONE* **7**: e41547.

Gazi, A.D., Sarris, P.F., Fadouloglou, V.E., Charova, S.N., Mathioudakis, N., Panopoulos, N.J., and Kokkinidis, M. (2012) Phylogenetic analysis of a gene cluster encoding an additional, rhizobial-like type III secretion system that is narrowly distributed among *Pseudomonas syringae* strains. *BMC Microbiol.* **12**: 188.

Goehring, U.-M., Schmidt, G., Pederson, K.J., Aktories, K., and Barbieri, J.T. (1999) The N-terminal Domain of *Pseudomonas aeruginosa* Exoenzyme S Is a GTPase-activating Protein for Rho GTPases. *J. Biol. Chem.* **274**: 36369–36372.

Göhre, V. and Robatzek, S. (2008) Breaking the Barriers: Microbial Effector Molecules Subvert Plant Immunity. *Annu. Rev. Phytopathol.* **46**: 189–215.

Goldová, J., Ulrych, A., Hercík, K., and Branny, P. (2011) A eukaryotic-type signalling system of *Pseudomonas aeruginosa* contributes to oxidative stress resistance, intracellular survival and virulence. *BMC Genomics* **12**: 437.

Goodman, A.L., Kulasekara, B., Rietsch, A., Boyd, D., Smith, R.S., and Lory, S. (2004) A Signaling Network Reciprocally Regulates Genes Associated with Acute Infection and Chronic Persistence in *Pseudomonas aeruginosa*. *Dev. Cell* **7**: 745–754.

Gueguen, E., Durand, E., Zhang, X.Y., d' Amalric, Q., Journet, L., and Cascales, E. (2013) Expression of a *Yersinia pseudotuberculosis* Type VI Secretion System Is Responsive to Envelope Stresses through the OmpR Transcriptional Activator. *PLoS ONE* **8**: e66615.

Guo, M., Tian, F., Wamboldt, Y., and Alfano, J.R. (2009) The Majority of the Type III Effector Inventory of *Pseudomonas syringae* pv. tomato DC3000 Can Suppress Plant Immunity. *Mol. Plant. Microbe Interact.* **22**: 1069–1080.

Haapalainen, M., Engelhardt, S., Küfner, I., Li, C.-M., Nürnberger, T., Lee, J., *et al*. (2011) Functional mapping of harpin HrpZ of *Pseudomonas syringae* reveals the sites responsible for protein oligomerization, lipid interactions and plant defence induction. *Mol. Plant Pathol.* **12**: 151–166.

Hassett, D.J., Alsabbagh, E., Parvatiyar, K., Howell, M.L., Wilmott, R.W., and Ochsner, U.A. (2000) A Protease-Resistant Catalase, KatA, Released upon Cell Lysis during Stationary Phase Is Essential for Aerobic Survival of a *Pseudomonas aeruginosa oxyR* Mutant at Low Cell Densities. *J. Bacteriol.* **182**: 4557–4563.

Hayward, R.D., Cain, R.J., McGhie, E.J., Phillips, N., Garner, M.J., and Koronakis, V. (2005) Cholesterol binding by the bacterial type III translocon is essential for virulence effector delivery into mammalian cells. *Mol. Microbiol.* **56**: 590–603.

Hirano, S.S. and Upper, C.D. (1990) Population Biology and Epidemiology of *Pseudomonas Syringae*. *Annu. Rev. Phytopathol.* **28**: 155–177.

Ho, B.T., Basler, M., and Mekalanos, J.J. (2013) Type 6 Secretion System–Mediated Immunity to Type 4 Secretion System–Mediated Gene Transfer. *Science* **342**: 250–253.

Ho, B.T., Dong, T.G., and Mekalanos, J.J. (2014) A View to a Kill: The Bacterial Type VI Secretion System. *Cell Host Microbe* **15**: 9–21.

Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M.A., Trinidad, R.R.S., *et al*. (2010) A Type VI Secretion System of *Pseudomonas aeruginosa* Targets a Toxin to Bacteria. *Cell Host Microbe* **7**: 25–37.

Hsueh, P.-R., Teng, L.-J., Pan, H.-J., Chen, Y.-C., Sun, C.-C., Ho, S.-W., and Luh, K.-T. (1998) Outbreak of *Pseudomonas fluorescens* Bacteremia among Oncology Patients. *J. Clin. Microbiol.* **36**: 2914–2917.

Hsu, F., Schwarz, S., and Mougous, J.D. (2009) TagR promotes PpkA-catalysed type VI secretion activation in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **72**: 1111–1125.

Hueck, C.J. (1998) Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. *Microbiol. Mol. Biol. Rev.* **62**: 379–433.

Ishikawa, T., Rompikuntal, P.K., Lindmark, B., Milton, D.L., and Wai, S.N. (2009) Quorum Sensing Regulation of the Two *hcp* Alleles in *Vibrio cholerae* O1 Strains. *PLoS ONE* **4**: e6734.

Job, V., Matteï, P.-J., Lemaire, D., Attree, I., and Dessen, A. (2010) Structural Basis of Chaperone Recognition of Type III Secretion System Minor Translocator Proteins. *J. Biol. Chem.* **285**: 23224–23232.

Jouihri, N., Sory, M.-P., Page, A.-L., Gounon, P., Parsot, C., and Allaoui, A. (2003) MxiK and MxiN interact with the Spa47 ATPase and are required for transit of the needle components MxiH and MxiI, but not of Ipa proteins, through the type III secretion apparatus of *Shigella flexneri*. *Mol. Microbiol.* **49**: 755–767.

Journet, L., Agrain, C., Broz, P., and Cornelis, G.R. (2003) The Needle Length of Bacterial Injectisomes Is Determined by a Molecular Ruler. *Science* **302**: 1757–1760.

Kadurugamuwa, J.L. and Beveridge, T.J. (1995) Virulence factors are released from *Pseudomonas aeruginosa* in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion. *J. Bacteriol.* **177**: 3998–4008.

Kazmierczak, B.I. and Engel, J.N. (2002) *Pseudomonas aeruginosa* ExoT Acts In Vivo as a GTPase-Activating Protein for RhoA, Rac1, and Cdc42. *Infect. Immun.* **70**: 2198–2205.

Khajanchi, B.K., Sha, J., Kozlova, E.V., Erova, T.E., Suarez, G., Sierra, J.C., *et al*. (2009) N-Acylhomoserine lactones involved in quorum sensing control the type VI secretion system, biofilm formation, protease production, and in vivo virulence in a clinical isolate of *Aeromonas hydrophila*. *Microbiology* **155**: 3518–3531.

Kim, B.H., Kim, H.G., Kim, J.S., Jang, J.I., and Park, Y.K. (2007) Analysis of functional domains present in the N-terminus of the SipB protein. *Microbiology* **153**: 2998–3008.

Kochar, M., Upadhyay, A., and Srivastava, S. (2011) Indole-3-acetic acid biosynthesis in the biocontrol strain *Pseudomonas fluorescens* Psd and plant growth regulation by hormone overexpression. *Res. Microbiol.* **162**: 426–435.

Koo, J., Burrows, L.L., and Lynne Howell, P. (2012) Decoding the roles of pilotins and accessory proteins in secretin escort services. *FEMS Microbiol. Lett.* **328**: 1–12.

Kopaskie, K.S., Ligtenberg, K.G., and Schneewind, O. (2013) Translational regulation of *Yersinia enterocolitica* mRNA encoding a type III secretion substrate. *J. Biol. Chem.* jbc.M113.504811.

Koraimann, G. (2003) Lytic transglycosylases in macromolecular transport systems of Gram-negative bacteria. *Cell. Mol. Life Sci. CMLS* **60**: 2371–2388.

Kosarewicz, A., Königsmaier, L., and Marlovits, T.C. (2012) The blueprint of the type-3 injectisome. *Philos. Trans. R. Soc. B Biol. Sci.* **367**: 1140–1154.

Koster, M., Bitter, W., De Cock, H., Allaoui, A., Cornelis, G.R., and Tommassen, J. (1997) The outer membrane component, YscC, of the Yop secretion machinery of *Yersinia enterocolitica* forms a ring-shaped multimeric complex. *Mol. Microbiol.* **26**: 789–797.

Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., Kanamaru, S., van Raaij, M.J., Arisaka, F., *et al*. (2003) Three-dimensional structure of bacteriophage T4 baseplate. *Nat. Struct. Mol. Biol.* **10**: 688–693.

Krishnan, H.B., Natarajan, S.S., and Kim, W.-S. (2011) Distinct Cell Surface Appendages Produced by *Sinorhizobium fredii* USDA257 and S. fredii USDA191, Cultivar-Specific and Nonspecific Symbionts of Soybean ▿. *Appl. Environ. Microbiol.* **77**: 6240–6248.

Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., *et al*. (1998) Supramolecular Structure of the *Salmonella typhimurium* Type III Protein Secretion System. *Science* **280**: 602–605.

Kudryashev, M., Stenta, M., Schmelz, S., Amstutz, M., Wiesand, U., Castaño-Díez, D., *et al*. (2013) In situ structural analysis of the *Yersinia enterocolitica* injectisome. *eLife* **2**: e00792.

Kvitko, B.H., Ramos, A.R., Morello, J.E., Oh, H.-S., and Collmer, A. (2007) Identification of Harpins in *Pseudomonas syringae* pv. tomato DC3000, Which Are Functionally Similar to HrpK1 in Promoting Translocation of Type III Secretion System Effectors. *J. Bacteriol.* **189**: 8059–8072.

Lackner, G., Moebius, N., and Hertweck, C. (2011) Endofungal bacterium controls its host by an *hrp* type III secretion system. *ISME J.* **5**: 252–261.

Lara-Tejero, M., Kato, J., Wagner, S., Liu, X., and Galán, J.E. (2011) A Sorting Platform Determines the Order of Protein Secretion in Bacterial Type III Systems. *Science* **331**: 1188–1191.

Lee, D.G., Urbach, J.M., Wu, G., Liberati, N.T., Feinbaum, R.L., Miyata, S., *et al*. (2006) Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol.* **7**: R90.

Lee, J., Klüsener, B., Tsiamis, G., Stevens, C., Neyt, C., Tampakaki, A.P., *et al*. (2001) HrpZPsph from the plant pathogen *Pseudomonas syringae* pv. phaseolicola binds to lipid bilayers and forms an ion-conducting pore in vitro. *Proc. Natl. Acad. Sci.* **98**: 289–294.

Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., *et al*. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl. Acad. Sci.* **106**: 4154–4159.

LeRoux, M., Leon, J.A.D., Kuwada, N.J., Russell, A.B., Pinto-Santini, D., Hood, R.D., *et al*. (2012) Quantitative single-cell characterization of bacterial interactions reveals type VI secretion is a double-edged sword. *Proc. Natl. Acad. Sci.* **109**: 19804–19809.

Lertpiriyapong, K., Gamazon, E.R., Feng, Y., Park, D.S., Pang, J., Botka, G., *et al*. (2012) *Campylobacter jejuni* Type VI Secretion System: Roles in Adaptation to Deoxycholic Acid, Host Cell Adherence, Invasion, and In Vivo Colonization. *PLoS ONE* **7**: e42842.

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.

Lilic, M., Vujanac, M., and Stebbins, C.E. (2006) A Common Structural Motif in the Binding of Virulence Factors to Bacterial Secretion Chaperones. *Mol. Cell* **21**: 653–664.

Li, M., Le Trong, I., Carl, M.A., Larson, E.T., Chou, S., De Leon, J.A., *et al*. (2012) Structural Basis for Type VI Secretion Effector Recognition by a Cognate Immunity Protein. *PLoS Pathog* **8**: e1002613.

Lindeberg, M., Cartinhour, S., Myers, C.R., Schechter, L.M., Schneider, D.J., and Collmer, A. (2006) Closing the Circle on the Discovery of Genes Encoding Hrp Regulon Members and Type III Secretion System Effectors in the Genomes of Three Model *Pseudomonas syringae* Strains. *Mol. Plant. Microbe Interact.* **19**: 1151–1158.

Lindeberg, M., Cunnac, S., and Collmer, A. (2012) *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. *Trends Microbiol.* **20**: 199–208.

Lin, J.-S., Wu, H.-H., Hsu, P.-H., Ma, L.-S., Pang, Y.-Y., Tsai, M.-D., and Lai, E.-M. (2014) Fha Interaction with Phosphothreonine of TssL Activates Type VI Secretion in *Agrobacterium tumefaciens*. *PLoS Pathog* **10**: e1003991.

Liu, H., Coulthurst, S.J., Pritchard, L., Hedley, P.E., Ravensdale, M., Humphris, S., *et al*. (2008) Quorum Sensing Coordinates Brute Force and Stealth Modes of Infection in the Plant Pathogen *Pectobacterium atrosepticum*. *PLoS Pathog* **4**: e1000093.

Li, Y.-R., Che, Y.-Z., Zou, H.-S., Cui, Y.-P., Guo, W., Zou, L.-F., *et al*. (2011) Hpa2 Required by HrpF To Translocate *Xanthomonas oryzae* Transcriptional Activator-Like Effectors into Rice for Pathogenicity. *Appl. Environ. Microbiol.* **77**: 3809–3818.

Loquet, A., Sgourakis, N.G., Gupta, R., Giller, K., Riedel, D., Goosmann, C., *et al*. (2012) Atomic model of the type III secretion system needle. *Nature* **486**: 276–279.

Lossi, N.S., Manoli, E., Förster, A., Dajani, R., Pape, T., Freemont, P., and Filloux, A. (2013) The HsiB1C1 (TssB-TssC) Complex of the *Pseudomonas aeruginosa* Type VI Secretion System Forms a Bacteriophage Tail Sheathlike Structure. *J. Biol. Chem.* **288**: 7536–7548.

Lozupone, C.A. and Knight, R. (2007) Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci.* **104**: 11436–11440.

Luna, E., Pastor, V., Robert, J., Flors, V., Mauch-Mani, B., and Ton, J. (2010) Callose Deposition: A Multifaceted Plant Defense Response. *Mol. Plant. Microbe Interact.* **24**: 183–193.

Ma, A.T., McAuley, S., Pukatzki, S., and Mekalanos, J.J. (2009) Translocation of a *Vibrio cholerae* Type VI Secretion Effector Requires Bacterial Endocytosis by Host Cells. *Cell Host Microbe* **5**: 234–243.

MacIntyre, D.L., Miyata, S.T., Kitaoka, M., and Pukatzki, S. (2010) The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc. Natl. Acad. Sci.* **107**: 19520–19524.

Ma, L.-S., Lin, J.-S., and Lai, E.-M. (2009) An IcmF Family Protein, ImpLM, Is an Integral Inner Membrane Protein Interacting with ImpKL, and Its Walker A Motif Is Required for Type VI Secretion System-Mediated Hcp Secretion in *Agrobacterium tumefaciens*. *J. Bacteriol.* **191**: 4316–4329.

Ma, L.-S., Narberhaus, F., and Lai, E.-M. (2012) IcmF Family Protein TssM Exhibits ATPase Activity and Energizes Type VI Secretion. *J. Biol. Chem.* **287**: 15610–15621.

Marlovits, T.C., Kubori, T., Lara-Tejero, M., Thomas, D., Unger, V.M., and Galán, J.E. (2006) Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* **441**: 637–640.

Marlovits, T.C., Kubori, T., Sukhan, A., Thomas, D.R., Galán, J.E., and Unger, V.M. (2004) Structural Insights into the Assembly of the Type III Secretion Needle Complex. *Science* **306**: 1040–1042.

Matson, J.S. and Nilles, M.L. (2001) LcrG-LcrV Interaction Is Required for Control of Yops Secretion in *Yersinia pestis*. *J. Bacteriol.* **183**: 5082–5091.

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.

Matz, C., Moreno, A.M., Alhede, M., Manefield, M., Hauser, A.R., Givskov, M., and Kjelleberg, S. (2008) *Pseudomonas aeruginosa* uses type III secretion system to kill biofilm-associated amoebae. *ISME J.* **2**: 843–852.

McCaw, M.L., Lykken, G.L., Singh, P.K., and Yahr, T.L. (2002) ExsD is a negative regulator of the *Pseudomonas aeruginosa* type III secretion regulon. *Mol. Microbiol.* **46**: 1123–1133.

Meyer, D., Cunnac, S., Guéneron, M., Declercq, C., Gijsegem, F.V., Lauber, E., *et al*. (2006) PopF1 and PopF2, Two Proteins Secreted by the Type III Protein Secretion System of *Ralstonia solanacearum*, Are Translocators Belonging to the HrpF/NopX Family. *J. Bacteriol.* **188**: 4903–4917.

Michiels, T., Vanooteghem, J.C., Rouvroit, C.L. de, China, B., Gustin, A., Boudry, P., and Cornelis, G.R. (1991) Analysis of virC, an operon involved in the secretion of Yop proteins by *Yersinia enterocolitica*. *J. Bacteriol.* **173**: 4994–5009.

Mikkelsen, H., McMullan, R., and Filloux, A. (2011) The *Pseudomonas aeruginosa* Reference Strain PA14 Displays Increased Virulence Due to a Mutation in *ladS*. *PLoS ONE* **6**: e29113.

Miyata, S., Casey, M., Frank, D.W., Ausubel, F.M., and Drenkard, E. (2003) Use of the *Galleria mellonella* Caterpillar as a Model Host To Study the Role of the Type

III Secretion System in *Pseudomonas aeruginosa* Pathogenesis. *Infect. Immun.* **71**: 2404–2413.

Miyata, S.T., Kitaoka, M., Brooks, T.M., McAuley, S.B., and Pukatzki, S. (2011) *Vibrio cholerae* Requires the Type VI Secretion System Virulence Factor VasX To Kill *Dictyostelium discoideum*. *Infect. Immun.* **79**: 2941–2949.

Miyata, S.T., Unterweger, D., Rudko, S.P., and Pukatzki, S. (2013) Dual Expression Profile of Type VI Secretion System Immunity Genes Protects Pandemic *Vibrio cholerae*. *PLoS Pathog* **9**: e1003752.

Montagner, C., Arquint, C., and Cornelis, G.R. (2011) Translocators YopB and YopD from *Yersinia enterocolitica* Form a Multimeric Integral Membrane Complex in Eukaryotic Cell Membranes. *J. Bacteriol.* **193**: 6923–6928.

Morita-Ishihara, T., Ogawa, M., Sagara, H., Yoshida, M., Katayama, E., and Sasakawa, C. (2006) *Shigella* Spa33 Is an Essential C-ring Component of Type III Secretion Machinery. *J. Biol. Chem.* **281**: 599–607.

Moscoso, J.A., Mikkelsen, H., Heeb, S., Williams, P., and Filloux, A. (2011) The *Pseudomonas aeruginosa* sensor RetS switches Type III and Type VI secretion via c-di-GMP signalling. *Environ. Microbiol.* **13**: 3128–3138.

Mota, L.J. (2006) Type III secretion gets an LcrV tip. *Trends Microbiol.* **14**: 197–200.

Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., *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*. *Nat. Cell Biol.* **9**: 797–803.

Mucyn, T.S., Yourstone, S., Lind, A.L., Biswas, S., Nishimura, M.T., Baltrus, D.A., *et al*. (2014) Variable Suites of Non-effector Genes Are Co-regulated in the Type III Secretion Virulence Regulon across the *Pseudomonas syringae* Phylogeny. *PLoS Pathog* **10**: e1003807.

Mueller, C.A., Broz, P., Müller, S.A., Ringler, P., Erne-Brand, F., Sorg, I., *et al*. (2005) The V-Antigen of *Yersinia* Forms a Distinct Structure at the Tip of Injectisome Needles. *Science* **310**: 674–676.

Murdoch, S.L., Trunk, K., English, G., Fritsch, M.J., Pourkarimi, E., and Coulthurst, S.J. (2011) The Opportunistic Pathogen *Serratia marcescens* Utilizes Type VI Secretion To Target Bacterial Competitors. *J. Bacteriol.* **193**: 6057–6069.

Natale, P., Brüser, T., and Driessen, A.J.M. (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochim. Biophys. Acta BBA - Biomembr.* **1778**: 1735–1756.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V. a. P., *et al*. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**: 799–808.

Niemann, G.S., Brown, R.N., Mushamiri, I.T., Nguyen, N.T., Taiwo, R., Stufkens, A., *et al*. (2013) RNA Type III Secretion Signals that require Hfq. *J. Bacteriol.* JB.00024–13.

Noble, R.C. and Overman, S.B. (1994) *Pseudomonas stutzeri* infection a review of hospital isolates and a review of the literature. *Diagn. Microbiol. Infect. Dis.* **19**: 51–56.

Nwachukwu, S.U. (2001) Bioremediation of Sterile Agricultural Soils Polluted with Crude Petroleum by Application of the Soil Bacterium, *Pseudomonas putida*, with Inorganic Nutrient Supplementations. *Curr. Microbiol.* **42**: 231–236.

Ogino, T., Ohno, R., Sekiya, K., Kuwae, A., Matsuzawa, T., Nonaka, T., *et al*. (2006) Assembly of the Type III Secretion Apparatus of Enteropathogenic *Escherichia coli*. *J. Bacteriol.* **188**: 2801–2811.

Olive, A.J., Kenjale, R., Espina, M., Moore, D.S., Picking, W.L., and Picking, W.D. (2007) Bile Salts Stimulate Recruitment of IpaB to the *Shigella flexneri* Surface, Where It Colocalizes with IpaD at the Tip of the Type III Secretion Needle. *Infect. Immun.* **75**: 2626–2629.

Ortiz-Martín, I., Thwaites, R., Macho, A.P., Mansfield, J.W., and Beuzón, C.R. (2010) Positive Regulation of the Hrp Type III Secretion System in *Pseudomonas syringae* pv. phaseolicola. *Mol. Plant. Microbe Interact.* **23**: 665–681.

De Pace, F., Boldrin de Paiva, J., Nakazato, G., Lancellotti, M., Sircili, M.P., Guedes Stehling, E., *et al*. (2011) Characterization of IcmF of the type VI secretion system in an avian pathogenic *Escherichia coli* (APEC) strain. *Microbiol. Read. Engl.* **157**: 2954–2962.

De Pace, F., Nakazato, G., Pacheco, A., Paiva, J.B. de, Sperandio, V., and Silveira, W.D. da (2010) The Type VI Secretion System Plays a Role in Type 1 Fimbria Expression and Pathogenesis of an Avian Pathogenic *Escherichia coli* Strain. *Infect. Immun.* **78**: 4990–4998.

Pallen, M.J., Beatson, S.A., and Bailey, C.M. (2005) Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perpective⋆. *FEMS Microbiol. Rev.* **29**: 201–229.

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.

Pell, L.G., Kanelis, V., Donaldson, L.W., Howell, P.L., and Davidson, A.R. (2009) The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc. Natl. Acad. Sci.* **106**: 4160–4165.

Pieper, R., Huang, S.-T., Robinson, J.M., Clark, D.J., Alami, H., Parmar, P.P., *et al*. (2009) Temperature and growth phase influence the outer-membrane proteome and the expression of a type VI secretion system in *Yersinia pestis*. *Microbiology* **155**: 498–512.

Pier, G.B., Grout, M., and Zaidi, T.S. (1997) Cystic fibrosis transmembrane conductance regulator is an epithelial cell receptor for clearance of *Pseudomonas aeruginosa* from the lung. *Proc. Natl. Acad. Sci.* **94**: 12088–12093.

Podladchikova, O., Antonenka, U., Heesemann, J., and Rakin, A. (2011) *Yersinia pestis* autoagglutination factor is a component of the type six secretion system. *Int. J. Med. Microbiol.* **301**: 562–569.

Preston, G.M. (2007) Metropolitan Microbes: Type III Secretion in Multihost Symbionts. *Cell Host Microbe* **2**: 291–294.

Pukatzki, S., Kessin, R.H., and Mekalanos, J.J. (2002) The human pathogen *Pseudomonas aeruginosa* utilizes conserved virulence pathways to infect the social amoeba *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci.* **99**: 3159–3164.

Pukatzki, S., Ma, A.T., 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.* **104**: 15508–15513.

Pukatzki, S., Ma, A.T., 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. U. S. A.* **103**: 1528–1533.

Quinaud, M., Plé, S., Job, V., Contreras-Martel, C., Simorre, J.-P., Attree, I., and Dessen, A. (2007) Structure of the heterotrimeric complex that regulates type III secretion needle formation. *Proc. Natl. Acad. Sci.* **104**: 7803–7808.

Radics, J., Königsmaier, L., and Marlovits, T.C. (2014) Structure of a pathogenic type 3 secretion system in action. *Nat. Struct. Mol. Biol.* **21**: 82–87.

Radja Commare, R., Nandakumar, R., Kandan, A., Suresh, S., Bharathi, M., Raguchander, T., and Samiyappan, R. (2002) *Pseudomonas fluorescens* based bio-formulation for the management of sheath blight disease and leaffolder insect in rice. *Crop Prot.* **21**: 671–677.

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.

Rezzonico, F., Binder, C., Défago, G., and Moënne-Loccoz, Y. (2005) The Type III Secretion System of Biocontrol *Pseudomonas fluorescens* KD Targets the Phytopathogenic Chromista *Pythium ultimum* and Promotes Cucumber Protection. *Mol. Plant. Microbe Interact.* **18**: 991–1001.

Rietsch, A. and Mekalanos, J.J. (2006) Metabolic regulation of type III secretion gene expression in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **59**: 807–820.

Rietsch, A., Vallet-Gely, I., Dove, S.L., and Mekalanos, J.J. (2005) ExsE, a secreted regulator of type III secretion genes in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* **102**: 8006–8011.

Robinson, J.B., Telepnev, M.V., Zudina, I.V., Bouyer, D., Montenieri, J.A., Bearden, S.W., *et al*. (2009) Evaluation of a *Yersinia pestis* mutant impaired in a thermoregulated type VI-like secretion system in flea, macrophage and murine models. *Microb. Pathog.* **47**: 243–251.

Rodríguez, H. and Fraga, R. (1999) Phosphate solubilizing bacteria and their role in plant growth promotion. *Biotechnol. Adv.* **17**: 319–339.

Roest, H.P., Mulders, I.H.M., Spaink, H.P., Wijffelman, C.A., and Lugtenberg, B.J.J. (1997) A *Rhizobium leguminosarum* Biovar trifolii Locus Not Localized on the Sym Plasmid Hinders Effective Nodulation on Plants of the Pea Cross-Inoculation Group. *Mol. Plant. Microbe Interact.* **10**: 938–941.

Roine, E., Saarinen, J., Kalkkinen, N., and Romantschuk, M. (1997) Purified HrpA of *Pseudomonas syringae* pv. tomato DC3000 reassembles into pili. *FEBS Lett.* **417**: 168–172.

Ross, J.A. and Plano, G.V. (2011) A C-Terminal Region of *Yersinia pestis* YscD Binds the Outer Membrane Secretin YscC. *J. Bacteriol.* **193**: 2276–2289.

Russell, A.B., Hood, R.D., Bui, N.K., LeRoux, M., Vollmer, W., and Mougous, J.D. (2011) Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**: 343–347.

Russell, A.B., LeRoux, M., Hathazi, K., Agnello, D.M., Ishikawa, T., Wiggins, P.A., *et al*. (2013) Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature* **496**: 508–512.

Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014) Type VI secretion system effectors: poisons with a purpose. *Nat. Rev. Microbiol.* **12**: 137–148.

Russell, A.B., Singh, P., Brittnacher, M., Bui, N.K., Hood, R.D., Carl, M.A., *et al*. (2012) A Widespread Bacterial Type VI Secretion Effector Superfamily Identified Using a Heuristic Approach. *Cell Host Microbe* **11**: 538–549.

Saad, M.M., Staehelin, C., Broughton, W.J., and Deakin, W.J. (2008) Protein-Protein Interactions within Type III Secretion System-Dependent Pili of *Rhizobium* sp. Strain NGR234. *J. Bacteriol.* **190**: 750–754.

Salmond, G.P.C. and Reeves, P.J. (1993) Membrance traffic wardens and protein secretion in Gram-negative bacteria. *Trends Biochem. Sci.* **18**: 7–12.

Sato, H. and Frank, D.W. (2004) ExoU is a potent intracellular phospholipase. *Mol. Microbiol.* **53**: 1279–1290.

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

Schoehn, G., Di Guilmi, A.M., Lemaire, D., Attree, I., Weissenhorn, W., and Dessen, A. (2003) Oligomerization of type III secretion proteins PopB and PopD precedes pore formation in *Pseudomonas*. *EMBO J.* **22**: 4957–4967.

Schraidt, O., Lefebre, M.D., Brunner, M.J., Schmied, W.H., Schmidt, A., Radics, J., *et al*. (2010) Topology and Organization of the *Salmonella typhimurium* Type III Secretion Needle Complex Components. *PLoS Pathog* **6**: e1000824.

Schraidt, O. and Marlovits, T.C. (2011) Three-Dimensional Model of *Salmonella*'s Needle Complex at Subnanometer Resolution. *Science* **331**: 1192–1195.

Schwarz, S., Hood, R.D., and Mougous, J.D. (2010) What is type VI secretion doing in all those bugs? *Trends Microbiol.* **18**: 531–537.

Schwarz, S., West, T.E., Boyer, F., Chiang, W.-C., Carl, M.A., Hood, R.D., *et al*. (2010) *Burkholderia* Type VI Secretion Systems Have Distinct Roles in Eukaryotic and Bacterial Cell Interactions. *PLoS Pathog* **6**: e1001068.

Sears, C.L. (2005) A dynamic partnership: Celebrating our gut flora. *Anaerobe* **11**: 247–251.

Sexton, J.A., Miller, J.L., Yoneda, A., Kehl-Fie, T.E., and Vogel, J.P. (2004) *Legionella pneumophila* DotU and IcmF Are Required for Stability of the Dot/Icm Complex. *Infect. Immun.* **72**: 5983–5992.

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.

Shanahan, P., O'Sullivan, D.J., Simpson, P., Glennon, J.D., and O'Gara, F. (1992) Isolation of 2,4-Diacetylphloroglucinol from a Fluorescent Pseudomonad and Investigation of Physiological Parameters Influencing Its Production. *Appl. Environ. Microbiol.* **58**: 353–358.

Sheng, L., Lv, Y., Liu, Q., Wang, Q., and Zhang, Y. (2013) Connecting type VI secretion, quorum sensing, and c-di-GMP production in fish pathogen *Vibrio alginolyticus* through phosphatase PppA. *Vet. Microbiol.* **162**: 652–662.

Shneider, M.M., Buth, S.A., Ho, B.T., Basler, M., Mekalanos, J.J., and Leiman, P.G. (2013) PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500**: 350–353.

Silby, M.W., Winstanley, C., Godfrey, S.A.C., Levy, S.B., and Jackson, R.W. (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol. Rev.* **35**: 652–680.

Silverman, J.M., Austin, L.S., Hsu, F., Hicks, K.G., Hood, R.D., and Mougous, J.D. (2011) Separate inputs modulate phosphorylation-dependent and -independent type VI secretion activation. *Mol. Microbiol.* **82**: 1277–1290.

Sorg, I., Wagner, S., Amstutz, M., Müller, S.A., Broz, P., Lussi, Y., *et al*. (2007) YscU recognizes translocators as export substrates of the *Yersinia* injectisome. *EMBO J.* **26**: 3015–3024.

Sory, M.P., Boland, A., Lambermont, I., and Cornelis, G.R. (1995) Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the cyaA gene fusion approach. *Proc. Natl. Acad. Sci.* **92**: 11998–12002.

Sperandio, D., Decoin, V., Latour, X., Mijouin, L., Hillion, M., Feuilloley, M.G.J., *et al*. (2012) Virulence of the *Pseudomonas fluorescens* clinical strain MFN1032 towards *Dictyostelium discoideum* and macrophages in relation with type III secretion system. *BMC Microbiol.* **12**: 223.

Spreter, T., Yip, C.K., Sanowar, S., André, I., Kimbrough, T.G., Vuckovic, M., *et al*. (2009) A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nat. Struct. Mol. Biol.* **16**: 468–476.

Srinivasa Rao, P.S., Lim, T.M., and Leung, K.Y. (2003) Functional Genomics Approach to the Identification of Virulence Genes Involved in *Edwardsiella tarda* Pathogenesis. *Infect. Immun.* **71**: 1343–1351.

Srinivasa Rao, P.S., Yamada, Y., Tan, Y.P., and Leung, K.Y. (2004) Use of proteomics to identify novel virulence determinants that are required for *Edwardsiella tarda* pathogenesis. *Mol. Microbiol.* **53**: 573–586.

Stavrinides, J., Ma, W., and Guttman, D.S. (2006) Terminal Reassortment Drives the Quantum Evolution of Type III Effectors in Bacterial Pathogens. *PLoS Pathog* **2**: e104.

Suarez, G., Sierra, J.C., Sha, J., Wang, S., Erova, T.E., Fadl, A.A., *et al*. (2008) Molecular characterization of a functional type VI secretion system from a clinical isolate of *Aeromonas hydrophila*. *Microb. Pathog.* **44**: 344–361.

Subtil, A., Parsot, C., and Dautry-Varsat, A. (2001) Secretion of predicted Inc proteins of *Chlamydia pneumoniae* by a heterologous type III machinery. *Mol. Microbiol.* **39**: 792–800.

Sukhan, A., Kubori, T., Wilson, J., and Galán, J.E. (2001) Genetic Analysis of Assembly of the *Salmonella enterica* Serovar Typhimurium Type III Secretion-Associated Needle Complex. *J. Bacteriol.* **183**: 1159–1167.

Sun, J. and Barbieri, J.T. (2003) *Pseudomonas aeruginosa* ExoT ADP-ribosylates CT10 Regulator of Kinase (Crk) Proteins. *J. Biol. Chem.* **278**: 32794–32800.

Sun, K., Bröms, J., Lavander, M., Gurram, B.K., Enquist, P.-A., Andersson, C.D., *et al*. (2014) Screening for inhibition of the *Vibrio cholerae* VipA-VipB interaction identifies small molecule compounds active against type VI secretion. *Antimicrob. Agents Chemother.* AAC.02819–13.

Tan, K.S., Chen, Y., Lim, Y.-C., Tan, G.-Y.G., Liu, Y., Lim, Y.-T., *et al*. (2010) Suppression of Host Innate Immune Response by *Burkholderia pseudomallei* through the Virulence Factor TssM. *J. Immunol.* **184**: 5160–5171.

Thijs, I.M.V., Keersmaecker, S.C.J.D., Fadda, A., Engelen, K., Zhao, H., McClelland, M., *et al*. (2007) Delineation of the *Salmonella enterica* Serovar

Typhimurium HilA Regulon through Genome-Wide Location and Transcript Analysis. *J. Bacteriol.* **189**: 4587–4596.

Thorslund, S.E., Ermert, D., Fahlgren, A., Erttmann, S.F., Nilsson, K., Hosseinzadeh, A., *et al*. (2013) Role of YopK in *Yersinia pseudotuberculosis* Resistance against Polymorphonuclear Leukocyte Defense. *Infect. Immun.* **81**: 11–22.

Troisfontaines, P. and Cornelis, G.R. (2005) Type III Secretion: More Systems Than You Think. *Physiology* **20**: 326–339.

Tummler, B., Wiehlmann, L., Klockgether, J., and Cramer, N. (2014) Advances in understanding *Pseudomonas. F1000Prime Rep.* **6**:

Vallis, A.J., Finck-Barbancon, V., Yahr, T.L., and Frank, D.W. (1999) Biological Effects of *Pseudomonas aeruginosa* Type III-Secreted Proteins on CHO Cells. *Infect. Immun.* **67**: 2040–2044.

Veenendaal, A.K.J., Hodgkinson, J.L., Schwarzer, L., Stabat, D., Zenk, S.F., and Blocker, A.J. (2007) The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Mol. Microbiol.* **63**: 1719–1730.

Veesler, D. and Cambillau, C. (2011) A Common Evolutionary Origin for Tailed-Bacteriophage Functional Modules and Bacterial Machineries. *Microbiol. Mol. Biol. Rev.* **75**: 423–433.

Viboud, G.I. and Bliska, J.B. (2001) A bacterial type III secretion system inhibits actin polymerization to prevent pore formation in host cell membranes. *EMBO J.* **20**: 5373–5382.

Wagner, S., Königsmaier, L., Lara-Tejero, M., Lefebre, M., Marlovits, T.C., and Galán, J.E. (2010) Organization and coordinated assembly of the type III secretion export apparatus. *Proc. Natl. Acad. Sci.* **107**: 17745–17750.

Wang, X., Wang, Q., Xiao, J., Liu, Q., Wu, H., Xu, L., and Zhang, Y. (2009) *Edwardsiella tarda* T6SS component evpP is regulated by esrB and iron, and plays essential roles in the invasion of fish. *Fish Shellfish Immunol.* **27**: 469–477.

Wang, Y., Sun, M., Bao, H., Zhang, Q., and Guo, D. (2013) Effective Identification of Bacterial Type III Secretion Signals Using Joint Element Features. *PLoS ONE* **8**: e59754.

Wang, Y., Zhang, Q., Sun, M., and Guo, D. (2011) High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles. *Bioinformatics* **27**: 777–784.

Weber, B., Hasic, M., Chen, C., Wai, S.N., and Milton, D.L. (2009) Type VI secretion modulates quorum sensing and stress response in *Vibrio anguillarum*. *Environ. Microbiol.* **11**: 3018–3028.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci.* **95**: 6578–6583.

Whitney, J.C., Chou, S., Russell, A.B., Biboy, J., Gardiner, T.E., Ferrin, M.A., *et al*. (2013) Identification, Structure, and Function of a Novel Type VI Secretion Peptidoglycan Glycoside Hydrolase Effector-Immunity Pair. *J. Biol. Chem.* **288**: 26616–26624.

Williams, S.G., Varcoe, L.T., Attridge, S.R., and Manning, P.A. (1996) *Vibrio cholerae* Hcp, a secreted protein coregulated with HlyA. *Infect. Immun.* **64**: 283–289.

Woese, C.R. (1987) Bacterial evolution. *Microbiol. Rev.* **51**: 221–271.

Xiao, Y., Heu, S., Yi, J., Lu, Y., and Hutcheson, S.W. (1994) Identification of a putative alternate sigma factor and characterization of a multicomponent regulatory cascade controlling the expression of *Pseudomonas syringae* pv. syringae Pss61 *hrp* and *hrmA* genes. *J. Bacteriol.* **176**: 1025–1036.

Yahr, T.L., Vallis, A.J., Hancock, M.K., Barbieri, J.T., and Frank, D.W. (1998) ExoY, an adenylate cyclase secreted by the *Pseudomonas aeruginosa* type III system. *Proc. Natl. Acad. Sci.* **95**: 13899–13904.

Yang, CH., Young, T., Peng, MY., and Weng, MC. (1996) Clinical spectrum of *Pseudomonas putida* infection. *J. Formos. Med. Assoc. Taiwan Yi Zhi* **95**: 754–761.

Yang, H., Tan, Y., Zhang, T., Tang, L., Wang, J., Ke, Y., *et al*. (2013) Identification of Novel Protein-Protein Interactions of *Yersinia pestis* Type III Secretion System by Yeast Two Hybrid System. *PLoS ONE* **8**: e54121.

Yang, L., Xu, B., He, W., and Zhang, L. (2014) The HrpW protein of *Lonsdalea quercina* N-5-1 has pectate lyase activity and is required for full bacterial virulence. *J. Basic Microbiol.* n/a–n/a.

Zahrl, D., Wagner, M., Bischof, K., Bayer, M., Zavecz, B., Beranek, A., *et al*. (2005) Peptidoglycan degradation by specialized lytic transglycosylases associated with type III and type IV secretion systems. *Microbiology* **151**: 3455–3467.

Zarivach, R., Deng, W., Vuckovic, M., Felise, H.B., Nguyen, H.V., Miller, S.I., *et al*. (2008) Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS. *Nature* **453**: 124–127.

Zhang, D., Souza, R.F. de, Anantharaman, V., Iyer, L.M., and Aravind, L. (2012) Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **7**: 1–76.

Zhang, W., Wang, Y., Song, Y., Wang, T., Xu, S., Peng, Z., *et al*. (2013) A type VI secretion system regulated by OmpR in *Yersinia pseudotuberculosis* functions to maintain intracellular pH homeostasis. *Environ. Microbiol.* **15**: 557–569.

Zheng, J., Ho, B., and Mekalanos, J.J. (2011) Genetic Analysis of Anti-Amoebae and Anti-Bacterial Activities of the Type VI Secretion System in *Vibrio cholerae*. *PLoS ONE* **6**: e23876.

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

Zoued, A., Durand, E., Bebeacua, C., Brunet, Y.R., Douzi, B., Cambillau, C., *et al*. (2013) TssK Is a Trimeric Cytoplasmic Protein Interacting with Components of Both Phage-like and Membrane Anchoring Complexes of the Type VI Secretion System. *J. Biol. Chem.* **288**: 27031–27041.

Chapter 2.

Title: Genomic analysis and experimental characterisation of the type VI secretion systems in *Pseudomonas* spp: novel clusters and putative effectors uncovered.

Publications associated with this chapter: Genomic analysis of the Type VI Secretion Systems in *Pseudomonas* spp.: novel clusters putative effectors uncovered.

Authors: M. Barret, F. Egan, E. Fargier, JP. Morrissey and F. O'Gara.

Abstract

Bacteria encode multiple protein secretion systems that are crucial for interaction with the environment and with hosts. In recent years, attention has focused on Type VI secretion systems (T6SSs), which are specialised transporters widely encoded in Proteobacteria. The myriad processes associated with these secretion systems could be explained by subclasses of T6SS each involved in specialised functions. To assess diversity and predict functions associated with different T6SSs, comparative genomic analysis of 34 *Pseudomonas* genomes was performed. This identified 70 T6SSs, with at least one locus in every strain, except for *P. stutzeri* A1501. By comparing 11 core internal proteins of the T6SS, it was possible to identify 5 main *Pseudomonas* phylogenetic clusters with strains typically carrying T6SS from more than one clade. Separate phylogenetic analysis of VgrG and Hcp, two core external proteins, yielded similar phylogenetic division. Using a combination of phylogenetic and meta-analysis of transcriptome datasets it was possible to associate specific subsets of VgrG and Hcp proteins to each *Pseudomonas* T6SS clade. Moreover, a closer examination of the genomic context of *vgrG* genes in multiple strains highlights a number of additional genes associated with these regions, which may represent new T6SS effectors. To understand the relevance of the presence of phylogenetically distinct T6SS families several virulence assays were undertaken. Experimental evidence herein shows that T6SS belonging to the same phylogenetic branches are playing different roles for different *Pseudomonas* strains.

Introduction

Gram negative bacteria rely on several secretion systems to influence their environment by translocating protein and DNA into host cells and the extracellular milieu. These secretion systems can range from simple transporters to multi-component complexes. Type III (T3SS), Type IV (T4SS) and, more recently, Type VI (T6SS) secretion systems have received considerable attention because they are specialised in mediating the delivery of effectors directly into the host cytoplasm via a needle-like apparatus.

The Type VI secretion machinery is the product of approximately 15 conserved genes which are generally found together inside a genomic locus (Cascales, 2008). The mechanism of T6S has yet to be fully elucidated but a putative model of T6SS assembly has been proposed (Bönemann *et al*., 2010). Briefly, stacked, tubular hemolysin coregulated protein (Hcp) hexamers form a 4 nm wide conduit for the passage of effectors from the cytoplasm to the environment or into another host cell (Mougous *et al*., 2006). Hcp proteins interact with valine glycine repeat (Vgr) proteins, which could pierce the outer bacterial membranes with the help of the TssE protein, which potentially disrupts the peptidoglycan layer (Bönemann *et al*., 2010). Outgrowth of the Hcp tubules is energised by the ATPase ClpV, which produces a conformational change by disassembling the IglA-IglB complex surrounding the Hcp rings (Bröms *et al*., 2009; Bönemann *et al*., 2009). The proteins IcmF and DotU act as associated inner membrane-spanning structural proteins that anchor the secretion system in the cell membrane (Zheng and Leung, 2007; Ma *et al*., 2009), whereas the lipoprotein TssJ extends into the periplasm from the outer membrane and interacts with IcmF (Aschtgen *et al*., 2008). Although part of the T6S apparatus, Hcp and VgrG proteins are also secreted by bacteria with a functional T6SS. Some VgrG and Hcp proteins, called evolved VgrG (Pukatzki *et al*., 2009) or Hcp (Blondel *et al*., 2009), have a C-terminal domain extension and therefore could also act as effectors (Pukatzki *et al*., 2007; Ma and Mekalanos, 2010; Suarez *et al*., 2010). Interestingly, numerous *hcp* and *vgrG* paralogues are scattered around the bacterial chromosome. This raises the questions as to how do Hcps and VgrGs evolve, from where are they acquired, and whether all of them are secreted?

T6SS are widespread in Proteobacteria, particularly among Gammaproteobacteria (Shrivastava and Mande, 2008), and are more frequent than T3SS and T4SS in marine

isolates (Persson *et al*., 2009). Like T3SS and T4SS, several findings suggest that T6SS have been acquired through horizontal gene transfer (HGT) (Bingle *et al*., 2008). Indeed, the T6SS gene loci are frequently found inside genomic islands gained by HGT. Moreover, some T6S apparatus proteins such as Hcp and VgrG exhibit structural homology to phage tail-associated proteins, which suggests a common ancestral origin (Mougous *et al*., 2006; Leiman *et al*., 2009; Pell *et al*., 2009). Hence, it is believed that these T6SS genomic islands have been spread among bacteria by bacteriophages. Interestingly, some T6SS proteins are still able to interact with early phage protein, as demonstrated for Fha2 (PA1665) of *Pseudomonas aeruginosa* PAO1 (Roucourt *et al*., 2009). A recent phylogenetic analysis performed on T6SS core components across a range of bacterial species has shown that the T6SS loci can be divided into 5 clusters (Boyer *et al*., 2009). Although these phylogenetic clusters have probably evolved to adapt to various environments, it is difficult to find any correlation between clusters and ecological niches (Schwarz, West, *et al*., 2010).

The presence of multiple T6SS clusters in individual bacterial strains suggests that these secretion systems perform different roles for the bacterial cell (Bingle *et al*., 2008). Several phenotypes such as increased or attenuated virulence against human cells (Parsons and Heffron, 2005; Pukatzki *et al*., 2007; Robinson *et al*., 2009; Suarez *et al*., 2010), animals (Potvin *et al*., 2003; Burtnick *et al*., 2010), plants (Liu *et al*., 2008; Wu *et al*., 2008; Lesic *et al*., 2009), fish (Wang *et al*., 2009) and bacteria (Hood *et al*., 2010; MacIntyre *et al*., 2010; Schwarz, Hood, *et al*., 2010; Schwarz, West, *et al*., 2010) have been associated with the T6SS. More general physiological roles, such as biofilm formation (Southey-Pillig *et al*., 2005; Aschtgen *et al*., 2008) and quorum-sensing regulation (Weber *et al*., 2009) have also been linked to T6SS. Different T6SS in a given strain may secrete different sets of effectors, or, as with the T3SS (Cornelis, 2006), the myriad of processes associated with T6SS could also be explained by secretion of specific subsets of effectors by one T6SS. Thus far, these hypotheses cannot be tested as only a few T6S effectors have been identified (Zheng and Leung, 2007; Hood *et al*., 2010).

This chapter focuses on bioinformatics and experimental approaches to elucidate the functions of the T6SS within the genus *Pseudomonas*. Pseudomonads have a remarkable ecological and metabolic diversity, and are of interest as agents of plant disease (*P. syringae*), plant growth promotion (*P. fluorescens*) or bio-remediation (*P.*

putida). Reflecting the importance of this genus, many *Pseudomonas* genome sequencing projects are currently being undertaken or have been recently completed. As previous *in silico* studies included relatively few *Pseudomonas* genomes, this chapter identifies T6SS loci in all 34 *Pseudomonas* species sequenced to date and establishes their evolutionary relationship. *P. aeruginosa* has emerged as one model organism for T6SS studies (Filloux *et al*., 2008). *P. aeruginosa* possess 3 different loci named HSI-I to III which perform different functions (Mougous *et al*., 2006; Lesic *et al*., 2009). Whereas HSI-I may be involved in interbacterial interactions through secretion of Tse2 (Hood *et al*., 2010), HSI-II and III could be linked to virulence towards animals and plants (Lesic *et al*., 2009). The correlation between evolutinary relatonship and fucntion was assessed using several experimental models. Through a combination of phylogeny and transcitpome analysis, the association of *vgrG* and *hcp* paralogues with each T6SS cluster was assessed. Furthermore, a suite of genes associated with these paralogues was identified, and possibly includes new T6S effectors.

Methods

Bioinformatic procedures

Pseudomonas *genomic data acquisition*

Information about the current status of *Pseudomonas* genome sequencing projects was obtained from the Genomes Online Database (GOLD, updated on April 7, 2010). Among these sequencing projects, 33 having sequence data publicly available were selected for genome identification of T6SS loci. A preliminary draft of the strain *P. fluorescens* F113 (R. Rivilla, D. Dowling and F. O'Gara, unpublished) was also included in this analysis. The genomes analysed covered 8 different species. Genome accession numbers and information about the *Pseudomonas* genome sequencing projects utilised in this work are detailed in **Table 2.3**.

In silico *identification of T6SS loci*

Nucleotide and amino acid sequences of ORFs representing T6SS components were obtained from public sequence databases. The ORFs of *P. aeruginosa* PAO1 (Stover *et al.*, 2000), and "outgroup species" representing each branch in T6SS phylogenetic trees previously described (Blondel *et al*., 2009; Boyer *et al*., 2009) were used as baits in sequential BLASTN, BLASTX and BLASTP searches to identify homologues in the 34 *Pseudomonas* genomic sequences. A T6SS locus was defined as a gene cluster encoding at least five core components. A systematic analysis of gene content and gene architecture of the identified T6SS gene clusters was performed in 20 finished or permanent draft genomes. The remaining genomes correspond to unfinished projects containing more than one contig; therefore, an exhaustive ORF-by-ORF analysis of T6SS genetic architecture was not possible. In such cases, we only determined the presence/absence of T6SS core components in unassembled contigs.

Genomic islands analysis

The T6SS loci sequences of seven representative strains (**Table S2.1**) were examined for sequence composition bias such as aberrant G+C percentage or dinucleotide frequency. The dinucleotide frequency analysis calculates the genomic dissimilarity values δ* (the average dinucleotide relative abundance difference)

between T6SS loci sequences and the associated genome sequence using a web based application, δρ –web (Passel *et al*., 2005). GC% and dinucleotide frequencies of each T6SS locus were assessed in 5 kb windows and compared to the overall chromosomal signature. As the *Pseudomonas* genome is a flexible genome with numerous genes acquired by HGT (Gross and Loper, 2009), loci were arbitrarily defined as being of heterologous origin when the percentage of the genomic fragments with lower genomic dissimilarity was above 80% and when GC% was above or below 90% of the genomic fragments. Presence of insertion elements, flanking direct repeats, and proximity of tRNA was also assessed at the vicinity of T6SS loci.

Phylogenetic analyses

In the case of T6SS loci, the prevalence of each COG defined by Boyer and colleagues (Boyer *et al*., 2009) was analysed for every locus. COGs with frequencies higher than 90% were considered as *Pseudomonas* T6SS core components. Phylogenetic analyses were performed on the amino acids sequences from each of the 11 selected "core" COGs (COG0542, COG3455, COG3515, COG3516, COG3517, COG3518, COG3519, COG3520, COG3521, COG3522 and COG3523). Maximum-likelihood (ML) trees with 1000 bootstrap replicates were built with PhyML (Guindon and Gascuel, 2003) using the WAG amino acid substitution model of evolution (Whelan and Goldman, 2001) and four categories of substitution rates. To test the homogeneity between trees, the split distances were determined by TOPD/FMTS (Puigbò *et al*., 2007) then an average-linkage method of clustering was applied on these distances. In order to compute the super-tree, for each of the individual trees a Matrix Representation with Parsimony (MRP) was built using Mesquite V.2.72 then concatenated into a super matrix. The super-tree was built with the Pars program from PHYLIP (Felsenstein, 2005) with 1000 bootstrap replicates.

Beside this phylogenetic super tree construction, individual phylogenetic trees from 180 VgrG and 163 Hcp amino acids sequences were also generated by Maximum-likelihood by the method outlined above.

Comparative analysis of transcriptomes

All *P. aeruginosa* transcriptome datasets publicly available were retrieved from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) when available, or directly from publications. T6SS inducing or repressing conditions were defined for each locus when at least 50% of the genes within that locus were differentially regulated. Based on these criteria, 19 transcriptome datasets were considered (**Table S2.3**). Genes were clustered according to their expression profiles using the *MultiExperiment Viewer Software v4.5.1 (*http://www.tm4.org/mev/*).* Genes were grouped either with Pearson correlation or Euclidian distance as distance for K-means clustering. The optimal number of expression clusters was chosen after a *figure-of-merit* algorithm. For robustness, only genes present in expression clusters obtained with Euclidian distance and Pearson correlation were defined as T6SS co-regulated genes.

Experimental procedures

Bacterial strains and plasmids

Bacterial strains and plasmids are listed in **Table 2.1**. Bacterial strains were routinely cultured in Luria-Bertani (LB) medium in aerobic conditions (shaking at 200rpm). *P. aeruginosa* and *Escherichia coli* were grown at 37 °C and *P. fluorescens* F113 was grown at 30 °C.

Antibiotics were used at the following concentrations: gentamycin 30 $\mu g.ml^{-1}$, trimethoprim 800 $\mu g.ml^{-1}$ chloramphenicol $\mu g.ml^{-1}$ for *Pseudomonas* gentamycin 10 $\mu g.ml^{-1}$, kanamycin 50 $\mu g.ml^{-1}$ and chloramphenicol 30 $\mu g.ml^{-1}$ for *E. coli*.

Generation of mutants, complemented strains

The following procedure was employed to generate unmarked in-frame deletion of *tssB* genes in both F113 and PA14, and the *vgrG* islands PA14_43080-1000 and PA14_44900-44930. Upstream and downstream DNA fragments flanking the gene of interest were PCR-amplified (approximately 500 bp for *tssB* deletion and 800 bp for *vgrG* island deletion) and fused together by cross-over PCR using a complementary tag (**Table 2.2**). The resulting PCR product was cloned in the suicide plasmid

pEX18Gc (Hoang *et al*., 1998) digested with the appropriate restriction enzymes given in **Table 2.2**. The plasmids were introduced in F113 or PA14 by triparental mating and transconjugants were selected on LB gentamycin 30 $\mu g.ml^{-1}$. Single cross-over insertion in the chromosome was verified by PCR. Then, using LB supplemented with sucrose (10% w/v), bacterial strains were cultured overnight, serially diluted and plated. The second-cross-over event was verified by PCR amplification from the flanking regions.

To generate complementation constructs, DNA fragments containing the complete *PA14_43090* and F113 *tssB3* coding regions as well as their predicted endogenous ribosome binding sites were amplified using primers listed in **Table 2.2**. The DNA fragments were ligated into the *Xho*1-*Eco*R1 digested pBBR1MCS-5 or pBBR1MCS-1 plasmids (Kovach *et al*., 1995) and introduced in the corresponding mutant strains (**Table 2.2**).

Plasmid and genomic DNA isolations, electrophoresis and restriction enzyme digestions were performed using standard protocols (Green, 2012). Polymerase chain reactions were carried out using GoTaq® Green Master Mix or Pfu DNA polymerase using manufacturer's recommended conditions (Promega, Madison, USA). Plasmids were mobilised into *Pseudomonas* by triparental mating using the helper plasmid pRK2013 (Figurski and Helinski, 1979). Nucleotide sequencing was carried out by *GATC* Biotech *(Konstanz, Germany).*

Table 2.1 Strains and plasmids used in this study

Strains/Plasmids	Characteristics	References
E. coli DH5α	Lab strain	Green, 2012
E. coli HB101	Lab strain	Green, 2012
P. fluorescens F113	Sugarbeet , Ireland	(Shanahan *et al*., 1992)
P. aerugionsa PA14	Clinical isolate	(He *et al*., 2004)
P. fluorescens F113 Δ*tssB1*	In-frame deletion of *tssB1*	This study
P. fluorescens F113 Δ*tssB2*	In-frame deletion of *tssB2*	This study
P. fluorescens F113 Δ*tssB3*	In-frame deletion of *tssB3*	This study
P. fluorescens F113 Δ*tssB1-3*	In-frame deletion of *tssB1*, *tssB2* and *tssB3*	This study
P. aerugionsa PA14 Δ*tssB2*	In-frame deletion of *tssB2*	This study
P. aerugionsa PA14 Δ*tssB3*	In-frame deletion of *tssB3*	This study
P. aerugionsa PA14 Δ*tssB2-3*	In-frame deletion of *tssB2* and *tssB3*	This study
P. aerugionsa PA14 ΔPA14_43080-100	In-frame deletion of PA14_43080-100	This study
P. aerugionsa PA14 ΔPA14_44900-30	In-frame deletion of PA14_44900-30	This study
pRK2013	Helper plasmid, Mob+, Tra+, KmR	Figurski and Helinski, 1979
pBBR1MCS-1	Broad host range vector, chloramphenicol resistance	Kovach *et al*., 1995
pBBR1MCS-5	Broad host range vector, gentamicin resistance	Kovach *et al*., 1995
pBBR1MCS-5_*tssB3*	pBBR1MCS5 carrying the *tssB3* coding region	This study
pBBR1MCS-1_PA14_43090	pBBR1MCS1 carrying thevPA14_43090 coding region	This study
pMS402	Promoterless *luxCDABE* vector, TpR KmR	Duan *et al*., 2003
pMS402-tssA1_F113	pMS402 with the F113 *tssA1* predicted promoter	This study
pMS402-tssA2_F113	pMS402 with the F113 *tssA2* predicted promoter	This study
pMS402-tssB3_F113	pMS402 with the F113 *tssB3* predicted promoter	This study
pMS402-tssA1_PA14	pMS402 with the PA14 *tssA1* predicted promoter	This study
pMS402-tssA2_PA14	pMS402 with the PA14 *tssA2* predicted promoter	This study
pMS402-tssB3_PA14	pMS402 with the PA14 *tssB3* predicted promoter	This study
Tn:TssM1	PA14 with transposon insert in *tssM* from HSI-I	(Liberati *et al*., 2006)
Tn:TssM2	PA14 with transposon insert in *tssM* from HSI-II	(Liberati *et al*., 2006)
Tn:TssM3	PA14 with transposon insert in *tssM* from HSI-III	(Liberati *et al*., 2006)
Tn:03320	PA14 with transposon insert in PA14_03320	(Liberati *et al*., 2006)
Tn:18970	PA14 with transposon insert in PA14_18970	(Liberati *et al*., 2006)
Tn:29390	PA14 with transposon insert in PA14_29390	(Liberati *et al*., 2006)
Tn:33970	PA14 with transposon insert in PA14_33970	(Liberati *et al*., 2006)
Tn:33980	PA14 with transposon insert in PA14_33980	(Liberati *et al*., 2006)
Tn:43090	PA14 with transposon insert in PA14_43090	(Liberati *et al*., 2006)
Tn:44900	PA14 with transposon insert in PA14_44900	(Liberati *et al*., 2006)
Tn:67230	PA14 with transposon insert in PA14_67230	(Liberati *et al*., 2006)
Tn:69520	PA14 with transposon insert in PA14_69520	(Liberati *et al*., 2006)
Tn:69550	PA14 with transposon insert in PA14_69550	(Liberati *et al*., 2006)

Table 2.2. *Primers used in this study*

Name	Sequence (5' to 3')	Purpose
tssB1upFF_*Bam*H1	GGATCCAAGCCATCGAACAACAGGTC	*tssB1* upstream fragment of F113
tssB1upRF	AGCAGGCCCAGCAACAGGTCACGTTGCTGCTCCTGGTC	*tssB1* upstream fragment of F114
tssB1downFF	ACCTGTTGCTGGGCCTGCTAAGAAATCATCATGAAGGC	*tssB1* downstream fragment of F113
tssB1downRF_*Xba*1	TCTAGATCTTGAAGATCGGGCTCTG	*tssB1* downstream fragment of F113
tss1BflankFF	GCCTGACGTCCGATATCACC	validation of tssB1 deletion
tss1BflankRF	GTGGCGACTGGTCGAAGTAG	validation of tssB1 deletion
tssB2upFF_*Eco*R1	GAATTCGGGTTATCGAATTGCGCTGG	*tssB2* upstream fragment of F113
tssB2upRF	AGCAGGCCCAGCAACAGGTGCCTTCTTTGGCCATGACT	*tssB2* upstream fragment of F113
tssB2downFF	ACCTGTTGCTGGGCCTGCTTACTGGATGAGCTGGGCCT	*tssB2* downstream fragment of F113
tssB2downRF_*Xba*1	TCTAGACTGACCGTATTCAGCGGTGT	*tssB2* downstream fragment of F113
tss2BflankFF	ACAGTCTCCACCAAGGCGA	validation of tssB2 deletion
tss2BflankRF	GCGACACTGGAGACGTACTG	validation of tssB2 deletion
tssB3upFF_*Bam*H1	GGATCCGTTATAGCTGGCGAAGACCA	*tssB3* upstream fragment of F113
tssB3upRF	AGCAGGCCCAGCAACAGGTGACTCTGTCGAGCTTGTGC	*tssB3* upstream fragment of F113
tssB3downFF	ACCTGTTGCTGGGCCTGCTCCTGATCGCGCGCATCAT	*tssB3* downstream fragment of F113
tssB3downRF_*Xba*	TCTAGAGCGCTCTGGTCGAATTCG	*tssB3* downstream fragment of F113
tss3BflankFF	TCATCTCGCGACTGTCGAAC	validation of *tssB3* deletion
tss3BflankRF	GTGTAGTCACCCACCAGCAG	validation of *tssB3* deletion
tssB2upFP_*Eco*R1	GAATTCCCCATCTCCAGAACGACAGT	*tssB2* upstream fragment of PA14
tssB2upRP	AGCAGGCCCAGCAACAGGTGCTACCGAGCCTTCTTTGG	*tssB2* upstream fragment of PA14
tssB2downFP	ACCTGTTGCTGGGCCTGCTTGAGCCACCCCTAGCCAAG	*tssB2* downstream fragment of PA14
tssB2downRP_*Xba*1	TCTAGACACCGAACTGGCCGTATTCG	*tssB2* downstream fragment of PA14
tss2BflankFP	CTGGATCAGCGTCCATGTC	validation of tssB2 deletion
tss2BflankRP	GCTGGAGACGTATTGCATCA	validation of tssB2 deletion
tssB3upFP_*Bam*H1	GGATCCACCAGCTCCAGGCTCCATAC	*tssB3* upstream fragment of PA14
tssB3upRP	AGCAGGCCCAGCAACAGGTCAGCTTGTGCTGCGTACTCT	*tssB3* upstream fragment of PA14
tssB3downFP	ACCTGTTGCTGGGCCTGCTGCCTGCAGGAAATCAAAGC	*tssB3* downstream fragment of PA14
tssB3downRP_*Xba*1	TCTAGAGGTCGTTCTGCAGTTCGTTC	*tssB3* downstream fragment of PA14

tss3BflankFP	GAACGACTCCGACTCGATG	validation of *tssB3* deletion
tss3BflankRP	TCGAACTCGACCGCCTTCTC	validation of *tssB3* deletion
pmstssA1FF_*Xho*1	CTCGAGCAAGCGATAAGTACTACACCAAA	F113 *tssA1* promoter fusion
pmstssA1RF_*Bam*H1	GGATCCACATCCACTTGGATCGATCTC	F113 *tssA1* promoter fusion
pmstssA2FF_*Xho*1	CTCGAGTCACATTGAAGCAGTCTTGAGG	F113 *tssA2* promoter fusion
pmstssA2RF_*Bam*H1	GGATCCTTGTGAGATGGGCAGTTTTG	F113 *tssA2* promoter fusion
pmstssB3FF_*Xho*1	CTCGAGCTTCCTTTTCTCAGCGGTCA	F113 *tssA3* promoter fusion
pmstssB3FF_*Bam*H1	GGATCCATCGCATTCCCTTGCTGAT	F113 *tssA3* promoter fusion
pmstssA1FP_*Xho*1	CTCGAGCCGAGGGATTTCGGTTC	PA14 *tssA1* promoter fusion
pmstssA1RP_*Bam*H1	GGATCCGGCGATGCGTTCGAGTT	PA14 *tssA1* promoter fusion
pmstssA2FP_*Xho*1	CTCGAGCCTCGGGCAACTTCA	PA14 *tssA2* promoter fusion
pmstssA2RP_*Bam*H1	GGATCCAAGGCGTCGGCTGAT	PA14 *tssA2* promoter fusion
pmstssB3FP_*Xho*1	CTCGAGACAGCTCCGCATCCTTC	PA14 *tssA3* promoter fusion
pmstssB3FP_*Bam*H1	GGATCCCCCTGATCATCGCAGTC	PA14 *tssA3* promoter fusion
vgrG43uF_*Eco*R1	GAATTCAGTTGCTCCTTGATCGGTAC	upstream fragment of *vgrG* (PA14_43080) island
vgrG43uR	AGCAGGCCCAGCAACAGGTCACCTAACAATTCTCATGC	upstream fragment of *vgrG* (PA14_43080) island
vgrG43downF	ACCTGTTGCTGGGCCTGCTCATCGCCGAAGTGGATAACC	downstream fragment of *vgrG* (PA14_43080) island
vgrG43downR_*Xba*1	TCTAGAGCTTCGTAATGCTTTCTGTCG	downstream fragment of *vgrG* (PA14_43080) island
43flaF	GCGGTATCGAGTTTTTCCAG	validation of *vgrG* (PA14_43080) island deletion
43flaR	CGAGGCAGTGATTTCTGTCA	validation of *vgrG* (PA14_43080) island deletion
vgrG44uF_*Eco*R1	GAATTCGCTGGTGAAAGCCATCGAAG	upstream fragment of *vgrG* (PA14_44900) island
vgrG44uR	AGCAGGCCCAGCAACAGGTACTATCACACGACGCAGCAC	upstream fragment of *vgrG* (PA14_44900) island
vgrG44downF	ACCTGTTGCTGGGCCTGCTCGTTCGATATGCCGAACGAC	downstream fragment of *vgrG* (PA14_44900) island
vgrG44downR_*Xba*1	TCTAGATCGGCGTTGATCAGGAAG	downstream fragment of *vgrG* (PA14_44900) island

44flaF	TCATGGTCGGCAACGTG	validation of *vgrG* (PA14_44900) island deletion
44flaR	CCCATCATCACCGGCATG	validation of *vgrG* (PA14_44900) island deletion
tssB3mcs5F_*Xho*1	CTCGAGACGGTGATCCGTACCACCGAACG	tssB3 complementation
tssB3mcs5R_*Eco*R1	GAATTCCGAGGGTCTCGTGCTGGTT	tssB3 complementation
43090mcs1F_*Xho*1	CTCGAGAACGTCCGGGTCAACTC	PA14_43090 complementation
43090mcs1R_*Eco*R1	GAATTCTCATGGGGTCAGGTCCT	PA14_43090 complementation

Construction of promoter gene fusions and promoter fusion analysis

The promoter regions *of tssA1* and *tssA2* were predicted using BPROM (Softberry Inc., NY, USA). As no predicted promoter was found upstream of *tssA3* in either PA14 or F113, the promoter region of *tssB3* was used instead of *tssA3* for promoter fusions assays monitoring the expression of this T6SS locus. These predicted promoter regions were amplified by PCR using genomic DNA as template sequence. These were ligated into the *Bam*HI and *Xho*I digested plasmid pMS402 (Duan *et al*., 2003) which contains the promoterless *luxCDABE* operon, and introduced to the F113 or PA14 wild type strains.

For measuring promoter activity cells from overnight cultures were inoculated to a final optical density at 600 nm of 0.05 in LB medium. Luminescence from promoter-*lux* reporter fusions was measured on a Tecan GENios microplate reader. All the data presented in this chapter are mean of three independent biological replicates.

Infection of Galleria mellonella

Galleria mellonella larvae were obtained from Livefood UK (Axbridge, England) and stored in the dark until use. Bacterial strains from overnight cultures were washed twice in PBS and resuspended in PBS and serially diluted to an OD_{600nm} of 0.000001. Ten µL of bacterial suspension (approximately 10 bacterial cells) were injected into the haemocoel using a sterile Hamilton syringe and a 30 gauge disposable needle. Insects were placed in a 9.0 cm Petri dishes lined with 8.5 cm Whatman paper and incubated at 30°C in the dark. As a control 10 insects were injected with sterile PBS. Insects were individually examined for the production of pigmentation and the time

of death, when moths no longer responded to touch, was recorded until five days post-inoculation.

Caenorhabditis elegans *synchronization and virulence assays*

The bacterivorous nematode *Caenorhabditis elegans* wild-type Bristol strain N2 was obtained from the *Caenorhabditis* Genetics Center (Minneapolis, MN, USA). *C. elegans* were maintained under standard culturing conditions at 22°C on nematode growth medium (NGM: 3 $mg.ml^{-1}$ NaCl, 2.5 $mg.ml^{-1}$ peptone, 17 $mg.ml^{-1}$ agar, 5 $\mu g.ml^{-1}$ cholesterol, 1 mM $CaCl_2$, 1 mM $MgSO_4$, 25 mM KH_2PO_4, H_2O to 1 litre) agar plates with *E. coli* OP50 as a food source (Sulston and Hodgkin, 1988). Synchronous cultures of worms were generated after worm adult population exposure to a sodium hypochlorite/sodium hydroxide solution as previously described (Stiernagle, 1999). The resulting eggs were incubated at 22 °C on an *E. coli* OP50 lawn until the worms reached the L4 (48 hours) life stage (confirmed by light microscopy).

Bacterial lawns used for *C. elegans* survival assays were prepared by spreading 50 µl of *P. aeruginosa* strains on 35 mm NGM conditioned Petri dishes supplemented with 0.05 $mg.ml^{-1}$ 5-fluoro-2'-deoxyuridine. This nucleotide analogue blocks the development of the *C. elegans* next generation via the inhibition of DNA synthesis, thus preventing the offspring from the experimental animals. The plates were incubated overnight at 30 °C and then placed at room temperature for 4 h. Fifteen to twenty L4 synchronised worms were harvested with M9 solution (3 $mg.ml^{-1}$ KH_2PO_4, 6 $mg.ml^{-1}$ $NaHPO_4$, 5 $mg.ml^{-1}$ NaCl, and 1 mM $MgSO_4$, H_2O to 1 litre), placed on the 35 mm assay Petri dishes and incubated at 22 °C. Worm survival was scored at 1 hour, 24 hours and each subsequent day, using an Axiovert S100 optical microscope (Zeiss, Oberkochen, Germany) equipped with a Nikon digital Camera DXM 1200F (Nikon Instruments, Melville, NY, USA). The worms were considered dead when they remained static without grinder movements for 20 s. The results are expressed as the percentage of living worms. The results are the average of three independent biological replicates.

Epithelial cell internalisation

A549 cells were grown in minimal Eagle's medium (MEM) (Sigma-Aldrich, U. K.) supplemented with 10 % foetal bovine serum (FBS), 50 units.ml^{-1} pen-strep and 2 mM L-glutamine (both Gibco). Cells were incubated at 37 °C in a humidified 5% CO2 atmosphere. For infection assays, bacteria were grown overnight in standard conditions and then sub-cultured in infection medium (MEM supplemented with FBS only) shaking at 37 °C for 3 hours. For bacterial invasion bacteria were added at an infection (MOI) of 10:1. After 1 hour of infection, ceftazimide (1 mg.ml^{-1}) and gentamicin (2 mg.ml^{-1}) (both Sigma-Aldrich) were added to cells for a further 2 hours, following which cells were washed twice with PBS and lysed using 0.1 % Triton X-100. Intracellular bacteria were enumerated by serial dilutions and plate counts. Assays were carried out in duplicate and are the average of 3 biological reps. Error bars are given as standard deviation.

Dictyostelium discodeum *predation assay*

Dictyostelium discodeum were maintained on a lawn of *Klebsiella aerogenes* on SM agar (Formedium). For the predation assay *D. discodeum* were scraped off the plate and resuspended in HL5 media (Formedium). Approximately 100 *D. discodeum* cells were mixed with 300 µl of *K. aerogenes* at OD600nm 0.5 and either 15 µl of PA14 at OD600nm 1 to 1.0 X 10^{-4} or 15 µl F113 at OD600nm 5 to 5.0 X 10^{-2}. This mix was then spread on SM agar plates which were incubated at 23 °C for 5 days. For assays with the complemented strain, 15 µl of F113 was added at a concentration of OD600nm 0.5. Plaque formation was scored on a scale of 1-10 (arbitrary units).

Results

Identification of T6SS gene clusters in *Pseudomonas*

T6SS loci in 33 publically available *Pseudomonas* genomes and a draft genome of *P. fluorescens* F113 were identified by sequential TBLASTN, BLASTX and BLASTP searches using genes of the 3 T6SS loci in *P. aeruginosa* PAO1, and representatives of the T6SS core components as defined by Boyer and colleagues (Boyer *et al*., 2009) from non-Pseudomonad bacteria used in a previous phylogenetic study on T6SS (Blondel *et al*., 2009), as baits. T6SS loci were defined when at least 5 homologues to the bait sequences (e-value < 10^{-5}) were found in the same genomic location.

With the exception of *P. stutzeri*, all *Pseudomonas* strains analysed possess at least one putative T6SS locus (**Table 2.3**), with multiple T6SSs being present in 27 of the strains. A maximum of 3 T6SS exist in any one strain, as is the case for *P. fluorescens* F113 and all *P. aeruginosa* strains.

Table 2.3. T6SS clusters in *Pseudomonas*. The strain *P. fluorescens* F113 highlighted in bold has been sequenced recently and is currently under annotation. Asterisks: only presence or absence of T6SS loci could be performed in unfinished genomes. Therefore, we cannot rule out the possibility of the presence of additional T6SS clusters in these draft genomes.

Strains	**Status**	**NCBI Reference Sequence**	**Locus 1**	**Locus 2**	**Locus 3**	**Locus 4A**	**Locus 4B**	**Locus 5**	**Total**	**VgrG**	**Hcp**
P. aeruginosa PAO1	Complete	NC_002516	1	0	1	1	0	0	3	10	5
P. aeruginosa UCBPP-PA14	Complete	NC_008463	1	0	1	1	0	0	3	11	6
P. aeruginosa PA7	Complete	NC_009656	1	0	1	1	0	0	3	7	3
P. aeruginosa LESB58	Complete	NC_011770	1	0	1	1	0	0	3	9	5
P. aeruginosa 2192	Draft	NZ_AAKW00000000	1	0	1	1	0	0	3	7	5
P. aeruginosa C3719	Draft	NZ_AAKV000000000	1	0	1	1	0	0	3	9	5
P. aeruginosa PACS2	Draft	NZ_AAQW00000000	1	0	1	1	0	0	3	9	4
P. aeruginosa PAb1	Draft	NZ_ABKZ000000000	1	0	1	1	0	0	3	6*	6*
P. entomophila L48	Complete	NC_008027	1	0	0	0	0	0	1	10	8
P. fluorescens Pf-5	Complete	NC_004129	0	0	1	0	0	0	1	3	1
P. fluorescens PfO-1	Complete	NC_007492	1	0	0	1	0	0	2	10	4
P. fluorescens SBW25	Complete	NC_012660	1	0	1	0	0	0	2	3	3
***P. fluorescens* F113**	**Draft**		**1**	**0**	**1**	**1**	0	**0**	**3**	**8**	**3**
P. mendocina ymp	Complete	NC_009439	1	0	1	0	0	0	2	4	2
P. putida KT2440	Complete	NC_002947	2	0	0	0	1	0	3	4	5
P. putida GB-1	Complete	NC_010322	1	0	0	0	1	0	2	4	5
P. putida W619	Complete	NC_010501	0	1	0	0	1	0	2	3	3
P. putida F1	Complete	NC_009512	1	0	0	0	1	0	2	3	2

Strains	**Status**	**NCBI Reference Sequence**	**Locus 1**	**Locus 2**	**Locus 3**	**Locus 4A**	**Locus 4B**	**Locus 5**	**Total**	**VgrG**	**Hcp**
P. stutzeri A1501	Complete	NC_009434	0	0	0	0	0	0	0	0	0
P.savastanoi pv. *savastanoi* NCPPB 3335	Draft	NZ_ADMI01000000	1	0	0	0	1	0	2*	3*	6*
P. syringae pv. *aesculi* 2250	Draft	NZ_ACXT00000000	0	0	0	0	1	0	1*	4*	3*
P. syringae pv. *aesculi* NCPPB3681	Draft	NZ_ACXS00000000	1	0	0	0	1	0	2*	1*	5*
P. syringae pv. *oryzae* 1_6	Draft	NZ_ABZR00000000	1	0	0	0	1	0	2*	4*	7*
P. syringae pv. *syringae* B728a	Complete	NC_007005	0	0	0	0	1	0	1	5	4
P. syringae pv. *syringae* FF5	Draft	NZ_ACXZ00000000	1	0	0	0	1	0	2*	2*	8*
P. syringae pv. *syringae* 642	Draft	NZ_ACXZ00000000	1	0	0	0	1	0	2*	4*	6*
P. syringae pv. *tabaci* 11528	Draft	NZ_ACHU00000000	1	0	0	0	1	0	2*	4*	6*
P. syringae pv. *tomato* DC3000	Complete	NC_004578	2	0	0	0	0	0	2	6	3
P. syringae pv. *tomato* T1	Draft	NZ_ABSM00000000	1	0	0	0	1	0	2	4	7
P. syringae pv. *tomato* K40	Draft	NZ_ADFY00000000	1	0	0	0	1	0	2*	4*	6*
P. syringae pv. *tomato* Max13	Draft	NZ_ADFZ00000000	1	0	0	0	1	0	2*	4*	6*
P. syringae pv. *tomato* NCPPB 1108	Draft	NZ_ADGA00000000	1	0	0	0	1	0	2*	4*	5*
Pseudomonas sp. UK_4	Draft	NZ_ACOQ00000000	1	0	0	0	0	0	1*	0*	0*

Pseudomonas T6SS phylogeny

Genes with homologues in a variety of organisms are often classified according to the COG database (Tatusov *et al*., 2000). 13 COGs were found in over 90% of *Pseudomonas* T6SS loci, and were considered to be the T6SS core components. This is in agreement with previous phylogenetic and experimental results (Zheng and Leung, 2007; Boyer *et al*., 2009). In order to study the evolutionary relationship between the T6SS of *Pseudomonas,* these COGs were used for phylogenetic analysis. As two of these COGs, COG3157 (Hcp) and COG3501 (VgrG), are often encoded in multiple copies within a genome far from any T6SS loci, and are "external" proteins which may come in contact with the target organism and thus be subject to higher evolutionary pressure, these COGs were analysed separately.

Individual trees were constructed for each of the core components in T6SS loci, and these were congruent according to split distances (TOPD/FMTS). A super-tree was then built from the individual trees of the 11 "internal" core components **(Figure 2.1).**

A previous phylogenetic analysis has split the T6SS into 5 clusters (Boyer *et al*., 2009). Notably, the subtree for cluster 4 from the previous study is not supported by our analysis, and is instead split into two clusters, which are termed cluster 4A and 4B for the sake of consistency with the previous nomenclature. Cluster 1 exhibits some divergence and can be divided into two subclusters, 1.1 and 1.2. T6SS of all cluster types are present in the *Pseudomonas* genus, except cluster 5 which in our tree is comprised solely of outgroup species. Previous independent analyses have already reported the presence of three T6SS clusters, termed cluster 1 (or HSI-II), cluster 3 (or HSI-I) and cluster 4 (or HSI-III) in *Pseudomonas* (Bingle *et al*., 2008; Boyer *et al*., 2009). However this is the first report of a T6SS belonging to cluster 2 in the *Pseudomonas* genus. This is found in *P. putida* W619, and it is also present in a *Pseudomonas*-related strain, uncultured proteobacterium QSI (Williamson *et al*., 2005).

Cluster 1 is the most common cluster in the analysed strains, where cluster 1.2 is exclusive to *P. putida*. Cluster 3, is found only in *P. fluorescens, P. mendocina* and *P. aeruginosa,* and it has been demonstrated to mediate interbacterial killing in the latter (Hood *et al*., 2010). The cluster 4A is encoded in *P. aeruginosa* and some *P. fluorescens* strains, whereas cluster 4B is found in *P. putida* and *P. syringae* species.

In our data, multiple T6SS loci within the same genome are usually of a different phylogenetic cluster, with only two strains, namely *P. syringae* pv. *tomato* DC3000 and *P. putida* KT2440 possessing two copies of the same T6SS cluster. This would suggest that these clusters play different functional roles, and possibly provide a competitive advantage in certain niches. Moreover, a closer examination of these T6SS families which occur more than once in a single genome reveals that two T6SS core genes in *P. putida* KT2440 (PP4083 and PP4075) have frameshift mutations while one core gene from *P. syringae* pv. *tomato* DC3000 (PSPTO_2542) is disrupted by a transposase, suggesting that in each case only one of the loci encodes a functional T6SS. This is consistent with T6SS of the same family playing redundant roles.

While analysing the data however, it should be considered that not every T6SS gene cluster identified *in silico* is necessarily functional. Indeed, some other *Pseudomonas* T6SS loci are also incomplete with multiple gene deletions. The most striking example comes from *P. fluorescens* SBW25 cluster 1.1, where five T6SSs core components are missing.

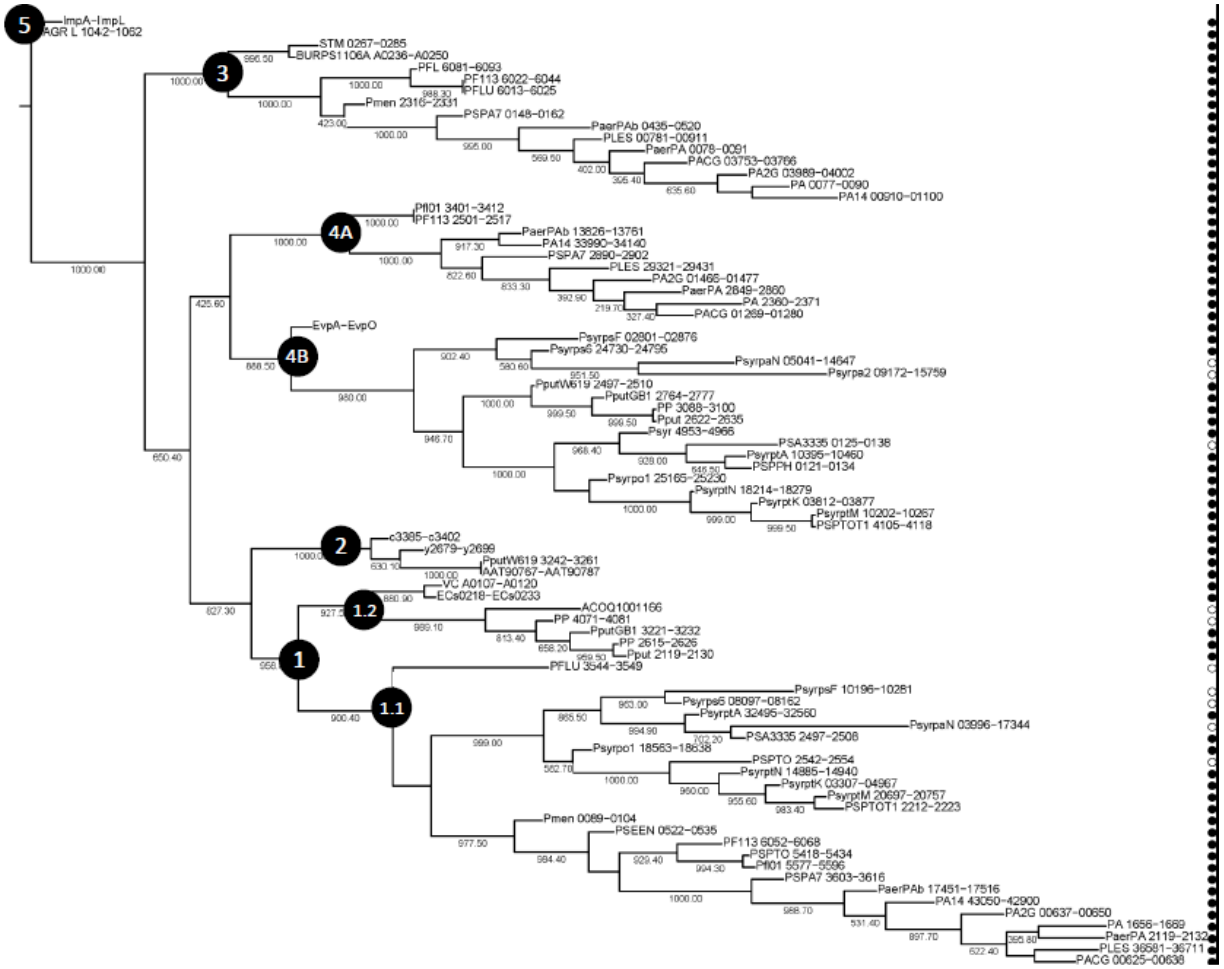


Figure 2.1. Phylogenetic distribution of T6SS clusters in *Pseudomonas* species. Maximum-likelihood trees with 1000 bootstrap replicates were built with PhyML for each 'core' protein. In order to compute the supertree, for each of the individual trees a matrix representation with parsimony (MRP) was built using Mesquite V.2.72 and then concatenated into a supermatrix. The supertree was built with the Pars program from PHYLIP with 1000 bootstrap replicates. T6SS cluster nomenclature (Boyer *et al*., 2009) is used to show the major phylogenetic clusters (1, 2, 3 and 5). However, cluster 4 was not supported as a single clade in this work and was therefore divided into 4A and 4B. The phylogenetic cluster 1 may be subdivided into two sub-clusters represented in the nodes labelled 1.1 and 1.2. Black circles to the right of the figure indicate a complete T6SS locus, whereas white circles represent T6SS loci in which at least one core component is missing or mutated.

The multiple T6SS loci that are present in some genomes are usually of different phylogenetic origins. This suggests that they are gained by independent horizontal gene transfer instead of gene duplication (Bingle *et al*., 2008). Hence, an analysis of genomic dissimilarities between *Pseudomonas* T6SS loci and the associated genomes was performed on seven representative strains using the program δρ (Passel *et al*., 2005). The dinucleotide frequencies and GC% of each *Pseudomonas* T6SS locus were similar to that of the genomes in which they presently reside **(Table S2.1**). Therefore, these T6SS loci have probably been subject to amelioration over an extended period of time.

Accessory gene context varies among *Pseudomonas* T6SS clusters

While evidence here and elsewhere suggests that there are 13 core components of the T6SS, additional genes are present in many T6SS loci **(Figure 2.2)**. Multiple accessory elements have been found in every *Pseudomonas* T6SS cluster. Some of these have known functions, such as the post-translational regulation roles of the Fha, PpkA and PppA proteins. Other non-core proteins could have roles in binding the T6SS apparatus to the membrane or cell wall. It has previously been reported that the core T6SS proteins are sometimes modified by the addition of a peptidoglycan-binding domain, or that a separate protein containing a peptidoglycan-binding domain can be found in some T6SS loci, and that this is correlated with T6SS phylogeny (Aschtgen, Thomas, *et al*., 2010). For example, the SciZ protein is a non-core T6SS protein which is necessary for efficient functioning of the particular T6SS where it is encoded (Aschtgen, Gavioli, *et al*., 2010). The phylogenetically related T6SS of *P. putida* W619 and Q also possesses a homologue of this gene, suggesting a specific role of this protein for cluster 2.

All cluster 1 loci encode the Fha protein, which is innolved in T6SS activation, but those of cluster 1.2 lack the phosphatase and kinase genes of this post-transregulatory mechanism (Mougous *et al*., 2007). This may mean the regulation is absent or altered in these bacteria, and indeed from BLASTP comparisons it is clear that the Fha proteins from cluster 1.1 are much more similar to the Fha proteins in cluster 3 than to the the Fha proteins in cluster 1.2.

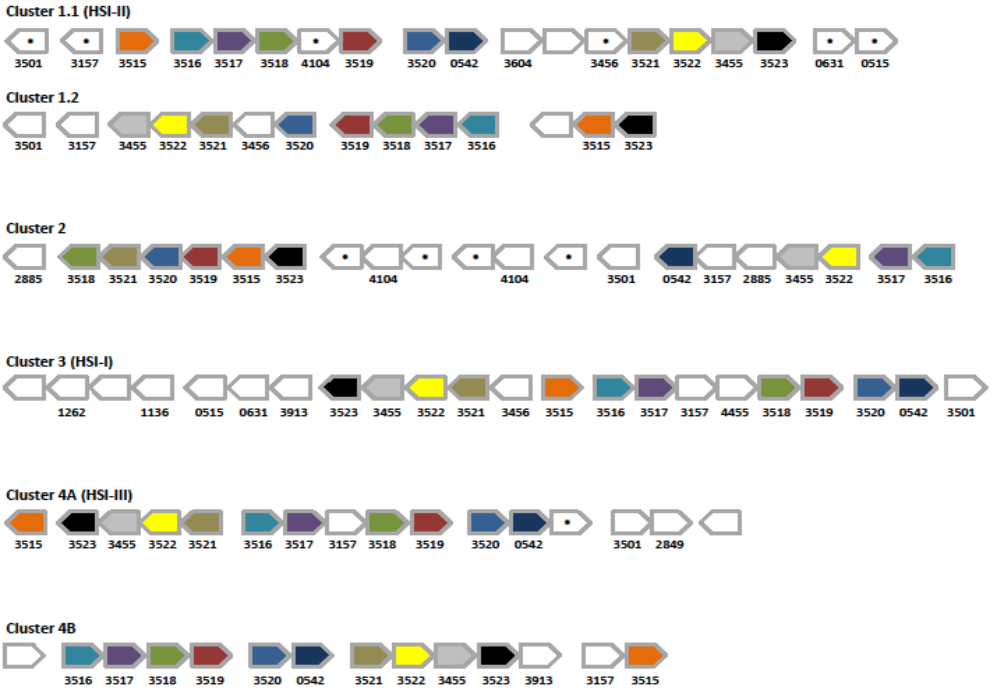


Figure 2.2. Genomic organization of the *Pseudomonas* T6SS clusters. Genes are represented as blocked arrows showing the direction of their transcription. Numbers represent the COG number. A unique colour is assigned to T6SS core component genes used in the phylogeny. Asterisks indicate that genes are not always conserved among all strains.

Some specific regulatory components of T6SSs (Bernard *et al.*, 2010) are only encoded in Pseudomonads. For instance, the posttranslational regulatory elements TagR which promotes efficient phosphorylation of the kinase PpkA leading to activation of HSI-I (Mougous *et al.*, 2007; Hsu *et al.*, 2009), is only present in cluster 3 of *P. aeruginosa* and *P. fluorescens*. There are also hypothetical proteins from various loci of the same family which have some degree of homology **(Table S2.2)**. In some cases these have predicted activity e.g. ADP-ribosylation, which may indicate that these are effector molecules. Taken together, these data suggest that each T6SS cluster encodes specialised secretion machinery, well conserved through evolution, but over time the various T6SS become optimised for expression or activation in the particular conditions where they are useful during the lifestyle of a particular bacterial species.

Phylogenetic relationship between Hcps/VgrGs and T6SS clusters

According to the current model of the T6SS, VgrG and Hcp proteins are believed to interact and form part of the extracellular secretion apparatus. A very interesting feature of the genetics of T6SS is the findings that, in addition to VgrG and Hcp genes being often found inside T6SS loci, many T6SS-possessing bacteria, including all *Pseudomonas* strains in this study, possess numerous VgrG and Hcp paralogues that are encoded elsewhere in the genome (**Table 2.3**). Whereas VgrG and Hcp proteins located inside T6SS locus are likely to be associated with their respective secretion systems, the role(s) of these "orphan" Hcp and VgrG proteins is less clear. It is not known whether they are recruited by specific T6SSs or whether they have other functions.

In order to ascertain whether "orphan" VgrG and Hcp proteins can be linked to specific T6SS clusters, the evolutionary relationships that exist among VgrG/Hcp paralogues were investigated. Phylogenetic analyses of 180 VgrGs and 163 Hcps from all *Pseudomonas* genomes as well as "outgroup species" representing each branch in T6SS phylogenetic trees were performed. Hcps of *Pseudomonas* can be divided into 6 clusters (1.1, 1.2, 2, 3, 4A, 4B). The naming of the clusters reflects that some Hcp are within T6SS loci and are assumed to belong to that T6SS system. The division of Hcp is very similar to the phylogenetic division of the T6SS (**Figure 2.3**), which suggests that every gene in each Hcp clade may be specifically recruited by the corresponding T6SS cluster. Such correlation between VgrGs and T6SS phylogenetic clusters is also clear for cluster 2, 3, 4A and 4B (**Figure 2.4**). However, a subset of orphan VgrGs, which group with cluster 1 are present in strains where this cluster is absent (e.g. PFL2812 from *P. fluorescens* Pf-5). This discrepancy might be due to the loss of cluster 1.1 from *P. fluorescens* Pf-5 or by an acquisition of just this *vgrG* through HGT.

According to the phylogenetic analysis of VgrGs and Hcps, it is believed that 3 VgrGs (VgrG1, PA0095 and VgrG4) and 1 Hcp (Hcp1) are associated with locus 3 (or HSI-I) whereas VgrG3 and Hcp3 are part of locus 4A (or HSI-III) of *P. aeruginosa*. Among these proteins, VgrG1, VgrG4 and Hcp1 have been already isolated from the supernatant of a *P. aeruginosa* strain with artificially induced cluster 3 T6SS (Mougous *et al*., 2006; Hood *et al*., 2010). The other orphan Hcps and VgrGs of *P. aeruginosa* seem to be related to locus 1.1 (HSI-I).

Orphan VgrGs and Hcps are frequently encoded together in the same genomic location, consistent with the view that they interact as part of the T6SS. In some cases, however, genes encoding these proteins were found singly, not in proximity to each other or a T6SS locus. A closer examination of the phylogenetic trees shows that Hcps without a partner VgrG often grouped together (**Figure 2.3**), whereas no such pattern is seen for VgrGs (**Figure 2.4**). Therefore, it is possible that these particular Hcp proteins are no longer part of a T6SS and are linked to other biological functions, and it is of note that the sequence of the HilE protein involved in regulation of gene expression in *Salmonella enterica* is somewhat similar to Hcp (Blondel *et al*., 2009).

The VgrG phylogeny also highlights three distinct groups of VgrG proteins containing C-terminal extensions. Because these extensions could possibly function as effector-domains, these VgrGs have been named evolved-VgrGs (Pukatzki *et al*., 2009). The first evolved-VgrG group contains PA0262 orthologues, which are only found in *P. aeruginosa* genomes. The second group is restricted to one VgrG of *P. fluorescens* Pf0-1 (Pfl01_2329) and five VgrGs of *P. entomophila* L48. Finally, the third group is shared by three proteins (Psyr_4974, Psyrpol_ 25120 and Psyrpa2_10189) each belonging to one *P. syringae* strain. A motif search was systematically performed with INTERPRO scan to predict a specific function for VrgG C-terminal extensions but none of the domains shares homology with domains associated with biological functions.

Cargo genes are associated with orphan *hcp-vgrG*

Some *vgrG* and *hcp* genes of *P. aeruginosa* PAO1 have been reported to be located inside genomic islands (Wilderman *et al*., 2001; Ernst *et al*., 2003). To see if this trend was applicable to all orphan *vgrG* and *hcp* genes of *P. aeruginosa* PAO1, the GC% and dinucleotide frequencies of each genomic region associated with *vgrG* or *hcp* were assessed in 5 kb windows and compared to the overall chromosomal signature **(Table 2.4)**. Five out of the ten *vgrG* regions analysed showed atypical sequence composition when compared to the genome in which they presently reside. Therefore, acquisition of these five "*vgrG* islands" by HGT seems to be relatively recent.

We reasoned that these "*vgrG* islands" could have been acquired along with other genes (hereinafter referred to as cargo genes). The VgrG phylogenetic analysis and the

conserved genome organization of *P. aeruginosa* were then used to determine whether orphan *vgrGs* were always linked to cargo genes. Remarkably, each orphan *vgrG* was associated with a set of specific cargo genes in all strains of *P. aeruginosa* carrying that *vrgG* gene **(Figure 2.5a)**. The function(s) of each cargo gene was predicted using BLASTP and INTERPRO scan. It was also confirmed using the gene neighbourhoods tool on IMG (http://img.jgi.doe.gov/) that many of these cargo genes are present in in T6SS loci or similar *vgrG* islands in other bacterial genera **(Figure 2.5b)**. Although most cargo genes encode hypothetical proteins or putative lipoproteins, genes encoding predicted lipases/esterases are also frequently linked to *vgrG* in these islands. These proteins are identified by the presence of a phospholipase D domain (e.g. PldA or PA5089) or a PGAP1 domain (e.g. PA1510).

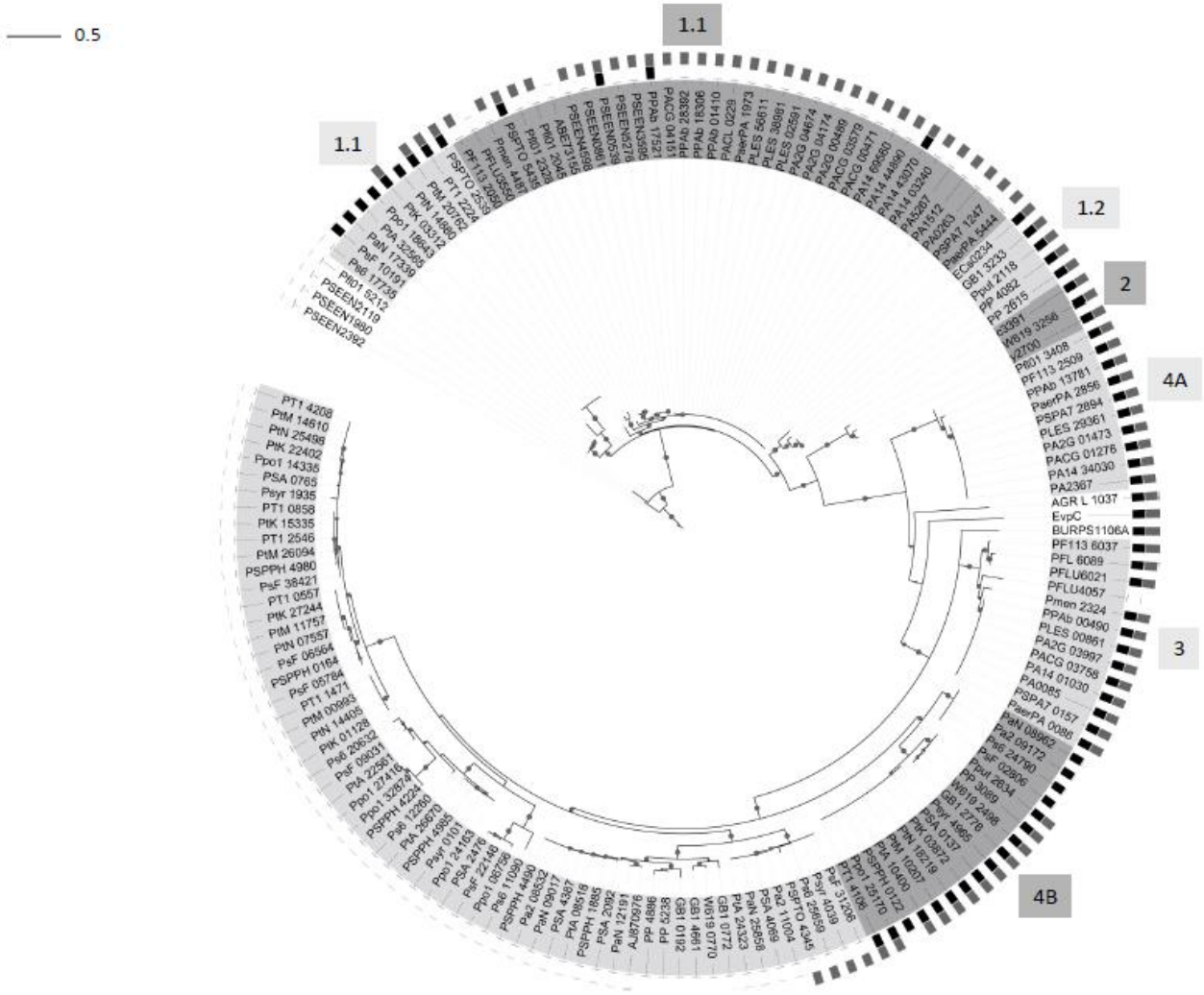


Figure 2.3. Phylogenetic distribution of Hcp proteins within *Pseudomonas* species. A distance tree (maximum-likelihood) was calculated from 163 Hcp protein sequences of *Pseudomonas* spp. Black circles indicate branches with bootstrap support values of 75% (1000 replicates). Black rectangles indicate that Hcp proteins are encoded inside a T6SS locus. Dark grey rectangles indicate that *hcp* is linked to *vgrG*. The scale bar indicates genetic distance.

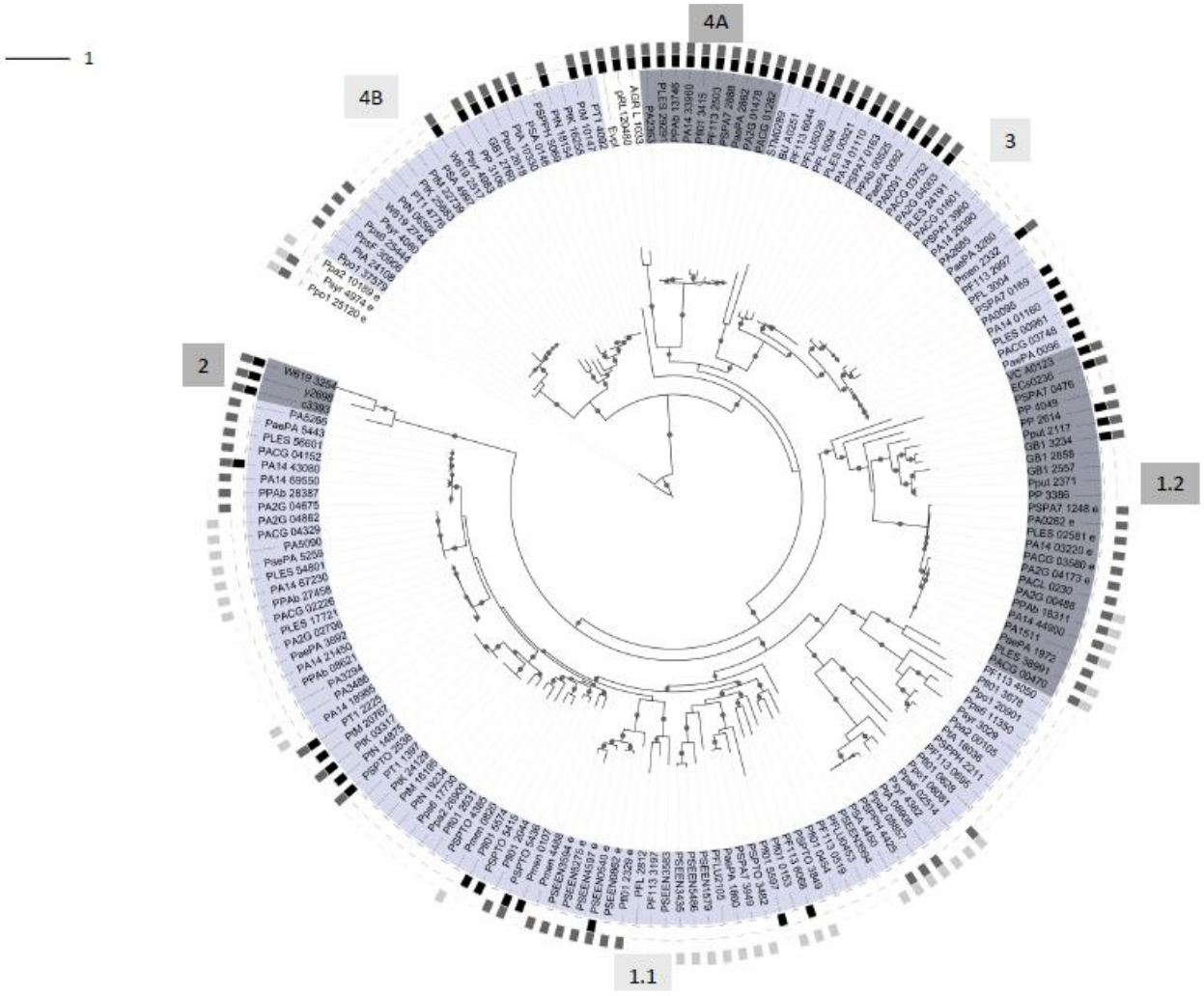


Figure 2.4. Phylogenetic distribution of VgrG proteins within *Pseudomonas* species. A distance tree (maximum-likelihood) was calculated from 180 VgrG protein sequences of *Pseudomonas* spp. Black circles indicate branches with bootstrap support values .75% (1000 replicates). Black rectangles indicate VgrG proteins encoded inside a T6SS locus. Dark grey rectangles indicate that *vgrG* is linked to *hcp*, whereas light grey rectangles indicate that a gene encoding a lipase is found in the vicinity of *vgrG*. Evolved VgrG proteins are labelled with the letter e. The scale bar indicates genetic distance.

PGPA1 is an endoplasmic reticulum (ER) membrane protein which is involved in inositol deacylation of glycosylphosphatidylinositol (GPI)-anchored proteins (Tanaka *et al*., 2004), which is important for efficient ER-to-Golgi transport of GPI-anchored proteins. Thus the presence of the PGPA1 domain could mean that this protein interferes with eukaryotic protein trafficking. Two other *P. aeruginosa* specific lipases, PA3487 (or PldA) and PA5089, encoded within VgrG islands possess phospholipase D (PLD) domains, and PLD activity has been demonstrated for PldA (Wilderman *et al*., 2001). Bacterial PLDs are often associated with virulence as they specifically hydrolyse phosphatidylcholine, although phosphatidylinositol-specific PLDs have also been identified (Liscovitch *et al*., 2000). PldA is involved in virulence in a chronic pulmonary infection model (Wilderman *et al*., 2001). However, during its characterization, this enzyme, which lacks the type 2 signal sequence, was not detected in culture supernatants under the conditions used (Wilderman *et al*., 2001). Interestingly, the fact that PA1510 and PA5089 also lacked a canonical hydrophobic signal peptide, suggested that these three proteins could possibly be secreted by the T6SS.

The presence of gene encoding lipases in the vicinity of *vgrG* was also assessed in the other *Pseudomonas* species. A total of 34 genes encoding lipases or esterases were encoded in the vicinity of 118 orphan VgrGs indicating that the association is found in multiple *Pseudomonas* species (**Figure 2.4**). The lipases/esterases can be separated into several different clades, but the same lipases can be associated with *vgrG* genes of different phylogenetic clusters e.g. members of the same lipase clade are found encoded adjacent to a *vgrG* gene from cluster 4 in *P. syringe* B728A a *vgrG* gene form cluster 1 in several *P. aeruginosa* strains.

The fact that multiple VgrGs and Hcps could be associated with one specific secretion apparatus could possibly explain the myriad of phenotypes associated with T6SS. Indeed, one may propose that several subsets of effectors, including different Hcp and VgrG proteins, are associated with particular T6SS under certain conditions. However, we cannot rule out the possibility that some orphan Hcp and VgrG have evolved non-T6SS specialised functions.

Co-regulation of orphan *vgrG*/*hcp* and T6SS clusters

Previous studies have demonstrated that there is often co-regulation of the genes encoding the structural components of secretion systems and the genes encoding substrates of those secretion systems (Cornelis, 2006; Alvarez-Martinez and Christie, 2009). To ascertain whether such links could be established for T6SS and VgrG/Hcp proteins, we carried out a meta-analysis of selected *P. aeruginosa* transcriptome datasets for each HSI-locus. Transcriptome datasets were analysed when at least 50 % of the structural T6SS genes were differentially expressed (**Table S2.3**). Expression profiles of all *P. aeruginosa* genes were generated using either Pearson correlation or Euclidian distance as distance for K-means clustering. A list of gene co-regulated with the T6SS structural genes was obtained for each locus (**Table S2.4**).

Table 2.4. Genomic dissimilarities between *P. aeruginosa* PAO1 VgrG regions and the associated genome Dinucleotide frequency and mol% G+C of each VgrG genomic region were calculated using the software dr-web (van Passel *et al.*, 2005). Rows in bold indicate that the regions differ significantly in mol% G+C and dinucleotide frequency compared to the PAO1 genome signature. References indicate that these regions have been already reported to be located inside genomic islands.

***vgrG* islands**	**Chromosome position**	**Locus**	**GC%**	**% genomic fragments with lower GC%**	**1000 x δ* (genomic dissimilarity)**	**% genomic fragments with lower δ***	**Mobile elements**	**References**
PA0091								
	112506-117524	PA0091-0095	66	33	59	81	none	
PA0095								
	117524-122541	PA0095-0101	67	50	17	1	none	Ernst *et al.*, 2003
	122541-127541	PA0101-0105	67	50	36	32	none	
PA0262								
	286154-291154	PA0254-0260	63	7	37	33	Putative integrase	Ernst *et al.*, 2003
	291154-296154	PA0260-0262	65	24	34	26	none	
	296154-301154	**PA0262-0265**	**62**	**5**	**95**	**98**	**tRNA-Arg**	
PA1511								
	1638379-1643379	PA1508-1511	60	3	58	79	none	
	1643379-1648379	PA1511-1515	63	10	49	62	none	
PA2373								
	2621552-2626757	PA2371-2374	64	12	57	79	none	Ernst *et al.*, 2003
	2626757-2631962	PA2374-2378	68	75	46	58	none	
PA2685								
	3030752-3035752	**PA2683-2685**	**70**	**92**	**60**	**82**	**RHS element**	
	3035752-3040752	PA2685-2688	67	45	36	31	none	
PA3294								
	3681584-3686584	**PA3290-3294**	**58**	**2**	**104**	**99**	none	
	3686584-3691584	PA3294-3297	67	50	42	47	none	
PA3486								
	3895233-3900233	PA3482-3486	67	44	39	38	none	Wildermann *et al.*, 2001
	3900233-3905233	PA3486-3487	66	44	42	47	none	
PA5090								
	5724473-5729473	**PA5085-5090**	**58**	**2**	**87**	**97**	**TPR-like**	Ernst *et al.*, 2003
	5729473-5734473	PA5090-5093	69	79	23	6	none	
PA5266								
	5925602-5930602	**PA5263-5266**	**61**	**4**	**65**	**87**	none	Ernst *et al.*, 2003
	5930602-5935602	PA5266-5270	66	27	36	31	none	

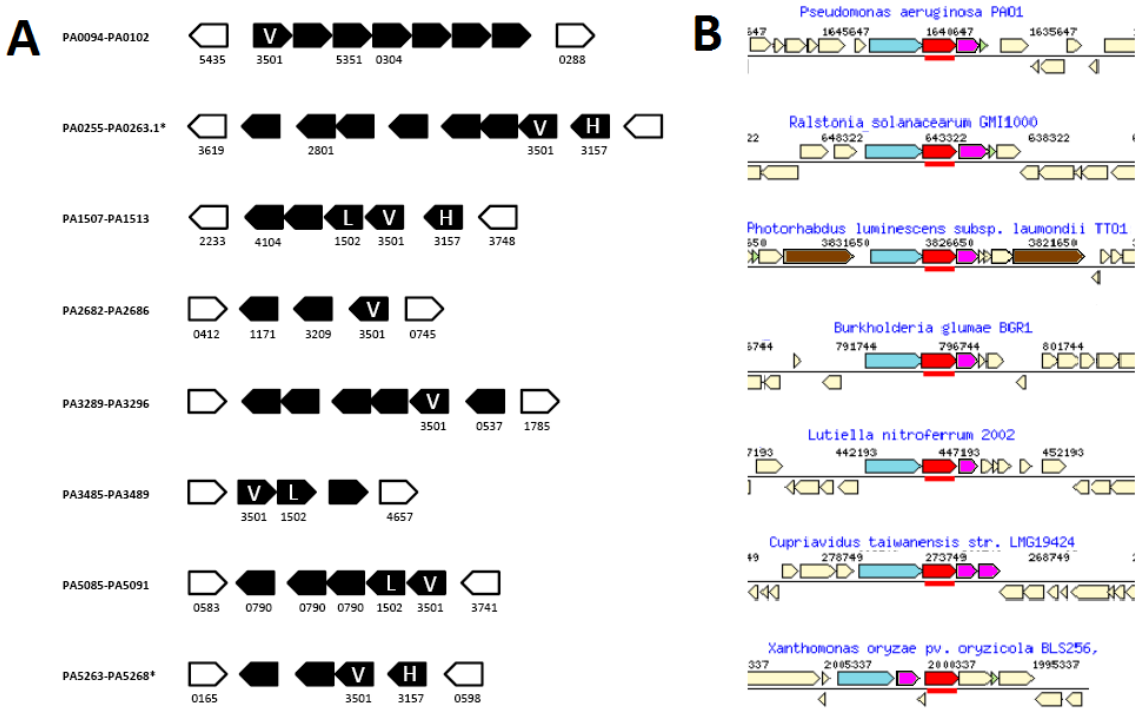


Figure 5. **(A)** *vgrG* islands of *P. aeruginosa* PAO1. **(B)** genomic context suggests these are T6SS-related proteins in other bacteria. Homologues of PA1510 (red) in other bacteria occur next to a *vgrG* (blue) gene or with T6SS in a T6SS locus with core T6SS genes such as *icmF* (brown).

A total of 36, 90 and 96 genes were co-regulated with the phylogenetic clusters 1, 3 and 4a, respectively (**Table S2.4**). We specifically examined whether orphan *vgrG* and *hcp* genes showed patterns of co-regulation. Two genes encoding VgrG1 and VgrG4 (PA2685) along with 2 genes encoding Tse proteins (PA1844 and PA3484) were co-regulated with the phylogenetic cluster 3 (or HSI-I), PA2373 was co-regulated with phylogenetic cluster 4A (or HSI-III) and PA3294 and PA3486 with phylogenetic cluster 1 (HSI-II). The other 5 orphan *vgrG* genes, PA0095, PA0262, PA1511, PA5090 and PA5266, were unassigned to any particular cluster. Among the 5 genes encoding Hcp proteins, PA0085 was linked with phylogenetic cluster 3, PA0263 with cluster 1.1 (or HSI-II), PA2367 with cluster 4A whereas PA1512 and PA5267 were unassigned. Taken together these results show that some orphan VgrGs or Hcps are co-regulated

with a T6SS phylogenetic cluster. Crucially, linkage of VgrG or Hcp with certain T6SS clusters based on expression analysis is in accordance with linkage based on phylogeny (**Table 2.5**). The lack of co-regulation between T6SS loci and some orphan VgrGs/Hcps could possibly be explained by (*i*) the conditions used for transcriptome experiments; (*ii*) the high degree of identity between some *hcp* and *vgrG* paralogues, which could mask the expression pattern of specific genes during array hybridization; (*iii*) or the fact that these orphan genes are not part of T6SS.

Some genes linked with *vgrG* islands (e.g. PA0097, PA2684 and PA3487 or *pldA*) were also co-regulated with some T6SS phylogenetic clusters, which could possibly suggest that these genes are also related to T6SS. Interestingly, *pldA* associated with the orphan *vgrG* PA3486, is co-regulated with cluster 1.1. This might suggest that the phospholipase D PldA could possibly be secreted by this T6SS.

Table 2.5. Relationship between *P. aeruginosa* T6SS locus and VgrGs/Hcps. Relationship between orphan VgrGs/Hcps and associated *Pseudomonas aeruginosa* T6SS locus are summarised in this table. PC: relationship based on VgrG or Hcp phylogenetic clusters (**Figure 2.3** and **Figure 2.4**). EC: relationship based on expression cluster (**Table S2.4**). Y indicates that the corresponding protein is associated with cluster 1, 3 or 4A.

VgrG/Hcp	Cluster 1 (HSI-II)		Cluster 3 (HSI-I)		Cluster 4A (HSI-III)		Biological evidence
	PC	EC	PC	EC	PC	EC	
PA0085 (Hcp1)			Y				Mougous *et al*., 2006
PA0091 (VgrG1)			Y	Y			Hood *et al*., 2010
PA0095 (VgrG)			Y				none
PA2685 (VgrG4)			Y	Y			Hood *et al*., 2010
PA0262 (VgrG)	Y						none
PA0263 (Hcp)	Y	Y					none
PA1511 (VgrG)	Y						none
PA1512 (Hcp)	Y						none
PA2367 (Hcp3)					Y	Y	none
PA2373 (VgrG3)					Y	Y	none
PA3294 (VgrG)		Y					none
PA3486 (VgrG)		Y					none
PA5090 (VgrG)							none
PA5266 (VgrG)	Y						none
PA5267 (Hcp)	Y						none

Experimental evidence suggests that targets of the T6SS are not dictated by phylogeny

The T6SS loci of *P. fluorescens* F113 and *P. aeruginosa* are quite similar in terms of homology and gene order. To assess if these loci were expressed at similar levels under laboratory conditions, the promoter regions from these loci were inserted into the PMS402 vector containing the promoterless *lux* cassette.

While expression of HSI-I and HSI-II was consistently low in *P. fluorescens* F113, HSI-III was expressed at a high level regardless of growth phase, indicating that its expression may be constitutive, as is the case for the T6SS in *Serratia marcescens* (**Figure 2.6**) (Murdoch *et al*., 2011). Constitutive expression of the T6SS may indicate that the specific function of this T6SS is often or urgently required by these bacteria in their particular niche.

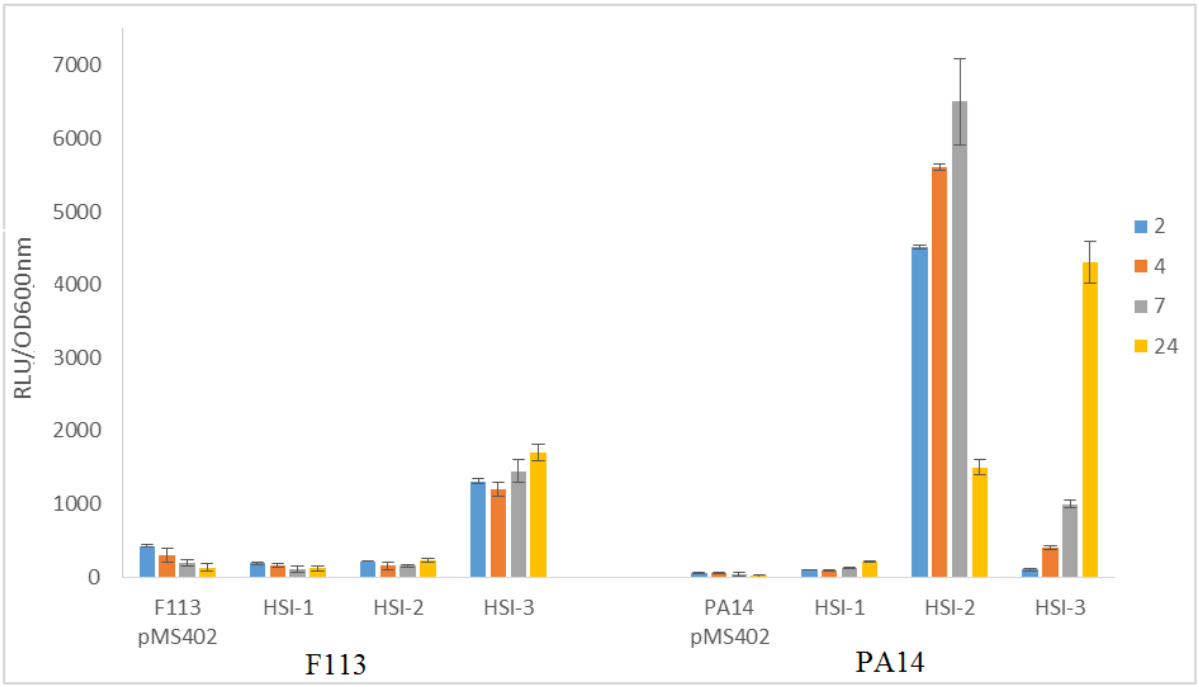


Figure 2.6. Promoter activity of genes from each T6SS loci from F113 and PA14 growing in standard growth conditions was assessed after 2, 4 7 and 24 hours of growth. The Y axis displays the relative light units (RLU) divided by the OD600 nm of the culture. Error bars represent standard deviation of three biological replicates.

Similar to F113, the HSI-I locus of PA14 was poorly expressed under standard growth conditions. However, the HSI-II locus was strongly induced, but the level of promoter activity decreased after the onset of stationary phase. The HSI-III locus showed some activity during the log phase growth but was highly expressed after the onset of stationary phase. These data indicate that the T6SS loci belonging to the same cluster are not necessarily expressed under the same conditions.

Although, a systematic prediction of each T6SS cluster function will require more studies on individual strains, some correlation between function and phylogenetic cluster is evident from available data. For example, the phylogenetic cluster 1 has been involved in virulence of *Aeromonas hydrophila*, *Vibrio cholerae* and *Burkholderia thailandensis* towards eukaryotic cells (Pukatzki *et al*., 2007; Schwarz, West, *et al*., 2010; Suarez *et al*., 2010). In *P. aeruginosa* this cluster, called HSI-II, is also involved in bacterial virulence towards plants and animals (Lesic *et al*., 2009).

It has already been reported that *P. aeruginosa* uses conserved virulence mechanisms in multiple hosts (Pukatzki *et al*., 2002). Virulence towards mice and *A. thaliana* previously reported to be mediated by HSI-II and HSI-III. To see if these T6SS were also determinants of *Pseudomonas* virulence against other eukaryotes, T6SS deletion mutants were tested for virulence towards the social amoeba *Dictyostelium discodeum.* This is a model organism for *P. aeruginosa* pathogenesis, and for *V. cholerae* T6SS pathogenicity assays (Pukatzki *et al*., 2002, 2006).

TssB is one of two proteins which forms the T6SS sheath and is essential for a functional T6SS (Zheng and Leung, 2007). To generate strains with non-functional T6SS, mutants lacking the *tssB* gene were constructed in PA14. Importantly, neither the single *tssB* mutants, nor the PA14 *ΔtssB2 ΔtssB3* double mutant displayed altered growth rates compared to the wild type strain.

The *ΔtssB2 ΔtssB3* double mutant was tested for increased plaque formation in a *D. discodeum* predation assay following the protocol of Sperandio *et al* (Sperandio *et al*., 2012), but no effect was seen (**Figure 2.7**). As F113 would also encounter amoeba in its niche, a strain of F113 lacking all *tssB* genes was engineered for use in F113 *D. discodeum* grazing assays **(Figure 2.8).** Plates inoculated with the *ΔtssB1-3* triple mutant showed increased plaque counts in contrast to the wild type F113 strain, which did not permit plaques to form **(Figure 2.8a)**. Individual deletion mutants in T6SS genes

were then tested, and the *ΔtssB3* strain was shown to have greater susceptibility than the wild type strain **(Figure 2.8b)**. The *tssB3* mutant was complemented with by expressing the *tssB3 in trans* using the expression vector pBBR5MCS. Plaque formation was decreased in plates containing the complemented mutant compared to plates containing the mutant strain carrying the empty pBBR5MCS vector **(Figure 2.8c) (Table 2.6)**. Indeed, the level of plaque formation seen in plates containing the complemented mutant was more similar to the level seen in plates containing the wild type strain in this assay. While this trend was consistent within each biological replicate, there was some variation between the levels of plaque formation observed between the biological replicates.

Given the high levels of homology evident between T6SS in F113 and PA14, the involvement of the T6SS from only one of these bacteria in virulence towards this model, suggests that T6SS of the same cluster do not always play the same roles, even within members of the same genus.

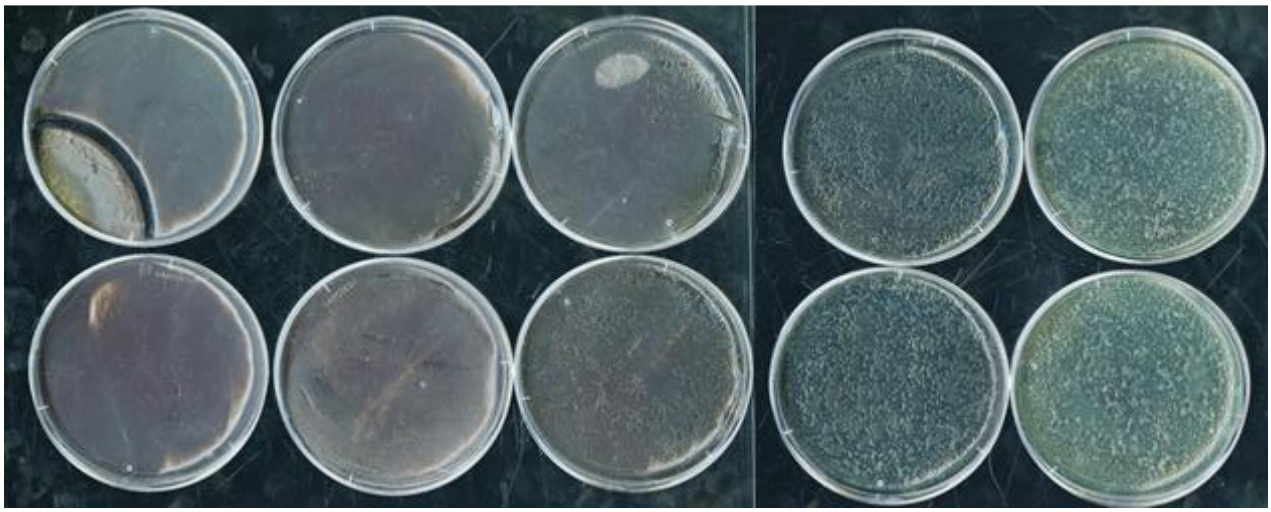


Figure 2.7. *Dictyostelium* predation assays with PA14 (top series) and PA14 mutant (bottom series). *P. aeruginosa* were added at inoculation densities ranging from OD600nm 1.0 to 1.0 X 10^{-4} (from left to right). Plaques do not appear in plates inoculated with PA14 at inoculation densities of OD600nm 1.0 X 10^{-2} or higher. No difference is observed in plaque formation between the PA14 wild type and the *ΔtssB2-3* mutant.

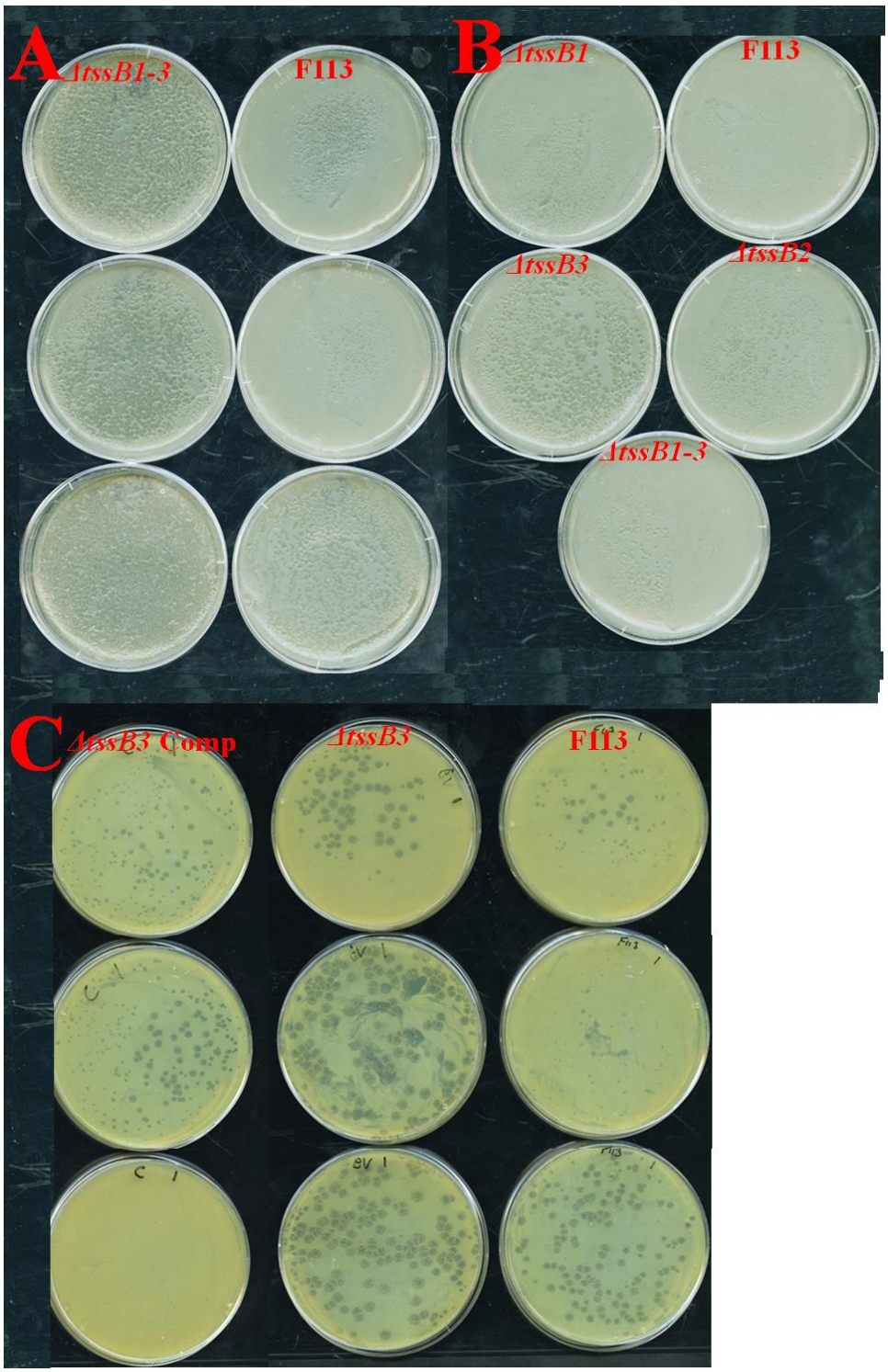


Figure 2.8. (A) *Dictyostelium* predation assays with F113 wild type (right column) and the *ΔtssB1-3* mutant (left column). F113 were added at inoculation densities ranging from OD600nm 5.0 to 5.0 X 10^{-2} (from top to bottom. **(B)** *Dictyostelium* predation assays with F113 wild type and mutant strains. Anticlockwise from top right: F113, *ΔtssB1, ΔtssB3, ΔtssB2, ΔtssB1-3*. For all assays an inoculation density of OD600nm to 5.0 X 10^{-1} was used. **(C)** *Dictyostelium* predation assays with the complemented *ΔtssB3* mutant (left), the *ΔtssB3* mutant carrying the empty vector (centre) and the

F113 wild type (right). For all assays an inoculation density of OD600nm to 5.0 X 10^{-1} was used. Three representative technical replicates are shown.

Table 2.6. Plaque formation by *Dictyostelium* on plates containing *K. aerogenes* and various F113 strains was scored from 1-10. Aggregate scores from 3 biological replicates each containing three technical replicates shows that the greater plaque formation occurred in plates containing the *ΔtssB3* strain carrying the empty vector relative to plates containing the wild type strain and the complemented *ΔtssB3* mutant.

	F113 pBBR1MCS-5	*ΔtssB3pBBR1MCS-5*	*ΔtssB3* pBBR1MCS-5_tssB3
Rep 1	1 0 2	4 5 6	1 2 4
Rep 2	2 1 6	5 5 8	1 4 0
Rep 3	3 2 0	5 7 6	2 4 6
Aggregate score	17	56	24

Different roles for T6SS in different strains.

A recent report indicated that PAO1 virulence towards the nematode *C. elegans* and full invasion of both human HeLa cells and Calu-3 lung epithelial cells (Sana *et al*., 2012).

The contribution of HSI-II *ΔtssB2* and HSI-III to invasion of lung epithelial cells by PA14 was assessed using a gentamicin protection assay. After co-culture with A549 airway epithelial cells neither *ΔtssB2*, *ΔtssB3* or even the *ΔtssB2 ΔtssB3* double mutant displayed decreased levels of invasion compared to the wild type strain **(Figure 2.9**). This is in contrast to the results record by Sana and colleagues. In fact, there was a statistically insignificant increase in invasion by the T6SS mutants. This may be explained by the use of airway epithelial cells instead of HeLa cells or by strain-differences between the different *P. aeruginosa* strains PAO1 and PA14.

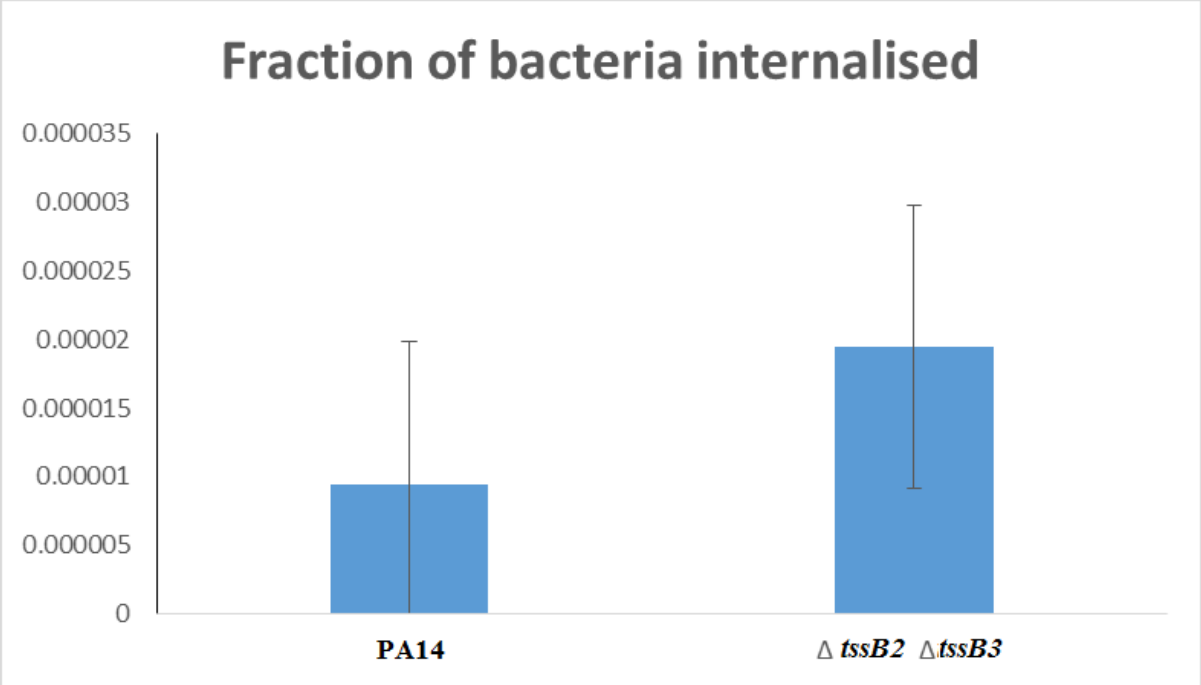


Figure 2.9: Internalisation of PA14 and PA14 *ΔtssB2 ΔtssB3* into A549 airway epithelial cells as measured by gentamicin protection assay reveals a statistically insignificant two-fold difference. Error bars represent standard deviation of three biological replicates.

The *ΔtssB2, ΔtssB3 and ΔtssB2 ΔtssB3* strains were also tested for impaired killing of *C. elegans* by PA14. The experiments were done following the protocol used by Sana and colleagues (Sana *et al*., 2012), although as PA14 kill *C. elegans* more efficiently than PAO1, this experiment took place over eight days instead of 16. None of the mutants displayed statistically different levels of killing to the wild type, though with a P. value of 0.07, the difference between wild type and the HSI-II mutant is trending on significance (**Figure 2.10).**

Taken together, the results from PA14 interactions with airway epithelial cells and *C. elegans* indicate that the roles of T6SS from the same family may not be consistent even between members of the same species.

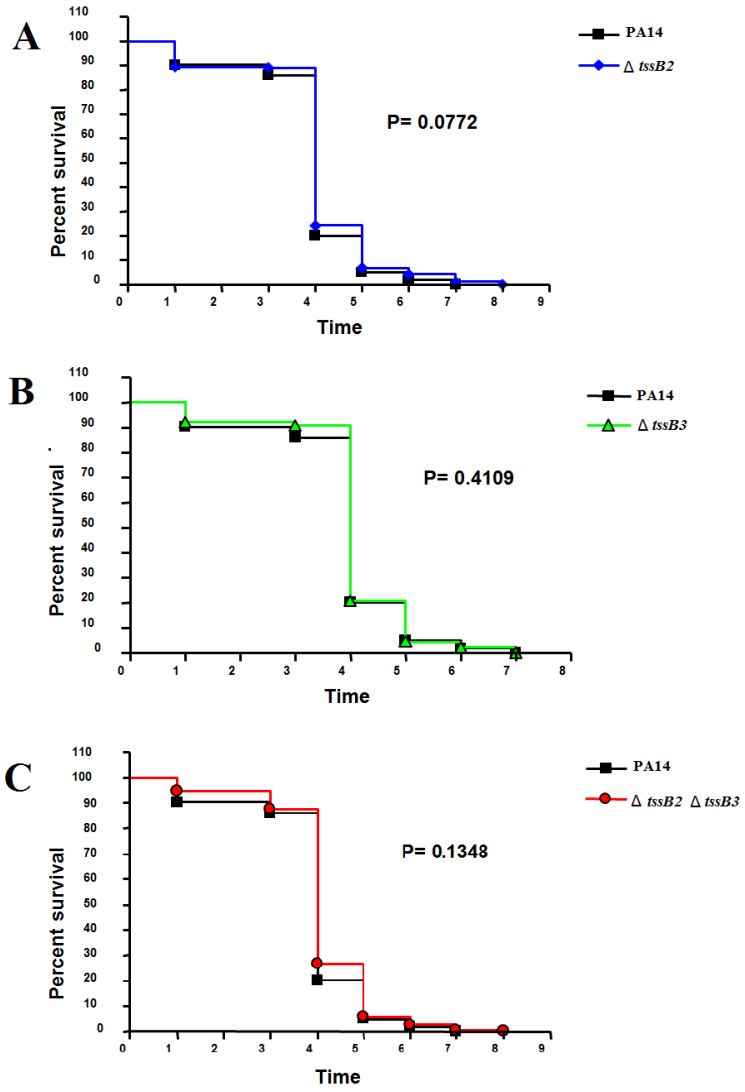


Figure 2.10. The survival of *C. elegans* consuming PA14 and T6SS mutant strains *ΔtssB2* **(A)**, *ΔtssB3* **(B)** *and ΔtssB2 ΔtssB3* **(C)** was similar, in contrast to the results previously reported for *P. aeruginosa* PAO1 (Sana *et al*., 2012).

Phenotypic characterisation of *vgrG* islands

The *vgrG* islands are a potential source of new T6SS-associated genes. The bioinformatic evidence for this is strong, but experimental confirmation is lacking. The colony morphologies of PA14 strains with transposon insertions in orphan *vgrG* genes were similar to that of the wild type, but disruption of PA14_44900 resulted in a slightly slower growth rate **(Figure 2.11)**. To confirm that this was not due to a secondary mutant the entire *vgrG* island PA14_44900-44930 was deleted. This deletion mutant also grew slowly **(Figure 2.11)**.

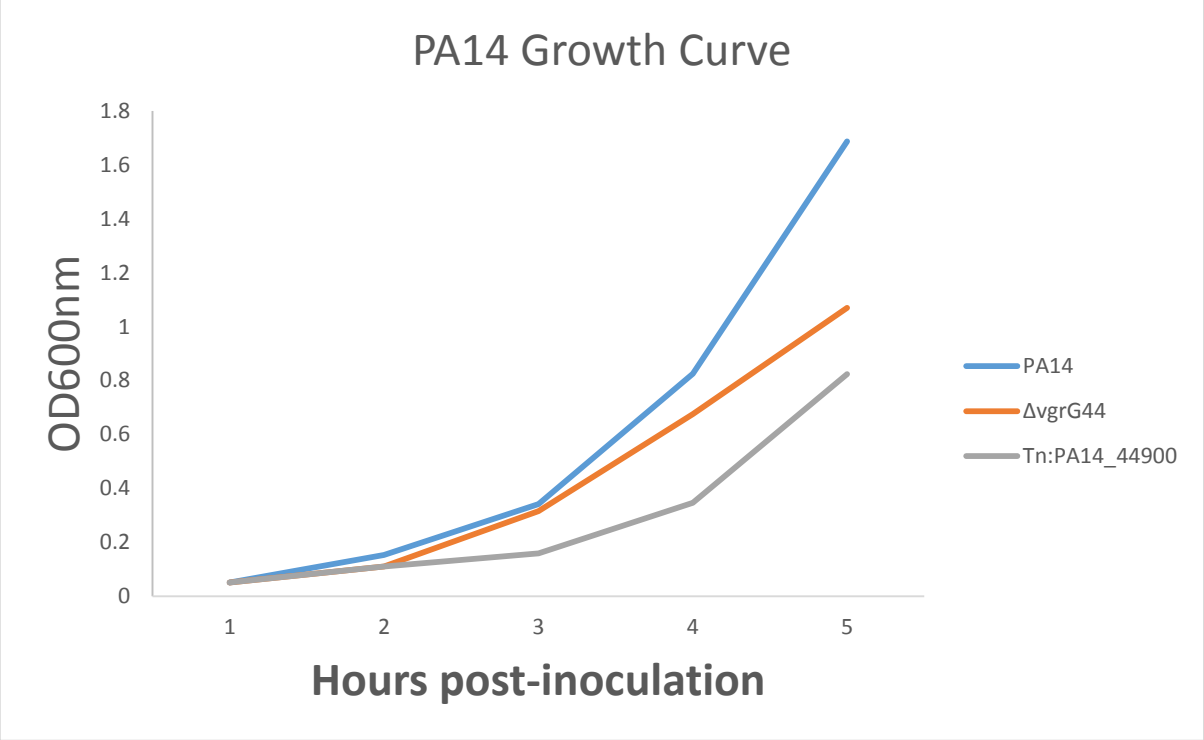
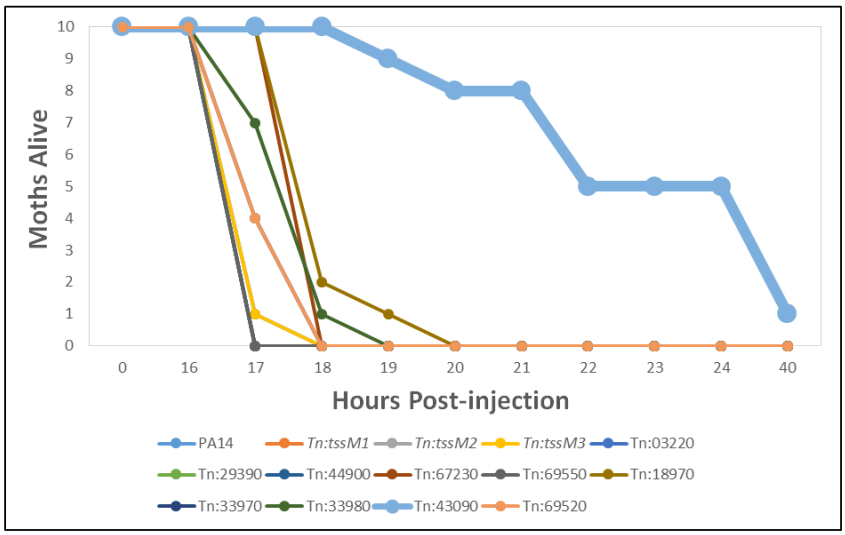


Figure. 2.11. Deletion of the entire *vgrG* island (PA14_44900-44930) or transposon disruption of the *vgrG* gene PA14_44900 led to reduced growth. Experiment was performed under standard conditions: shaking in LB at 37 °C.

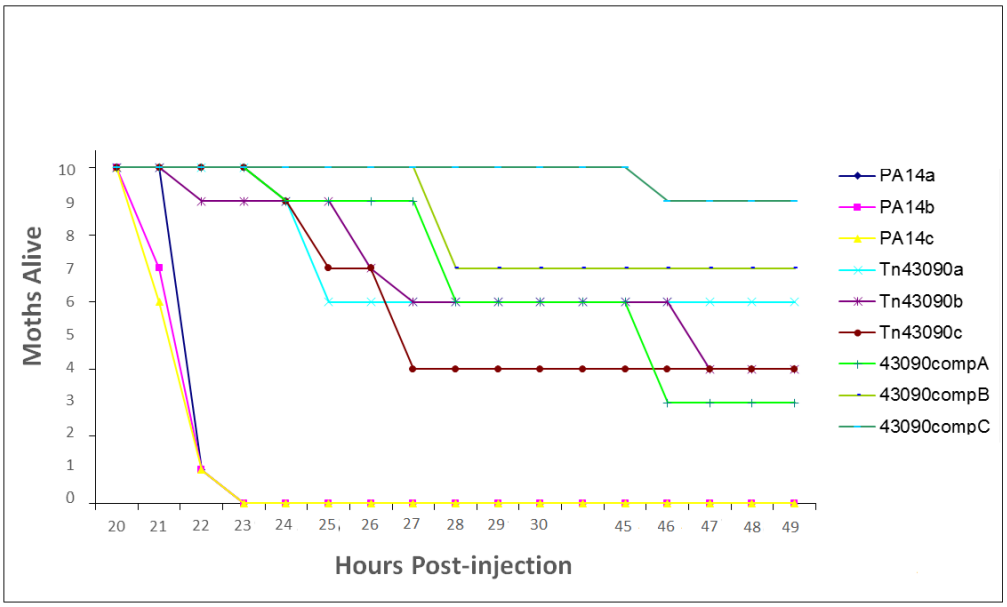
To test these genes for involvement in T6SS-phenotypes a range of PA14 transposon mutants were tested in another model organism for *P. aeruginosa* virulence: the wax moth *Galleria Mellonella* (Hilbi *et al*., 2007). For consistency, transposon mutants in the core T6SS genes *tssM* were used instead of the previously obtained *ΔtssB* strains in this assay.

A strain containing an insertion in the gene PA14_43090 was shown to be impaired in virulence **(Figure 2.12)**. This strain also had a modest growth defect (data not shown). A clean deletion mutant which lacked the entire *vgrG* island containing PA14_43090 was constructed, but neither it nor a *tssB2*-3 double mutant displayed impaired virulence and both grew at the same rate as the wild type PA14. Moreover, expressing the PA14_43090 on the pBBR1MCS plasmid did not complement the virulence or growth defects of the transposon mutant, indicating that the result with the transposon mutant may be due to a secondary mutation. No strain in the transposon bank carries an insertion in *vgrG* gene PA14_21450, so this gene was not tested in these assays.

A


Moths Alive
Hours Post-injection
PA14
Tn:tssM1
Tn:tssM2
Tn:tssM3
Tn:03220
Tn:29390
Tn:44900
Tn:67230
Tn:69550
Tn:18970
Tn:33970
Tn:33980
Tn:43090
Tn:69520

B


Moths Alive
Hours Post-injection
PA14a
PA14b
PA14c
Tn43090a
Tn43090b
Tn43090c
43090compA
43090compB
43090compC

C

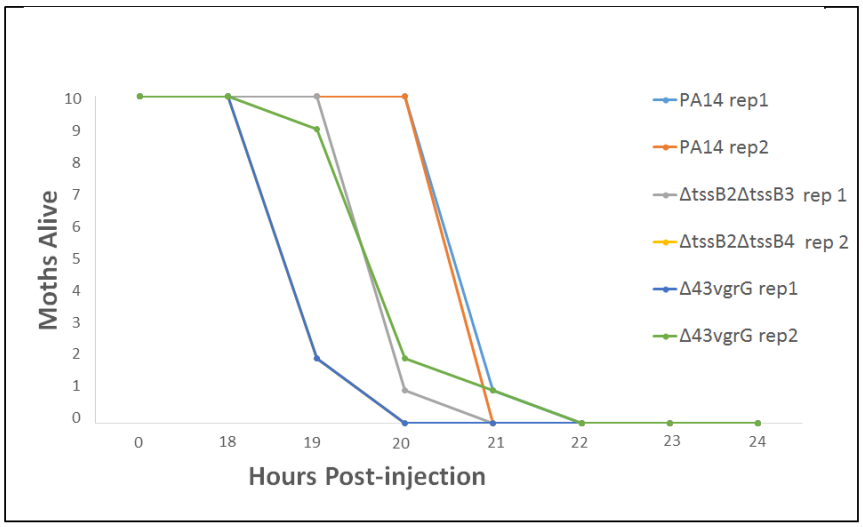

Moths Alive
Hours Post-injection
PA14 rep1
PA14 rep2
ΔtssB2ΔtssB3 rep 1
ΔtssB2ΔtssB4 rep 2
Δ43vgrG rep1
Δ43vgrG rep2

Figure 2.12. (A) PA14 strains whose vgrG island genes had been disrupted by transposons were assessed for virulence towards the wax moth. Disruption of PA14_43090 led to slower killing in this assay. **(B)** The virulence defect of this strain was not complementable by expressing the PA14_43090 gene on the plasmid pBBR1MCS1. **(C)** Deletion of the entire *vgrG* island PA14_43080-PA14_43100, which contains the PA14_43090 gene, did not result in a virulence defect.

Discussion

This study demonstrates the prevalence of T6SS in *Pseudomonas* spp. Every *Pseudomonas* genome sequenced (except *P. stutzeri* A1501) possesses at least one putative T6SS. Among these strains, five different T6SS clusters, having distinct evolutionary origins, have been found. Whereas phylogenetic clusters 1, 3 and 4 have already been reported inside the *Pseudomonas* genus (Boyer *et al*., 2009), the present work identified a novel *Pseudomonas* T6SS locus related to cluster 2 and demonstrated a clear separation of the previously reported cluster 4 into the distinct clusters 4A and 4B.

The division of T6SS into distinct phylogenetic families may reflect differences in the role or mechanism of these T6SS. The specificity of the relationship between phylogeny T6SS accessory genes supports the hypothesis that there are mechanistic differences between T6SS families.

If phylogeny is useful for predicting the function(s) associated with different T6SSs, it is most likely that that is due to T6SS of different phylogenetic origins being optimised to interact with certain cell types, or with certain sets of effectors. The cluster 3 of *P. aeruginosa* has been shown to target a toxin to other bacteria, which suggests a role for this system in bacterial-bacterial interaction (Hood *et al*., 2010). Whether this is the case with each cluster 3 T6SS outside *Pseudomonas* or even within *Pseudomonas* is unclear, as the bacterial toxin is unique to *P. aeruginosa* genomes, but it is possible to imagine that these effectors are targeted to the T6SS apparatus using a cluster specific secretion signal. Unfortunately, no other T6SS effectors have as yet been identified within the *Pseudomonas* genus, which hampers further understanding of any substrate specificity of these secretion systems.

Experimental results reported in this chapter show that different T6SS families do not necessarily share common roles **(Figure 2.13)**. For example, HSI-III T6SS of F113 is important in resistance to *D. discodeum* grazing, but the closely related HSI-III of PA14 has no role in this interaction. Moreover, expression of these loci varies between F113 and PA14.

While there are differences between the roles of T6SS, certain families could still be optimised to target broader categories of target organism. Further experimental work

will be required before it can be determined whether a relationship between T6SS phylogeny and role exists.

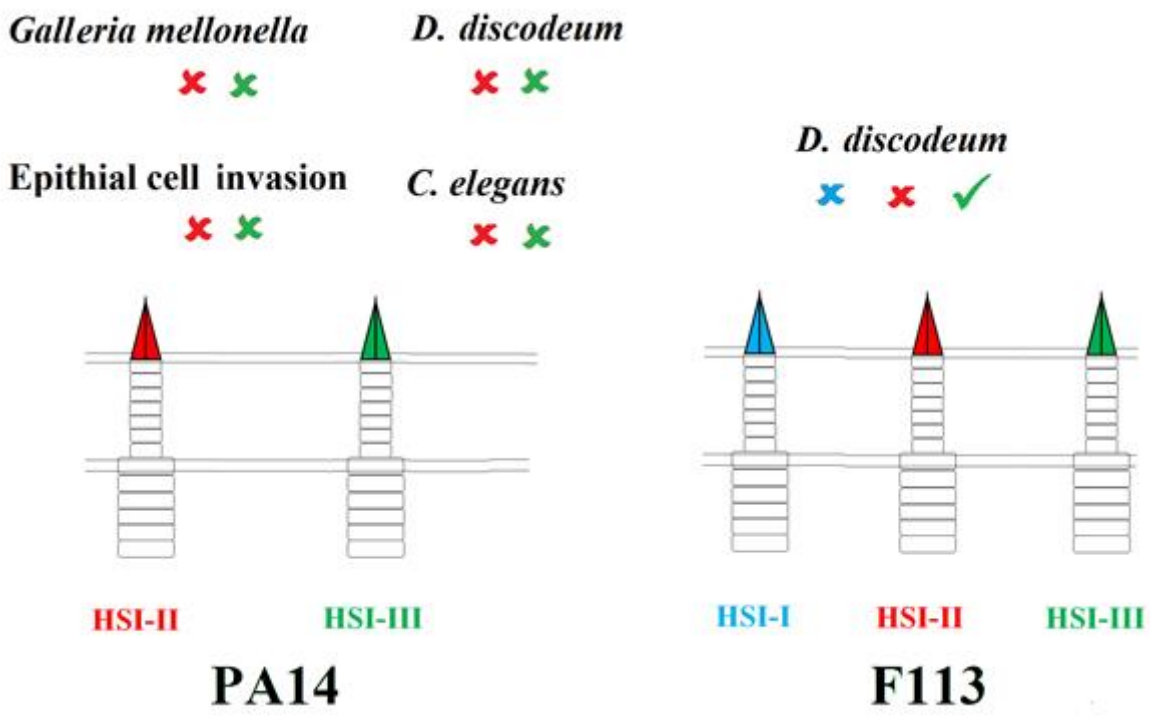


Figure 2.13. The use of the T6SS to mediate bacterial interactions with model organisms varies within members of the same genus and same species. *P. fluorescens* F113 can influence the grazing activity of *D. discodeum* in a HSI-III dependent manner. However, this T6SS is seemingly not involved in interactions between *P. aeruginosa* PA14 and *D. discodeum*, or with the model organisms *G. mellonella* and *C. elegans*, nor does it increase the uptake of the bacteria into epithelial cells.

Though these phylogenetically similar T6SS do not share common functions in the virulence assays reported here, these T6SS may have the same capacity to target certain cell types, but are regulated differently according to the needs to bacterium possessing them.

The previously reported involvement of the HSI-II T6SS of *P. aeruginosa* PAO1 in internalisation of the bacterium into epithelial cells and in virulence towards *C. elegans* was not evident when these virulence assays were done using the strain PA14. Given the close similarity between the T6SS loci in these strains it would be surprising if the

capabilities of the T6SS differed between these strains. Rather, this is more likely to reflect differences in the lifestyles of these bacteria. PA14 is known to be a more virulent strain than PAO1 (Lee *et al.*, 2006). Therefore, these contrasting results may be due these different strains adapting slightly different pathogenic strategies.

A novel observation from this study was the identification of several conserved gene arrangements consisting of 2 to 8 genes containing *vgrG* and often *hcp* (**Figure 2.5**). These "*vgrG* islands" are frequently observed in other bacterial species and occasionally cargo genes from these *vgrG* islands can be found within a T6SS locus from another genome. Therefore, these genes are likely to be involved in T6SS dependent phenotypes and may be encode T6SS effectors.

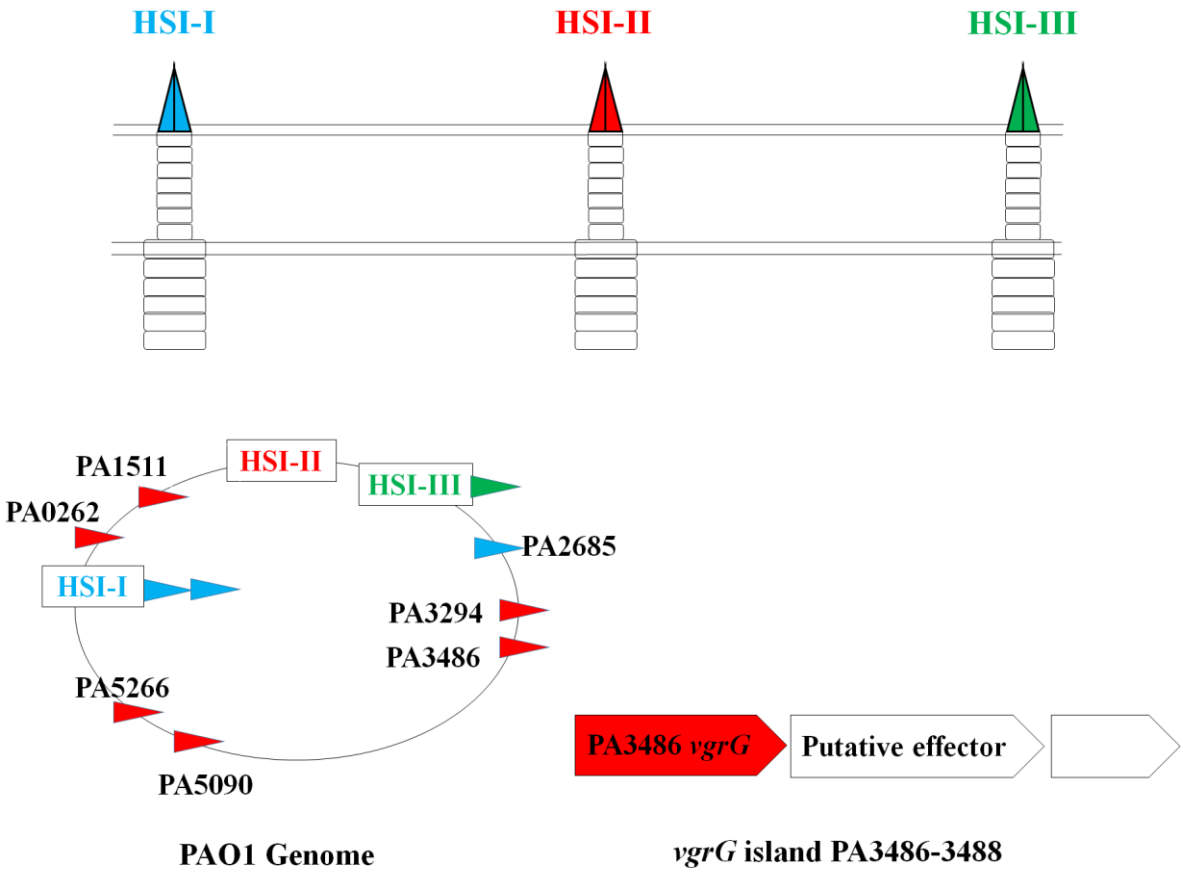


Figure 2.14. A model for orphan VgrG and putative effector association with specific T6SS. Orphans *vgrG* genes not linked to any T6SS locus are scattered around the PAO1 chromosome. Based on their phylogenetic clustering these genes can be linked to certain T6SS families: HSI-I (blue), HSI-II (red) and HSI-III (green). As putative effector genes are encoded within *vgrG* islands it is possible that these effectors are secreted specifically by certain T6SS. Shown is the representative *vgrG* island PA3486-PA3488.

Thus far, only one *vgrG* island has been associated with a biological function: recognition of self by *Proteus mirabilis* (Gibbs *et al*., 2008). It is important to note, however, that the involvement of the T6SS in this phenotype remains to be demonstrated. Common among this *vgrG* islands are genes encoding proteins with potential lipase or esterase activity.

As *vgrG* genes from these islands could be linked to certain T6SS families based on their phylogeny, and it is possible that the cargo genes of these *vgrG* islands also link to the same T6SS family **(Figure 2.14)**. Thus not only are new T6SS effectors predicted by this study, so is their association with certain T6SS subsets.

References

Alvarez-Martinez, C.E. and Christie, P.J. (2009) Biological Diversity of Prokaryotic Type IV Secretion Systems. *Microbiol. Mol. Biol. Rev.* **73**: 775–808.

Aschtgen, M.-S., Bernard, C.S., Bentzmann, S.D., 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.

Aschtgen, M.-S., Gavioli, M., Dessen, A., Lloubès, R., and Cascales, E. (2010) The SciZ protein anchors the enteroaggregative *Escherichia coli* Type VI secretion system to the cell wall. *Mol. Microbiol.* **75**: 886–899.

Aschtgen, M.-S., Thomas, M.S., and Cascales, E. (2010) Anchoring the type VI secretion system to the peptidoglycan: TssL, TagL, TagP... what else? *Virulence* **1**: 535–540.

Attila, C., Ueda, A., and Wood, T.K. (2008) PA2663 (PpyR) increases biofilm formation in *Pseudomonas aeruginosa* PAO1 through the psl operon and stimulates virulence and quorum-sensing phenotypes. *Appl. Microbiol. Biotechnol.* **78**: 293–307.

Bernard, C.S., Brunet, Y.R., Gueguen, E., and Cascales, E. (2010) Nooks and Crannies in Type VI Secretion Regulation. *J. Bacteriol.* **192**: 3850–3860.

Bingle, L.E., Bailey, C.M., and Pallen, M.J. (2008) Type VI secretion: a beginner's guide. *Curr. Opin. Microbiol.* **11**: 3–8.

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**: 354.

Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H., and Mogk, A. (2009) Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J.* **28**: 315–325.

Bönemann, G., Pietrosiuk, A., and Mogk, A. (2010) Tubules and donuts: a type VI secretion story. *Mol. Microbiol.* **76**: 815–821.

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.

Bröms, J.E., Lavander, M., and Sjöstedt, A. (2009) A Conserved α-Helix Essential for a Type VI Secretion-Like System of *Francisella tularensis*. *J. Bacteriol.* **191**: 2431–2446.

Burtnick, M.N., DeShazer, D., Nair, V., Gherardini, F.C., and Brett, P.J. (2010) *Burkholderia mallei* Cluster 1 Type VI Secretion Mutants Exhibit Growth and Actin

Polymerization Defects in RAW 264.7 Murine Macrophages. *Infect. Immun.* **78**: 88–99.

Cascales, E. (2008) The type VI secretion toolkit. *EMBO Rep.* **9**: 735–741.

Chugani, S. and Greenberg, E.P. (2007) The influence of human respiratory epithelia on *Pseudomonas aeruginosa* gene expression. *Microb. Pathog.* **42**: 29–35.

Cornelis, G.R. (2006) The type III secretion injectisome. *Nat. Rev. Microbiol.* **4**: 811–825.

Duan, K., Dammel, C., Stein, J., Rabin, H., and Surette, M.G. (2003) Modulation of *Pseudomonas aeruginosa* gene expression by host microflora through interspecies communication. *Mol. Microbiol.* **50**: 1477–1491.

Ernst, R.K., D′Argenio, D.A., Ichikawa, J.K., Bangera, M.G., Selgrade, S., Burns, J.L., *et al*. (2003) Genome mosaicism is conserved but not unique in *Pseudomonas aeruginosa* isolates from the airways of young children with cystic fibrosis. *Environ. Microbiol.* **5**: 1341–1349.

Felsenstein, J. (2005) Using the quantitative genetic threshold model for inferences between and within species. *Philos. Trans. R. Soc. B Biol. Sci.* **360**: 1427–1434.

Figurski, D.H. and Helinski, D.R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci.* **76**: 1648–1652.

Filiatrault, M.J., Wagner, V.E., Bushnell, D., Haidaris, C.G., Iglewski, B.H., and Passador, L. (2005) Effect of Anaerobiosis and Nitrate on Gene Expression in *Pseudomonas aeruginosa*. *Infect. Immun.* **73**: 3764–3772.

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.

Gibbs, K.A., Urbanowski, M.L., and Greenberg, E.P. (2008) Genetic Determinants of Self Identity and Social Recognition in Bacteria. *Science* **321**: 256–259.

Goodman, A.L., Kulasekara, B., Rietsch, A., Boyd, D., Smith, R.S., and Lory, S. (2004) A Signaling Network Reciprocally Regulates Genes Associated with Acute Infection and Chronic Persistence in *Pseudomonas aeruginosa*. *Dev. Cell* **7**: 745–754.

Green, M.R. (2012) Molecular cloning: a laboratory manual 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Gross, H. and Loper, J.E. (2009) Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat. Prod. Rep.* **26**: 1408–1446.

Guindon, S. and Gascuel, O. (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **52**: 696–704.

He, J., Baldini, R.L., Déziel, E., Saucier, M., Zhang, Q., Liberati, N.T., *et al*. (2004) The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 2530–2535.

Hilbi, H., Weber, S.S., Ragaz, C., Nyfeler, Y., and Urwyler, S. (2007) Environmental predators as models for bacterial pathogenesis. *Environ. Microbiol.* **9**: 563–575.

Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M.A., Trinidad, R.R.S., *et al*. (2010) A Type VI Secretion System of *Pseudomonas aeruginosa* Targets a Toxin to Bacteria. *Cell Host Microbe* **7**: 25–37.

Hsu, F., Schwarz, S., and Mougous, J.D. (2009) TagR promotes PpkA-catalysed type VI secretion activation in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **72**: 1111–1125.

Kang, Y., Nguyen, D.T., Son, M.S., and Hoang, T.T. (2008) The *Pseudomonas aeruginos*a PsrA responds to long-chain fatty acid signals to regulate the fadBA5 β-oxidation operon. *Microbiology* **154**: 1584–1598.

Kovach, M.E., Elzer, P.H., Steven Hill, D., Robertson, G.T., Farris, M.A., Roop II, R.M., and Peterson, K.M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**: 175–176.

Lee, D.G., Urbach, J.M., Wu, G., Liberati, N.T., Feinbaum, R.L., Miyata, S., *et al*. (2006) Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol.* **7**: R90.

Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., *et al*. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl. Acad. Sci.* **106**: 4154–4159.

Lequette, Y., Lee, J.-H., Ledgham, F., Lazdunski, A., and Greenberg, E.P. (2006) A Distinct QscR Regulon in the *Pseudomonas aeruginosa* Quorum-Sensing Circuit. *J. Bacteriol.* **188**: 3365–3370.

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.

Liberati, N.T., Urbach, J.M., Miyata, S., Lee, D.G., Drenkard, E., Wu, G., *et al*. (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 2833–2838.

Liscovitch, M., Czarny, M., Fiucci, G., and Tang, X. (2000) Phospholipase D: molecular and cell biology of a novel gene family.

Liu, H., Coulthurst, S.J., Pritchard, L., Hedley, P.E., Ravensdale, M., Humphris, S., *et al*. (2008) Quorum Sensing Coordinates Brute Force and Stealth Modes of Infection in the Plant Pathogen *Pectobacterium atrosepticum*. *PLoS Pathog* **4**: e1000093.

Ma, A.T. and Mekalanos, J.J. (2010) In vivo actin cross-linking induced by *Vibrio cholerae* type VI secretion system is associated with intestinal inflammation. *Proc. Natl. Acad. Sci.* **107**: 4365–4370.

MacIntyre, D.L., Miyata, S.T., Kitaoka, M., and Pukatzki, S. (2010) The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc. Natl. Acad. Sci.* **107**: 19520–19524.

Ma, L.-S., Lin, J.-S., and Lai, E.-M. (2009) An IcmF Family Protein, ImpLM, Is an Integral Inner Membrane Protein Interacting with ImpKL, and Its Walker A Motif Is Required for Type VI Secretion System-Mediated Hcp Secretion in *Agrobacterium tumefaciens*. *J. Bacteriol.* **191**: 4316–4329.

Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., *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*. *Nat. Cell Biol.* **9**: 797–803.

Murdoch, S.L., Trunk, K., English, G., Fritsch, M.J., Pourkarimi, E., and Coulthurst, S.J. (2011) The Opportunistic Pathogen *Serratia marcescens* Utilizes Type VI Secretion To Target Bacterial Competitors. *J. Bacteriol.* **193**: 6057–6069.

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.

Passel, M.W.J. van, Luyf, A.C.M., Kampen, A.H.C. van, Bart, A., and Ende, A. van der (2005) δρ-Web, an online tool to assess composition similarity of individual nucleic acid sequences. *Bioinformatics* **21**: 3053–3055.

Pell, L.G., Kanelis, V., Donaldson, L.W., Howell, P.L., and Davidson, A.R. (2009) The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc. Natl. Acad. Sci.* **106**: 4160–4165.

Persson, O.P., Pinhassi, J., Riemann, L., Marklund, B.-I., Rhen, M., Normark, S., *et al*. (2009) High abundance of virulence gene homologues in marine bacteria. *Environ. Microbiol.* **11**: 1348–1357.

Potvin, E., Lehoux, D.E., Kukavica-Ibrulj, I., Richard, K.L., Sanschagrin, F., Lau, G.W., and Levesque, R.C. (2003) In vivo functional genomics of *Pseudomonas aeruginosa* for high-throughput screening of new virulence factors and antibacterial targets. *Environ. Microbiol.* **5**: 1294–1308.

Puigbò, P., Garcia-Vallvé, S., and McInerney, J.O. (2007) TOPD/FMTS: a new software to compare phylogenetic trees. *Bioinformatics* **23**: 1556–1558.

Pukatzki, S., Kessin, R.H., and Mekalanos, J.J. (2002) The human pathogen *Pseudomonas aeruginosa* utilizes conserved virulence pathways to infect the social amoeba *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci.* **99**: 3159–3164.

Pukatzki, S., Ma, A.T., 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.* **104**: 15508–15513.

Pukatzki, S., Ma, A.T., 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. U. S. A.* **103**: 1528–1533.

Pukatzki, S., McAuley, S.B., and Miyata, S.T. (2009) The type VI secretion system: translocation of effectors and effector-domains. *Curr. Opin. Microbiol.* **12**: 11–17.

Robinson, J.B., Telepnev, M.V., Zudina, I.V., Bouyer, D., Montenieri, J.A., Bearden, S.W., *et al*. (2009) Evaluation of a *Yersinia pestis* mutant impaired in a thermoregulated type VI-like secretion system in flea, macrophage and murine models. *Microb. Pathog.* **47**: 243–251.

Roucourt, B., Lecoutere, E., Chibeu, A., Hertveldt, K., Volckaert, G., and Lavigne, R. (2009) A procedure for systematic identification of bacteriophage–host interactions of *P. aeruginosa* phages. *Virology* **387**: 50–58.

Sana, T.G., Hachani, A., Bucior, I., Soscia, C., Garvis, S., Termine, E., *et al*. (2012) The Second Type VI Secretion System of *Pseudomonas aeruginosa* Strain PAO1 Is Regulated by Quorum Sensing and Fur and Modulates Internalization in Epithelial Cells. *J. Biol. Chem.* **287**: 27095–27105.

Schuster, M., Hawkins, A.C., Harwood, C.S., and Greenberg, E.P. (2004) The *Pseudomonas aeruginosa* RpoS regulon and its relationship to quorum sensing. *Mol. Microbiol.* **51**: 973–985.

Schuster, M., Lostroh, C.P., Ogi, T., and Greenberg, E.P. (2003) Identification, Timing, and Signal Specificity of *Pseudomonas aeruginosa* Quorum-Controlled Genes: a Transcriptome Analysis. *J. Bacteriol.* **185**: 2066–2079.

Schwarz, S., Hood, R.D., and Mougous, J.D. (2010) What is type VI secretion doing in all those bugs? *Trends Microbiol.* **18**: 531–537.

Schwarz, S., West, T.E., Boyer, F., Chiang, W.-C., Carl, M.A., Hood, R.D., *et al*. (2010) *Burkholderia* Type VI Secretion Systems Have Distinct Roles in Eukaryotic and Bacterial Cell Interactions. *PLoS Pathog* **6**: e1001068.

Shanahan, P., O'Sullivan, D.J., Simpson, P., Glennon, J.D., and O'Gara, F. (1992) Isolation of 2,4-Diacetylphloroglucinol from a Fluorescent Pseudomonad and Investigation of Physiological Parameters Influencing Its Production. *Appl. Environ. Microbiol.* **58**: 353–358.

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**: e2955.

Southey-Pillig, C.J., Davies, D.G., and Sauer, K. (2005) Characterization of Temporal Protein Production in *Pseudomonas aeruginosa* Biofilms. *J. Bacteriol.* **187**: 8114–8126.

Sperandio, D., Decoin, V., Latour, X., Mijouin, L., Hillion, M., Feuilloley, M.G.J., *et al*. (2012) Virulence of the *Pseudomonas fluorescens* clinical strain MFN1032 towards *Dictyostelium discoideum* and macrophages in relation with type III secretion system. *BMC Microbiol.* **12**: 223.

Starkey, M., Hickman, J.H., Ma, L., Zhang, N., Long, S.D., Hinz, A., *et al*. (2009) *Pseudomonas aeruginosa* Rugose Small-Colony Variants Have Adaptations That Likely Promote Persistence in the Cystic Fibrosis Lung. *J. Bacteriol.* **191**: 3492–3503.

Suarez, G., Sierra, J.C., Erova, T.E., Sha, J., Horneman, A.J., and Chopra, A.K. (2010) A Type VI Secretion System Effector Protein, VgrG1, from *Aeromonas hydrophila* That Induces Host Cell Toxicity by ADP Ribosylation of Actin. *J. Bacteriol.* **192**: 155–168.

Tanaka, S., Maeda, Y., Tashima, Y., and Kinoshita, T. (2004) Inositol Deacylation of Glycosylphosphatidylinositol-anchored Proteins Is Mediated by Mammalian PGAP1 and Yeast Bst1p. *J. Biol. Chem.* **279**: 14256–14263.

Tart, A.H., Wolfgang, M.C., and Wozniak, D.J. (2005) The Alternative Sigma Factor AlgT Represses *Pseudomonas aeruginosa* Flagellum Biosynthesis by Inhibiting Expression of fleQ. *J. Bacteriol.* **187**: 7955–7962.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**: 33–36.

Teitzel, G.M., Geddie, A., Long, S.K.D., Kirisits, M.J., Whiteley, M., and Parsek, M.R. (2006) Survival and Growth in the Presence of Elevated Copper: Transcriptional Profiling of Copper-Stressed *Pseudomonas aeruginosa*. *J. Bacteriol.* **188**: 7242–7256.

Tian, Z.-X., Mac Aogáin, M., O'Connor, H.F., Fargier, E., Mooij, M.J., Adams, C., *et al*. (2009) MexT modulates virulence determinants in *Pseudomonas aeruginosa* independent of the MexEF-OprN efflux pump. *Microb. Pathog.* **47**: 237–241.

Vallet, I., Diggle, S.P., Stacey, R.E., Cámara, M., Ventre, I., Lory, S., *et al*. (2004) Biofilm Formation in *Pseudomonas aeruginosa*: Fimbrial cup Gene Clusters Are Controlled by the Transcriptional Regulator MvaT. *J. Bacteriol.* **186**: 2880–2890.

Waite, R.D. and Curtis, M.A. (2009) *Pseudomonas aeruginosa* PAO1 Pyocin Production Affects Population Dynamics within Mixed-Culture Biofilms. *J. Bacteriol.* **191**: 1349–1354.

Wang, X., Wang, Q., Xiao, J., Liu, Q., Wu, H., Xu, L., and Zhang, Y. (2009) *Edwardsiella tarda* T6SS component evpP is regulated by esrB and iron, and plays essential roles in the invasion of fish. *Fish Shellfish Immunol.* **27**: 469–477.

Weber, B., Hasic, M., Chen, C., Wai, S.N., and Milton, D.L. (2009) Type VI secretion modulates quorum sensing and stress response in *Vibrio anguillarum*. *Environ. Microbiol.* **11**: 3018–3028.

Whelan, S. and Goldman, N. (2001) A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol. Biol. Evol.* **18**: 691–699.

Wilderman, P.J., Vasil, A.I., Johnson, Z., and Vasil, M.L. (2001) Genetic and biochemical analyses of a eukaryotic-like phospholipase D of *Pseudomonas aeruginosa* suggest horizontal acquisition and a role for persistence in a chronic pulmonary infection model. *Mol. Microbiol.* **39**: 291–304.

Williamson, L.L., Borlee, B.R., Schloss, P.D., Guan, C., Allen, H.K., and Handelsman, J. (2005) Intracellular Screen To Identify Metagenomic Clones That Induce or Inhibit a Quorum-Sensing Biosensor. *Appl. Environ. Microbiol.* **71**: 6335–6344.

Wolfgang, M.C., Jyot, J., Goodman, A.L., Ramphal, R., and Lory, S. (2004) *Pseudomonas aeruginosa* regulates flagellin expression as part of a global response to airway fluid from cystic fibrosis patients. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 6664–6668.

Wolfgang, M.C., Lee, V.T., Gilmore, M.E., and Lory, S. (2003) Coordinate Regulation of Bacterial Virulence Genes by a Novel Adenylate Cyclase-Dependent Signaling Pathway. *Dev. Cell* **4**: 253–263.

Wood, L.F. and Ohman, D.E. (2009) Use of cell wall stress to characterize σ22 (AlgT/U) activation by regulated proteolysis and its regulon in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **72**: 183–201.

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.

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

Zheng, P., Sun, J., Geffers, R., and Zeng, A.-P. (2007) Functional characterization of the gene PA2384 in large-scale gene regulation in response to iron starvation in *Pseudomonas aeruginosa*. *J. Biotechnol.* **132**: 342–352.

Chapter 3

Title: Bioinformatics analysis of T6SS effectors and related components: insights into T6SS mechanism and roles.

Research from this chapter contributed to the following publication:

Title: Tle Distribution and Diversity in Metagenomic Datasets Reveals Niche Specialisation (in submission)

Authors: F. Egan, J.F. Reen and F. O'Gara

Abstract

Less than a decade after its discovery, much progress has made on the mechanism and roles of the Type VI Secretion System (T6SS). However, several questions remain open. Some non-essential genes associated with this system have no known functions, and seemingly little attempt has been made to resolve this experimentally. However, the wealth of genomic data allows for a bioinformatic investigation of their functions within the T6SS. Similarly, while the roles of the T6SS have been elucidated using model organisms, genomic and metagenomic data analysis is needed to understand how this system affects the complex interactions that govern the dynamics of microbial communities in the wider environment. In this chapter the phylogenetic distribution of the T6SS and related genes was investigated and the genomic context of non-essential T6SS genes was used to develop hypotheses about mechanisms of effector targeting by the T6SS. In addition, the prevalence of T6SS related genes, including of a key T6SS effector superfamily known as type six lipase effectors (Tle), was studied in over 2000 metagenomic datasets representing diverse ecosystems and host niches. Differential overall abundance of Tle, and of particular Tle families, in different environments strongly supports the hypothesis of Tle niche specialisation and suggests that these effectors may play important niche-specific roles.

Introduction

The Type Six Secretion System (T6SS), has emerged as a highly significant player in the bacterial interactome. Widely considered as a weapon of interbacterial warfare, T6SS has also been shown to be involved in virulence towards a very wide range of eukaryotic organisms (Filloux, 2013; Russell *et al*., 2014). However, T6SS is not just involved in killing and several reports link it with general physiological functions such as attachment/biofilm formation, the general stress response, osmotolerance and maintaining pH homeostatis by secretion of H+ ions (Enos-Berlage *et al*., 2005; Aschtgen *et al*., 2008; Weber *et al*., 2009; Gueguen *et al*., 2013; Zhang *et al*., 2013). In addition, roles for the T6SS in activities such as cell-cell signalling and phage defence have been proposed (Gibbs *et al*., 2008; Russell *et al*., 2014). Moreover, as functional characterisation of these systems has largely been achieved through of the use of pathogenicity assays in a range spectrum of model organisms, their role outside of virulence may be underreported (Schwarz *et al*., 2010). Indeed, how frequently T6SS is used in the environments and for what roles is still uncertain. However, the data from genomic and experimental studies in conjugation with data from the growing amount of metagenome data, may allow significant insights into the physiological roles of these secretion systems in their natural context i.e. within communities.

Previous approaches along these lines were limited to certain environments and to the T6SS itself, which has several different roles (Persson *et al*., 2009). Ultimately, the functionality of the T6SS depends on what it secretes. Until recently, very few substrates for the T6SS had been identified, but now our understanding of T6SS effectors is much more complete. Not only have T6SS effector superfamilies been identified, but new studies have revealed how T6SS effectors associate with the apparatus and how they might be delivered to target cells. T6SS effectors are hypothesised to be associated with the T6SS in several different ways with varying degrees of evidence (Shneider *et al*., 2013).

The archetypical effector is secreted through the Hcp tubules, and it is the Hcp protein itself which acts as a chaperone to stabilise these effector molecules and localise them to the T6SS apparatus (Silverman *et al*., 2013). In addition, a subset of the secreted structural T6SS components Hcp and VgrG possess C-terminal

extensions which can have effector activity (Blondel *et al*., 2009; Pukatzki *et al*., 2009).

Structural studies indicate that PAAR proteins, which are often encoded within T6SS loci, localise to the tip of the trimeric VgrG needle where they are hypothesised to "sharpen" the needle (Shneider *et al*., 2013). Many hypothetical PAAR genes code for N-terminal or C terminal extensions which presumably have effector activity. Indeed, such activity has been demonstrated for the effector Tse6 in *P. aeruginosa* (Whitney *et al*., 2014). In addition, PAAR proteins on the needle of the syringe may facilitate binding of other effectors, and it is possible that, at least in some cases, the extensions of the PAAR proteins serve to bind other effectors, instead of, or in addition to, having effector activity. The function of PAAR as a secretion signal was predicted from genomic analysis previously (Zhang *et al*., 2012).

It is likely that other effectors may be able to bind directly to VgrG proteins, as is proposed for Tle2VC (Dong *et al*., 2013). Other non-core T6SS proteins are often found encoded in T6SS loci or *vgrG* islands, such as COG5435, which is hypothesised to be an adapter molecule for effector association with T6SS (Zhang *et al*., 2012). Other proteins often encoded in T6SS loci or *vgrG* islands may fulfil this or other roles related to effector secretion.

As the theoretical basis for the association of T6SS effectors with the T6SS is being discovered, so too are the effectors, including not just individual effectors but also whole effector families or superfamilies. It was hypothesised in Chapter 2 that a suite of T6SS effectors were encoded in *vgrG* islands. Lipase genes within these islands which were identified as possible effectors were confirmed to be secreted in a T6SS dependent manner subsequently (Dong *et al*., 2013; Russell *et al*., 2013). These T6S lipase effectors (Tle) are now considered to be one of three identified T6SS effector superfamilies, the others being T6S amidase effectors (Tae) and T6S glycosidase hydrolase effectors (Tge) (Russell *et al*., 2014).

While Tae and Tge target the bacterial cell wall, Tle are phospholipases which target the cell membrane, allowing them to antagonise both prokaryotic and eukaryotic cells. Indeed, the first example of a Tle effector secreted in a T6SS-dependent manner was from *Vibrio cholerae*, where a T6SS effector protein was found to bind to *vgrG* and was necessary for efficient killing of amoeba (Dong *et al*., 2013).

Of these effector superfamilies, Tle appears to be most widespread in genomic sequences. In fact, the Tle superfamily is composed of 5 different families which have little detectable sequence homology and are likely to have evolved independently (Russell *et al.*, 2013). This suggests there is a strong selective pressure to have virulence factors which target the cell membrane. Similarly to the T6SS itself, multiple Tle genes can be encoded in the same genome, and these are usually of distinct phylogenetic origins. For example, five Tle genes are present in *Pseudomonas aeruginosa* PA14 (Tle1, Tle3, Tle4 and two copies of Tle5). The existence of multiple Tle families and the fact that several can be found within the same genome suggests that each family may have some degree of specialisation.

Another important class of T6SS effectors are the Rhs/YD repeat proteins (Koskiniemi *et al.*, 2013). In gram negative cells Rhs appears to be exported by the T6SS while in gram positive bacteria it appears to be exported by the general secretory pathway. Similarly to other interbacterial effectors, a cognate immunity protein is encoded adjacently to these Rhs proteins, providing protection for the producing cell against self-intoxication and intoxication by sister cells. Rhs proteins have a conserved core with variable C terminal domains with toxin activity. C-terminal re-assortment is responsible for the diversity of Rhs toxicity (Jackson *et al.*, 2009). Many Rhs proteins contain PAAR sequences, which presumably are responsible for Rhs binding to VgrG.

With new effectors have been and are being experimentally characterised in some model organisms, questions remain as to whether they are strictly T6SS associated or are secreted in other manners, as is the case for Rhs, and whether they have different targets or roles in other contexts. Moreover, the association of many effectors with T6SS is not fully understood. Indeed, several Pfam/COG families with unknown roles are often encoded within operons containing putative or confirmed T6SS effectors. In Chapter 2, genomic analysis was used to formulate hypotheses about the T6SS, several of which have been confirmed (Russell *et al.*, 2013). Genomic analysis by others has also provided useful insights into and hypotheses concerning the T6SS (Boyer *et al.*, 2009; Zhang *et al.*, 2012). In this study T6SS-associated genes were analysed at the genomic level to formulate new hypotheses about T6SS functioning, and to provide the foundation for a study on the roles of the T6SS within environments based on metagenomic data. Importantly, T6SS distribution does not

mirror Tle effector distribution, and certain Tle families are more abundant in certain niches, suggesting niche-specific selection.

Methods

General Bioinformatics

Tle sequences were retrieved form genomic and metagenomic datasets on the IMG and M/IMG (http://img.jgi.doe.gov/) servers by sequential rounds of BLASTP analysis using the bait sequences listed in the supplementary data. Gene neighbourhood and co-occurrence analyses were done using the tools of the same name on the IMG database.

Generation of consensus motifs

Genomic sequences for each Tle were aligned using MAFFT E-INS-I using the BLOSSUM62 substitution matrix (Katoh and Standley, 2013) and metagenomic sequences were added to the genomic alignment using the add fragments option on the MAFFT server. Consensus sequences were derived using WebLogo 3 (Crooks *et al*., 2004), after generation of a non-redundant set of sequences using CD-HIT (Fu *et al*., 2012) to eliminate sequences showing 100% identity.

Construction of Phylogenetic trees

For trees outlining the split between Tle5a and Tle5b, Tle5 sequences and Tli5 sequences were retrieved from IMG genomic datasets by BLASTP searches. For phylogenetic trees of Tle with niche location, sequences were retrieved by BLASTP searches on over 500 genomes with isolation metadata available on the IMG database. Sequences were aligned using the MAFFT E-INS-I using the BLOSSUM62 substitution matrix (Katoh and Standley, 2013). Maximum likelihood trees of Tle5 (left) and Tli5 (right) proteins were generated with MEGA 5 using WAG amino acid substitution model with 1000 bootstrap replicates (Tamura *et al*., 2011).

Results

Phylogenetic distribution of T6SS and its effectors

To assess the taxonomic distribution of T6SS and its effectors in genomes, the frequency of the several genes which are known or thought to be associated with the T6SS was assessed in proteobacterial and non-proteobacterial genomes on the IMG database. These genes were assessed using the COG and Pfam designations. The following COGs were used: COG3157 (Hcp), COG3209 (Rhs), COG3501 (VgrG,) COG3516 (TssB), COG3517 (TssC), COG4104 (common PAAR), and COG5435. The following Pfam were used: Pfam09937 (TagA), Pfam08786 Pfam13503, Pfam05954 (VgrG), Pfam05488 (common PAAR), Pfam13665 (rare PAAR, also known as TagD), Pfam10106 (subset of VgrG with extensions), Pfam05638 (Hcp) and Pfam05593 (Rhs). Pfam09937 (TagA), Pfam08786/COG5435 and Pfam13503 represent putative T6SS-associated genes of unknown function.

Since Pfam designation differs from COG designation as it is based on protein domains rather than overall protein sequence, the results are not directly comparable between COG and Pfam. Some COG have Pfam counterparts e.g. COG5435 and Pfam08786 both represent the same gene. Other Pfams in this study have no COG counterpart, such as Pfam09937 and Pfam13503. Similarly, two distinct PAAR families exist and are designated Pfam05488 and Pfam13665, with the former being also designated COG4104 and the latter not having COG representation. Other genes which have COG identifiers are associated with multiple Pfam domains. For example, COG3501, representing VgrG, is composed of two Pfam domains, of which Pfam05954 was chosen for this study. In addition, a subset of VgrG proteins have a C terminal domain which is represented by Pfam10106.

The prevalence of the Tle superfamily was also assessed. As no identifiers (such as COG of Pfam domains) are specific to Tle genes, Tle encoding sequences were detected using the program BLASTP on the IMG database.

As over half of sequenced genomes belong to Proteobacteria, it was not surprising that most hits were of proteobacterial origin, but even when this is taken into consideration, nearly all hits were much more common in Proteobacteria. When the number of hits for each COG or Pfam was normalised by the number of genomes surveyed, we see that the only genes which occur with appreciable frequency outside

Proteobacteria are Rhs and VgrG (**Figure 3.1 and Figure 3.2**). Similarly, the Tle superfamily was strongly associated with Proteobacteria, occurring at 0.15 copies per proteobacterial genome compared to 0.005 copies per non-proteobacterial genome.

Gene neighbourhood analysis of the T6SS core components TssB and TssC from genomes on the IMG database confirmed that these are almost always encoded with T6SS loci and are therefore reliable proxies for T6SSs abundance. From the analysis, it was seen that while T6SS are particularly enriched in Proteobacteria, they are also present in Acidobacteria, Gemmatimonadetes, Nitrospirae, Planctomycetes and Verrucomicrobia. Some of the incidences of these T6SS-associated genes in non-Proteobacteria are due to this, but the effect on overall numbers is negligible. As expected, the genomic distribution of Tle is similar to the previously reported genomic distribution of T6SS, as both are also occasionally found in *Acidobacteria* and *Planctomycetes* as well as Proteobacteria (Zhang *et al.*, 2012). However, they also occur in Bacteroidetes. A very small number of TssB genes were present in Bacteroidetes, but in these instances they were encoded in a locus which contains with *vgrG*, *Rhs*, an ATPase gene. All Tle detected in Bacteroidetes are found within these loci, which encode another phage-derived secretion system known as the *Photorhabdus* virulence cassette (PVC) (Yang *et al*., 2006; Penz *et al*., 2010). The appearance of T6SS core genes in these PVC loci is quite rare, but they can include *tssM* and *tssB* (data not shown). The involvement of VgrG, Rhs, Tle and perhaps very occasionally other T6SS proteins, in a different secretory pathway underscores the similarities of various secretion systems.

Both *vgrG* and *rhs* genes appear in other contexts too. Of 303 *rhs* genes in finished non-proteobacterial genomes, just over 4% are in genera which contain a T6SS while 25% occur in Actinobacteria and 37% occur in Firmicutes with 17% occurring in Bacteroidetes. That there are high numbers of Rhs-related proteins outside of Proteobacteria is understandable as these are known to also mediate interbacterial killing in gram positive bacteria and likely are part of the PVC (Koskiniemi *et al*., 2013).

Only 10% of non-proteobacterial *vgrG* genes are found in genomes with a T6SS locus, while 20% are found in Bacteroidetes, where they are likely part of the PVC. It is not clear what function the remaining 70% of *vgrG* have. These are primarily found in Actinobacteria and Firmicutes in loci which contain phage genes such as

COG3497 (85%) and an ATPase (75%), suggesting it is a prophage which may have similarity with the T6SS. Some of these loci also have phosphatase and *fha* genes, reminiscent of the post-translational control system which regulates some T6SS (Mougous *et al*., 2007). While the other T6SS associated genes are infrequently found outside of bacteria possessing the T6SS, when they do occur there is often a link to phage related genes.

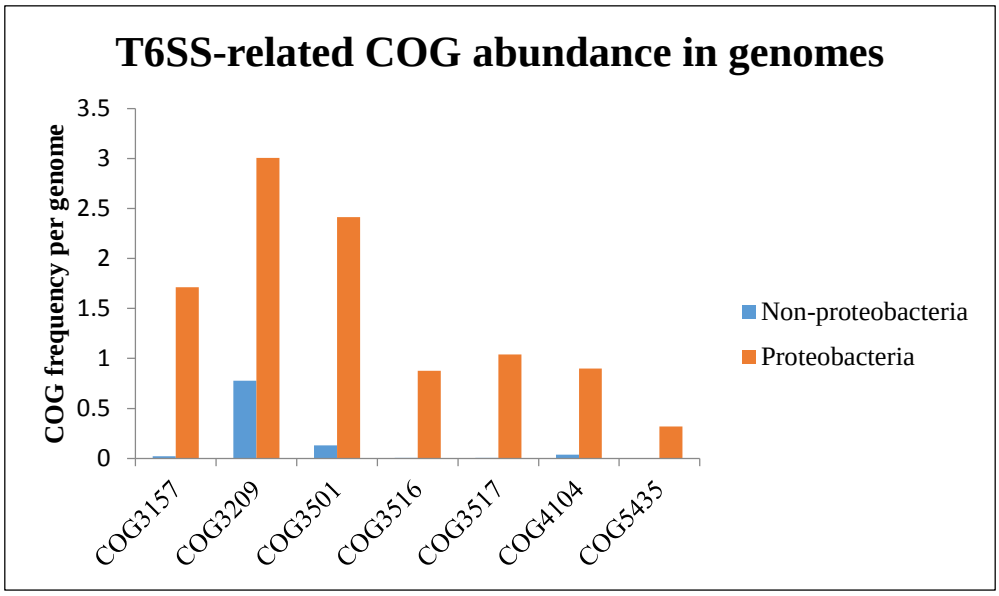


Figure 3.1: Abundance of T6SS-associated COGs in Proteobacterial and non-proteobacterial genomes. Total COG incidence was normalised by the number of genomes assessed: approximately 7500 non-proteobacterial genomes and 6000 proteobacterial genomes. Only COG3209 (Rhs) occurs often outside Proteobacteria.

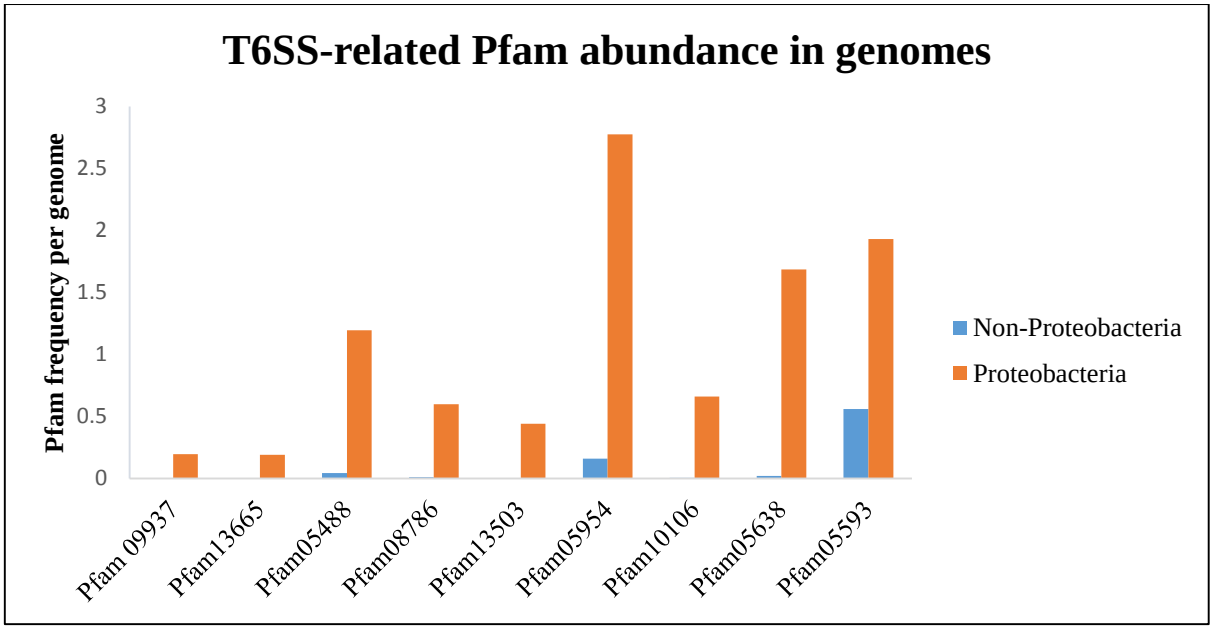


Figure 3.2: Abundance of T6SS-associated Pfams in proteobacterial and non-proteobacterial genomes. Total Pfam incidence was normalised by the number of genomes assessed: approximately 7500 non-proteobacterial genomes and 6000 proteobacterial genomes. Only Pfam05593 (Rhs) occurs often outside Proteobacteria.

The Pfam abundance is broadly similar to the abundance of COG as most of the putative T6SS-associated Pfam are rare outside of Proteobacteria, suggesting they are exclusive to the T6SS. However, the Pfam05954 is more common than COG3501 as due to the evolutionary origin of VgrG some phage related genes also belong to this Pfam. Conversely, not all of the diverse COG3209 genes are annotated as Pfam05593, so the incidence of COG3209 is higher. Ultimately, these proteins are not exclusive to the T6SS.

Gene co-occurrence gives insight into protein interactions.

Proteins which interact are often encoded adjacently or in some cases encoded within the same gene due to gene fusion (Dandekar *et al.*, 1998; Pukatzki *et al.*, 2007; Boyer *et al.*, 2009). To understand the associations between some of these new effectors or T6SS core components which might be responsible for their association with the T6SS or with each other, cassette searches were performed on the IMG database to identify where two or more proteins are encoded in the same genomic locus. Given the basis of the Pfam and COG annotations, these are analysed separately (**Figure 3.3 and Table 3.1**).

The values reported are likely to be underestimates, because not every gene may have correctly been assigned a COG or Pfam value, and as discussed above, Pfam05954 is found in some phage-related genes which are not *vgrG*. Another factor which results in lower co-occurrence values is the fact that in draft genomes some genomic loci are not complete. For example, COG3516 and COG3517 co-occur 91% and 77% of the time in all genomes analysed but in finished genomes this rises to 96% and 90%, respectively. As would be expected from the direct association of TssB and TssC in the T6SS sheath, these genes have the highest rate of co-occurrence. For similar reasons Hcp and VgrG might be expected to have a high rate of co-occurrence. However, this is not the case. The high number (80%) of *vgrG* genes with no adjacent *hcp* is not altogether surprising as *vgrG* islands which lack *hcp* are common, which may be expected if these islands were a source of additional effectors which were not transported through the Hcp tubule. However, our current understanding of T6SS does not allow for so many *hcp* genes (70%) to occur without a partner *vgrG*. Some *hcp* genes within T6SS loci may be too far from *vgrG* genes for to be identified as co-occurring, but at least 25% of Hcp is found without any VgrG or TssB genes in its vicinity. In Chapter 2, it was highlighted that *hcp* without a partner *vgrG* gene are phylogenetically similar, possibly indicating some Hcp are not associated with the T6SS. The frequency with which *hcp* occur in the absence of other T6SS genes across all proteobacterial genomes is consistent with this hypothesis.

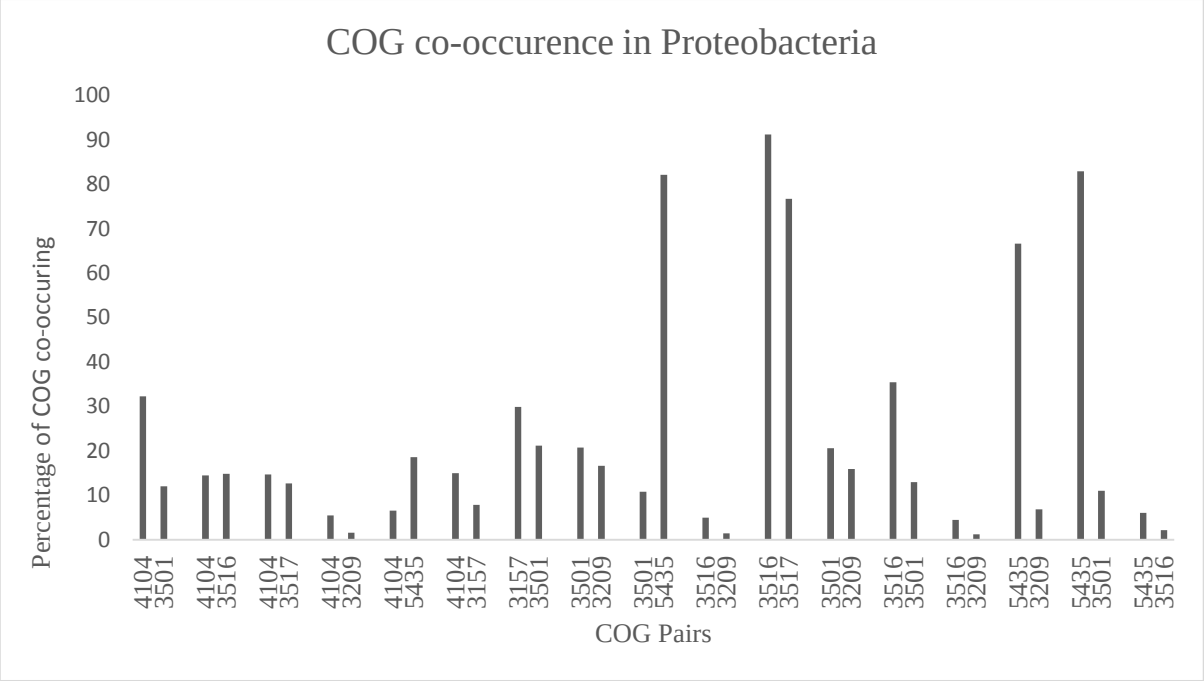


Figure 3.3. COG co-occurrence: The frequency with which two COGs are encoded within the same genomic location was assessed using the cassette search tool on IMG. The blue bars represent what percentage of one COG is found in the vicinity of another. For example, 33% of the COG 4104 genes co-occur with COG3501 genes, while 13% of COG3501 gene co-occur with COG4104 gene.

Despite being rarely encoded near the T6SS core component TssB, Cog5435 co-occurs with *rhs* and *vgrG* genes with high frequency. Conversely, the majority of *vgrG* and *rhs* genes do not co-occur with COG5435, suggesting that COG5435 acts on VgrG/Rhs in a manner dispensable for general T6SS activity. It has previously been hypothesised that COG5435 is an adapter between an effector and the T6SS apparatus, based on its prevalence among *vgrG* islands (Zhang *et al*., 2012). A three gene operon consisting of a PAAR interbacterial toxin, its cognate immunity gene, and COG5435 is consistent with this hypothesis (Whitney *et al*., 2014). This gene is also represented by the Pfam08786, which belongs to the Mog1/Pspb superfamily, but the relevance of this is unclear. Mog1 is known to be both a component and modulator of the cardiac sodium channel, and it is possible that Pfam08786 may have a similar role T6SS protein complex in bacterial cells (Wu *et al*., 2008).

PAAR subsets are associated with distinct accessory components.

Like Pfam08786 (COG5345), TagA (Pfam09937) and Pfam13503 have co-occurrence profiles which suggest they may be adapter proteins, as they are encoded next to PAAR domains frequently **(Table 3.1)**. Moreover, the association seems to be specific for either of the two PAAR subsets. The less common PAAR (Pfam13665) and *tagA* strongly co-occur, with reciprocal values above 70%. This suggests that these proteins rely on each other for function. Conversely, it co-occurs less frequently with the Pfam05448 PAAR proteins. When the prevalence of both PAAR sub-families in the database is taken into account, *tagA* is over 100 times more likely to co-occur with the Pfam13665 than with Pfam05448.

In contrast, Pfam13503 and Pfam08786 both co-occur more often with PAAR family Pfam05448, with values of 25% and 40%, respectively. When the greater abundance of this PAAR family is taken into account as above, these proteins co-occur 57 and 6 times more frequently with this PAAR subset. This suggests that these proteins are adapter proteins for this specific PAAR. As Pfam13503 and Pfam08786 co-occur with the same PAAR, they would also be expected to co-occur with each other to some degree. However, they are 50 times less likely to co-occur than would be expected on this basis. This may indicate that they have the same function and are not needed simultaneously, or that they have mutually exclusive functions which interfere with each other.

The existence of two PAAR subfamilies and the specificity of their association with other Pfams outlined above, suggested that these proteins have distinct binding sites. Work on the PAAR binding site shows that the Pfam05488 subset bind to two hydrophobic residues on the B sheet of VgrG (Shneider *et al*., 2013). Perhaps because not enough binding pairs have been characterised, this site does is conserved enough to preclude binding sites other VgrG which are predicted to associate with the alternative PAAR subfamily.

The subset of VgrG with a C-terminal Pfam10106 domain also have interesting co-occurrence profiles with PAAR genes. While *vgrG* genes in general are slightly (1.6 times) more likely to occur with the Pfam13665 PAAR subset, *vgrG* genes encoding a Pfam10106 C terminal domain are 10 times more likely to co-occur with Pfam05488 PAAR genes. Interestingly, Pfam10106 has the same level of co-

occurrence with Pfam13503 and the common PAAR (15%), but never co-occurs with Pfam08786, despite the fact that they both co-occur with Pfam05488. This may indicate that these VgrG proteins may bind more often to the more common PAAR subset, and the nature of this protein-protein interactions results in a situation where only Pfam13503 can fulfil the adapter role and Pfam08786 cannot. Alternatively, it may suggest that Pfam13503 and Pfam08786 have different roles, and where *vgrG* genes containing the Pfam10606 domain are concerned, the action of Pfam08786 proteins is redundant or even prohibitive.

Tle association with the T6SS apparatus.

Investigation of co-occurrence profiles of putative T6SS adapter proteins/chaperons suggested interesting hypothesis about which protein-protein interactions occur in the T6SS. However, it was not clear whether stand-alone effectors i.e. effectors which are not encoded as C-terminals domains of T6SS components, associate with the T6SS in this manner. To address this question, we analysed co-occurrence of the Tle effector superfamily with T6SS components. As Tle are large proteins it was hypothesised that they were not secreted though the T6SS apparatus but delivered to target cells by association with the VgrG protein. Indeed, Tle2 from *Vibrio* has been shown to associate with its cognate VgrG protein in an immuno-precipitation assay (Dong *et al*., 2013). However, this could be due to both proteins binding to a third protein such as PAAR, so how Tle associate with the T6SS is an open question. As mentioned, Pfam designations for Tle families are not entirely accurate, as they are assigned to lipase proteins which do not appear to be effectors, and are not assigned to all members of the Tle family. However, in some cases it is possible to roughly estimate Tle co-occurrence using Pfam domains, though not for the Tle5 family **(Table 3.2)**. From gene neighbourhood analysis of Tle BLASTP results, it was seen that over 50% of Tle hits were encoded adjacently to a *vgrG* gene (data not shown). Therefore, the lower levels of co-occurrence between Tle and VgrG suggest that all the co-occurrence values reported in **Table 3.2** may be underestimates. All Tle analysed are infrequently encoded among the minor PAAR sub-family, but only Tle2 was infrequently encoded among the major PAAR sub-family. Despite this, like the other Tle families, Tle2 was encoded often among Pfam13503. However, these

proteins may still associate with PAAR. For example, the Pfam13503 member VasW and PAAR are both required for full T6SS activity, despite not co-occurring in *V. cholera* genome (Miyata *et al*., 2013; Shneider *et al*., 2013). What is striking is that while the putative adapter Pfam13503 is commonly associated with Tle, the other putative adapter Pfam08786 is not at all, again suggesting these proteins have very distinct roles or possibly mutually exclusive roles. Interestingly, structural modelling suggests both Tli4 and a subset of the Tli5 proteins are similar to the Pfam08786 (**Table 3.3**). The implications of this are unclear, but it may indicate that these proteins are targeted to the T6SS in the same manner as the Pfam08786 proteins. While these immunity proteins are thought to work as defence against attack, the local phospholipase activity of Tle associated with the T6SS could prove detrimental for the attacking cell. Another remarkable observation is that apart from Tle2, the majority of the Tle associated with *vgrG* are associated with *vgrG* subset containing the Pfam10106 domain, despite the fact that this Pfam appears in less than 25% of proteobacterial VgrG. This may reflect that this domain is necessary for Tle binding, either directly or indirectly. Fusions can be indicative of protein-protein associations. Some VgrG and Hcp, termed evolved VgrG or Hcp, have C-terminal extensions with effector activity (Pukatzki *et al*., 2007; Blondel *et al*., 2009), and indeed some Tle1 genes are actually evolved VgrG and Hcp proteins carrying lipase domains, while other Tle1 lipase domains can be found as extensions of PAAR proteins. In addition, some Tle2 hits were to PAAR containing proteins. As association with Hcp, PAAR, and VgrG covers all current models for how substrates are secreted by the T6SS, it is not very informative on the manner of Tle interaction with the T6SS apparatus. Instead, this result possibly reflects that effector evolution is often driven by the addition of an effector domain to existing T6SS-associated protein.

Table 3.1. The co-occurrence of genes represented by various Pfam domains was assessed in over 6000 Proteobacteria. The table shows the percentage of the Pfam (left) which co-occur with other Pfam (top) is shown. For example, 73% of TagA are encoded in the vicinity of a gene containing a Pfam13665 domain whereas 75% of Pfam13665 are encoded in the vicinity of a *tagA* gene.

	Feature	**Pfam09937** TagA	**Pfam13665** Uncommon PAAR	**Pfam05488** Common PAAR	**Pfam08786** Possible adapter	**Pfam13503** Possible adapter	**Pfam05954** VgrG	**Pfam10106** VgrG subset	**Pfam05638** Hcp	**Pfam05593** Rhs
Pfam09937	TagA		73	5	3	0	68	1	41	5
Pfam13665	Uncommon PAAR	75		5	3	0	60	1	38	5
Pfam05488	Common PAAR	1	1		20	9	35	9	13	21
Pfam08786	Possible adapter	1	1	41		0	35	0	5	40
Pfam13503	Possible adapter	0	0	24	0		61	23	23	4
Pfam05954	VgrG	5	4	15	8	10		17	18	15
Pfam10106	VgrG subset	0	0	16	0	15	71		16	11
Pfam05638	Hcp	5	4	9	2	6	29	6		7
Pfam05593	Rhs	1	1	13	12	1	21	4	6	

Table 3.2. Co-occurrence of Tle gene with selected T6SS genes. As co-occurrence with TagA or the minor PAAR subset was less than 1% for all Tle genes these data were omitted from the final table.

		Pfam05488	Pfam13503	Pfam10106	Pfam05954	Pfam05638	Pfam05593
Tle1	Pfam09994	11.21412804	16.86534216	14.61368653	18.32229581	8.653421634	0.971302428
Tle2	Pfam01764	0.982197667	11.23388582	0.368324125	9.883364027	3.49907919	0.675260896
Tle3	Pfam11678	23.80952381	0	47.20496894	41.82194617	6.832298137	0.207039337
Tle4	Pfam07819	6.481481481	0.854700855	26.13960114	24.35897436	7.905982906	0

Conserved islands are not predicted by genomic co-occurrence

Due to horizontal gene transfer the same genomic regions can be distributed through a range of phylogenetically diverse bacteria. One such genomic region, PA0095-PA0101, was identified in Chapter 2, and contains several of the T6SS associated genes which were used for cassette searches above e.g. TagA, VgrG, and Pfam13665. In some cases additional T6SS such as Tle2 genes appear to have been incorporated into this genomic region. However, several of the genes of this region do not seem to be associated with other T6SS genes except specifically when within this island **(Figure 3.4).**

Moreover, except for PA0096, these genes are members of larger COG families, suggesting that genes have general physiological functions but are can be associated with T6SS under specific conditions. Unfortunately, the possible roles of these proteins in a secretion process are not clear, and structural modelling with Phyre2 does not suggest that these proteins have any similarity to T6SS components **(Table 3.3)**. It is possible that COGs which have a general physiological function in the bacterial cell have a detrimental effect on target cell physiology, but they may simply be associated with regulation of the T6SS or cell repair after a firing event.

The TagA protein is encoded within this island. The *tagA* genes are sometimes found to be fused the *tagB* gene, which contains a pentapeptide repeat domain, and this occurs in yet another widespread T6SS-associated genomic island (Shalom *et al*., 2007). Pentapeptides are common proteins which are not usually associated with the T6SS. These domains are thought to mediate protein-protein interactions, which may indicate that TagA may associate with effectors or T6SS components in this manner. In loci containing *tagA*/*tagB* fusions (such as *BPSS1504)* the Pfam13665 PAAR gene is also present. Therefore these two separate islands have several shared proteins. One protein which is lacking any homologues within this island is the hypothetical protein encoded by PA0096. Instead, another protein, TagC, is encoded within this island. Despite having very little sequence similarity, both TagC and PA0096 (which both co-occur with Pfam09937 and Pfam13665) are modelled with high confidence (>90%) to the crystal structure of the bacteriophage p2 membrane-piercing protein gpV, which is analogous to the gp5-like domain of at the C terminal of unevolved VgrG, and, like the common PAAR, is stabilised by a metal ion (Browning *et al*., 2012; Shneider *et al*., 2013). Based on this, it is possible these

TagC/PA0096-like proteins have a similar membrane piercing function to PAAR, and can work in conjugation with the rare PAAR subset.

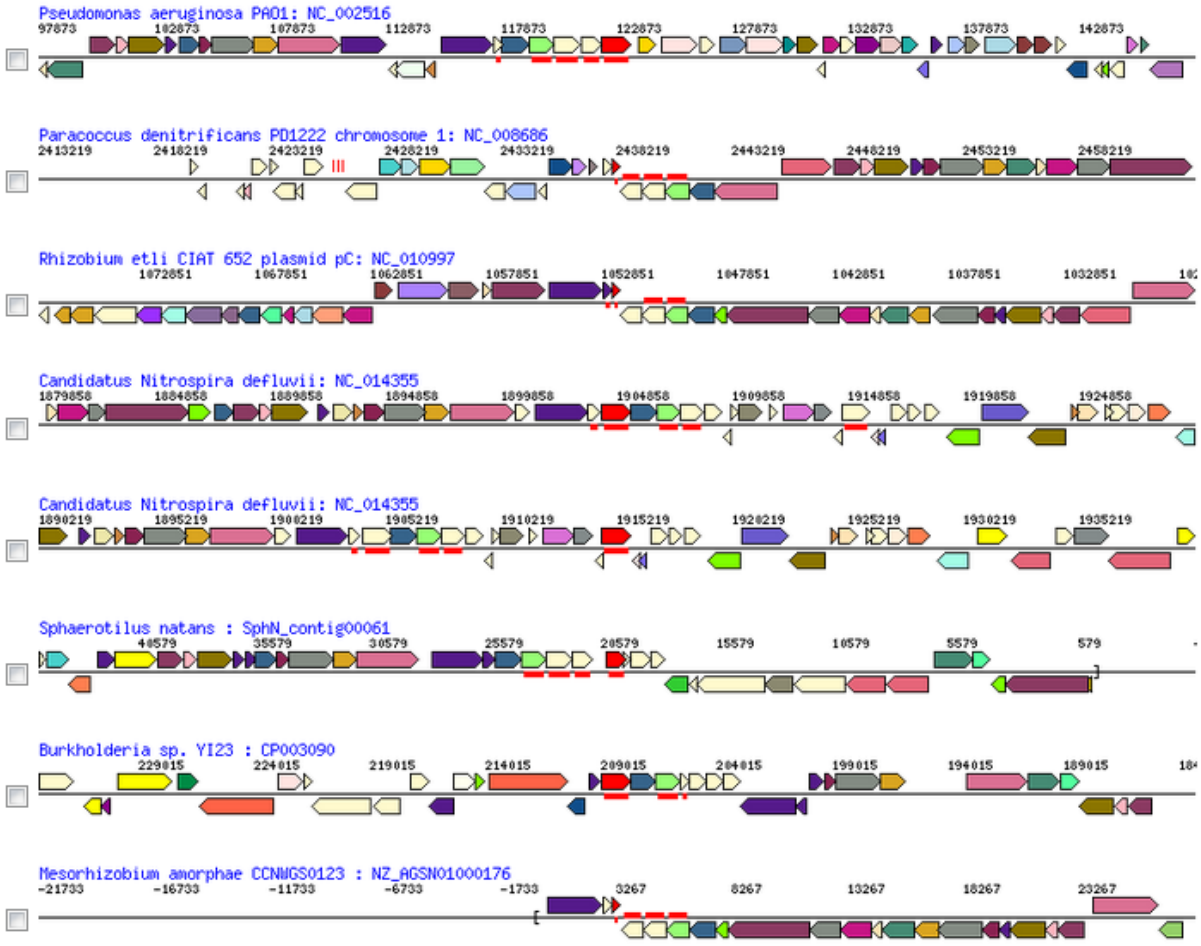


Figure 3.4. Co-occurrence of genes suggests association. Homologues from different bacteria are coloured similarly. From left to right in *P. aeruginosa* PAO1: Purple: VgrG, White (red underlining): PA0096, Blue: TagA, Green: Cog0304, White (red underlining): Minor PAAR. White (red underlining): PA0100, Red: PA0101.

Table 3.3. Structural modelling of selected T6SS proteins using Phyre2.

Protein	Phyre2 result	Confidence %	Identity %	Length %
Pfam09937 TagA	Molybdenum-containing oxidoreductases-like dimerisation domain	71.4	27	10
Pfam08786	PA0094 (Mog1/PsbP superfamily)	100	100	98
Pfam13503	ArsR-like transcriptional regulators	47	20	19
TagC	Crystal structure of the bacteriophage p2 membrane-piercing protein2 gpV	93	19	66
PA0096	Crystal structure of the bacteriophage p2 membrane-piercing protein2 gpV	99.8	14	85
PA0098	3-oxoacyl-[acyl-carrier-protein] synthase 2	100	14	99
PA0100	Two-component system sensor histidine kinase/response	99	12	93
PA0101	PBS lyase HEAT-like repeat	99.9	13	57
Tle1	Crystal structure of a functionally unknown protein Lpg2422 from *Legionella*	97	21	27
Tle2	Crystal structure of lipase from *Gibberella zeae*	100	30	27
Tle3	*A. fulgidus* lipase with fatty acid fragment and calcium	98	20	19
Tle4	Crystal structure of the gene lin2722 products from *Listeria. innocua*	99.5	31	24
Tle5a	Phospholipase D	100	15	40
Tle5b	Phospholipase D	100	24	51
Tli1	Structure of the C-terminal domain of human thrombospondin-2	39	24	26
Tli2	SelT/SelW/SelH selenoprotein domain	65	19	18
Tli3	Benzalacetone synthase	86	16	10
Tli4	PA0094 (Mog1/PsbP superfamily)	98	11	36
Tli5a	PA0094 (Mog1/PsbP superfamily)	86	9	32
Tli5b	Crystal structure of AlgK from *Pseudomonas fluorescens* WCS374	100	16	91
Tli5b	Hcp	99	26	64

T6SS frequencies in metagenomes

The use of model organisms to study T6SS has provided many insights into the functioning of T6SS, but larger question of the roles of the T6SS within communities cannot be answered in such a reductive manner. The recent dramatic increase in metagenome sequences being made publically available provides an ideal opportunity to address this question. The IMG database now contains over 2000 metagenomes, many of which can be assigned to the broader environmental categories: Marine, Fresh, Human, Arthropod, Rhizosphere/Soil, and Engineered. These environments were examined for T6SS-related gene distribution. As T6SS is mainly associated with proteobacteria, the number of T6SS genes was normalised by dividing by the predicted number of proteobacterial genes in the metagenome.

T6SS abundance in metagenomes was assessed via the frequency of COG3516, which represents the T6SS core component TssB that is exclusive for the T6SS **(Figure 3.5)**. To assess whether identifying genes by COG was a reliable approach using metagenomic data, the genomic context of 318 COG3516 hits from metagenomes was analysed in contigs greater than 1 kb in length using the gene neighbourhood tool of IMG/M. A total of 295 contigs (92.8%) encode an additional T6SS related gene. Some of the remaining 23 contigs may represent T6SS loci, as 18 have sequence on only one side of the COG3516 gene, two contain only COG3516 and three encode hypothetical protein either side of COG3516. This suggests that searching for functions by COG on IMG database is accurate approach which does not yield high degrees of false positives.

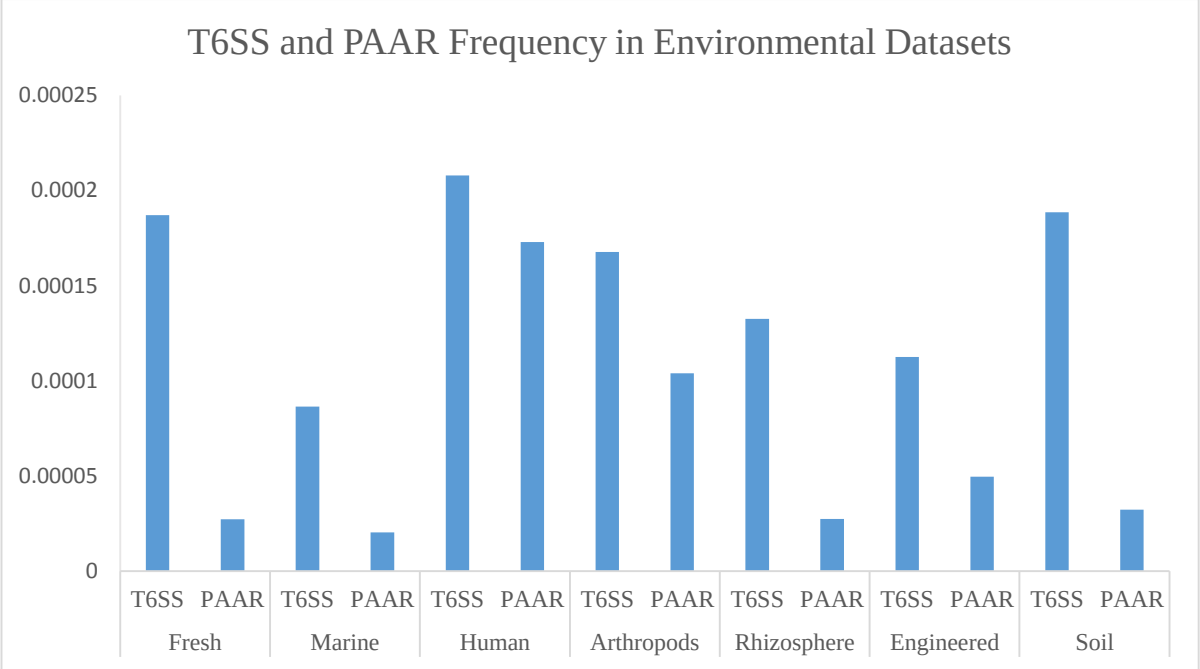


Figure 3.5. T6SS-Associated Genes in Environmental Categories: The abundance of T6SS and PAAR was assessed by COG occurrence in environmental categories normalised by the predicted proteobacterial gene counts.

The results suggest that environments are not selecting for different numbers of T6SS. As mentioned, T6SS numbers are not necessarily indicative of T6SS-dependent killing. The abundance of PAAR genes (COG4104), which are possibly the sites of binding for many effectors, in different environments shows more variation **(Figure 3.5)**. In particular it is high in human and arthropod metagenomes, while being occurring at similar levels in the remaining environments. Similar numbers of PAAR in rhizosphere and bulk soil was not expected as the level of biological activity and competition is higher in this niche. In fact T6SS is slightly more abundant in the bulk soil than the rhizosphere. A previous study of T6SS frequency in aquatic metagenomes found higher numbers in productive waters (Persson *et al*., 2009). PAAR genes are only one of several alternatives facilitating effector secretion by the T6SS. For this reason, and because its phylogenetic distribution seems to indicate it very often associated with the T6SS, the abundance of an effector superfamily, Tle, was studied.

No Pfam or COG domains were both specific and exclusive to the various Tle families. Therefore, to avoid false positive and false negatives, databases were searched for Tle

sequences using the program BLASTP on the IMG database. To establish the veracity of this method and generate bait sequences, multiple rounds of BLASTP analysis were performed using genomic datasets on IMG. Results were limited to hits encoded adjacent to a *vgrG* gene using the gene neighbourhood tool in IMG. After several iterations a list of baits representing the fewest amount of diverse Tle needed to obtain all Tle hits from genomic data was generated **(Supplementary File 3a)**. In absolute terms, 2078 Tle were identified in metagenomes and the numbers of Tle in an environmental category ranged from 31 in soil to 1059 in human. In relative terms the numbers of Tle ranged from approximately 1 per 4000 prokaryote genes in arthropod metagenomes to 1 per 700,000 prokaryote genes in soil metagenomes.

Environment type determines Tle numbers

Given the diversity of the environmental categories studied it is perhaps not surprising that differences in total Tle frequency were evident between these metagenome sets (**Figure 3.6)**. However, the extent of these differences is quite striking. It is particularly notable that the arthropod, rhizosphere and human metagenomes contain many *Tle* genes, while Tle is relatively infrequent in aquatic environments and in the bulk soil. This suggests that Tle function is important within these certain niches.

This may simply reflect the fact that Tle are involved in interactions with eukaryotes, or that Tle abundance may reflect the level of activity/competition within these niches. Indeed, previous analysis of aquatic metagenomes found that secretion system genes were much more prevalent in productive waters (Persson *et al*., 2009). However, a role for Tle proteins in the marine and soil ecosystems cannot be ruled out. Other factors may also contribute to a lower Tle abundance. The lack of Tle in aquatic environments might be due to the lack of activity within this niche or could possibly reflect that open water niches in aquatic environments are not conducive to a contact-dependent method of bacterial killing. Greater sampling of sediments or sites of high bacterial concentrations from these environments may result in greater Tle representation in aqueous environments, though in the limited amount of sediment-based aquatic metagenomes Tle genes are not more abundant

One caveat to these results is that certain sites can dominate an environmental category. For example, of the freshwater genomes, roughly 45% of the prokaryote

genes and 90% of T6SS comes from one site; Wetland microbial communities from Twitchell Island in the Sacramento Delta. Despite this, only 12% of freshwater Tle comes from this source. A closer look at the contribution of the various metagenomes to the Tle frequency in their environmental category reveals interesting observations. Fungus-associated arthropod genomes make up about 5.6% of total arthropod metagenome DNA, but can account for at least 25% of all Tle, suggesting that they may be involved in bacterial-fungal interactions.

Within the human metagenome category, Proteobacteria are much more common in the oral niche than in the stool**.** As secretion systems are generally more frequent in Proteobacteria, the contrasting abundance of this phylum in these two niches suggests that more bacterial killing occurs in the mouth. Tle frequency being significantly higher in the mouth than in the gut, even allowing for the differential proteobacterial representation. Indeed, the latter niche contributes less than 20% of the amount of Tle that would be expected based on its proteobacterial gene count. Whether Tle should be considered overrepresented in the mouth or underrepresented in the gut is a matter of perspective, but these data certainly support a role for Tle proteins in contributing to the bactericidal activity in the oral niche.

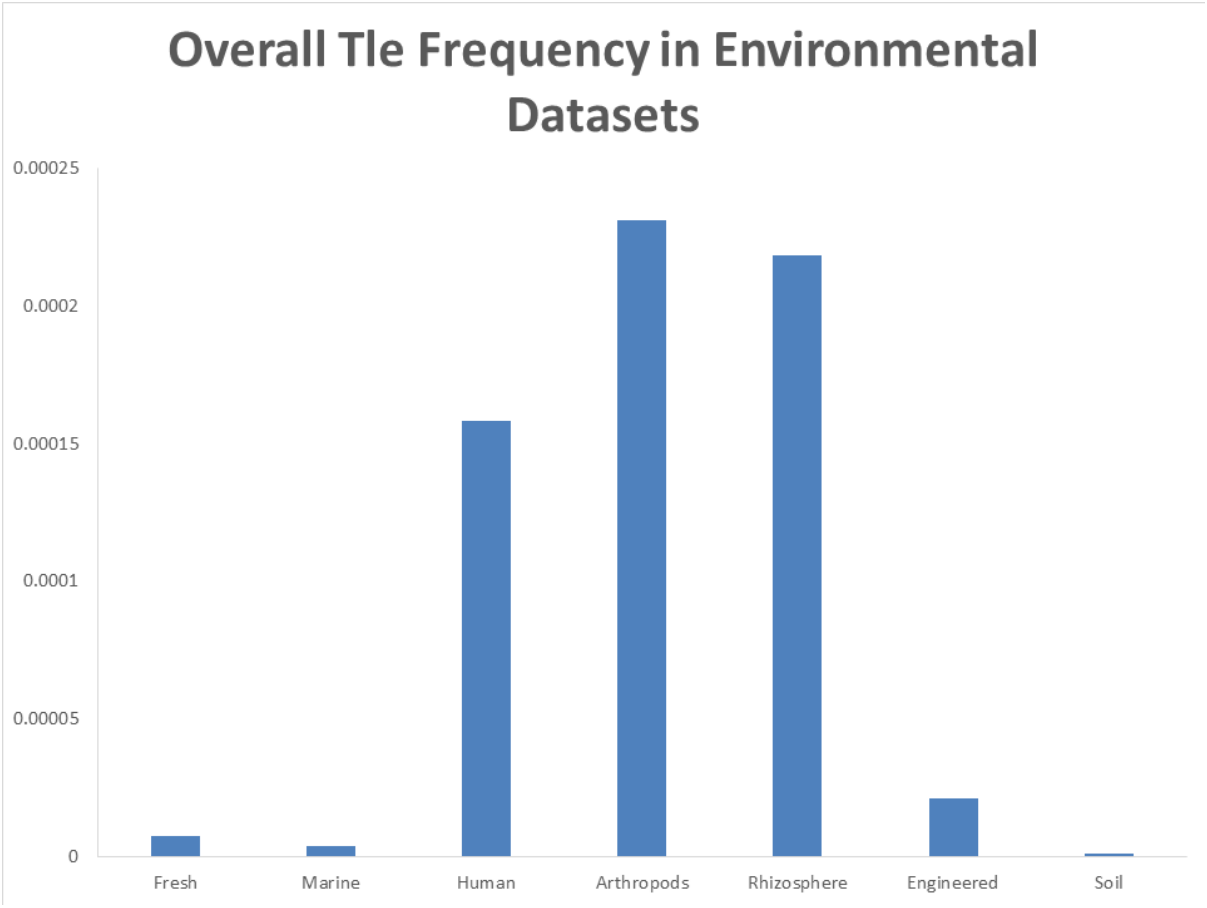


Figure 3.6. Total Tle abundance (TLE/proteobacterial gene count) in environmental datasets shows variation, with high frequency of Tle correlated with host-associated metagenomes.

Tle abundance differs from T6SS abundance

One of the key reasons for analysing at the level of effector molecules instead of secretion systems was the possibility that Tle abundance would be more informative than general T6SS abundance. While Tle function would appear to be specific, the T6SS itself is more functionally diverse The T6SS could have many functions, and therefore its abundance might not necessarily reflect Tle abundance. As shown in **Figure 3.7**, the abundance of T6SS does not reflect the abundance of Tle, as T6SS numbers are relatively similar across the metagenomes datasets. The human metagenomes have highest amount of T6SS, but T6SS is only 2.4 times more abundant in this environment than in the marine metagenomes where it is least frequent.

A corollary to the disparity between T6SS and Tle abundance is that T6SS does not seem to be enriched by the environments which might be expected to have greater

available energy or the presence of a eukaryotic host. For example, the rhizosphere and host-associated arthropod metagenomes have proportionally fewer T6SS per prokaryotic genes than bulk soil and engineered metagenomes respectively, in spite of the fact that the latter two environments would not be expected to contain a higher proportion of eukaryotes. The discrepancy between T6SS and Tle abundance may suggest that T6SS may be playing a role not related to killing in the environments where Tle is underrepresented. Alternatively, killing could be achieved using a different set of effector molecules.

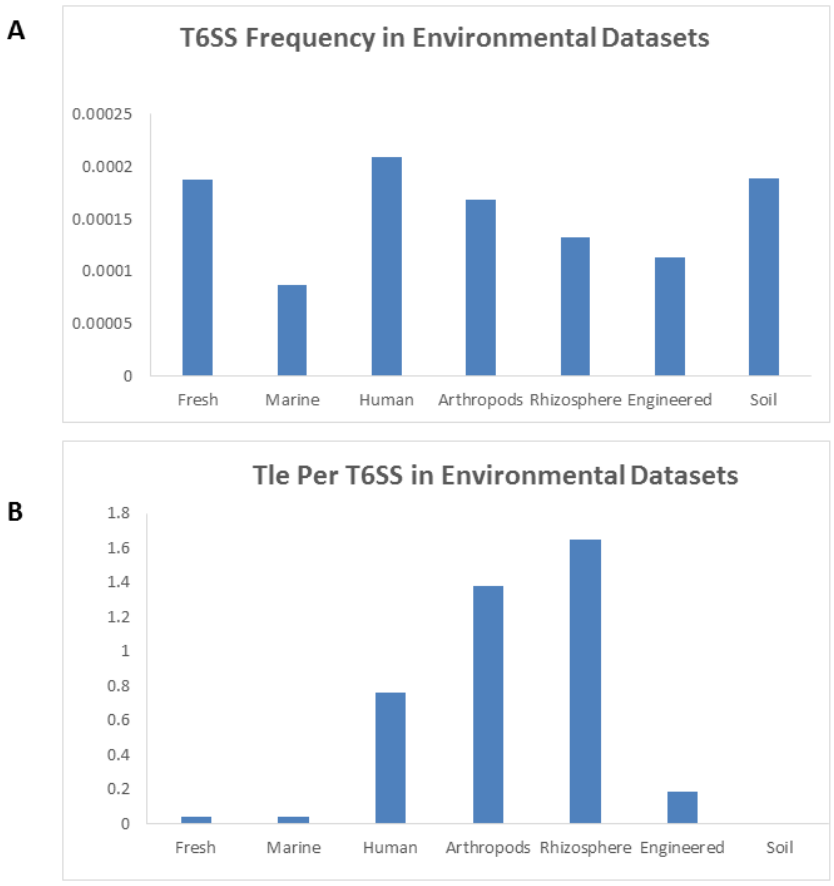


Figure 3.7. (A) T6SS abundance in environmental datasets was calculated similarly to Tle abundance (T6SS gene COG3516/proteobacterial gene count). **(B)** Comparison of T6SS and Tle frequency shows that T6SS and Tle abundance are dissimilar. Due to the limited variation in T6SS numbers, Tle genes per T6SS are highest in host-associated metagenomes, mirroring the overall Tle distribution.

Divergent Tle family representation suggests niche-specific specialisation of Tle families

As Tle genes are often present on horizontally transferred *vgrG* islands they are not necessarily constrained by the general evolution of the bacteria they reside in (Barret *et al*., 2011). For example, *P. aeruginosa* PA7 has Tle2 but other *P. aeruginosa* strains such as PAO1 do not. Does the evolution of five different families of Tle represent some degree of specialisation? As mentioned, Tle families do have unique sites of activity for phospholipid cleavage (Russell *et al*., 2013). If the evolution of five separate families occurred because each family was in some way specialised, and therefore more useful in certain niches, it follows that various environments would show variation in the frequency of Tle families. If the contrary is true, and no niche is exerting selective pressure to favour particular Tle families, all families should be expected to occur with similar frequencies.

While Tle occur with similar frequencies in some niches, other niches show large variations (**Figure 3.8)**. Distribution in arthropod metagenomes is consistent, with only a 3 fold difference between the number of the least common and most common Tle families. The rhizosphere has a more disparate Tle distribution, with the most abundant family (Tle2) being 10 times more abundant that the least abundant family (Tle4). Though Tle are generally uncommon in aquatic environments, Tle1 and Tle2 are relatively enriched in freshwater and marine datasets, respectively.

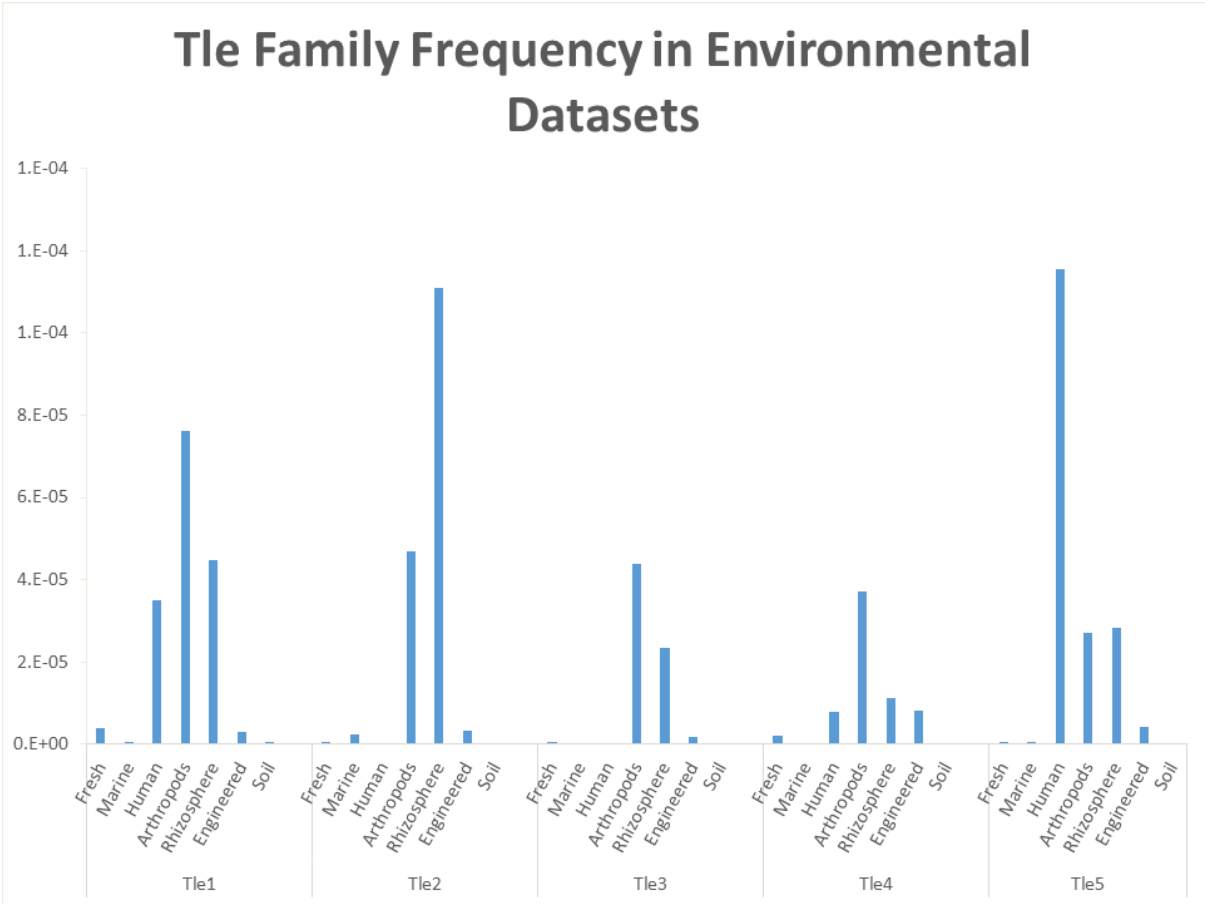


Figure 3.8. The frequency of Tle families in metagenomic data from various environmental categories shows marked variation in Tle family distribution in several areas. The highest numbers of three families occur in arthropods, while Tle2 and Tle5 are most common in the rhizosphere and in the human microbiome. Data is again displayed as Tle hits per proteobacterial genes; Tle family abundance across all metagenomes, weighted by proteobacterial abundance per contributing environment, reflects aggregate length of bars per Tle family.

The most striking variation in Tle family abundance is found within the human metagonomes. Tle2 and Tle3 are completely absent but Tle1 and especially Tle5 are highly abundant. Initially, the lack of Tle2 and Tle3 families from human metagenomes was surprising because they occur in genomes of bacteria which have been isolated from, or have some association with, humans. However, this discrepancy may be explained by the specific sites which have been sequenced for metagenome analysis. Some human-associated bacteria possessing these Tle families, such as

Masilla timonae, were isolated from areas of the bodies which are not represented in the available metagenome datasets (Lindquist *et al*., 2003).

It is interesting to note the general lack of Tle in the gut is largely explained by the paucity of the Tle5 family, where the stool niche makes up less than 8% of the Tle5 numbers it would be expected to based on proteobacterial gene count. In contrast, Tle5 is highly represented in the mouth. As this is a highly specific niche we assessed the divergence of Tle sequences found in this site. It is clear from BLAST analyses that most of the Tle5 in this niche are homologous to Tle from the *Haemophilus* and *Aggregatibacter* genera. Experimental evidence will be required to determine whether the frequency of Tle5 contributes to the prevalence of these bacteria within the mouth, or whether any other Tle family would serve equally as well.

As Tle are likely to be horizontally transferred, the differential abundance of Tle families could also be due to a founder effect. To test this hypothesis, the niche location of Tle within phylogenetic trees of Tle sequences was noted **(Figure S3a-e)**. Tle phylogeny was not well correlated with niche, except in cases of Tle from highly relates strains in the same niche. This is incongruent with a potential founder effect.

Furthermore, instances where members of the same species independently acquired members of the same Tle family, suggests a degree of selective pressure to obtain these genes. *P. aeruginosa* strains have several Tle genes, but closely related species such as *P. resinovorans* or *P. thermotolerans* have few or no Tle, which may be due the fact that *P. aeruginosa* live in several different niches. Other Tle profiles from genomes are in agreement with the metagenomic data. Tle2 is very frequent in both rhizosphere metagenomes and in soil-dwelling *P. fluorescens* species, while being much less common in other *Pseudomonas* species. Tle1 and Tle2 are enriched in aquatic metagenomes and also in *Vibrio* species.

While our data was normalised by proteobacterial gene count, some Tle1 sequences from metagenomes were homologous to lipases found in the phylum Bacteroidetes. This was initially surprising, as several T6SS core genes were not identified in this phylum. In fact, these lipases appear to be genuine effectors of another phage-derived secretion system, similar to the *Photorhabdus* virulence cassette/anti-feeding island (Yang *et al*., 2006; Penz *et al*., 2010; Zhang *et al*., 2012). BLASTP searches suggest that approximately half of the human-associated Tle1 sequences (roughly 10% of

overall Tle sequences from human metagenomes) are from Bacteroidetes, possibly a reflection of the prevalence of Bacteroidetes in the mouth. Tle1 abundance in human metagenomes is therefore somewhat overestimated in our data, and should be approximately half of the reported figure. However, BLASTP searches suggest very few of the Tle1 sequences from other metagenomes are from Bacteroidetes, and that metagenomic sequences from other Tle families were overwhelming proteobacterial in origin.

Divergence within Tle families

During the course of the BLAST analysis of Tle5 it became clear that there were several Tle proteins which were quite different from the majority of the Tle5 sequences. From genomic data a phylogenetic tree was constructed, clearly showing a divergence between two branches of Tle5 proteins, which are referred to here as Tle5a and Tle5b. This split between Tle5 sequences is congruent with a previous phylogenetic analysis of Tle5 (Russell *et al*., 2013). The lack of similarity between these branches is such that members of the Tle5a sub-cluster, which includes the previously characterised PldA/Tle5 from PAO1, are actually more homologous to phospholipases from eukaryotes than they are to their Tle5b counterparts. Indeed, the very few Tle5 from the marine environment all appear to be fungal lipases instead of genuine Tle proteins. Such false positives are infrequent elsewhere in other environments. Tle5a are much less common both in genome and metagenome datasets.

An examination of the genomic context of Tle5 from genome sequences shows that this divergence is also evident in the adjacently encoded immunity genes. Tle5b genes are encoded next to putative immunity genes with Sel1 domains, and these are relatively highly conserved, while the less common Tle5a genes are encoded next to more divergent immunity proteins which have little homology to the Sel1-containing proteins. Protein structure can be similar in the absence of sequence homology, but Tli5a and Tli5b were predicted to be different structures by modelling tool Phyre2 (Kelley and Sternberg, 2009). Therefore, it might be useful to consider Tle5 as composed of the two sub-groups Tle5a and Tle5b **(Figure 3.9)**.

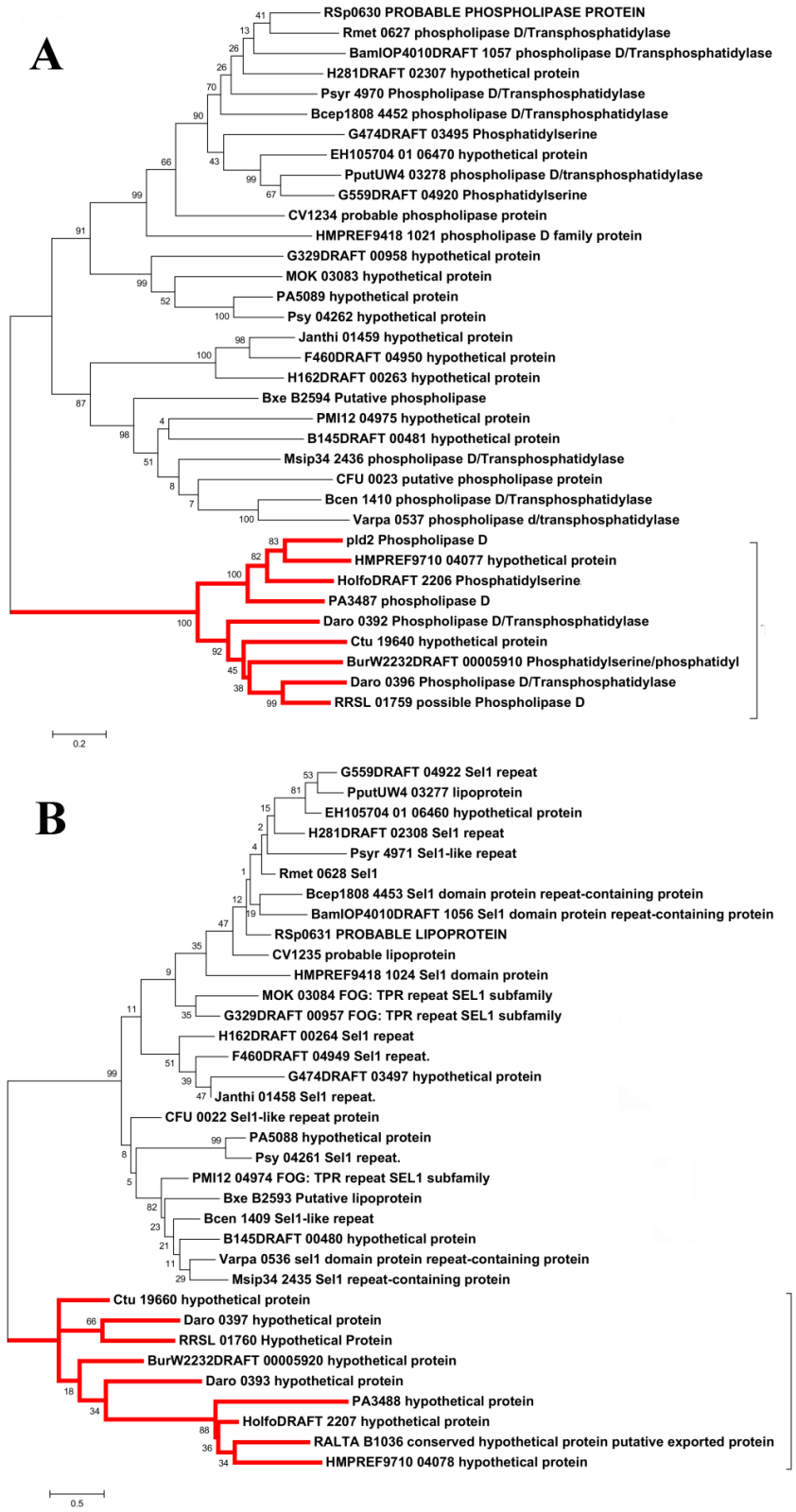


Figure 3.9. Maximum likelihood trees of Tle5 **(A)** and Tli5 **(B)** proteins from genome sequences were generated using WAG amino acid substitution model with 1000 bootstrap replicates. The Tle5a proteins and the corresponding immunity genes are coloured in red, and group separately to the Tle5b and Tli5b proteins. As Tle5b proteins are much more common than Tle5a proteins, only representatives of Tle5b were used in the tree.

Sequence analysis reveals conservation of motifs with residue divergence in all families of Tle in diverse ecological niches.

In the event that Tle niche specialisation were to occur, selective pressure might be expected to manifest itself in the amino acid sequences of Tles from different environments. Therefore, a comparison of Tle amino acid sequences from available metagenome datasets was undertaken, focusing on the niche-specific conservation of residues. The five Tle lipase families contain either the GxSxG motif or dual HxKxxxD motifs. Several additional conserved motifs exist in each specific family and, in line with our expectations, the areas of greatest conservation were observed in domains predicted to have lipase activity **(Figure 3.10).** However, inter-familial alignments of Tle proteins did not identify motifs that were common to all families, although Tle1-4 do share the GxSxG motif.

The lack of any conserved motifs between Tle families supports the hypothesis that these families have unique features. These unique features may be amenable to selection by specific niches, resulting in the differential Tle family abundance reported above. To investigate this, sequences of the same Tle family from different niches were compared, and analysed for the occurrence of changes in highly conserved amino acid residues. Several niche-specific conservations were identified, including residue 2596 in the Tle5 alignment, where the polar uncharged amino acid Asparagine (N) dominates in sequences from human metagenomes but the positively charged Histidine dominates in sequences from Arthropod metagenomes (**Figure 3.11**). More frequent were residues which were quite conserved in sequences from one environment, but not in sequences from another environment (**Figure 3.11**). Due to the partial nature of many metagenome sequences, in many positions there is limited coverage, which may mask other niche-specific residue changes.

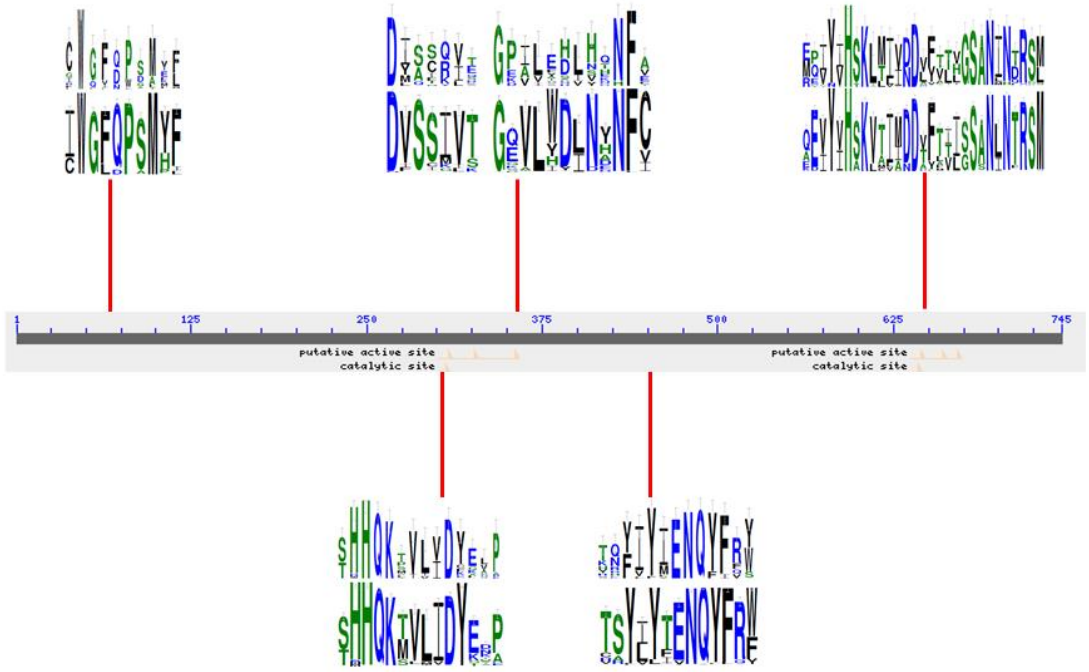


Figure 3.10. Weblogo showing consensus motifs on Tle5 protein sequences. Upper logos represent sequences from Arthropod metagenomes, while the lower logos represent sequences from Human metagenomes. Genomic sequences for each Tle were aligned using MAFFT E-INS-I using the BLOSSUM62 substitution matrix (Katoh and Standley, 2013) and metagenomic sequences were added to the genomic alignment using the add fragments option on the MAFFT server. Consensus sequences were derived using WebLogo 3 (Crooks *et al*., 2004), after generation of a non-redundant set of sequences using CD-HIT to eliminate sequences showing 100% identity.

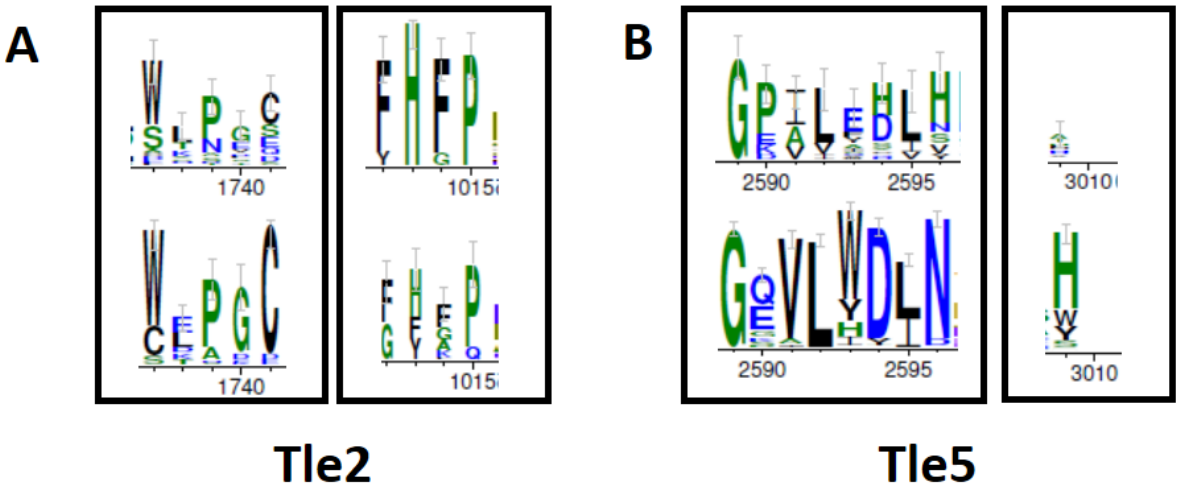


Figure 3.11. Consensus sequences for two regions in **(A)** Tle2 and **(B)** Tle5 show differences in conserved residues from sequences found in different environments. For Tle2 sequences from arthropod metagenomes (top) are compared with sequences form rhizosphere metagenomes (bottom). For Tle 5 sequences from arthropod metagenomes (top) are compared with sequences from human metagenomes.

Discussion

The T6SS are mainly associated with Proteobacteria, though have been reported to occur in other phyla (Bingle *et al*., 2008). Work in this chapter has expanded the number of phyla in which T6SS are known to occur. Of the analysed T6SS-related proteins, only VgrG and Rhs are found commonly outside Proteobacteria, indicating that they are involved in T6SS-independent processes. In these cases they are still encoded among prophage-related genes, underlining the evolutionary link between T6SS and phage. In addition, the presence of Tle1 sequences among Bacteroidetes encoding a secretion apparatus similar to the T6SS suggests several secretion systems had similar origins, and that there is a corresponding similarity between substrates they can export.

Despite progress in understating the roles of T6SS core proteins, many accessory proteins have no known function. In this chapter, bioinformatics analysis of effectors and accessory proteins possibly associated with effector export, suggests certain protein-protein interactions, and these may be important in the export process. Co-occurrence profiles presented in this chapter suggest that certain accessory components interact, and particularly the interactions between PAAR proteins and other accessory components is interesting as it suggest that these proteins may have a role in effector association with the T6SS. Inferences from gene co-occurrence profiles are speculative, but the levels of co-occurrence observed are quite compelling, particularly TagA co-occurrence with the rare PAAR subset being 100 times more frequent than TagA co-occurrence with the major PAAR subset. While not at the same magnitude, the co-occurrence between the predicted adapter proteins Pfam08786 and Pfam13503 and the major PAAR subset is also quite strong. Conversely, poor co-occurrence is also suggestive, and is the basis for our hypothesis that Pfam08786 and Pfam13503 proteins may have redundant or inhibitory functions.

While genomic evidence suggests that these proteins have key roles in the T6SS, this must be confirmed by experimental evidence. While many of these Pfam domains have not been characterised experimentally the roles of VasW from *V. cholerae* and IdsC from *P. mirabilis*, which both are members of the Pfam13503, though not identified as such in these papers, have been investigated (Gibbs *et al*., 2008; Miyata

et al., 2013). IdsC, like other genes in its *vgrG* island including *vgrG*, is necessary for identification, and subsequent merging, with similar bacteria. However, it is not involved directly in this recognition, and therefore seems necessary for the delivery of the effectors that are (Gibbs *et al*., 2008). VasW is encoded in a *vgrG* island with effector protein VasX. Deletion of VasW does not impair Hcp secretion but abrogates VasX secretion and therefore VasX-mediated killing of competitor bacteria. Another Pfam13503 is encoded in this genome and associated with the same T6SS, but clearly does not compensate for the loss of VasX, which may indicate that the association between adaptor and effector is specific to individual adapter/effector pairs. This may mean that these proteins are chaperones which localise effectors to the T6SS. VasW deletion has the same effect on Hcp secretion as VasX deletion in that it only abrogates it when VgrG-3 is also deleted (Miyata *et al*., 2013). This suggests that VasX, VasW or both these proteins play a structural role in the T6SS, and that these proteins are co-dependent for localisation to the T6SS apparatus. This is consistent with a chaperoning function for Pfam13503, though alternative hypotheses may be proposed.

While the analysis of core and accessory components is vital to understanding of the T6SS, it should not be overlooked that other possible T6SS-related genes do not belong to either category. The existence of an island (encoded by PA0095-PA0101 in *P. aeruginosa* PAO1) encoded across several bacteria genera is just one of several possible case studies which shows several proteins which are not considered to be accessory components but are conserved in islands which appear to be related to the T6SS. While it is not clear what relevance all these genes may have in T6SS process, at least one of these may be (PA0096) involved in secretion based on structural modelling which suggests it has a role in membrane piercing.

While questions about the mechanism of effector export remain, the distribution and ultimately the impact of these effectors on environments is unknown. Analysis of the distribution of a widespread effector superfamily is a novel way of examining levels of T6SS-mediated competition and killing within environments, and potentially more accurate than the method employed by previous studies which simply measure secretion system abundance (Persson *et al*., 2009).

Though the T6SS and the PAAR proteins are present at relatively similar levels within different environments Tle numbers vary significantly, being much more abundant in host-associated environments. Tle abundance possibly reflects the level of activity/competition within niches. One host-associated environment which is known to be highly competitive is the rhizosphere (Nannipieri *et al*., 2007) and indeed, Tle is much more frequent in this niche than in the bulk soil. Other factors may also contribute to a lower Tle abundance. The lack of Tle in aquatic environments might be due to the lack of activity within this niche or could possibly reflect that open water niches in aquatic environments are not conducive to a contact-dependent method of bacterial killing. Greater sampling of sediments or sites of high bacterial concentrations from these environments may result in greater Tle representation in aqueous environments, though in the limited amount of sediment-based aquatic metagenomes Tle genes are not more abundant.

As well as differences in aggregate Tle diversity between niches, there is differential abundance in members of each Tle family. This differential abundance, the presence of family and niche specific polymorphisms, and the presence of different Tle families within the same genomes suggests there is specialisation of the various Tle families and various environments are exerting selective pressure to favour certain families **(Figure 3.12).**

Even within families, there is considerable divergence, and this is especially evident in Tle5 sequences which may be split into Tle5a and Tle5b. Possibly this split, and divergence in other Tle families, represents evolutionary adaptions to different environments.

If evolution is favouring certain Tle families in certain niches and driving divergence within families, it follows that certain amino acids are being selected at certain residues. Several examples of different residues associated with niches were identified. Whether this variation was due to genetic drift in these species or some selective pressure driving adaption will remain unclear until further functional and bioinformatic studies are completed. Although the data in this chapter do not support a role for a founder effect, attempts at demonstrating the convergence of residues to particular niche-associated polymorphisms is currently restricted by limitations in the available metagenomic datasets, where individual species or genera dominate the

available genetic information from each niche. While convergent evolution would be the logical outcome of selective pressure at certain residue sites, more sequencing data, and possibly experimental data, are needed before this can be demonstrated. However, data in this chapter is congruent with the hypothesis that selective pressure within different environments may manifest in divergent Tle sequences.

The wealth of metagenomic data now available presents an opportunity to find ways to study ecosystems directly. Environmental factors can be diverse even within niches and as sampling levels increase, it will further inform the analysis and conclusions drawn by this study.

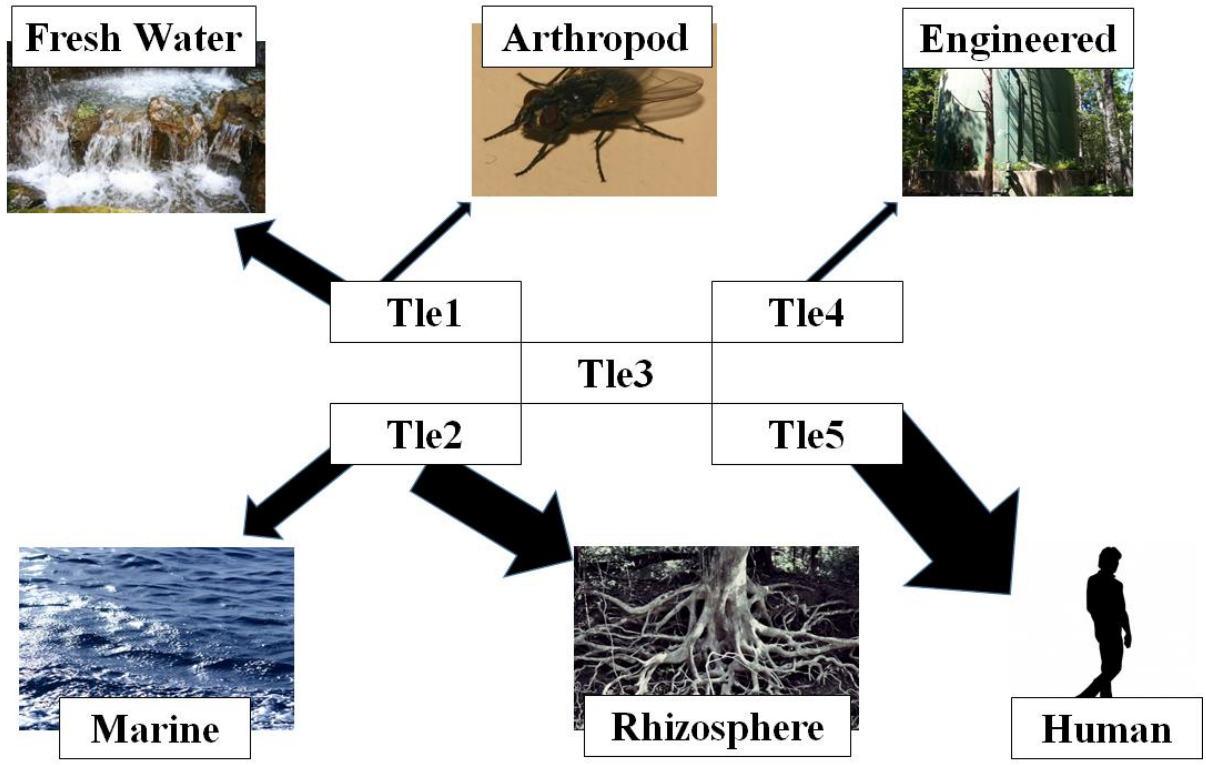


Figure 3.12. The distribution of the various Tle families varies with different environments, suggesting that there is niche specific selection. The degree of selection is reflected by the size of the arrows, with the strongest selection being for Tle5 in the human metagenomes and Tle2 in the rhizosphere metagenomes.

References

Aschtgen, M.-S., Bernard, C.S., Bentzmann, S.D., 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.

Barret, M., Egan, F., Fargier, E., Morrissey, J.P., and O'Gara, F. (2011) Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: novel clusters and putative effectors uncovered. *Microbiology* **157**: 1726–1739.

Bingle, L.E., Bailey, C.M., and Pallen, M.J. (2008) Type VI secretion: a beginner's guide. *Curr. Opin. Microbiol.* **11**: 3–8.

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**: 354.

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.

Browning, C., Shneider, M.M., Bowman, V.D., Schwarzer, D., and Leiman, P.G. (2012) Phage Pierces the Host Cell Membrane with the Iron-Loaded Spike. *Structure* **20**: 326–339.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004) WebLogo: A Sequence Logo Generator. *Genome Res.* **14**: 1188–1190.

Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**: 324–328.

Dong, T.G., Ho, B.T., Yoder-Himes, D.R., and Mekalanos, J.J. (2013) Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. *Proc. Natl. Acad. Sci.* **110**: 2623–2628.

Enos-Berlage, J.L., Guvener, Z.T., Keenan, C.E., and McCarter, L.L. (2005) Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*. *Mol. Microbiol.* **55**: 1160–1182.

Filloux, A. (2013) The rise of the Type VI secretion system. *F1000Prime Rep.* **5**:

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

Gibbs, K.A., Urbanowski, M.L., and Greenberg, E.P. (2008) Genetic Determinants of Self Identity and Social Recognition in Bacteria. *Science* **321**: 256–259.

Gueguen, E., Durand, E., Zhang, X.Y., d' Amalric, Q., Journet, L., and Cascales, E. (2013) Expression of a Y*ersinia pseudotuberculosis* Type VI Secretion System Is Responsive to Envelope Stresses through the OmpR Transcriptional Activator. *PLoS ONE* **8**: e66615.

Jackson, A.P., Thomas, G.H., Parkhill, J., and Thomson, N.R. (2009) Evolutionary diversification of an ancient gene family (rhs) through C-terminal displacement. *BMC Genomics* **10**: 584.

Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**: 772–780.

Kelley, L.A. and Sternberg, M.J.E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**: 363–371.

Koskiniemi, S., Lamoureux, J.G., Nikolakakis, K.C., Roodenbeke, C. t'Kint de, Kaplan, M.D., Low, D.A., and Hayes, C.S. (2013) Rhs proteins from diverse bacteria mediate intercellular competition. *Proc. Natl. Acad. Sci.* **110**: 7032–7037.

Lindquist, D., Murrill, D., Burran, W.P., Winans, G., Janda, J.M., and Probert, W. (2003) Characteristics of *Massilia timonae* and *Massilia timonae*-Like Isolates from Human Patients, with an Emended Description of the Species. *J. Clin. Microbiol.* **41**: 192–196.

Miyata, S.T., Unterweger, D., Rudko, S.P., and Pukatzki, S. (2013) Dual Expression Profile of Type VI Secretion System Immunity Genes Protects Pandemic *Vibrio cholerae*. *PLoS Pathog* **9**: e1003752.

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

Nannipieri, P., Ascher, J., Ceccherini, M.T., Landi, L., Pietramellara, G., Renella, G., and Valori, F. (2007) Microbial diversity and microbial activity in the rhizosphere. *Suelo (Argentina)* **25**: 89–97.

Penz, T., Horn, M., and Schmitz-Esser, S. (2010) The genome of the amoeba symbiont "*Candidatus* Amoebophilus asiaticus" encodes an afp-like prophage possibly used for protein secretion. *Virulence* **1**: 541–545.

Persson, O.P., Pinhassi, J., Riemann, L., Marklund, B.-I., Rhen, M., Normark, S., *et al*. (2009) High abundance of virulence gene homologues in marine bacteria. *Environ. Microbiol.* **11**: 1348–1357.

Pukatzki, S., Ma, A.T., 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.* **104**: 15508–15513.

Pukatzki, S., McAuley, S.B., and Miyata, S.T. (2009) The type VI secretion system: translocation of effectors and effector-domains. *Curr. Opin. Microbiol.* **12**: 11–17.

Russell, A.B., LeRoux, M., Hathazi, K., Agnello, D.M., Ishikawa, T., Wiggins, P.A., *et al*. (2013) Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature* **496**: 508–512.

Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014) Type VI secretion system effectors: poisons with a purpose. *Nat. Rev. Microbiol.* **12**: 137–148.

Schwarz, S., Hood, R.D., and Mougous, J.D. (2010) What is type VI secretion doing in all those bugs? *Trends Microbiol.* **18**: 531–537.

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.

Shneider, M.M., Buth, S.A., Ho, B.T., Basler, M., Mekalanos, J.J., and Leiman, P.G. (2013) PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500**: 350–353.

Silverman, J.M., Agnello, D.M., Zheng, H., Andrews, B.T., Li, M., Catalano, C.E., *et al*. (2013) Haemolysin Coregulated Protein Is an Exported Receptor and Chaperone of Type VI Secretion Substrates. *Mol. Cell* **51**: 584–593.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **28**: 2731–2739.

Weber, B., Hasic, M., Chen, C., Wai, S.N., and Milton, D.L. (2009) Type VI secretion modulates quorum sensing and stress response in *Vibrio anguillarum*. *Environ. Microbiol.* **11**: 3018–3028.

Whitney, J.C., Beck, C.M., Goo, Y.A., Russell, A.B., Harding, B.N., De Leon, J.A., *et al*. (2014) Genetically distinct pathways guide effector export through the type VI secretion system. *Mol. Microbiol.* **92**: 529–542.

Wu, L., Yong, S.L., Fan, C., Ni, Y., Yoo, S., Zhang, T., *et al*. (2008) Identification of a New Co-factor, MOG1, Required for the Full Function of Cardiac Sodium Channel Nav1.5. *J. Biol. Chem.* **283**: 6968–6978.

Yang, G., Dowling, A.J., Gerike, U., ffrench-Constant, R.H., and Waterfield, N.R. (2006) *Photorhabdus* Virulence Cassettes Confer Injectable Insecticidal Activity against the Wax Moth. *J. Bacteriol.* **188**: 2254–2261.

Zhang, D., Souza, R.F. de, Anantharaman, V., Iyer, L.M., and Aravind, L. (2012) Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **7**: 1–76.

Zhang, W., Wang, Y., Song, Y., Wang, T., Xu, S., Peng, Z., *et al*. (2013) A type VI secretion system regulated by OmpR in *Yersinia pseudotuberculosis* functions to maintain intracellular pH homeostasis. *Environ. Microbiol.* **15**: 557–569.

Chapter 4

Title: The role of SPI-1 T3SS in *P. fluorescens* F113 rhizosphere competence

Publications associated with this chapter: Characterization of the SPI-1 and Rsp type three secretion systems in *Pseudomonas fluorescens* F113.

Authors: M. Barret, F. Egan, J. Moynihan, JP. Morrissey, O. Lesouhaitier and F. O'Gara.

Abstract

The T3SS has been shown to be involved in virulence towards both plants and animals, but occurrence of this system encoded in the genomes of strains which are not known pathogens has led to new roles being discovered for T3SS. These include roles in mutualistic interactions and roles involving single-celled organisms. The biocontrol strain *P. fluorescens* F113 has to compete with a range of organisms in the rhizosphere, and has two separate T3SS encoded in its genome – one belonging to the Hrp1 family and one belonging to the SPI-1 family. The Hrp1-T3SS of a related *P. fluorescens* strain has been shown to manipulate the plant defence response, and an equivalent role for this system in the life cycle of *P. fluorescens* F113 seems probable. In contrast, the role of the SPI-1 T3SS in this species is unknown. Results in this chapter indicate that this system is expressed in response to contact with amoeba and that this T3SS is important for allowing F113 to avoid amoeboid grazing. This may represent an important factor in the rhizocompetence of F113 and other strains of *P. fluorescens*. In addition, two new putative effectors of this system have been identified.

Introduction

The non-flagellar type III secretion system (T3SS) is one of six specialised secretion systems identified in gram-negative bacteria. It mediates host microbe interactions by translocating molecules, termed effectors, from the bacterial cell into a eukaryotic target cell. These interactions can range from mutualism to parasitism (Preston, 2007). The T3SS consists of three main parts; a structural apparatus embedded in the bacterial membranes, cytoplasmic proteins controlling export, and an extracellular needle or pilus for effector delivery. In addition, T3SS effectors (T3Es) often associate with the structural apparatus via chaperones and are released from the chaperones before export due to the activity of a T3SS-specific ATPase.

Phylogenetic analyses of the proteins comprising the structural apparatus have split the T3SS into seven different families (Chlamydial, Ysc, Rhizobial, Hrp1, Hrp2, SPI-1 and SPI-2) (Abby and Rocha, 2012). Though many of the structural and cytoplasmic proteins of the T3SS are highly conserved, these families show differences in their host range, extracellular structure and translocator suite.

Sequencing of the genome of the plant growth promoting rhizobacterium (PGPR) *P. fluorescens* F113 revealed the presence of two T3SSs in this organism, belonging to the Hrp1 and SPI-1 families.

The Hrp1 system was initially described in phytopathogens such as *P. syringe* DC3000, but later was found to be common in range of non-pathogenic rhizobacteria (Alfano and Collmer, 1997; Rezzonico *et al*., 2004). In these strains it is also referred to as the Rsp (Rhizosphere secreted protein) T3SS (Preston *et al*., 2001). The Rsp T3SS was found to be expressed during rhizosphere colonisation by *P. fluorescens*. (Rainey, 1999). More recently, an Rsp T3SS, encoded in the genome of *P. brassicacearum* Q8r1-96 was found to be involved in modulating the plant immune response (Mavrodi *et al*., 2011). As the Rsp T3SS from *P. fluorescens* F113 and *P. brassicacearum* Q8r1-96 are highly similar and the effector molecules responsible for this phenotype are also found in F113, it is likely that this system plays a similar role in *P. fluorescens* F113. However, as described below, recent developments have seen that one T3SS can target a diverse array or eukaryotes and the Hrp1 T3SS may not restricted to plant-bacterial interactions as another strain of *P. fluorescens* was show to kill amoeba in a Hrp T3SS-dependent manner (Sperandio *et al*., 2012).

In contrast to the finding that an Rsp T3SS is found in the F113 genome, the presence of the SPI-1 T3SS in a strain of *P. fluorescens* was surprising. This T3SS is best known for its role in allowing enteric pathogens such as *Salmonella* and *Shigella* to invade mammalian epithelial cells and is known to be involved in the virulence of other mammalian pathogens, such as *C. violaeceum* (Galán, 1996; Miki *et al*., 2010). Recently however, other functions for this T3SS have been described. It is needed for the replication of the symbiont *Sodalis glossinidius* within insect cells and for the survival of *Pantoea stewartii* in the flea beetle (Dale and Moran, 2006; Correa *et al*., 2012). In *Salmonella*, it is also involved in suppression of the plant immune system during plant colonisation and may a have a further role in *Salmonella*-amoeba interactions (Tezcan-Merdol *et al*., 2004; Schikora *et al*., 2011). In addition to these new discoveries, genome sequencing is revealing that this T3SS is present in many bacteria which are not mammalian pathogens, such as the plant pathogens *Erwinia* and *Xanothomonas* (Triplett *et al*., 2006; Alavi *et al*., 2008).

In addition to F113, *P. fluorescens* genomes which have been recently sequenced contain the SPI-1 T3SS. The apparent remnants of the system are also present in SBW25, suggesting this was present in an ancestral *P. fluorescens* genome and in some cases has been lost. Though clinical strains of *P. fluorescens* exist, the pathogenic potential of *P. fluorescens* is generally considered to be quite low.

The presence of the SPI-1 T3SS in commensal rhizobacteria raises several questions. Is this system expressed at all and what is its ecological significance in the life cycle of *P. fluorescens* F113? In this work, the transcriptional activator of the SPI-I T3SS in *P. fluorescens* F113 was identified and genes upregulated its expression were detected, including two putative new effectors. The potential roles of this T3SS in interactions with other soil organisms was investigated. The results demonstrate that the SPI-1 T3SS of F113 is induced in contact with amoeba and is involved in resistance to amoeboid predation.

Methods

Bacterial strains and media

Bacterial strains and plasmids are listed in **Table 4.1**. *P. fluorescens* F113 was routinely maintained at 30 °C on LB agar, or LB broth or in M9 minimal medium with sodium citrate (0.3%, w/v) as a sole carbon source, with shaking (180 rpm). *Escherichia coli* strains were grown at 37°C shaking (180 rpm) in LB medium. When appropriate, antibiotics were supplemented at the following concentrations : for *E. coli* gentamycin 10 $\mu g.ml^{-1}$, ampicillin 50 $\mu g.ml^{-1}$ and kanamycin 50 $\mu g.ml^{-1}$; for *P. fluorescens* gentamycin 25 $\mu g.ml^{-1}$, tetracycline 20 $\mu g.ml^{-1}$, *piperacillin* 25 $\mu g.ml^{-1}$, trimethoprim 1000 $\mu g.ml^{-1}$ and gentamycin 10 $\mu g.ml^{-1}$ with tetracycline 10 $\mu g.ml^{-1}$.

DNA techniques

Plasmid and genomic DNA isolations, electrophoresis and restriction enzyme digestions were performed using standard protocols (Green, 2012). Polymerase chain reactions were carried out using GoTaq® Green Master Mix or Pfu DNA polymerase using *manufacter's recommended conditions* (Promega, Madison, USA). Plasmids were mobilised into *P. fluorescens* F113 by triparental mating using the helper plasmid pRK2013 (Figurski and Helinski, 1979). Nucleotide sequencing was carried out by *GATC* Biotech *(Konstanz, Germany).*

Table 4.1. Strains and plasmids used in this study

Strains/Plasmids	Characteristics	References
E. coli DH5α	Lab strain	Green, 2012
E. coli HB101	Lab strain	Green, 2012
P. fluorescens F113	Sugarbeet , Ireland	Shanahan *et al*., 1992
P. fluorescens FG9	F113 gacAΩ::mini-Tn5-lac	Delany, 1999
P. fluorescens F113 Δ*rscU*	In-frame deletion of *rscU* (PSF113_5599)	This study
P. fluorescens F113 Δ*spaS*	In-frame deletion of *spaS* (PSF113_1788)	This study
pRK2013	Helper plasmid, Mob+, Tra+, KmR	Figurski and Helinski, 1979
pME6032	IPTG-inducible expression vector	Mao *et al*., 2002
pME6032ExoUFLAG	pME6032 carrying PSF113_5498 with C-terminal FLAG tag	This study
pBBR1MCS-4	Broad host range vector, AmpR	Kovach *et al*., 1995
pBBR1MCS-5	Broad host range vector, GmR	Kovach *et al*., 1995
pBBR1MCS-*hilA*	pBBR1MCS5 carrying the complete *hilA* coding region (PSF113_1794)	This study
pBBR1MCS-*rscU*	pBBR1MCS5 carrying the complete *rscU* coding region (PSF113_5599)	This study
pBBR1MCS-*spaS*	pBBR1MCS5 carrying the complete *spaS* coding region (PSF113_1788)	This study
pPROBE-GT	Promoterless *gfp* vector, GmR	Miller *et al.*, 2000
pPROBE-*hilA*	pPROBE-GT with a 160 bp *hilA* promoter fragment	This study
pMS402	Promoterless *luxCDABE* vector, TpR KmR	Duan *et al*., 2003
pMS402-*iglA3*	pMS402 with the *iglA3* promoter fragment	This study
pMS402-*hilA*	pMS402 with the *hilA* promoter fragment	This study

Construction of promoter gene fusions and promoter fusion analysis

The promoter region *hilA* was predicted using BPROM (Softberry Inc., NY, USA). DNA sequence including the predicted *hilA* promoter and transcriptional start site were amplified from F113 genomic DNA. The *hilA* promoter was cloned into the *Bam*HI-*Kpn*I and *Bam*HI-*Eco*RI digested pPROBE-GT plasmid (Miller *et al*., 2000), which contains a promoterless *gfp* gene. In addition, the putative *hilA* promoter region was PCR amplified and cloned into the *Bam*HI and *Xho*I digested plasmid pMS402 (Duan *et al*., 2003) which contains the promoterless *luxCDABE* operon.

Cells from overnight cultures were inoculated to a final optical density at 600 nm of 0.05 in LB medium. Fluorescence from promoter *gfp* reporter fusion was measured on a MWG Sirius HT reader at an excitation wavelength of 490 nm and an emission of 510 nm. Luminescence from promoter-*lux* reporter fusions was measured on a Tecan GENios microplate reader. Relative luciferase counts were obtained from a Xenogen

IVIS 100 Imaging System using strains carrying the empty vector pMS402 for background levels corrections. All the data presented in this manuscript are mean of three independent biological replicates.

Expressions of *gfp* reporter fusions were monitored from aliquots of 200 µL harvested directly from the well. Expression of strains bearing *lux* reporter fusions in contact with amoeba were assessed as follows: bacterial strains from overnight cultures were washed twice in PAS and spot-inoculated on PYG agar (15 $g.l^{-1}$) at an optical density at 600 nm of 0.1. Bacterial drops were allowed to dry for 20 min before the addition of 1×10^4 amoeba at contact, 0.5 cm or 1 cm from the edge of the bacterial spot. After 48 hours, luminescence signals were registered using the IVIS 100 Imaging System.

Gene expression analysis

Total RNA was extracted from *P. fluorescens* F113 strains grown at optical density at 600 nm of 1.2 in LB medium using Trizol® (Invitrogen). Contaminant genomic DNA was removed by RQ1 DNase (Promega) and RNA was purified with the Qiagen RNeasy Mini kit (Quiagen) according to the manufacturer's instructions. RNA concentration and integrity was determined with a Nanodrop ® and denaturing agarose gels, respectively. All RNA samples were stored at −80°C.

The genome sequence of *P. fluorescens* F113 (Redondo-Nieto *et al*., 2012) was used to design a customised high-density oligonucleotide array (Roche NimbleGen, Madison, WI, USA). This microarray contains 76811 individual probes of 60 bp representing 5934 genes including CDS, non-coding RNA, rRNA, and tRNA. With the exception of 9 genes, each individual gene is represented by 6 different probes, which are replicate twice on the array.

The cDNA synthesis, hybridisation, scanning and microarray data analysis were performed by Roche NimbleGen (Reykjavík, Iceland). Moderated *t*-test with the Benjamini-Hochberg correction for False Discovery Rate (FDR) was used to identify genes showing differential expression patterns ($P < 0.05$). Microarray data analysis has only been performed on one independent biological replicate for each F113 strain. Therefore only selected differentially expressed genes that have been validated by qRT-PCR are discussed in the text.

cDNA synthesis was performed with the Superscript III Reverse Transcriptase kit (Invitrogen) using 0.5 µg μl^{-1} of random primers. Reverse Transcription (RT) PCRs were carried out at 25 cycles with the housekeeping gene, *proC,* as a control. For qRT-PCR analysis, cDNA synthesised from 3 independent biological RNA samples were amplified using the Qiagen QuantiFast SYBR Green PCR kit (Quiagen). The expression stability of four housekeeping genes (*leuC*, *proC, purF* and *rbfA*) was calculated using geNorm (Vandesompele *et al*., 2002). Based on M score, the gene *rbfA* was chosen for data normalisation.

Generation of T3SS mutants, complemented strains, overexpression constructs

The following procedure was employed to generate unmarked in-frame deletion of *rscU* and *spaS,* in the F113 wild type. Upstream and downstream DNA fragments flanking the gene of interest were PCR-amplified (approximately 500 bp each) and fused together by cross-over PCR using a complementary tag (**Table 4.2**). The resulting PCR product was cloned in the suicide plasmid pEX18Gc (Hoang *et al*., 1998) digested with *Eco*RI and *Xba*I. The plasmids were introduced in F113 by triparental mating and transconjugants were selected on LB gentamycin 25 $\mu g.ml^{-1}$. Single cross-over insertion in the chromosome was verified by PCR. Then, using LB supplemented with sucrose (10% w/v), bacterial strains were cultured overnight, serially diluted and plated. The second-cross-over event was verified by PCR amplification from the flanking regions and sequencing. The Δ*rscU* F113 and Δ*spaS* F113 mutants are deleted for 919 (from codons 33 to 339) and 697 (from codons 87 to 320) bp, respectively.

To generate complementation and overexpression constructs, DNA fragments containing the complete *rspL, rscU, spaS* and *hilA* coding regions as well as their predicted endogenous ribosome binding sites were amplified using primers listed in **Table 4.2**. Each DNA fragment was ligated into the *Eco*RI-*Xba*I digested pBBR1MCS-5 plasmid (Kovach *et al*., 1995) and introduced in the different F113 strains (**Table 4.1**).

Putative effector PSF113_5498 (ExoU) was amplified and FLAG-tagged at the C-terminal end of the sequence using the primers in **Table 4.2** and was ligated in to the

pME-6032 vector by after digestion with *Kpn*1 and *Xho*1 (Mao *et al*., 2002). These constructs were introduced into different F113 strains by triparental mating.

Table 4.2. Primers used in this study.

Name	Sequence (5' to 3')	Purpose
hilA_F	GCGAATTCAGGCTCAATCGCGCTTCA	*hilA* overexpression
hilA_R	GGTCTAGATCATGCCAGGCGCCTCT	*hilA* overexpression
hilAprom_R	GGATCCCTGTGGGTAACAGGCTCAGG	*hilA* promoter fusion pms402
hilAprom_F	CTCGAGCTTCCCAGTGTGCTTAGGTTT	*hilA* promoter fusion pms402
hilAPPeco	GAATTCCTGTGGGTAACAGGCTCAGG	*hilA* promoter fusion Pprobe
hilAPPbam	GGATCCCTTCCCAGTGTGCTTAGGTTT	*hilA* promoter fusion Pprobe
spaSudelFeco	GAATTCATGAACGCTTTCTCGGTGTC	*spaS* upstream fragment
spaSudelR	AGCAGGCCCAGCAACAGGT CGGCAGCAACACCTTGAG	*spaS* upstream fragment
spaSddelF	ACCTGTTGCTGGGCCTGCT GCTACAGCTTCCTCCAGGTG	*spaS* downstream fragment
spaSddelRxba	TCTAGAGAAAGAACTTCTCGGCATCG	*spaS* downstream fragment
spaSflankU	AAGACCCTGGTCATCAGTGC	validation of *spaS* deletion
spaSflankD	TGGCCGGTGTACAACATTGG	validation of *spaS* deletion
spaScomp2Fkpn	GGTACCAAGGGCAGACGTTCAAGG	*spaS* complementation
spaScompRcla	ATCGATTCACGCCTGCTCAACCTG	*spaS* complementation
hrcUudelFeco	GAATTCGATGACGATCTTGCTGCTG	*hrcU* upstream fragment

hrcUudelR	AGCAGGCCCAGCAACAGGT TGCTCAAGTCCTGGCTTTGG	*hrcU* upstream fragment
hrcUddelF	ACCTGTTGCTGGGCCTGCT CCATATTTATCGGGTGTTGC	*hrcU* downstream fragment
hrcUddelRxba	TCTAGACTCATGAAGGAGTCCGGCGT	*hrcU* downstream fragment
hrcflankU	ATTTACCCCGTCGCTTACCT	validation of *hrcU* deletion
hrcflankD	TCCATCAACTGCTTCTGACG	validation of *hrcU* deletion
hrcUcompFkpn	GGTACCCCTGCTTCACCTGTTGCTG	*hrcU* complementation
hrcUcompRcla	ATCGATCGACCTCAACGCTGGTTC	*hrcU* complementation
exoUFLAGF	GGTACCAGGTACAACGAACGAACCA	inducible expression of *exoU*-FLAG
exoUFLAGR	CTCGAGTTACTTATCGTCGTCATCCTTGTAATCTTTGAACTGCATGT	inducible expression of *exoU*-FLAG
rbfA_280F	TTGCACTTCCACTACGACGA	qRT-PCR validation
rbfA_396R	CTTGGTGTCTTCGGGTTCAG	qRT-PCR validation
1783_305F	GGAAAACAATCGCTGTTCCA	qRT-PCR validation
1783_401R	TCGATGACAGCAACAACAGG	qRT-PCR validation
hilA_F380	TTCGCAATCTGGATGAGATG	qRT-PCR validation
hilA_R561	ACGTCCGCTCAAGCAATAAT	qRT-PCR validation
prgK_298F	GAAAAGGCCCGCTTGTATTC	qRT-PCR validation
prgK_405R	GGCATCCAGGTCGTAACTGA	qRT-PCR validation
1802_270F	GCAGGTTTTGGTTCGTTTGA	qRT-PCR validation
1802_394R	ACACATCGCGCATAAAGCTC	qRT-PCR validation
4041_720F	GCGTTCCATCACCTTCGTAA	qRT-PCR validation
4041_829F	ACGTGTAGAACGCCATTTGC	qRT-PCR validation

Protein Extraction and Western Blot

Strains carrying the pME6032 vector containing FLAG-tagged *exoU* were grown overnight in M9 at 23 °C and transferred to 30 °C and supplemented with 1 mM IPTG upon reaching OD 0.8. After 3.5 hours incubation, cells were pelleted by centrifugation (20000 *g* for 10 minutes). Supernatant was incubated in 10% TCA overnight and proteins were precipitated with acetone washing. Proteins were separated using a 12% SDS 1D gel, transferred to a nitrocellulose membrane and detected by sequential incubation with anti-FLAG (Sigma) and Goat anti-Rabbit antibodies (DakoCytomation), and binding was detected using a chemi-luminescence kit (Thermo-scientific).

Caenorhabditis elegans *synchronisation and virulence assays*

The bacterivorous nematode *Caenorhabditis elegans* wild-type Bristol strain N2 was obtained from the *Caenorhabditis* Genetics Center (Minneapolis, MN, USA). *C. elegans* were maintained under standard culturing conditions at 22°C on nematode growth medium (NGM: 3 $mg.ml^{-1}$ NaCl, 2.5 $mg.ml^{-1}$ peptone, 17 $mg.ml^{-1}$ agar, 5 $\mu g.ml^{-1}$ cholesterol, 1 mM $CaCl_2$, 1 mM $MgSO_4$, 25 mM KH_2PO_4, H_2O to 1 litre) agar plates with *E. coli* OP50 as a food source (Sulston and Hodgkin, 1988). Synchronous cultures of worms were generated after the adult population was exposed to a sodium hypochlorite/sodium hydroxide solution. The resulting eggs were incubated at 22 °C on an *E. coli* OP50 lawn until the worms reached the L4 (48 hours) life stage (confirmed by light microscopy).

Bacterial lawns used for *C. elegans* survival assays were prepared by spreading 50 µl of *P. fluorescens* strains on 35 mm NGM conditioned Petri dishes supplemented with 0.05 mg ml^{-1} 5-fluoro-2'-deoxyuridine. This nucleotide analogue blocks the development of the *C. elegans* next generation via the inhibition of DNA synthesis, thus preventing offspring from the experimental animals. The plates were incubated overnight at 30 °C and then placed at room temperature for 4 h. Fifteen to twenty L4 synchronised worms were harvested with M9 solution (3 $mg.ml^{-1}$ KH_2PO_4, 6 $mg.ml^{-1}$ $NaHPO_4$, 5 $mg.ml^{-1}$ NaCl, 1 mM $MgSO_4$, H_2O to 1 litre), placed on the 35 mm assay Petri dishes and incubated at 22 °C. Worm survival was scored at 1 hour, 24 hours and each subsequent day, using an Axiovert S100 optical microscope (Zeiss, Oberkochen,

Germany) equipped with a Nikon digital Camera DXM 1200F (Nikon Instruments, Melville, NY, USA). The worms were considered dead when they remained static without grinder movements for 20 s. The results are expressed as the percentage of living worms. The results are the average of three independent biological replicates.

Bacterial-amoeba interactions

Acanthamoeba polyphaga was a kind gift from Dr M. Sanchez-Contreras (Universidad Autónoma de Madrid, Spain). Amoeba cultures were routinely maintained at 25 C in 75 cm^2 tissue culture flasks filled with 20 ml of PYG medium (proteose peptone 20 g. l^{-1}, yeast extract 5 g. l^{-1}, glucose 10 g. l^{-1}, in Page's amoeba saline solution). *A. polyphaga* was sub-cultured every 5 days by tapping the flask vigorously and resuspended at a 1:10 dilution in fresh media.

For coculture experiments, cells were harvested by centrifugation (300 *g*, 5 min), washed once in Page's amoeba saline (PAS) solution and resuspended in PAS. Cells were stained with trypan blue and counted with a hemocytometer. *A. polyphaga* were inoculated at a concentration of 2×10^5 cells.ml^{-1} in a 6-well-plate. Amoebae were allowed to adhere for 120 min before the addition of bacterial cells. F113 strains from overnight cultures were washed twice in PAS and inoculated at a multiplicity of infection of 100:1. Grazing experiments were performed with a single bacterial strain in each well. For selective feeding assay, bacterial strains were co-inoculated at a ratio of 1:1 (multiplicity of infection of 100:1 for each strain). In order to distinguish strains in mixed treatments, the empty vectors pBBR1MCS4 and pBBR1MCS5, which carry different selection markers, were introduced in the bacterial strains by triparental mating. As a control, bacteria were also inoculated into PAS. At different time interval, cells were scraped from the well and serial dilution and plating were performed on selective LB medium supplemented with gentamycin 25 µg.ml^{-1} or piperacillin 25 µg.ml^{-1}. The results are the average of four independent biological replicates. Statistical analysis was performed using two-tailed student's *t*-test for unpaired data.

Infection of Galleria mellonella

Galleria mellonella larvae were obtained from Livefood UK (Axbridge, England) and stored in the dark until use. F113 strains from overnight cultures were washed twice PBS and resuspended in PBS at an OD_{600nm} of 1. Ten µL of bacterial suspension (approximately 5×10^6 bacterial cells) were injected into the haemocoel using a sterile Hamilton syringe and a 30 gauge disposable needle. Insects were placed in a 9.0 cm Petri dishes lined with 8.5 cm Whatman paper and incubated at 30°C in the dark. As a control 10 insects were injected with sterile PBS. Insects were individually examined for the production of pigmentation and the time of death was recorded until five days post-inoculation. Insects were considered dead when they no longer responded to touch.

Results

Identification of HilA regulon.

Previous studies of the T3SS in other strains have shown that the transcriptional activator HilA induces expression of genes of the SPI-1. To confirm that this finding also applicable to F113 and to identify putative effectors, F113 strains expressing *hilA* on the broad host range plasmid pBBR1MCS5, and a strain carrying the empty vector were compared via transcriptome. According to the microarray data, 51 genes were upregulated by *hilA* expression. Of these, 24 encode proteins involved in the assembly or functioning of the SPI-I system, while two others (PSF113_1802 and PSF113_4041) contain a T3 secretion signal and are likely new SPI-I effectors. As only one sample was analysed by the array, the expression of the three SPI-I core genes and both putative effectors was analysed by qRT-PCR and their upregulation by *hilA* was confirmed **(Table 4.3)**.

These novel putative T3SS effectors have little homology to known proteins. Modelling with Phyre2 suggests that PSF113_1802 may have a role in nucleotide binding as it has moderate homology with histones and viral nucleotide binding proteins. PSF113_4041 is modelled with 48% confidence to alpha-14 giardin, an annexin-like protein which remodels the cytoskeleton (Pathuri *et al*., 2009). While bioinformatics hypotheses need to be confirmed with experimental evidence, nucleotide binding and cytoskeletal remodelling are two common activities of T3SS effectors (Dean, 2011).

The SPI-1 T3SS is functional

Though not upregulated by expression of *hilA*, another putative T3SS effector is encoded in the genome of F113, with 38% identity to the effector ExoU of *P. aeruginosa*. In *P. aeruginosa* this effector is secreted by a T3SS belonging to the Ysc family, which is generally considered to be animal-targeting T3SS family. To test whether this ExoU homologue is a genuine substrate of the T3SS the vector pBBR1MCS carrying *hilA*, the activator of the SPI-1 system, and the inducible vector PME6032 carrying a FLAG-tagged version of the putative effector ExoU, were introduced into the *P. fluorescens* F113 wild type and *spaS* mutant backgrounds. The SpaS protein is a homologue of the YscU switch protein which is

necessary for T3SS activity. This allowed the expression of *exoU* in strains where T3SS expression was constitutive but secretion was either functional (wild type) or defective (Δ*spaS*).

The secreted protein profiles of these strains show that ExoU is secreted in a SPI-1 dependent manner, as secretion only occurs in the wild type, but not the mutant background (**Figure 4.1**). The use of the pBBR1MCS vector to induce T3SS expression meant that complementation of the secretion defect of the Δ *spaS* mutant was not attempted. However, as outlined below, another phenotype of this strain was complemented by expression of *spaS in trans* so it is very unlikely the secretion defect was due to a secondary mutation.

Table 4.3. Gene regulated by T3SS transcriptional activator HilA according to microarray and qRT-PCR data. Putative effectors are highlighted in bold.

Locus	**Gene name**	**Function**	**HilA/MCS5 (log2)**	
			array	**qRT-PCR**
PSF113_1783		Type III secretion protein	6.5	6.4 ± 0.3
PSF113_1794	*hilA*	Type III secretion transcriptional activator HilA	6.1	10.6 ± 0.4
PSF113_1799	*prgK*	Type III secretion bridge	5.9	8.2 ± 0.3
PSF113_1802		Hypothetical protein	7.0	7.1 ± 0.2
PSF113_4041		Hypothetical protein	5.3	6.3 ± 0.2

The SPI-1 T3SS in induced by contact with amoeba

To identify the inducing conditions for SPI-1 activity, the predicted *hilA* promoter region was cloned into the pProbe and pMS402 vectors, which allow detection of promoter activity via GFP and luciferase production, respectively. In LB broth *hilA* is expressed at a basal level, and was not affected by the addition of quorum sensing molecules PQS or HHQ (**Figure 4.2**).

However, a promoter fusion assay on solid media revealed that *P. fluorescens* F113 responds to the presence of *Acanthamoeba polyphaga* by inducing expression of the *hilA* gene. This was initially done using the pProbe vector and subsequently confirmed using the Pms402 vector (**Figure 4.2**). Amoeba inoculated 1 cm away from the bacteria did not induce *hilA* expression, but did at a distance of 0.5 cm from or in contact with the bacteria.

The Gac regulatory system has been shown to be involved in the survival of *P. fluorescens* CHAO when co-cultured with amoeba (Jousset *et al*., 2009). In addition, the GacA orthologue of *Salmonella enterica*, SirA, positively regulates *hilA* expression (Ahmer *et al*., 1999). Therefore *hilA* promoter activity was measured in a previously obtained *gacA* transposon mutant (Delany, 1999). Expression of *hilA* is reduced in this background compared to the wild type.

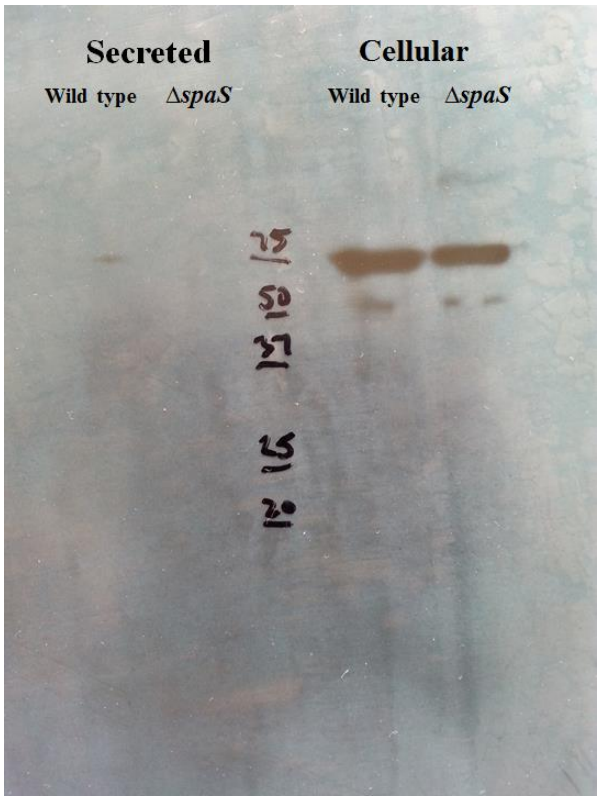


Figure 4.1. Anti-ExoU immuno-blot of **s**ecreted and cellular proteins fractions from wild type and Δ*spaS* strains constitutively expressing *hilA*. 2500ng of protein were loaded in the secreted fractions and 7500ng of protein were loaded in the cellular fractions. ExoU is not detected in the Δ*spaS* secretome but is present in the wild type secretome. As the cellular fractions of both strains contain ExoU this is not due a lack of ExoU production in the Δ*spaS* strain but because ExoU is secreted into the culture media in a SPI-1 T3SS-dependent manner.

Contribution of T3SS to interactions with organisms in the rhizosphere.
To assess the potential role(s) of SPI-1 and Hrp1 T3SSs in interactions with organisms present in the rhizosphere, in-frame deletions of the *spaS* and *hrcU* genes were generated. Mutations of *spaS* or *hrcU* homologues abolish the secretion of needle and translocator subunits, resulting in non-functional T3SSs (Izoré *et al*., 2011).

When *P. fluorescens* was co-cultivated with *A. polyphaga* the *P. fluorescens* F113 were consumed by the amoeba, with a bacterial survival rate of about 60% and 30% after one and two days co-incubation, respectively. Neither of the mutants in the T3SSs showed any difference in survival compared to the wild type, and nor did a strain where *gacA* was inactivated by a transposon. The latter result was surprising as the involvement of the Gac regulatory system in *P. protegens* CHAO in avoidance of amoeboid grazing has been demonstrated (Jousset *et al*., 2009). However, in that study that *gacS* mutants of *P. protegens* CHAO are preferentially grazed by amoeba when co-cultivated with wild type cells. Therefore, the fitness of the T3SS mutants was tested in a selective feeding assay where mutant and wild type cells were mixed in a 1:1 ratio and co-incubated with amoeba. The mutant and wild type strains contained empty pBBR1MCS vectors, and these vectors were also used for complementation. Competing strains carried vectors with different antibiotic resistance markers. This allowed for numeration of the strains after the interaction with amoeba.

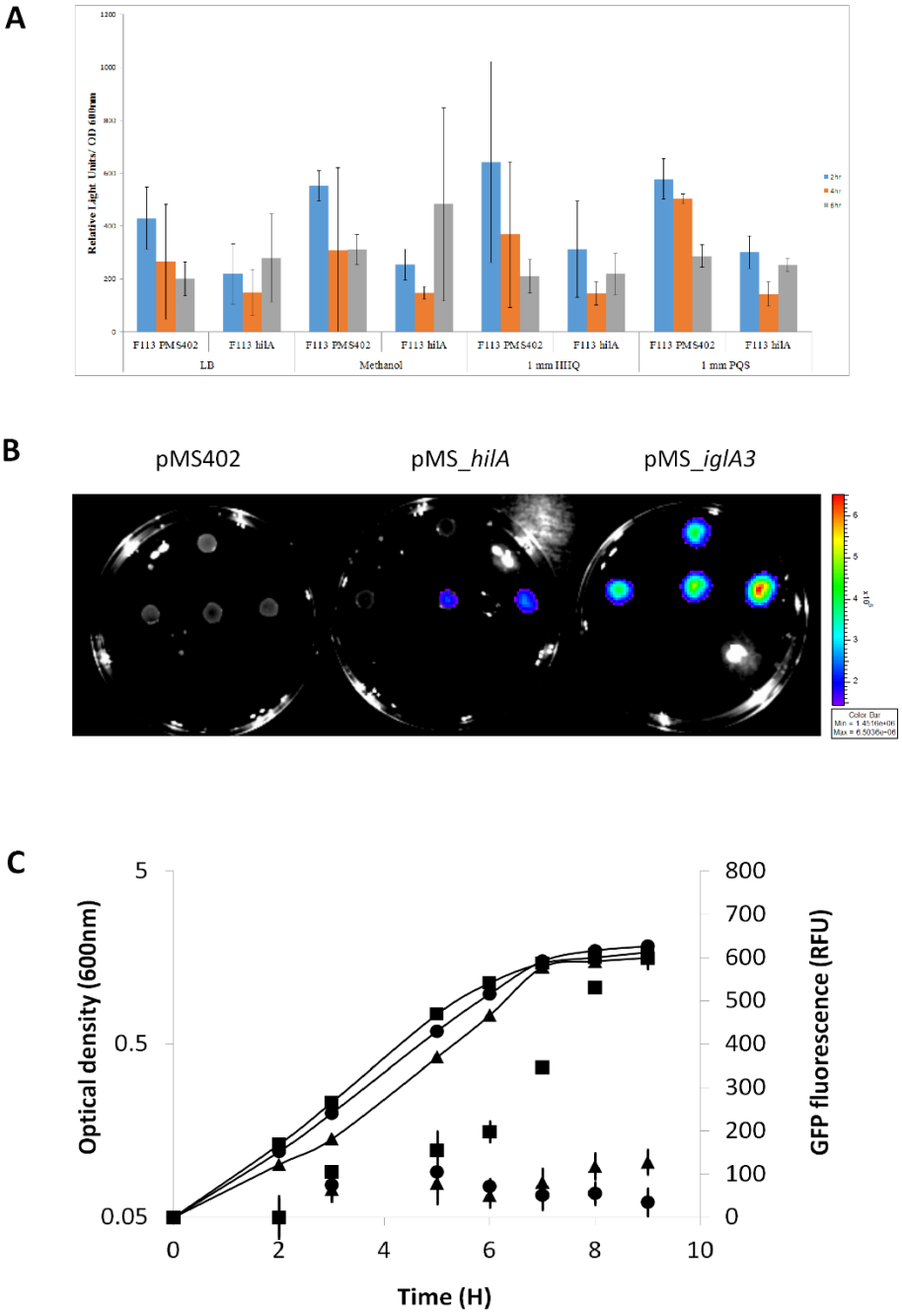


Figure 4.2 (A) Light production by strains carrying the PMS402 with the *hilA* promoter or no promoter. Expression of *hilA* is not induced by the presence of signalling molecules HHQ or PQS. As these molecules are dissolved in methanol, an equal concentration of methanol was used for cultures lacking the molecules. Error bars represent 95% confidence intervals.

(B) Composite photographic and luminescence images of *hilA* and *iglA3* (*tssB3*) promoter fusions *P. fluorescens* strains carrying the promoterless vector pMS402, pMS-*hilA* or pMS-*iglA3* were spot-inoculated on PYG medium supplemented with trimethoprim. Four spots corresponding to F113 strains without (upper spot) or with amoebae inoculated at 1 cm (left), 0.5 cm (centre) or at contact (right) of the bacterial spot are presented on each plate. Pictures were taken after 48 h of interactions. The log

scales on the right of each image represent luminescence in photons per second per square centimetre per steradian.

(C) Transcriptional fusions were grown in LB at 30 °C shaking. Growth curves (principal y axis) and relative fluorescens units (secondary y axis) were monitored in F113 pPROBE-GT (circle), F113 pPROBE-*hilA* (square) and in FG9 pPROBE-*hilA* (triangle). Each value is the mean of three samples obtained from three independent experiments.

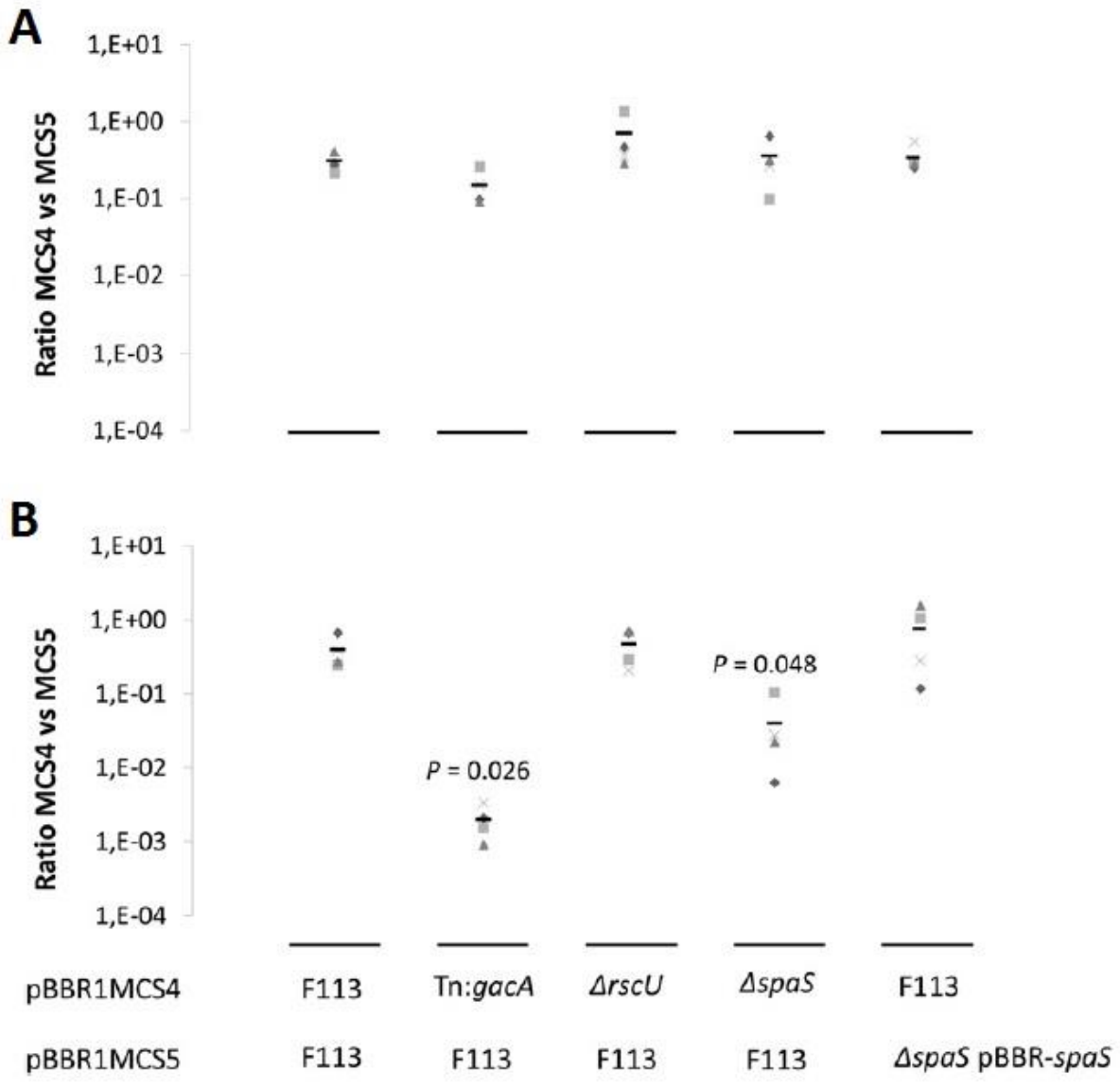


Figure 4.3. Relative fitness of mutant in PAS lacing amoeba **(A)** or PAS containing amoeba **(B)**. Fitness is expressed as ratio (log10 transformed) between strains carrying the vector pBBR1MCS4 and strains carrying the vector pBBR1MCS5. Experiments have been performed on four independent replicates.

While strains with a non-functional Hrp1 T3SS showed no increased susceptibility to grazing, the ratio of wild type to *spaS* mutant after 48 hours of co-incubation with amoeba was roughly 25:1, and this phenotype was complementable by expression of vector-borne *spaS* **(Figure 4.3)**. The effect was more pronounced when using the *gacA* mutant, where the final ratio of wild type to mutant was approximately 100:1. Importantly, competing strains incubated in PAS solution displayed comparable survival rates in the absence of amoeba.

The effect of SPI-I T3SS of *P. fluorescens* F113 on the mortality of another bacteriovorous model organism, *Caenorhabditis elegans* was tested to see if the T3SS was used for interactions with other grazers. Under slow-killing conditions (Tan *et al*., 1999), the *gacA* negative strain FG9 is severely defective in its ability to kill *C. elegans*, confirming previous results obtained with *P. protegens* CHA0 (Neidig *et al*., 2011) (**Figure 4.4**). Unexpectedly, worms exposed to the *spaS* mutant exhibited an increase in mortality rate relative to control worms feeding on *P. fluorescens* F113 ($P < 0.0002$) (**Figure 4.4 (B)**). The trans-complemented strain presented a virulence activity comparable to the wild type. A possible explanation for the increase in *C. elegans* mortality observed when the SPI-1 system is non-functional is that when *C. elegans* is exposed to the SPI-I needle or SPI-1 effectors, it mounts a defence response, enabling it to survive for a longer period of time. To test if the defence response of *C. elegans* was specific to the SPI-I T3SS or a more general reaction to T3SS protein, the survival of *C. elegans* exposed to a Δ*rscU* strain was also assessed. The RscU protein is the functional equivalent of the SpaS protein for the Rsp-1 T3SS. Although, the same effect is observed for the Δ*rscU* strain, this phenotype is not restored by trans-complementation (**Figure 4.4 (C)**).

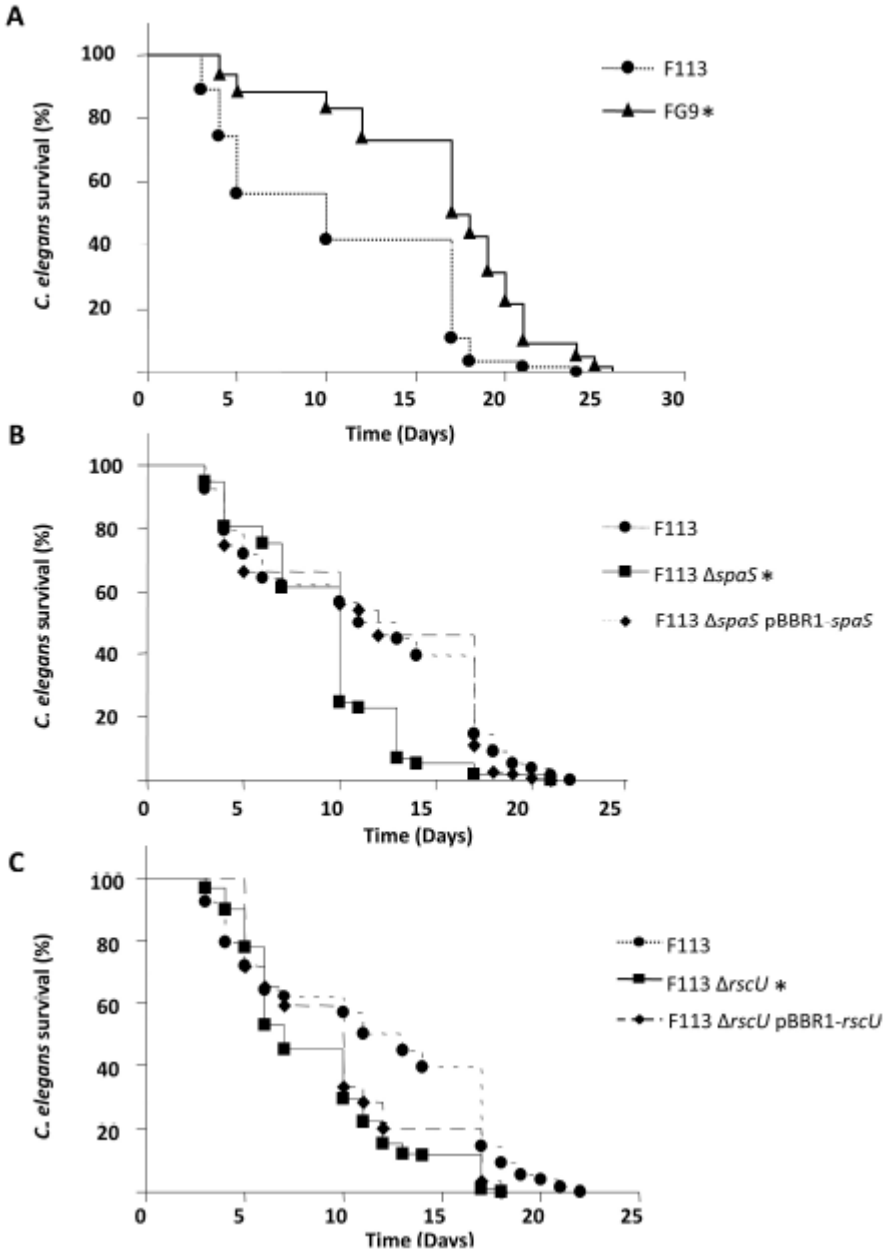


Figure 4.4. Kaplan-Meier survival plots of worms fed with *P. fluorescens* F113 wild type and **(A)** *gacA* mutant, **(B)** spaS mutant and complemented mutant and **(B)** *rscU* mutant and complemented mutant. Each value reported for the assay is the mean of measurements of nine samples from three independent experiments. Asterisks indicates significant difference ($P<0.05$) between F113 wild-type and the corresponding mutant tested, as assessed by pairwise comparisons (log rank test).

That *P. fluorescens* F113 is likely to compete with insects in its niche is reflected in the fact that the genome encodes several putative insecticidal toxins. Indeed, other strains of *P. fluorescens* such as CHAO and Pf-5 possess virulence towards *Galleria mellonella*, though this is not mediated through the T3SS (Péchy-Tarr *et al*., 2008). However, *Galleria mellonella* killing by *P. aeruginosa* is mediated through T3SS (Miyata *et al*., 2003). Moreover, the T3SS of *Pantoea stewartii* is required for the persistence of this bacteria in flea beetle-guts, and this T3SS belongs to the SPI-1 family (Correa *et al*., 2012). At an inoculum density of 5 x 10^6 bacterial cells, no significant difference in toxicity towards the larvae was observed between the different F113 strains (**Figure 4.5**), which suggests that the SPI-1 T3SS is not involved in the killing of *G. mellonella*.

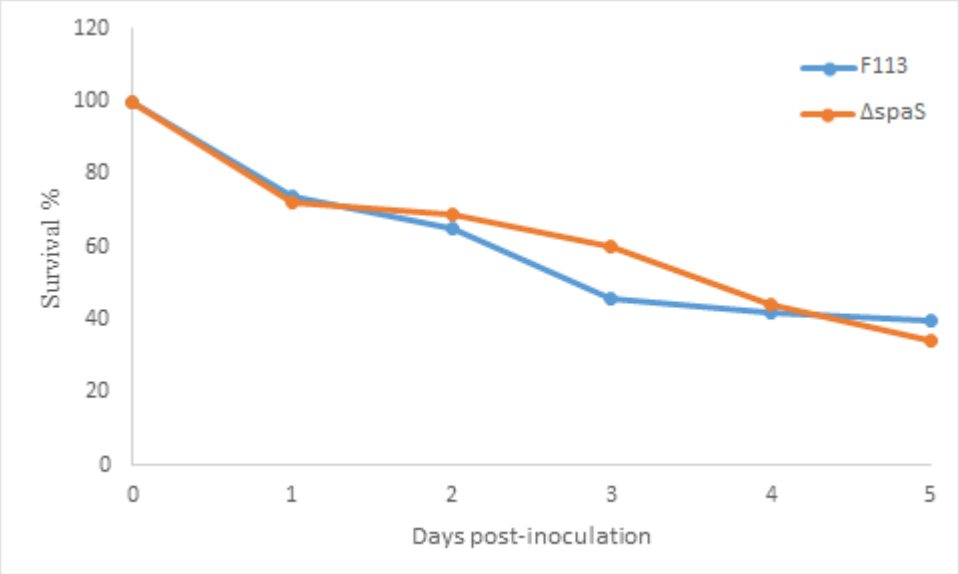


Figure 4.5. Toxicity of *Pseudomonas fluorescens* F113 and the *spaS* mutant to larvae of *Galleria mellonella*. The percentage of survival is shown for 30 larvae treated with F113 (blue) and Δ*spaS* (red) strains. Bacteria were injected at 5 x 10^6 cells per larvae.

Discussion.

The rhizosphere is an intensely competitive environment involving interactions with a range of prokaryotes and eukaryotes. The finding that *P. fluorescens* F113 possesses two different T3SS underlines the importance for bacteria of being able to influence interactions with different eukaryotes to their own advantage in this niche. The Hrp-1 T3SS is also present in *P. fluorescens* Q8r1-96 where it is used to modulate the plant immune response *(Mavrodi et al.*, 2011). This is also likely to be the function of this system in F113 as both F113 and Q8r1-96 share identical Hrp-1 T3SS effector repertoires and their Hrp-1 structural genes have a high level of sequence identity. The other T3SS, belonging to the SPI-1 family, appears to be important in allowing F113 to avoid amoeboid grazing **(Figure 4.6)**.

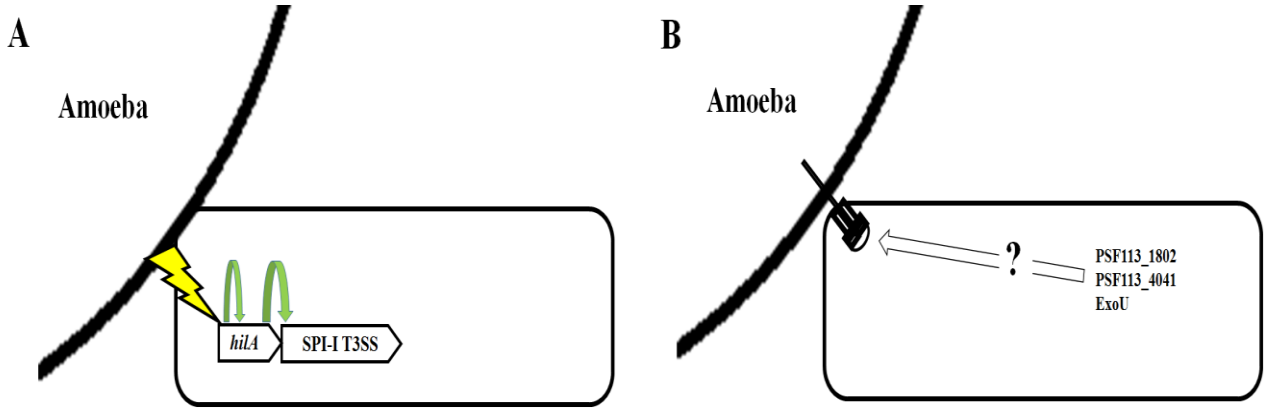


Figure 4.6. (A) Contact between the amoeba and F113 induces expression of the *hilA* gene, resulting in upregulation of the T3SS via an unknown signal. **(B)** The presence of the completed SPI-1 T3SS enhances F113 survival as the amoeba preferentially consumes T3SS mutant strains. The mechanism behind this phenotype is unknown, but may involve the activities of putative T3SS effectors PSF113_1802, PSF_4041 or ExoU.

There is a 25 fold difference in survival between the SPI-1 mutant and wild type strains when co-cultured with the amoeba, suggesting that this system plays a role in the rhizosphere competence of F113.

Interestingly, the grazing assays show F113 are a viable food source for *A. polyphaga* and these bacteria are unable to either kill the amoeba or survive within them. Therefore, amoeba killing is not a necessary component of grazing avoidance. Indeed, in a mixed population, such as is found in the rhizosphere, influencing the grazing of amoeba to extent seen might be beneficial or almost as beneficial as amoeba killing. In fact, leaving the amoeba alive to graze on competing bacteria may be ultimately be a successful strategy for F113 to colonise its niche.

As mentioned, *Salmonella* species have multiple uses for their SPI-1 T3SS, including activity against mammalian cells and amoeba, with the activity seen to date being much more dramatic in the case of the former. In light of this it may be that the lack of amoeba killing is due to the activity against amoeba being a secondary function of this system, and alternate roles for this system await discovery.

However, it is possible that the lack of killing in our experiments is due to the strain of *Acanthamoeba* used, and other strains or even other genera of bacterivores would be susceptible to T3SS-dependent killing by F113.

In previously published assays showing differential survival of the wild type strain and *gacS* mutants of *P. protegens* CHAO in a mixed population predated by amoeba, the effect is attributed to the production of secondary metabolites under the regulation of the *Gac* system (Jousset *et al*., 2009). It is not known how T3SS expression is induced by contact with amoeba, but the finding that *hilA* expression is reduced in a *gacA* transposon mutant suggests that the global *P. fluorescens* response to the presence of amoeba may be mediated through the *Gac* system and that the activation of T3SS has been integrated into this regulatory pathway. In competitive feeding assays the survival of a *gacA* transposon mutant was lower than that of an *spaS* mutant. Therefore, it is likely that F113 uses both secondary metabolite production and its SPI-1 T3SS to avoid grazing. While inactivation of *Gac* genes in *P. protegens* CHAO increases susceptibility to amoeba grazing, the increase is not a great as seen in the Gac mutant in *P. fluorescens* F113. The lack of any T3SS in *P.*

protegens is likely the reason for this difference between the strains in their interactions with the bacterivore.

While the presence of the SPI-1 T3SS is of benefit to F113 during encounters with *A. polyphaga*, it seems to be detrimental in interactions with another bacterivore: *C. elegans*. This is somewhat analogous to the paradigm of compatible and incompatible interactions between phytopathogens and plants, where with one host the use of the T3SS results in infection but with another it limits infection as the plant mounts a defence response (Göhre and Robatzek, 2008). The phenotype observed is possibly due to a *C. elegans* induced defence response, but it is not clear whether this is a specific response to the secreted effectors, or a general response to PAMPS such as the SPI-1 needle or translocators.

Several hypotheses may be put forward to explain the enhanced survival of F113 compared to the SPI-1 mutant. Given that many eukaryotes, including amoeba, are susceptible to SPI-1 dependent killing by other organism, it is possible that the amoeba have evolved to actively seek other prey after perception of the SPI-1 needle or translocators, or secreted effectors which contain T3SS secretion signals. Alternatively, the phagocytic process may be physically impaired by the presence of the SPI-1 needle or translocators. Another possibility is that the phenotype may depend on the action of the translocated effectors, which may interfere with the phagocytic mechanism while not having any lethal effects on the cell, or may display limited general cytotoxicity which is not sufficient to kill the amoeba, but can result in an impaired phagocytosis.

Research into the roles of the SPI-1 T3SS and the underlying mechanisms behind such SPI-1 dependent phenotypes outside the paradigm of mammalian infection is hampered by the lack of known effectors. In this respect, the identification of two novel putative effectors, PSF113_1801 and PSF113_4041, is an important result. These effectors may be responsible for the resistance to grazing phenotype. Further assays will determine what involvement these effectors may have in this potentially widespread and important phenotype among rhizobacteria.

References

Abby, S.S. and Rocha, E.P.C. (2012) The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems. *PLoS Genet* **8**: e1002983.

Ahmer, B.M., van Reeuwijk, J., Watson, P.R., Wallis, T.S., and Heffron, F. (1999) *Salmonella* SirA is a global regulator of genes mediating enteropathogenesis. *Mol. Microbiol.* **31**: 971–982.

Alavi, S.M., Sanjari, S., Durand, F., Brin, C., Manceau, C., and Poussier, S. (2008) Assessment of the Genetic Diversity of *Xanthomonas axonopodis* pv. phaseoli and Xanthomonas fuscans subsp. fuscans as a Basis To Identify Putative Pathogenicity Genes and a Type III Secretion System of the SPI-1 Family by Multiple Suppression Subtractive Hybridizations. *Appl. Environ. Microbiol.* **74**: 3295–3301.

Alfano, J.R. and Collmer, A. (1997) The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J. Bacteriol.* **179**: 5655–5662.

Correa, V.R., Majerczak, D.R., Ammar, E.-D., Merighi, M., Pratt, R.C., Hogenhout, S.A., *et al*. (2012) The Bacterium *Pantoea stewartii* Uses Two Different Type III Secretion Systems To Colonize Its Plant Host and Insect Vector. *Appl. Environ. Microbiol.* **78**: 6327–6336.

Dale, C. and Moran, N.A. (2006) Molecular Interactions between Bacterial Symbionts and Their Hosts. *Cell* **126**: 453–465.

Dean, P. (2011) Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol. Rev.* **35**: 1100–1125.

Delany, I.R. (1999) Genetic analysis of the production of the antifungal metabolite 2,4-diacetylphloroglucinol by the biocontrol strain *Pseudomonas fluorescens* F113. PhD Thesis National University of Ireland, Cork, Ireland.

Duan, K., Dammel, C., Stein, J., Rabin, H., and Surette, M.G. (2003) Modulation of *Pseudomonas aeruginosa* gene expression by host microflora through interspecies communication. *Mol. Microbiol.* **50**: 1477–1491.

Figurski, D.H. and Helinski, D.R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci.* **76**: 1648–1652.

Galán, J.E. (1996) Molecular genetic bases of *Salmonella* entry into host cells. *Mol. Microbiol.* **20**: 263–271.

Göhre, V. and Robatzek, S. (2008) Breaking the Barriers: Microbial Effector Molecules Subvert Plant Immunity. *Annu. Rev. Phytopathol.* **46**: 189–215.

Green, M.R. (2012) Molecular cloning: a laboratory manual 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Izoré, T., Job, V., and Dessen, A. (2011) Biogenesis, Regulation, and Targeting of the Type III Secretion System. *Structure* **19**: 603–612.

Jousset, A., Rochat, L., Péchy-Tarr, M., Keel, C., Scheu, S., and Bonkowski, M. (2009) Predators promote defence of rhizosphere bacterial populations by selective feeding on non-toxic cheaters. *ISME J.* **3**: 666–674.

Kovach, M.E., Elzer, P.H., Steven Hill, D., Robertson, G.T., Farris, M.A., Roop II, R.M., and Peterson, K.M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**: 175–176.

Mao, W., Warren, M.S., Black, D.S., Satou, T., Murata, T., Nishino, T., *et al*. (2002) On the mechanism of substrate specificity by resistance nodulation division (RND)-type multidrug resistance pumps: the large periplasmic loops of MexD from *Pseudomonas aeruginosa* are involved in substrate recognition. *Mol. Microbiol.* **46**: 889–901.

Mavrodi, D.V., Joe, A., Mavrodi, O.V., Hassan, K.A., Weller, D.M., Paulsen, I.T., *et al*. (2011) Structural and Functional Analysis of the Type III Secretion System from *Pseudomonas fluorescens* Q8r1-96. *J. Bacteriol.* **193**: 177–189.

Miki, T., Iguchi, M., Akiba, K., Hosono, M., Sobue, T., Danbara, H., and Okada, N. (2010) *Chromobacterium* pathogenicity island 1 type III secretion system is a major virulence determinant for *Chromobacterium violaceum*-induced cell death in hepatocytes. *Mol. Microbiol.* **77**: 855–872.

Miller, W.G., Leveau, J.H.J., and Lindow, S.E. (2000) Improved *gfp* and *inaZ* Broad-Host-Range Promoter-Probe Vectors. *Mol. Plant. Microbe Interact.* **13**: 1243–1250.

Miyata, S., Casey, M., Frank, D.W., Ausubel, F.M., and Drenkard, E. (2003) Use of the Galleria mellonella Caterpillar as a Model Host To Study the Role of the Type III Secretion System in *Pseudomonas aeruginosa* Pathogenesis. *Infect. Immun.* **71**: 2404–2413.

Neidig, N., Paul, R.J., Scheu, S., and Jousset, A. (2011) Secondary Metabolites of *Pseudomonas fluorescen*s CHA0 Drive Complex Non-Trophic Interactions with Bacterivorous Nematodes. *Microb. Ecol.* **61**: 853–859.

Pathuri, P., Nguyen, E.T., Ozorowski, G., Svärd, S.G., and Luecke, H. (2009) Apo and Calcium-Bound Crystal Structures of Cytoskeletal Protein Alpha-14 Giardin (Annexin E1) from the Intestinal Protozoan Parasite Giardia lamblia. *J. Mol. Biol.* **385**: 1098–1112.

Péchy-Tarr, M., Bruck, D.J., Maurhofer, M., Fischer, E., Vogne, C., Henkels, M.D., *et al*. (2008) Molecular analysis of a novel gene cluster encoding an insect toxin in plant-associated strains of *Pseudomonas fluorescens*. *Environ. Microbiol.* **10**: 2368–2386.

Preston, G.M. (2007) Metropolitan Microbes: Type III Secretion in Multihost Symbionts. *Cell Host Microbe* **2**: 291–294.

Preston, G.M., Bertrand, N., and Rainey, P.B. (2001) Type III secretion in plant growth-promoting *Pseudomonas fluorescens* SBW25. *Mol. Microbiol.* **41**: 999–1014.

Rainey, P.B. (1999) Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. *Environ. Microbiol.* **1**: 243–257.

Redondo-Nieto, M., Barret, M., Morrisey, J.P., Germaine, K., Martínez-Granero, F., Barahona, E., *et al*. (2012) Genome Sequence of the Biocontrol Strain *Pseudomonas fluorescens* F113. *J. Bacteriol.* **194**: 1273–1274.

Rezzonico, F., Défago, G., and Moënne-Loccoz, Y. (2004) Comparison of ATPase-Encoding Type III Secretion System hrcN Genes in Biocontrol Fluorescent Pseudomonads and in Phytopathogenic Proteobacteria. *Appl. Environ. Microbiol.* **70**: 5119–5131.

Schikora, A., Virlogeux-Payant, I., Bueso, E., Garcia, A.V., Nilau, T., Charrier, A., *et al*. (2011) Conservation of *Salmonella* Infection Mechanisms in Plants and Animals. *PLoS ONE* **6**: e24112.

Sperandio, D., Decoin, V., Latour, X., Mijouin, L., Hillion, M., Feuilloley, M.G.J., *et al*. (2012) Virulence of the *Pseudomonas fluorescens* clinical strain MFN1032 towards *Dictyostelium discoideum* and macrophages in relation with type III secretion system. *BMC Microbiol.* **12**: 223.

Tan, M.-W., Mahajan-Miklos, S., and Ausubel, F.M. (1999) Killing of *Caenorhabditis elegan*s by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. *Proc. Natl. Acad. Sci.* **96**: 715–720.

Tezcan-Merdol, D., Ljungström, M., Winiecka-Krusnell, J., Linder, E., Engstrand, L., and Rhen, M. (2004) Uptake and Replication of *Salmonella enterica* in *Acanthamoeba rhysodes*. *Appl. Environ. Microbiol.* **70**: 3706–3714.

Triplett, L.R., Zhao, Y., and Sundin, G.W. (2006) Genetic Differences between Blight-Causing *Erwinia* Species with Differing Host Specificities, Identified by Suppression Subtractive Hybridization. *Appl. Environ. Microbiol.* **72**: 7359–7364.

Vandesompele, J., Preter, K.D., Pattyn, F., Poppe, B., Roy, N.V., Paepe, A.D., and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**: research0034.

Chapter 5

General Discussion

Ideas in this chapter contributed to the following publication:

Title: The SPI-1-like Type III secretion system: more roles than you think

Authors: F. Egan, J.F. Reen and F. O'Gara

The overall aim of this work was to examine the functions and mechanisms of the T3SS and T6SS, using an integrated bioinformatic and experimental approach. Within *Pseudomonas* these secretion systems are particularly prevalent, perhaps as these bacteria live in various niches associated with a broad diversity of prokaryotic and eukaryotic taxa. As the function of these secretion systems is ultimately dependent on the activities of their secreted substrates, finding effectors secreted by these systems is of utmost importance.

VgrG islands are a previously unknown source of effectors

T3SS has been a topic of research for decades, and many of its effectors have been identified from a range of model organisms (Dean, 2011). In contrast, T6SS is a more recently described system. Notwithstanding this, the T6SS field has seen rapid progress in many areas, including the identification and characterisation of effectors. **Table 5.1** shows experimentally validated and putative T6SS effectors in PAO1, all of which were identified post 2010. These include the Tge proteins, Tae proteins, Tde proteins Rhs proteins and the Tle proteins (Koskiniemi *et al*., 2013; Ma *et al*., 2014; Russell *et al*., 2014).

Many T6SS effectors are now known to be encoded in horizontally transferred operons which include *vgrG* genes, called *vgrG* islands, which are encoded away from T6SS loci. These islands were first identified by the bioinformatics approaches outlined in Chapter 2, and correctly hypothesised to be a common source of T6SS effectors throughout bacteria possessing the T6SS. The common lipase genes within the *vgrG* islands in particular were considered likely to encode T6SS effectors in Chapter 2. This hypothesis has now been confirmed with the identification of the Tle superfamily (Russell *et al*., 2013). In addition, several other effectors are now known to be encoded in various *vgrG* islands e.g. VasX (Miyata *et al*., 2011).

The *vgrG* islands are widespread, diverse and horizontally transmitted. Therefore, not only are they potentially a source of new effectors in an extensive range of bacteria, but they also represent a key mechanism of diversification of T6SS function.

An interesting hypothesis in Chapter 2 was that these effectors were specifically or preferentially secreted by certain phylogenetic families of T6SS.

Table 5.1. Genes in *P. aeruginosa* PAO1 which have been shown or predicted to be T6SS effectors. Based on conserved elements or sequence homology several of these genes have been shown to be encoded in T6SS loci or *vgrG* islands in multiple genera of bacteria. Despite their inclusion in this table, there have been no reported attempts to prove that the Tle3 or Tle4 proteins are functional T6SS effectors. In addition PA3907 has only been annotated as a toxin gene recently due to advances in bioinformatic prediction and its association with T6SS is based on the presence of a PAAR gene within the same operon.

Gene	Gene name	Conserved Element	T6SS associated in multiple genera
PA0092	Tsi4		No
PA0093	Tse4	PAAR	Yes
PA0094		Pfam08786	Yes
PA0095	VgrG		Yes
PA0096			Yes
PA0097	TagA	Pfam09937	Yes
PA0098			Yes
PA0099		PAAR	Yes
PA0100			Yes
PA0101			Yes
PA0259	Tli3		Yes
PA0260	Tle3		Yes
PA0261			Yes
PA0262	VgrG		Yes
PA0263	Hcp		Yes
PA1508		PAAR	Yes
PA1509	Tli4		Yes
PA1510	Tle4		Yes
PA1511	VgrG		Yes
PA1512	Hcp		Yes
PA1844	Tse1		No
PA1845	Tsi1		No
PA2683	Tsi6		
PA2684	Tse6	Rhs	Yes
PA2685	VgrG		Yes
PA2702	Tse2		No
PA2703	Tsi2		No

PA3904		PAAR	Yes
PA3905		Pfam13503	Yes
PA3906			No
PA3907		Pfam15648 - Tox-REase-5	No
PA3908		Pfam15579 - Imm32	No
PA3290	Tle1		Yes
PA3291	Tli1a		Yes
PA3292	Tli1b		Yes
PA3293		Pfam13503	Yes
PA3294	VgrG		Yes
PA3484	Tse3		No
PA3485	Tsi3		No
PA3486	VgrG		Yes
PA3487	Tle5a	PldA	Yes
PA3488	Tli5a		Yes
PA5086	Tli5b		Yes
PA5087	Tli5b		Yes
PA5088	Tli5b		Yes
PA5089	Tle5b		Yes
PA5090	VgrG		Yes
PA5264			Yes
PA5265			Yes
PA5266	VgrG		Yes
PA5267	Hcp		Yes

T3SS and T6SS can be subdivided into distinct phylogenetic families.

One of the notable features of T3SS and T6SS is that they can be subdivided into separate phylogenetic families. Indeed, work in Chapter 2 identified a new family within the T6SS phylogeny. Phylogenetic analyses have highlighted seven different families of T3SS and have been useful for unravelling the evolutionary history of the system, as well as highlighting the relationship between the T3SS family and (a) the specificities of its translocation process as well as (b) broader categories of target organisms. Indeed, the difference in the extracellular components of plant and animal-related T3SS is striking. These differences and much of the early experimental evidence suggested that target specificity for plants or animal cells was

absolute. However, recent reports show that the SPI-I T3SS of *Salmonella* has an additional function in the colonisation of plants, despite lacking the pilus (Schikora *et al.*, 2008; Shirron and Yaron, 2011). Interestingly, the clinical isolate *P. fluorescens* MFM1032 use a T3SS which probably belongs to the Hrp1 family to lyse macrophages and interfere with the growth of amoeba (Sperandio *et al*., 2012). Unfortunately, in the absence of a genome sequence it is not clear whether this strain possesses the plant-associated pilus and translocator repertoire.

Although extremely diverse, previously described Type Three effectors are composed of less than 40 motifs or domains which have the potential to interfere with conserved eukaryotic cellular processes (Dean, 2011). Therefore, it may not be surprising that in some situations, certain T3SS families have roles which are not classically associated with this secretion system. In addition, the increase in genomic sequences shows that certain T3SS families are encoded in a wider range of bacteria than previously thought.

One such example is the presence of the SPI-1 system in a PGPR strain such as *Pseudomonas fluorescens* F113. Work in Chapter 4 showed that this system is important for allowing the bacterium to avoid amoeboid grazing in the context of a mixed bacterial community. New sequencing data is revealing that the SPI-1 T3SS in not uncommon in other rhizobial strains. Thus, the SPI-1 T3SS could be an important factor in the rhizocompetence of bacterial strains associated with plant roots.

While the mechanism underpinning this phenotype is currently unknown, two novel putative T3SS effectors were identified in this strain in Chapter 4. Future experimental work could determine whether the action of these uncharacterised effectors is responsible for the observed phenotype.

Other roles for the SPI-I system outside of mammalian pathogens have been reported recently (Correa *et al*., 2012), and it is likely that more roles for each T3SS family remain to be discovered as the known host range of these T3SS expands. However, current evidence suggests that the phylogenetic grouping of the T3SS families reflects a genuine split between the plant and animal-targeting capabilities of certain T3SS, with rare exceptions.

With T6SS, the relevance of different phylogenetic families is unclear. Certainly, the fact that accessory gene content is associated specifically with certain T6SS families, reported here and elsewhere, suggests that phylogenetic differentiation can be associated with mechanistic differences (Boyer *et al*., 2009). One such example is the post-translational regulatory pathway involving Fha. Another difference is the presence of an enhancer binding protein which upregulates expression of its cognate T6SS locus. However, from the sizeable body of T6SS research, it now appears that there is no meaningful correlation between phylogenetic family of T6SS and target organism. Indeed, several T6SS and even individual effectors have multiple targets. The most notable example of the latter is the Tle family, which is active against both eukaryotic and prokaryotic membranes (Dong *et al*., 2013; Russell *et al*., 2013; Jiang *et al*., 2014). This is consistent with the experimental evidence provided in Chapter 2 of this thesis.

While *P. fluorescens* F113 and *P. aeruginosa* PA14 possess very similar T6SS, only F113 appears to use its T6SS to target the social amoeba *Dictyostelium discodeum*. Furthermore, while a *P. aeruginosa* PAO1 mutant with a defective HSI-II system was impaired in virulence towards the model nematode *C. elegans* and in internalisation into human epithelial cells, these phenotypes were not observed using the corresponding PA14 mutant (Sana *et al*., 2012). In fact, recent reports show that a PA14 HSI-II mutant is internalised into human epithelial cells at a higher rate than wild type PA14 (Jones *et al*., 2014). This is in agreement with the trend, albeit not considered statistically significant, from the internalisation assays in Chapter 2.

While the correlation between target and T6SS family does not appear to exist, the existence of several distinct T6SS families provides a useful scaffold upon which to dissect the associations with the diversity of machinery and effector proteins. An interesting feature of T6SS is that while the core genes are typically encoded in one genomic locus, several copies of the *vgrG* and *hcp* genes could be found scattered throughout the chromosome. Phylogenetic analysis presented in Chapter 2 linked these orphan *vgrG* genes with particular T6SS families. Following on from the discovery of *vgrG* islands, it was hypothesised that effectors within these *vgrG* islands could also be linked to certain T6SS families.

Bioinformatic evidence suggests a range of protein-protein interactions of currently uncharacterised T6SS associated genes may be responsible for effector recruitment.

The possible association of effectors with certain T6SS phylogenetic families raises the question: what mechanism could be responsible for such an association? The recruitment of effectors to the T6SS apparatus could be mediated by effector binding of the T6SS apparatus, of the VgrG protein, or of specific adapters/chaperones. Several T6SS associated genes of unknown function were considered possible adapter proteins. If they co-occurred with certain subsets of effectors or T6SS structural proteins it would provide a basis for experimental work which attempts to elucidate their function.

Perhaps the strongest hypothesis from the co-occurrence profiles was that the separate subfamilies of the PAAR proteins have distinct interacting partners: Pfam08786 and Pfam13503 interact with the major subfamily and Pfam, and TagA interacts with the minor subfamily. As PAAR proteins are thought to be involved in effector recruitment, it is possible that the interactions with TagA, Pfam08786 and Pfam13503 are important for the purpose of effector recognition or localisation to the T6SS apparatus. Another interesting hypothesis from the analysis of co-occurrence profiles was that the interactions of the major PAAR family with the Pfam08786 and Pfam13503 domains were either/or interactions, as these two genes very rarely co-occurred. Putative associations between various Tle families and the subset of VgrG proteins containing the Pfam10106 domain, the major PAAR proteins and the Pfam13503-domains were also suggested by co-occurrence profiles.

How certain effectors are recruited to the T6SS apparatus and whether this recruitment is specific for certain systems have been subject to much research in recent years. The HSI-I locus of *P. aeruginosa* is phylogenetically linked to three VgrG proteins: VgrG1a-c. While any of these VgrG proteins are sufficient for T6SS activity, secretion of some effectors is specifically dependent on the presence of an adjacently encoded VgrG protein (Hachani *et al*., 2014). This is in agreement with the hypothesis presented in Chapter 2, which linked effectors in *vgrG* islands to T6SS via their cognate VgrG. Indeed, such specificity may be due to the activities of putative adapter proteins as hypothesised in Chapter 3, as the *vgrG* islands

containing two of these proteins also contain predicted adapter proteins. Determining if, and indeed how, putative adapters interact with VgrG proteins, and whether these interactions are exclusive to certain VgrG, could provide vital information regarding the mechanism through which proteins are targeted to and selectively secreted by the T6SS.

Another effector, PldA, which is a member of the Tle5a family, is encoded in a *vgrG* island, the VgrG protein of which was predicted to associate with HSI-II. PldA was later shown to be secreted in a HSI-II-dependent manner (Russell *et al*., 2013). In contrast, despite being encoded adjacent to a *vgrG* gene (PA5090) belonging to the HSI-II family, Tle5b is in fact secreted in a HSI-III dependent manner. This would suggest that either the adjacent VgrG is not involved in secretion of this effector, or that this VgrG does not associate with the T6SS apparatus it was linked to in Chapter 2.

Interestingly this Tle5b protein was shown to interact with several phylogenetically distinct VgrG proteins in a bacterial two-hybrid assay (Jiang *et al*., 2014). This suggests there is no role for adapter proteins in recruitment of Tle5b proteins to the T6SS, and that this protein can be secreted by a range of T6SS. However, such assays do not necessarily mirror the conditions inside the cell whether a multiprotein complex is formed through numerous protein-protein interactions. Therefore, this result needs to be strengthened by additional evidence.

The role of putative adapter proteins remains unresolved. While it is understood that evolved VgrG and Hcp proteins are playing dual roles as structural proteins and effectors, it is unclear whether other effectors or putative accessory proteins are also important in the functioning of the T6SS. *Vibrio cholerae* strains lacking both the effector VasX and VgrG-3, which is not encoded in the vicinity of VasX, are secretion deficient (Miyata *et al*., 2013). However, single mutants retain their secretion ability. Further work is needed to understand the nature of the interactions at the tip of the T6SS apparatus which have resulted in such a phenotype. The protein-protein interactions hypothesised in Chapter 3 provide a strong theoretical basis for further work on this question.

Co-occurrence profiles as a tool for future research

While it will be interesting to see if the hypotheses based on the co-occurrence reported in Chapter 3 are accurate, there is no reason why this approach should be restricted to the T6SS field. Indeed, the underlying foundation of such an approach, the genomic co-occurrence of genes encoding interacting proteins, was established several years before the discovery of the T6SS (Dandekar *et al*., 1998). Furthermore, it has been used to identify linkages between proteins to some degree already. The "neighbourhood" criterion which is partly the basis for predicted protein-protein interactions on the STRING database is essentially based on this principle (Szklarczyk *et al*., 2011). However, comparisons in STRING are based on gene homology instead of domains such as Pfam, therefore broader co-occurrence profiles, such as those identified in Chapter 3, are missed.

As the data available to bioinformaticians increases it becomes increasingly necessary to automate large data mining projects. There is certainly scope for a database which automates the gene co-occurrence process outlined in Chapter 3, and uses this information to predict protein-protein interactions or proteins which participate in similar processes. As was demonstrated in Chapter 3, it will be important that such approaches can counter the bias introduced by the presence of unfinished genomes. In addition, there is a further bias introduced by the disproportional representation of a phylogenetically limited range of bacteria within genome databases, which must be overcome, perhaps by using phyletically restricted datasets.

Tle distribution in metagenomes

Combining experimental evidence from model organisms to the growing database of genomic sequences brings us closer to establishing what roles the T6SS is playing within diverse ecological niches. However, this question can only be fully addressed with large scale metagenomic analysis. The bias towards certain bacterial genera or species in genome sequences merely reinforces this. In Chapter 3, metagenomic data available on the IMG website was mined for the prevalence of T6SS genes. This approach has been used previously to augment an analysis on the prevalence of secretion systems within marine environments (Persson *et al*., 2009). However, the

analysis in Chapter 3 compared T6SS prevalence across a range of metagenomes from different environments, not just from marine metagenomes. While T6SS was initially characterised in pathogenic interactions it has now shown to deliver anti-bacterial effectors to competing micro-organisms. In addition, there is an increasing appreciation of the potential for co-operative behaviour and physiological roles to be mediated by the T6SS.

Therefore, an important innovation in this chapter was to also analyse the distribution of the Tle effector superfamily to look at a specific T6SS activity. Indeed, the distribution of Tle did not mirror the general T6SS distribution. Whereas T6SS levels were relatively consistent in all environments, Tle levels were quite varied. In particular, host-associated environments displayed much greater frequencies of Tle. This could reflect that host-associated niches support greater communities giving scope for more complex interactions, while bacteria in environments with less available energy are likely to be involved in less complex activities (Hoehler and Jørgensen, 2013). Furthermore, the observation of differential representation of each Tle family within the metagenome datasets suggests the possibility of some degree of niche-related specialisation. The existence of several separate families of Tle also supports the idea that each family is in some way specialised. While the Tle superfamily is considered to consist of five families (Russell *et al*., 2013), analysis presented in Chapter 3 suggests that Tle5 consists of two sub-families. This is based on phylogenetic differences in Tle proteins and differences in the corresponding immunity genes. Indeed, this hypothesis is supported by a recent report showing that the immunity proteins Tli5a and Tli5b confer immunity only against members of the Tle5a and Tle5b sub-families, respectively (Jiang *et al*., 2014). In fact, the split is so great between the subfamilies that a member of Tle5b sub-family was recently reported to be a novel virulence factor lacking any cognate immunity gene, as the only other characterised Tle5/Tli5 pair belonged to the other subfamily (Russell *et al*., 2013; Lery *et al*., 2014).

As new metagenomic data from different environments becomes available, these data will need to be interrogated to further support the hypothesis of niche specialisation. Furthermore, as metagenomic analysis can only identify the potential activity of communities, such analysis could be augmented by the inclusion of meta-transcriptome data which more accurately determines the activity of the certain

genes within the environment. Subsequent metagenomic and metatransciptomic analysis to investigate the role of T6SS within communities should also examine effector prevalence in this datasets instead of concentrating solely on the T6SS structural genes. It would be advantageous if these analyses include T6SS effector superfamilies other than Tle, and indeed such analyses may have to incorporate new Tle families should they be discovered, as these would obviously not have been detected by BLAST analysis using existing Tle proteins as bait sequences.

Why do the T3SS and T6SS both exist?

Two of the most significant experimental findings in this thesis are the involvement of the T6SS and T3SS in the resistance of *P. fluorescens* F113 to predation by two different amoeba: *D. discodeum* and *Acathamoeba polyphaga* **(Figure 5.1)**. The similarity of the roles of T3SS and T6SS in F113 encapsulates a fundamental question: why do both secretion systems exist? *D. discodeum* and *A. polyphagia* belong to the same phylum, yet are targeted by the different secretion systems in F113. Indeed, some of the earliest work on each of these secretion systems involved macrophage killing, in a T3SS-dependent manner by *Salmonella* species but in a T6SS-dependent manner by *Vibrio* species (Pukatzki *et al*., 2006; Rytkönen *et al*., 2007).

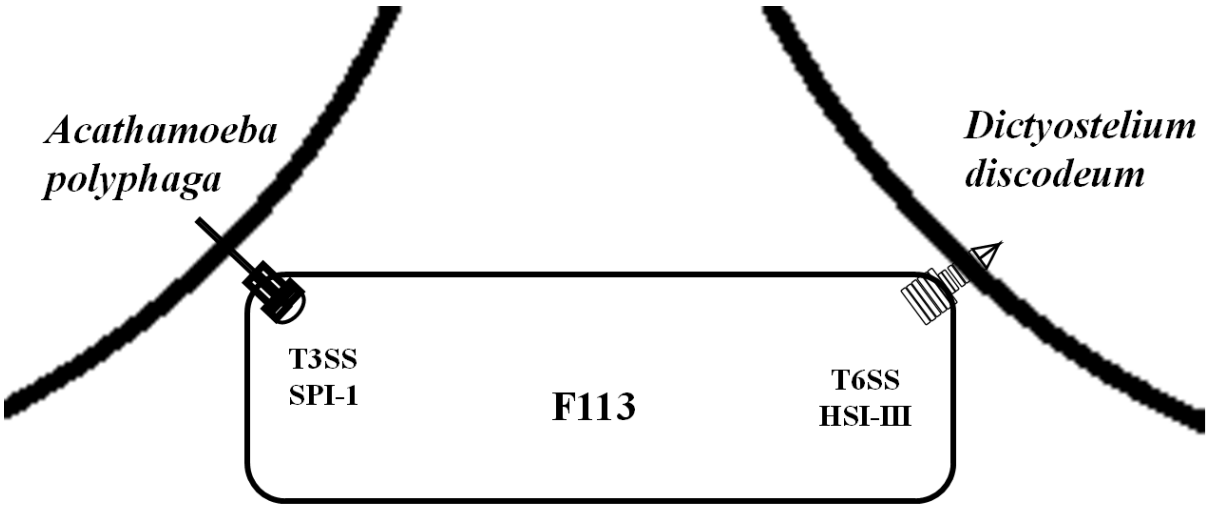


Figure 5.1. *P. fluorescens* F113 uses the SPI-1 T3SS to interact with *A. polyphaga* to avoid grazing and the HSI-III T6SS to inhibit plaque formation by *D. discodeum*.

The T6SS and T3SS are essentially systems for the transport of biomolecules through the dual membranes of the gram negative cell, and the membranes or cell walls of target cells. When viewed so simplistically they appear redundant. Yet, both evolved, and are sometimes encoded within the same genome. The reasons for why T6SS and T3SS both exist are perhaps revealed by the target they do not have in common: bacteria.

T3SS have never been shown to deliver molecules into other bacterial cells. In contrast, despite being initially characterised by its interactions with eukaryotes, there has been a gradual realisation that the T6SS is very likely to be mainly dedicated to interbacterial interactions. The theoretical framework for this view includes its similarity to bacteriophage and its much greater prevalence in genomes and metagenomes than the T3SS, particularly in genomes of bacteria which are not known to often associate with eukaryotic organisms (Leiman *et al*., 2009; Persson *et al*., 2009; Schwarz *et al*., 2010; Russell *et al*., 2014). Importantly, such a view is now supported by a myriad of examples in the scientific literature (as reviewed by Russell, Peterson, *et al*., 2014). T6SS-mediated pathogenesis towards higher eukaryotes seems to be the exception rather than the rule.

Prokaryotes have means to protect themselves from attack, such as pyocin release and even T6SS counter attack (Basler *et al*., 2013; Cornforth and Foster, 2013). However, this does not compare with the sophisticated immune systems of higher eukaryotes. In many cases it may be necessary for the pathogen to manage such a response. Several examples where the T3SS does just that have been reported in the literature. Despite displaying decreased levels of cellular invasion, SPI-I mutants provoke a stronger pro-inflammatory cytokine response *(*Pavlova *et al., 2011)*. In addition, it is well known that phytopathogens have an extensive repertoire of T3SS effectors dedicated to modulating the plant immune response *(Guo et al., 2009)*. The only example where T6SS appears to be involved in such a process is in the greater inflammation elicited by *Helicobacter hepaticus* T6SS mutants, by an unknown mechanism (Chow and Mazmanian, 2010).

Manipulating the immune response or the cellular regulation of a host cell is a delicate process. Several T3SS effectors influence the expression of eukaryotic transcription factors through the regulation of Rho proteins, while a T3SS family

from *Xanthomonas* has been shown to induce transcription of cellular genes by mimicking eukaryotic transcription factors (Kay *et al*., 2007; O'Grady *et al*., 2007). Perhaps for this reason, regulation of T3SS secretion appears more precise than regulation of the T6SS. For example, the YopK protein is a translocated effector which decreases levels of subsequent effector translocation. The *yopK* mutant displays greater cytotoxicity, but is less virulent to the host (Thorslund *et al*., 2013). T3SS effector ExoS from *P. aeruginosa* has a similar role in regulating subsequent effector secretion *(*Cisz *et al.*, 2008*)*. Though it is important to remember that T3SS has been the subject of more extensive investigation, no comparable regulatory feedback mechanism has been identified in T6SS.

Furthermore, non-pathogenic relationships between bacteria and eukaryotes can be quite complex. In several cases T3SS has been shown to play a key role in maintaining symbiotic relationships (Dale and Moran, 2006; Preston, 2007). Mediating such an intricate relationship would require a high degree of sophistication and control of effector secretion, perhaps not conducive to a secretion system like the T6SS, which appears to be less finely controlled.

The T3SS is often described as a molecular syringe, and this term was also once applied to the T6SS. However, given the association of effectors with the Hcp tubule or the VgrG trimer and the underlying mechanism of secretion being based on the contractile ability of the T6SS sheath, the T6SS is better described as a crossbow delivering a single poisoned arrow (Brunet *et al*., 2013). Not only is this a more apt comparison, it might be the clearest way of understanding the difference between, and the need for, both the T3SS and T6SS. T3SS evolved from flagella, a secretion system designed to secrete components for an organelle which was still under construction. As it came to mediate bacterial-host interactions, this level of sophistication may have beneficial. In contrast, T6SS appears to have been an adaption from the bacteriophage and which was useful in targeting other bacteria in a more explosive manner. Its targets may have now expanded to include eukaryotes, and but it still lacks a level of fine control comparable to the T3SS.

Concluding remarks

Since the discovery of T3SS and T6SS, huge progress has been made in understanding these systems, both in terms of mechanism and function. For example, in Chapter 2 *in silico* analysis led to the discovery of a widespread group of operons known as *vgrG* islands which are a common source of T6SS effectors. However, much of this progress has come from experimental work with model organisms. Indeed, this study has experimentally demonstrated T6SS and T3SS-dependent phenotypes in *P. fluorescens* F113, which may have significant roles for F113 in the colonisation of its niche. As both of the secretion systems involved in these phenotypes are found in other PGPR strains, this may be a common determinant of survival in the rhizosphere. Therefore, this knowledge has potential applications in the development of biocontrol strains which can contribute to sustainable agricultural practices.

However, a limitation of such experimental work is that conclusions from these studies may not reflect how bacteria within communities use these secretion systems and how they shape community structure. Metagenomic data provides the opportunity to answer such questions, if interrogated properly.

Bacteria have expanded to fill every niche, diversified into every lifestyle. The myriad uses of their widespread secretion systems reflect this. Results here and elsewhere show the T6SS can be split into phylogenetically distinct families. However, families do not appear to be linked with distinct roles. This illustrates a simple truth: the phenotype is determined by effectors. Therefore, examination of metagenomic data should be based on effector prevalence, not secretion system abundance. The approach outlined in Chapter 3 for the study of T6SS and Tle influence on environments will provide a template for future work in this area. In a pioneering bioinformatic study, Zhang and colleagues showed that diverse effectors could be identified by commonalities in their domain architecture (Zhang *et al*., 2012). Comprehensive metagenomic analysis of effectors based on this principle could help realise the vast potential of metagenomics to illuminate life in microbial communities.

An intriguing result in Chapter 3 further suggests overarching analysis of effectors may be a key area of future research. The distribution of T6SS-associated genes such

as *vgrG* exceeded the phylogenetic distribution of the T6SS itself, and these genes were found among other genes which are reminiscent of prophage-derived secretion systems. It is only in recent years that the frequency and influence of interbacterial interactions are being appreciated and that the nature of these interactions are being identified. It would be surprising if other secretion systems, phage-derived or not, which mediate direct interactions are not common throughout prokaryotes, waiting to be discovered, their influence waiting to understood, and even harnessed.

References

Basler, M., Ho, B.T., and Mekalanos, J.J. (2013) Tit-for-Tat: Type VI Secretion System Counterattack during Bacterial Cell-Cell Interactions. *Cell* **152**: 884–894.

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.

Brunet, Y.R., Espinosa, L., Harchouni, S., Mignot, T., and Cascales, E. (2013) Imaging Type VI Secretion-Mediated Bacterial Killing. *Cell Rep.* **3**: 36–41.

Chow, J. and Mazmanian, S.K. (2010) A Pathobiont of the Microbiota Balances Host Colonization and Intestinal Inflammation. *Cell Host Microbe* **7**: 265–276.

Cisz, M., Lee, P.-C., and Rietsch, A. (2008) ExoS Controls the Cell Contact-Mediated Switch to Effector Secretion in *Pseudomonas aeruginosa*. *J. Bacteriol.* **190**: 2726–2738.

Cornforth, D.M. and Foster, K.R. (2013) Competition sensing: the social side of bacterial stress responses. *Nat. Rev. Microbiol.* **11**: 285–293.

Correa, V.R., Majerczak, D.R., Ammar, E.-D., Merighi, M., Pratt, R.C., Hogenhout, S.A., *et al*. (2012) The Bacterium *Pantoea stewartii* Uses Two Different Type III Secretion Systems To Colonize Its Plant Host and Insect Vector. *Appl. Environ. Microbiol.* **78**: 6327–6336.

Dale, C. and Moran, N.A. (2006) Molecular Interactions between Bacterial Symbionts and Their Hosts. *Cell* **126**: 453–465.

Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**: 324–328.

Dean, P. (2011) Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol. Rev.* **35**: 1100–1125.

Dong, T.G., Ho, B.T., Yoder-Himes, D.R., and Mekalanos, J.J. (2013) Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. *Proc. Natl. Acad. Sci.* **110**: 2623–2628.

Guo, M., Tian, F., Wamboldt, Y., and Alfano, J.R. (2009) The Majority of the Type III Effector Inventory of *Pseudomonas syringae* pv. tomato DC3000 Can Suppress Plant Immunity. *Mol. Plant. Microbe Interact.* **22**: 1069–1080.

Hachani, A., Allsopp, L.P., Oduko, Y., and Filloux, A. (2014) The VgrG proteins are "A la carte" delivery systems for bacterial type VI effectors. *J. Biol. Chem.* jbc.M114.563429.

Hoehler, T.M. and Jørgensen, B.B. (2013) Microbial life under extreme energy limitation. *Nat. Rev. Microbiol.* **11**: 83–94.

Jiang, F., Waterfield, N.R., Yang, J., Yang, G., and Jin, Q. (2014) A *Pseudomonas aeruginosa* Type VI Secretion Phospholipase D Effector Targets Both Prokaryotic and Eukaryotic Cells. *Cell Host Microbe* **15**: 600–610.

Jones, C., Hachani, A., Manoli, E., and Filloux, A. (2014) An rhs Gene Linked to the Second Type VI Secretion Cluster Is a Feature of the *Pseudomonas aeruginosa* Strain PA14. *J. Bacteriol.* **196**: 800–810.

Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007) A Bacterial Effector Acts as a Plant Transcription Factor and Induces a Cell Size Regulator. *Science* **318**: 648–651.

Koskiniemi, S., Lamoureux, J.G., Nikolakakis, K.C., Roodenbeke, C. t'Kint de, Kaplan, M.D., Low, D.A., and Hayes, C.S. (2013) Rhs proteins from diverse bacteria mediate intercellular competition. *Proc. Natl. Acad. Sci.* **110**: 7032–7037.

Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., *et al*. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl. Acad. Sci.* **106**: 4154–4159.

Lery, L.M., Frangeul, L., Tomas, A., Passet, V., Almeida, A.S., Bialek-Davenet, S., *et al*. (2014) Comparative analysis of *Klebsiella pneumoniae* genomes identifies a phospholipase D family protein as a novel virulence factor. *BMC Biol.* **12**: 41.

Ma, L.-S., Hachani, A., Lin, J.-S., Filloux, A., and Lai, E.-M. (2014) *Agrobacterium tumefaciens* Deploys a Superfamily of Type VI Secretion DNase Effectors as Weapons for Interbacterial Competition In Planta. *Cell Host Microbe* **16**: 94–104.

Miyata, S.T., Kitaoka, M., Brooks, T.M., McAuley, S.B., and Pukatzki, S. (2011) *Vibrio cholerae* Requires the Type VI Secretion System Virulence Factor VasX To Kill Dictyostelium discoideum. *Infect. Immun.* **79**: 2941–2949.

Miyata, S.T., Unterweger, D., Rudko, S.P., and Pukatzki, S. (2013) Dual Expression Profile of Type VI Secretion System Immunity Genes Protects Pandemic *Vibrio cholerae*. *PLoS Pathog* **9**: e1003752.

O'Grady, E., Mulcahy, H., Adams, C., Morrissey, J.P., and O'Gara, F. (2007) Manipulation of host Kruppel-like factor (KLF) function by exotoxins from diverse bacterial pathogens. *Nat. Rev. Microbiol.* **5**: 337–341.

Pavlova, B., Volf, J., Ondrackova, P., Matiasovic, J., Stepanova, H., Crhanova, M., *et al*. (2011) SPI-1-encoded type III secretion system of *Salmonella enterica* is required for the suppression of porcine alveolar macrophage cytokine expression. *Vet. Res.* **42**: 16.

Persson, O.P., Pinhassi, J., Riemann, L., Marklund, B.-I., Rhen, M., Normark, S., *et al*. (2009) High abundance of virulence gene homologues in marine bacteria. *Environ. Microbiol.* **11**: 1348–1357.

Preston, G.M. (2007) Metropolitan Microbes: Type III Secretion in Multihost Symbionts. *Cell Host Microbe* **2**: 291–294.

Pukatzki, S., Ma, A.T., 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. U. S. A.* **103**: 1528–1533.

Russell, A.B., LeRoux, M., Hathazi, K., Agnello, D.M., Ishikawa, T., Wiggins, P.A., *et al*. (2013) Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature* **496**: 508–512.

Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014) Type VI secretion system effectors: poisons with a purpose. *Nat. Rev. Microbiol.* **12**: 137–148.

Rytkönen, A., Poh, J., Garmendia, J., Boyle, C., Thompson, A., Liu, M., *et al*. (2007) SseL, a *Salmonella* deubiquitinase required for macrophage killing and virulence. *Proc. Natl. Acad. Sci.* **104**: 3502–3507.

Sana, T.G., Hachani, A., Bucior, I., Soscia, C., Garvis, S., Termine, E., *et al*. (2012) The Second Type VI Secretion System of *Pseudomonas aeruginosa* Strain PAO1 Is Regulated by Quorum Sensing and Fur and Modulates Internalization in Epithelial Cells. *J. Biol. Chem.* **287**: 27095–27105.

Schikora, A., Carreri, A., Charpentier, E., and Hirt, H. (2008) The Dark Side of the Salad: *Salmonella typhimurium* Overcomes the Innate Immune Response of *Arabidopsis thaliana* and Shows an Endopathogenic Lifestyle. *PLoS ONE* **3**: e2279.

Schwarz, S., Hood, R.D., and Mougous, J.D. (2010) What is type VI secretion doing in all those bugs? *Trends Microbiol.* **18**: 531–537.

Shirron, N. and Yaron, S. (2011) Active Suppression of Early Immune Response in Tobacco by the Human Pathogen *Salmonella Typhimurium*. *PLoS ONE* **6**: e18855.

Sperandio, D., Decoin, V., Latour, X., Mijouin, L., Hillion, M., Feuilloley, M.G.J., *et al*. (2012) Virulence of the *Pseudomonas fluorescens* clinical strain MFN1032 towards *Dictyostelium discoideum* and macrophages in relation with type III secretion system. *BMC Microbiol.* **12**: 223.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., *et al*. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**: D561–D568.

Thorslund, S.E., Ermert, D., Fahlgren, A., Erttmann, S.F., Nilsson, K., Hosseinzadeh, A., *et al*. (2013) Role of YopK in *Yersinia pseudotuberculosis* Resistance against Polymorphonuclear Leukocyte Defense. *Infect. Immun.* **81**: 11–22.

Zhang, D., Souza, R.F. de, Anantharaman, V., Iyer, L.M., and Aravind, L. (2012) Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **7**: 1–76.

Table S2.1. Genomic dissimilarities between selected *Pseudomonas* T6SS loci and the associated genomes. The dinucleotide frequency and mol% G+C of each T6SS locus region was calculated using the software dr-web (van Passel *et al*., 2005). Rows in bold indicate that the regions differ significantly in mol% G+C and dinucleotide frequency compared to the genome signature.

T6SS cluster	Chromosome position	Locus	GC%	% genomic fragments with lower GC%	1000 x δ* (genomic dissimilarity)	% genomic fragments with lower δ*	Mobile elements
***P. aeruginosa* PAO1 (66.7% GC)**							
	1798463-1803626	**PA1651-1656**	**63**	**8**	**66**	**89**	**none**
cluster 1.1 (HSI-II)	1803626-1808789	PA1656-1660	64	12	43	49	none
cluster 1.1 (HSI-II)	1808789-1813970	PA1660-1663	67	55	33	25	none
cluster 1.1 (HSI-II)	1813970-1819142	PA1663-1669	66	30	33	26	none
cluster 1.1 (HSI-II)	1819142-1824314	PA1669-1671	68	73	50	66	none
	1824314-1829486	PA1671-1679	65	24	24	8	none
	77387-82404	PA0064-0070	68	64	38	37	none
cluster 3 (HSI-I)	**82404-87421**	**PA0070-0074**	**70**	**91**	**60**	**83**	**none**
cluster 3 (HSI-I)	87421-92438	PA0074-0077	71	98	53	72	none
cluster 3 (HSI-I)	92438-97455	PA0077-0079	71	97	53	73	none

cluster 3 (HSI-I)	97455-102474	PA0079-0084	68	61	41	43	none
cluster 3 (HSI-I)	102474-107489	PA0084-0089	69	82	43	46	none
cluster 3 (HSI-I)	107489-112506	PA0089-0091	68	71	65	88	none
cluster 3 (HSI-I)	112506-117524	PA0091-0095	66	33	59	81	none
	117524-122541	PA0095-0101	67	49	17	1	none
	2600732-2605937	PA2353-2359	69	84	30	17	none
cluster 4A (HSI-III)	2605937-2611142	PA2359-2361	68	72	83	97	none
cluster 4A (HSI-III)	2611142-2616347	PA2361-2366	68	75	55	76	none
cluster 4A (HSI-III)	2616347-2621552	PA2366-2371	70	96	51	68	none
cluster 4A (HSI-III)	2621552-2626757	PA2371-2374	64	12	57	79	none
	2626757-2631962	PA2374-2378	68	75	46	58	none
***P. fluorescens* Pf-5 (63.3% GC)**							
	6890249-6895202	PFL6070-6074	66	82	36	26	none
cluster 3	6895202-6900155	PFL6074-6078	67	94	27	10	none
cluster 3	6900155-6905108	PFL6078-6081	68	96	54	65	none

cluster 3	6905108-6910061	PFL6081-6083	68	97	56	69	none
cluster 3	6910061-6915014	PFL6083-6088	64	45	38	32	none
cluster 3	6915014-6919967	PFL6088-6092	66	84	65	82	none
cluster 3	6919967-6924920	PFL6092-6094	66	76	61	77	none
cluster 3	6924920-6929873	PFL6094-6096	63	32	37	30	none
	6929873-6934826	PFL6096-6105	60	35	49	56	Integration host factor (PFL6102)
***P. fluorescens* Pfl0-1 (60.5% GC)**							
	6237319-6242365	Pfl01_5570-5573	57	8	36	37	none
cluster 1.1	6242365-6247411	Pfl01_5573-5577	63	80	44	58	none
cluster 1.1	6247411-6252457	Pfl01_5577-5580	62	78	74	93	none
cluster 1.1	**6252457-6257503**	**Pfl01_5580-5584**	**64**	**96**	**90**	**97**	**none**
cluster 1.1	6257503-6262549	Pfl01_5584-5590	60	36	51	75	none
cluster 1.1	6262549-6267595	Pfl01_5590-5596	60	27	58	84	none
cluster 1.1	6267595-6272639	Pfl01_5596-5598	57	10	52	73	none
	6272639-6277683	Pfl01_5598-5605	61	41	37	41	Integration host factor (Pfl01_5600)

	3872850-3877953	Pfl01_3395-3400	60	35	29	18	none
cluster 4A	**3877953-3883056**	**Pfl01_3400-3403**	**64**	**93**	**68**	**91**	**none**
cluster 4A	3883056-3888159	Pfl01_3403-3408	62	74	48	67	none
cluster 4A	3888159-3893262	Pfl01_3408-3412	65	99	51	73	none
cluster 4A	3893262-3898368	Pfl01_3412-3416	56	7	45	62	none
	3898368-3903471	Pfl01_3416-3423	62	28	35	36	none
***P. putida* KT2440 (61.5% GC)**							
	3471896-3476877	**PP3084-3088**	**65**	**96**	**64**	**80**	**Excinuclease ABC subunit A (PP3087)**
cluster 4B	3476877-3481858	PP3088-3090	63	64	55	68	none
cluster 4B	3481858-3486839	PP3091-3093	65	90	60	76	none
cluster 4B	3486839-3491820	PP3093-3096	65	96	45	48	none
cluster 4B	3491820-3496800	PP3096-3100	62	49	89	95	none
	3496800-3501780	PP3100-3101	57	9	54	65	none
	2984362-2989308	PP2610-2613	57	10	57	70	none
cluster 1.2	2989308-2994254	PP2613-2617	60	23	22	3	none

cluster 1.2	2994254-2999200	PP2617-2623	61	34	47	49	none
cluster 1.2	2999200-3004146	PP2623-2627	60	20	34	22	none
cluster 1.2	3004146-3009091	PP2627-2628	61	27	41	38	none
	3009091-3014036	PP2628-2634	64	87	86	93	none
	4591831-4596179	PP4067-4071	62	39	50	53	none
cluster 1.2	4596179-4600527	PP4071-4072	61	32	44	39	none
cluster 1.2	4600527-4604875	PP4072-4077	59	14	49	51	none
cluster 1.2	4604875-4609223	PP4077-4080	62	41	40	32	none
cluster 1.2	4609223-4613571	PP4080-4084	59	17	26	6	none
	4613571-4617919	PP4084-4085	59	15	74	86	none
***P. putida* W619 (61.4%)**							
	3568589-3573595	PputW619_3235-3241	59	14	43	45	none
cluster 2	3573595-3578601	PputW619_3241-3246	58	14	42	42	none
cluster 2	3578601-3583607	PputW619_3246-3248	61	28	49	58	none
cluster 2	**3583607-3588613**	**PputW619_3248-3251**	**52**	**2**	**71**	**89**	**none**

cluster 2	**3588613-3593619**	**PputW619_3251-3254**	**51**	**1**	**75**	**91**	**none**
cluster 2	3593619-3598625	PputW619_3254-3257	57	8	44	46	none
cluster 2	3598625-3603632	PputW619_3257-3261	56	5	31	18	none
	3603632-3608639	PputW619_3261-3268	55	4	56	72	ISPsy14, transposition helper protein (PputW619_3262)
	2767266-2772239	**PputW619_2494-2497**	**67**	**99**	**67**	**86**	**Excinuclease ABC subunit A (PputW619_2496)**
cluster 4B	2772239-2777212	PputW619_2497-2500	63	73	61	79	none
cluster 4B	2777212-2782185	PputW619_2500-2503	64	89	80	93	none
cluster 4B	2782185-2787158	PputW619_2503-2506	65	96	51	62	none
cluster 4B	2787158-2792131	PputW619_2506-2510	62	55	62	94	none
cluster 4B	2792131-2797104	PputW619_2510-2515	57	9	49	58	none
	2797104-2802077	PputW619_2515-2520	56	6	40	38	none
***P. syringae* pv. *tomato* DC3000 (58.4% GC)**							
	6143471-6148512	**PSTPO5409-5412**	**54**	**9**	**62**	**85**	**ISPsy5 transposase (PSTPO5411)**
	6148512-6153553	**PSTPO5413-5415**	**49**	**1**	**95**	**98**	**none**
cluster 1.1	6153553-6158594	PSTPO5415-5418	59	54	34	29	none

cluster 1.1	6158594-6163635	PSTPO5418-5422	58	35	64	87	none
cluster 1.1	6163635-6168676	PSTPO5422-5425	59	47	73	93	none
cluster 1.1	**6168676-6173717**	**PSTPO5425-5428**	**54**	**9**	**60**	**83**	**none**
cluster 1.1	**6173717-6178758**	**PSTPO5428-5433**	**53**	**6**	**62**	**85**	**ISPssy, transposase (PSTPO5428)**
cluster 1.1	6178758-6183798	PSTPO5433-5436	57	25	46	60	none
cluster 1.1	6183798-6188839	PSTPO5436-5438	57	29	40	44	none
	6188839-6193880	**PSTPO5438-5442**	**48**	**1**	**97**	**99**	**ISPssy, transposase (PSTPO5440)**
	6193880-6198921	PSTPO5442-5447	55	16	38	41	ISPsy5 transposase (PSTPO5445)
	2796308-2801756	PSTPO2533-2538	56	16	42	50	none
cluster 1.1	2801756-2807204	PSTPO2538-2542	53	8	45	58	ISPssy, transposase (PSTPO2541)
cluster 1.1	2807204-2812652	PSTPO2538-2547	52	4	41	46	none
cluster 1.1	2812652-2818100	PSTPO2547-2552	56	21	42	52	none
cluster 1.1	2818100-2823548	PSTPO2552-2554	57	27	35	32	none
	2823548-2828996	PSTPO2554-2561	61	86	17	2	none

***P. syringae* pv. *syringae* B728a (59.2% GC)**

	5866000-5871000	Psyr4948-4952	59	35	26	13	none
cluster 4B	5871000-5876000	Psyr4952-4957	59	34	89	97	none
cluster 4B	5876000-5881000	Psyr4957-4960	60	62	56	78	none
cluster 4B	5881000-5886000	Psyr4960-4962	61	66	42	50	none
cluster 4B	5886000-5891000	Psyr4962-4966	59	44	58	81	none
	5891000-5896000	Psyr4966-4971	51	2	51	70	none

Table S2.2a. HSI-I- loci in *Pseudomonas* species. The *fha* gene in *P. aeruginosa* PA14 is split over two ORF. It is not known if this is still functional.

HSI-I		*P. aeruginosa*							*P. mendocina*	*P. fluorescens*			
		PAO1	PA14	PA7	LesB	PA2192	PA3719	PACS2	Pmen	PFF5	Pfl01	SBW25	F113
COG	**Genes**	PA	PA14	PSPA7	PLES	PA2G	PACG	PaerPA	Pmen	PFL	Pfl01	PFLU	PSF113
	tagQ	PA0070	00820	0141	00711	03982	03773	01000071	2309	6074	-	6006	+
COG1262	***tagR***	PA0071	00830	0142	00721	03983	03772	01000072	2310	6075	-	6007	+
COG4591	***tagS***	PA0072	00850	0143	00731	03984	03771	01000073	2311	6076	-	6008	+
	tagT	PA0073	00860	0144	00741	03985	03770	01000074	2312	6077	-	6009	+
COG0515	***ppkA***	PA0074	00875	0145	00751	03986	03769	01000075	2313	6078	-	6010	+
COG0631	***pppA***	PA0075	00890	0146	00761	03987	03768	01000076	2314	6079	-	6011	+
COG3913	***tagF***	PA0076	00900	0147	00771	03988	03767	01000077	2315	6080	-	6012	+

COG3523	***tssM1***	PA0077	00910	0148	00781	03989	03766	01000078	2316	6081	-	6013	+
COG3455	***tssL1***	PA0078	00925	0149	00791	03990	03765	01000079	2317	6082	-	6014	+
COG3522	***tssK1***	PA0079	00940	0150	00801	03991	03764	01000080	2318	6083	-	6015	+
COG3521	***tssJ1***	PA0080	00960	0151	00811	03992	03763	01000081	2319	6084	-	6016	+
COG3456	***fha1***	PA0081	00970/80	0152	00821	03993	03762	01000082	2320	6085	-	6017	+
COG3515	***tssA1***	PA0082	00990	0154	00831	03994	03761	01000083	2321	6086	-	6018	+
COG3516	***tssB1***	PA0083	01010	0155	00841	03995	03760	01000084	2322	6087	-	6019	+
COG3517	***tssC1***	PA0084	01020	0156	00851	03996	03759	01000085	2323	6088	-	6020	+
COG3157	***hcp1***	PA0085	01030	0157	00861	03997	03758	01000086	2324	6089	-	6021	+
									2325				+
									2326				+
COG4455	***tagJ1***	PA0086	01040	0158	00871	03998	03757	01000087	2327		-		
COG3518	***tssE1***	PA0087	01060	0159	00881	03999	03756	01000088	2328	6090	-	6022	+
COG3519	***tssF1***	PA0088	01070	0160	00891	04000	03755	01000089	2329	6091	-	6023	+
COG3520	***tssG1***	PA0089	01080	0161	00901	04001	03754	01000090	2330	6092	-	6024	+
COG0542	***clpV1***	PA0090	01100	0162	00911	04002	03753	01000091	2331	6093	-	6025	+
COG3501	***vgrG1***	PA0091	01110	0163	00921	04003	03752	01000092	2332	6094	-	6026	+
		PA0092	01120	0165	00931	04004	03751	01000093			-		
COG4104		PA0093	01140	0167	00941	04005	03750	01000094		6209, 6096	-		
PspB/Mog1 structure		PA0094	01150	0168	00951	04006	03749	01000095		6099, 6094	-		
COG3501	**vgrG**	PA0095	01160	0169	00961	?	03748	01000096	-	-			

Table S2.2b (part i). HSI-II- loci in *P. aeruginosa*, *P. mendocina* and *P. fluorescens* species. Gene loci containing decimal points represent genes which are not present in official annotations.

HSI-II		*P. aeruginosa*							*P. mendocina*	*P. fluorescens*		
		PAO1	PA14	PA7	LesB	PA2192	PA3719	PACS2	Pmen	Pfl01	SBW25	F113
COG	**Genes**	PAO1	PA14	PSPA7	PLES	PA2G	PACG	PaerPA	Pmen	Pfl01	SBW25	F113
			43100.1									
COG3209			Rhs									
			43090									
			43080							5597		
COG3157			Hcp									
	transposase											
COG3515	***tssA2***	PA1656	43050	3616	36711	00637	00625	01002119	0089	5596	3549	+
PFAM11319												
COG3516	***tssB2***	PA1657	43040	3615	36701	00638	00626	01002120	0090	5595	3548	+
COG3517	***tssC2***	PA1658	43030	3614	36691	00639	00627	01002121	0091	5594	3547	+
COG3518	***tssE2***	PA1659	43020	3613	36681	00640	00628	01002122	0092	5593	3546	+
COG4104	***PAAR***	PA1659.1	43010		36671.1	00640.1	00628.1		0093	5592		
									0094	5591		
										5590		
	transposase											
										5589		
										5588		
										5587		
COG3519	***tssF2***	PA1660	43000	3612	36671	00641	00629	01002123	0095	5586	-	+
COG3520	***tssG2***	PA1661	42990	3611	36661	00642	00630	01002124	0096	5585	-	+
COG0542	***clpV2***	PA1662	42980	3610	36651	00643	00631	01002125	0097	5584	-	+
	sfa2	PA1663	42970	3609	36641	00644	00632	01002126	0098	5583	-	+
	orfX	PA1664	42960	3608	36631	00645	00633	01002127	0099	5582	-	+

COG3456	***fha2***	PA1665	42950	3607	36621	00646	00634	01002128	0100	5581	-	+
COG3521	***tssJ2***	PA1666	42940	3606	36611	00647	00635	01002129	0101	5580	-	+
COG3522	***tssk2***	PA1667	42920	3605	36601	00648	00636	01002130	0102	5579	-	+
COG3455	***tssL2***	PA1668	42910	3604	36591	00649	00637	01002131	0103	5578	3545	+
COG3523	***tssM2***	PA1669	42900	3603	36581	00650	00638	01002132	0104	5577	3544	+
COG0631	***stp1***	PA1670	42890	3602	36571	00651	00639	01002133	0105	5576	3543	+
COG0515	***stk1***	PA1671	42880	3601	36561	00652	00640	01002134	0106	5575	3542	+
									0107	5574		
	glycoside hydrolase								0108			

Table S2.2b (part ii). HSI-II- loci in *P. syringae* and *P. putida* species. Gene loci containing decimal points represent genes which are not present in official annotations. Two similar full length glycoside hydrolase genes occur in close proximity in *P. putida* W619.

HSI-II		*P. syringae*							*P. putida*				
		PEL48	DC3000	DC3000	PsyrT1	Pst11528	NCPPB	PsyrFF5	W619	KT2440	KT2440	GB1	Ppf1
COG	**Genes**	PSEEN	PSPTO	PSPTO	PSPTOT1	PsyrptA	NCPPB	PsyrFF5	PputW619	PP	PP	PputGB1	Pput
COG3209												3234	
				5437									
			2537		2226								
			2538	5436	2225	32570				2614	4083		2117
COG3157		0539	2539	5435	2224	32565				2615	4082	3233	2118
	transposase		2541										
COG3515	***tssA2***	0522	+	5434	2223	32560	+	+	3246	2626	4072	3222	2129
PFAM11319										2625	4073	3223	2128

COG3516	***tssB2***	0523	2543	5433	2222	32555	+	+	3261	2624	4074	3224	2127
COG3517	***tssC2***	0524	2544	5432	2221	32550		+	3260	2623	4075	3225	2126
COG3518	***tssE2***	0525	2545	5431	2220	32545	+	+	3242	2622	4076	3226	2125
COG4104	***PAAR***			5430					3249, 3252	2610			
				5429.2									
				5429.1									
	transposase			5428									
				5427.2									
				5427.1									
COG3519	***tssF2***	0526	2546	5427	2219	32540	+	+	3245	2621	4077	3227	2124
COG3520	***tssG2***	0527	2547	5426	2218	32535	+	+	3244	2620	4078	3228	2123
COG0542	***clpV2***	0528	2548	5425	2217	32530	+		3255	2628			
	sfa2	0529	2549	5424	2216	32525	+	+					
	orfX	0530	2550	5423	-								
COG3456	***fha2***	0531	-	5422	-					2619		3229	2122
COG3521	***tssJ2***	0532	2551	5421	2215	32520	+	+	3243	2618	4079	3230	2121
COG3522	***tssk2***	0533	2552	5420	2214	32515	+	+		2617	4080	3231	2120
COG3455	***tssL2***	0534	2553	5419	2213	32510	+		3258	2616	4081	3232	2119
COG3523	***tssM2***	0535	2554	5418	2212	32495	+		3247	2627	4071	3221	2130
COG0631	***stp1***	0536	-	5417									
COG0515	***stk1***	0537	-	5416									
			2555		2211					2628			2131
		0540		5415					3254		4049		
	glycoside hydrolase			5413					3251,3253				

Table S2.2c (part i). HSI-III- loci in *P. aeruginosa* and *P. fluorescens* species. Gene loci containing decimal points represent genes which are not present in official annotations. ORFs corresponding to PSPA7_2897 are present in several *P. aeruginosa* and *P. fluorescens* strains but this ORF is only annotated as a gene in *P. aeruginosa* PA7. Due to the incomplete nature of the *P. fluorescens* F113 draft genome it is not always clear if genes are complete; such cases are noted with a question mark.

HSI-III		*P. aeruginosa*							*P. fluorescens*	
		PAO1	PA14	PA7	LesB	PA2192	PA3719	PACS2	Pfl01	F113
COG	**Genes**	PAO1	PA14	PSPA7	PLES	PA2G	PACG	PaerPA	Pfl01	F113
Phospholipase-D	***sfa3***	PA2359	34150	PSPA7_2903	29441	01465	01268	01002848		
COG3515	***tssA3***	PA2360	34140	PSPA7_2902	29431	01466	01269	01002849	3401	+
	hcp									
	ompA									
COG3913	***tagF***									
COG3523	***tssM3***	PA2361	34130	PSPA7_2901	29421	01467	01270	01002850	3402	+
COG3455	***tssL3***	PA2362	34110	PSPA7_2900	29411	01468	01271	01002851	3403	+
COG3522	***tssK3***	PA2363	34100	PSPA7_2899	29401	01469	01272	01002852	3404	+
		+	+	PSPA7_2897	+	+	+	+	3404.1	+
COG3521	***tssJ3***	PA2364	34080	PSPA7_2898	29391	01470	01273	01002853	3405	?
ADP-ribosylation										
COG3516	***tssB3***	PA2365	34070	PSPA7_2896	29381	01471	01274	01002854	3406	+
COG3517	***tssC3***	PA2366	34050	PSPA7_2895	29371	01472	01275	01002855	3407	+
COG3157	***hcp3***	PA2367	34030	PSPA7_2894	29361	01473	01276	01002856	3408	+
COG3518	***tssE3***	PA2368	34020	PSPA7_2893	29351	01474	01277	01002857	3409	+
COG3519	***tssF3***	PA2369	34010	PSPA7_2892	29341	01475	01278	01002858	3410	+
COG3520	***tssG3***	PA2370	34000	PSPA7_2891	29331	01476	01279	01002859	3411	+

COG0542	***clpV3***	PA2371	33990	PSPA7_2890	29321	01477	01280	01002860	3412	+
		PA2372								
									3413	
				PSPA7_2889					3414	
			33980		29311					?
			33970		29301					
						1477.1	1280.1	01002860.1		
						1477.2	01281.2	01002861		
PspB/Mog1 structure										
COG3501	***vgrG3***	PA2373	33960	PSPA7_2888	29291	01478	01282	01002862	3415	+
		PA2374	33940	PSPA7_2887	29281	01479	01283	01002863	3416	+

Table S2.2c (part ii). HSI-III- loci in *P. syringae* and *P. putida* species. Gene loci containing decimal points represent genes which are not present in official annotations.

HSI-III		*P. syringae*							*P. putida*			
		PsyrT1	Pst11528	Psyr1448A	2250	B728	NCPPB	PsyrFF5	W619	KT2440	GB1	Ppf1
COG	**Genes**	PSPTOT1	PsyrptA	PSPPH	2250	Psyr	PsyrpaN	PsyrFF5	PputW619	PP	PputGB1	Pput
Phospholipase-D				0117		4970						
				4090, 0118		4083, 4969		+				
		4104				4967						
COG3515	***tssA3***	4105	10395	0121	+	4966	+	+	2497	3088	2777	2635
COG3157	***hcp***	4106	10400	0122		4965			2498	3089	2776	2634
	ompA	4107	10405	0123		4964	+		2499	3090	2775	2633
COG3913	***tagF***	4108	10410	0124		4963		+	2500	3090.1	2774	2632
COG3523	***tssM3***	4109	10415	0125		4962			2501	3091	2773	2631
COG3455	***tssL3***		10420	0126		4961			2502	3092	2772	2630
COG3522	***tssK3***		10425	0127		4960			2503	3093	2771	2629
COG3521	***tssJ3***		10430	0128		4659			2504	3094	2770	2628
									2515			
									2514			
ADP-ribosylation		4099	10385, 10350			4977			2513	3101	2761	
			10340			4979			2512	PP_3100.2		2620
		4093	10335							3103	2762	
		4119	10465	0135	+	4952	+	+	2511	PP_3100.1	2763	2621
COG3516	***tssB3***	4118	10460	0134	+	4953	+	+	2510	3100	2764	2622
COG3517	***tssC3***	4117	10455	0133	+	4954	+	+	2509	3099	2765	2623
COG3157	***hcp3***											
COG3518	***tssE3***	4116	10450	0132		4955			2508	3098	2766	2624
COG3519	***tssF3***	4115	10445	0131		4956	+	+	2507	3097	2767	2625
COG3520	***tssG3***	4114	10440	0130	+	4957			2506	3096	2768	2626
COG0542	***clpV3***	4113	10435	0129		4958	5016	+	2505	3095?	2769	2627

		1769, 4091	10325	4984		2518	3106.1		2617
PspB/Mog1 structure		4090	10320	4985		2519	3107		2616
COG3501	***vgrG3***	4092		4974/ 4983	+	2517	3106	2760	2618

Table S2.2d. HSI-IV- locus in *P. putida***.**

HSI-IV	**Genes**	W619
COG		PputW619
COG3515	***tssA4***	3246
COG3516	***tssB4***	3261
COG3517	***tssC4***	3260
COG3518	***tssE4***	3242
COG4104	***PAAR***	3249, 3252
COG3519	***tssF4***	3245
COG3520	***tssG4***	3244
COG0542	***clpV4***	3255
COG3521	***tssJ4***	3243
COG3522	***tssk4***	3259
COG3455	***tssL4***	3258
COG3523	***tssM4***	3247
		3254
glycoside hydrolase		3251,3253

Table S2.3. Studies where >50% of the genes in one or more T6SS in *P. aeruginosa* are differentially regulated.

Comparison	**HSI-I (locus 3)**	**HSI-II (locus 1.1)**	**HSI-III (locus 4A)**	**References**
PA2663 mutant vs PAO1	<50	<50	**73**	(Attila *et al*., 2008)
T3SS and rhamnolipids mutant + cells vs PAO1 alone	**50**	<50	<50	(Chugani and Greenberg, 2007)
PAO1 cells vs PAO1	<50	<50	**80**	(Chugani and Greenberg, 2007)
PAO1 aerobic + nitrate vs no nitrate	<50	**69**	<50	(Filiatrault *et al*., 2005)
retS mutant vs PAK	**88**	<50	<50	(Goodman *et al*., 2004)
PA2206 mutant vs PAO1	**66**	<50	**53**	Haynes *et al*., unpublished
PA2206 mutant + overexpression PA2206 vs PAO1	**63**	<50	**60**	Haynes *et al*., unpublished
psrA mutant vs PAO1	<50	**81**	<50	(Kang *et al*., 2008)
PAO1 vs *qscR* mutant OD 0.8	<50	<50	**53**	(Lequette *et al*., 2006)
lasI rhlI mutant +3OC12-HSL	<50	**88**	<50	(Schuster *et al*., 2003)
lasI rhlI mutant +C4-HSL +3OC12-HSL	<50	**94**	<50	(Schuster *et al*., 2003)
Wild type vs *lasR rhlR* mutant	<50	**94**	<50	(Schuster *et al*., 2003)
OD 2.0 wt vs *rpoS* mutant	<50	**100**	**100**	(Schuster *et al*., 2004)
RSCV vs wt	<50	**81**	<50	(Starkey *et al*., 2009)
algT mucA mutant vs *mucA* mutant	**78**	**100**	**67**	(Tart *et al*., 2005)
Cu adapted	**53**	**56**	<50	(Teitzel *et al*., 2006)
mexEF mutant + overexpression MexT vs PAO1	**72**	<50	<50	(Tian *et al*., 2009)
mvaT mutant vs WT	<50	**56**	<50	(Vallet *et al*., 2004)
Biofilm anaerobic vs aerobic	**56**	<50	<50	(Waite and Curtis, 2009)
TSB+5mM EGTA / TSB+5mM CaCl2	**63**	**75**	<50	(Wolfgang *et al*., 2003)
LB+10mM NTA / LB+10mM NTA +20mM CaCl2	**75**	<50	<50	(Wolfgang *et al*., 2003)
WT (10% MPM 1) vs WT (M63)	<50	<50	**53**	(Wolfgang *et al*., 2004)
WT (10% MPM 2) vs WT (M63)	<50	<50	**93**	(Wolfgang *et al*., 2004)
D-cycloserine for 60 min	**69**	<50	<50	(Wood and Ohman, 2009)
PA2384 mutant vs PAO1 OD=2.1	<50	**94**	<50	(Zheng *et al*., 2007)

Table S2.4. Genes co-regulated with T6SS in *P. aeruginosa*

Locus Tag	Gene Name	Product Name	Locus 1	Locus 3		Locus 4A	Transmembrane helices	Export signal	PhoA screen
			PAO1	PAO1	Pf-5	PAO1			
PA0263	***hcpC***	**secreted protein Hcp**	**Y**						
PA1656	***tssA2***	**hypothetical protein**	**Y**						
PA1657	***iglA2***	**conserved hypothetical protein**	**Y**						
PA1658	***iglB2***	**conserved hypothetical protein**	**Y**						
PA1659	***tssE2***	**hypothetical protein**	**Y**						
PA1660	***tssF2***	**hypothetical protein**	**Y**						
PA1662	***clpV2***	**probable ClpA/B-type protease**	**Y**						
PA1663	***sfa2***	**probable transcriptional regulator**	**Y**						
PA1664	***orfX***	**hypothetical protein**	**Y**					**type II export signal**	
PA1665	***fha2***	**hypothetical protein**	**Y**						
PA1666	***tssJ2***	**hypothetical protein**	**Y**					**type II export signal**	
PA1667	***tssk2***	**hypothetical protein**	**Y**						
PA1668	***dotU2***	**hypothetical protein**	**Y**				**Y**		
PA1669	***icmF2***	**hypothetical protein**	**Y**				**Y**	**type I export signal**	**Y**
PA2066		hypothetical protein	Y						
PA2068		probable major facilitator superfamily (MFS) transporter	Y				Y		
PA2300	*chiC*	chitinase	Y						
PA2593		hypothetical protein	Y						
PA3294		**hypothetical protein**	**Y**						
PA3327		probable non-ribosomal peptide synthetase	Y						
PA3328		probable FAD-dependent monooxygenase	Y						
PA3329		hypothetical protein	Y						
PA3330		probable short chain dehydrogenase	Y						
PA3331		cytochrome P450	Y						
PA3332		conserved hypothetical protein	Y						
PA3333	*fabH2*	3-oxoacyl-[acyl-carrier-protein] synthase III	Y						
PA3334		probable acyl carrier protein	Y						

PA3335		hypothetical protein	Y					
PA3336		probable major facilitator superfamily (MFS) transporter	Y			Y	type I export signal	
PA3486	***vgrG***	**conserved hypothetical protein**	**Y**					
PA3487	***pldA***	**phospholipase D**	**Y**			**Y**		
PA4142		probable secretion protein	Y			Y		
PA4210	*phzA1*	probable phenazine biosynthesis protein	Y					
PA4211	*phzB1*	probable phenazine biosynthesis protein	Y					
PA4217	*phzS*	flavin-containing monooxygenase	Y					
PA4917		hypothetical protein	Y					
PA0070	***tagQ***	**hypothetical protein**		**Y**	**Y**	**Y**	**type II export signal**	**Y**
PA0071	***tagR***	**hypothetical protein**		**Y**			**type I export signal**	**Y**
PA0072	***tagS***	**hypothetical protein**		**Y**		**Y**		
PA0073	***tagT***	**probable ATP-binding component of ABC transporter**		**Y**	**Y**			
PA0074	***ppkA***	**serine/threonine protein kinase PpkA**		**Y**				
PA0075	***pppA***	**PppA**		**Y**				
PA0076	***tagF***	**hypothetical protein**		**Y**	**Y**			
PA0077	***icmF1***	**IcmF1**		**Y**		**Y**		
PA0078	***dotU1***	**hypothetical protein**		**Y**		**Y**		
PA0079	***tssK1***	**hypothetical protein**		**Y**				
PA0080	***tssJ1***	**hypothetical protein**		**Y**	**Y**		**type II export signal**	
PA0082	***tssA1***	**hypothetical protein**		**Y**	**Y**			
PA0083	***iglA1***	**conserved hypothetical protein**		**Y**	**Y**			
PA0084	***iglB1***	**conserved hypothetical protein**		**Y**	**Y**			
PA0085	***hcp1***	**Hcp1**		**Y**	**Y**			
PA0086	***tagJ1***	**hypothetical protein**		**Y**				
PA0087	***tssE1***	**hypothetical protein**		**Y**	**Y**			
PA0088	***tssF1***	**hypothetical protein**		**Y**	**Y**			
PA0089	***tssG1***	**hypothetical protein**		**Y**	**Y**			
PA0090	***clpV1***	**ClpV1**		**Y**				
PA0091	***vgrG1***	**VgrG1**		**Y**				
PA0092		hypothetical protein		Y				

PA0093		hypothetical protein	Y	Y		Y		
PA0094		hypothetical protein	Y					
PA0097		hypothetical protein	Y					
PA0126		hypothetical protein	Y				type II export signal	Y
PA0169		hypothetical protein	Y					
PA0170		hypothetical protein	Y					
PA0171		hypothetical protein	Y					
PA0277		conserved hypothetical protein	Y				type II export signal	
PA0563		conserved hypothetical protein	Y			Y		
PA0989		hypothetical protein	Y				type I export signal	
PA1069		hypothetical protein	Y					
PA1394		hypothetical protein	Y					
PA1395		hypothetical protein	Y				type II export signal	
PA1396		probable two-component sensor	Y			Y		
PA1791		hypothetical protein	Y			Y		
PA1844	***tse1***	**hypothetical protein**	**Y**					
PA1845	*tsi1*	hypothetical protein	Y				type I export signal	
PA2233	*pslC*	PslC	Y		Y		type II export signal	
PA2234	*pslD*	PslD	Y		Y		type II export signal	
PA2235	*pslE*	hypothetical protein	Y		Y	Y		
PA2236	*pslF*	hypothetical protein	Y		Y			
PA2237	*pslG*	probable glycosyl hydrolase	Y		Y	Y	type I export signal	
PA2464		hypothetical protein	Y	Y			type I export signal	
PA2536		probable phosphatidate cytidylyltransferase	Y	Y		Y		
PA2537		probable acyltransferase	Y	Y		Y		
PA2538		hypothetical protein	Y	Y		Y		
PA2539		conserved hypothetical protein	Y	Y		Y	type I export signal	
PA2540		conserved hypothetical protein	Y					
PA2581		hypothetical protein	Y				type II export signal	
PA2684	*rhs*	conserved hypothetical protein	Y					
PA2685	***vgrG4***	**conserved hypothetical protein**	**Y**					

PA2703	*tsi2*	hypothetical protein	Y				
PA2774		hypothetical protein	Y	Y	Y		
PA2775		hypothetical protein	Y	Y	Y		
PA2792		hypothetical protein	Y			type I export signal	Y
PA2793		hypothetical protein	Y			type II export signal	
PA3021		hypothetical protein	Y			type I export signal	
PA3060	*pelE*	PelE	Y		Y	type I export signal	
PA3061	*pelD*	PelD	Y		Y		
PA3062	*pelC*	PelC	Y			type II export signal	Y
PA3484	***tse3***	**hypothetical protein**	**Y**				
PA3485	*tsi3*	hypothetical protein	Y			type II export signal	
PA3619		hypothetical protein	Y			type I export signal	
PA3661		hypothetical protein	Y			type II export signal	
PA3716		hypothetical protein	Y			type II export signal	
PA3722		hypothetical protein	Y				
PA3727		hypothetical protein	Y	Y			
PA3728		hypothetical protein	Y				
PA3729		conserved hypothetical protein	Y	Y	Y		
PA3730		hypothetical protein	Y		Y		
PA3794		hypothetical protein	Y		Y		
PA3850		hypothetical protein	Y				
PA4033		hypothetical protein	Y		Y		
PA4317		hypothetical protein	Y		Y		
PA4318		hypothetical protein	Y		Y		
PA4320		hypothetical protein	Y		Y		
PA4487		conserved hypothetical protein	Y			type I export signal	
PA4488		conserved hypothetical protein	Y			type I export signal	
PA4489		conserved hypothetical protein	Y	Y		type I export signal	Y
PA4490		conserved hypothetical protein	Y	Y		type I export signal	
PA4491		conserved hypothetical protein	Y		Y		Y
PA4625		hypothetical protein	Y			type I export signal	

PA5033		hypothetical protein	Y			type I export signal	
PA5112	*estA*	esterase EstA	Y			type I export signal	Y
PA5113		hypothetical protein	Y		Y	type I export signal	Y
PA5114		hypothetical protein	Y		Y		
PA5136		hypothetical protein	Y			type I export signal	
PA5441		hypothetical protein	Y			type I export signal	Y
PA2231	*pslA*	PslA		Y	Y		
PA2232	*pslB*	PslB		Y			
PA2238	*pslH*	PslH		Y			
PA2239	*pslI*	PslI		Y			
PA2240	*pslJ*	PslJ		Y	Y	type I export signal	
PA2242	*pslL*	hypothetical protein		Y	Y		
PA2243	*pslM*	hypothetical protein		Y			
PA2244	*pslN*	hypothetical protein		Y			
PA2245	*pslO*	hypothetical protein		Y	Y		
PA2246	*bkdR*	transcriptional regulator BkdR		Y			
PA2247	*bkdA1*	2-oxoisovalerate dehydrogenase (alpha subunit)		Y			
PA2248	*bkdA2*	2-oxoisovalerate dehydrogenase (beta subunit)		Y			
PA2249	*bkdB*	branched-chain alpha-keto acid dehydrogenase		Y			
PA2250	*lpdV*	lipoamide dehydrogenase-Val		Y			
PA2265		gluconate dehydrogenase		Y			
PA2276		probable transcriptional regulator		Y			
PA2279	*arsC*	ArsC protein		Y			
PA2290	*gcd*	glucose dehydrogenase		Y	Y	type I export signal	
PA2303	*ambD*	AmbD		Y			
PA2304	*ambC*	AmbC		Y			
PA2305	*ambB*	AmbB		Y			
PA2306	*ambA*	AmbA		Y	Y		
PA2330		hypothetical protein		Y			
PA2331		hypothetical protein		Y			
PA2363	***tssK3***	**hypothetical protein**		**Y**	**Y**		

PA2365	***iglA3***	**conserved hypothetical protein**	**Y**			
PA2366	***iglB3***	**uricase PuuD**	**Y**			
PA2367	***hcp3***	**hypothetical protein**	**Y**			
PA2368	***tssE3***	**hypothetical protein**	**Y**			
PA2369	***tssF3***	**hypothetical protein**	**Y**			
PA2370	***tssG3***	**hypothetical protein**	**Y**			
PA2371	***clpV3***	**probable ClpA/B-type protease**	**Y**			
PA2372		**hypothetical protein**	**Y**			
PA2373	***vgrG3***	**conserved hypothetical protein**	**Y**			
PA2374		**hypothetical protein**	**Y**			
PA2375		hypothetical protein	Y	Y		
PA2378		probable aldehyde dehydrogenase	Y			
PA2380		hypothetical protein	Y			
PA2381		hypothetical protein	Y			
PA2383		probable transcriptional regulator	Y			
PA2384		hypothetical protein	Y			
PA2385	*pvdQ*	3-oxo-C12-homoserine lactone acylase PvdQ	Y		type I export signal	Y
PA2386	*pvdA*	L-ornithine N5-oxygenase	Y			
PA2393		probable dipeptidase precursor	Y	Y		
PA2394	*pvdN*	PvdN	Y		type I export signal	
PA2395	*pvdO*	PvdO	Y		type I export signal	Y
PA2396	*pvdF*	pyoverdine synthetase F	Y			
PA2397	*pvdE*	pyoverdine biosynthesis protein PvdE	Y	Y		
PA2398	*fpvA*	ferripyoverdine receptor	Y		type I export signal	
PA2399	*pvdD*	pyoverdine synthetase D	Y			
PA2400	*pvdJ*	PvdJ	Y			
PA2402		probable non-ribosomal peptide synthetase	Y			
PA2403		hypothetical protein	Y	Y		
PA2404		hypothetical protein	Y	Y		
PA2405		hypothetical protein	Y		type I export signal	
PA2406		hypothetical protein	Y	Y		

PA2407		probable adhesion protein	Y		type I export signal	
PA2408		probable ATP-binding component of ABC transporter	Y			
PA2409		probable permease of ABC transporter	Y	Y		
PA2410		hypothetical protein	Y		type I export signal	Y
PA2411		probable thioesterase	Y			
PA2412		conserved hypothetical protein	Y			
PA2413	*pvdH*	L-2,4-diaminobutyrate:2-ketoglutarate 4-aminotransferase, PvdH	Y			
PA2414		L-sorbosone dehydrogenase	Y		type II export signal	Y
PA2415		hypothetical protein	Y	Y		
PA2416	*treA*	periplasmic trehalase precursor	Y		type II export signal	
PA2417		probable transcriptional regulator	Y			
PA2418		hypothetical protein	Y			
PA2419		probable hydrolase	Y			
PA2423		hypothetical protein	Y			
PA2424	*pvdL*	PvdL	Y			
PA2425	*pvdG*	PvdG	Y			
PA2426	*pvdS*	sigma factor PvdS	Y			
PA2433		hypothetical protein	Y		type I export signal	Y
PA2443	*sdaA*	L-serine dehydratase	Y			
PA2444	*glyA2*	serine hydroxymethyltransferase	Y			
PA2445	*gcvP2*	glycine cleavage system protein P2	Y	Y		
PA2446	*gcvH2*	glycine cleavage system protein H2	Y			
PA2449		probable transcriptional regulator	Y			
PA2452		hypothetical protein	Y		type I export signal	Y
PA2453		hypothetical protein	Y		type II export signal	
PA2454		hypothetical protein	Y			
PA2455		hypothetical protein	Y			
PA2456		hypothetical protein	Y		type I export signal	
PA2491		probable oxidoreductase	Y			
PA2494	*mexF*	multidrug efflux transporter MexF	Y	Y		
PA2496		hypothetical protein	Y			

PA2501		hypothetical protein	Y	Y
PA2512	*antA*	anthranilate dioxygenase large subunit	Y	
PA2513	*antB*	anthranilate dioxygenase small subunit	Y	
PA2514	*antC*	anthranilate dioxygenase reductase	Y	

Supplementary data 3.1. List of bait sequences used in BLAST analyses of metagenomes.

>Tle1a

MLAVQLKRTVNINNDVANNQPSCKQDCRDVVNISVFFDGTGNNEDADKKEKKWSNPARLWRNARTH
SDKNEAKNDYAIYVSGVGTRFNAELNIFQRAISDFQDHYSLLGMGVGLGGARRLDYGEDQLNDALKQ
VLIFNAQKAEKDVKKYVGEKKNYSFAEVDKNLKDHRLIKKINISVFGFSRGAALARAFTNQFIGQCESN
CDGLTYGQGKYPIEFKFLGIFDTVASFGLPATNLSNSLPFLERDLVVDERVQNCSHYIAGNELRFAFPVD
VIHKDNKLANPNWKEVVYPGVHSDVGGGYEPDSQGVNDNFARIPLKHMLDDAVQAGVRMYSYEELQ
QNFKDLFKEQFEIQPDSQKYYDAVKAATPSQGSVQEQIKGCMKLYYSAYGTIARAGKELSVSERVRQE
NKFREYIPVGPSDMATEMQRLKKLKEATASKKDGFNIFRVFSPVSTAYEYMISIEDWQFESWNTNVSD
DIKKFYLNYVHDSKYGFLSNVEPFSYFRQRRVYESRRSASGEERDKKAAEQKVTCSAPKQEIPQTQFID
AYEQAQLTENFLNAS

>Tle1b

MSDRCIPCEKSKNWIELDFRDENNQSYEGFDVIIEDASGAIQTVNLTSGINHIEGIASGPVKVTIDTQTLID
VVEDRDKRLDSETSLVPEFAKEALGGPEQNQSKKYLHATLGDLWSNNPDYEFLEEHNKGKLGNVTFV
HNESYVIEVLDFTKQEIRFGVFFDGTGNNSFNADFGAMCEDEASVPDEVQDICDDLSEFSEKNRGSYDN
SVTNIGRLSKAYDSEKENIFSIYMEGVGTKALEEDQRFPDMGLGVGDRGVISISQLAGEKLVALIIENKI
NHNLSKFVFDIFGFSRGAATARHFANEIHKKVSGLVYEKLQLNSSSKVSINFIGLYDTVAAIGCIKDLLD
TTDENNHGINLYLPKDIANKVVHFTAHHECRDNFALNSVAPEHQEIALLGVHSDIGGGYKDPKDSVLI
MKPYLRTISEDDWHAAQNFNNKAIEKIEEMKQEFLPYVVSSSCFSQNNVVKTRLTNKHYRFRATLSMT
RDISPDLQLLSFNLMHKIATDNNVPLLTNTLIFDGRLAELYLCYEDTYNHSKNISGLSIIKEVPESLSSYV
LKHYVHCSDHWALDSGMYTMKPKINQYGERERQVWLHKKQKGYPF

>Tle1c

MPNFGFHIAPTHPVAGRLTYDSKKLSENILKQQSDERVFSRAQEQKRLSEGDVVGGAPCCKAIHITLGF
DGTNNNDKADGSSVSPSCSNVARLIHASIGSGDDINSRGIFKYYCPGVGTVFPDIKEFTPSNMGLIGAEG
GENRINWGLVQLVDALFYTLLKSRLKLNDVQGLVEEMSTNWTVSTLTGGLLENGEKKRRAALEPKLK
ELEEKLRQRQNSGQKPHILAMRLYIYGFSRGAAEARAFANWLQELTRVSDADGRVEYRFAGLPISIEFL
GLFDTVAAVGLADSAPFAAGHMDWADDTMRLPDEALSQCLPTILPEDCSFLKRCVHLVSCHEQRASFP
LDSIRRRDMDANGRRTGPSCYRKWTVEYAYPGVHSDVGGGYGVGNQGKAVGGSEFLLSQIALQHMY
AEAFEAGAPLQVPAPAVHPDFHEEWRVMVPKIEAEFSVSEELATRFNAWQAQAKAGPLEEVIRRETAL
ITAWRIDRYAGGLRNKAFFANVPPDMPEAQQKAWEALHKRRSREYAAAQQGEPLPPMSAAEQAEWD
RNVALIGGEDQLRDLRVEKQFDPPLDQRQLLGAAAEFAHDYKGDWGVLDDGMTVGGVIDLLLGGTV
FLINEEDEAEEYSQIHRDGSARYHQLFSAPDRVAPGQEKLVALFDEQVHDSRAWFMNTSAIGPREPFTD
YFRYRLVHFDNESNKRLSVLATAGRVVGVGVMLASVGLSVKRRDPRMLLGLFLPSLARPLLSGKVGL
PEISAFDPLTGIALPMVGGAALDNLRAFTREPGDKVEQIGQLPPPPPLAVAAVQSPALQQVLLAQQTVE
ALKARDLGSLAGLVAKAELTQAPAAATPAWLAEAKQALQDMGTEQAQPPGPGSAPGWLKRGKDLM
ESL

>Tle1d

MSTMKPDVVWYPPLFPENGRLPTQAALVGQNCKQQDSHELAYRNELCQAAGRLVEPPCCKTLHISLFF
DGTGNNLNNDLYIDVPKHPTNIARLFRATIGQGHAGGVQGHAGELVDMEGTSGNKYYKYYIPGVGTP
FPEVADLNYSTPGLVFATYGEERINWALIRIIDALRRTLGKPALSDNESWKAVKEMSFGALDQYRPIVF
RQLLKDLEPELKLALKQPEPGKPKLLGIKLYVYGFSRGAAAARAFVNWLVELLPNGFRKGTKPDLCLT
LEGANWKVPLSVEFLGLLDTVASVGVAHIAPLAEGHMGWADNAMELPDNGLIKNCVHLVSAYEQRL
CFPLDSICRHDGTYPSYAREMVYPGMHSDIGGGYMPGDQGKAVANNDGYLLSQIALHDIHAAAFFAG
APLKVLKDALPDDLKKDTWRIMPVELSHEFDIAPDMIQRFNAWRELTLNLPSSLEKINDENAARYEPVR
ASVNVITALEDQITWITAWRIGRYASGNYTKQRFYIDAAANGMDKDSDPKIREASEQARDIKQRKVEK
TRRKMQESAEKDTYIPYPPGPKDFDPALGQTQLRQAAEEFREDYQLKQRAVTGNWIYKGMDKLSNLV
YTFSDDDVSGEYIKIKKEANEKYSMLFPPQGESSNEQQPGGLVRALFDDQVHDSRAWFMQATLDAREP
WGSYFLYRMIYFGELMSKNITPLVKYGRFIEGRFKDTVPGVDFIIERAEKGGAIMGENSPPEIKQKISVV
NSSADIPLTMLSDLPLQPALTNEPAELIAWQQQSREAEKLASTKAAIASLWP

>Tle1e

MAELERLGPDPFALTLEEKANCMASYMEAVMGQCTREFRLSFFFDGTNNNKYRDAPRQAHSNVARLF
DIFDISPDQQAIYVPGIGTPFEKEIGDTGRGNHARAGLGAGWGGEARINWALLQITDVLYKRYYSDDLS
KVMGSTDIALVRQISSDLSLGLANVRTVGRDEVEQLANATNWEMLTSAAGTVLKPPRHEERRALLRER
REFLSQKLKALIAGRKPKMLRIRLSVFGFSRGAAEARVFANWLKDALDDDMTLAGVPVSFDFLGIFDT
VASVGVANSTKVATGHSGWGEEEFLRIPSYVKRTVHLVSAHEVRGSFPLDKAVADNCLELAYPGVHT
DVGGAYQPGDQGRGCNADGQPDDSNKLSQVTLAKMYREAVAAGVPLNPLAKGLTQRIKDSLKISPDL
IKAYNDYVDAVNPLIRKRGGGTTGAAQVQYGLYLRWRRLRLTGGAQAFESQPFFKRAQHYGAQCAN
DLSVANALLREEAKDLETRENDPAYSDGWMRRVLVVLPITQGTAMADRLKQTVWGDKVREWREVK
AYWNDTSPLDPRIVRICDDYIHDSRAWFKPFGAPSDSVWRMRQQARLEQLKQQDAAWKQVAADLNR
DLIGTLKRYESTANGGEGEVPPNPVVGQDRIDLDRYLKEGSVPMETKGREPSSMWGYLRWRTHFAPTP
TLGERAQAAWDEVAAVPGKVARKAKDVAHDAEARLANTARKLLDAGQESLERAAKDALQHFLGSG
GVPRL

>Tle1f

MSEIFQTHPVGPTPDILPVLPQCFLATEKCEIDLQIGIFFDGTGNNQDWDEADGCNPGLGKTQREARKDS
NVARLFRAYPEQPEWGISRVYVPGVGTPFPEIGENAYATFGPSMGRGGDARINFALLKLLDFLHKAVR
RKISPFVDEVILALCRNAKRRLIARTDTYSPVAGGTQDIQALDAVGMRTEGGLLCDFFGARPQAEAFYR
RQTQALAEKIANTSKPKLVDITIDVFGFSRGATQARVFCNWLKAMMDGNRLCGVPARIRFLGIFDTVAS
VGLPSSAGGDGHFSWATPDALRVPPSIQCVHYVAMHENRTSFPLDSLRGPDGALPENHKQFALPGTHS
DVGGGYLPGSQGRSPSGQPDDMLSQIPLELMYQAARTAQVPLDKTLTLDGQYDPFALHPNLRQAHAA
FMQAERPQATTARWLLSYLAWRYQIRDVYATELSWVTRARASTPNDFDDLTGANQTLLADIAVLEGE
SDKAMKFLRAATRLNPSALAGLEAASTVRRAALASEAQDVLTHIKAHPILGTPENPEHLTPQAYLFAN
YMHDSYAGFRPFDQPLYWAACIDPVPTSWESQGYLRFRRVYQGTGKPVTHAPPTEQQLLEAEQRERE
AQLQTIQSTIDMGKVFSGPW

>Tle2a

MDSFNYCVQCNPEENWLELEFRSENDEPIDGLLVTITNQSAPSNTYTQTTSSGKVLFGKIAAGEWRASV
SQASLLTEVEKYASRKEGQESPVKKRAAAELDAADKDTKQYRFTTIGDFWDEAPKDEFLQKQHKGID
VNASAEKAGFRLSHNQTYVFEIKALRSYMPVIIDTDEFNLVNSYTFALLSKLAYATNDFNRDDGKTIDN
QGAISTVISQLKRKERPTYSGDLQAKWLLEEIPYSKALSAQYYAEDDVGSEGYIIFNDELAIIGVRGTEP
YFQSKKPPVDNTKFKIIKAASGMAAVIADKIESATDSPGMKDLIITDLDAAQIAPEEFGGTYVHRGFYQY
TMALLSLMEKDLGLHKIKKFYCCGHSLGGAGALLISALIKDSYHPPVLRLYTYGMPRVGTRSFVERYQ
NILHYRHVNNHDLVPQIPTVWMNTDVSEGFHVLDVFKSRVDLMRKMLTDDDDDNYQHHGHLSQLLT
YNSNNQVLLTPKQTQVTMLDLANLATNDSVAMVDGLSDASIVEHGMEQYIPNLFEQLTALSDESLMV
HYQRAISALEQEIATLQQSYLTVKQAWIESIGNGTPTMNIGRLMSEMHSINKLIENRNKIRGELRQIVSDP
QRMPATKFLISQQTLPDEIKVQIR

>Tle2b

MVLIKTERSDRTNKIIVILNDGYARCSALNLFVRHRHDSLALKSSRCYVNSPLRTLRGEANSTVKTNFQS
NGYGYRHAVIGELCDKAPNIEKWEQARFGLPHYHFPKEDEFNGLKFMGEDFNQRHVIEVCPFRAWSLI
LNHTPDYDMVNAYNLGLMSLLVYKDETMVDPDEIEDMREFINTPDTTTSFFYQQCFDLSKVPVMTDG
YDYPAIVTDVPFKERYSPAIYLDLQQVENAPKFEHDTQMFYVENDTQIIAAWRGTASVRDALTDATYR
PIPCPKSILSTQNAKVHKGFLEAYQCVEKYFVQKIEKLKSSAGNKNKNLYICGHSLGGALALLHSSELR
DNNPLLYTYGSPRVFTGSGVNALASVNHFRHVNDADTVTSVPFDTNMDNWLFNVYGPLGTVFGFGW
TQATLPTIPLQKALPDIGEVYYHHGNTVSFFEARQIGERKIPSKGGRENDFVGKERWRMNGNYKFYLVP
DIANKLNGQLEQEQKNYIHAIKADRTLIDEIFPKGNNPSLDTIMTMATDHSMASKYQVIMNNRLLSMLE
PTKAPKLSASKQRFEELIHHKDAYPLNAQRNTTFIALENALKNEVKRSTGKNDILKQALERYKDEIHEI

>Tle2c

MHEKLSPQVAAELAALPYKFIRKAPISFDIISRNISNLFEVSLSDGAIKGTSGGLPFQLFNKTYGFAITAKG
KNDNNKDHHVIAIRGTKVMVPDWMSNLNIGFATGPNNKDVHAGFMKAFNSLIPSFEDYINKHNPKHV
YCVGHSLGGALATLTASWLQSKYGISTSIYTFGAPRVGNQSFAIQIKEFIPTYRVTHGMDPVPWIPLWPF
MHADDEYLLAGNKGARFSLGAHSMVEPTPGYINTAGSYKSYDDIANKISCLSARQRIKLDYKNRTQAR
FSAEWAQKISNALMSILVDAGYSISALLNVAGSCTMTLYDVLARTIHKIVEISKSKKEDVKGLLGHMLV
FAGQYHDLGSLTVALIRSVFEIMTRALYKMAGHALQKSI

>Tle2d

MTSFEKDLQSPLGSRMLTCPAGGKWTSFQLIDEFCSGEPYAGLAYIVTDSEGHKYTGRLDEAGKGKVN
NHFAGPVTLLFDQAYEGKEKLYSYLQGRPHYPLKITELQVRAEATHYLNPNATRTRERPEIADGGDYF
QVEVRHLVRHASHLPPEVYRDYPLDSGCAAIMREHGKLGVALMPQRHTILEVRPLRALRPILSTAPAFC
ALNLYQLALMATLSYCPFGQKPKVPCAETRSVKFLLQPSVGNWFGDALPKGEELWKVDSAQTKAYYP
LYEDVPYSARLEIIPFDPNLYAVNRHTPERDPEHPASVHFLDDIGSKDSTDTQAFITHNDELILIAVRGTA
EIVADGLRDADALQVPFAEGEGQVHRGFYEAAKKAAAFAVNYLEKFYTGQTLLICGHSLGGAITLLLA
EMLRRRPEGYKIQLYTYGAPRAGDADFVKGAADLVHHRMVNHNDPVPSVPGSWMNTKADIYGTGAA
LTFVNVPLGLSVFVAGITNWTGEAYDHHGRLRHAMPVEFGRQQVSSILWEPGCDTITQHAACDVAIRQ
RHGLPDRPTLLKQIFDAGHHSMTGAYIPACWAVLRRWQEAQESNRTLVTEREFALVESALQRITDQLR

RQRSNLPGRPDSYVRSRKNIVEALNHEIENIRTTRERLASLRHRRLTTTDVYGSLAEQPERLAESLPRWR
LHPENLATEQLAMAPEAAEDDPLLVTLYGHRIGAPHTFDIDSII

>Tle2e

MEIIGNNTCVTSLRPKYKAVQCRGEQHKYWLELQFVDENNTPVSGLNVRLEYHPLASAKDVEMWRKS
GYDYNPTPPPNPPAGVTDSQGVVRFDDLYWIAVDVNTDGQPLADEMEKRPLGLRRNPNSQPVSNNMF
RPETRDPKWRSDVQEKAEVAGYIHHYVTIGELCDQLPEIEGWAEPEPPKFHFPPGRSLKGTEIARDALD
KRHVIEVCPFRAWVLALHDTKEYDLANALNLSIMADLVYAAEVKNPTIDYFFRQKCQDLSCLPQLVEY
PNYFHMLAVDVPFRERYQPPLYMNTGEGALGEGDTRLFTVECTSHVLVSWCGTDSLLNVQTDVSFGP
KRCPAELGGMGDVHGGFLEAYQLPKRKFGDKLDAVKDSLGKRKTLFVCGHSLGGALALLYAAEMKA
FNPVLYTYGMPRTFSRLAAYLLRGIIHYRHANDNDTVTQIPPEVDLDSELYEKLGWLGDKLGFDWVTT
SAIGLLPRTFGGNAGVDMGITQERDPYWHHGKTVLFLRAEQSVMRSGERNALWIGGGTVSYTEKAAV
KLYLVPALNEECLTSTREHQAAFIQCLAPGELEKRFPKGENTASSGWFSKPSNHSMAHRYLPYIHNQVL
ELADPARDMDRKAMRAMFREGVEQKANFSNPDEVERNREFLALQDMLPVALRLTQAEETGKNALLR
FAAVTEEKVELSQ

>Tle2f

METRSLRVGREDKYSTVRKLAEQEEYEYQYITIGKLCDKLPTITPNWEDKKKPPAYHFPDTRFTGLTVK
KLNCRHVLEICPFRAWSLVLHHIPDYSVVNAYNLVLMSILSYADYDTNGDGQATGSISHFFYQQMLEL
SRVPYQINDKKFQPIVKDVPFHQRYTDVEFIDSSQGAEAKKDTQLFFAASSVEVIVAWRGTASLLDVVT
DVTYKPINLTCDIKIPCSGIVSNGLVHQGFLDAFEVIDNNDVESDFYLLSKIIEDKELYICGHSLGGALALI
HAAKLIDYNPILYTYGMPRTFTLSAMNELDGIPHYRHVNENDAVPSVPPERELDNHFYKLWGPLGYVF
GYSWSVIKTLIPFGNEVFCHHGEVVHFFKTESVLEWLEERNPRAKIKIRKRLSETAKLFLVPCLSSQSDE
NAKANQQEFIQKIDRESKDKLFPKHGNPELKNALSGGDHPSGKYAEYIGDRIMELYDTTNDGFYKKEQ
DVFVFQMEKYKNDIPQSEYNRNKSFLAVDKLLVTTLNVTKSLPESDIAMQRYCLKKDGL

>Tle2g

MNKMQGCVDCNNFVYSFDIELVDELEQPISDIDYEILLERNNLVLAKGKTDAQGRISVEGLPSLPLRLV
LDTAMFLEEMQEPQRHLRLGRTSADSTVKPRAEQQGREYSYATVGQLIKELPTIPGWPEDRDLPAFHFP
DQAPKGWSIRPRGKEHQHRKLTIEVCPFRAWVLMLHNGPEYSLVNGYNLALMSVLAYAGAERDDHG
SKAGALKPYVEQVLTHLGRLPFKVNDQYLAPVVKDVPFSERYRHYNFIDTNNDKMGKVGHTQLLYLV
SDEHAVVAWRGTEHGKDWLTDIYGIQGNANPILPSGKIHLGFKNAYQELSANKNLDDAKKIMEQHLL
GKPLFICGHSLGGALALIHAAEHRAEKPQLYTYGMPRVFDRTAVAHLGDVVHYRHINNNDLITAVPHP
KWSTVRIQVALSTGLLPYPTPTKYSPEVRRNLDSQDYQHHGNIIHFTNFSMDYDTGPGYQNMGWMAS
YGKTIKSLQIKSILAPTLIGANSQKAYELLIQQYSTEVLDNPDRKEVNPFDHPSSLYARYLGDRLYDRLC
KDQEVQSHFEVQAKKIKQAAYSSVDKNKPAQLDALLQADDLVLDQSLLPIASQPIETNGLKRCYLYAM
QLPKEAIRSEQAEVAKETQQLKEQVNNDANYKMHRNASYLDPGDRATLPKLDDSVKDSIDLVDKQAK
GMKKMASLMQ

>Tle3a

MRAPKFKHFAPKVDALQEVPLRYESNPFTLELGRLIDGPENEGKPVNRVLNPLPCLVIFVHGVNSEGEW
YNAAEQHICAGLNDRLERWGTQVELISSVPEDGDPMYTYDIKHPLKRTRPHFEGGQNAANSNYLFAKS
PVIRFYWGFKARKGDYGRPSAHDYRDDHHPELRDYPVALDEDDAWGGGPFQNGTSGLWHMFRKDK
GFTLDFQHVNPITDRYLTECPPRTYYVHAARRLAHLIATVRRNSPHETINIVSHSQGTMVSTLATLMLK
EQGVRGPEALFVCNSPFSLNEPGGLMECAQFGNDYVVTQAARFATLKAVADVVKEAGEHADAITEAN
QQAQCLCAPARASLLNREEFGKFYVYANPHDRVMGASVLRSIGWRALTQAEHGRLGASNLLVRMFAE
NIEAGKGGYHYAACPKSFAITKDDDGKEIHMERAEFWIPKSEKIGGVYGVYSPPDKAEEAVHINAPTVP
WLTDQALRRHMEQAKNTKGKPAWESLAGWDGRLKDFHEYRGYLKPEKNGGDVRDVAAYGQVYGP
NYEVVRMEVDVYGKRYPVYKTTAEAQQDLIKTAHNLTNHSTLPSNEFVVRGVMAWDLAMGLNQSY
TDVAYWTYLKALADWKCSDPYFLENDEGHQPEPGEPPPGVDRTLAYPQREPSVASAEPAVTGYACPQ
TGWWQSNEAGAIASERRQFIREGQIMPKVAIAGQPSLWQRIKGERPTWSTATVWTLASNESAIAGSGA
GGVVS

>Tle3b

MSDTKTPSPRIVGEHQAILNCNRPQNAFVNTDVQNRFPCTTILIHGVNDLGTDFGTVEGGLCEGLNDRL
GRTDFKGAEYSHGRMANDPSQVSVADMMKNMDDVIYRRQESEETRSPLIPFYWGLRVGKEDLPRDPN
KETVNGQYVDRFGNRLDEHRARNGGFFANATNNIPDMFDSNFKGGMATKILDRMQGDPTHPLREAEN
RHYMVLAARRLAALVRQIRLIDPDGTVNIIAHSQGTLISLLAQAYLVEGLIPNQCGPADRPADTLVLIDS
PYSLSEEFMDRLLQRGDQQQTTYARAKTLANLVGHVASGKYPTPPFDQLKYEPGRDNFGVVGPTWDP
AQARRLSGPQGNDCVVFAERDNRGKVYMYFSPEDATVGLSGVNGMGCSGLPDSIDVRAAQQGSKPET
INLLSAAFRQRVFTRRLRQGKPVQVGTPPGNFTIREKGETSHGLPGGFTTWIKSAQISVGTQRNINGEAL
TPPFDPEMEGNVLPGTAATPLSKNNDGEFAPGKQAIDQLEAEIALSTNSGEGALQALPAQTIDWPTSED
GTLPTSTEVEASLNAGKGPDDQCKVSRVVSTVPSSPGHIVVYRQETLNEAKVRLMNQHKGESSYHSAV
MSGRRNHRCATAFDVSVGQARALDDPDWATLLRAVADWRTPMSRIDKLTKAHTNLDEQTLHIIRANC
EYYDQGNFPGEDVIPKAFPAGVVSETIEMRTDEINKQVRARSPHHMHG

>Tle3c

MQDAASDGGGLDPRTPSGETSSGGSDDSQNRYLVANGNGVMTLSQSDYLCVMQMPLPGIVIFVHGVN
SEGEWFEASEKGLCAGLNTRLGRTDDDLAHHGPQAGQMAPMHYMNSLTDDGFINPQLTSKTYLQPDE
NTYSPVIHFRWGYKANAEELKRYGESIFLNEQNYWGGGPFANGCGSLPDLWNAGVDPRIFGFINVQGI
NPTQRPLFETPPRMYNVLGALRLAKLIGSIRKAQKDVPITVVCHSQGNMVGIAAAFLGDRLSKDAGVR
CVADTYVLANPPYSLLGTIKGDSSPFMEQWTERNTKDSHGNRGRVESQARVKTLKAFLELIAKSKDAE
DERDPLDFRMTMANLQSRADGKPAYDPDYDRQQYGVQNHTYGRVTLYSCPHDQVISAAPVQGMGW
RGLHPDEVALVKGSSLFTQRVFATKFKVGQASPVTYRTWEDDWRYGKGSTDLFFYPTSPPAKYNLVK
SLRGNQEWYGKVATIAIAPALYIVTMATSALGMFRVSAEPPKKCGTTQEGWRVVCDAPALPEPFFPQA
LRYGALVHVKDGTAESDFNEDNDPVAAARNKNKTSINTADPYDTYKPKGKHAAEMEPKGDANSEAS
QRYEDHAIARMEARRDNNPEWVKDGVVIGEAGPNAQMPAGYEKWRNKQIVHILDAAVNNPTNHSTI
MTNEKHARLALAYDVAIGVCRISQQRLNDFRIEADWRFGHQIGKYNPNIVYSEYFRYGTMKKVSLTN
WVTNDAEATLPDSIDDTRQGGTFLGIGEVA

>Tle3d

MAGEKKKPYCPRIVSENDVGVHKSLGLEAQTCRVGIPRPMPGIVILVHGVNDVGEAYQNQEKGIIAGL
KKRLNRPDFYAHEWKDYRITVPGRSPIIPFYWGYKPVTREEYMGDQQRYRNEVSKLQDSAHLPFDAY
QEDNADKKAELGNDGKGRFKYQNDNFKNALDVNFAKGGGTFANATTNIPDMLGPGAGGVALAAAG
FSTLYMNGGDFTHPIFPNPHRIYQFFAAQRLADLILTIRRDLTTANDVINIVAHSQGTIITMLANMLVKQ
AGYDPVNCVILNHSPYSLEGRLSEDIQPGHHQTDNARQTTFKNFCALMATQYKGGEISEANLLAMEGA
CTLRKASDSPLRQDEKYRRDNNGKVYNYFCPNDGTVSLKNIQGIGWRGIPRNIASAIPNLYQRVFYQH
GQTGIAPTGKPFELPPSHKGDAEYSSLTNTAYTAHDVIVNGEELPQPFVFKLQGQDNHPDNDPKTSDKP
YTAYIDPDSPDAYISYSAKAYAIKRTRSATYSLSRYASLSWRPGHILTADELKMESYERKVEVIHGEVIG
TKDFSQLTLTWLKPREELENEWQKADPVGYSQHSSIVMSEFAPSHAMAFDLAIGPCRSFDYKAGKFWE
DLLHRADWRDPLNGNIAAKEYYRTGRLPGAQTKNHMNKPDEVLPHGNYGVLNEFNNATRVRPSRDL
AAGNQEVANLQWEMPKPKSDRELGIGNQRSALDI

>Tle3e

MSEMIHKNIRNPKFLSEDDIGRPVSAGAMTIQNCAVMRPLPCIVILVHGVNDVGEAYQDQDTGICAGLN
TRLGRQDIHPHEWKTHEFMISDADDNVAKLTCSVQDHTCIGVANRSPVIPFYWGYKPVDHDTWEQDQ
KQYRKDLLEKHHNADLPYDTYRENDAVKIRQHDNQNIDNLNNWLDLTSAKGGGTFANATTCIPDMFG
PGASGRILEAVGLFKSRSDLENHGDWSHPIYQNPHRIYQAYAARRLADLILDIRQNPLTQNDTINIVAHS
QGTIITMLANMWVQAEGAAPADCVILNHSPYALENRWLENAMPGNQQTRQGRQKTLSNFCRLMSQN
ALYQPGAISHDSAYLQQLHDIGCISLNKSPWNDPLYNRNNFGKVYNYFCPNDQVVSMSPIQGFGWRGI
PDKIKAQSGDNLYQRVFCKDVVVGDKTAFHFEMPARQPDDSADTGYSFTDVTVNAPVLPDPFVFKLM
AQGKGYKAALSGNDPKIAKAAMKAERFPAECIDVPACAQFQYLDNGVALNEAQLEELRSQYPWDIVS
GIYHELNGYFRQLVLLRRMTDKELDRAIKMDTTFSQHSSIVTNIEVAQKSTAFDLAIGQCQAFEYQEFW
EGLLLKADWRREMNPDPSVKKYYLDGVLPKAVKALMNKPEREKGMPMGDFGVLNDYGARVRPNSP
GTNLTDVLQWEMPAPQI

>Tle4a

MACNTECGCKPAVIMLVFIPGIMGSRLKNTETGDSIWDPTAGGGFHESSAMASAKEKRKQELVQMQSL
DDDSVFEGLSKWFSRNVDFKVSFAYATEIVPQVARKLWRIPRIKDWVFSGPKGRKKLISNEDEKDAYG
RQEDLTSVDEGTEAYFKMYTSVSSQQIAEKTFRGWGSVLWASYGGFLKWLQGCEWDFKADYPGSSFP
VYAVGYNWMLSNRVAGDRIKDKIHGFVAEAVSEGADKNNIKIILITHSMGGFAARSAVILSGLKDEIEL
VIHGAMPTDGAPQLYQEFRTGSTGASKLVLGKNAADTTAVAGFCQGGLELLPNHAYKMKSDSEMKC
GKASHEWLYYNNGVGANKLLPVEYGDKIYDFYRHFDRWYSMVQPALLAPEMKAEELTKDKLDFYTE
EFNKRIDDCKEYHKALAGDYFPTTIVLYSTSDKSHDCSVWQGGNNLDEQPVESWKTISNENNDWFFGD
GTVTLMASDAEIKRARQIKQRQREALVSHQFMPPMTVAVEKYTLGAKLVDGDGTVNAGAGEGLPIDK
VRAIAIEATEEHQDFYNCPKVRTLVVDQIKAALPDIHDKITRC

>Tle4b

MSGFEPTRLVGTTWDDDGNDVAQSVLTGQNMKVRALCLTTPDVVVPILFVPGIMGTRLKVTGRDKGP
AWFPPESTWESLVLAFKHLVRTPADRQRLLNPDTTEVDEDGPASPDDTSKTLLSLAPGNTDAERIKWR

GWGQLHADSYLGILSMLETSMAMIFDPASQGTLLTSRWKELVMDRQDAAKLGAQKPFVPLEEDHLRD
AAELLYPVHAVGYNWLQSNKVSGQRLAGEIERITAYYRSKGKSCEQVILVTHSMGGLVARACAQLPG
MAERILGVVHGVMPAIGAPATYKRIRAGFEGMAQVVLGRNAADCTAVMANAPGPLELLPTAQYKTW
TNQGERHWLRTSYRAIGQRGMPEEMDSFLGEGDPYAEIYLNNTLDWWKLVREELIDPAGKEDKERAR
SENKVLAREDAKSSDFEQFAKRMDKALTLHQQIEDQYHPNTYAYYAADPQQPAWNEINWKCGPMVP
GDPAKAQLAEDDLNGMLELRFGEHHRRFFTLQDGTGPGDGTVPAESGGAPKPHVVQLFKHEGKQKSH
DSYDHQFSYNAKIAQAVTLYSIIRIVSSSTHFKTPSGDKCV

>Tle4c

MIFIPGVMGSLLINRNTGQEIWNPPNETNRAAIAFGWNGYFSDASERETKYDTSSAVVSTLGAIVTTGC
DLTADEARRRGWGSVHSKSYHGSLAWLERELNHPMKAGKLLGAWASGDPKGEDFTLNPVIGTAPSSY
GANGPGIDPLKVDSDEFKKFCQYRYPVYAIGYNWLQSNLLSAEDAVKGVDYVDRKTHKKYHLMGIKE
ICAENKVKKAIIITHSMGGLVARMASVIAGYDDLMYGVIHGAQPATGAPLAARRFRAGAEGEGAILNK
VLVGRNAPEFTAITVNAPGPLELLPMPDYHDYEPWWVVRDMKRQEVLALPKKNDTLKLFTSQAWYG
LVPNQSDAMLDPAGIAKQRLQTTQPGMTLQKNFLTTMQDVVDRQHQLINNYHPNTFAFYGNGPLRRS
SEVDPTKGGTAKAPTKQETTAPADKLLTFGRIYWRGDFPAGTTEEDLMKAELLSDDQKGQVRIRVKGL
VCTLFAEIMDPNFPQERGLIFGDGTVPSWSGESVARGLKAGVPGEKAKGVQIAFTQNGFDHQGCFGHP
WAKWATLYSMVKVAKSIDVPV

>Tle4d

MSTTMHMDRADKNKVGTAEIEDRRVIPVIFLPGIMGTNLKENSEKKNAKAVWRYDSETSLLGWTLPSS
GPKERKRLLHPDRVKIDNRGLIPMSPDVQKHIDLVIATQYPSNIADKEEMERFTKALLKAEEEDHPEKK
LFGSRRDRGWGEVAAPSYGAFLDALQTALYRDKPTKKDNTLSAIYQKLVDEPLGLEFGVDSLDETYLE
VMKLYQFPVHVVGYNWLGSNILSAKRLGEEVDRIIDKYQTQGMKCSKVILVTHSMGGLVARYYSECL
EGRDKIYGIVHGVMPSVGAAATYTRMKRGTENTESGLAGAVGYLTSHILGRNAAEMVAIGSQSSGALE
LLPANDYGDDWLKVVDRDGSTLTLPADLPVGEDGKKENCLYSGLYLNREKWWKLMDENLLNPFNTT
LNHTQIEDDWKIYENLIMTCIKPFHENIFNKYHPHTFSFYGKVKDKEIPEAHRTQETAVWSGALASGQIG
ETPLAPKMNDERLYLSEVSAVRTVKDELSPEEQAWAFKGDNDVVYGWVGQRFTLRNACENGDGTVP
LRAGQIVHENIRERLAVQVLHEAAYRDTLSQAFALRSIIKIAQQVNYDEKMVFA

>Tle4e

MSSEPLEPNQDVIIPRSRDSLGRPVYKAQLTRTDNQSEKVALIRQTAPLPVIFIPGIMGTNLRNKADKSEV
WRPPNGLWPMDDLFASIGALWTWAWRGPKARQELLKAEQVEVDDQGTIDVGQSGLSEEAARLRGW
GKVMRSAYNPVMGLMERRLDNIVSRRELQAWWNDEALSPPGDQGEEQGKVGPIDEEELLRASRYQFD
VWCAGYNWLQSNRQSALDVRDYIENTVLPFYQKECGLDPEQMRRMKVILVTHSMGGLVARALTQLH
GYERVLGVVHGVQPATGSSTIYHHMRCGYEGIAQVVLGRNAGEVTAIVANSAGALELAPSAEYREGR
PWLFLCDAQGQVLKDIDGKPRAYPQNQDPYEEIYKNTTWYGLVPEQNSQYLDMSDKKEGLRVGPRD
NFEDLIDSIANFHGELSAAGYHSETYAHYGADDSRHSWRDLIWKGDPTPLETPGATLNDDENGTYNSW
FRRGLPTIVQGPLETGNPLDASGSGGDETVPTDSGQAPALAGVKASFRHGSKGKGQANTKRGYEHQES
YNDARAQWAALYGVIKITQLADWHPNDKGGT

>Tle4f

MADTSQMPRIAPLSYNARGNPVHTWTLTPSHITDPVHCILPPDGVLPVIFVPGIMGSNLKSKPEKGKGK
NKAAPVWRLDAGFMGKNMWLALNWINKKAGIRQQLLHPARVEVDNKGAVPERSVGTVIVPPEPDRK
KTILALTKRYEERGWGEVSETSYHAFLLWLEDTLNSEFMPHKWPQFDIRPEHLHTRSVEPGPVQVTELK
PEMPIAMPGLGANLAAQLPSIISDELAARGGYRMPVHACGYNWLDSNEEAASRLAVRIDELMQQYGR
NCQQVILVTHSMGGLVARRCGQLPGMADKIAGVVHGVMPAIGAPVAYRRCKVGMRDEDPIAGAVIGP
TGQEVTAVFAQAPGALQLLPTQDYAPGWLRLIDERGAPAMPRQPVKDPYEEIYLRRDRWWGLLREEW
LAPKDGDPITWETFAINIKKASSFHQRIAGSYHPQTYVYYGNDDKYPSFETITWEMRRGSRLNGPYASS
PDAFTVSSLQMPEVRDDGRSPLYVGGQAESVMPPRGDPNAPVKTVQTSYWELHCRMQDGAGDGTVP
VSSGRAPIRQIRQGSVRQQVQAPGFDHEASYANPLTQQFTLYSLIKIAAKAKRPLCAG

>Tle4g

MVNKSIAQSGSATSPTTNKAGTYVEVPLQDTIPIIFVPGIMGSNIFNTALNKPVWKLGNGGGMVGTIYS
QMQKSPATLQNELDPLNTRVDTSGDIKVDSRLKLTEKTLRERYWGTVHWDSYGGILTYLQMVLNNVD
LNEKPIYGGGVGGGYAVIHQMKQQEAVYEWKSLLNQGESKKWTPQEPFISINQQEIEHLKKFHFPVYA
MGYNWLQSSENGAATVAQKLDKIKQEYGTRFHKFIIVTHSMGGLLTRRLVQLRGGDIAGVVHGVMPA
EGAAAAYRRLVAGSSEGGLVSAVAANVIGKNTEHVTAVLANAPGGLELLPSKAYNNGRPWLFLNGSG
LMNGKDLQTNVVSLPKRDPYEEIYKADGVWWEMVKEELVDPANMVKKSNPNKSVKNIYKLKINEVK
NFHDKITNKYHPCTYVNYGHDPKHSSFGTLTWTLDRPLRGLTAEQMKKLPRATAKEIGTYRQKIVQEQ
MKKIKEGKGADANNRDLALENNGIRYISLVSGNLGVFSISNQNAPGDGTVPYQSGCAPLKQAGVKQVF
KMTGFDHQGSYNNTQVKRSVLYSIVKIIKENNIQPKYR

>Tle5a

MGAPDKQVYQIVGGDQKAAKTATSAPGCFVHGKDVFSEPVKGNAVQFFVTGTEYFDNVARAIEGAQS
SVFITGWQVNFDVVLTGKKTLWNCLRTAVRNGASVYVMPWMSPKVGVDTGDLETALTVIQLNAGLP
SPRAFVLPAVSQCDQPGALGIAFSHHQKLVVIDNKFAYVGGIDLAYGRRDDGKYSLKAEGRQGSEFYN
SCVPAIHSLSSVEQTAYLTRAELVAACFDNKAGRAAQFFLSAPMKPLAGAMDAYSSASDKIKDVNKQI
SDWWVTSDVVPEFVRKAQDKVIDAAQETAADASKWAYQQLGTTLQTKVEKLREYGGAQVADATTA
LMAWLNGATLDSLPPSLLQNTADTIQAFVMRLVLALQAEGSQRKQCYANLEKLGKLLPAGGKCPDSS
VQPRMPWHDVHCRIEGPSVYDLSRNFVRRWNGVALQYERSDGKTVDALLRHLGIAARLKAPRIGAAH
RPVRSKAQPGSCWVQVLRSAPKKMRVAEAAGEADKTTPSVAESSCLSAMLKNIEGASHFIYIEGQFFQS
DYGSTMIGNEEADGGAPVSGPMHALMDVKGSPGYQKYAAQLGILGVPPGQIYKSLKWSQIDDVQRDI
RGGGADFVNDLKRVMATQAQIAGFSALGPSQKSLKNPICKALGDRITRAIYDGKPFHVYMVLPVHPEG
TLDTINIMTQQHLTMQSLVFGSHSLVNRIRRALLAMKYVRERKMEVKRARDAAESAKIEDLNRNIRQG
EWQKYLTLLNLRNWDVLHGRPVTEQIYVHSKLLIADDRVVVLGSANINDRSQLGDRDSELAVVIHDD
KAMSVRLDGQLVQPVGTFAHGLRKALWKKHFGLMGGVAPATELGLPSILDGPGDLDTWQAVQRVAQ
TNAEAYDKAFPYVPRSKSRSSVWPTWDDSNGKLRGYMPFNERFWREERPRDDSFTWDAKQILSESTPS
GVRGFIVALPVMWTLGENNDSKMNLTTLAHVESEDEVNRERQTASRSPSQSGQADAYS

>Tle5b

MANINKSITTPIAQNHTSSAMACLPWFVCNSEYSPSAAGFRPLVNGRDAFAAVYKAIADAKHSIDIICW
GFQPSMYFLRDGTGKRIGELLEQKGKEGVKVRLLCWRDPFYMAELSENNMPGNDFITPLKRALPKWV
YDHVSMLGRDYQTDEERVFDINWYQRANLNNVTNPVSHSVTDRIIEDTVNLPLNVVAALSKPFTSGSN
PVLHKIAGFPNIELATRGFDLSVRSEIVGRISGYGTMDGWSGGTKATTSAAMGGEPSHHQKMVLVDYD
DSDLATGFVMGHNMLDQYWDTDKHSYQAMAPSKGRNGPYPWQDISSRVTGPILQYLNKNFCEAWD
DATGQKLAKSREHLAGIHKIRNDAPDDVSVMAQIVRTQSEKNKREIQVMYLNAVNNATQCIFIQNQYF
RWNDLALKIKDVAKAHAKAGRDFGRDGSIYLFVITNSSDDAVGNGTVRTYQMLESLGCGNSMPGVA
QLQRSDDYAAQQKDLEAQLAAQQSIQHKFAAQGVDINAASLQGAMNFYQDNMAKQQELQQALNEL
KRKKQLNAHIDPDAPPLDIEGLKVQVCTLVAPDSPGGNWLPVYVHAKLMTIDDAFMTIGSANVNLRS
MNVDSELNICHENSVVTKSLRQRLWSMHAGDKGAQEDFFKSFMAWHDIVEQNKVNKTKNLPPEKSL
VKFVRTSAKRTYAD

>Tle5c

MADSILKSQNPTSVAIDEQSRAAKSSAQWLLEKRKDIAPITYENHLEFLVCGEEGFANIANDLKLAKGS
VNLCCWGFDPGMELIREKGETKWPRGQTYGQLLEEVANKGVTVRLMVWYDETASAKQNNVPGYSD
TGHNPITHLPAFLVGTSKSPYDSKDRQQYCIEWWKTHRPKGNCSGTNKHLLVTFRSVDSADAAHALA
DEEFKPQDGILPTNERPLMTDYPTHHQKPIVIDYDYNNGYKAVGYVMGLNSVTDYWDTTKHAIDDSL
RELLASKTLEAEQAHEQATLDQAPTKGYKHVRPYQDYACRLYGPALKRVYDNFALDWNHYAPDTHR
MPELTEIPSNIQKVPGNANHAVQIVRTHPRDKDKTIKELYIHASSWARNYIYMENQYFFYPTFAQKLKSI
RQDHWKAWAEKSGRPSSEMGILYMFIVIPHPEDDGMIPRTYDMLTELGASDPTANNPAMSNQRTYDK
ENPRPQDYPDAKTVSYKRDTGVGPVISEIKVLDHPSVQELASTYGIKVSIARLRASGPIPGHDMAYREIY
IHSKLMIVDDVFITMGSANMNQRSMTADSEINIGATGLDYAAPLRERIFKLHSGGDISGSGDFKEIPNVF
EDWNRRMAKNRNVWKSGTAALKGFLLPFEDHRATGTMHAQVTIPSTSDTSALT

>Tle5d

MSQNDIISPVATVDTSECAIITPWFVENTEYPPVPATYMPLVNGEEAFRSVHEAIANAKKTVDIICWGFQ
PSMYFIRDGKAPCIGELLVSIAKEKNVQVRILGWEAPFNSAGVAGESNVPGKDIIRFGDRNGQSAEDKQ
HAYDRWWFSRFSNSGEWIKAINGPIVSPAQPTRNSPILTPVFVGRGFDLLERAEIAYQELLYSVDSGLSG
MSAAVMAIAPTHHQKTVLVDYELPDDAVGFVMGHNMLDEYWDKDAHSSKRRPPALGREFPDRDTW
APNTGPRGKWPRQDMSCRVTGPLLEHLHQNFAQAWNRETEENLLLSRNAKCVAKKLKPLAGNGDLV
MAQLLRTQMREGRRDIEKLYLKSVNNATQYIYIENQYFRWPPLAEAIKENIKAQKKAGRIPEKDGYLY
LFVVSNVTDEGIGAGTANTQRMLDSLGRGDTIPELTRLRRMEELKAKMYEPSPYPALSDEGIAEGFRKR
AEYEKAAGELKKKVADDFMVPIDGLKVHICSLVAKDSPAEAWMPTYVHSKIMLINDVFTTHGSANINT
RSMEVDSEMNIAHECAKVTKALRKRLWNIHTGGQGVQDDPADAFDAWNKLIKRNHDNQYGEEKKTP
ETPLIQFLYDKTTLKDKD

>Tle5e

MSQNKQQISTTEGKLCATVNFNWFLKDAEKFENGTRSEPVPATFKALVNGKEAFEELHERIENAQHSID
IAIWGFQPSMHFKRDGKSPCIGDLLIQKALEGKKVRILVWSLPGNIQTFSEANLGNKPGVWLKDKVEGV
TSEQVDYDRWWYEAIQGELDEVIVNAKTDGIVHVWEAHEIEKHEKLVEFTKSPKRTNLIYKNRKVAPQ
NEDFKPRILPDGRKVNHSFKDTELPDGKGTLTDGSYDFALKKFKSHHQKTVLIDYEDPDLAVGFVLEH

NMVDNYWDDSNHSLKTTLPNKGKNSPTPLQDVSSIVTGQVLWDINHNFCQSWDRQNNKQWGKDPVD
IGITGKRQSFTRDHYQPNPSLVDDSKLVMAQIVRTYDQPNIEDIMKVYLKNIKQTTSYIYTENQYFRFPP
LVREFISHWETIKNNGRTEGPIHWFTVTNSSDEGIGAGTYTTNEMFKLLGKQDVMPGVARNIKLNELEA
QLGMAKKREFRLYNESMKAPTAEGKTVAAAEFEKNQQEIQRIEKEIGEIKEKQQIEEKKTQEAGKAGS
AKKEGELNQIESSELGQEEPNLTKELGYEISDTPGIKAHICTLMPKDENGKYVHTYKKNGKDTPAEVYV
HSKVTIMDDVFTVISSANLNTRSMQVDTELGIIMECADVAEGLRKRLWDLHTNKNSAANPDDMHDYV
VAVEAFDQWLDLINKNKQAKEKNETPTCALREFYRADPTVSRSD

>Tle5f

GVKVRVICWYMHLAQFMENPTPGDNTISWFEKGRQNRDDEQRKLDQDWYLKARLKNTGANNGVTE
LIQREWDLLWYGRDKTLVIDNIEFVTRGFGFIERVEIAWQLLVNGEDQNGSINNKFQSTVAMSAAPTHH
QKMVLIDYESPDEAVGFVMGHNTLDAYWDKDDHSYTRRGPREGRNGQTPRQDISSKVTGPILAHLND
NFCEAWDRETGGDLLGARAALFSQLKPRRDAGAAIMAQLLRTQQQDGVHDIRELYLKATNNVTNYIY
MENQYFRWEPIASAIKTAVQKQQIGGRSHAEHGSIHVFVVTNSSDEGMGDGSMNTYRMLDSLGRKDV
LPGIAKLERNEALDRELDAAKLKTTNARQSLAALDDPRIRQAYINPKLLENLKIERRAEVEAAQQRQEE
LEATIPFKKSKAINFTEIQGLKVHICTLVAPDSPPDDWMPVYIHSKLTIIDDVFTTLGSANINIRSMEVDSE
LNICHEHPELSKKLRQQLWNIHTNHMGAQDDVEDAFDIWLDIIEKNKSRQRTNKGESIFKGLPPYASLIE
FYRESVTRTNLD

>Tle5g

MSENKPTVIPATESKLCATVNFNWFLKDAEKFENGTRSQPVPATFKALVNGKEAFEELHNRIEKAKSCI
DIAIWGFQPSMHFKRDGESPCIGDLLIQKALEGVKVRILVWSMWGNIQTFQEANLGNRPGVFKGREKG
VTDAQAGYDRWWFEAIRGNLDEVIVNNRAEGVIHVWNVDEIEKHKKLTEFTKSPKRTNLVYKNRHVE
KQKSNYKDKTLPEEAQRIHQHNFKDTTLPDGSYNFALKYAPSHHQKTVLIDYDKPEIAVGFVLEHNMV
DNYWDTDEHSLKTTAPNRGKNSPTPLQDVSGIVTGRVLWDINHNFCQSWDRENNKQWGDDEVETNL
MVKRKGFTREHYSYENNLWKEAWRSWRKLCAPMTNRDIEDIYEGLFKHIKQTTSYLYTGNSIFRFRAS
EXFIAHWESLKNCGRAEGPIHWFTVTNSSDDGIGKGTYTTNEMFKLLGRQEVMPNVAREIRQQELEFE
LGKRKVNQAVLFNRIMTLPTAEGKAHAQREYRANEQEITRLEKEIADVKAKQREVEKRQEVQRTEGS
GDKENKENNKNDETNSELNTIENRELSKEETNLTKALGYEFSDTPGIKAHICTLMPKDENGNYVHRYK
DKKGKELPAEVYVHSKVTIMDDVFTVISSANLNTRSMQVDTELGIIMECNEVAEGLRKRLWDLHTNK
NGAANPDDMHDYEVAKRAFNMWKDLLESNKRAQNDGTFPDCALREFYRADPTVSRSD

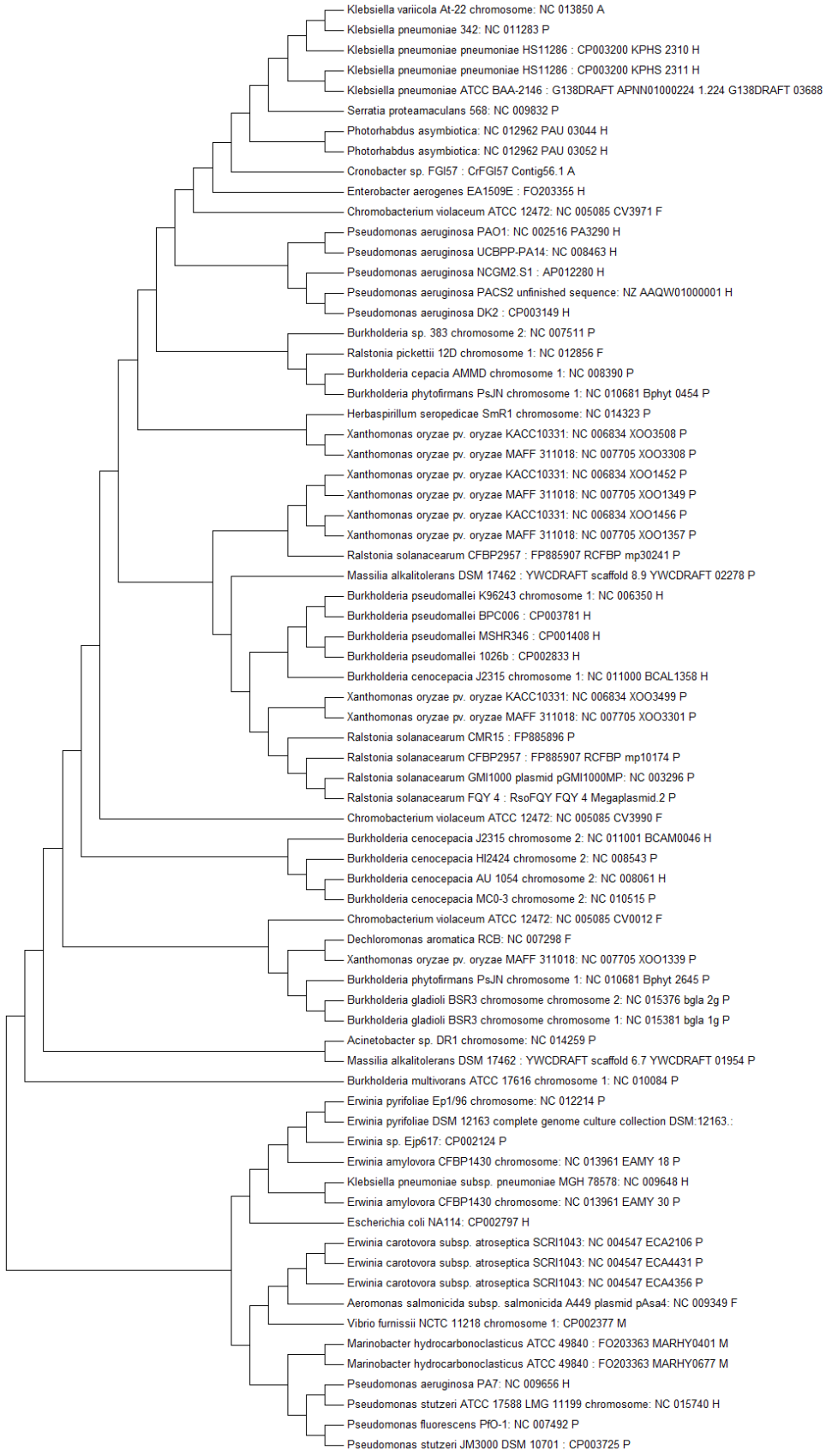


Figure S3.1. Maximum likelihood tree of Tle1 sequences from bacterial strains isolated from arthropod (A), fresh water (F), human (H), marine (P) and plant niches.

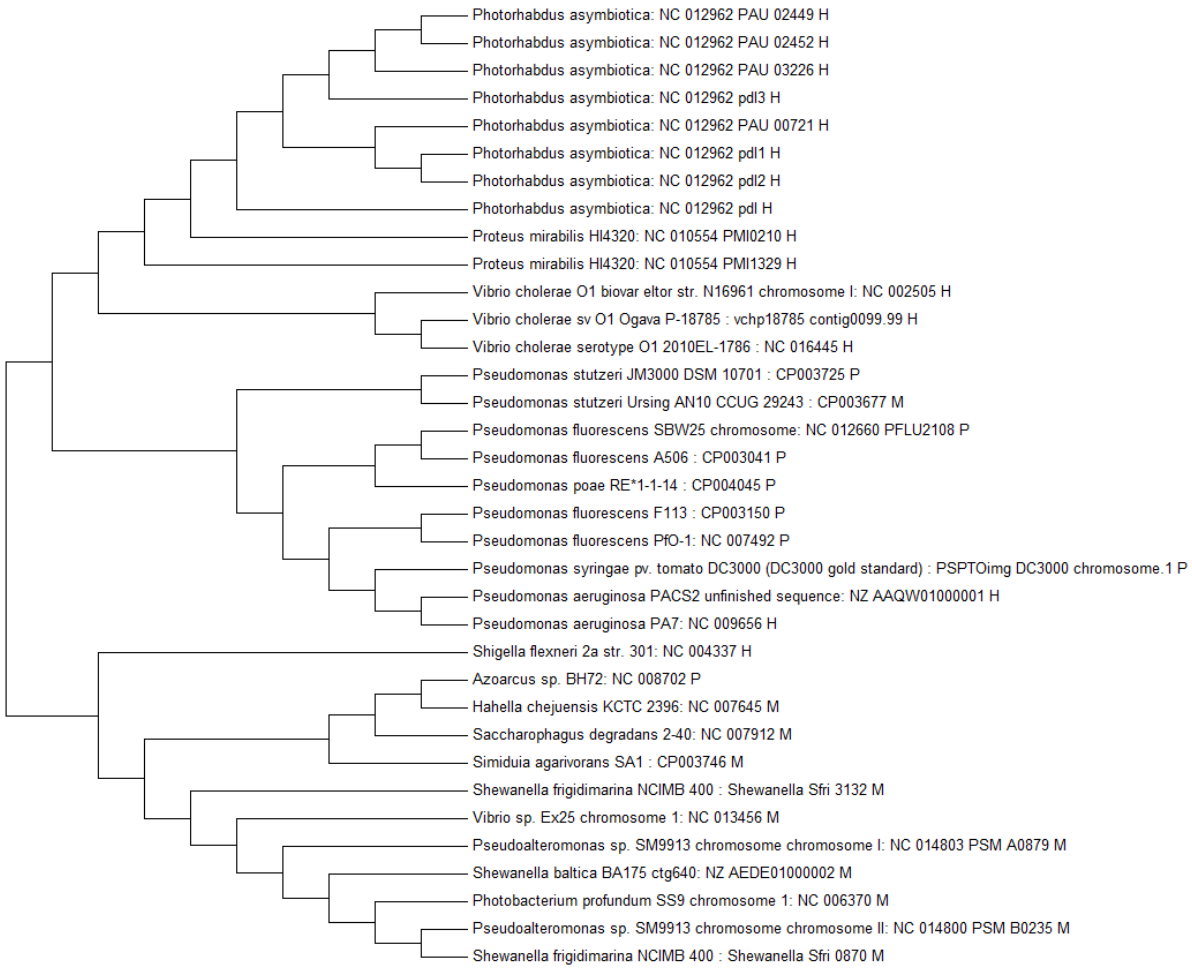


Figure S3.2. Maximum likelihood tree of Tle2 sequences from bacterial strains isolated from arthropod (A), fresh water (F), human (H), marine (P) and plant niches.

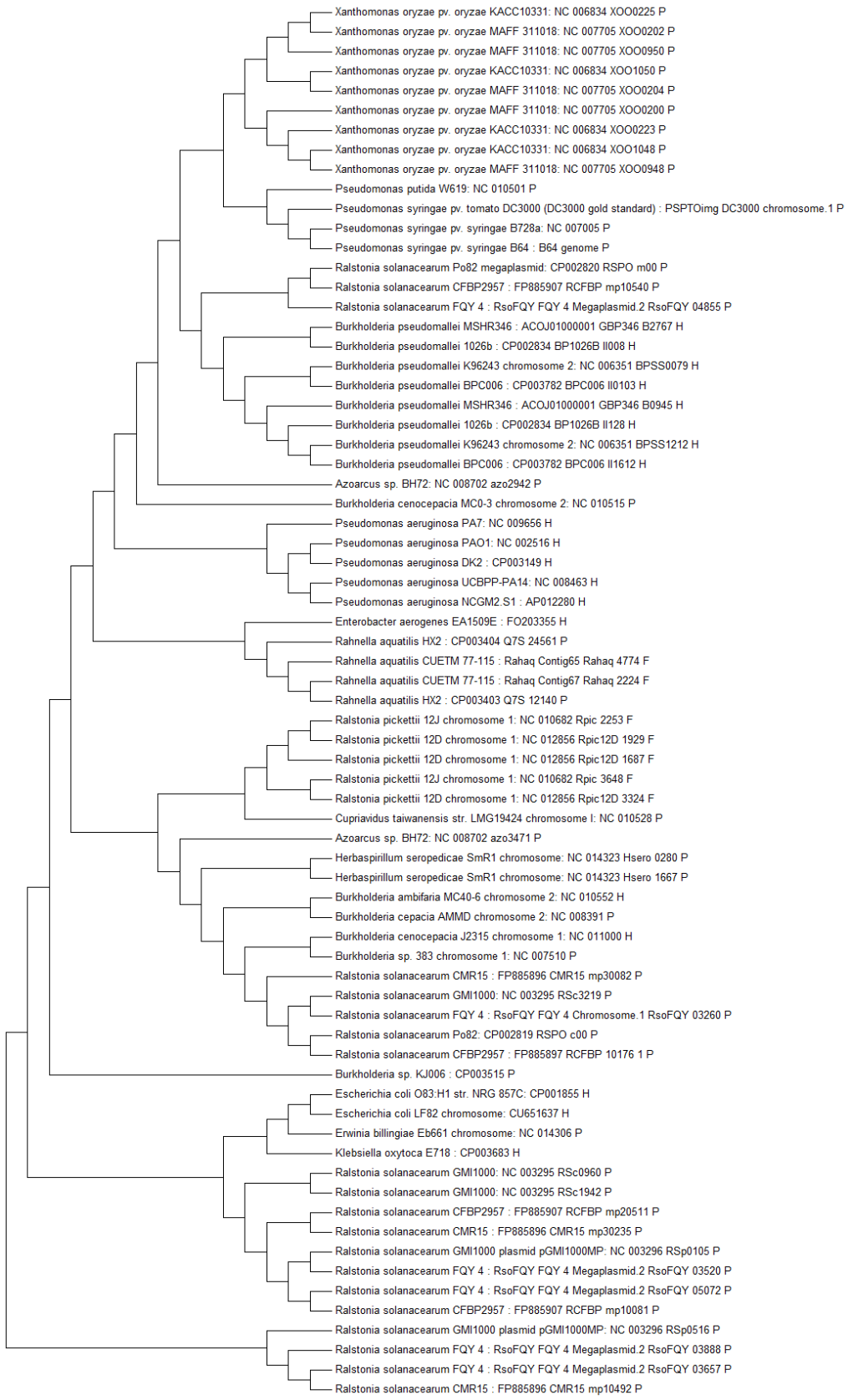


Figure S3.3. Maximum likelihood tree of Tle3 sequences from bacterial strains isolated from arthropod (A), fresh water (F), human (H), marine (P) and plant niches.

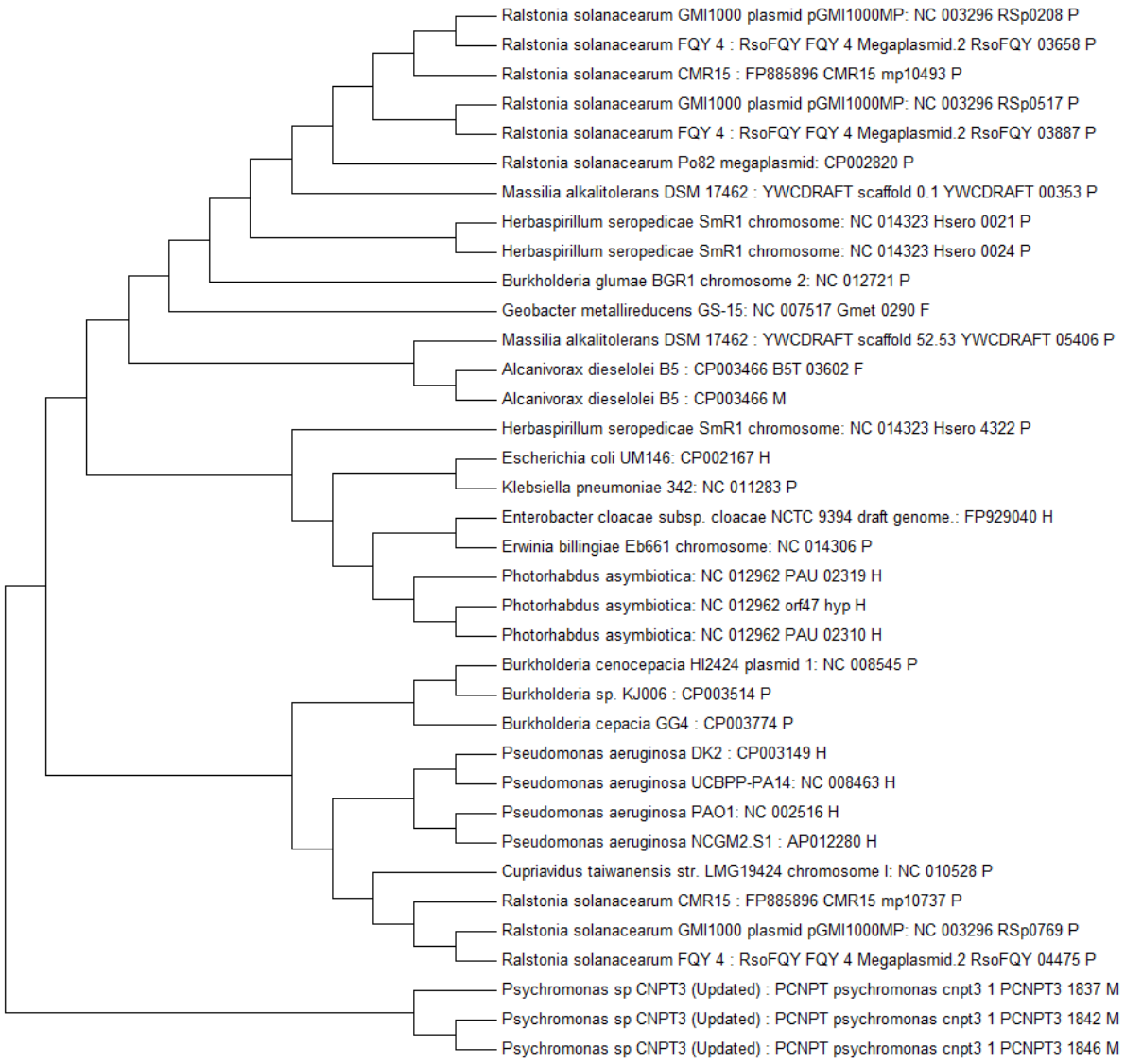


Figure S3.4. Maximum likelihood tree of Tle4 sequences from bacterial strains isolated from arthropod (A), fresh water (F), human (H), marine (P) and plant niches.

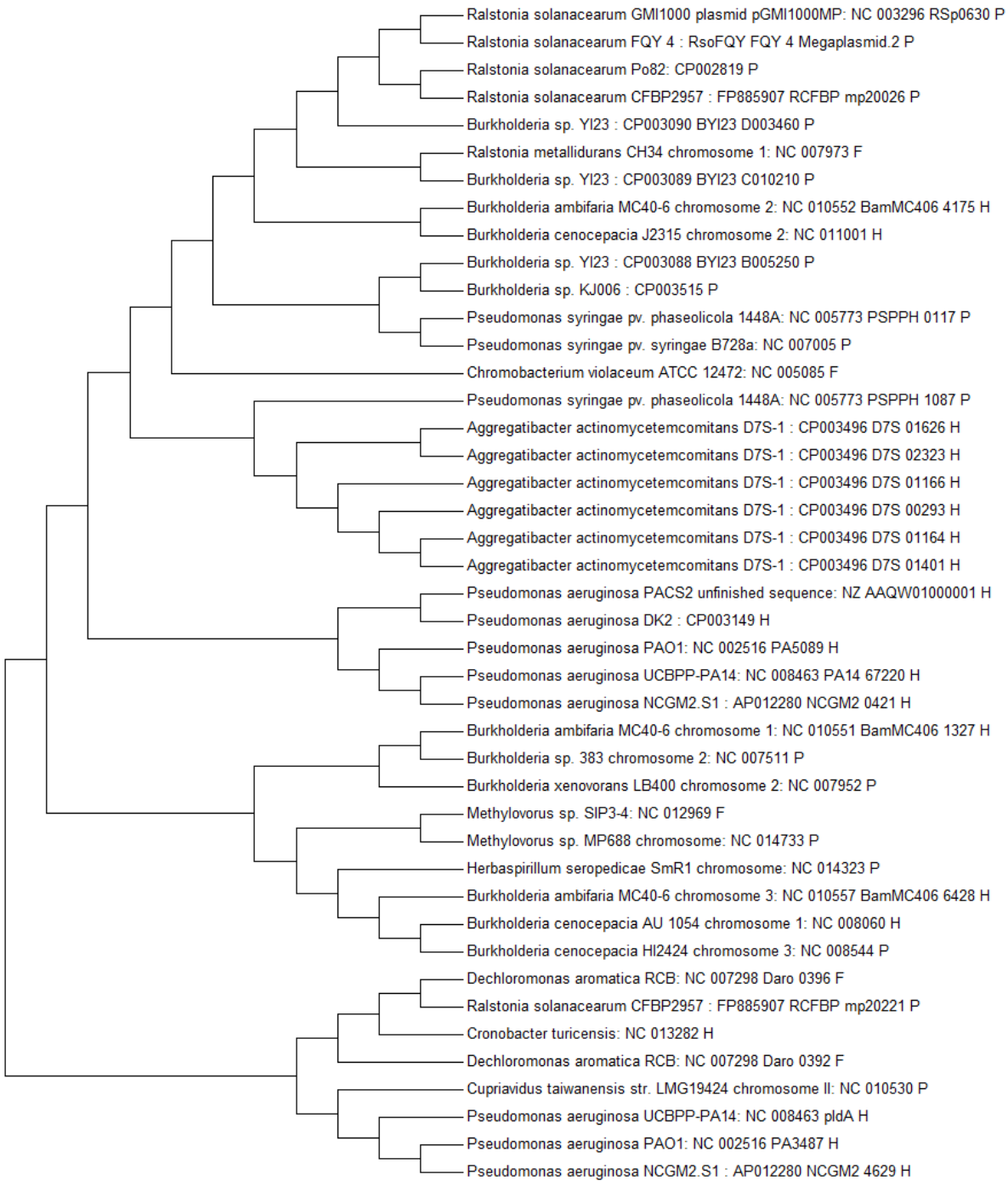


Figure S3.5. Maximum likelihood tree of Tle5 sequences from bacterial strains isolated from arthropod (A), fresh water (F), human (H), marine (P) and plant niches.

Acknowledgements

This thesis represents my failures, missed opportunities, mistakes, and crippling character flaws and bad luck, but anyone reading this right now is looking at a body of work I am proud of. And though I worked hard at it (eventually), it also represents the hard work and support of several other people. First and foremost I want to thank my supervisor Prof. Fergal O'Gara for (mis!)placing so much faith in me and giving the opportunity to work on this project.

I'd like to thank all the senior researchers and staff in BIOMERIT for their help during this thesis, Marlies, Claire, Bernie and John O'Halloran. I'd especially like to thank Dr. Matthieu Barret for helping me get started when I first arrived. I think we had a great partnership, and it was fantastic to work with you. Merely being thanked in my acknowledgement section is wholly inadequate for Dr. Jerry Reen, but as theses don't have an apology section it will have to do. Thanks for being the positive to my negative this last year. I know you're working too hard, but I really think it's going to pay off for you.

In my time here I've had highs (meeting the Bulgarian minister for science) and lows (Tuesdays 12 until much later than 1), but most importantly I had the pleasure to share them with a great bunch of lads, right through from Elizabetha to Jose and his brother Javier. Most have left now, so a special mention to the ones that stuck around to see the last exhausted horse finally stumble over the finish line. Yvonne, it was a great to work in a lab with you and your gorgeous buns. Also I liked your baking. And of course Dave, who whenever I was having doubts and focusing on the negatives, reminded me of what really matters most – the proper cleaning and storage of the spoons. There's been too many passed through BIOMERIT to talk about everyone who has left, but I'd like to mention, in order of which they appear in my acknowledgement section, Eileen (probably the most generous person I know, excluding myself), Jemma (I used those invasion assays in my thesis, now you can truly say you've made a contribution to science) Conan (a wordy adversary), Christina, Katelin, Bob, Helen from Texas (how you struggled against the handicap of two X chromosomes was an inspiration to us all) and Sharon (after writing this thesis I'm even more impressed with how easy you seemed to find it). Even if our paths never cross again, I know that I'll carry your names me with always as we scream through the infinite void of space on this pale blue dot we call love.

They probably will never read it, but I'm grateful to Olivier Lesouhaitier and the group in Rouen, especially Thibaut who helped me settle in France and adapt to a different timezone/milk. Also thanks to Dr. Marcel Jansen and Aoife Coffee for their help with my plant experiments.

I'd also like to mention my friend Pat Murphy. Although you know I started my PhD several weeks before Pat Murphy, so I could never really have learned anything directly from him. Still I think, in a way, in a very meaningful way…

And of course, I'd like to thank the fans for their support, by which I mean my family for allowing me to fulfil my dream of living as a teenager well into my 20s.

I'm very grateful that right now we're living in an environment that allows people to pursue scientific goals and provides the support for research. I think the best and only meaningful thanks I can give for this is to advocate for funding for basic research.