Temperature- and host-dependent transcriptional responses in the entomopathogenic bacterium, *Yersinia entomophaga* MH96

A thesis presented in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy
in
Genetics

at Massey University, Albany Campus,
New Zealand.

Amber Rose Paulson

2020

Abstract

Yersinia entomophaga MH96 is a virulent pathogenic bacterium that is infective towards a broad range of insects and is under development as a biopesticide. MH96 produces insecticidal toxin complex called Yen-TC that is secreted at temperatures of 25 °C and below and has been shown to be the primary virulence factor (VF) during *per os* challenge against the New Zealand grass grub, *Costelytra giveni* and other agricultural pests (Hurst et al., 2011a, 2019).

New insights into the pathobiology of MH96 during insect infection were gained from the *in vivo* transcriptome, including identification of a core secreted weaponry of co-expressed/co-secreted VFs, including Yen-TC and other exoenzymes; however, many other diverse types of VFs, including toxins, effectors, fimbriae, secretion systems, efflux pumps, iron acquisition, stress response and metabolic adaptation were also identified as highly expressed under *in vivo* conditions. A small DNA-binding protein, Yen6, was shown to be under thermoregulation at the transcriptional level and host-dependent-regulation at the post-transcriptional level and contributed to virulence during intrahemocoelic infection of *Galleria mellonella* at 37 °C.

The *in vivo* transcriptome of Δ*yen6* and *in vitro* DNA-binding specificity analysis provided evidence that Yen6 is a novel LytTR-containing regulator that activates a ribose uptake/metabolism gene cluster, *rbsD-xylG-rbsC-xylF-rbsK-ccpA*, and represses a fructose uptake/metabolism gene cluster, *IIA-fruK-IIB* and a gene for RNA-binding protein *yhbY* during infection at 37 °C. Another small DNA-binding protein, Yen7, was also implicated as a potential temperature-dependent activator of Yen-TC component genes and over-expression of *yen7* resulted in restored secretion by MH96 at 37 °C; however, deletion of *yen7* did not abrogate Yen-TC production. Experimental investigations into potential regulatory linkages between Yen6 and *yen7* were undertaken, and evidence to date does not support Yen6 as transcriptional repressor of *yen7*.

A 17.5 Kb unstable element within the genome of MH96 with linkages to Yen-TC and toxin secretion, motility and cell shape was identified. Overall the findings

presented in this thesis represent the most detailed investigation of MH96 pathogenesis to date, reinforcing MH96 as one of the most highly entomopathogenic bacteria known to humankind; yet suggesting MH96 has possibly maintained at least one core thermoregulatory mechanism more typical of an opportunistic pathogen.

Acknowledgements

Much appreciation to supervisor Dr. Mark Hurst and co-supervisors Dr. Maureen O'Callaghan, Dr. Xue-Xian Zhang and Dr. Paul Rainey for affording me the opportunity to carry out my Ph.D. research on MH96 and all support and helpful feedback provided throughout my program. This research is aligned to and was supported by the Next Generation Biopesticides Programme, funded by the New Zealand Ministry for Business, Innovation and Employment (Contract C10X1310).

Thank you to the Canadian National Science and Engineering Council for Postgraduate Scholarship-Doctoral Program. Thank you, Universities New Zealand, - Te Pōkai Tara for the Commonwealth Scholarship. Thank you, New Zealand Microbiological Society, for supporting this work with a travel grant and student talk award. Thank you to New Zealand Institute of Advanced Study at Massey University for supporting this work. I wish to thank the examiners for providing comprehensive reviews of the thesis as well as thoughtful feedback.

Thank you to AgResearch Limited Ph.D. students, research associates, scientists and technicians on the Forage Science - Soil Science Team for assistance with many aspects of my project, including sharing equipment, donating supplies, providing advice or training. Thank you to AgResearch Ltd. Statistician Paul Maclean for statistical analysis of phenotype micoarray data. Thank you to Marion Schoof and Calloghan Innovation for expression and purification of Yen6 recombinant protein. Thank you to Dr. Naren at University School of Natural and Computational Sciences at Massey University for conducting electrophoretic mobility shift assays. Appreciation to Dr. Peter Fineran at Otago University for generously providing the pVIK107 vector. Appreciation also to fellow Massey distance students Bruce Borquist and Jill Sorensen for support during the writing process.

Final thanks to mom and dad for all your support over the years, as well as my partner, Michael Ashbee for making sacrifices to come to New Zealand so I could work on this project and for your help and support.

Table of Contents

List of illustrations

List of tables

List of supplementary illustrations

List of supplementary tables

List of R source code

Table of Abbreviations

Abbreviation	Definition
ADP	adenosine diphosphate
AHL	N-acyl homoserine lactone
ALA	aminolevulinic acid
AMP	adenosine monophosphate
apoLP-III	apolipophorin III
asRNA	anti-sense ribonucleic acid
ATP	Adenosine triphosphate
β gal	β galactosidase
bin	binary toxin
BLAST	Basic Local Alignment Search Tool
BLASTp	protein BLAST search against protein sequence
BLASTx	translated nucleotide BLAST search against protein sequence
Bo.	*Bordetella*
Bt	*Bacillus thuringiensis*
Ca.	*Caenorhabditis*
CBP	chitin binding protein
cDNA	complementary deoxyribonucleic acid
CDT	cytolethal distending toxin
CFU	colony forming units
CP	cell pellet
CPM	counters per million
CS	cell supernatant
CSP	cold-shock protein
Csr	Carbon storage regulator
DBM	diamondback moth
DE	differentially expressed
Dhc	heavy chain Dynein
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DOPA	droxyphenylalanine
DRS	direct repeat sequence
EDTA	ethylenediaminetetraacetate
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
EPEC	enteropathogenic *Escherichia coli*
Er.	*Erwinia*
EST	expressed sequence tag
FCM	fuzzy c-means
FDR	false discovery rate
Fe-S	iron-sulfur cluster
G6PD	glucose-6-phosphate dehydrogenase
GFP	green fluorescent protein
GH	glycosyl hydrolase
GI	genomic island
GlcNAc	N-acetyl glucosamine

Abbreviation	Definition
GMP	guanosine monophosphate
GOI	gene of interest
GSH	glutathione
h	hour
Hcp	hemolysin co-regulated protein
Hf	high-fidelity
HPAI	high pathogenicity island
HPI	hours post-infection
ICE	integrative and conjugative element
IF3	initiation factor 3
IPGT	isopropyl β-d-1-thiogalactopyranoside
LB agar	LB agar (Merck)
LB broth	Luria broth base (Invitrogen)
LC-ESI-MS/MS	liquid chromatography electrospray ionization ion trap-tandem mass spectrometry
LCMO	lytic chitin monooxygenase
LD_{50}	lethal dose 50
LPS	lipopolysaccharide
LRR	leucine-rich repeat
LTTR	LysR-type transcriptional regulator
MCF	makes caterpillar floppy
MCS	multiple cloning site
MES	2-(N-morpholino) ethanesulfonic acid
MGE	mobile genetic elements
min	minute
MncRNA	micro RNA-like non-coding ribonucleic acid
mRNA	messenger ribonucleic acid
MU	Miller unit
MW	molecular weight
NADPH	nicotinamide adenine dinucleotide phosphate oxidase
NCBI	National Center for Biotechnology Information
ncRNA	non-coding ribonucleic acid
NEB	New England Biolabs
NRPS	Nonribosomal peptide synthetase
OD	optical density
OMP	outer membrane protein
ORF	open reading frame
Pa.	*Paenibacillus*
PAGE	polyacrylamide gel electrophoresis
PAI	pathogenicity island
PCA	principle component analysis
PCR	polymerase chain reaction
PG1	phylogroup 1
PKS	polyketide synthase
PM	peritrophic membrane
PO	phenoloxidase
POI	protein of interest

Abbreviation	Definition
PPP	pentose phosphate pathway
PPT	phosphopantetheinyl transferase
proPO	prophenoloxidase
Ps.	*Pseudomonas*
PTS	phosphotransferase system
QS	quorum sensing
RBS	ribosomal binding site
RDF	recombination directionality factor
Rhs	rearrangement hotspot
RIN	RNA integrity number
RLE	relative log expression
RNA	ribonucleic acid
RNA-seq	whole transcriptome shotgun sequencing
RNS	reactive nitrogen species
ROS	reactive oxygen species
rpm	rotations per minute
RR	response regulator
rRNA	ribosomal ribonucleic acid
RTX	repeats-in-toxin
Sa.	*Salmonella*
Sc.	*Streptococcus*
SCOTS	selective capture of transcribed sequences
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sh.	*Shigella*
SOD	superoxide dismutase
Sp.	*Spodoptera*
St.	*Staphylococcus*
T1SS	type 1 secretion system
T2SS	type 2 secretion system
T3SS	type 3 secretion system
T4SS	type 4 secretion system
T5SS	type 5 secretion system
T6SS	type 6 secretion system
T7SS	type 7 secretion system
T8SS	type 8 secretion system
TB	terrific broth
tBLASTx	translated nucleotide BLAST search against nucleotide sequence
TC	toxin complex
TC-associated factors	insecticidal toxin complex-associated
TCA	tricarboxylic acid cycle
TCRS	two-component response regulator
Tm	annealing temperature
TMS	trimethylation

Abbreviation	Definition
TPS	two-partner secretion
tRNA	transfer ribonucleic acid
U.K.	United Kingdom
UQ	upper quartile
USD	United States dollar
UTR	untranslated region
VF	virulence factor
VFDB	virulence factor database
VgrG	valine glycine repeat protein G
VIP	vegetative insecticidal protein
wHTH	winged helix-turn-helix
Wi.	*Wigglesworthia*
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside
Yen-TC	*Yersinia entomophaga* insecticidal toxin complex

Chapter 1. Introduction

1.1 Thesis Statement

This thesis investigates temperature- and host-dependent regulatory responses controlling the production of virulence factors (VFs) in the entomopathogenic bacterium *Yersinia entomophaga* MH96, using both high-throughput *in vivo* transcriptomics and molecular microbiology approaches. Entomopathogenic bacteria have evolved pathogenic strategies to infect insect hosts, which rely on the ability to regulate a coordinated response to environmental cues encountered during infection. Genome-wide responses to changes in temperature and *in vivo* growth conditions by MH96 are characterized in this thesis, followed by experimental work defining a thermoregulator that switches carbon metabolism during infection at 37 °C and contributes to intrahemocoelic virulence against *Galleria mellonella*. The thesis also reports on the presence of an excisable element within the genome of MH96 that contributes to secretion, cell shape and motility.

1.2 Thesis contribution to knowledge

- Provides a method of *in vivo* transcriptomics for bacterial pathogens from the hemolymph of model host, *Galleria mellonella*, which has broad potential application in pathogenic bacteriology;
- Offers insights into the pathobiology of MH96 by genome-wide *in vivo* transcriptome investigations, representing the first entomopathogenic bacterium *in vivo* transcriptome to be reported;
- Gives evidence for the presence of complex and layered regulatory responses to temperature and host-specific factors by an insect pathogen that occur at both transcriptional and post-transcriptional levels;

- Represents the first investigations into thermoregulatory mechanisms of MH96's main weapon, secreted insecticidal toxin complex called Yen-TC, by putative transcriptional regulators Yen6 and Yen7;

- Highlights novel LytTR-containing transcriptional regulator Yen6 as an important temperature- and host-dependent small DNA-binding protein that can activate ribose uptake/metabolism genes and repress genes for fructose uptake/metabolism and an RNA-binding protein involved in protein translation during infection at 37 °C; and

- Provides the first report of an unstable island containing phage-related holin-endolysin secretion system and citrate fermentation clusters, that is associated with secretion, motility and cell shape in MH96.

1.1 Major research motivators – biopesticide development, thermoregulation of Yen-TC and studying pathogenic bacteria *in vivo*

Y. entomophaga MH96 represents a virulent entomopathogen that is known to secrete Yen-TC, a potent toxin that is required to penetrate the midgut allowing for successful infection of insect hosts (Hurst et al., 2011a, 2014; Marshall et al., 2012). Temperature has also been shown to have major effects on virulence by *per os* challenge against *G. mellonella* and Yen-TC secretion by MH96 as well (Hurst et al., 2011a, 2015). In addition to Yen-TC component genes, which are encoded on a pathogenicity island (PAI) called PAI_{YE96}, the genome encodes numerous other putative VFs, including toxins and secretion system machinery, many of which are also located on regions of the genome believed to be acquired by horizontal gene transfer (HGT) (Hurst et al., 2016). Despite noticeable phenotypic differences of MH96 at 25 and 37 °C during intrahemocoelic infection of *G. mellonella,* temperature shifts were not found to affect virulence level (Hurst et al., 2015). Based on these results, this thesis reports on further investigations into the *in vivo* transcriptome and putative VFs encoded by MH96, probing molecular host- and temperature-dependent regulatory mechanisms controlling virulence during intrahemocoelic infection of *G. mellonella* at 25 and 37 °C.

There are three primary drivers for the research presented in this thesis. Firstly, this thesis contributes to the identification of potential biopesticide agents with novel modes of actions, especially by using *in vivo* transcriptomics to identify genes for putative VFs that responded to conditions specifically within an insect host. A second major motivator for research conducted as part of this thesis is the current focus on understanding thermoregulatory mechanisms controlling insecticidal TC production by pathogenic bacteria, including *Yersinia* spp. (Fuchs et al., 2008; Starke and Fuchs, 2014; Starke et al., 2013). Finally, this thesis was also driven by the need to better understand genome-wide responses of pathogenic bacteria to the host environment and how host- and temperature-dependent signals influence virulence and metabolic adaptation during infection.

1.1.1 Entomopathogenic bacteria – important source of biopesticides

With an ever-growing human population, global food security is facing serious and impending threats including climate change, reduced soil productivity and loss of crop production due to pests, such as weeds, diseases and animals. Among animal pests, arthropods (mainly insects and mites) are responsible for the most global crop damage (18 – 26 %), with an estimated total annual loss of more than $470 billion USD (Culliney, Thomas, 2014). Arthropods are typically managed by application of synthetic pesticides, which is considered less effective compared to weed management that can be achieved by both herbicides and mechanical control (Oerke, 2006). Arthropod pests are also a challenge to manage because modern industrial agricultural systems are more prone to infestations compared to traditional farming methods; therefore, intensive production systems must receive more excessive pesticide application, which can select for resistance among pest populations (Culliney, Thomas, 2014; Oerke, 2006). So, to meet future global food security needs, the development of more sustainable and efficient management tools for arthropod pests must be a research priority.

Due to non-target effects on natural predators and pollinators as well as risks to human health and the environment, growing social pressure has led nations to impose tighter regulations on the use of chemical pesticides, especially endocrine disruptors (García et

al., 2017; Li et al., 2016). In addition, development of alternative strategies to control arthropod pests is also needed to combat insect resistance to pesticides; for example, Diamondback moth (DBM), *Plutella xylostella*, has become one of the world's most economically damaging agricultural pests, now demonstrating resistance to 95 different chemical and biological pesticide agents (according to the Arthropod Pesticide Resistance Database as of 2019). Additionally, symbiont-mediated resistance to chemical pesticides has also been described, where populations of hemipteran insects have gained resistance to chemical pesticides by acquiring (from the field) symbiotic bacteria of the genus *Burkholderia*, that can detoxify or metabolize pesticidal agents (Kikuchi et al., 2012; Tago et al., 2015).

Derived from natural sources like bacteria, viruses or fungi, biopesticides represent a promising avenue for control of arthropod pests, especially since these agents tend to have specific targets, rather than broadly affecting both target and non-target species. Well-known biopesticide products based on *Bacillus thuringiensis* (*Bt*) consist of spore-forming *Bt* strains that produce pore-forming Cry and Cyt toxins and vegetative insecticidal proteins (VIPs) (Bravo et al., 2011; Federici, 2005; Glare et al., 2012; Jurat-Fuentes and Jackson, 2012). *Bt* has been commercially available as a sprayable biopesticide for over 70 years, and Cry toxins and VIPs have been transformed into several different insect-resistant transgenic crops (Bravo et al., 2011; Jurat-Fuentes and Jackson, 2012; Lacey et al., 2015). However, just like synthetic pesticides, insects have been shown to develop resistance to *Bt* when selection is applied both in the laboratory and field (Farias et al., 2014; Gassmann, 2016; Pickett et al., 2017; Reisig and Kurtz, 2018). Therefore, there is a growing need to use integrated management approaches, including development of novel biopesticide agents with diverse and novel modes of action (Bravo et al., 2011).

The Gammaproteobacteria class of bacteria contains many entomopathogenic species, including the triphasic nematode symbionts *Photorhabdus luminescens* and *Xenorhabdus nematophila*, as well as insect pathogenic *Pseudomonas* (*Ps.*) *entomophila, Serratia entomophila* and MH96, all of which are known to produce a diversity of insecticidal toxins and other insect-specific VFs. This thesis is focussed on identifying novel VFs in MH96 (Hurst et al., 2011b), which is a highly virulent insect pathogen and currently

under development as biopesticide, but the methods described here could be widely applicable to any potential biological control agent. Here we describe a method for the identification of putative insect-specific VFs from genome-wide *in vivo* transcriptome sequencing.

1.1.2 Thermoregulation of insecticidal TC from entomopathogenic bacteria

Understanding regulatory and secretory mechanisms associated with insecticidal TCs produced by pathogenic bacteria represents an active area of research and recent reports have characterized thermoregulatory mechanisms governing insecticidal TC expression in *Y. enterocolitica* W22703 (Starke and Fuchs, 2014; Starke et al., 2013) and, more recently, secretion of insecticidal TC by domesticated phage-related holin-endolysin secretion system that is co-located with the insecticidal TC genes on a PAI called *tc*-PAI^{Ye} (Springer et al., 2018a). The occurrence of insecticidal TCs associated with diverse mobile genetic elements (MGEs) such as plasmids or PAIs is widespread among genomes of pathogenic bacteria including *S. entomophila*, *P. luminescens, P. asymbiotica* (Duchaud et al., 2003a; ffrench-Constant et al., 2003; Waterfield et al., 2004), *X. nematophila* (Morgan et al., 2001), *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica*, *Y. mollaretii, Y. frederiksenii, Ps. syringae* pv. tomato, *Fibrobacter succinogenes* and *Treponema denticola* (Dobrindt et al., 2004, 2015; Dodd et al., 2006; Fuchs et al., 2008) and MH96 (Hurst et al., 2011a). This is likely the result of a complicated evolutionary history of high rates of HGT and recombination of insecticidal TC genes among diverse pathogens.

Extensive structural investigations have also provided insights into the molecular mechanism of toxin translocation by insecticidal TCs in *P. luminescens* (Gatsogiannis et al., 2013, 2016; Meusch et al., 2014) and MH96 (Busby et al., 2012, 2013b; Landsberg et al., 2011; Piper et al., 2019) revealing a versatile toxin delivery system. Despite the obvious importance of insecticidal TCs in pathogenesis against insects, there has been limited study into regulatory mechanisms of insecticidal TCs that have evolved in entomopathogens. To this end, this thesis aims to further characterize temperature- and host-dependent regulatory mechanisms controlling Yen-TC production in MH96, including the role of two hypothetical regulators, Yen6 and Yen7.

1.1.3 Coordinated virulence responses of bacterial pathogens to host- and temperature-dependent cues

Coordinated responses of VFs to environmental signals encountered within the host is essential to pathogenic bacteria, but much of our current understanding of such global response networks has been identified under specific *in vitro* conditions, including metal limitation (Andrews et al., 2003; Litwin and Calderwood, 1993; Morrissey et al., 2004), shifts in temperature (Lam et al., 2014; Ono et al., 2005; White-Ziegler and Davis, 2009) and pH (Foster, 1999; Olson, 1993) or exposure to oxidative stress (Fang et al., 2016; Van Der Straaten et al., 2004) for example, but the actual signals that the pathogen responds to during infection are not that well-known. Furthermore, in the environment, microbes are expected to exist in much more dynamic and complex systems compared to *in vitro* conditions experienced during laboratory culture conditions, and there is a growing need to understand how pathogens respond and adapt to *in vivo* environments, especially in the development of novel antimicrobials/biopesticides as well as understanding how host-dependent factors are driving the evolution of pathogenesis and antibiotic resistance among microbes. So, along these lines, this thesis investigates both temperature- and host-dependent regulation of VFs of MH96 to identify important regulatory networks involved in virulence.

Temperature is a common regulatory cue for pathogenic bacteria; however, thermoregulatory mechanisms can be complex, exerting multi-level control at transcriptional and post-transcriptional levels (Lam et al., 2014). For example, the expression of pYV-encoded T3SS and Yop effector genes by human-pathogenic *Yersinia* spp. (i.e., *Y. pestis*, *Y. enterocolitica* and *Y. pseudotuberculosis*) is controlled by transcriptional regulator LcrF (VirF) (Böhme et al., 2012; Han et al., 2004; Marceau, 2005). In *Y. pseudotuberculosis*, transcription of *lcrF* is repressed at temperatures lower than 37 °C by nucleoid-associated protein YmoA, which is heat-labile at 37 °C (Böhme et al., 2012). The messenger RNA of *lcrF* is also thermally regulated by the presence of an RNA thermometer that obscures the Shine-Dalgarno sequences at temperatures lower than 37 °C due to secondary structure formation (Böhme et al., 2012).

Adding to the complexity, another heat-labile regulator, RovA, is also known from *Yersinia* spp. as a transcriptional repressor of genes for VFs (Cathelyn et al., 2006, 2007; Nagel et al., 2001; Wagner et al., 2013) and tolerance to host-induced stress and colonization (Heroven et al., 2007). Like LcrF, RovA is also heat-labile and undergoes a conformational change at 37 °C, resulting in de-repression of its regulon upon infection of mammalian host (Herbst et al., 2009; Quade et al., 2012). RovA is also thermoregulated at the protein-level by enhanced proteolytic cleavage by at 37 °C (Herbst et al., 2009). In addition to the thermoregulatory control of important VFs by LcrF, RovA and YmoA in *Yersinia* spp., temperature is also known to be a key regulatory factor that controls motility (Ding et al., 2009; Kapatral et al., 1996, 2004) as well as insecticidal TC (Starke and Fuchs, 2014; Starke et al., 2013) and Yst exotoxin production (Delor and Cornelis, 1992; Mikuliskis et al., 1994). In MH96, temperature is also known also affect the production of Yen-TC components and secretion under *in vitro* conditions (Hurst et al., 2011a), and considering temperature is such a key regulatory cue among other pathogenic *Yersinia* spp., further investigations into temperature-dependent regulation of virulence by MH96 were considered warranted.

Pathogenic bacteria have not only evolved to sense physiological changes in the environment, such as temperature, pH, or iron availability, but also host-specific factors including host-derived hormones (Karavolos et al., 2013; Lesouhaitier et al., 2019; Moreira et al., 2016; Patt et al., 2018) or antimicrobial peptides (AMPs) (Bader et al., 2005; Otto, 2009), cytokines (Luo et al., 1993; Porat et al., 1991; Wu et al., 2005) and glutathione (Reniere, 2018; Reniere et al., 2015; Wong et al., 2015) produced by the immune system of the host. Furthermore, pathogenic bacteria also use nutritional sensing in order to modulate virulence during infection; for example, a carbon and nitrogen nutrient source found in mammalian skin, urocanate, is emerging as a cue for bacterial pathogenesis (Zhang et al., 2014). Similarly, ethanolamine- (Anderson et al., 2015; Gonyar and Kendall, 2014; Kendall et al., 2012; Luzader et al., 2013) and sugars, such as fucose can also potentiate virulence and colonization by intestinal pathogens that sense the break-down products of other gut associated bacteria (Autieri et al., 2007; Pacheco et al., 2012). Sensing of the amino acid L-proline by entomopathogenic *P. luminescens* and *X. nematophila* was found to regulate metabolism, proton-motor force and virulence (Crawford et al., 2010), which makes sense because L-proline is known

as a major constituent of the hemolymph of *G. mellonella* (Killiny, 2018). Recognition of host through interkingdom signaling is often a critical step for pathogenesis (Kendall and Sperandio, 2016; Miller et al., 2007; Rohmer et al., 2011); however, the underlying mechanisms related to host recognition by bacterial pathogens, especially entomopathogens remains mainly uncharacterized. While this thesis does not necessarily identify specific host-factors recognized by MH96 during insect infection, this work does characterize the host-dependent effects on virulence regulation in MH96 and then builds on these findings to provide further insights into the complex molecular mechanisms governing pathogenesis of MH96.

1.2 Thesis structure

The introductory and literature review components of this thesis are located within this chapter (Chapter 1) and Chapter 2, respectively. Chapter 1 provides general introduction to the research topic and background context motivating the research presented in this thesis, while Chapter 2 provides an overview of VFs and virulence regulatory mechanisms of entomopathogenic bacteria with specific focus on MH96 and temperature-dependent regulation of insecticidal TCs. The general molecular materials and methods are provided in Chapter 3. This thesis contains four data chapters that are presented in draft manuscript format, each including a brief manuscript-style intro, summary methods, results, discussion and summary. The four data chapters included in this thesis are as follows:

Chapter 4: *In vivo* transcriptome provides insights into putative host- and temperature-dependent response of entomopathogenic bacterium, *Yersinia entomophaga* MH96 during infection of *Galleria mellonella*;

Chapter 5: Targeted investigations into putative VFs of MH96; consisting of:

Part A. Phenotyping of several deficient mutant strains by protein visualization and virulence testing by intrahemocoelic infection of *G. mellonella*; and

Part B. Molecular characterization of Yen6 and Yen7, including regulatory influences on *chi1* (Yen-TC proxy), Yen-TC component levels and secretion;

Chapter 6: Characterization of the Yen6 regulon and molecular phenotyping of Δ*yen6* and Δ*cspA123*/HCUI$_{YE96}$ by RNA-seq; and

Chapter 7: Characterization of an excisable holin-endolysin secretion system and citrate fermentation island in the genome of MH96.

The last chapter of this thesis (Chapter 8) provides a final discussion, including synopsis of all combined data to form conceptualized models of MH96 pathogenic strategies at 25 and 37 °C. The final chapter also considers limitations and future perspectives related to the research presented in this thesis.

1.3 Research implications

Advanced understanding of MH96 pathobiology may drive future research-based development of safer and more sustainable biopesticide products, First, this thesis describes a high-throughput method to identify putative insect-specific toxins and effectors that could be developed as biopesticide products with novel modes of action or enhancers for pre-existing products to support adaptive management approaches; and second, by determining putative VFs that respond to the *in vivo* environment encountered at mammalian host temperature (i.e., 37 °C), these methods also identify putative VFs which may support mammalian infection and could be considered as potential candidates for targeted deletion in genetically modified biopesticide products as a means to reduce risks of non-target infections.

As *G. mellonella* emerges as an important model host for pathogenic testing for human pathogenic bacteria, the methodologies described within this thesis, for both *in vivo* transcriptomics and *in vivo* β-galactosidase reporter assay from within the hemolymph of *G. mellonella* are predicted to have broad applicability among the field of pathogenic bacteriology. It is also predicted that with increasing genomic resources for *G. mellonella* becoming available, these methods will also foster enhanced dual-seq approaches that may provide new insights into regulatory networks controlling virulence and in-host metabolism by opportunistic pathogens, including medically important human pathogens such as *Escherichia coli*, *Sa. enterica*, and *Yersinia spp*.

A major finding reported in this thesis was the discovery of LytTR-containing regulator, Yen6, which was shown to contribute to virulence of MH96 against *G. mellonella* at 37 °C by modulating central carbon metabolism towards the pentose phosphate pathway and repression of a gene for the RNA-binding protein YhbY. While much work was undertaken as part of this thesis to experimentally determine regulatory linkages between Yen6 and the putative Yen-TC activator, *yen7*, limited evidence was found to support a direct regulatory connection between Yen6 and *yen7*, even though both loci are co-located on PAI_{YE96}. As a result, additional investigations will be required to unravel the host- and temperature-dependent regulatory mechanisms controlling production and secretion of Yen-TC and co-secreted factors by MH96.

Finally, this thesis reports an unstable element (unexpectedly discovered from the transcriptome of Δ*cspA123*/$\Delta HCUI_{YE96}$ strain) that can actively excise from the genome of MH96, and when excised MH96 undergoes a conspicuous phenotypic change, with markedly reduced protein secretion and Yen-TC production, rounder cell shape and reduced motility. These results have important implications on future research, especially due to the high rate of excision in MH96, which has major phenotypic consequences that may affect formulation for development or efficacy of whole-cell biopesticide applications.

Chapter 2. Virulence factors produced by entomopathogenic bacteria - focus on *Yersinia entomophaga* and temperature-dependent regulation of insecticidal toxin complexes

2.1 Defining virulence factors of entomopathogenic bacteria

Similar to human pathogens, entomopathogenic bacteria have evolved specialized virulence factors (VFs) that are essential for gaining: entry into the insect body, colonization and proliferation within host tissue(s) that are not normally colonized, circumvention of host defences, and bioconversion of host tissues into nutrients (Hill, 2012; Vallet-Gely et al., 2008). VFs contribute to pathogenesis by facilitating attachment, invasion and/or damage to host cells and tissues, avoidance or disruption of normal host immune response and acquisition of essential nutrients, such as iron, while inside the host (Hill, 2012; Korves and Colosimo, 2009; Wu et al., 2008). While many proposed strategies to categorize or define virulence factors have been presented in the past, there is not a single globally accepted method for defining VFs (Hill, 2012; Korves and Colosimo, 2009; Vallet-Gely et al., 2008; Wassenaar and Gaastra, 2001; Wu et al., 2008). In this work, any gene product that has been shown to contribute to pathogenesis (i.e., deficient mutant strain has reduced virulence compared to wild-type) was considered a VF. Subsequently, putative VFs were defined when a gene demonstrated significantly higher expression *in vivo* compared to *in vitro* growth conditions and either A) shared significant sequence homology to a VF identified in another system (primarily assigned from the virulence factor database (VFDB)), or B) were located within a specific genomic region of interest containing several other known or putative VFs (e.g., PAI_{YE96} or rearrangement hotspot (Rhs)-associated regions).

Using this very broad definition of a VF, this work considers a much larger breadth of gene products with diverse roles in pathogenesis (Table 2.1), than have been undertaken in previous 'VF bioprospecting' efforts with entomopathogens, which typically are more focused on secreted toxins and effectors. Therefore, it is important to also distinguish between primary VFs produced by entomopathogenic bacteria (i.e., secreted toxins and effectors) and other virulence-associated life-style genes, which are

considered in this thesis as secondary VFs (see definition in Wassenaar and Gaastra, 2001). Primary VFs are generally toxins or other effectors that are secreted and directly act on host cells and tissues to promote pathogenesis and are generally not present in non-pathogenic strains, while secondary VFs comprise all other gene products that contribute a pathogens ability to proliferate and cause disease within a host. The identification and characterization of novel primary VFs produced by entomopathogenic bacteria represent an important source for discovery of new insecticidal biopesticide agents, but secondary VFs, including secretions systems, defensive proteases, secondary metabolites and virulence regulatory mechanisms also represent key areas of research with respect to bacterial entomopathogens. Below is a broad overview of different types of VFs that have been characterized from entomopathogenic bacteria.

Table 2.1. Categories of virulence factors considered in this study.

Category	Definition	Examples
Primary – secreted and directly act against host cells/tissues		
Toxin	Proteins that are secreted and directly target host cells/tissues, giving rise to pathogenesis.	Insecticidal toxin complex, vegetative insecticidal protein or Cry toxin.
Effector	Proteins that are translocated directly into host cells by type 3 or 6 secretion systems.	*Yersinia* spp. Yop or *Salmonella* spp. Spa type 3 secretion system and hemolysin co-regulated protein (Hcp) or polymorphic rearrangement hot-spot (Rhs) toxins.
Secondary – all other non-primary factors contributing to pathogenesis (i.e., gene abrogation results in attenuated virulence).		
Adherence	External structures used to attach to host cells or tissues.	Fimbriae or filamentous hemagglutinins.
Defence response	Factors used to detoxify against antimicrobial compounds or host-induced stress.	Superoxide dismutase, efflux pumps, iron-sulfur cluster.
Horizontal gene transfer	Factors that promote lateral gene acquisitions or mobilization.	Conjugative pili or integrases.
Iron acquisition	Systems used by pathogenic bacteria to acquire iron in the host environment.	Siderophores, hemolysins and iron-specific transporters.
Metabolic adaptation	Enzymes and transporters involved in bioconversion of host-derived nutrients.	Amino acid transporters or chitinases.
Mobility	Factors related to motility.	Flagella or chemotaxis.
Outer membrane	Factors related to modification of outer membrane/capsule often used to evade host-immune response.	Lipopolysaccharide or outer membrane proteins.
Regulation	Regulatory factors controlling pathogenic response to host.	Two-component sensors, quorum sensing or diguanylate cyclase.
Secretion system	Multi-component systems used for translocation and secretion of toxins and effectors by bacteria.	Type 2, 3, and 6 secretion systems.
Unclassified	Factors that could not be classified specific categories listed above.	Non-ribosomal peptide synthetase or tellurium resistance genes.

2.1.1 Toxins

Among primary VFs produced by entomopathogenic bacteria, toxins have been best characterized and developed as biopesticide agents and there are many different types of insecticidal toxins, ranging in complexity, structure and mode of action, but all toxins cause destruction of host cells and tissues. Historically, research on insecticidal toxins has primarily focused on pore-forming Cry and Cyt toxins of *Bt;* both toxins interact with host cell receptors, which triggers a conformational change and insertion into the host cell membrane resulting in the formation of a pore (Parker and Feil, 2005). Cry and Cyt toxins are not just restricted to *Bt* though; a Cyt-like toxin with insecticidal activity against the pea aphid was identified from the plant pathogenic bacteria, *Dickeya dadantii* (Grenier et al., 2006).

In addition to Cry and Cyt toxins of *Bt,* binary (Bin) toxins are another crystallized insecticidal protein produced by entomopathogenic bacteria. Bin toxins are comprised of two components, BinA and BinB, with the former component being considered the main toxic effector, while the role of the BinB component is the binding of host midgut epithelial membrane (Srisucharitpanit et al., 2013, 2014). The Bin toxin produced by *Lysinibacillus sphaericus* was shown to have insecticidal activity against mosquitoes in the genera *Culex* and *Anopheles*, where it is thought to interact with the midgut epithelial cell membrane receptors, causing pore formation and subsequent cell destruction (Tangsongcharoen et al., 2015). Interestingly, the nematode symbionts, *Photorhabdus luminescens* and *Xenorhabdus nematophila* also encode genes for Bin toxins, PirAB and XaxAB respectively; however, these are not crystallized like in *L. sphaericus*. PirB shares similarities with a beetle protein, leptinotarsin (i.e., a known potent neurotoxin) (Crosland et al., 2005) and the pore forming domain of *Bt* Cry toxin (ffrench-Constant et al., 2007). PirAB has oral activity against mosquito larvae (Ahantarig et al., 2009) but also has a toxic effect when injected into the greater wax moth, *Galleria mellonella* (Ahantarig et al., 2009), while XaxAB has hemolytic activity against insect hemocytes and mammalian red blood cells (Ribeiro et al., 2003).

Insecticidal TCs, otherwise known as ABC toxins, have been identified in a number of important entomopathogens, including *P. luminescens* (Waterfield et al., 2001), *X.*

nematophila (Morgan et al., 2001), *Bt* (Blackburn et al., 2011), *Serratia entomophila* (Hurst et al., 2000) and MH96 (Hurst et al., 2011a). Other pathogenic species of *Yersinia* also encode insecticidal TCs, some of which are not known to be pathogenic against insects (Alenizi et al., 2016; Fuchs et al., 2008). The TCs of entomopathogens are orally active against a broad range of insects and are tripartite toxins, assembled from TcA, TcB and TcC sub-units, with maximum toxicity achieved in heteromultimeric form (A_5BC) (Lang et al., 2010; Waterfield et al., 2001). More detail of one of the best described insecticidal TCs, Yen-TC from MH96, is provided in section 2.3.1 below.

Further to TCs, pore-forming Cry/Cyt and Bin toxins mentioned above, entomopathogenic bacteria are known to harbour genes that encode many other toxins, including the apoptotic makes caterpillar floppy (MCF) cytotoxin encoded by *P. luminescens* (Daborn et al., 2002), repeats-in-toxin (RTX) exoproteins produced by Gram-negative bacteria (Daborn et al., 2001), cytolethal distending toxins produced by pathogenic Proteobacteria (Vodovar et al., 2006) and vegetative insecticidal proteins (VIPs), which are known as the other main toxin produced by *Bt* (Zhu et al., 2006). Work continues to characterize the diversity of toxins produced by insect pathogenic bacteria and identify new toxins with different targets and modes of action for biopesticide applications.

2.1.2 Host cell adhesion

Another requirement of a successful pathogen is the ability to interact with host cells and tissues. For this reason, entomopathogens must produce adhesive factors to attach to host cells. Such adhesion VFs range from simple monomeric adhesion molecules to more sophisticated multimeric proteins including attachment fimbriae and highly complex secretion systems (Pizarro-Cerdá and Cossart, 2006). The genome of *P. luminescens* contains a number of putative genes that are likely involved in adhesion, including two adhesins, one calcium dependent cadherin-like adhesion molecule and several fimbrial clusters (Duchaud et al., 2003b). The genome of another insect pathogen, *Pseudomonas* (*Ps.*) *entomophila,* contains several putative genes that likely contribute to host cell adhesion, including two autotransporter proteins with a pertactin-like domain, two filamentous hemagglutinin and one putative surface

adhesion protein (Vodovar et al., 2006). Other surface molecules, such as invasins, can facilitate attachment and invasion of host cells by manipulation of actin filaments at the site of entry (Pizarro-Cerdá and Cossart, 2006), yet this phenomenon has not been well studied in entomopathogenic bacteria, although the genome of *P. luminescens* does contain one putative invasin (Heermann and Fuchs, 2008).

2.1.3 Secretion systems

Secretion systems of Gram-negative pathogenic bacteria are another important class of VFs, especially since toxins and effectors must be secreted and/or translocated into host cells to work. Secretion of exotoxins and other effectors is facilitated by type 1, 2 and 5 secretion systems (T1SS, T2SS and T5SS), while the intracellular delivery of toxins, secondary metabolites, proteases and other virulence determinants is carried out by type 3, 4, and 6 secretion systems (T3SS, T4SS and T6SS) (Costa et al., 2015). Type 7 secretion systems have been shown to be important for virulence of *Mycobacterium* (Daleke et al., 2012) and secretion of a nuclease toxin that targets competitor bacteria in *Staphylococcus aureus* (Cao et al., 2016). The type 8 secretion system is involved in the secretion and assembly of usher/chaperone fimbriae (Barnhart and Chapman, 2006; Remaut et al., 2008). A type 9 secretion system has also been described, and was shown to be involved in gliding motility and chitin utilization among Bacteroides (Kharade and McBride, 2014; McBride and Zhu, 2013). The presence of different types of secretion systems can vary among entomopathogenic bacteria; for example, the genome of *P. luminescens* encodes several secretion systems, including four T1SS, one T2SS and T3SS (Waterfield et al., 2002), while *Ps. entomophila* does not encode any T3SS or T4SS secretion systems (Vodovar et al., 2006). Due to the non-constitutive activity of secretion systems among Gram negative bacteria, it has been suggested that specialized adhesion molecules targeting host-specific receptors may trigger activation of secretion systems in the host environment (Gerlach and Hensel, 2007).

Similar to multi-drug efflux pumps, T1SSs consist of three components structures that span both the inner and outer membrane and use adenosine triphosphate (ATP) energy to secretes proteins directly into the extracellular milieu (Delepelaire, 2004; Kanonenberg et al., 2013). The T1SS secretes unfolded proteins containing signal

sequences, such as C-terminal glycine rich repeats (Delepelaire, 2004) or N-terminal encoded C39 peptidase domain (Kanonenberg et al., 2013). Many entomopathogenic bacteria including *P. luminescens*, *X. nematophila* and *Ps. entomophila* produce toxins that belong to a larger family of RTX exoproteins, which are all secreted via the T1SS (Linhartová et al., 2010). Such toxins include MCF in *P. luminescens* and *X. nematophila* (Daborn et al., 2002; Kim et al., 2017) as well as metalloprotease PrtA in *P. luminescens* and *X. nematophila* (Bowen et al., 2003; Caldas and Cherqui, 2002) and homologous metalloprotease AprA in *Ps. entomophila* (Lee et al., 2018; Liehl et al., 2006). In *P. asymbiotica*, both the genes encoding PrtA and the T1SS exporter complex, as well as *mcf1*, which encodes a MCF toxin were all found to be under thermoregulation during *in vitro* growth, with significantly higher expression at 37 compared to 28 °C (Mulley et al., 2015); however, temperature-dependent differences in the levels of these T1SS-secreted toxins by *P. asymbiotica* have yet to be determined under *in vivo* conditions.

While T1SS use a single step to secrete proteins across both membranes, T2SS and T5SS require a two-step process for protein secretion because these take advantage of "Sec-dependent" mechanisms. First, unfolded substrate proteins containing N-terminal signal sequences are directed across the cytoplasmic membrane by either the SecYEG translocon (conserved cytoplasmic transportation machinery) or Tat (twin-arginine-translocation) system. Once inside the periplasmic space, either unfolded (from SecYEG translocon) or folded (from Tat transporter) exoproteins can then be transported across the outer membrane by T2SS (Nivaskumar and Francetic, 2014) or T5SS autotransporter, two-partner systems (TPS) or trimeric autotransporter adhesins (Chauhan et al., 2019; Leo et al., 2012; Leyton et al., 2012). The structure and function of T2SS and T5SS completely differ however; the T2SS structure is comprised of 12-16 different components spanning both inner and outer membranes (Nivaskumar and Francetic, 2014), while T5SS secretion pore (C-terminal β-barrel translocator domain) is fused into a larger, multi-domain polypeptide that also encodes a passenger domain (i.e., the exoprotein) and N-terminal Sec-dependent signal sequence (Leo et al., 2012; Leyton et al., 2012). Another key difference between T5SS and T3SS is the requirement for chaperones during periplasmic transit of the TPS and outer membrane-bound β-barrel assembly machinery (Leyton et al., 2012).

Among entomopathogenic bacteria, T2SS have been identified from the genomes of *P. luminescens, X. bovienii, Y. entomophaga* and *Ps. entomophila* (McQuade and Stock, 2018), but their functional role in secretion during pathogenesis has not been a focus of much research to date. On the other hand, several VFs produced by entomopathogenic bacteria are known to be secreted using T5SS mechanisms (McQuade and Stock, 2018). For example, *Ps. entomophila* produces a putative autotransporter sharing similarities with a hemolysin produced by *Y. pestis* (Vodovar et al., 2006). Furthermore, a number of different TPS -secreted VFs are encoded by *P. luminescens* TT01, including the gene for hemolysin PhlA, which was shown to be expressed during insect infection using green fluorescent protein (GFP)-reporter assays, but was not required for virulence against *Manduca sexta* (Brillard et al., 2002). A homologous hemolysin produced by *X. nemtatophila*, XhlA, was shown to contribute to virulence against *M. sexta*, possibly by destruction of host hemocytes (Cowles and Goodrich-Blair, 2005).

T3SS are commonly found in Gram-negative bacterial pathogens (Galán et al., 2014), where in mammalian pathogens they have been shown to facilitate attachment to intestinal cells and active transport of toxic effectors into the host cell, resulting in cellular death (Tseng et al., 2009). Recent studies suggest that the target of T3SS in entomopathogens may not be the epithelial cells of the insect gut, but instead the phagocytic immune cells (hemocytes) (Fauvarque et al., 2002; Walker et al., 2013). For example, *P. luminescens* uses a T3SS to inhibit phagocytosis by the immune cells of *M. sexta* (Brugirard-Ricaud et al., 2005). *Y. entomophaga* encodes two T3SS systems that may also target host cells during infection; T3SSYE1 is arranged in a similar gene order to the T3SS of *Salmonella* (*Sa.*) *enterica* serovar Typhimurium SP1 1 (involved in entry of intestinal cells) and located downstream from a putative intestinal cellular attachment intimin protein, suggestive of a role in invasion of host mid-gut epithelial cells (Hurst et al., 2016). In contrast the MH96-encoded, T3SSYE2 shares the same gene order with the *Ysa* system of *Y. enterocolitica* biotype 1B (Hurst et al., 2016), which was demonstrated to facilitate intracellular replication of *Y. enterocolitica* within *Drosophila melanogaster* S2 cells (Walker et al., 2013).

Like the T3SS, the T6SS is also a cell envelope spanning multi-component secretion machine that is used to translocate effectors into target cells (Ho et al., 2014; Zoued et

al., 2014). The T6SS is a contractile injection system, sharing structural components homologous to contractile bacteriophage tails (Böck et al., 2017; Leiman et al., 2009). T6SS have also been widely identified from the genomes of *P. luminescens*, *P. asymbiotica*, *X. nematophila*, *X. bovienii*, *Ps. entomophila* and MH96 using bioinformatics, but only few functional studies have been carried out (McQuade and Stock, 2018). *Ps. entomophila* encodes two predicted T6SS that are proposed to target eukaryotic hosts because these loci share similarities with divergent *Pseudomonas* species known to infect humans (Sarris and Scoulica, 2011). In *X. bovienii*, strains that encode a greater number or T6SS could out-compete strains encoding fewer T6SS during *in vitro* intra-species competition assays (McMullen et al., 2017). Temperature-dependent expression of T6SS has been observed in *Y. pestis*, where genes for T6SS structural proteins were more highly expressed at 26 compared to 37 °C, indicating a possible role for T6SS in the flea vector (Pieper et al., 2009).

2.1.4 Defense against host immune responses

Entomopathogenic bacteria are known to produce secondary metabolites and proteolytic enzymes that enhance virulence by interfering with the host immune response or cause direct damage to host cells/tissues. For example, *P. luminescens* is able to avoid melanisation by the host immune system by producing a small-molecule antibiotic, (*E*)-1,3-dihydroxy-2-(isopropyl)-5-(2-phenylethenyl)benzene, which inhibits the key melanisation enzyme, phenol oxidase (Eleftherianos et al., 2007). *P. luminescens* also produces a number of metalloproteases, including PrtS, which causes melanization when injected into *G. mellonella* (Held et al., 2007). Furthermore, a serralysin metalloprotease present in *S. marcescens* culture supernatant was shown to inhibit attachment capability of silkworm hemocytes and murine macrophages *in vitro* (Ishii et al., 2014). In another insect pathogen, *Ps. entomophila* supresses the host immune response by producing a metalloprotease that can deactivate host-derived antimicrobial peptides (AMPs) (Lee et al., 2018; Liehl et al., 2006).

Production of secondary metabolites by non-ribosomal peptide synthetases (NRPS) or polyketide synthases (PKS) has also been described from entomopathogenic bacteria. Secondary metabolites produced by NRPS and PKS system in the honey bee specific

pathogen *Paenibacillus* (*Pa.*) *larvae* have antibacterial and antifungal activities and are thought to play a role in eliminating bacterial competitors (Garcia-Gonzalez et al., 2014a, 2014b; Müller et al., 2015; Sood et al., 2014) or possibly even have cytotoxic effects against host cells (Sood et al., 2014).

Another important aspect of survival within the host environment is resistance to various physiological stressors that can be encountered during infection, especially oxidative stress resulting from exposure to reactive oxygen species (ROS) generated by the immune system of the host. In order to cope with such stresses, many pathogenic bacteria can produce antioxidants including reactive thiols, such as glutathione or iron-sulfur cluster or superoxide dismutase, that provide resistance against harmful ROS, reactive nitrogen species (RNS) and other damaging hydroxyls (Amich et al., 2013; Broxton and Culotta, 2016; Reniere, 2018; Song et al., 2013). In general, bacteria have evolved mechanisms to maintain cellular homeostasis, which include orchestrated responses to environmental stress, such as the SOS, stringent, cold-, heat- and acid-shock responses. Maintaining homeostasis during infection also imperative for pathogenic bacteria, and not surprisingly both cold- and heat-shock proteins have been shown to contribute to virulence of bacterial pathogens (Eshwar et al., 2017; Michaux and Giard, 2016; Stewart and Young, 2004). Furthermore, the SOS response in *Staphylococcus aureus* was shown to promote increased horizonal transfer of VFs encoded within PAIs by induction of helper prophages that mediate the lateral spread of VFs (Maiques et al., 2006; Penadés et al., 2015; Úbeda et al., 2005).

2.1.5 Outer membrane

The modification of outer membrane by pathogenic bacteria represents another important aspect of virulence, because the outer membrane of the bacteria can have direct contact with host cells and tissues and contribute to intoxication or adherence to host cells. With respect to outer membrane, a well-known example, is the surface molecule lipopolysaccharide (LPS), which is an immune system elicitor of Gram-negative bacteria and has been shown to be a VF in at least one entomopathogenic bacterium (Rodríguez-Segura et al., 2012). Also, the LPS of *X. nematophila* has been shown to suppress the phenol oxidase cascade when injected into *G. mellonella*

(Dunphy and Webster, 1991). Other important outer membrane associated VFs are outer membrane protein/porins (OMPs), which have not been extensively studied among entomopathogenic bacteria; however, in *X. nematophila* an OMP called OpnS contributed competitive advantage for growth within an insect host (Van Der Hoeven and Forst, 2009).

2.1.6 Iron acquisition

Another obstacle to persistence within the host environment is the uptake of iron and other limiting nutrients, which can be scarce within the host environment. Some entomopathogenic bacteria are known to encode genes for iron scavenging molecules, such as siderophores and iron-storage proteins (i.e., ferritin and bacterioferritin) transferrins (Andrews et al., 2003; Ratledge and Dover, 2000) and these iron acquisition systems are considered VFs (Ratledge and Dover, 2000). The diversity of putative iron acquisition systems found in the genome of *P. luminescens* provides evidence of the importance of iron acquisition in invertebrate hosts (Crennell et al., 2000; Duchaud et al., 2003b; Heermann and Fuchs, 2008). In *Bt*, expression of a gene for a hemolysin HlyII is activated by an iron sensing ferric uptake regulator, Fur (Tran et al., 2013). These findings suggest that at time points during the infection when iron is scarce (i.e., host phagocytic cells have sequestered iron as a natural host defence), the pathogen responds by producing HlyII, which acts on hemocytes, resulting in cell lysis and release of the sequestered iron.

2.1.7 Degradative enzymes and nutrient acquisition

Except for iron, not much attention has focused on the acquisition of other essential nutrients by entomopathogenic bacteria during infection of the insect host. Recently, untargeted metabolomic profiling showed that the hemolymph of *G. mellonella* contained a rich source of nutrients in the form of amino acids, organic acids and to a lesser extent fatty acids (Killiny, 2018). A primary source of carbon that was abundant in the hemolymph was in the form of a glucose-glucose disaccharide called trehalose, which is a common sugar known to be important in insects as a carbon storage (Elbein et al., 2003), source of glucose energy (Reyes-DelaTorre et al., 2012) and cryo-protectant (Saeidi et al., 2013; Wang et al., 2010). In addition to trehalose, chitin represents another abundant source of carbon and nitrogen among arthropods (as well as crustaceans, fungi and parasitic nematodes), making it the second most abundance biopolymer in nature (Gooday, 1990; Souza et al., 2011). Accordingly, entomopathogens have evolved metabolic processes that facilitate bioconversion of chitin via the chitinolytic pathway which requires the activity of glycoside hydrolases (i.e., GH18 and 19 family chitinase or GH20 family chitobiase) that initiate the degradation of chitin by hydrolysis of the (1→4)-β-glycoside bond yielding β-1,4-*N*-acetylglucosamine $(GlcNAc)_x$ residues (Adrangi and Faramarzi, 2013; Beier and Bertilsson, 2013; Berini et al., 2018; Toratani et al., 2008; Uchiyama et al., 2003). The chitinolytic pathway can also involve enzymes in the AA10 family, called lytic chitin monooxygenases (LCMOs), otherwise known as chitin-binding proteins (CBPs), which contribute to chitin degradation by making breaks in the polymer chain through oxidative cleavage which increases accessibility for chitinases (Vaaje-Kolstad et al., 2005a, 2010).

Within the insect host, chitin also plays an important role as a component of the peritrophic membrane (PM), which is comprised of chitin fibrils attached to glycoproteins and proteoglycans and forms a physical barrier between the gut lumen and the mid-gut epithelia (Erlandson et al., 2019; Kelkenberg et al., 2015). In this context, chitinolytic enzymes secreted by entomopathogenic bacteria may also be considered primary VFs because the main outcome of their activity is degradation of the PM, which allows further access to underlying midgut epithelia, where the pathogen can then release additional VFs targeting host-specific ligands for adhesion.

Some entomopathogenic bacteria, such as *S. marcescens* and *S. proteamaculans*, secrete a chitinolytic machinery (Suzuki et al., 1998; Vaaje-Kolstad et al., 2005a, 2013), which consists of multiple chitinases (Suzuki et al., 2002), a chitobiase (Toratani et al., 2008; Uchiyama et al., 2003), and a lytic chitin monooxygenase/chitin-binding protein (LCMO/CBP) (Suzuki et al., 1999). MH96 also secretes several putative and known chitinolytic factors *in vitro* at 25 °C, including three chitinases, two (Chi1 and Chi2) that form the outer portions of the Yen-TC holotoxin (Busby et al., 2012; Landsberg et al., 2011; Piper et al., 2019) and LCMO/CBP (CbpA) (supplemental Figure S18). *Pa. larvae* produces a LCMO/CBP known as *Pl*CBP49 during infection of honey bees and when the gene for *Pl*CBP49 was deleted, the deficient strain was almost completely avirulent and could no longer degrade the PM (Garcia-gonzalez et al., 2014).

2.1.8 Regulation of virulence factors in response to host by entomopathogenic bacteria

Entomopathogenic bacteria regulate expression of key VFs in response to cues from the host environment, including host-specific nutrients or iron availability (Crawford et al., 2010; Fang et al., 2016; Tran et al., 2013). In *P. luminescens* and *X. nematophila*, L-proline (a free amino acid found in the hemolymph of *G. mellonella*) was shown to induce the production of virulence factors and secondary metabolites as well as affect proton motor force (Crawford et al., 2010). Also, in *Bt* the gene for hemolysin, HlyII, was shown to be regulated by iron availability *in vivo* through the global regulator Fur (previously described in section 2.1.6). Other than these two examples, the host-specific environmental signals that control gene regulation in entomopathogenic bacteria have not yet been widely characterized but important host signals including temperature, pH, nutrient and oxygen limitation, envelope stress and iron availability likely induce coordinated virulence responses, as shown in other pathogenic bacteria (Fang et al., 2016; Foster, 1999; Lam et al., 2014; Reniere, 2018).

One well-known regulatory mechanism used by pathogenic bacteria to respond to extracellular signals during infection are two-component regulatory systems (TCRSs), which usually are comprised of a membrane-bound histidine kinase (the sensor) and cytosolic response regulator (Zschiedrich et al., 2016). The sensor components of TCRSs

have been shown to sense temperature and light, as well as respond to changes in the concentration of Mg^{+}, O^{-2}, specific amino and carboxylic acids as well as presence of antimicrobial peptides produced by the host (Zschiedrich et al., 2016). When activated, the kinase drives phospho-relay, resulting in conformational change of the DNA binding domain of the response regulator, that acts to regulate appropriate response genes (Capra and Laub, 2012; Zschiedrich et al., 2016). TCRSs are widely distributed among pathogenic bacteria and some have been shown to be required for virulence in entomopathogens, like *Ps. entomophila*, where the TCRS GacS/GacA was shown to be a master virulence regulator that activates two small ncRNAs which then sequester two RNA-binding proteins, RmsA1 and RmsA2. In the absence of GacS/GacA activation, the RmsAs is not sequestered and can then repress the translation of target genes, including several important VFs (Dieppois et al., 2015; Liehl et al., 2006; Opota et al., 2011; Vallet-Gely et al., 2010). Similarly, the TCRS in *P. luminescens*, PhoP/PhoQ, was shown to be required for virulence in the host *Spodoptera* (*Sp.*) *littoralis* and deletion of *phoP* had impacts on several components of the cell envelope (Derzelle et al., 2004).

Quorum sensing (QS) is also emerging as important regulatory mechanism of pathogenic bacteria, including entomopathogens. QS is a process of cell-cell communication in bacteria, where gene expression is modified depending on cell densities, which is sensed through the accumulation of auto-induced signalling molecules (Rutherford and Bassler, 2012). First described in a squid symbiont, *Vibrio fischeri*, the LuxR QS system, which binds an acyl-homoserine lactone (AHL) signal molecule, was found to control bioluminescence in the light organs of the host in a cell-density-dependent manner (Antunes et al., 2010; Schuster et al., 2013; Verma and Miyashiro, 2013). The LuxR/LuxI QS sensing system is common among Gram-negative pathogenic bacteria and has been shown to be involved in the regulation of VFs in *Ps. aeruginosa* and *S. marcescens* for example (Jimenez et al., 2012; Thomson et al., 2000). In *P. luminescens*, the presence of large clusters of LuxR solos (quorum sensing signal receivers without the presence of autoinducer) containing N-terminal PAS4-domains, has led to the hypothesis that the PAS4-domain may bind insect hormones, since this domain shares similarity with insect juvenile hormone receptors in *D. melanogaster* (Brameyer et al., 2014; Subramoni et al., 2015).

Regulation of virulence by another important and ubiquitous secondary messenger, cyclic di-GMP has also been linked to insecticidal activity (Fu et al., 2018; Kalia et al., 2013; Mielich-Süss and Lopez, 2015; Slamti et al., 2016) as well as biofilm formation (Mielich-Süss and Lopez, 2015), cell-cell aggregation (Tang et al., 2016) and motility (Fu et al., 2018) in the entomopathogen, *Bt.* When cyclic di-GMP was over-expressed, over 4 % of genes were found to be differentially expressed based on transcriptome analysis, including genes related to unexpected pathways such as protein translation and nucleotide metabolism (Fu et al., 2018). Other research found that levels of cyclic di-GMP regulate chemotaxis and adhesion, via a cyclic di-GMP-sensing riboswitch located within the 5′ untranslated region (UTR) of collagen adhesion protein (Tang et al., 2016). Cyclic di-GMP is also known to regulate motility and biofilm formation among other pathogenic bacteria, and recently *Sa. enterica* serovar Typhimurium was shown to modulate levels of cyclic di-GMP in response to carbon sources such as glucose, GlcNAc, sialic acid and L-arginine (Mills et al., 2015). Cyclic di-GMP is synthesized by diguanylate cyclases, which are also recognized as important contributors to virulence (Tamayo, 2019; Tischler and Camilli, 2012). Furthermore, the diguanylate cyclase gene, *rrp1*, has been found to be essential for survival of the causative agent of Lyme disease, *Borrelia burgodorferi*, within the tick vector, where cyclic di-GMP appears to control a catabolic switch related to glycerol transport and metabolism (He et al., 2011).

A complex interplay between the regulation of virulence and flagellar motility is also recognized in *P. luminescens* and *X. nematophila* (Givaudan and Lanois, 2017). For example, in *X. nematophila*, the flagellar master regulators, *flhD* and *fliZ* were both shown to be involved in the regulation of virulence (Givaudan and Lanois, 2000, 2017; Lanois et al., 2008). DNA methylation by deoxyadenosine methylase (Dam) is another regulatory factor in *P. luminescens* and *X. nematophila* that was also shown to be an important contributor to motility and virulence, where *dam* overexpression resulted in impaired motility and effects on virulence against *Sp. littoralis* (Payelleville et al., 2017, 2018). This section provides some examples of how entomopathogenic bacteria respond to dynamic environmental stresses through a variety of regulatory mechanisms; however, a full review of this fascinating topic is considered well beyond the scope of this thesis.

2.2 Focal entomopathogen: *Yersinia entomophaga* MH96

The research reported in this thesis focused on *Y. entomophaga* MH96, which has been shown to be highly virulent against a broad range of insect pests during oral challenges due to the primary actions of a secreted insecticidal toxin called Yen-TC, which is required to penetrate the midgut allowing for successful infection (Hurst et al., 2011a, 2014, 2019; Marshall et al., 2012). While Yen-TC may be the best-studied VF produced by MH96, a Yen-TC-deficient mutant was still found to be as virulent as wild-type during infection of *G. mellonella* by injection (Hurst et al., 2015), which demonstrated that additional VFs must also contribute to pathogenesis during intrahemocoelic infection. Furthermore, the genome of MH96 was found to encode genes for numerous different toxins, secretion systems and other putative VFs suspected to have hemocylic activity (Hurst et al., 2016). For these reasons, MH96 represented an opportunity to identify novel VFs with hemocoelic activity during infection of *G. mellonella* and expand the understanding of pathogenic strategies utilized by MH96 during intrahemocoelic infection and the cytotoxic potential of MH96 against insect cells beyond Yen-TC. MH96 was considered an ideal subject for utilization of *in vivo* transcriptomics to identify genes for putative VFs that respond to the host environment, especially because the genome contains many (likely functionally redundant) putative VFs, which makes high-throughput transposon mutant screening for attenuated virulence particularly difficult in this species. A detailed review of MH96, including current phylogenetetics, understanding of pathobiology and insecticidal TC production and secretion is provided below.

2.2.1 Phylogenetic context

MH96 belongs to the genus *Yersinia* (order Enterobacteriaceae within the class Gammaproteobacteria and family Yersiniaceae). The genus currently contains 18 species, of which three, *Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica* can cause disease in humans; *Y. pestis* is the infamous causative agent of plague while the latter two species are known to cause sporadic outbreaks of enteritis in humans (McNally et al., 2016). Other species of *Yersinia*, including *Y. rhodei*, *Y. aldovae*, *Y. bercovieri*, *Y.*

frederiksenii, *Y. intermedia*, *Y. kristensenii* and *Y. moolaretti* are considered *enterocolitica*-like opportunistic pathogens causing diarrheal illness among susceptible individuals (Agbonlahor, 1986; Cafferkey et al., 1993; Loftus et al., 2002) and some strains of *Y. kristensenii* and *Y. frederiksenii* were found to have high invasion potentials in Caco-2 cells (Imori et al., 2017). Furthermore, a different strain of *Y. entomophaga*, IP36721 (not MH96), was recently isolated from human urinary tract but was not associated with pathology (Le Guern et al., 2018).

Although, the three *Yersinia* species that cause disease in humans have been the primary focus of research, the entire genus is emerging as a model for studying pathogen evolution (McNally et al., 2016). Phylogenetic analysis based on maximum-likelihood of 84 concatenated housekeeping genes revealed that the genus *Yersinia* contains 14 discrete 'species clusters' and that pathogenicity has independently evolved several times over evolutionary timescales (McNally et al., 2016; Reuter et al., 2014). Comparative genomics of *Yersinia* spp. has revealed key drivers of the evolution of pathogenesis among the group (and perhaps more broadly among Enterobacteriaceae) including gene loss/gain and genome re-arrangement (McNally et al., 2016).

In addition to the human pathogenic species, there are other pathogenic species among the genus *Yersinia*, including *Y. ruckeri*, the causative agent of red-mouth disease in salmonids (Ewing et al., 1978) and MH96, which is the focal organism in this thesis. Entomopathogenic properties of other *Yersinia* species have been reported, with *Y. frederiksenii* and *Y. intermedia* found to be insecticidal against *M. sexta* and *G. mellonella* (Springer et al., 2018b). There is also emerging evidence that some *Yersinia* may be zoonotic, including the highly heterogenous group *Y. enterocolitica*, which contains some strains with insecticidal and/or nematocidal activity (Alenizi et al., 2016; Bresolin et al., 2006; Spanier et al., 2010). Among *Y. enterocolitca* are six distinct phylogroups, of which one, phylogroup 1 (PG 1), is known to be non-pathogenic against mouse infection model (Bottone, 1997) and lacks the virulence plasmid pYV that encodes key VF required for successful human infection (Bottone, 1999). Recently PG 1 *Y. enterocolitca* was shown to be highly virulent against the insect *G. mellonella*, with a median lethal dose (LD_{50}) of only 10 colony forming units (CFUs) (Alenizi et al., 2016).

So, from an ecological evolutionary perspective, it is this high propensity for virulence among species of *Yersinia*, which includes a spectrum of pathogenic strategies ranging from opportunism (i.e., *Y. enterocolitca*) to highly-specialized zoonotic pathogens (i.e., *Y. pestis*) that make this group particularly useful for understanding drivers of pathogen evolution.

2.2.2 *Yersinia entomophaga*, MH96 – current state of knowledge

MH96 was isolated from a larva *Costelytra giveni* (Coleoptera: Scarabaeidae) cadaver (Hurst et al., 2011b), which is an endemic and economically significant pasture pest in New Zealand. Development of MH96 as a biopesticide has proven consistent pathogenesis by *per os* challenge against *C. giveni*, as well as a wide range of coleopteran, lepidopteran and orthopteran species (Hurst et al., 2011a, 2014, 2019; Marshall et al., 2012). The recently published MH96 draft genome consists of a single chromosome of ~ 4.3 Mb in length, encoding 4,225 predicted proteins and sharing 93.8 % sequences similarity to *Y. nurmii* type strain APN3a-cT (Hurst et al., 2016). The draft genome confirms the highly virulent nature of MH96, which encodes a diverse array of putative VFs, including toxins, effector proteins, secretion and iron acquisition systems, proteolytic enzymes and adhesion structures (Hurst et al., 2016). The extensive repertoire of putative VF encoding genes in the genome of MH96 supports the observation that MH96 is highly pathogenic against a broad range of hosts, probably through the utilization of a diverse pathogenic strategies to invade, colonize, proliferate within the host and cause disease.

The genome of MH96 contains regions likely acquired by horizontal gene transfer (HGT). Compared to *Y. ruckeri* 29473 and *Y. enterocolitica* 8081, the genome of MH96 contains two unique regions that encode a number of putative VFs including toxins components, iron acquisition systems and genes related to adhesion (Hurst et al., 2016). Also thought to be acquired by HGT, the genome of MH96 contains five Rhs-associated regions, each encoding a single Rhs-containing protein. Rhs-associated region 1 is located within the MH96 pathogenicity island (PAI) called PAI_{YE96}, that also encodes Yen-TC component genes and is further described in section 2.3 below. The Rhs domain resides with the C subunit of the toxin, forming one of the capped ends of the

β-barrel sheath structure (Busby et al., 2013b). The role of the remaining four Rhs-associated regions (2-5) in MH96 virulence have yet to be characterized. It is suspected that Rhs-associated regions 2, 3, and 4 contain operons involved in the production and delivery of hemocoelic active toxins, LopT, Spt4 and a hypothetical protein, respectively. LopT is known to target eukaryotic cells (Brugirard-Ricaud et al., 2004) and Spt4 shares amino acid sequence similarity with a eukaryotic transcription elongation factor domain (Hurst et al., 2016). Furthermore, Rhs-associated region 3 also appears to contain all necessary T6SS component genes and Rhs-associated region 5 is predicted to be an effector island. (Hurst et al., 2016). Further characterization of the Rhs-associated regions in MH96 and their role in intrahemocoelic infection, specifically the third region containing the putative T6SS, is an interesting direction for future research.

2.3 Insecticidal toxin complexes and Yen-TC

High molecular weight (~ 1 million Daltons) insecticidal TCs, otherwise known as 'ABC toxins', were first discovered in the nematode symbiont *P. luminescens* and were shown to have oral toxicity against *M. sexta* (Bowen and Ensign, 1998; Bowen et al., 1998). Subsequent investigations have shown that insecticidal TCs have oral activity against a broad range of insects and are produced by many other entomopathogens including, *X. nematophila* (Sergeant et al., 2003), *Bt* (Blackburn et al., 2011), *S. entomophila* (Hurst et al., 2000) and MH96 (Hurst et al., 2011a) as well as non-entomopathogenic bacteria. Insecticidal TCs are composed of three protein subunits, designated TcA, TcB and TcC, which must be assembled into the Yen-TC holotoxin for maximum toxicity (Lang et al., 2010; Waterfield et al., 2001). The genomic organization of the genes encoding the TC components may be complex; for example, the genome of *P. luminescens* contains multiple homologous *tcA*, *tcB* and *tcC* component genes, which are primarily located on large tandom arrays or PAIs (Waterfield et al., 2002). In contrast, three *Serratia* species and *Y. frederiksenii* encode the *sepABC*-like *tc* genes on a virulence plasmid (Dodd et al., 2006; Hurst et al., 2000).

In MH96, the Yen-TC operon is located within 32 kb PAI_{ye96} and consists of two *tcA* (*yenA1* and *yenA2)*, one *tcB* (*yenB)*, two *tcC* (*yenC1* and *yenC2*) homologous genes and two chitinase genes (*chi1* and *chi2*) that flank *yenA1* and *yenA2* (Hurst et al., 2011a) (Figure 2.1). Yen-TC holotoxin is the main virulence determinant of MH96 and exposure of *C. giveni* and the diamond-back moth, *Plutella* (*Pl.*) *xylostella* to secreted Yen-TC by oral challenge resulted in dissociation of the midgut (Hurst et al., 2011a; Marshall et al., 2012). The TcA-subunit is a pentameric structures decorated by two chitinases, Chi1 and Chi2 (Busby et al., 2012; Landsberg et al., 2011; Piper et al., 2019) that are thought to increase the pathogenicity of *Y. entomophaga* against a broad range of insect hosts due to endochitinase activity, (Busby et al., 2012; Hurst et al., 2011a; Landsberg et al., 2011) as well as performing other important structural and host-surface lectin binding functions (Piper et al., 2019) .

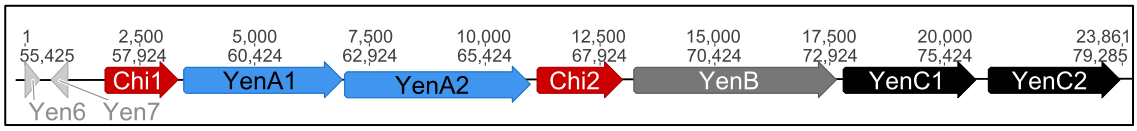


Figure 2.1: Genomic organization of insecticidal toxin complex, Yen-TC components and putative regulators Yen6 and Yen7, from *Yersinia entomophaga* MH96. Arrows represent open reading frames that are color-coded for function; light grey = regulatory, red = chitinase, blue = YenA component, dark grey = YenB component and black = YenC component. Top scale is in bp and bottom scale is with reference to MH96 genome (GCA_001656035.1).

Electron microscopy (EM)-derived structure of Yen-TC previously showed that YenA1 and YenA2 combine to form the TcA-subunit (Landsberg et al., 2011). Similarly, cryo-EM of the insecticidal TcA component from *P. luminescens* supports a syringe-like mechanism of toxin translocation across the host membrane, in which the TcA homologue forms a trans-membrane pore by interacting with receptors on the surface of host cells (Gatsogiannis et al., 2013, 2016; Meusch et al., 2014; Piper et al., 2019). More recent cryo-EM visualization of the Yen-TC TcA-subunit inserted into the liposome determined the pore forming apparatus is formed by YenA2, including the neuraminidase-like domain pore-closing loop, while a portion of YenA2 and YenA1 form the inner longitudinal pore structure (Piper et al., 2019). Cryo-EM structure also confirmed that in addition to the pore forming apparatus, Chi1 and Chi2 also interact

with host surface receptors and conformational change of Chi1 upon binding is critical to pore formation (Piper et al., 2019). This study identified that Yen-TC, Chi1 and Chi2 all had significant binding activity against a range of structures, including GlcNAc, the natural substrate of chitin (Piper et al., 2019).

With respect to the TcBC-subunit, x-ray crystallography-derived structures of Yen-TC revealed that the BC complex forms a large hollow beta-barrel structure that encases the toxic portion of the C protein and is capped on either end by a beta-propeller and the Rhs -associated core domain of the C protein on the B and C ends, respectively. The repetitive elements of the Rhs domain (YD repeats) correspond to a stand-turn-strand motif, which form the extended β-sheet tube-like shell extending from the N-terminal region of the C subunit into the B subunit (Busby et al., 2013b).

The currently proposed mode of action of Yen-TC involves recognition of host insect midgut epithelial cells surface ligands by the YenA2, Chi1 and Chi2 resulting in Chi1 conformational change exposing the pre-pore apparatus which becomes inserted into the lipid bilayer of the host cell, followed by formation of the trans-membrane pore, through which the TcC toxin-subunit is translocated into the host cell cytoplasm. As such, the YenA2, Chi1 and Chi2 components are thought to contribute to host-specificity through interaction with host-specific surface ligands (Piper et al., 2019). To this end, encoding multiple TcA orthologs may contribute to *P. luminescens'* ability to kill a broad range of insect hosts. Similarly, the TcA components encoded by *X. nematophila* have been shown to have varied activity against different lepidopteran pests (Sergeant et al., 2003). Meusch *et al*. (2014) found that the TcA of *P. luminescens* contains four putative receptor-binding domains, which share structural features with the receptor-domains of anthrax and diptheria toxins.

In MH96, the TcB and TcC homologues were shown to be the sole determinants of the toxicity by *per os* challenge against *Pl. xylostella* through histopathological and quantitative dose response analysis (Landsberg et al., 2011); however, all three protein sub-units (TcA, TcB and TcC) are required for maximum toxicity (ffrench-Constant, Richard Waterfield, 2005). The TccC3 and TccC5 components of *P. luminescens* were identified as adenosine diphosphate (ADP)-ribosyltransferases, which can act on host-cell microtubules and Rho-GTPases (Lang et al., 2010). Exposure of *G. mellonella*

hemocytes and HeLa cells to the these TcCs resulted in major modifications of host cell cytoskeleton, including clustering of actin and inhibition of phagocytosis (Lang et al., 2010). Hurst et al., (2015) also reported host cell invasion as well as ruffling of *G. mellonella* hemocytes in response to Yen-TC.

2.3.1 Insecticidal TCs in non-entomopathogenic *Yersinia* species

Several non-entomopathogenic species of *Yersinia* also encode the genes for an insecticidal TC, located within the '*Yersinia* TC-pathogenicity island' (*tc*-PAI^{Ye}), including *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica* (biotype 2-5 strains only), *Y. mollaretii* and *Y. similis* (Fuchs et al., 2008) (Figure 2.2). The 20 kb *tc*-PAI^{Ye} is missing from *Y. enterocolitica* biotype 1B (highly pathogenic) and most 1A (generally non-pathogenic) strains (Bresolin et al., 2006; Fuchs et al., 2008; Spanier et al., 2010). The *tc*-PAI^{Ye} region is embedded in a common genomic backbone, flanked by a genes for putative LysR-like DNA regulator and DNA gyrase modulator, 5' and 3', respectively (Fuchs et al., 2008). The organization of the *tc*-PAI^{Ye} region is similar among strains, although frameshift mutations in some strains are present among *tcA* and *tcB* component genes (Fuchs et al., 2008). In *Y. pestis* CO92, it is thought that a frame-shift mutation in a *tcB* gene represents an adaptive gene-loss that contributed to an evolutionary advantage for persistence within in the flea vector due to reduced insecticidal activities (Parkhill et al., 2001). Homologous *tcC* genes are also encoded outside of *tc*-PAI^{Ye} among a few strains of *Y. pestis* and *Y. pseudotuberculosis* IP32953 (Fuchs et al., 2008). Other genes located within *tc*-PAI^{Ye} include a gene for LysR-like DNA regulator *tcaR2*, and four highly conserved phage-associated genes are located between the *tcB* and *tcC* homologues. The four genes in the phage-associated region were shown to encode a lysis cassette in *Y. enterocolitica* strain W22703, consisting of a holin, endolysin, endopeptidase and two spanins (Springer et al., 2018a).

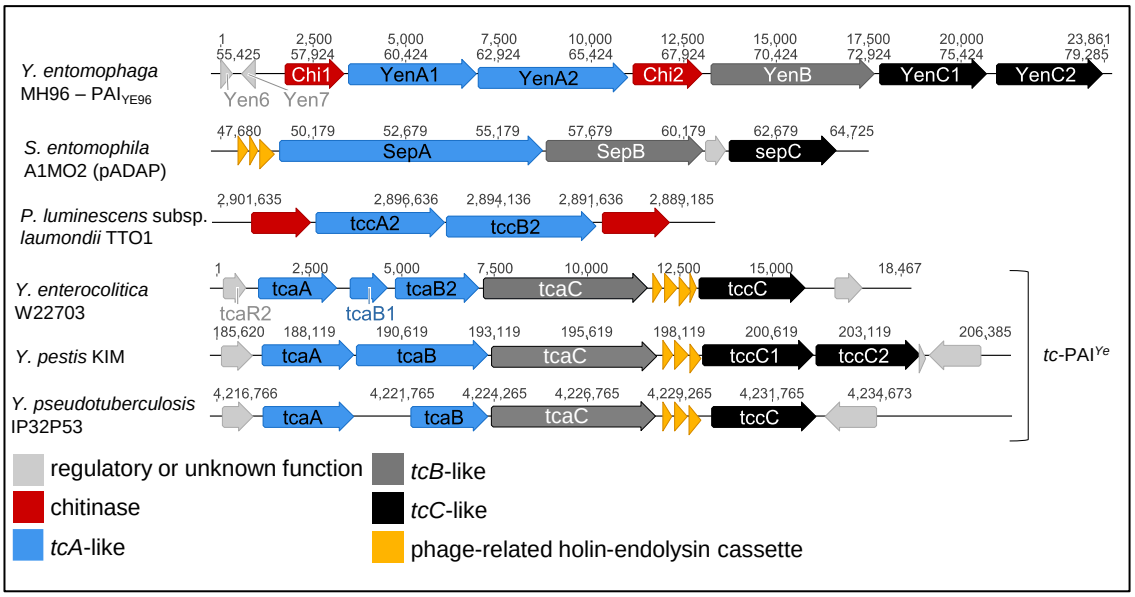


Figure 2.2: Comparison of insecticidal toxin complex components from *Yersinia entomophaga* MH96, *Serratia entomophila* A1MO2, *Photorhabdus luminescens* subspecies laumondii TTO1, *Y. enterocolitica* W22703, *Y. pestis* KIM and *Y. pseudotuberculosis* IP32953.

There is evidence that TCs may contribute to mammalian infections in *Y. enterocolitica*. It was shown that *Y. enterocolitica* strain T83 mutants with attenuated *tc* genes had reduced ability to colonize the intestinal tracts of mice compared to wild-type (Tennant et al., 2005). Transient production of TC proteins from *Y. pseudotuberculosis* strain IP32953 and *Y. pestis* strain KIM+ resulted in actin and nuclear rearrangements in cultured human gut (Caco-2) and mouse fibroblast cells (NIH3T2), respectively. Similar results were also found when recombinant TC proteins were applied topically to these cell lines, suggesting that insecticidal TCs have specificity against mammalian cells as well as insect cells (Hares et al., 2008).

The *tc* genes of mammalian pathogenic *Yersinia* have also been shown to have contribute to virulence against insects. For example, the homologous *tcA* gene of *Y. enterocolitica* is required for insecticidal activity (Bresolin et al., 2006). Furthermore, crude extract of *E. coli* strain BL21 with heterologously expressed *tc* genes of *Y. pseudotuberculosis* IP32953 were shown to be toxic against larval *M. sexta* (Pinheiro and Ellar, 2007). The insecticidal TCs produced by *Y. pseudotuberculosis* were also shown to be acutely toxic to fleas (Erickson et al., 2007). Furthermore, it was recently found that PG 1 *Y. enterocolitica* (this phylogroup is non-pathogenic to mammals and lacks major virulence plasmid pYV), is highly pathogenic to *G. mellonella* by intrahemocoelic

injection (LD_{50} of 10 cells) (Alenizi et al., 2016). Presently, it is unclear why some non-insect pathogenic species of *Yersinia* possess insecticidal *tc* genes, but this may suggest that insects serve some yet unknown purpose in the lifecycle of these pathogenic species or relate to some yet unknown environmental niche.

2.3.2 Temperature-dependent regulation of TCs in *Yersinia*

The regulation of the insecticidal *tc* genes has been shown to be temperature-dependent in some human-pathogenic *Yersinia* species. For example, expression of *tcaA* is strongly induced at low temperatures (10 – 15 °C) but repressed at mammalian body temperature (37 °C) in *Y. enterocolitica* W22703 (Bresolin et al., 2006). The expression of *tcaA* and *tcaB* homologues in two *Y. pestis* strains was upregulated at 26 compared to 37 °C, but this effect was not observed for *tccC* (Han et al., 2004; Motin et al., 2004). In contrast, expression of *tcaABC* in *Y. pseudotuberculosis* strain IP32953 was observed over a range of temperatures from 15 °C to 37 °C, visualized by sodium dodecyl sulphate-polyacrylamide (SDS-PAGE) (Pinheiro and Ellar, 2007).

A mechanism of temperature-dependent regulation of TCs has been described in *Y. enterocolitica* W22703, where transcription of *tcaA* and *tcaB* is silenced at 37 °C, but highly induced at lower temperatures (Bresolin et al., 2006). A LysR-type thermolabile regulator, TcaR2, was shown to induce *tcaA* and *tcaB* expression (Starke et al., 2013). Another LysR-type regulator, TcaR1, was shown to act as a repressor of *tcaR2* (Starke et al., 2013). The nucleoid-associated protein and H-NS in complex with YmoA were also shown to globally repress TC gene expression in W22703, by binding regions upstream and downstream of *tcaA* and *tcaB* at body temperature (Starke and Fuchs, 2014). The global role of H-NS and its associated YmoA in the temperature-dependent regulation of the *tc* genes in W22703 is not surprising, as H-NS has been shown to be global regulator of horizontally acquired virulence factors and plasmids broadly among the Enterobacteriaceae (Hüttener et al., 2015; Landick et al., 2015; Navarre et al., 2007).

2.3.3 Effects of temperature on Yen-TC production and virulence of MH96

A key characteristic of MH96 is constitutive production and secretion of Yen-TC components into liquid culture when grown at temperatures of 25 °C or lower, which can be visualized by SDS-PAGE (Figure 2.3) (Hurst et al., 2011a). While toxin components can be visualized in MH96 cell pellets (CP) grown at 25, 30 and 37 °C, more YenA1 and YenA2 (at least) are produced at 25 °C compared to higher temperatures. The toxicity of culture supernatant (CS) from MH96 grown at 25, 30 and 37 °C were tested *per os* against *C. giveni*, which confirmed the CS were only toxic when MH96 was grown at 25 °C and not at higher temperatures. In a different study, Hurst *et al*. (2015) determined that the rate of mortality of the wax moth, *G. mellonella,* when challenged *per os* with MH96, was significantly reduced from 93 % to 16 % when the host was incubated at 25 °C compared to 37 °C. Furthermore, very few (zero to less than 1,000) MH96 cells per larva were recovered from macerates of larvae provided a diet containing 2.4 x 10^8 CFUs, when incubated at 37 °C for 24 hours.

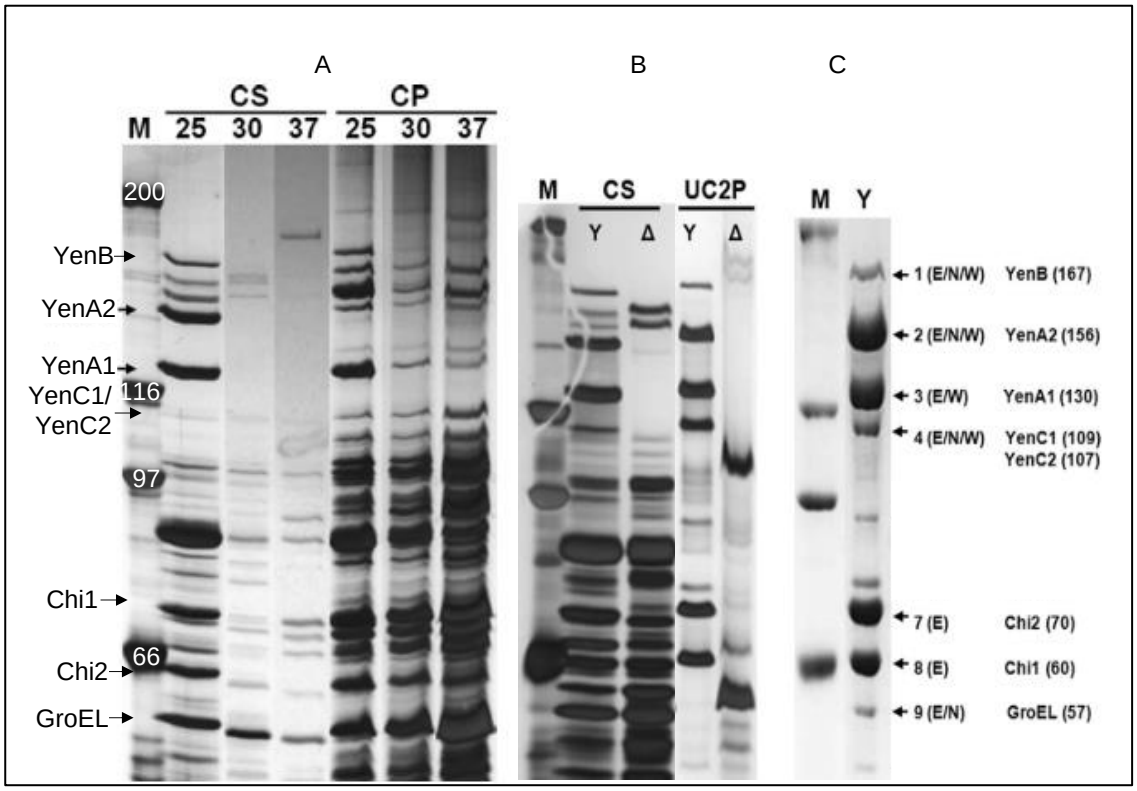


Figure 2.3: *Yersinia entomophaga* protein production and secretion at 25, 30 and 37 °C, including identification of Yen-TC components. Modified from Hurst et al. (2011). (A) SDS-polyacrylamide gel (silver stained) of Y. entomophaga MH96 culture supernatant (CS) and cell pellets (CP) grown at 25, 30 and 37 °C. Lane M contains Bio-Rad broad range protein standard. (B) SDS-polyacrylamide gel (silver stained) of MH96 (Y) and the ΔYen-TC mutant (Δ) CS and pellet after ultracentrifugation and application to a step gradient (UC2P) grown at 25 °C. (C) SDS-polyacrylamide gel (Coomassie brilliant blue stained) showing bands that were assessed with either LC-ESI-MS/MS (E), N-terminal sequence analysis (N) or western blot (W).

In the same study, MH96 was also shown to be highly virulent when injected directly into the hemocoel of *G. mellonella*, with median lethal dose (LD_{50}) of ~ 3 cells after four days, regardless of whether the insects were incubated at 25 or 37 °C. Unlike the *per os* challenge, the four-day LD_{50} was comparable to wild-type when *G. mellonella* was injected with the ΔTC (Yen-TC deficient strain) and incubated at 25 °C, which demonstrated that MH96 must deploy additional hemocoel-active factors, in addition to Yen-TC, during intrahemocoelic infection of *G. mellonella* (Hurst et al., 2015). Observations of green fluorescent protein (GFP)-tagged MH96 budding inside of hemocytes was observed at 24 hours using fluorescence microscopy. Further microscopic investigations using a Yen-TC-linked GFP-reporter strain, called YenA1, revealed temperature-dependent differences in cell morphology and density-

dependent fluorescence when injected into the hemocoel of *G. mellonella.* When insects were maintained at 25 °C, the Yen-TC GFP-reporter strain fluoresced at a minimum cell density of 2 x 10^7 CFU/ml and cells took on irregular forms (i.e., bulging, spherical and filamentous), while at lower cell densities the shapes of cells were smaller and more uniform in shape. At 37 °C, fluorescence occurred at a minimum cell density of ~3 x 10^9 CFU/ml but was more sporadic and weaker and the cells were smaller compared to 25 °C.

In both *in vitro* (Hurst et al., 2011b) and *in vivo* (Hurst et al., 2015) investigations described above, temperature-dependent effects on Yen-TC were observed and temperature shift from 25 to 37 °C reduced virulence of MH96 by *per os* challenge against insects due to reduced secretion/toxin production/*tc* gene expression. Therefore, one major goal of this thesis was to use *in vivo* RNA-seq to unravel temperature- and host-dependent regulatory networks potentially involved in thermoregulation of Yen-TC component genes in MH96 as well as experimentally determine the role of two putative regulators, Yen6 and Yen7 (co-located on PAI_{YE96} with Yen-TC component genes). A detailed introduction to RNA-seq is provided in Chapter 4 (see section 4.1.1).

2.4 Specific Objectives

By understanding the host- and temperature-dependent transcriptional responses of MH96, this thesis aimed to fill several important gaps in scientific knowledge as outlined below.

1. A general lack of genome-wide transcriptional investigations of entomopathogenic bacteria during insect infection (see Chapter 4, section 4.1.1 for more detailed review), and more specifically to identify the currently unknown hemocoelic VFs produced by MH96 during intrahemocoelic infection of *G. mellonella*.

The first data chapter of this thesis (Chapter 4) addressed these gaps by providing one (if not the first) *in vivo* transcriptome of an entomopathogenic bacteria therefore greatly increases our knowledge of how entomopathogenic bacteria, like MH96, respond during insect infection and also provided a proof of principle, that *in vivo* transcriptomics combined with functional enrichment can be used to effectively identify hemocoelic factors that respond to the host environment.

2. While the structure of MH96's main weapon, Yen-TC has been the focus of much research, there has not been any investigations into the thermo or host-dependent regulatory mechanisms governing production of Yen-TC by MH96 to date.

High resolution transcriptional insights into the host- and temperature-dependent effects on MH96 expression of Yen-TC is provided in Chapter 4 of this thesis, which also reports the identification co-expression of several gene clusters encoding toxins, exoenzymes or secretion systems, some of which are also known to be co-secreted with Yen-TC components by MH96.

The first molecular investigations into the direct regulatory linkages between *yen6, yen7* and Yen-TC production and secretion are provided Chapter 5 of this thesis. A key area of focus included in this data chapter was the determination of whether Yen6 is a thermoregulatory repressor of *yen7* and/or Yen-TC production and whether Yen7 was a thermoregulatory activator of Yen-TC production/secretion. These preliminary investigations identified the regulation of Yen-TC is complex, including transcriptional and post-transcriptional mechanisms.

3. First characterization of a novel LytTR-containing virulence regulator that responds to both temperature- and host-dependent cues during further implicating the interconnectivity of metabolism and virulence in pathogenic bacteria.

A key finding of this thesis was the importance of a Yen6 during intrahemocoelic infection of *G. mellonella* at 37 °C, which could be attributed to increased expression of *yen6* at 37 °C (reported in Chapter 4) and increased translation of Yen6 under *in vivo* conditions (reported in Chapter 5). Based on these findings Chapter 6 of this thesis reports the transcriptome of Δ*yen6* under *in vivo* conditions at 37 °C that revealed the Yen6 regulon, which was further validated by electrophoretic mobility shift assays (EMSA). These findings combined with phenotypic investigations of *yen6* impact on carbon utilization by MH96 supports Yen6 as a key regulator that potentially alters central carbon metabolism during oxidative stress. These findings contribute to a growing body of evidence supporting the entanglement of metabolism and virulence in pathogenic bacteria and expand the diversity of known LytTR-containing transcriptional regulators from pathogenic bacteria.

4. Horizontal gene transfer is a major driver in pathogenic bacteria evolution (more detailed review in Chapter 7, section 7.1.1) but limited knowledge of mobile genetic elements in the genome of MH96, and among the genomes of entomopathogenic bacteria is currently lacking.

The first report of an unstable genomic element from the genome of MH96 encoding a holin-endolysin secretion system and citrate fermentation cluster is provided in the final data chapter of this thesis (Chapter 7). Importantly, further bioinformatic and molecular investigations linked this element to secretion of Yen-TC and also shared regulatory elements suggesting cross-talk between PAI_{YE96} and the unstable element. These findings contribute to our understanding of the role of HGT in the evolution of entomopathogenic genome evolution.

The general molecular materials and methods provided in this chapter (Chapter 3) represent generic methods that were used to generate the data presented in this thesis. Some additional methods for more specific applications, such as *in vivo* RNA-seq, electrophoretic mobility shift or β-galactosidase assays for example, are provided within summary method sections of each data chapter, as required.

3.1 Bacterial strains, plasmids and growth conditions

Standard microbiological culture techniques, including sterile practices were carried out under Physical Containment level 2 as required by the AgResearch Containment and Transitional Facility Manual (Version 6.8), including the Invertebrate Containment Standard Operating Procedures (Version 1.0). All strains used in this study are listed in supplementary Table S1 (targeted mutagenesis) and Table S2 (*lacZ*-reporter strains). All plasmids used in this study are provided in supplementary Table S3 (targeted mutagenesis) and Table S4 (arabinose induction and lacZ-reporter). *Yersinia entomophaga* MH96 and *Escherichia coli* strains were routinely grown at 25 or 30 °C and 37 °C, respectively. All cultures were grown at 250 rpm in Luria Broth Base (Invitrogen) broth (LB broth) or on LB agar Miller (LB agar) (Merck), unless otherwise stated. Antibiotic concentrations used for *E. coli* derivative strains were ampicillin 100 μg/ml, chloramphenicol 30 μg/ml, kanamycin 50 μg/ml, spectinomycin 100 μg/ml and tetracycline 30 μg/ml. Antibiotic concentrations used for MH96 derivative strains were ampicillin or carbenicillin 400 μg/ml, chloramphenicol 90 μg/ml, kanamycin 50 μg/ml, spectinomycin 100 μg/ml and tetracycline 30 μg/ml.

3.2 Bacterial enumerations

Viable plate counts for bacterial enumerations were conducted using serial (ten-fold) dilutions series. Dilutions were made using phosphate buffer solution (PBS; 137 mM NaCl, 2.7 mM KCl and 10 mM phosphate, pH 7.4 at 25 °C; Sigma) or sterile MilliQ water and 50 μl of dilution were spread on agar plates using a disposable spreader, unless otherwise stated. Only plates containing between 30 and 300 colony forming

units (CFUs) were enumerated. Enumerations were conducted using LB agar plates unless otherwise specified. Optical density (OD) of culture was measured at 600 nm using Smart Spec Spectrometer (BIO-RAD).

3.3 DNA Manipulation

3.3.1 Plasmid DNA isolation by alkaline lysis

Plasmid DNA isolation by alkaline lysis (or 'minipreparation') was routinely carried to prepare plasmids for diagnostic restriction, restriction/ligation for cloning and transformation as well as for polymerase chain reaction (PCR) template. Cells from 1 mL of overnight cultures were pelleted by centrifugation and resuspended in 150 µl of 25 mM Tris-HCl (pH 8.0), 10 mM ethylenediaminetetraacetate (EDTA) and 0.9 % glucose and then mixed with 200 µl of 200 mM NaOH and 1 % SDS in MilliQ water and incubated at 37 °C for 4 min. Next 150 µl of 3M sodium acetate (pH 5.2) was added and mixed vigorously to shear genomic DNA and the mixture was incubated on ice for at least 5 min. The mixture was centrifuged for 10 min at 16,100 x g and the supernatant decanted into 1 ml of 100 % ethanol and mixed by inverting the tube five times. The DNA was pelleted by centrifugation at 16,100 x g for 5 minutes and the ethanol was removed. The DNA pellet was air-dried for 30 min at 37 °C and then resuspended in MilliQ water.

3.3.2 Polymerase chain reaction for targeted mutagenesis

All primers used to generate constructs for targeted mutagenesis by PCR are provided in supplementary Table S5. All methods related to commercial kits were completed as recommended by the manufacturer unless otherwise stated. Genomic DNA template from MH96 and derivative strains were purified from 1 ml overnight cultures using PrepMan Ultra Sample Preparation Reagent (Applied Biosystems) and diluted 100-fold. Plasmid DNA template for antibiotic markers was isolated by minipreparation from 1 ml overnight culture and diluted 1,000-fold.

PCR for antibiotic markers, sequence validation and construct synthesis (not including overlap-extension PCR) were performed using Platinum *Taq* DNA polymerase (Invitrogen) using a standard two-step program: 95 °C for 3 min; then 95 °C for 25 s, 55 °C for 25 s, 72 °C for 1 kb/min repeated 5 times; then 95 °C for 25 s, 50 °C for 25 s, 72 °C for 1 kb/min repeated 30 times; then 72 °C for 10 min on either a BioRad C1000 Touch Thermal Cycler or Eppendorf Mastercycler Gradient. Optimization using gradient PCRs were performed as required using the above-mentioned program but testing a range of two-step annealing temperatures between 50/55 and 60/65 °C, as required.

3.3.3 Constructs for homologous recombination

Targeted mutagenesis constructs were made either using naturally occurring restriction sites found within genes of interest (GOI) or using overlap-extension PCR. Amplicons of approximately 2 kb were generated containing these GOI, which were used as a base for the homologous recombination construct by gene interruption (further described in section 3.3.7 below and Chapter 5, section 5.2.1). Most mutagenesis constructs were generated by over-lap extension PCR using Phusion High-Fidelity DNA polymerase (New England Biolabs (NEB)) to generate error-free overlap-extension PCR products (further described in section 3.3.4 below and Chapter 5, section 5.2.1).

3.3.4 Overlap-extension PCR

All amplicons used to synthesize overlap-extension PCR constructs (except for some selective markers) were generated using Phusion High-Fidelity DNA polymerase (NEB). Custom annealing temperatures (Tm) for each primer pair were calculated using the NEB Tm Calculator (v.1.10.3; http://tmcalculator.neb.com). Regions of approximately 1 kb directly 5′ and 3′ to the GOIs were amplified to include 21 bp overlap for selectable markers using the following PCR program: 98 °C for 30 s; then 98 °C for 10 s, Tm for 30 s, 72 °C for 1 kb/30 s repeated 5 times; then 98 °C for 30 s; then 98 °C for 10 s, (Tm – 5 °C) for 30 s, 72 °C for 1 kb/30 s repeated 25 times; then 72 °C for 10 min). Overlap-extension PCR was then used to fuse the specified selectable marker

between the 5′ and 3′ regions using the following PCR protocol with nested primer pairs:

Primerless reactions:

Amplicons to be fused (i.e., 5′ and 3′ regions flanking GOI regions and selectable marker) were mixed in approximately equal molar with dNTPs, polymerase and high-fidelity (HF)-buffer (total reaction volume of 20 µl). The primerless cycle consisted of 98 °C for 30 s; then 98 °C for 10 s, Tm for 30 s, 72 °C for 1 kb/30 s repeated 15 times; then 72 °C for 10 min.

Overlap-extension reaction:

Next 5 µl of the above mixture was used as template for the overlap-extension reaction, containing dNTPs, polymerase, HF-buffer and nested primers, total reaction volume of 50 µl. The above PCR program was repeated 25 times, followed by a final 72 °C elongation.

Finally, overlap-extension products were purified using the High Pure PCR Product Purification Kit (Roche) resuspended in water (instead of TE buffer supplied with the kit) and then A-tailed using Klenow Fragment (3′ → 5′ exo-) (NEB) in a 50 µl reaction volume. Following incubation at 37 °C for 30 min, the A-tailed constructs were purified again using the High Pure PCR Product Purification Kit.

3.3.5 Cloning of targeted mutagenesis constructs

To facilitate cloning of targeted mutagenesis constructs into suicide vectors, the constructs were first ligated into pGEM-T vector and incubated overnight at room-temperature. The bacterial strain *E. coli* (DH10B) was grown overnight in 3 ml LB broth. A 250 ml conical flask containing 50 ml LB broth was inoculated with 1 % overnight culture and grown for approximately 1 hour and 45 minutes to an OD_{600} between 0.6 – 0.7. The culture was incubated on ice for 20 minutes then centrifuged at 4 °C for 10 minutes at 2,268 x g. The cells were resuspended in 25 ml ice-cold 0.1 M $CaCl_2$. Centrifugation was repeated to pellet the cells again, which were next

resuspended in 1 mL ice-cold 0.1 M $CaCl_2$ and incubated on ice for at least one hour prior to transformation.

Ligation reactions (1-3 μl) were incubated on ice for 5 minutes and then transformed into 100 μl chemically competent DH10B by a 30 second heat-shock in a 42 °C heat bath. Transformed cells were incubated in 1 ml broth at 37 °C and then plated on selective media. Clones were screened using ampicillin and either spectinomycin or kanamycin for overlap extension PCR products. Amplicons containing GOIs *rovA* and *vipB*, were ligated into pGEM-T and transformed into DH10B and clones were screened on ampicillin plates containing 100 ug/ml X-gal (5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside) for blue/white selection (pGEM containing an insert in the multiple cloning site (MCS) results in an interruption of the *lacZ* gene, so colonies are white instead of blue on X-gal as the strain will not produce β-gal).

3.3.6 Validation of targeted mutagenesis constructs

Clones carrying plasmids with correct insertion sizes were confirmed by restriction enzyme digestion (often with *EcoRI* flanking the TA cloning site pGEM-T) and visualization on 1 % agarose gel by electrophoresis and such 'diagnostic restriction digestions' were carried out in a total volume of 10 μL using NEB enzymes and reagents. Once correct sized inserts were confirmed by digestion, plasmids were isolated from 3 ml overnight culture using the High Pure Plasmid Isolation Kit (Roche) and the inserts were sequenced using M13F/M13R primers by Macrogen Korea. The sequence data were validated in Geneious (v.R10) by comparison of sequence and base-call quality ab1 chromatograms against the MH96 reference genome to ensure correct nucleotide sequence of all inserts.

3.3.7 Gene interruption by restriction and ligation of selective marker

The selective marker spectinomycin was used to interrupt GOIs *rovA* and *vipB* by naturally occurring *ClaI* and *BgIII* restriction sites, respectively. Restriction digests were carried out using NEB enzymes and reagents in 100 μl total reaction volumes and incubated at 37 °C for 1.5 hours. The DNA was then precipitated by ethanol

precipitation; briefly the DNA were incubated for two min at room temperature with two volumes 100 % ethanol and one-tenth volume 3 M sodium acetate (pH 5.2), then DNA were pelleted at 15,700 x g for 5 min, decanted and then dried for ~ 15 - 20 minutes at 37 °C and resuspend in 7 µl MilliQ water. Next, 2 µl of the digested DNA was visualized on 1 % agarose gel by electrophoresis. Digested PCR products (insert DNA) were mixed with linearized plasmid DNA in an approximate ratio of 1:3 in a total reaction volume of 10 µl. Ligations were incubated overnight at room temperature with T4 DNA ligase (Invitrogen) and supplied buffers and stored at -80 °C (as required). The ligation reaction was transformed into competent DH10B *E. coli* cells as described above in Section 3.3.5 and screened using ampicillin and spectinomycin. Clones were confirmed by diagnostic restriction as described above (section 3.3.6).

3.3.8 Ligation of suicide plasmids

Following sequence validation, pGEM-T with targeted mutagenesis constructs were digested (often by *EcoRI* from the pGEM-T MCS) to release the construct, which were then ligated into linearized suicide vectors pJP5608 (kan^R) or pJP5603 (tet^R) by ligation using T4 DNA Ligase (Invitrogen) as previously described in Section 3.3.7.

3.3.9 Electroporation of *E. coli*

The suicide plasmids pJP5603 and pJP5608 contain the R6K-based origin of replication, requiring host strain to supply *pir*-encoded π protein of plasmid R6K for replication (Penfold and Pemberton, 1992). To maintain the suicide plasmids, ligation reactions containing suicide vector and targeted mutagenesis constructs (from pGEM-T) were transformed into *E. coli* EC100D (pir^+) by electroporation.

E. coli strain EC100D was grown overnight in 3 ml LB broth. A 250 ml conical flask containing 50 ml LB broth was inoculated with 1 % overnight culture and grown for approximately 1 hour and 45 minutes to an OD_{600} between 0.6 – 0.7. Electrocompetent cells were made at either room temperature or 4 °C (Tu et al., 2016). The culture was incubated for 20 min then centrifuged at 2,604 x g to pellet the cells and resuspended in 50 ml MilliQ water. The cells were pelleted and resuspended four more times with

increasing g-forces (3,214 x g, 3,889 x g, 4,177 x g and 4,629 x g) and decreasing volumes (40 ml, 30 ml, 20ml and ~ 800 µl), respectively. MilliQ water containing 10 % glycerol was used for the final three resuspension steps and single reaction aliquots were stored in pre-chilled microcentrifuge tubes at -80 °C.

Transformation of electro-competent cells was carried out using standard methods described by (Dower et al., 1988), except some transformations were done at room-temperature, depending if the cells were also made at room-temperature or not (Tu et al., 2016). A 1 µl aliquot of ligation mix was added to 40 µl electro-competent cells and transferred into an electro-cuvette. Cells were subjected to 2.5 kV charge for approximately 5.2 microseconds using a *E. coli* Pulser (BioRad, Hercules, CA). Cells were elaborated in 1 ml LB broth at 37 °C for 1 hour.

Recovered cells were plated on selective media containing either tetracycline or kanamycin to select for the suicide plasmid and either the kanamycin or spectinomycin to select for mutagenesis construct. Resistant clones were screened for sensitivity to ampicillin (to ensure loss of pGEM-T cloning vector) and validated for correct insert size by digestion and visualization on agarose gel by electrophoresis as previously described (Section 3.3.5).

3.3.10 Chemical transformation of *E. coli* ST18

Next, validated suicide plasmids carrying targeted mutagenesis constructs were chemically transformed into an *E. coli* donor strain (ST18). ST18 is a derivative of S17*λpir* that is deficient for the *hemA* gene, rendering it defective in tetrapyrrole biosynthesis (Thoma and Schobert, 2009). Subsequently, for ST18 to survive, it must be grown in media containing 5-aminolevulinic acid (ALA), which allows counterselection on media lacking ALA.

E. coli ST18 were grown overnight in 3 ml LB broth supplemented with ALA 50 µg/ml. A 250 ml conical flask containing 50 ml LB broth was inoculated with 1 % overnight culture and grown for 1 h and 45 min until OD ~ 0.6 – 0.7 was reached. Cells were incubated at 4 °C for 1 hour. Cells were pelleted by centrifugation at 2,219 x g for 10 min at 4 °C and then resuspended in 10 ml Solution A and incubated on ice for 20

minutes. The cells were pelleted again by centrifugation at 2,219 x g for 10 min at 4 °C and resuspended in 1.5 ml of Solution A with glycerol. Single-reaction aliquots (50 μl) were stored in pre-chilled tubes at -80 °C.

Solution A	9.9 ml 1 M $MnCl_2$ 49.5 ml 1 M $CaCl_2$ 198 ml 50 mM 2-(N-morpholino) ethanesulfonic acid (MES) 742.6 ml MilliQ water
Solution A with glycerol	10 ml 1 M $MnCl_2$ 50.1 ml 1 M $CaCl_2$ 200.5 ml 50 mM MES 300.6 ml 50 % Glycerol 438.6 ml MilliQ water

Validated mutagenesis plasmids were then transformed into ST18 by chemical transformation as previously described in Section 3.3.5, except heat shock was administered for two minutes and cells were recovered in 1 ml LB broth supplemented with ALA 50 μg/ml.

3.4 Targeted mutagenesis by homologous recombination

Targeted mutagenesis by homologous recombination was achieved by conjugation of ST18 derivatives to MH96 at 30 °C for 6 – 7 h with ALA 50 μg/ml supplemented LB agar. Post-conjugation, transconjugants were then grown for 24 h on LB agar without ALA and appropriate antibiotics to select for genomic integration of the marker. All putative mutants were then screened for loss of suicide plasmid on media containing either kanamycin or tetracycline. Genomic template was extracted from putative mutants and screened by PCR, as described in section 3.3.2, using validation primers (external to the region combined) confirming correct integration. These PCR products were shipped to Macrogen, purified and then sequenced to validate mutant strains reported here. The sequence data were validated as described in section 3.3.5 to confirm proper integration of the selectable marker and correct sequence covering both points of recombination.

3.5 Protein visualization and identification

3.5.1 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

For visualization of crude protein extracts, wild-type and mutant strains were grown in 3 ml LB broth for approximately eight hours with appropriate antibiotics. Overnight cultures were diluted to 2 % in flasks containing 50 ml LB broth and grown at either 25 or 37 °C with 200 rpm shaking overnight (~ 18 hours) with appropriate antibiotic. In the morning, 1 ml culture were pelleted by centrifugation for 10 minutes at 8,000 x g and the cell supernatant was filter sterilized (0. 22 μm). Serial dilutions of the overnight culture were made in PBS (Sigma) and CFUs were enumerated to confirm consistent cell density among the overnight cultures.

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to (Laemmli, 1970) on one-dimensional slab gels (1.0 mm thick) containing 0.1 % SDS. Cell pellet (CP) were diluted 1:10 in water and 25 μl of diluted CP or undiluted cell supernatant (CS) samples were denatured at 95 °C for five minutes with 8 μl of 4 x loading buffer (125mM Tris/HCl (pH 6.8), 20 % glycerol (v/v), 3.8 % (w/v) SDS, 0.1 % (w/v) bromophenol blue and 10 % (v/v) 2-mercaptoethanol). Mid-range proteins were separated by size on hand-cast or Novex pre-cast tris-glycine with 10 % polyacrylamide under reducing conditions with 200 V for 55-60 minutes in SDS-PAGE running buffer (25 mM Tris/HCl, 200 mM glycine, 0.1 % (w/v) SDS, pH 8.3). Precision Plus Protein standard (Bio-Rad) was used as a marker. Smaller sized denatured proteins were mobilized on Novex pre-cast 10 % tricine (Life Technologies) under reducing conditions (added 0.4 μl β-mercaptoethanol to the Novex tricine sample buffer) by applying 125 V for 75 - 90 min in Novex 1 X tricine running buffer. Separated proteins were stained with silver as described by (Blum, H., Beier, H., Gross, 1987) but with the following modified incubation times: fixative: 30 min; methanol washes x 3: 10 min each; 0.02 % sodium thiosulphate pretreatment: 1 min; distilled water washes x 3: 20 sec each; stain: 10 min; distilled water washes x 2: 20 sec each; development: 1 – 10 min; stop development: 10 min; and then placed in distilled water for longer-term storage. Visualization of CS and CPs from mutants demonstrating an observed protein-level phenotype was repeated at least twice to confirm the phenotype

by SDS-PAGE, except for Δ*cpbA*. The size and location of Yen-TC components visualized by SDS-PAGE were inferred from previous analysis performed in the Hurst lab, including protein identification using a combination of approaches including liquid chromoatogrpahy-electrosparay ionization ion trap-tandem mass spectromoetry (LC-ESI-MS/MS), N-terminal sequences analysis and/or western immunoblot approaches (Hurst et al., 2011) and unpublished data (supplemental Figure S18).

3.5.2 Liquid chromatography-tandom-mass spectrometry

Specific unknown protein samples for LC-ESI-MS/MS analysis were separated by SDS-PAGE as described above and stained with 0.1 % (w/v) Coomassie Brilliant Blue R-250, 45 % (v/v) methanol and 10 % (v/v) acetic acid for 30 – 60 min at room temperature (approximately 22 °C). The gels were destained with MilliQ water with 40 % (v/v) methanol and 10 % (v/v) acetic acid until al bands were visable. The gel was stored in MilliQ water over-night at room-temperature (23 °C) and excised bands were stored at -20 °C.

LC-ESI-MS/MS analysis were carried by Ancy Thomas and Evelyne Maes of the AgResearch Proteomics Platform. De-stained proteins of interest (POIs) were reduced with 0.1 M tris(2-carboxyethyl) phosphine (Fluka Chemie, GmbH, Buchs, Germany), alkylation with 20 µl of 0.15 M iodoacetamide (Sigma, St. Louis, MO, USA) and digested for 18 hours with 1 µg of TPCK-trypsin (Promega, Madison, WI, USA) in presence of 10 % Acetonitrile. After digestion the peptides were dried and resuspended in 50 µl of 0.1 % FA prior to injection on the mass spectrometer.

LC-ESI-MS/MS was performed on a nanoAdvance UPLC coupled to a Bruker amaZon speed ion ETD trap mass spectrometer equipped with a CaptiveSpray ion source operated at 1900 V. Five µl of sample was loaded on a C18 nano trap column (Bruker, 75 µm x 2 cm, C18AQ, 3 µm particles, 200 Å pore size). The trap column was then switched in line with the analytical column (Bruker Magic C18AQ, 100 µm x 15 cm C18AQ, 3 µm particles, 200 Å pore size). The column oven temperature was 50 °C. Elution was with a multipart linear gradient from 2 to 45 % B in 70 min. at a flow rate of 800 nL/min. Total run time including column re-equilibration was 90 min. Solvent A

was water with 0.1 % formic acid (FA) and 2 % acetonitrile (ACN); solvent B was ACN with 0.1 % FA and 2 % water. Auto MS/MS data acquisition was performed with 10 precursors selected in the m/z 350-1200 range during each cycle per second.

Proteins were identified by using the database search tool Peaks DB (v8.5) (Zhang et al., 2012). The raw data were refined by a built-in algorithm. The proteins/peptides were identified with the following parameters: a precursor mass tolerance of 0.3 Da and fragment mass tolerance of 0.6 Da were allowed, the Uniprot_*Yersinia entomophaga* database (v2018.08) was used, semi-trypsin was specified as digestive enzyme and up to 2 missed cleavages were allowed. Both oxidation (M) and deamidation (NQ) are chosen as variable modifications, carbamidomethylation (C) is defined as fixed modification and a maximum of 3 post translational modifications per sample were allowed. False discovery rate (FDR) estimation was made based on decoy-fusion and FDR was set to 1%.

3.6 *Galleria mellonella* intrahemocoelic bioassay

MH96 and its derivatives were grown overnight in 3 ml broth at 30 °C, with shaking at 250 rpm. A series of five dilutions were made with PBS with the aim of covering a range of dosages from 0 – 10 cells (1:20,000,000 to 1:400,000,000) and placed on ice.

Larval *Galleria mellonella* (Biosuppliers, Auckland) were maintained on a fresh diet of honey, glycerol, plain baby rice cereal (Farex, New Zealand) and live granular yeast (Pam's, New Zealand) for no longer than one week. Healthy larvae of similar lengths (~ 20 mm) and weights (0.15 – 0.25 g) were selected and immobilized on ice for up to 1 h.

Larvae were injected with 10 µl inoculum containing MH96 strain culture dilutions or PBS (control) just below their third right leg with a 30-gauge needle on a 1 mL tuberculin syringe (BD) using a microinjector. Ten larvae were injected per dilution/strain and incubated at 25 or 37 °C. The larvae were checked daily for color changes, web production and mortality over a seven-day period. The bioassay was repeated three times per strain and temperature.

Median lethal dose (LD_{50}) for each strain/temperature combination were determined by fitting day-four mortality rates with binomial-type generalized linear models

(including logit link function) and the LD_{50} and its standard error were estimated using the R package MASS (Crawley, 2013) and plotted using ggplot2 (Wickham, 2016) in R (R Core Team, 2017). Dose responses were corrected for control mortality (~ 1 % control mortality for all bioassay data combined).

Chapter 4. *In vivo* transcriptome provides insights into putative host- and temperature-dependent response of entomopathogenic bacterium, *Yersinia entomophaga* MH96 during infection of *Galleria mellonella*

4.1 Introduction

4.1.1 Overview - *In vivo* transcriptomics of bacterial entomopathogens

Transcriptomics (or 'RNA-seq') refers to genome-wide characterization of gene expression by high-throughput sequencing of cDNA and over the last decade, transcriptomics has been used to study diverse bacteria in a variety of growth conditions. The use of *in vivo* transcriptomics in the study of pathogenic bacteria has been particularly useful for illuminating the molecular interplay between host and pathogen (Avican et al., 2015; Damron et al., 2016) as well as discovery of virulence-associated small non-coding RNA (sRNA) regulatory networks (Baddal et al., 2016; Barquist et al., 2016; Westermann et al., 2016; Yan et al., 2013) and riboregulators (Nuss et al., 2017). Initially, transcriptomic investigations of bacterial pathogens from within the host environment (*'in vivo'*) has been limited by a lack of methods to enrich for bacterial RNA from mixed samples containing substantially greater amounts of host-derived RNA (Westermann et al., 2017). More recently, the availability of dual host/microbe rRNA depletion kits (Baddal et al., 2016; Damron et al., 2016; Mavromatis et al., 2015; Rienksma et al., 2015; Thänert et al., 2017), fluorescence-activated cell sorting (Avraham et al., 2015; Westermann et al., 2016), laser capture microdissection (Vannucci et al., 2013) and/or bacterial-specific enrichment protocols (Bent et al., 2013b, 2013a; Humphrys et al., 2013; Mavromatis et al., 2015) has made dual-seq possible for pathogenic bacteria from within a variety mammalian host tissues and cell lines.

While methods to conduct *in vivo* transcriptomics are readily available, to our knowledge, *in vivo* transcriptomics has not yet been reported for an important arthropod pathogen or any other pathogenic bacteria that use arthropods as vectors. Such studies could potentially yield novel biopesticide agents or critical insights into possible control strategies for vector-borne diseases. As with any dual-seq study, the biggest hurdle is obtaining enough bacterial transcriptional signal from within the host cells/tissue, especially at early stages of infection when bacterial cell densities may be

relatively low. Another challenge is achieving adequate depletion of host insect rRNA, for which unlike for mammalian-host systems, there are no commercial kits currently available. Lastly, while a poly(A) enrichment strategy has been reported as a means to subtract insect host mRNA from mixed RNA samples (Kumar et al., 2016), this approach may result in confounding biases due to presence of polyadenylated messengers in bacteria (Mohanty and Kushner, 2006) or the presence of laterally-acquired genes in the host genome (Kumar et al., 2012).

Due to such technical challenges, there has been comparatively fewer reports of genome-wide expression using RNA-seq for entomopathogenic bacteria during infection of an insect or another invertebrate host compared to pathogens infecting mammalian hosts. To date, there has only been one reported study on the *in vivo* transcriptome of a bacteria that can infect insects, which investigated plant-pathogenic *Dickeya dadantii* (not normally known as an entomopathogen) during infection of the aphid host, *Acyrthosiphon pisum* (Costechareyre et al., 2013). In this study, no toxin-like genes were found to be highly expressed in the aphid compared to *in vitro* growth conditions, and all of four insecticidal toxins genes encoded by *D. dadantii* had higher expression *in vitro*. The study also found the expression of genes related to response to antimicrobial peptides (AMPs), efflux systems/transporters, motility (chemotaxis and flagella) and stress response (acid shock, anaerobiosis, a redox-sensitive transcriptional regulator gene (*soxR*) and Fe-S cluster assembly) responded to the *in vivo* growth conditions compared to *in vitro* (Costechareyre et al., 2013).

Other *in vivo* transcriptomics investigations of bacteria associated with invertebrate hosts has primarily focused on non-pathogenic associates, such as obligate endosymbiont *Wolbachia pipientis* infecting fruit flies and filarial nematodes (Kumar et al., 2012, 2016; Luck et al., 2017; Woolfit et al., 2015), obligate symbionts of agricultural pest Glassy-Winged Sharpshooter, "*Candidatus* Sulcia muelleri" and "*Candidatus* Baumannia cicadellinicola" (Bennett and Chong, 2017), citrus pathogen "*Candidatus* Liberibacter solanacearum" that is vectored by psyllid *Bactericera cockerelli* (Ibanez et al., 2014), gut-associated symbiont *Burkholderia insecticola* of the bean bug *Riptortus pedestris* (Ohbayashi et al., 2019) and obligate endosymbiont *Wigglesworthia* (*Wi.*) *glossinidia* of the important tsetse fly disease vector (Vigneron et al., 2017). Obligate

intracellular bacteria such as *W. pipientis* are particularly challenging with respect to enrichment for RNA-seq and most-definitely require specific enrichment approaches, such as host insect rRNA depletion and/or subtraction of polyadenylated mRNAs (Kumar et al., 2012, 2016) or bacterial mRNA enrichment (Luck et al., 2017). So far, the *in vivo* transcriptome investigations of *W. pipientis* within both fly and nematode hosts has identified putative ncRNAs (Woolfit et al., 2015) as well as the expression of genes from the *W. pipientis* genome that have been laterally integrated into the host genome (Kumar et al., 2016). Among glassy-winged sharpshooter endosymbionts, *Baumannia* and *Sulcia*, genes for the molecular chaperones DnaK and GroESL as well as genes related to nutritional synthesis (i.e., essential amino acid synthesis and B vitamins) were found to be expressed while in the host (Bennett and Chong, 2017). Similar to *Baumannia* and *Sulcia*, the obligate endosymbiont of tsetse flies, *Wi. glossinidia* was also found to over-express the molecular chaperone GroEL and genes related to transport and biosynthesis of B-vitamins (Vigneron et al., 2017). The transcriptome of *B. insecticola* from the midgut of the bean bug revealed that the symbiont produces all essential amino acids and B vitamins which are lacking in the host diet food and upregulates diverse transporters and metabolic pathways related uptake of sugars (i.e., ribose and rhamnose) and sulphur compounds (i.e., sulphate and taurine) and assimilation of host-derived nitrogenous waste products (i.e., allantoin and urea) (Ohbayashi et al., 2019). Recently, dual-seq was used to identify *Escherichia coli* genes with higher expression within the gut of *Caenorhabditis* (*Ca.*) *elegans* compared to *in vitro* growth conditions, which included several genes related to two-component regulatory systems (TCRS), lipopolysaccharide (LPS) biosynthesis and antibiotic resistance (Chan et al., 2019).

4.1.2 Host- and temperature dependent gene expression in *Photorhabdus* and *Xenorhabdus* species

While entomopathogenic bacteria have not been explored using dual-seq or *in vivo* seq approaches, *in vivo* expression of netmatode symbionts/entomopathogenic *P. temperata* and *X. koppenhoeferi* was achieved using selective capture of transcribed sequences (SCOTS) (An et al., 2009). The SCOTS approach aims to enrich for genes preferentially

expressed under specific growth conditions using subtractive hybridization of cDNA against biotinylated DNA followed by polymerase chain reaction (PCR) amplification, cloning and identification of cloned genes by sequence similarity (Wang et al., 2014b). In this example, An et al. (2009) used the SCOTS method to identify genes expressed *in vivo* by *P. temperata* and *X. koppenhoeferi* during infection of white grub, *Rhizotrogus majalis*. In this method *in vitro* and *in vivo* cDNA was hybridized against biotinylated bacterial genomic DNA that was pre-blocked with rRNA and transcripts explicitly produced *in vivo* where identified by additional hybridization against biotinylated bacterial genomic DNA, pre-blocked with the *in vitro* cDNA. Genes specifically produced by either *P. temperata* or *X. koppenhoeferi* were identified using a similar subtractive hybridization approach. From these investigations, hemolysin-related proteins and insecticidal TC genes were identified as expressed *in vivo* by both entomopathogens, while type 3 secretion system (T3SS) component genes were also identifed *in vivo* from *P. temperata* suggesting delivery of effectors via T3SS may be important in virulence in this species. Notably the work identified numerous host-induced stress response genes in both pathogens, including genes related to glutathione biosynthesis (important detoxification mechanism), DNA chaperone *dnaK* and well known stress repsonse genes *surA* (outer-membrane protein folding chaperone) and *uspB* (universial stress protein) (An et al., 2009).

An *in vitro* transcriptome study of opportunistic human pathogen/nematode symbiont/entompathogenic *P. asymbiotica* characterized genes with temperature-dependent responses at 28 and 37 °C. In this work very few primary virulence factors specifically repsonded to different temperatures, but a noticeable temperature-dependent effect was observered for genes related to carbohydrate and nitrogen metabolism (Mulley et al., 2015). At 37 °C several genes related to carbohydrate and nitrogen metabolism were reduced and several genes related to metabolism of peptides and amino acids were elevated, leading the authors to suggest *P. asymbiotica* may have evolved a temperature-dependent 'nutritional virulence' strategy within the human host. In this same study, the authors were able to identify proteins with temperature-dependent abundances using 2D-Fluorescence difference gele electrophoresis and LC-MS/MS. Using this approach, *P. asymbiotica* produced greater amounts of the molecular chaperones (GroEL, GroES, DnaK and ClpB), as well as

proteins related to oxidative stress response, quorum sensing, LPS biosynthesis, iron acquisition and a secreted asparaginase at 37 °C, while at 28 °C, a greater proportion of proteins with metabolic functions were identified. Such temperature-dependent metabolic shifts were confirmed using the Omnilog phenotype microarray system, which confirmed *P. asymbiotica* can survive on far greater diversity of sole carbon sources at 28 compared to 37 °C, which validated the results of the transcriptome and proteomics. Temperature was also identified as causing noticeable effects on utilization of amino acids and other sole sources of nitrogen as well as osmotic and pH tolerance (Mulley et al., 2015). Unlike, *Y. entomophaga* (MH96), *P. asymbiotica* was shown to be avirulent against *M. sexta* by low dose (1,000 cells) injection when incubated at 37 °C (Mulley et al., 2015), but both MH96 and *P. asymbiotica* are known to be highly virulent against insects at lower temperatures of 25 and 28 °C, respectively. Similar to the Mulley et al. (2015) study, this thesis also combine genome-wide changes at transcriptome and phenotype microarry data from MH96 under 25 and 37 °C to to further characterize nutritional strategies used by MH96 during infection.

4.1.3 Host- and temperature-dependent gene expression in pathogenic *Yersinia* species

While not the main focus here, dual-seq investigations on human pathogenic *Yersinia* species have been undertaken recently an *in vivo* transcriptome study of *Y. pseudotuberculosis* comparing samples from early and persistant stages of infection within mouse cecum identified characteristic expression signatures for each stage of infection. The pYV-encoded T3SS components and other virulence factors of *Y. pseudotuberculosis* were highly expressed only during early infection, and genes related to anaerobiosis, motility and defense against oxidative and acid stress were found to be more highly expressed during the persistant stage (Avican et al., 2015). In a different study, *in vivo* transcriptome analysis of *Y. entoercolitica* during extracellular and intracellular interactions with murine macrophages was used to identify Ysa T3SS, Yts2 T2SS and *tad* locus endoding a type IVb pilus, as potentially important virulence and survival within macrophages (Bent et al., 2015).

Generally temperature is widely known as a key environmental cue regulating virulence factors in pathogenic species of *Yersinia* and more broadly among pathogenic bacteria (Konkel and Tilly, 2000; Lam et al., 2014; Ono et al., 2005). Among *Yersinia species*, the nucleoid-associating *Yersinia* modulating protein, YmoA, and histone-like nucleoid-structuring (H-NS) protein are known to be key global regulators that are influenced by temperature (Madrid et al., 2007; McFeeters et al., 2007; Nieto et al., 2002) and act as important xenogeneic silencers (Navarre et al., 2007) that are involved in virulence regulation (Cornells et al., 1991; Madrid et al., 2001). Therefore, H-NS and YmoA would be expected to play a thermoregulatory role in MH96 as well, but were not the focus of this thesis. Despite attempts, a *Yersinia* spp. with a H-NS mutation has yet to be reported (including MH96) and a MH96 mutant deficient for *ymoA* was the target of a different research project.

The effect of temperature on genome-wide transcription has been previously investigated in *Y. pseudotuberculosis* using a tissue dual-seq approach (Nuss et al., 2017). This study revealed similar expression profiles of pYV-encoded virulence genes at 37 °C in both *in vitro* and *in vivo* conditions compared to lower temperatures, suggesting temperautre is a primary regulatory driver of pYV gene expression within the mammalian host. Another key finding of this study was higher *in vivo* expression of genes related to T3SS, iron acquisition and neutrifil- and nutritional stress-induced response were identified in lymphatic tissues during infection by *Y. pseudotuberculosis* (Nuss et al., 2017).

4.1.4 Advances in insect infection model – *Galleria mellonella*

The larva of the greater wax moth *G. mellonella* is emerging as a useful model host in the study of microbial infection and innate immunity (Champion et al., 2016; Desbois and McMillan, 2015; Tsai et al., 2016). Compared to traditional murine models, *G. mellonella* offers may advantages, including reduced husbandry, storage and training requirements, limited ethics concerns and ease of availability for use in virulence testing. Compared to other invertebrate models, such as *Caenorhabditis elegans* or *Drosophila melanogaster*, *G. mellonella* can be maintained at a range of temperatures, including that of mammalian hosts (37 °C). *G. mellonella* has been reported to be a

suitable model organism for virulence testing with a wide range of pathogenic bacteria including (but not limited to) *Listeria monocytogenes*, *Y. pseudotuberculosis*, *Campylobacter jejuni*, *Acinetobacter baumannii* and *Salmonella* (*Sa.*) *enterica* serovar Typhimurium (Champion et al., 2009; Kamal et al., 2011; Mukherjee et al., 2010; Peleg et al., 2009; Viegas et al., 2013).

G. mellonella can also be considered a good model host for investigation of human pathogens as it possesses an innate immune system, consisting of cellular and humoral components, some of which are considered relatively conserved between both insects and mammals (Champion et al., 2009; Desbois and McMillan, 2015; Junqueira, 2012; Mukherjee et al., 2010; Senior et al., 2011). The hemolymph of *G. mellonella* contains six different kinds of hemocytes that play various roles in immunity against pathogens and parasites (Boman and Hultmark, 1987). At least two different types of hemocytes (specifically plasmatocytes and granulocytes) are involved in phagocytosis, encapsulation and nodule formation (Pech and Strand, 1996; Tojo et al., 2000) and could be considered analogous to human macrophages and neutrophils. For example, the phagocytic cells of *G. mellonella* were found to produce a protein sharing homology to human calreticulin, which can also be found in the plasma membrane of neutrophils (Choi et al., 2002). Furthermore, investigations of the respiratory burst process in *G. mellonella* hemocytes also found conservation with human macrophages, since both share a number of homologous components, including the NADPH oxidase complex (Bergin et al., 2005) and translocated proteins involved the activation of the oxidative burst enzymatic cascade (Renwick et al., 2007).

In addition to the cellular component of the *G. mellonella* innate immune system, a range of different humoral factors are found in the hemolymph, including AMPs, pattern recognition molecules and reactive oxygen species (Tsai et al., 2016; Wojda, 2017). Some humoral components produced by *G. mellonella* have been studied in detail due to high homology with important innate immune factors produced by the mammalian innate immune system. For example many insects, including *G. mellonella*, produce an exchangeable lipid-binding protein called apolipophorin III (apoLp-III) that is present in the hemolymph and has been shown to play multiple roles in immune function, including pattern recognition, detoxification of bacterial endotoxins,

increased superoxide production by hemocytes and enhancement of other antimicrobial processes (Dunphy and Halwani, 1997; Niere et al., 1999; Park et al., 2005; Zdybicka-Barabas et al., 2013). ApoLp-III is homologous to the human apolipoprotein E, which has been shown to protect against LPS-induced septic shock caused by gram-negative bacteria (Van Oosten et al., 2002). As a recognition protein, apoLp-III has been shown to interact with both bacterial LPS and lipoteichoic acid (Halwani et al., 2000; Leon et al., 2006; Oztug et al., 2012) as well as fungal β-1,3 glucans (Whitten et al., 2014). Recently, pathogen-specific responses in re-localization of apoLp-III in the hemocytes and apoliophorin composition provided further support that apoLP-III is an important signalling molecule during early stages of immune response of *G. mellonella* and may coordinate immunogen-dependent responses by the innate immune system (Stączek et al., 2017).

Another important humoral component of the innate immune system of *G. mellonella* is prophenoloxidase (proPO), which is sequestered inside a specific type of hemocyte called oenocytoides (Schmit et al., 1977; Wojda, 2017). While few studies have focused on the phenoloxidase (PO) cascade in *G. mellonella* specifically, it has been better described in *Drosophila*, where upon pattern recognition or wounding a serine protease cascade activates proPO to PO, which is a main component of the melanin synthesis pathway (Cerenius et al., 2008; Tang, 2009). In response to an invading microbe, the innate immune system quickly responds by producing a dark pigment melanin which is used to encapsulate the pathogen. Another outcome of the PO cascade is the generation of highly-toxic intermediary compounds, including droxyphenylalanine (DOPA), quinons and free radicals, which are cytotoxic to host as well as the invader (Cerenius et al., 2008). As such, the PO cascade must be under tight regulation, and in *G. mellonella*, the activation of proPO was found to be regulated by apoLp-III as well as several other defence molecules (Park et al., 2005; Zdybicka-Barabas et al., 2014).

While *G. mellonella* has clearly found a role as a model host in virulence testing, there are still some challenges associated with variability in this host system. Currently most research is carried out on mixed-genotype larvae locally sourced from small growers that sell 'wax worms' as pet food and/or bait, which can lead to higher experimental variabilities requiring lots of biological replication. Recently however, a biotechnology

company in the U.K. has begun to supply research grade *G. mellonella*, called TruLarv (https://biosystemstechnology.com/products), which is a good indicator that the use of this model host is gaining wider scientific acceptance. Like any animal model, there are many other factors, including diet, handling, light-cycle, humidity, incubation temperature, pre-existing exposure to pathogens/parasites or pre-treatment with antibiotics or hormones for example, that may affect variability of *G. mellonella* model of infection. Pre-exposure to increased temperatures has been shown to result in an immune-priming effect (Mowlds and Kavanagh, 2008), while extended starvation (seven days with no food) was found to have a negative effect on *G. mellonella* immunity by reducing both cellular and immune responses (Banville et al., 2012). Some recommendations on standardized experimental approaches for study of bacterial diseases and antimicrobial drug tested have recently been published (Tsai et al., 2016) and were generally followed in this thesis.

Another challenge related to using *G. mellonella* as a model host is lack of genomic resources compared to other invertebrate model hosts such as *D. melanogaster* (Santos et al., 2015) or *C. elegans* (Yook et al., 2012), for example. Recently a draft shotgun-assembled genome for *G. mellonella* was published, but this resource does not yet include any gene annotations (Iglauer et al., 2018). An expressed-sequence tag (EST) library for immune genes of *G. mellonella* is however available (Vogel et al., 2011).

To date, no studies have used *in vivo* RNA-seq of a bacterial pathogen during infection of *G. mellonella.* This has been achieved for the human fungal pathogen *Candida albicans*, through enrichment of the *C. albicans* ORFome using biotinylated probes (Amorim-Vaz et al., 2015). In this study, the transcriptional response of *Caenorhabditis albicans* was surprisingly similar in both *G. mellonella* and mouse host at two different infection time-points, and infection-responsive genes were related to iron acquisition, stress response, adhesion and biofilm formation (Amorim-Vaz et al., 2015). In another study exploring *in vivo* transcriptome response of entomopathogenic fungi *Beauveria bassiana* at different stages of infection in *G. mellonella*, Chen et al. (2018) were able to characterize several discrete expression patterns among genes that responded to the *in vivo* conditions over-time, including genes related to host recognition/adhesion, host cuticle penetration and detoxification, evasion of host immune system and occupation

of the whole host, including limiting bacterial competition by production of antimicrobial compounds.

Here we describe a simplified method of enrichment of bacterial RNA from within the hemolymph of *G. mellonella* and demonstrate adequate enrichment and discovery of putative virulence factors (VFs) in MH96. This method was successfully used to identify genes encoding known and putative VFs under host- and temperature-dependent regulation, and to characterize functional categories of putative VFs that are expressed over the course of infection. This method was also able to provide a wealth of insight into a relatively understudied, yet highly pathogenic organism, and represents one (if not the only) reports of genome-wide expression of a pathogenic bacterium within an insect host at the time of publication.

4.2 *In vivo* and *in vitro* RNA-seq and phenotype microarray methods

4.2.1 RNA collection

4.2.1.1 *In vitro* samples

Wild-type MH96 was grown overnight in 3 ml Luria Broth Base (Invitrogen) (LB broth) at 25 °C with shaking (250 rpm). The next day, flasks containing 50 ml media were seeded with 1 % overnight culture and grown at either 25 or 37 °C with shaking until the appropriate cell densities were reached. *In vitro* samples (1 – 2 ml depending on cell density) were stabilized using RNAprotect Bacterial Reagent (Qiagen), according to the manufacture's guidelines and stored at -80 °C.

4.2.1.2 *In vivo* samples

Larval *G. mellonella* (Biosuppliers, Auckland, New Zealand) were maintained on a fresh diet of liquid honey, glycerol, Farex brand baby rice cereal and granular yeast for up to one week. Healthy larvae of similar lengths (> 20 mm) and weights (0.15 – 0.30 g) were selected and immobilized on ice.

MH96 was grown overnight (18 hours) in 3 ml nutrient broth with shaking (250 rpm) at 30 °C. Overnight culture was diluted in phosphate buffer solution (PBS) and placed

on ice. Larvae were surface sterilized with 70 % ethanol and then injected below the third right leg with 10 µl of inoculum using a 30-gauge ½" needle on a 1 ml tuberculin syringe (Terumo) with a micro-injector. Three larvae per treatment were mock-injected with PBS as a control.

For higher accuracy, inoculum from the microinjector was spread onto LB agar (Miller) to determine cell density. A significantly higher number of cells (five-hold higher) were administered for the early-infection samples compared to the middle- and late-infection samples at 25 °C to generate enough total RNA following extraction (see below for details on early-, middle-, and late-infection stages). Initially, a lower inoculum was administered for the 37 °C middle-infection samples; however, following the incubation period, the number of colony forming units (CFUs) were variable (ranging beyond a log value). Subsequently, the cell density in the inoculum was increased tenfold, which resulted in a more consistent cell density following *in vivo* incubation at 37 °C.

Larvae were placed in petri-dishes and incubated at either 25 or 37 °C within sealed plastic bags with moistened paper towel to maintain humidity. For RNA collection, hemolymph was obtained by puncturing the dorsal cuticle near the second segment with a 30-gauge ½" needle. Hemolymph was pooled from five to six individuals (depending on estimated cell density), yielding approximately 150 µl. Hemolymph samples from the early-infection time-point hemolymph were collected 75 – 90 minutes post infection. Hemolymph samples from the 25 and 37 °C middle-infection time-points were collected after 18 – 19 and 10 h post-infection, respectively. Hemolymph samples from the late-infection time-points were collected 26 – 30 h post infection.

Pooled hemolymph samples were immediately placed into a Micro centrifuge tube with 300 µl of RNAprotect Bacterial Reagent (Qiagen), vortexed for five seconds, incubated at room temperature for 5 min and then centrifuged at 300 x g for 5 min at 4 °C to pellet the hemocytes. The supernatant was carefully separated from the cell pellet by pipette. The supernatant was then pelleted by centrifugation for 10 min at 5,000 x g at ambient temperature (~ 22 °C). Following centrifugation, the supernatant was removed, and the samples were immediately stored at -80 °C.

To estimate the CFUs per larvae, serial dilutions from *G. mellonella* homogenate (5 larvae per treatment; surface sterilized with 70 % ethanol) were plated on K1 selective media and all orange colonies (*Yersinia*-positive) were counted within 24 hours. The weight of each larva was measured enabling the determination of CFUs per gram of larval homogenate. *In vivo* growth of MH96 at 25 °C was also measured by determining cell density per larva by dilution plating of *G. mellonella* homogenate (3 larvae per time point) to make a growth curve over 24 h.

4.2.2 RNA Extraction

Extraction of the RNA was completed within one week following sample collection. The samples were placed at room temperature and then re-suspended in 100 µL Ambion molecular biology grade Tris-EDTA (TE) buffer (10mM Tris-HCl, 1mM EDTA; pH 8.0) (Invitrogen) containing 15 mg/ml lysozyme (Roche or Sigma-Aldrich) and 10 µl proteinase K (Roche). The samples were incubated at room temperature for 10 min, with a 10 s vortex every two min to lyse the cells. RNA extraction was undertaken using the RNeasy mini kit (Qiagen) following the manufacture's guidelines. On-column DNAse (Qiagen) digestion was also completed followed by a second off-column DNAse digestion and subsequent RNA clean-up using the RNeasy mini kit. RNA was eluted two times in 30 µl RNAse free water, for a total volume of 60 µl. Total RNA extract was then precipitated at -20 C overnight in 100 % isopropanol, followed by two 70 % ethanol washes and then stored at -80 °C.

The RNA quality and quantity were assessed using the Nanodrop 2000 spectrophotometer (Thermo Scientific). The quality of the *in vivo* RNA samples were also visualized on a 1 % agarose gel containing 1 % v/v standard Janoa brand bleach, using the method described by Aranda *et al.* (2012). The *in vivo* RNA sample was used as PCR template for the amplification of the housekeeping gene *recA* (RecA_352F 5′ – TCTCAGCCAGATACCGGTGA; RecA_987R 5′- CAGCAACATTTCACGCAGCT).

It was difficult to achieve high enough total RNA from extracted from pooled hemolymph collected from only six larvae during early-infection, so the number of larvae needed to be doubled in order obtain the minimum amount of total RNA

required for Illumina sequencing (i.e., early-infection samples included hemolymph from twelve individual larvae). RNA samples with an estimated total RNA content of greater than 7.2 μg were diluted to 120 ng/μl concentration and 60 μl of this were added to RNAstable 1.5 micro-centrifuge tubes (Biomātrica, San Diego USA) following the manufacturer's guidelines. The entire sample was added to the RNA stable tube if it had less than 7.2 μg total RNA. Tubes were placed on the SpeedVac (Savant) without heat for 1.5 h.

4.2.3 Ribosomal RNA depletion, library preparation and sequencing

All work described in this section was completed by Macrogen Korea. Following recovery from the RNAstable tubes, RNA quantity and Integrity Number (RIN) were determined using the 2200 TapeStation (Agilent Technologies, Santa Clara, CA, USA). Host and bacterial ribosomal RNA was depleted from the *in vivo* samples using the ScriptSeq Complete Gold Kit (Epidemiology) (Illumina, Singapore) according to the manufacturer's guidelines then strand-specific cDNA libraries were prepared using the ScriptSeq RNA-Seq Library Preparation Kit (Illumina) following the standard protocol. The *in vitro* samples were prepared using the TruSeq stranded mRNA (Illumina) with the Ribo-Zero rRNA removal kit for Bacteria (Epicentre, Madison, WI, USA). To verify the size of the PCR enriched fragments, the template size distribution was assessed on the 2100 Bioanalyzer using a DNA 1000 chip (Agilent Technologies). The libraries were sequenced on the HiSeq2500 platform (Illumina) (Version HCS v2.2) to generate 101 bp paired-end reads.

4.2.4 Sequence processing and alignment

Sequence data were processed in a Linux environment on the AgResearch high-performance computing clusters. Flexbar (version 2.4) was used to filter adapter and barcode contamination and trim reads in 'ANY' mode (maintains the longer side of trimmed read regardless of direction) with minimum overlap of five, maximum uncalled base of one and minimum quality of threshold for trimming of 20 (Dodt et al., 2012). Adapter and barcode reference sequences used to trim the RNA-seq libraries are

provided in supplementary Table S6 Table S7. The bbsplit.sh function, of the BBMap package, was used to quantify and remove contaminating rRNA sequences with a minimum ratio of 0.5, a minimum hit number of one and a maximum insertion/deletion length of 500 nt (Bushnell, 2015). All libraries were aligned against full-length 5S, 16S and 23S rRNA sequence from the MH96 genome (GCA_001656035.1) and *in vivo* libraries were also mapped against *G. mellonella* partial 18S and 28S rRNA sequence retrieved from GenBank (Accessions: U65198.1, U65138.1, AF286298.1, AF423921.1 and X89491.1). *In vivo* libraries were further screened against the *G. mellonella* EST library (JG394435 – JG406465) (Vogel et al., 2011) to remove additional host sequence prior to alignment to the reference genome.

Remaining paired-end reads were aligned to the MH96 genome (downloaded from NCBI August 2018; GCA_001656035.1/ASM165603v1) using Rockhopper (v2.03) in verbose mode. Alignment stringencies were increased from default values, such that allowable mismatch threshold was decreased from 15 to 10 % of read length and the minimum seed was increased from 33 % to 50 % of read length. Any residual count data aligning to 5S, 16S and 23S rRNA sequence were not considered further in this analysis.

4.2.5 Preliminary Data Exploration

Code for *in vivo* RNA-seq analysis is provided in supplementary section: Source code 1. The preliminary expression data was investigated to determine whether the depth of sequencing was sufficient to quantify expression from a diversity of genes in the reference genome using the iNEXT R package (nboot = 50) (v2.0.18) (Hsieh et al., 2016). Also, variability between biological replicate samples was assessed by plotting the expression profiles in ggplot2 R package (v3.1.0) (Wickham, 2016). The biological coefficient of variance (BCV; a common measure of the average dispersion in gene expression values that is associated with biological variability) was calculated using the edgeR (v3.24.3) (Robinson et al., 2009). The RUVseq R package (v1.16.0) was used to generate relative log expression (RLE) plots (Risso et al., 2014). The data were also visualized using principle component analysis, as implemented in RUVseq to identify

possible outliers. The count data were normalized using upper-quartile normalization (Bullard et al., 2010).

4.2.6 Differential Expression Analysis

The normalized count data were filtered for non-expressed features (only features with greater than 5 aligned reads in at least two libraries were retained). The data were then converted to log2-counts-per-million (Log_2CPM) using the R package limma (v3.38.3) (Ritchie et al., 2015; Smyth, 2004) and the mean-variance relationship was modelled using the 'voom' method (Law et al., 2014). Linear models were fit to all features (protein-coding genes and non-coding RNAs) to estimate fold change and standard error and then the standard errors were smoothed by applying an Empirical Bayes approach.

Differentially expressed (DE) features between *in vivo* and *in vitro* libraries were defined by multiple testing with false discovery rate of 0.05, corrected using either the Benjamini and Hochberg (1995) or Benjamini and Yekutieli methods (2001). Two separate DE analyses were performed to answer the following biological questions:

1) Differential expression analysis 1
    a) During different growth stages (e.g., lag, exponential and stationary) at 25 °C, which genes respond more strongly to *in vivo* compared to *in vitro* growth conditions?
2) Differential expression analysis 2
    a) Which genes respond differently to 25 compared to 37 °C *in vitro*?
    b) Which genes respond differently to 25 compared to 37 °C *in vivo*?
    c) Which genes respond to temperature differently *in vivo* compared to *in vitro*?

Visualization of genome-wide temperature-dependent responses at 10^8 CFU/g or CFU/ml cell densities was completed with Perl software tool Circos (v0.69) (Krzywinski et al., 2009).

4.2.6.1 Cluster analysis and functional enrichment of putative *in vivo* virulence factors

Log_2CPM fold-changes were calculated for all DE features between *in vitro* and *in vivo* libraries, across cell densities at 25 °C then clustered using the fuzzy c-means (FCM) algorithm (Futschik, Matthias and Carlisle, 2005) using the MFuzz R package (v2.42.0) (Kumar and Futschik, 2007). Since these data included only three cell density points, optimization of the fuzzifier parameter, *m*, and the number of clusters, *k*, was not possible; neither the optimization method of Schwämmle and Jensen (2010) or assessment of cluster overlap and correlation for different parameter values (Israel et al., 2016) were not found to be informative for these data.

When using FCM to soft cluster expression data across a limited number of time-points, it is recommended by the package authors to qualitatively assessed a range of *m* and *k* values using functional information (Kumar and Futschik, 2007). To identify optimal *m* and *k* values for a limited number of cell densities, both predicted operon structure (output from Rockhopper) and functional annotations assigned using homology based search (Blastx, e-value cut-off = 1e-8) against the virulence factor database (VFDB) (accessed July, 2018, http://www.mgc.ac.cn/VFs/) (Chen et al., 2016) were considered in this analysis. Based on this qualitative assessment, twenty clusters with a fuzzifier value of 1.5 were identified. Host-specific and culture-specific factors were identified from clusters with median log_2CPM fold-change values of ≥ 2 and ≤ -2 for at least one cell density, respectively. These median log_2CPM fold-change cut-off values were found to capture genes with unique and highly differential expression between *in vivo* and *in vitro* libraries. Host-specific factors were then assigned as putative host-specific VFs based on sequence similarity to known VFs.

4.2.7 Phenotype microarray

Temperature has been shown to have a wide-spread effect on the metabolic capabilities of *Ps. asymbiotica* (Mulley et al., 2015), an emerging human pathogen that is entomopathogenic and symbiotic to nematodes. In an approach similar to Mulley et al., (2015), phenotype microarrays were undertaken according to manufacturer (Biolog, Inc.) guidelines (e.g., "PM procedures for *E. coli* and other GN Bacteria") to assess for temperature-dependent changes sole-carbon utilization by MH96. Fresh streaks of MH96 from -80 °C stocks were made onto LB agar and grown for 24 hours at 30 °C. A small loopful of each replicate strain was re-suspended in IF-0 solution containing 6 mM sodium thiosulfate and adjusted to 42 % turbidity using the Biolog turbidimeter. Samples were diluted 1:5 with Dye A in solution IF-0 and used to inoculate PM1 and PM2A plates with 100 µl per well. Plates were incubated at 25 and 37 °C in an OmniLog incubator and the formation of formazan was monitored every 15 min for 48 h.

Statistical analysis of phenotype microarray data was undertaken by Paul Maclean from Bioinformatics and Statistics team, AgResearch Ltd. (New Zealand). Omnilog data was extracted using the Omnilog software. Data from the "Meta Parameters by Well" sheet from the Omnilog software such as net slope and well annotation were added to the data files containing the actual readings for each well, from which the maximum values were calculated. The data for each carbon source and the respective negative control for each plate was plotted using the "xyplot" function from the lattice package (v 0.20-38) (Sarkar, 2008) in R (v3.6.1) (R Core Team, 2017).

Growth curves were detected using the following criteria:

1. > 5% slope as detected by the Omnilog software; and
2. Maximum value > 127 (largest negative control) to account for the rather inconsistent negative control samples.

Detected growth curves were checked against the plots (and some against the final pictures of the plates) and found to be conservative but reliable.

4.3 Results

4.3.1 RNA collection

In vitro samples of MH96 culture were collected from cell densities ranging from 3.2 x 10^7 to 4.4 x 10^9 CFU/ml grown at 25 or 37 °C (Table 4.1). Intrahemocoelic inoculum of MH96 used for *in vivo* treatments in *G. mellonella* ranged in cell density from hundreds to 10^7 cells per injection (Table 4.2).

Table 4.1: *Yersinia entomophaga* MH96 *in vitro* RNA samples, including cell density and optical density at time of collection (single replicates). CFU = colony forming unit and OD_{600} = optical density measured at 600 nm.

In vitro growth phase	Temperature (°C)	Mean Cell density (CFU/ml)	Mean OD_{600}	Culture time
Lag	25	3.2 x 10^7	0.138	4 ¾
Exponential	25	2.9 x 10^8	1.553	6
Stationary*	25	4.4 x 10^9	NA	~ 9
Exponential	37	3.3 x 10^8	1.026	3

* Stationary phase samples were collected separately by an AgResearch Research Associated, Mitchell Watson as part of a different project and combined with these data. An optical density was not recorded at time of collection so is not available (NA).

Table 4.2: Mean number of cells of *Yersinia entomophaga* MH96 from inoculum and hours-post infection of *Galleria mellonella* by intrahemocoelic injection corresponding to lag, exponential and stationary *in vivo* growth phases. (n = 3 replicate plates). CFU = colony forming units and SD = standard deviation.

In vivo growth phase	Temperature (°C)	Mean inoculum cells (CFU ± SD)	Hours-post infection
Lag	25	2 ± 0.1 x 10^7	1 ¼ - 1 ½
Exponential	25	157 ± 38	18 -19
Stationary	25	693 ± 281	26 - 30
Exponential	37	5 ± 0.3 x 10^3	10

When *G. mellonella* was infected with 345 cells and incubated at 25 °C, it took approximately 5 hours for the population to transition from an early-infection lag phase into exponential growth phase. The exponential growth phase occurred between

cell densities 1 x 10^3 to 1 x 10^9 CFU/g and lasted from 5 hours post-infection (HPI) until approximately 24 HPI (Figure 4.1). After 24 HPI the population entered stationary phase, reaching a cell density greater than 5 x 10^9 CFU/g after 48 HPI.

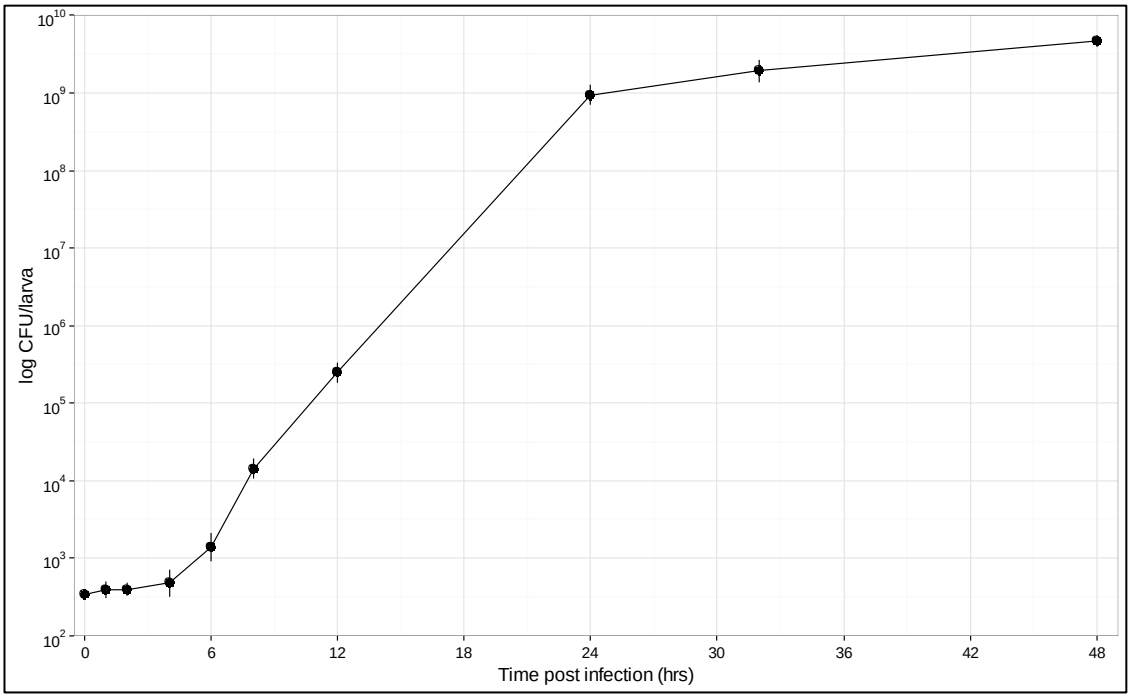


Figure 4.1: *In vivo* growth curve of *Yersinia entomophaga* MH96 in the hemolymph of *Galleria mellonella* at 25 °C. Larvae were injected with ~ 345 cells in 10 µl PBS and CFU/larvae were enumerated from three biological replicates per time point. Points represent mean density and vertical bars represent standard error.

Post-infection cell densities determined for the *in vivo* RNA ranged from 10^7 – 10^9 CFU/g (Table 4.3) and densities of around 5 x 10^8 and 10^9 CFU/g were considered to represent exponential and stationary growth phase, respectively. While the earliest infection sample was collected 90 min post-infection, the initial inoculum was relatively higher (2 x 10^7 CFU/g) compared to the inoculum administered for the *in vivo* growth curve, this treatment was still considered a representative of exponential growth phase due to the short exposure to the host environment, considering the logistical challenges in collecting RNA from lower density *in vivo* samples.

Table 4.3: Mean cell density of *Yersinia entomophaga* MH96 recovered from homogenized *Galleria mellonella* (n = 3 replicate homogenates) collected at the same growth phase/temperature as *in vivo* RNA samples. CFU = colony forming and SD = standard deviation.

In vivo growth phase	Temperature (°C)	Mean cell density per larvae (CFU/larvae ± SD)	Mean cell density per gram *G. mellonella* homogenate (CFU/g ± SD)
Lag	25	$1.1 \pm 0.3 \times 10^7$	$4.4 \pm 0.8 \times 10^7$
Exponential	25	$7.8 \pm 5.1 \times 10^7$	$4.4 \pm 3.0 \times 10^8$
Stationary	25	$1.1 \pm 0.8 \times 10^9$	$5.0 \pm 3.8 \times 10^9$
Exponential	37	$6.4 \pm 1.7 \times 10^7$	$3.1 \pm 0.9 \times 10^8$

4.3.2 Quality control – pre-rRNA depletion

RNA samples were found to have required 260/280 ratios, ranging from 1.80 - 2.04 (pure RNA is expected to have a 260/280 ratio of ~2.0) (Table 4.4). While the Nanodrop does not provide indication of quality, the RNA samples had enough total RNA content to permit high-throughput sequencing on the Illumina platform (typical minimum requirement of 3 μg).

Despite completion of both an on-column and off-column DNAse digestion, a few of the *in vivo* RNA samples produced very faint bands when used as a PCR template for amplification of a housekeeping gene, indicative of minimal amount of contaminating genomic DNA in the samples (data not shown). Due to the limited amount of total RNA present in some of the samples, a third DNAse digestion was not completed; however quality assurance completed using the RNA TapeStation confirmed adequate depletion of DNA in the samples (see below).

Table 4.4: Mean total RNA quantity and quality determined by spectrophotometer (*in vivo*: n = 4 and *in vitro*: n = 2). CFU = colony forming units and SD = standard deviation.

Growth phase/ Cell density (CFU/g or CFU/ml)	Temp-erature (°C)	Concen-tration ± SD (ng/µl)	Total ± SD (µg)	260/280 ± SD	260/230 ± SD
In vivo					
Lag / 10^7	25	111.2 ± 8.7	5.6 ± 0.4	1.83 ± 0.02	2.38 ± 0.05
Exponential / 10^8	25	190.4 ± 64.1	9.6 ± 3.3	1.87 ± 0.03	2.56 ± 0.05
Stationary / 10^9	25	321.6 ± 116.7	16.1 ± 5.8	1.93 ± 0.04	2.57 ± 0.03
Exponential / 10^8	37	225.1 ± 54.2	12.4 ± 3.0	1.90 ± 0.03	2.57 ± 0.01
In vitro					
Lag / 10^7	25	230.4 ± 13.6	11.5 ± 0.7	1.95 ± 0.00	2.42 ± 0.04
Exponential / 10^8	25	656.1 ± 39.9	39.4 ± 2.4	2.02 ± 0.04	2.58 ± 0.11
Stationary / 10^9	25	1691.1 ± 306.5	84.6 ± 15.3	2.10 ± 0.00	2.45 ± 0.01
Exponential / 10^8	37	871.6 ± 97.3	41.0 ± 1.2	2.05 ± 0.01	2.35 ± 0.04

Following recovery from the RNAstable tubes, the RNA 2200 TapeStation quality control analysis demonstrated that the RNA samples were of good quality based on RNA integrity number (RIN), which were all greater than or equal to 6.9 (RIN value of 7.0 or greater is typically the minimum cut-off) (Table 4.5). However, since the *in vivo* samples contained both bacteria and host RNA, the RIN value was considered somewhat unreliable as the RIN algorithm is unable to differentiate between rRNA in mixed eukaryotic/prokaryotic samples, potentially resulting in a quality underestimation. The electropherograms were also visually inspected as an additional quality control check, ensuring rRNA peaks were sharp and with relative areas under the curve as expected depending on the infection time-point. No obvious peaks larger than the 23S rRNA peak (i.e., contaminating gDNA) were found, indicating adequate DNAse depletion.

Table 4.5. Mean total RNA quantity and quality determined by TapeStation electrophoresis. CFU = colony forming unit, SD = standard deviation, RIN = RNA integrity number, n = number of samples.

Growth phase/Cell density (CFU/g or CFU/ml)	Temperature (°C)	n	Concentration ± SD (ng/µl)	RIN value ± SD
In vivo				
Lag / 10^7	25 °C	4	116.9 ± 7.6	8.1 ± 0.3
Exponential / 10^8	25 °C	4	131.4 ± 23.6	7.7 ± 0.3
Stationary / 10^9	25 °C	4	138.6 ± 5.9	7.7 ± 0.6
Exponential / 10^8	37 °C	4	154.2 ± 31.5	8.5 ± 0.1
In vitro				
Lag / 10^7	25 °C	2	106.4 ± 2.5	9.2 ± 0.1
Exponential / 10^8	25 °C	2	133.7 ± 5.7	9.3 ± 0.0
Stationary / 10^9	25 °C	2	140.2 ± 38.0	9.3 ± 0.1
Exponential / 10^8	37 °C	2	114.3 ± 1.7	9.1 ± 0.1

Two discrete 16S and 23S rRNA peaks were identified in the electropherogram of the exponential *in vitro* total RNA sample (Figure 4.2). Figure 4.3 shows an electropherogram obtained from a lag phase (10^7 CFU/g) *in vivo* RNA sample with a single rRNA peak comprised mixed host and bacterial rRNA. The electropherogram of an exponential phase (10^8 CFU/g) *in vivo* RNA sample is shown in Figure 4.4, in which three distinct rRNA peaks are easily distinguished; the unlabelled (middle) peak represents host rRNA (denatured insect rRNA migrates as a single band). Three distinctive rRNA bands were also identified from total *in vivo* RNA samples collected during exponential phase when visualized by electrophoresis on 1 % agarose gel. (Figure 4.5). The electropherograms from the stationary phase (10^9 CFU/g) *in vivo* RNA samples were similar the exponential phase and are not shown here.

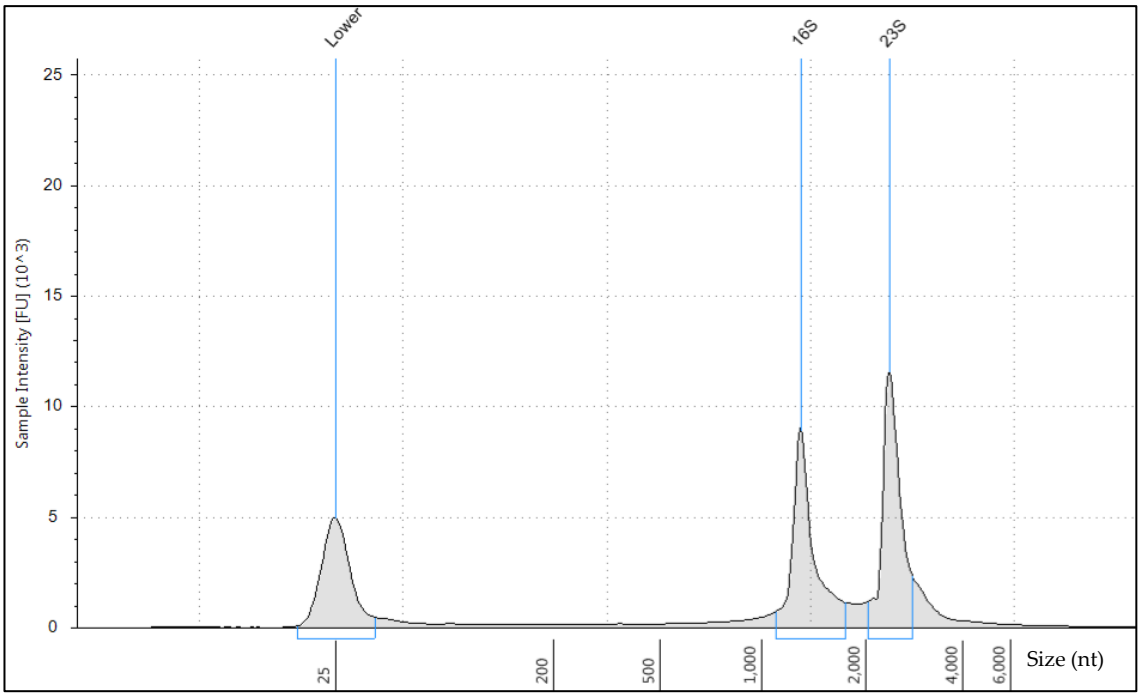


Figure 4.2. Exponential phase (10^8 CFU/ml) *in vitro* total RNA sample electropherogram. Graph has been cropped to show all peaks.

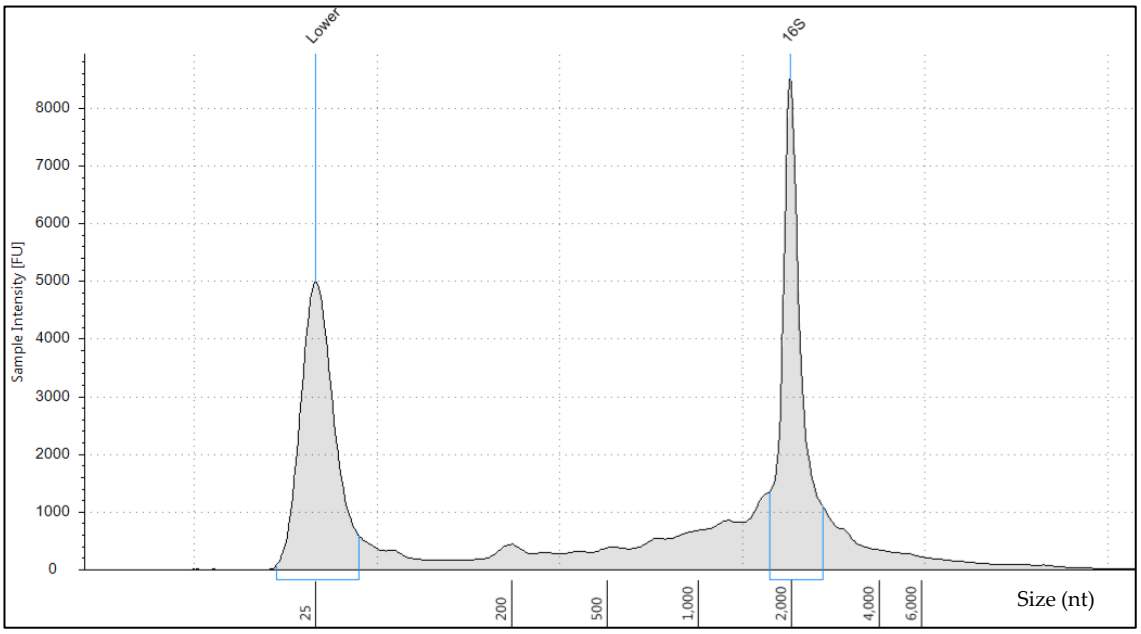


Figure 4.3. Lag phase (10^7 CFU/ml) *in vivo* total RNA sample electropherogram. Graph has been cropped to show all peaks.

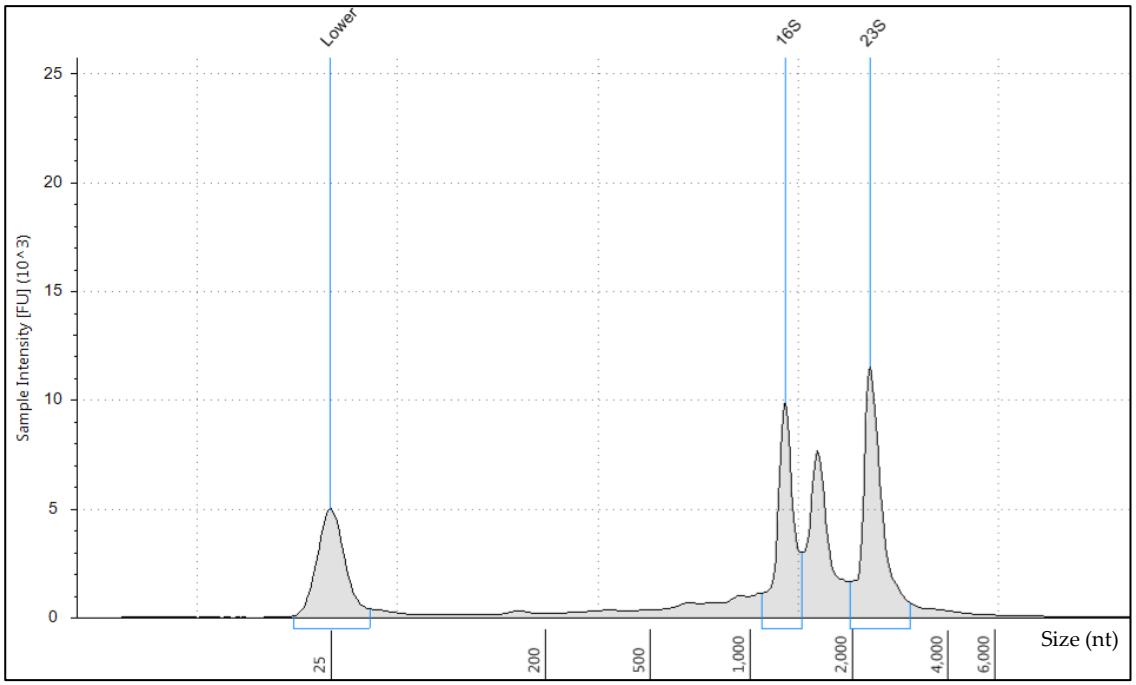


Figure 4.4. Exponential phase (10^8 CFU/ml) *in vivo* total RNA sample electropherogram. Graph has been cropped to show all peaks. The peak in between the prokaryotic 16S and 23S ribosomal RNA (rRNA) peaks represents insect rRNA which migrates as a single band when denatured.

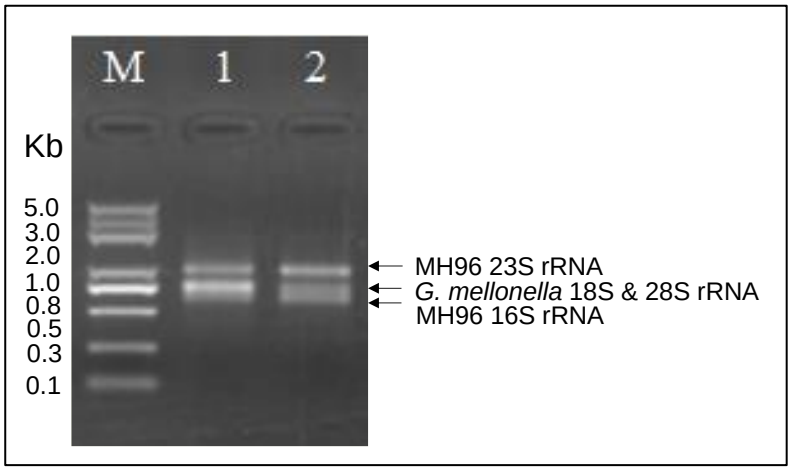


Figure 4.5. Example *in vivo* total RNA samples visualized on 1 % agarose gel by Novagene (MH96 grown to 10^8 CFU/g at 37 °C). M: Trans 2K marker. Lanes 1 and 2 are replicate total RNA samples pooled from hemolymph of *Galleria mellonella*.

4.3.3 Quality control – post rRNA depletion

Following the depletion of host/bacterial rRNA the samples were analysed on the RNA 2200 TapeStation to ensure adequate rRNA depletion and sample integrity prior to library preparation. The majority of samples were adequately depleted of host/bacteria rRNA (Figure 4.6), while two samples were inadequately depleted (Figure 4.7) and one

sample was degraded (Figure 4.8). Additional samples were submitted to ensure only non-degraded and adequately depleted samples were sequenced in this study.

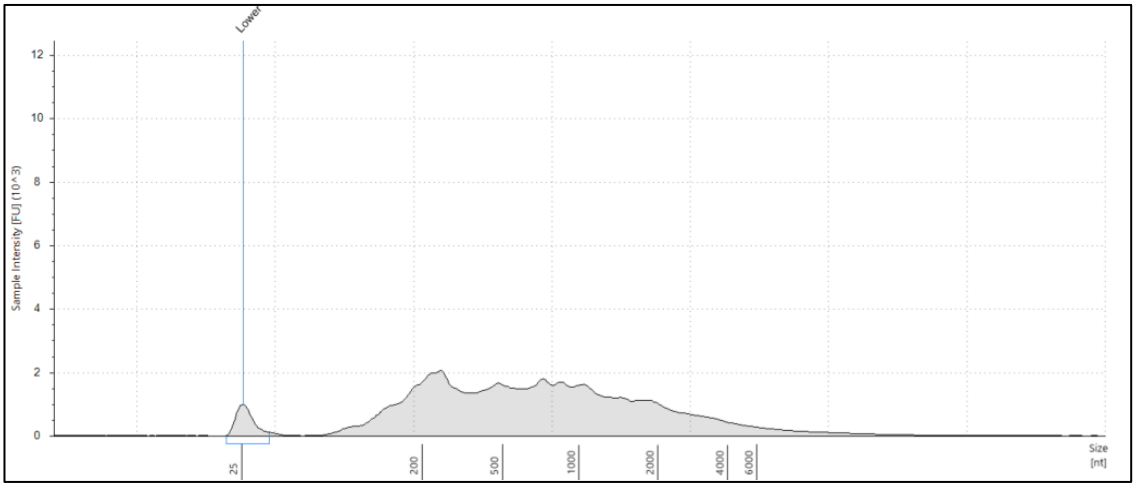


Figure 4.6. Example electropherogram of post-rRNA depletion stationary phase (10^9 CFU/ml) *in vivo* RNA showing adequate depletion.

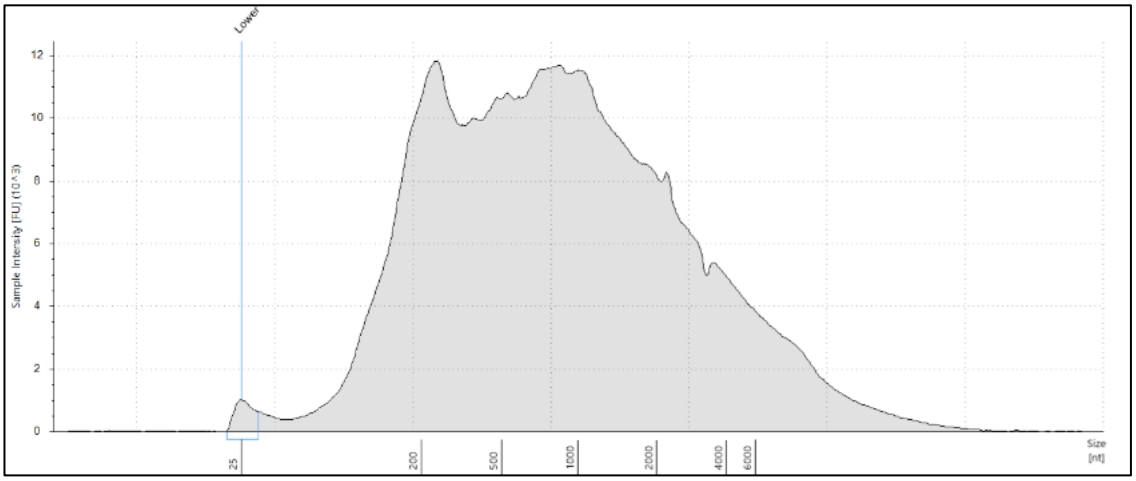


Figure 4.7. Example electropherogram of post-rRNA depletion stationary phase (10^9 CFU/ml) *in vivo* RNA showing inadequate depletion.

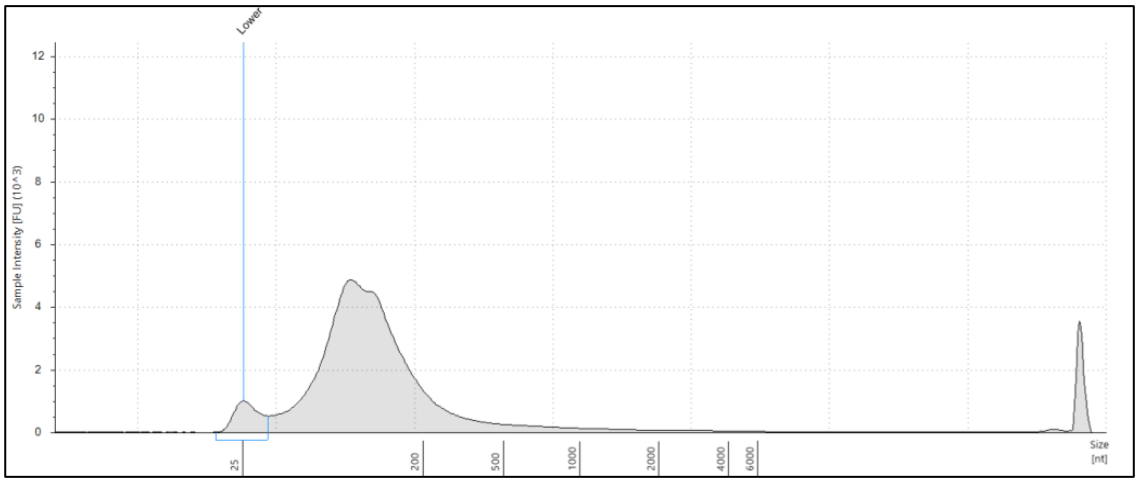


Figure 4.8. Example electropherogram of post-rRNA depletion stationary phase (10^9 CFU/ml) *in vivo* RNA showing degradation.

It was not possible to quantify the post-depletion RNA, since separate areas under the rRNA curves from host and bacteria could not be obtained from the Bioanalyzer. Instead, the proportion of remaining RNA concentration between pre- and post-rRNA depleted *in vivo* samples was used as a qualitative method (secondary to the electropherograms shown above) to assess the adequacy of rRNA depletion (Table 4.6).

Table 4.6: Dual post-rRNA depleted *in vivo* RNA mean sample concentration (n = 4). CFU = colony forming unit, SD = standard deviation, rRNA = ribosomal RNA.

Growth phase/cell density (CFU/g)	Temperature (° C)	Concentration rRNA-depleted RNA ± SD (ng/µl)	Mean proportion of post-/pre-rRNA depletion RNA concentration ± SD
Lag / 10^7	25	2.8 ± 0.7	2.4 ± 0.6 %
Exponential / 10^8	25	2.3 ± 1.5	1.7 ± 1.0 %
Stationary / 10^9	25	4.3 ± 1.0	3.1 ± 0.8 %
Exponential / 10^8	37	3.4 ± 0.4	2.3 ± 0.6 %

4.3.4 Sequence trimming/filtering

Paired-end sequencing data contained in gzip compressed fastq files were downloaded by file-transfer protocol from Macrogen Korea onto AgResearch servers. Base quality was good among all libraries, with greater than 90 % of bases having a minimum phred score of 30 (probability of incorrect base call 1 in 1,000 times) (Table 4.7). Very limited adapter trimming and quality filtering of the raw reads was required, with greater than 99 % of the sequence data being maintained post trimming (Table 4.8). *In vivo* samples were sequenced to a greater depth than *in vitro* samples to ensure

adequate coverage, which was estimated using rarefaction analysis from preliminary studies.

Table 4.7. RNA-seq raw read mean quantity and quality (*in vivo:* n = 4, *in vitro:* n = 2, standard deviation not shown). CFU = colony forming unit.

Growth phase / cell density (CFU/ml or CFU/g)	Temp-erature (°C)	Total read bases (Gb)	Total reads	GC content (%)	Q20* scores (%)	Q30** scores (%)
In vivo						
Lag / 10^7	25	6.564	64,985,540	43.70	97.01	95.01
Exponential / 10^8	25	3.738	37,012,275	48.96	97.58	95.84
Stationary / 10^9	25	5.781	57,244,437	48.80	96.37	93.70
Exponential / 10^8	37	3.727	36,901,330	48.25	95.66	92.67
In vitro						
Lag / 10^7	25	2.140	21,185,564	50.68	96.66	94.18
Exponential / 10^8	25	2.230	22,081,741	50.06	97.14	94.79
Stationary / 10^9	25	N/A***	17,565,070	N/A	N/A	N/A
Exponential / 10^8	37	1.862	18,430,803	49.28	97.39	95.29

*Q20 = Phred score of 20 (1 in 100 probability of incorrect base call)
**Q30 = Phred score of 30 (1 in 1,000 probability of incorrect base call)
***N/A these samples were processed at Macrogen at a later date, so the quality report is not available.

Table 4.8. Adapter removal and trimming results including mean trimmed, discarded, single-end and remaining reads (*in vivo*: n = 4, *in vitro:* n =2, standard deviation not shown).

Growth phase/ Cell density (CFU/g or CFU/ml)	Temp-erature (°C)	Trimmed low quality reads	Discarded reads	Single-end reads	Remaining reads/ (% of input)
In vivo					
Lag / 10^7	25	5,467,381	280,393	82,033	64,623,114 (99.5)
Exponential / 10^8	25	2,513,492	119,416	47,735	36,845,124 (99.6)
Stationary / 10^9	25	6,079,413	155,194	79,009	57,010,235 (99.6)
Exponential / 10^8	37	4,517,769	319,539	91,132	36,490,659 (99.2)
In vitro					
Lag / 10^7	25	1,920,767	15,295	0	21,170,269 (99.9)
Exponential / 10^8	25	1,689,817	53,169	8,315	22,02,258 (99.7)
Stationary / 10^9	25	1,401,048	12,688	0	17,552,382 (99.9)
Exponential / 10^8	37	1,462,654	13,373	0	18,417,430 (99.9)

While the post-rRNA depletion electropherograms indicated substantial depletion of both host and pathogen rRNA (Figure 4.6), the sequence data were assessed for

contaminating rRNA using bioinformatics approach (Table 4.9). The lag samples (10^7 CFU/g) were found to have between 18.0 and 29.3 % contaminating host rRNA, but less than 1 % microbial rRNA. Conversely, the remaining exponential (10^8 CFU/g) and stationary (10^9 CFU/g) *in vivo* samples contained much less host rRNA (2.0 – 0.1 %) but had varying amounts of microbial rRNA (6.9 – 43.6 %). The *in vitro* samples that were assessed contained less than 9 % microbial rRNA. Ribosomal RNA was not considered further in this analysis.

Table 4.9. Mean proportion of trimmed reads aligning to *Galleria mellonella* rRNA or EST library sequences or *Yersinia entomophaga* MH96 ribosomal RNA sequences using BBMap.

Growth phase / Cell density (CFU/g or CFU/ml)	Temp-erature (°C)	Reads align to *G. mellonella* rRNA	Reads align to *G. mellonella* EST	Reads align to *Y. entomophaga* rRNA
In vivo				
Lag / 10^7	25	23.0 %	4.8 %	0.2 %
Exponential / 10^8	25	0.4 %	0.8 %	26.0 %
Stationary / 10^9	25	1.1 %	0.7 %	27.0 %
Exponential / 10^8	37	0.3 %	1.4 %	22.1 %
In vitro				
Lag / 10^7	25	-	-	2.2 %
Exponential / 10^8	25	-	-	8.6 %
Stationary / 10^9	25	-	-	3.0 %
Exponential / 10^8	37	-	-	3.6 %

4.3.5 Alignment and preliminary data exploration

The filtered trimmed paired-end reads were aligned to the MH96 reference genome but only an average of 9 % trimmed reads from the lag (10^7 CFU/g) *in vivo* libraries aligned to the genome, while 56 - 59 % aligned from the remaining *in vivo* libraries. The *in vitro* libraries had the highest alignment rate (84 – 93 % of trimmed reads) (Figure 4.9).

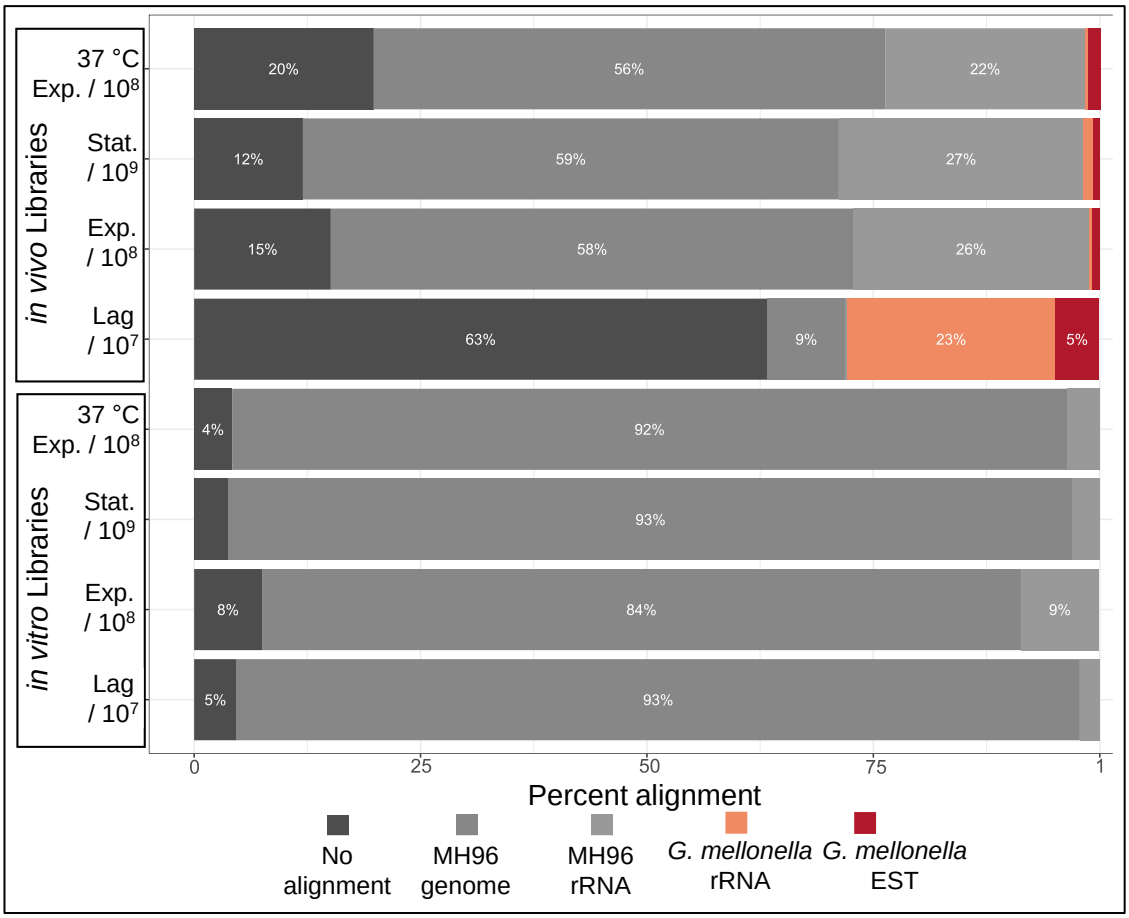


Figure 4.9. Mean proportion of paired-end reads attributed to host and pathogen based on average percent alignment to *G. mellonella* rRNA and EST sequence or *Y. entomophaga* MH96 genome or rRNA sequence (*in vivo:* n = 4, *in vitro*: n = 2). Libraries generated from treatments involving 37 °C incubation are indicated (all other treatment were incubated at 25 °C). The growth phase and estimated cell density (CFU/g and CFU/ml for *in vivo* and *in vitro*, respectively) are provided for each library (Exp. = exponential and Stat. = stationary). Only percentages of 4 % and greater are shown.

Of the reads aligning to the reference genome, most were located in protein coding regions (72 – 85 %); however, between 13 - 26 % of reads aligned to unannotated regions of the genome (Table 4.10). The proportion of reads aligning anti-sense to protein coding regions ranged from 1 – 2 % across all libraries representing 95 different anti-sense RNAs ranging in length from to 10 – 1, 187 nt. Rockhopper also predicted 713 multi-gene operons (ranging from 2 – 19 genes) and 478 non-coding RNAs (ranging from 10 – 415 nt length). Also predicted, were 1,528 and 1,065 5′- and 3′-UTRs, respectively.

Table 4.10: Mean percentage of trimmed paired-end reads aligning to of the reference genome, including protein-coding, tRNA and non-coding regions (*in vivo:* n = 4, *in vitro*: n = 2).

Growth phase/ cell density (CFU/g or CFU/ml)	Temp-erature (°C)	Total align-ments	Protein coding region	Anti-sense protein coding region	tRNA	Non-annotated regions
In vivo						
Lag / 10^7	25	9 %	77 %	2 %	1 %	22 %
Exponential / 10^8	25	58 %	74 %	2 %	1 %	24 %
Stationary / 10^9	25	59 %	75 %	1 %	0 %	24 %
Exponential / 10^8	37	56 %	72 %	1 %	1 %	26 %
In vitro						
Lag / 10^7	25	93 %	84 %	2 %	1 %	14 %
Exponential / 10^8	25	84 %	83 %	1 %	1 %	15 %
Stationary / 10^9	25	93 %	85 %	1 %	0 %	14 %
Exponential / 10^8	37	92 %	85 %	2 %	0 %	13 %

Rarefaction analysis determined the amount of paired-end reads aligning to the reference genome was sufficient to capture most expressed transcripts in the lag phase (10^7 CFU/g) *in vivo* libraries since only a slight improvement in total transcript detection would be expected if sequencing depth were increased (Figure 4.10). Rarefaction analysis was only conducted on the *in vivo* lag phase libraries since these contained the fewest number of reads/library.

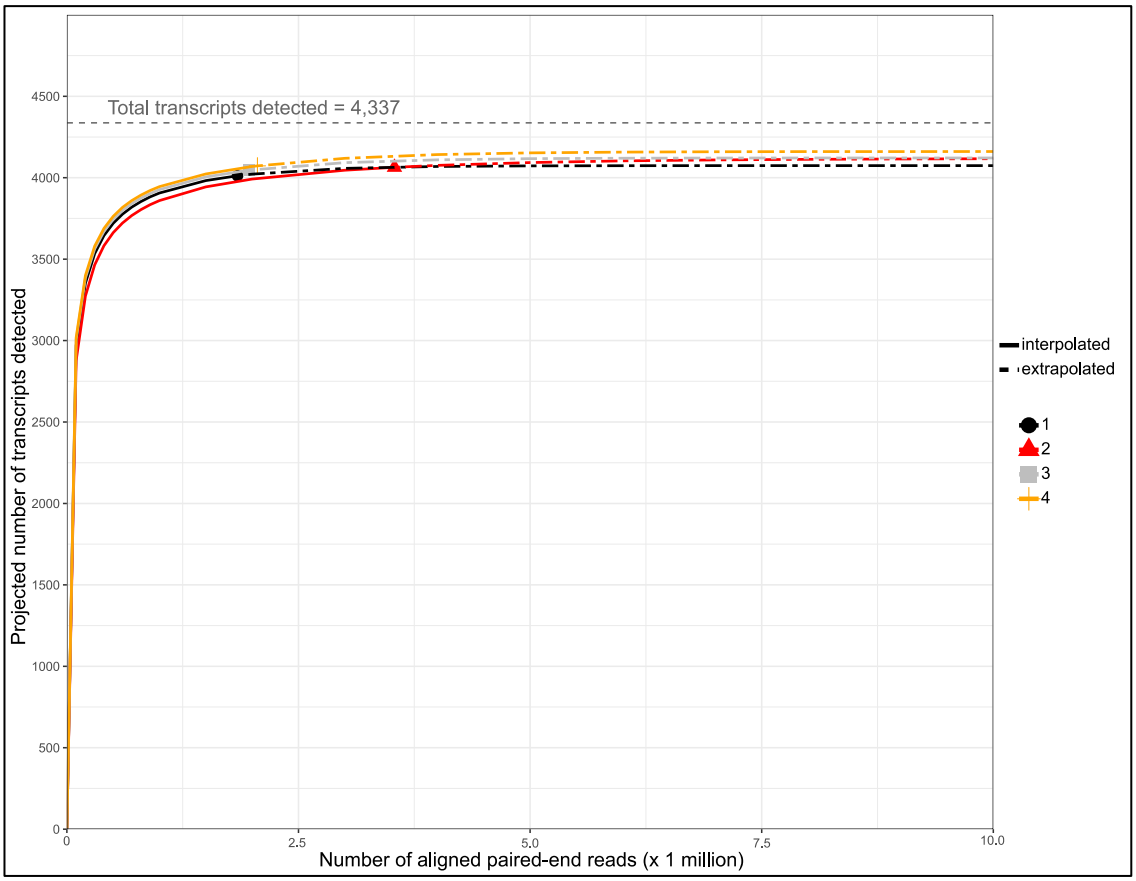


Figure 4.10. Predicted transcript detection based on the number of aligned paired-end reads for *in vivo* lag phase (10^7 CFU/g) libraries by rarefaction analysis. 1, 2, 3 and 4 represent individual libraries.

Relative log expression (RLE) plots were examined to identify any potentially unwanted technical variation arising from batch effects (i.e., samples collected on different days or by different researchers, etc.). The unnormalized RLE plot (Figure 4.11A) showed a strong effect of sequencing depth on the distribution of RLE values among all *in vivo* lag phase 25 °C libraries (consistently reduced sequence depth due to low cell density) and a single *in vivo* stationary phase 25 °C library (unintentionally sequenced to much greater depth than other samples). Application of upper quartile (UQ) normalization was used to reduce the effect of sequencing depth on the count data (Figure 4.11B). Following normalization, stationary (10^9 CFU/ml) *in vitro* libraries were found to have comparatively higher variation compared to the other libraries.

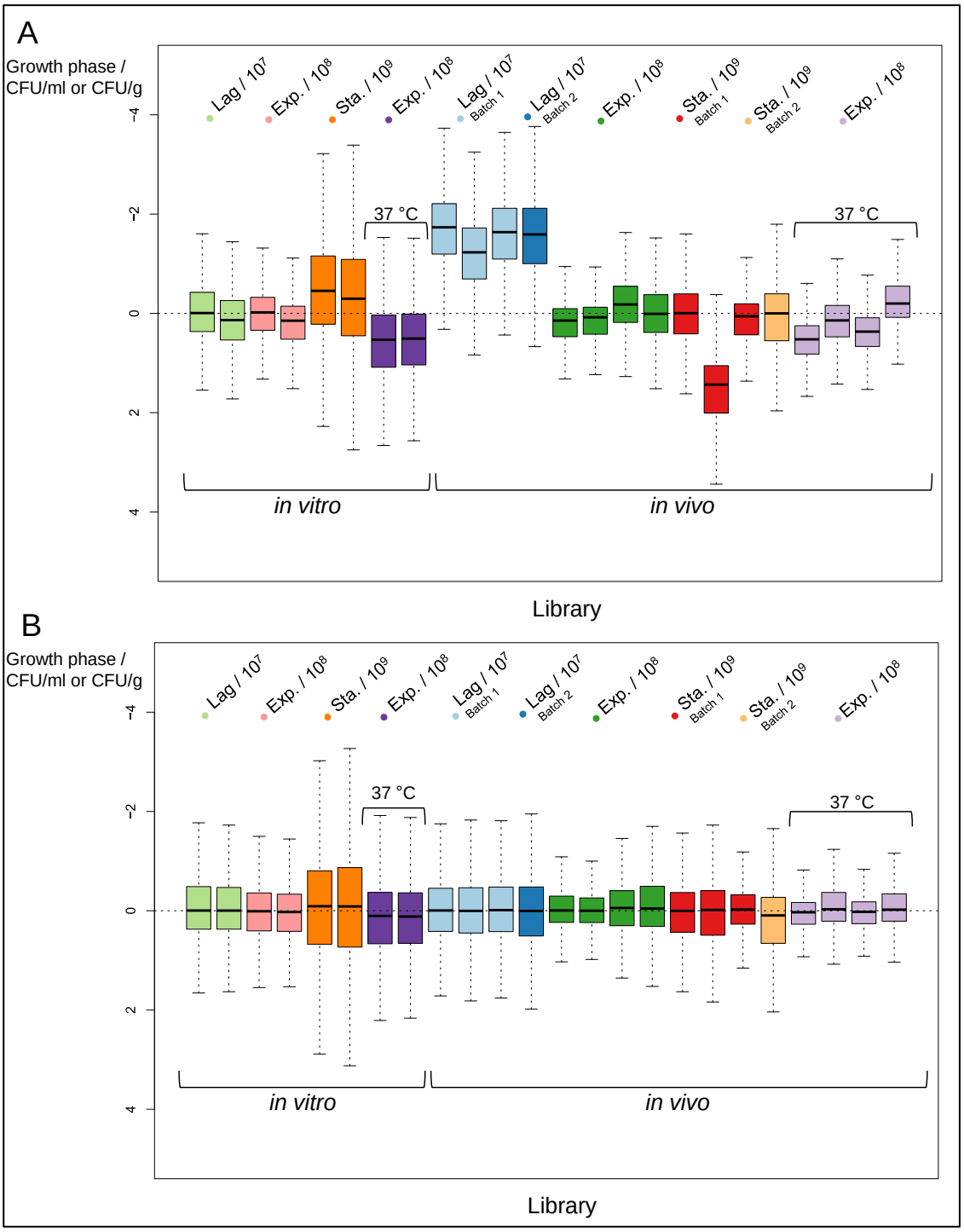


Figure 4.11. Relative log expression of *in vivo* and *in vitro* libraries represented as median boxplots. A) unnormalized and B) normalized by upper-quartile normalization. Libraries from 37 °C incubated treatments are noted; the remaining libraries were from 25 °C incubated samples. Individual collection batches are noted where required. Exp. = exponential and Sta. = stationary growth phases.

Principle coordinate analysis (PCA) of the normalized count data showed clear separation between *in vitro* and *in vivo* libraries (Figure 4.12). In general, each of the treatments clustered together, except one *in vivo* stationary (10^9 CFU/g) library

clustered more closely to the *in vivo* exponential (10^8 CFU/g) libraries. This replicate was identified as a biological outlier, likely resulting from cell-density variability within the pooled *in vivo* samples (i.e., the ranges of the exponential and stationary cell densities overlapped slightly) and was removed from the analysis. The libraries were more closely grouped by cell density and *in vivo*/*in vitro* status, rather than temperature because the *in vitro* exponential (10^8 CFU/g) libraries clustered very tightly, as did the *in vivo* exponential (10^8 CFU/g) libraries despite the samples being collected at different temperatures.

Based on regression analysis, count data were found to be more consistent among *in vitro* and *in vivo* 37 °C treatments compared to the 25 °C *in vivo* treatments (Figure 4.13 and supplementary Figure S2 - Figure S6). Count data from anti-sense RNA (asRNA) were found to be highly variable between some specific libraries in both *in vivo* and *in vitro* exponential 25 °C treatments (Figure S3 and Figure S6). Among *in vitro* libraries, the count data for protein coding genes were found to be highly variable between the two stationary 25 °C libraries (Figure S6) and this result was consistent with the result of the RLE analysis. Overall the entire RNA-seq data were found to include a relatively large degree of biological diversity, with a BCV of 34 % (generally accepted to range from 0.1 – 0.4, but there is not currently a published or standardized acceptable value range for bacterial transcriptome).

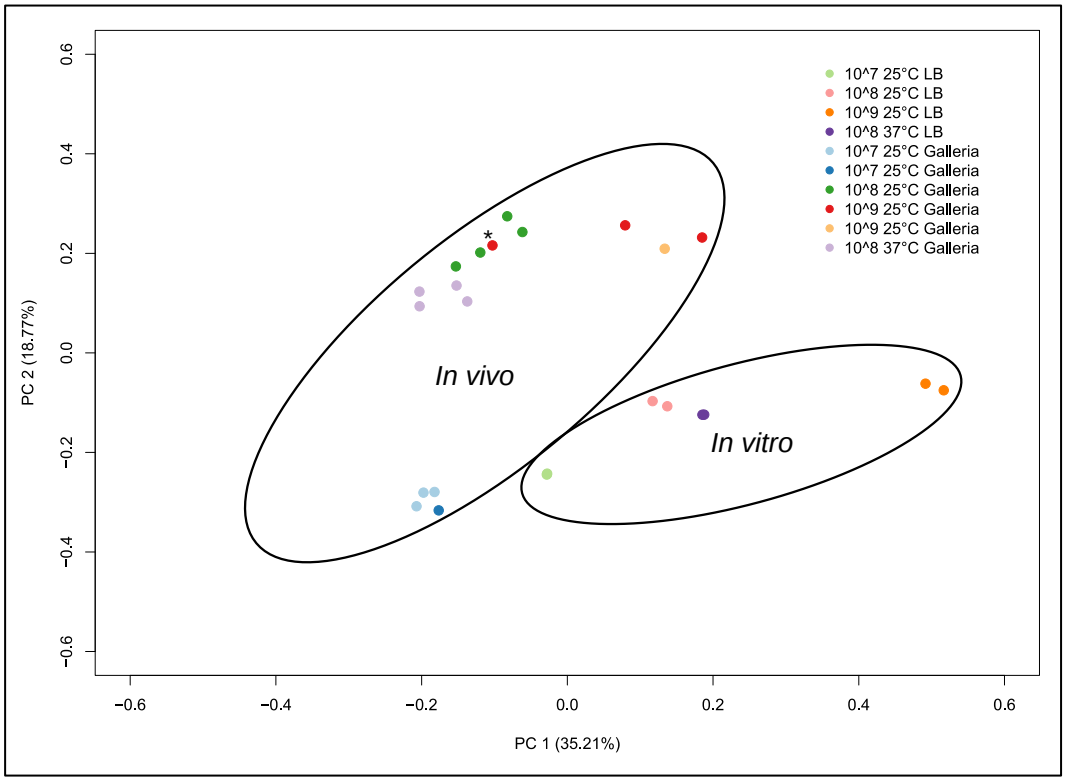


Figure 4.12. Principle component analysis of *in vivo* and *in vitro* library upper quartile normalized count data. * denotes outlier.

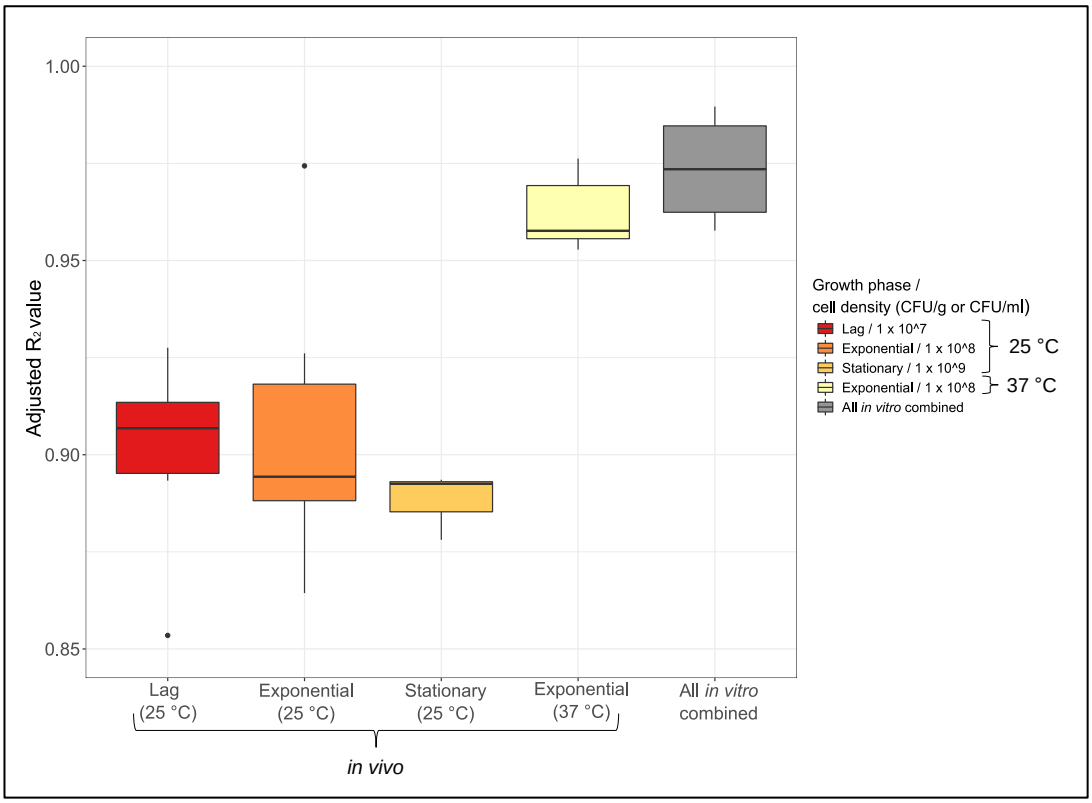


Figure 4.13: Expression variability represented as adjusted R_2 values for MH96 *in vivo* and *in vitro* libraries (*in vivo:* n = 4, *in vitro*: n = 8). Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) with whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively. Outliers are drawn as points.

4.3.6 Mean-variance transformation and differential expression

Transformation of the data using the mean-variance trend revealed a high degree of biological variation in the data because the fit-line asymptotes at a moderate level relatively early (Figure S1). If the data were from purely technical replicates, the fit line would steadily decrease as a function of the mean (no asymptote) since technical variation decreases with increasing count sizes while biological variation does not (Law et al., 2014; McCarthy et al., 2012). This variation correlated with the relatively large BCV.

From the 25 ^{0}C *in vivo* vs. *in vitro* contrasts, 2,397 (56.5 % of total) DE transcripts were identified (Figure 4.14). Among these DE transcripts, 93.1 % were protein coding genes,

while 4.6 and 1.3 % were from predicted non-coding RNA (ncRNA) and asRNA, respectively.

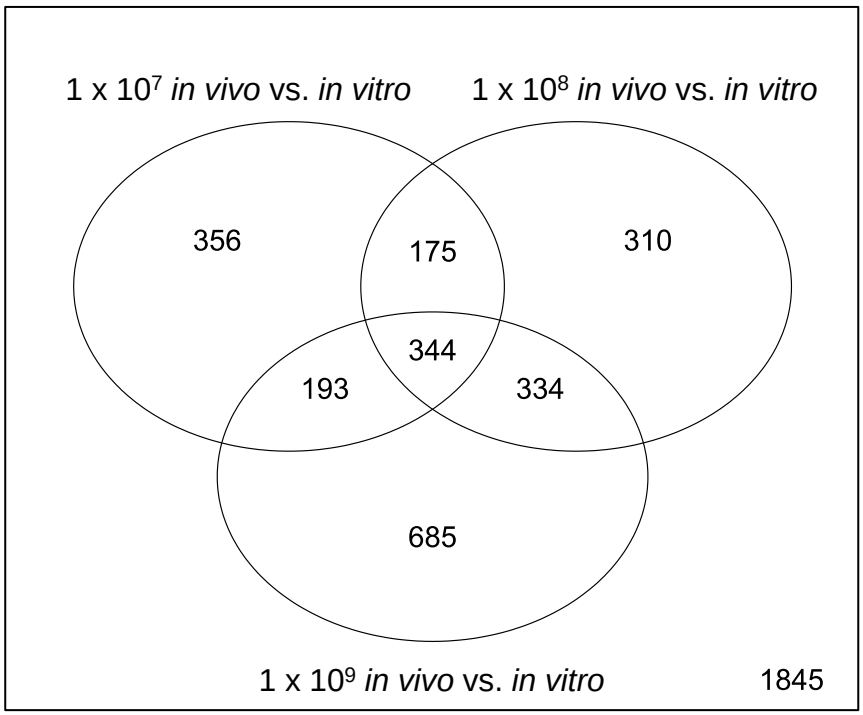


Figure 4.14. Differentially expressed transcripts from *in vivo* or *in vitro* libraries across 10^7, 10^8 and 10^9 CFU/ml or CFU/g cell densities. Significance was multiple-test corrected using the Benajmini and Yekutieli method with an adjusted p-value threshold of 0.05.

4.3.7 *In vivo* transcriptome analysis and identification of key putative VFs of *Y. entomophaga*

4.3.7.1 Fuzzy clustering, characterization of cluster type and functional enrichment for virulence factors using VFDB

The c-means fuzzy clustering algorithm were applied only to DE transcripts identified by comparing *in vivo* to *in vitro* libraries at 25 °C (i.e., 37 °C treatments were not included). The fold-change of $\log_2$ CPM values for the three cell densities (10^7, 10^8, and 10^9 CFU/g or CFU/ml) were used to generate 20 total clusters and 8 of these clusters contained a total of 798 host-specific transcripts with significantly higher expression *in vivo* compared to *in vitro* at one or more growth phases (Supplemental Figure S7). Among the host-specific transcripts, 28.7 % could be assigned a VF category from the VFDB reference database and were considered in subsequent functional enrichment analysis (Figure 4.15, Supplemental Table S8). Six clusters were determined to contain

transcripts that had stronger responses to *in vitro* growth conditions compared to *in vivo*, while the remaining six clusters generally contained transcripts that responded more subtly to *in vivo* or *in vitro* growth conditions (i.e., null clusters).

The eight *in vivo* clusters could be further organized into three main "cluster types" that were assigned by the expression profiles (Figure 4.15A). *In vivo* clusters that exhibited significantly higher *in vivo* expression across all three cell densities were identified as "Host All" type clusters (Figure S8 - Figure S10), while "Host Early" and "Host Late" clusters were characterized by having more density-dependent expression profiles, with significantly higher expression observed *in vivo* at either 10^7 or 10^9 CFU/g, respectively (Figure 4.15A, Figure S11 - Figure S15). Putative host-specific VFs were functionally classified using the following categories: adherence, defence response, horizontal gene transfer (HGT), iron acquisition, metabolic adaptation, mobility, outer membrane, regulation, secretion system and toxin/effector or unclassified (Figure 4.15B and Figure 4.16) according to (Chapter 2, Table 2.1). Together these factors represent the main consortia of genes that demonstrated a very strong transcriptional response to *in vivo* growth conditions at 25 °C and likely contribute to pathogenesis of MH96 against insects.

4.3.7.2 Functionally diverse putative virulence factors identified among the *in vivo* clusters

Diverse functional categories related to virulence were identified among the *in vivo* clusters, with genes related to iron acquisition, secretion systems, metabolism and outer membrane commonly found within several clusters and among all different cluster types (Figure 4.15). Notably, a large proportion of genes related to iron acquisition systems and the flagella were identified among Host Late *in vivo* type clusters (Figure 4.16). Among Host All clusters functional categories such as adherence and outer membrane where more prevalent compared to the other cluster types. An usher-chaperone fimbrial cluster (PL78_12465-12480) and co-expressed putative T3SS lipoprotein chaperone (PL78_14485) were the only constituents of one *in vivo* cluster and represent a potentially important adhesion used by MH96 to attach to the hemocytes of *G. mellonella*. Other potential adhesions identified using this *in vivo*

transcriptome approach included a chitin binding protein (PL78_08295) and a filamentous hemagglutinin (PL78_11060) that were identified among clusters containing genes with significantly higher expression *in vivo* at all cell densities. Several genes related to LPS and outer membrane proteins were also identified as having constitutively higher expression *in vivo* compared to *in vitro* conditions across all cell densities.

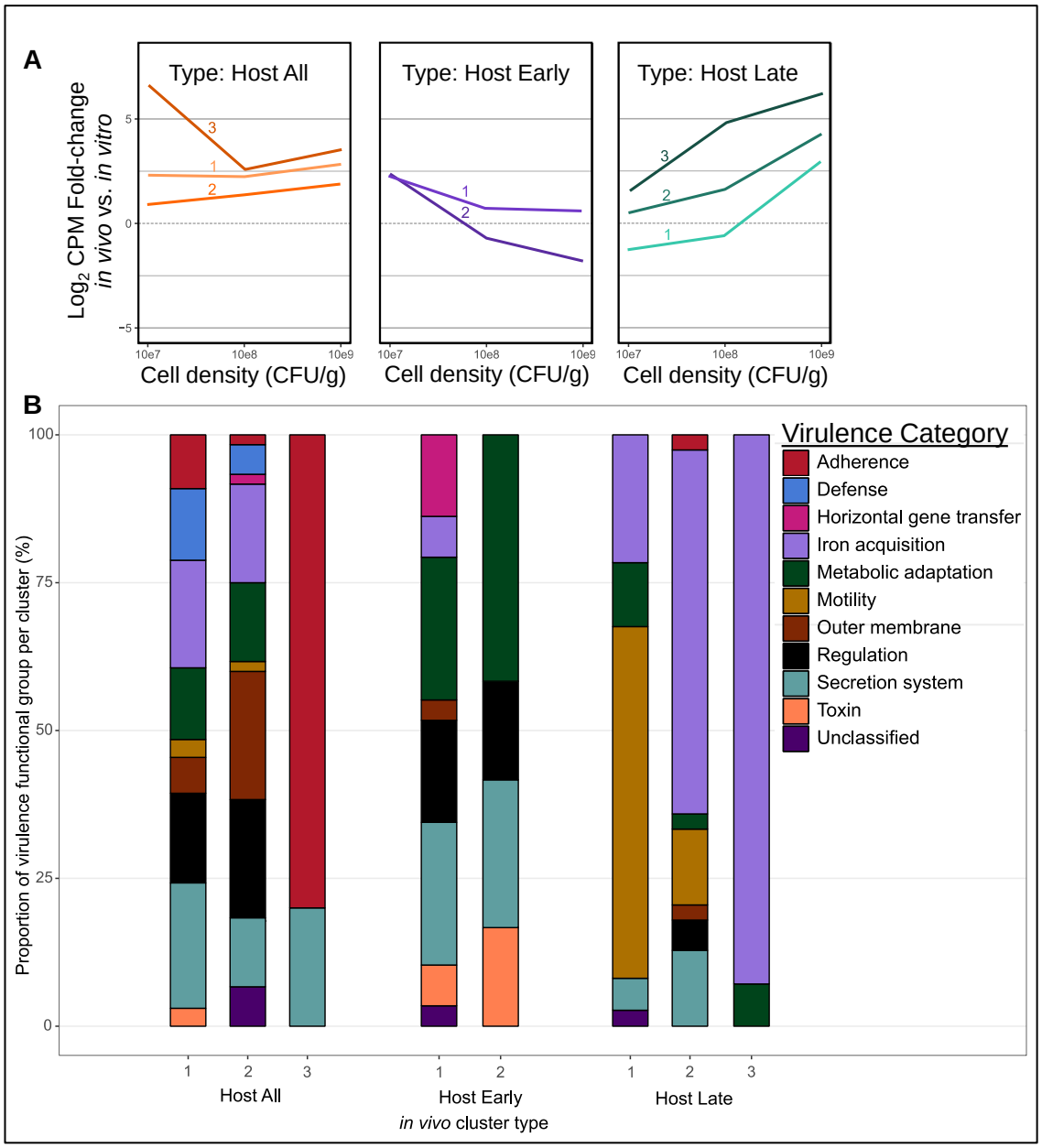


Figure 4.15: *In vivo* gene expression clusters considered to be putative host-specific virulence factors produced by *Yersinia entomophaga* MH96. A) Median $\log_2$ counts-per-million-fold-change between *in vivo* and *in vitro* libraries organized by *in vivo* cluster type. B) The proportion of genes organized by virulence function classification among *in vivo* clusters types. Refer to supplementary information (Supplementary Table S8, Supplementary figures Figure S7 to Figure S15).

Metabolism was one VF category that was found among all but one of the *in vivo* clusters. Genes related to metabolism of biotin (PL78_01660/PL78_01665, P78_05235 and PL78_05695), arginine (PL78_13305 and PL78_18495) and polypol (PL78_07325), phenylacetate catabolism (PL78_15405) and transportation of nitrate (PL78_05140), phosphate (PL78_12520), peptides (PL78_12040) and sugar (PL78_18960) were found among early type clusters. Metabolism related genes associated with late type clusters

included transport of phospholipids (PL78_07330), phosphates (PL78_05340) and sugars (PL78_11005 and PL78_12655) and synthesis of pyridoxal (PL78_09355) and purine (PL78_15720). Throughout all cell densities, metabolic genes were identified as highly expressed *in vivo*, including *citE* (PL78_17355) from the excisable holin-endolysin/citrate fermentation island HCUI$_{YE96}$ (see 0), transportation of sugar (PL78_12665 and PL78_07145) and urea (PL78_03840), nucleotide (PL78_06540, PL78_04775, PL78_15725 and PL78_11610) and lipid metabolism (PL78_18280) and synthesis of lipid (PL78_18340), tryptophan (PL78_18645), and L-cysteine (PL78_00505).

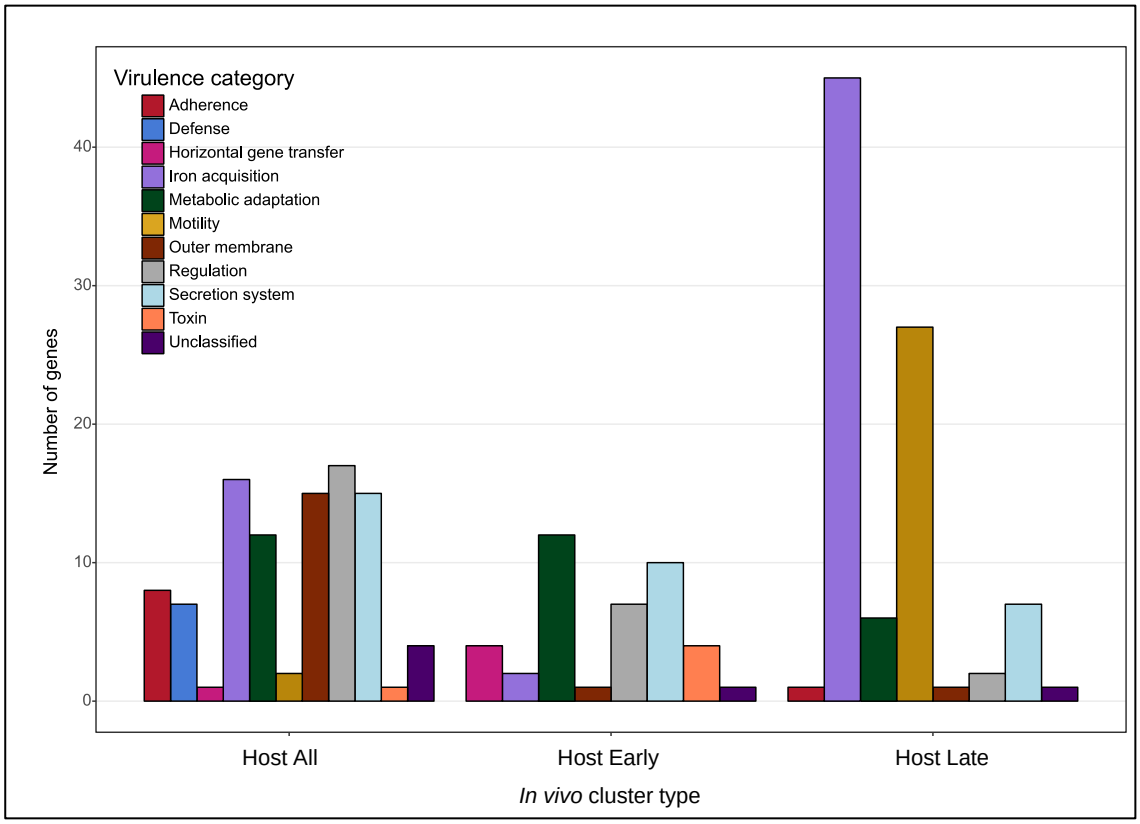


Figure 4.16: Number of genes assigned to each virulence category for putative virulence factors identified from *in vivo* cluster types.*In vivo* clusters combined by type: All = constitutively higher *in vivo* expression at 1 x 10^7, 1 x 10^8 and 1 x 10^9 CFU/g infection densities; Early = higher *in vivo* expression at 1 x 10^7 CFU/g infection density; Late = higher *in vivo* expression at 1 x 10^9 CFU/g infection density.

Genes sharing high sequence similarity to putative regulators related to virulence in other pathogen systems, were also a common occurrence among *in vivo* clusters. Specific types of transcriptional regulators, including two-component regulatory systems (TCRSs) (PL78_00360, PL78_03895 and PL78_14490) LysR-like (PL78_08895,

PL78_09565 and PL78_09850) and LacI-like (PL78_08875, PL78_13455 and PL78_16755) were frequently identified as highly expressed *in vivo* across all cell densities. Two genes encoding putative AraC-like regulators were found among the *in vivo* clusters, and PL78_17417, which is located on HCUI$_{YE96}$ (Chapter 6, section 6.3.1, Table 6.7), was also found to have significant transcriptional response to *in vivo* growth conditions during early infection. While not identified by VFDB annotation, the PAI$_{YE96}$-encoded *yen7* putative winged-helix-turn-helix (wHTH) regulator (PL78_03735) was also identified as having much higher expression *in vivo* across all cell densities compared to *in vitro* growth conditions.

Throughout the infection, MH96 also expressed genes related to ribosomal protein acetylation (PL78_13530), sigma factors (PL78_09636), methyltransferases (PL78_09665) and quorum sensing (PL78_18275), which also could have a role in regulating virulence. While not identified with a VFDB annotation, genes for other potentially important regulators found among *in vivo* clusters included a putative DNAse gyrase inhibitor (PL78_14045) and RNAse inhibitor (PL78_10660), which were more highly expressed *in vivo* throughout the infection. Another gene involved in ribosomal protein acetylation, PL78_04530, was also identified as highly expressed during early infection, along with two diguanylate cylases (PL78_04600 and PL78_18450), a TCRS (PL78_01960), LysR- (PL78_17190) and IclR-like (PL78_06535) regulators. Not as many putative virulence regulators were identified from the late type clusters, except for IscR- (PL78_15445) and LacI-like (PL78_12675) transcriptional regulators.

4.3.7.3 Identification of putative toxins, secretion systems and effectors with significantly higher *in vivo* expression

Several genes for putative toxins and effectors were identified among *in vivo* clusters Figure 4.17. The heat-stable enterotoxin *yenT* (PL78_03785, encoding in Rhs-associated region 1 of the MH96 genome) was the only toxin gene identified to be highly expressed *in vivo* across all cell densities; however, several other genes for putative toxins were identified from early type clusters, including putative cytolethal distending toxin (CDT) sub-components *cdtA* and *cdtB* (PL78_18440-5), *yenC3* (PL78_18780, a YenC homolog also known as *rhsA*) and putative vegetative insecticidal protein (VIP) component, *vip2* (PL78_16145). Beyond the genes for putative toxins that

were functionally annotated from VFDB, another potential toxin gene encoded on PAI_{YE96}, *yenU* (PL78_03775) was identified as highly expressed *in vivo* across all cell densities as well and expression of *chi1*, an integral component of the Yen-TC holotoxin, was also found to cluster among an early-type cluster but was not identified with a VFDB annotation. Unexpectedly, the genes for putative PirAB binary toxin (PL78_09590 and PL78_09595) were not identified as having higher expression *in vivo* and actually were found to have higher expression at 10^8 – 10^9 CFUs/ml *in vitro* at 25 °C.

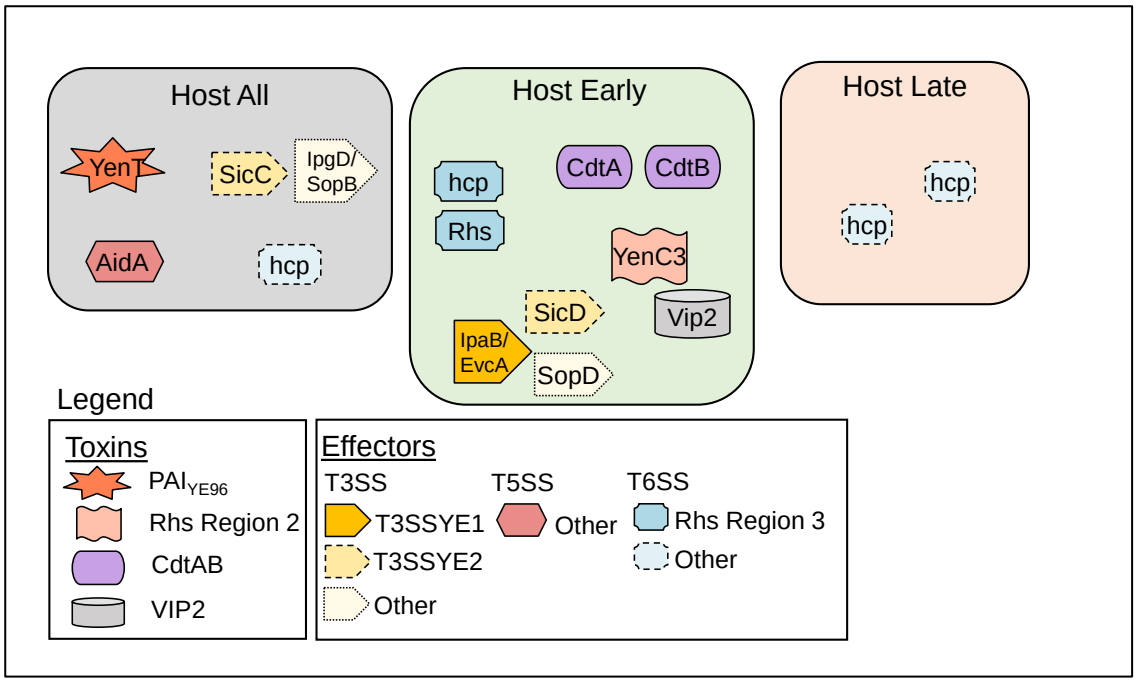


Figure 4.17: Functional enrichment of putative toxins and effectors of *Yersinia entomophaga* MH96 with significantly higher expression during infection of *Galleria mellonella.*

This *in vivo* transcriptome approach was also able to identify several putative effectors and secretion system components among all types of host-specific clusters. T3SSYE2 components and putative effectors SicC-like (PL78_14610/SipC/YspC) and IpgD/SopB-like (PL78_07345) were found to be highly expressed *in vivo* across all cell densities. In addition to this, several T3SSYE1 and T3SSYE2 components and effectors were found among early cluster types, including putative effectors SicD-like (PL78_14615/ SipD/ IpaD /YspD), IpaB/EvcA-like (PL78_18085) and SopD-like (PL78_14295). Furthermore, an AidA-like type 5 secretion system (T5SS) secreted adhesin (PL78_10240) and several

components of the T2SS were found among clusters with high *in vivo* gene expression profiles. While not identified by VFDB annotation, the holin-endolysin secretion system (PL78_17390 – PL78_17400) was also found to be associated with early *in vivo* cluster types.

Interestingly, although there were no genes related to type 6 secretion system (T6SS) structural components identified among *in vivo* clusters, five different putative T6SS effector genes were found associated with each of the *in vivo* cluster types. These T6SS effectors included rearrangement hot-spot (Rhs)-containing protein (PL78_01040) and five different hcp-type structural tube/effector proteins (PL78_02751, PL78_03598, PL78_03602, PL78_09211 and 00900). Also, the gene for putative T6SS effector Spt4 (PL78_00995) was found to have high *in vivo* gene expression across cell densities but was not assigned a VFDB annotation.

4.3.7.4 Evidence of response to stress and mobilization of horizontally-acquired genomic elements among *in vivo* clusters

Several efflux pumps (PL78_17050, PL78_17050, PL78_03990, PL78_14185 and PL78_07730) were highly expressed by MH96 during infection as well as xenobiotic transporter (PL78_05220) and serine dehydrogenase (PL78_03970) and were categorized under the virulence category of defence response. Several other putative stress response regulatory genes were noted as highly expressed by MH96 during *in vivo* growth conditions compared to *in vitro*; however, these were not identified by the VFDB annotations. These included the genes for three highly repetitive tandem cold-shock proteins (CSPs) - *cspA1* (PL78_17450), *cspA2* (PL78_18365), *cspA3* (PL78_18370) and a fourth cold-shock protein *cspB1* (PL78_15000) that were found to cluster with other genes that had constitutively higher *in vivo* expression across all cell densities. Other stress response genes were found to have greater expression only during early infection, including the exact duplicate of *cspA1*, *cspA4* (PL78_17451) and acid-shock and resistance proteins (PL78_14965 and PL78_17435). A sixth cold-shock protein gene, *cspC* (PL78_17485), was identified as having *in vivo* expression, but it was much higher during late infection compared to *in vitro* conditions.

Interestingly, four transposases (PL78_03345, PL78_04200, PL78_12162 and PL78_00880) and an integrase (PL78_18775) were identified among two different *in vivo* clusters and were classified under the virulence category of HGT. Several of these genes are located within previously recognized Rearrangement hot spot (Rhs)-associated regions of the genome (Hurst et al., 2016) including transposases, PL78_12162 and PL78_00880, located within Rhs3 (T6SS) and Rhs4 and integrase (PL78_18775) located within Rhs2. The VFDB annotations did not capture the full extent of HGT-related transcripts identified among the *in vivo* clusters (Figure 4.18). Among all *in vivo* clusters combined, genes associated with HGT, included 9 integrases, 11 transposases, 3 recombinases and 1 excisionase genes. Remarkably, 33 out of 77 tRNA loci were also highly expressed among *in vivo* clusters, including multiple tRNA-Ala, -Arg, -Asn, -Gln, -Gly, -Leu, -Ser and -Val loci (Figure 4.18). The largest number of tRNAs with higher expression *in vivo* were found to be associated with late cluster types. Similarly, a large number of phage-related transcripts (PL78_09145 – 09340) were identified from late cluster types but were not assigned functional annotations from VFDB.

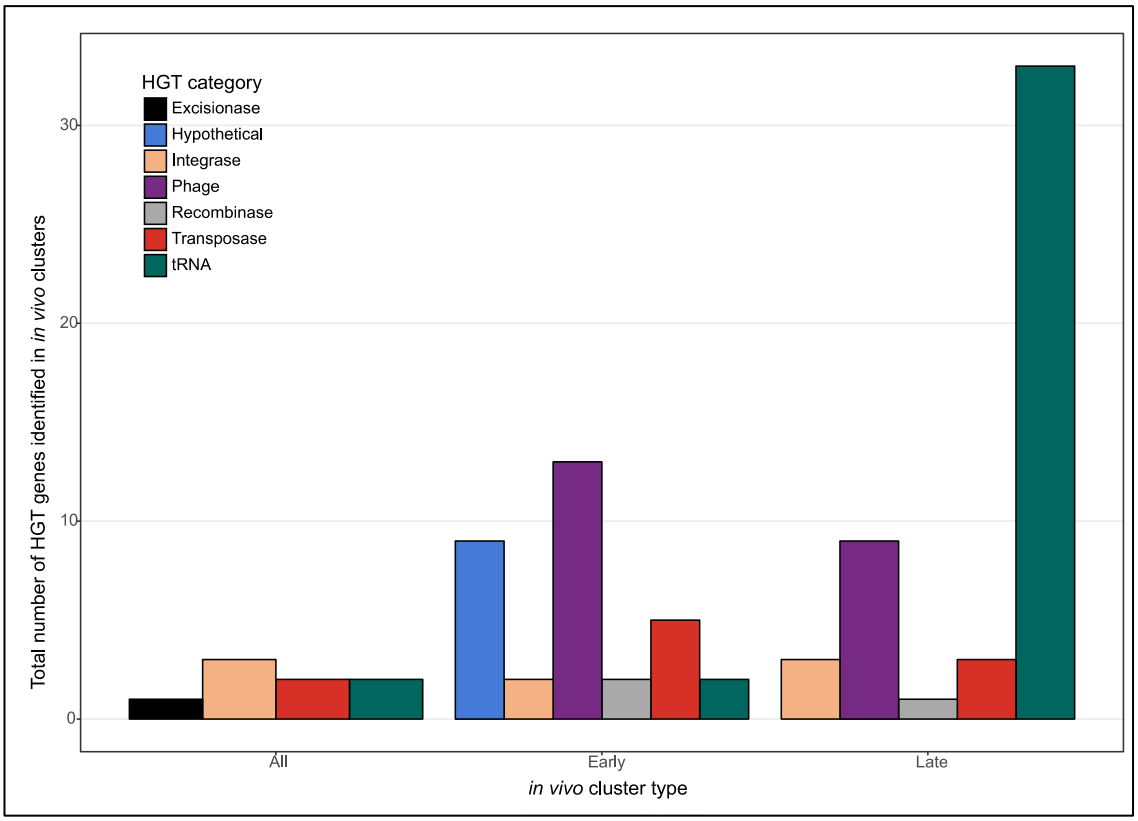


Figure 4.18: Number of genes identified from *in vivo* cluster types potentially involved in horizontal gene transfer. *In vivo* clusters combined by type: All = constitutively higher *in vivo* expression at 1 x 10^7, 1 x 10^8 and 1 x 10^9 CFU/g infection densities; Early = higher *in vivo* expression at 1 x 10^7 CFU/g infection density; Late = higher *in vivo* expression at 1 x 10^9 CFU/g infection density.

Several phage-related transcripts were also found to be more highly expressed at low cell density *in vivo*, including some transcripts from two different regions, including PL78_00090 to 00265 and PL78_06330 to 06360/PL78_19095 to 19210. These regions appear to encode putative prophage or bacteriophage elements; however, many of the predicted genes found within these regions encode hypothetical proteins of unknown function. One of these hypothetical genes, PL78_19155 shares significant sequence similarity (E-value 1e-13, 27 % similarity and 65 % coverage) with a secreted protein (SP3/LOC100164129) that was found to be highly secreted by host aphid, *Acyrthosiphon pisum* into specialized bacteriocyte organs that harbour their obligate endosymbiont *Buchnera aphidicola* (Shigenobu and Stern, 2013), which represents a good candidate for a potential gene that could be beneficial for insect colonization or pathogenesis by MH96. Only one other *Yersinia* isolate, *Yersinia* sp. FDAARGOS_228 (wildlife lymph node isolate from US, CP020409.2) was found to encode a gene with significant sequence homology (E-value 0.0, 85.9 % similarity and 99 % coverage) to PL78_19155,

which also appears to be associated with a prophage as well. Homologous sequences to PL78_19155 could also be widely detected among isolates of *Sa. enterica* subsp. *enterica* and *arizonae* within the GenBank nucleotide collection as of August 2019. The predicted protein secondary structure shares significant homology (98.3 % confidence/ 11 % i.d.) to the β-roll structure associated with the C-terminus of extracellular lipase, LipA, from *S. marcescens* (c2qubG) (Meier et al., 2007) using Phyre2 webtool and the β-roll domain of LipA shares similarity to the β-roll domain proteins, including serralysins.

4.3.8 Two- way factorial *in vivo*/*in vitro* – 25 vs 37 °C

A total of 2,269 transcripts representing 57.5 % of coding genes responded differently to 37 compared to 25 °C in the *in vitro* and *in vivo* treatments. Also, 1,353 (31.9 %) of all transcripts (including asRNA and ncRNA) responded differently to temperature depending on whether they were in the host or grown in culture (Figure 4.19). Of all DE transcripts, 94.2 % represented protein coding genes, while 4.5 and 1.3 % were from predicted ncRNA and asRNA, respectively.

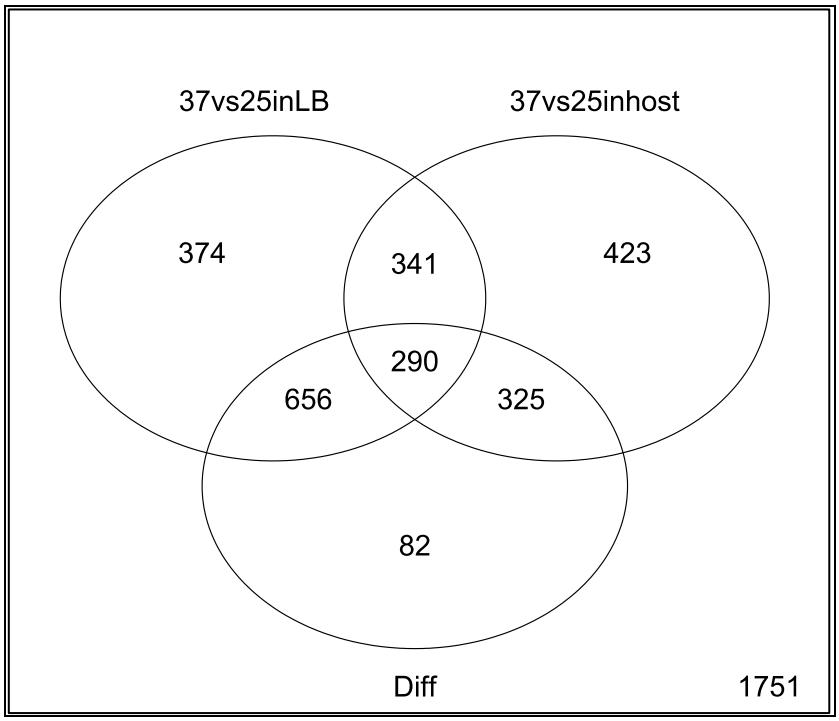


Figure 4.19: Number of differentially expressed transcripts from 25 versus 37 °C contrasts for *in vitro* and *in vivo* treatments at 1 x 10^8 CFU/g or CFU/ml and transcripts that respond differently to temperature *in vivo* or *in vitro* (Diff = interaction coefficient). Transcript significance was multiple-test corrected using the Benjamini and Hochberg method with an adjusted p-value threshold of 0.05.

Regions of interest from the draft genome (Hurst et al., 2016) were explored with respect to temperature-dependent responses, including unique regions 1 and 2, T2SS, T3SSYE1, T3SSYE2, T6SS (Rhs3), and Rhs-associated regions 2 – 5 as well as the flagellum (Figure 4.20). Most, if not all, of the flagellum and T6SS component genes were found to have significantly higher expression at 25 compared to 37 °C, in both *in vitro* and *in vivo* treatments (Table 4.11 and Table 4.12). genes for components of the T3SSYE1 were found to be more highly expressed at 25 compared to 37 °C *in vitro*, but expression was less affected by temperature *in vivo*. On the contrary, genes for components of the T2SS were more highly expressed at 25 compared to 37 °C *in vivo*, with expression less affected by temperature *in vitro*. Genes from the remaining regions of interest responded more variability to the different temperatures and *in vivo*/*in vitro* conditions than those regions described above.

Several genes for known and putative VFs were identified as sharing the same expression profile as the T2SS (i.e., significantly reduced expression only at 37 °C *in vivo*) (Table 4.13). These VFs include: all components of Yen-TC, Chi-SrfAB, peptidase

M66 (PL78_05495), RhsA/YenC3 (PL78_18780) and hypothetical protein (PL78_18785) (Rhs2-associated region), TC-associated phage-related region (PL78_17417-17386), chitinase (PL78_11910) and a hydrolase cluster (PL78_06535-06520) (Table 4.13). All these known and putative VFs share a distinctive co-expression pattern, and collectively the group will be referred to as "TC-associated factors". A gene for putative regulator, *yen7* (found directly upstream of Yen-TC component gene *chi1*), and a gene for IbpA-like heat-shock protein (PL78_12380) were also found to share the same expression profile as the TC-associated factors.

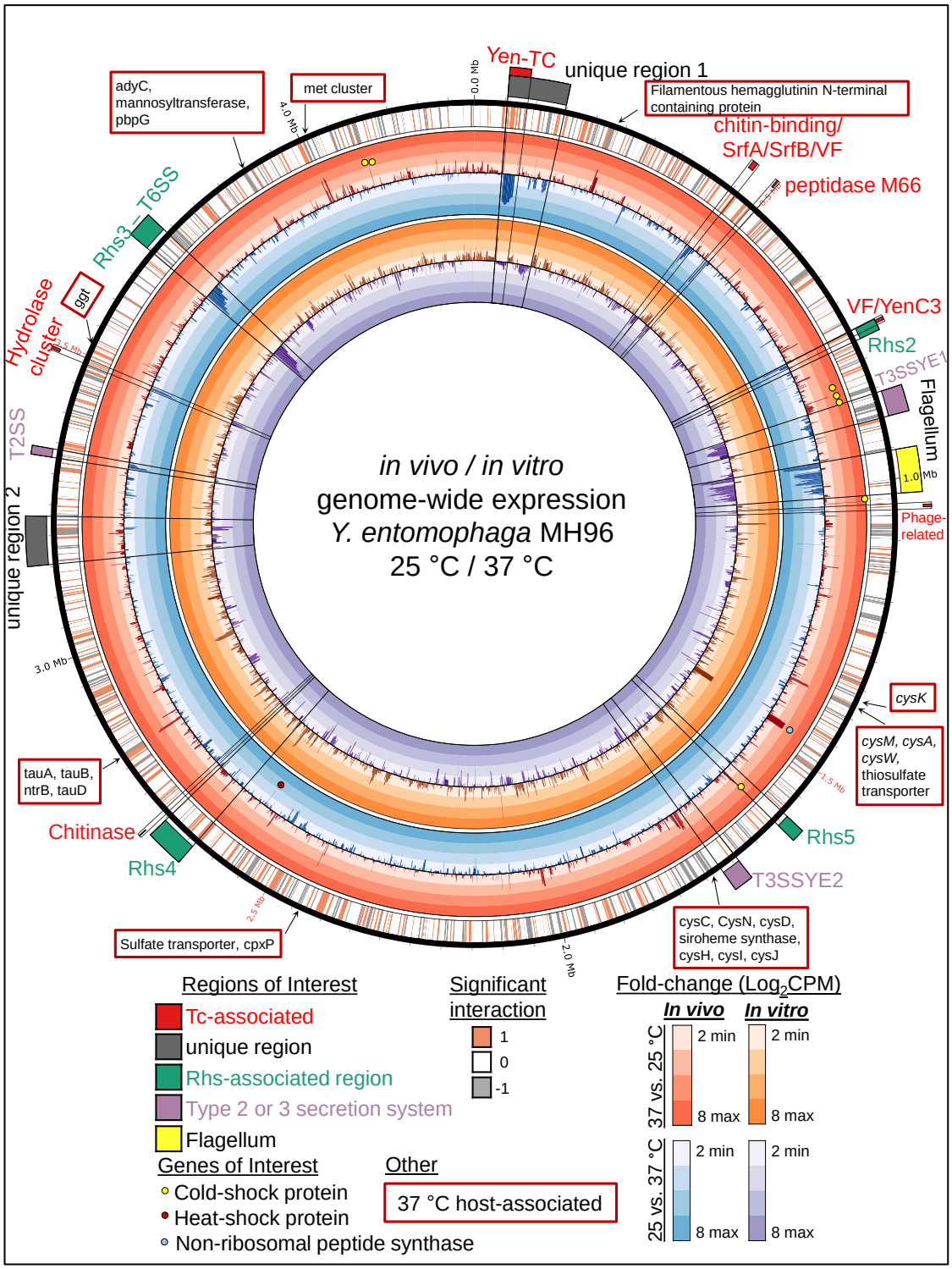


Figure 4.20: Circos plot of *Yersinia entomophaga* genome-wide transcriptional temperature response for *in vivo* and *in vitro* treatments. Further information related to flagella, type 6 secretion system (T6SS), Yen-TC and TC-associated factors is located in Table 4.11, Table 4.12 and Table 4.13 below, respectively. CPM = Counts-per-million. Mean fold-change Log_2CPM calculated with limma/voom method (*in vivo:* n =4, *in vitro*: n =2) on upper quartile normalized count data output from Rockhopper. Additional details related to genes with significantly higher expression *in vivo* at 37 compared to 25 °C are located in Supplementary Table S9.

Table 4.11: Mean 37 vs. 25 °C log_2CPM fold-change for the flagellum components encoded by *Yersinia entomophaga* MH96 (*in vivo*: . CPM = counts-per-million, VFDB = virulence factor database, F.p.value = moderated F *p*-value.

Operon	Locus tag	Gene product	37 vs. 25 °C		F.p.value	VFDB E-value
	Flagellum		*In vitro*	*In vivo*		
↓	PL78_17795	FliZ	-5.61	-6.53	2.49E-10	2.0E-119
↓	PL78_17790	FliA	-5.91	-6.33	3.99E-12	1.0E-165
	PL78_17785	FliB	-1.52	-2.44	4.77E-07	0.0E+00
	PL78_17780	FlgL	-6.07	-7.09	9.27E-14	6.0E-160
	PL78_17787	FlgL	-6.11	-7.09	3.86E-15	1.0E-155
	PL78_17786	FlgL	-6.36	-7.06	3.03E-12	7.0E-172
↓	PL78_17775	FlgL	-4.52	-6.33	1.49E-13	3.0E-99
↓	PL78_17776	hypothetical	-3.40	-3.89	2.57E-05	
↓	PL78_17770	FliD	-4.32	-4.06	5.18E-14	0.0E+00
↓	PL78_17765	FliS	-1.94	-2.85	8.12E-07	6.0E-83
↓	PL78_17760	FliT	-1.52	-2.79	9.19E-07	4.0E-62
	PL78_17740	endonuclease	-1.95	-5.23	1.59E-09	
	PL78_17735	FliE	-4.53	-4.50	3.56E-14	6.0E-57
↓	PL78_17730	FliF	-5.00	-5.28	9.64E-15	0.0E+00
↓	PL78_17725	FliG	-4.69	-4.76	5.09E-12	0.0E+00
↓	PL78_17720	FliH	-4.62	-5.13	1.82E-11	6.0E-99
↓	PL78_17715	FliI	-3.89	-5.20	6.78E-13	0.0E+00
↓	PL78_17710	FliJ	-3.62	-3.94	1.68E-09	2.0E-83
↓	PL78_17705	FliK	-3.41	-4.51	4.12E-11	4.0E-98
↓	PL78_17700	FliL	-4.12	-3.73	4.86E-13	3.0E-60
↓	PL78_17695	FliM	-3.92	-4.08	3.20E-15	0.0E+00
↓	PL78_17690	FliN	-4.30	-4.32	7.25E-11	6.0E-92
↓	PL78_17685	FliO	-4.17	-3.72	7.39E-10	2.0E-48
↓	PL78_17680	FliP	-3.47	-3.46	7.54E-07	5.0E-154
↓	PL78_17675	FliQ	-3.40	-2.63	3.63E-07	1.0E-34
↓	PL78_17670	FliR	-3.97	-2.03	1.00E-05	3.0E-117
↓	PL78_17665	FlgL	-1.85	-1.67	7.76E-07	4.0E-177
↓	PL78_17660	FlgK	-3.84	-4.06	1.18E-10	0.0E+00
↓	PL78_17655	FlgJ	-1.85	-2.46	7.45E-09	0.0E+00
↓	PL78_17650	FlgI	-2.68	-2.82	1.49E-09	0.0E+00
↓	PL78_17645	L-ring protein	-3.13	-3.59	1.57E-09	2.0E-110
↓	PL78_17640	FlgG	-4.18	-4.60	1.23E-10	1.0E-165
↓	PL78_17635	FlgF	-4.52	-5.27	8.76E-13	1.0E-155
↓	PL78_17630	FlgE	-4.73	-5.18	5.57E-15	0.0E+00
↓	PL78_17625	FlgD	-5.21	-4.91	9.68E-13	6.0E-101
↓	PL78_17620	FlgC	-4.94	-5.09	2.47E-13	1.0E-74
↓	PL78_17615	FlgB	-5.72	-5.70	6.01E-12	1.0E-89
	PL78_17610	FlgA	-3.03	-3.30	1.52E-13	8.0E-96
↓	PL78_17605	FlgM	-0.92	-1.63	9.79E-05	9.0E-41
↓	PL78_17600	FlgN	-1.36	-1.49	7.14E-04	3.0E-74
↓	PL78_17595	FlhE	-2.11	-1.23	9.22E-04	3.0E-49
↓	PL78_17590	FlhA	-1.77	-2.14	3.62E-07	0.0E+00
↓	PL78_17585	FlhB	-3.23	-3.96	1.04E-09	0.0E+00
	PL78_17580	MycP	-1.13	1.47	6.96E-06	
↓	PL78_17575	CheZ	-1.68	-1.26	7.10E-07	2E-122
↓	PL78_17570	PhoB	-1.19	-1.31	2.25E-06	1E-72
↓	PL78_17565	CheB	-1.27	-1.50	6.03E-07	0.0E+00
↓	PL78_17560	CheR	-2.18	-2.32	2.94E-09	0.0E+00
↓	PL78_17555	Tap (MCP-IV)	-2.95	-2.27	3.29E-09	2E-123
↓	PL78_17550	Tse	-4.21	-3.80	4.03E-12	0.0E+00
	PL78_17545	CheW	-1.66	-1.48	5.03E-05	8E-95
↓	PL78_17540	TorS	-1.62	-2.12	5.00E-07	0.0E+00
↓	PL78_17535	MotB	-2.28	-2.46	6.77E-09	2E-165
↓	PL78_17530	MotA	-4.00	-2.96	1.86E-09	1E-167
↓	PL78_17525	FlhC	-1.24	-1.27	4.21E-05	5E-138
↓	PL78_17520	FlhD	-1.10	-1.71	6.62E-07	2E-69

Fold-change (Log_2CPM)

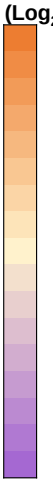


Table 4.12: Mean 37 vs. 25 °C log_2CPM fold-change for the type 6 secretion system components found in *Rhs*-associated region 3 encoded by *Yersinia entomophaga* MH96 (*in vivo*: n = 4, *in vitro*: n =2). CPM = counts-per-million, VFDB = virulence factor database, F.p.value = moderated F *p*-value.

Operon	Locus tag	Gene product	37 vs. 25 °C		F.p.value	VFDB E-value	Fold-change (Log_2CPM)
	T6SS		*In vitro*	*In vivo*			
	PL78_00900	Hcp	-6.99	-5.61	5.36E-15	4.0E-115	8
	PL78_00905	VipA	-7.20	-5.52	2.99E-14	2.0E-75	7
	PL78_00910	VipB	-7.07	-5.30	1.18E-16	0.0E+00	6
	PL78_00915	Lys	-6.67	-4.76	2.55E-13	7.0E-44	5
	PL78_00920	TssF	-5.95	-4.27	2.46E-13	0.0E+00	4
	PL78_00925	TssG	-5.96	-4.53	1.53E-10	6.0E-136	3
	PL78_00930	Type	-6.35	-4.36	2.71E-11	4.0E-123	2
	PL78_00935	TssJ	-5.47	-4.56	3.03E-08	7.0E-63	1
	PL78_00940	ImpJ	-5.16	-3.65	2.21E-11	0.0E+00	0
	PL78_00945	VasF	-5.18	-3.46	2.28E-11	1.0E-110	-1
	PL78_00950	ClpV	-4.99	-3.74	1.32E-10	0.0E+00	-2
	PL78_00955	Fis	-4.45	-3.10	2.68E-10	2.0E-17	-3
	PL78_00960	VasI	-4.89	-3.50	2.5E-08	2.0E-61	-4
	PL78_00965	ImpA	-4.95	-3.68	1.33E-09	7.0E-123	-5
	PL78_00970	IcmF	-4.59	-2.97	1.07E-09	0.0E+00	-6
	PL78_00975	ImpA,	-4.71	-2.70	2.83E-11	3.0E-115	-7
	PL78_00980	VgR	-4.54	-3.33	4.58E-10	0.0E+00	-8
	PL78_00985	hypothetical	-4.08	-3.20	9.47E-10		
	PL78_00990	Rhs PAAR	-1.86	-1.68	1.29E-09	2.0E-20	
	PL78_00995	Spt4	-0.13	-1.22	0.03		

In contrast, very few putative VF genes were found to be highly expressed at 37 °C, especially among the regions of interest except for the putative regulator *yen6* (located on PAI_{YE96} upstream of *yen7*), which was the most highly DE transcript between 37 compared to 25 °C *in vivo* from the entire genome. The gene for *yen6* also had significantly higher expression at 37 compared to 25 *in vitro*, but the expression of *yen6* was clearly enhanced by the *in vivo* growth conditions.

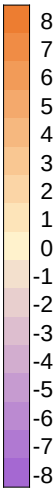
Among the top 100 most DE genes at 37 compared to 25 °C *in vivo* were putative adenylate cyclase (PL78_01795), porin OmpC (PL78_16930) and NRPS (PL78_15621) as well as a large number of genes related to non-organic and organic sulfur acquisition (Supplementary Tables Table S9Table S10). Furthermore, filamentous hemagglutinin N-terminal containing protein (PL78_04365, previously annotated hemolysin) and cognate two-partner secretion system (TPS) (PL78_04360), which is otherwise known as T5bSS and is currently annotated as secretion/activation protein ShlB/FhaC/HecB and another TPS (PL78_11055). MH96 encodes genes for two filamentous hemagglutinin N-terminal containing proteins, the aforementioned PL78_04365, and

PL78_11060, which was identified has having higher expression *in vivo* at 25 °C during early infection (Chapter 4, section 4.3.7.2 and supplementary Table S8). Also, genes for several CSPs (some identified previously among host-specific clusters at 25 °C), including three tandem *cspA1, cspA2 and cspA3, cpsA4*, two tandem *cspE2* and *cspB2* and an acid-resistance protein (PL78_17435), which all had significantly higher expression at 37 °C in the host, compared to 25 °C at 10^8 CFU/g cell density.

Table 4.13: Mean 37 vs. 25 °C log_2CPM fold-change for Yen-TC and TC-associated factors encoded by *Yersinia entomophaga* MH96 (*in vivo*: n = 4, *in vitro*: n =2). CPM = counts-per-million, VFDB = virulence factor database, F.p.value = moderated F *p*-value.

Operon	Locus tag	Gene product	37 vs. 25 °C		F.p.value	VFDB E-value
	PAI$_{Ye96}$		*In vitro*	*In vivo*		
	PL78_03730	Yen6	4.48	7.64	1.3E-12	
	PL78_03735	Yen7	0.27	-3.65	3.9E-08	
	Yen-TC					
↓	PL78_03740	Chi1	-0.53	-5.16	4.3E-10	
	PL78_03745	YenA1	-0.39	-5.50	6.1E-10	3.0E-09
	PL78_03750	YenA2	-0.57	-5.86	6.2E-10	
	PL78_03755	Chi2	-0.66	-5.64	1.4E-09	
↓	PL78_03760	YenB	-0.62	-5.33	5.7E-09	8.0E-92
	PL78_03765	YenC1	-0.82	-4.58	1.2E-07	0.0E+00
	PL78_03770	YenC2	-1.47	-4.31	8.8E-06	0.0E+00
	Chi-SrfAB					
	PL78_05310	CbpA, LCMO/CBP	-1.46	-5.56	1.3E-06	7.0E-12
↓	PL78_05315	SrfA	-2.31	-3.93	2.0E-10	
	PL78_05320	SrfB	-1.83	-3.01	2.4E-10	
	PL78_05325	hypothetical VF	-1.71	-2.25	2.1E-10	5.0E-75
	Peptidase					
	PL78_05495	peptidase M66	-2.94	-5.33	2.0E-06	0.0E+00
	Rhs2					
	PL78_18780	YenC3	-0.64	-1.80	2.0E-06	0.0E+00
	PL78_18785	hypothetical VF	-0.89	-5.22	1.2E-07	
	Tc Phage-associated					
	PL78_17417	AraC-like	-0.58	-3.82	1.9E-05	7.0E-16
↓	PL78_17400	hypothetical	-1.78	-7.06	2.2E-08	
	PL78_17395	peptidase M15	-1.94	-6.88	1.1E-08	
	PL78_17390	holin	-1.98	-6.28	2.5E-09	
	PL78_17385	PhoB-like	-0.94	-2.87	6.1E-06	
	PL78_17386	hypothetical	-0.89	-2.82	2.0E-06	
	Heat-shock protein					
	PL78_12380	IbpA, heat-shock protein	0.15	-3.54	8.3E-03	9.00E-29
	Chitinase					
	PL78_11910	chitinase	-1.73	-3.90	2.1E-03	
	T2SS					
	PL78_08975	response regulator	-0.96	-3.76	2.6E-03	2.0E-07
↓	PL78_08970	GspC	-1.14	-4.45	1.8E-05	
	PL78_08965	GspD	-0.61	-3.59	1.9E-08	0.0E+00
	PL78_08960	GspE	-0.61	-1.85	1.8E-05	0.0E+00
	PL78_08955	GspF	0.27	-1.70	2.1E-04	
	PL78_08950	GspG	-0.48	-1.37	0.02	3.0E-58
	PL78_08945	prepilin peptidase/Mtase	-0.26	-1.78	8.2E-04	2.0E-07
	PL78_08940	GspI	0.19	-1.23	0.05	2.0E-18
	PL78_08935	GspJ	-0.43	-1.27	7.2E-03	5.0E-22
	PL78_08930	GspK	0.13	-1.27	0.02	2.0E-08
	Hydrolase cluster					
	PL78_06535	IclR-like	-2.12	-5.34	2.9E-11	2.0E-09
↓	PL78_06530	oligopeptide transporter	-2.99	-4.33	5.5E-10	
	PL78_06525	putative hydrolase	-2.18	-3.78	4.0E-09	
	PL78_06520	putative lyase	-1.96	-2.82	1.3E-07	

Fold-change (Log_2CPM)



4.3.9 Phenotype microarray detect metabolic differences in sole carbon utilization by MH96 at 25 compared to 37 °C

In vivo RNA-seq comparing genome-wide expression of MH96 at 25 and 37 °C identified major shifts in gene expression, notably with wide-spread repression of Yen-TC and TC-associated factors, T6SS and flagella at 37 compared to 25 °C *in vivo* and these significant changes in gene expression are expected to have likely impacts on virulence and subsequent nutrient acquisition by MH96 within the host environment.

Using a phenotype microarray approach consistent changes in the abilities of MH96 to use several different sole-carbon sources at 25 compared to 37 °C were identified (Table 4.14). MH96 was able to utilize several amino acids (L-Glutamic acid, L-Glutamine, Gly-Glu and L-Threonine) as well as the carbohydrate uridine at 37 but not at 25 °C. MH96 was able to consistently utilization the polymer dextrin as a sole carbon source at 37 but not at 25 °C; however, MH96 was found to grow on dextrin in one of the 25° C replicate batch plates, making these findings slightly less reliable.

Table 4.14: Temperature-dependent differences in aerobic sole carbon source utilization by MH96 at 25 compared to 37 °C by phenotype microarray . (+) indicates positive carbon-source utilization per replicate plate, green cells = maximum value/minimum value > 2, yellow cells = maximum value/minimum value ≤ 2 and blank cells indicate no growth curve detected based on cut-off criteria.

		Temperature (°C)	
Carbon source	Chemical	25	37
amino acid	L-Glutamic acid		+ +
amino acid	L-Glutamine		+ +
amino acid	Gly-Glu		+ +
amino acid	L-Threonine		+ +
carbohydrate	Uridine		+ +
polymer	Dextrin	+	+ +
amino acid	L-Aspartic acid	+ +	
amino acid	L-Asparagine	+ +	
amino acid	L-Histidine	+ +	
carbohydrate	N-Acetyl-D-Glucosamine	+ +	
carbohydrate	D-Trehalose	+ +	
carbohydrate	D-Ribose	+ +	
carbohydrate	D-Fructose	+ +	
carbohydrate	α-D-Glucose	+ +	
carbohydrate	Sucrose	+ +	+
carbohydrate	D-Mannose	+ +	+ +
carbohydrate	D-Mannitol	+ +	+ +
carboxylic acid	L-Lactic acid	+ +	+ +

At 25 °C, MH96 was able to utilize a greater diversity of carbohydrates compared to 37 °C, including N-acetyl-D-glucosamine (GlcNAc), trehalose, ribose, fructose, glucose, sucrose and potentially mannose and mannitol, although the utilization of the latter two was found to be less reliable than the former. MH96 was also able to utilize several amino acids (i.e., L-aspartic acid, L-asparagine and L-histidine) at 25 °C but not at 37 °C, and there also appeared to be a possible temperature-dependent change in utilization of lactic acid (carboxylic acid), but this result was considered less reliable and would require further validation.

4.4 Discussion

4.4.1 Diverse putative VF encoded by MH96 respond to *in vivo* conditions, including a core insecticidal machinery regulated by host- and temperature-dependent factors.

This thesis provides a proof in principle, that as an exploratory tool, *in vivo* transcriptomics can be used for development of data-driven hypotheses and targeting of future molecular investigations of an entomopathogenic bacteria, an approach that may help drive novel biopesticide development. The application of *in vivo* transcriptomics to further characterize the host- and temperature-dependent regulatory mechanisms present in MH96 during infection of *G. mellonella* was useful in identification of putative insecticidal factors, such as toxins and T3SS and T6SS secreted effectors, as well as characterizing key regulatory themes involving Yen-TC and TC-associated factors, T3SSYE2, T6SS, flagella, stress response (including several cold-, acid- and heat-shock proteins and a NRPS), iron and sulfur acquisition. These results offer an excellent baseline on which further targeted investigations into the role of putative temperature-dependent virulence regulators in the pathobiology of MH96.

The *in vivo* transcriptome of MH96 identified diverse putative VFs that are likely to be important in pathogenesis during intrahemocoelic infection of *G. mellonella* at 25 °C. The *in vivo* transcriptome of MH96 provided genome-wide insights into gene expression of an entomopathogen during different growth stages within an insect hemocoel and under exposure to the immune system of the host. These methods produced high quality and robust RNA-seq data, revealing a complex transcriptome of both coding and diverse non-coding entities. Functional enrichment of putative VFs with significantly higher *in vivo* expression using the VFDB provided key insights into the pathogenesis process that occurs during different growth phases within the host. Using this approach several putative toxin and effector genes were found to have host-dependent expression, suggesting these primary VFs are important during infection of the insect hemocoel. Other insights included wide-spread host-dependent transcription of numerous highly-repetitive cold-shock protein genes and efflux pumps as well as genes related to HGT.

Temperature is known to have major phenotypic effects on MH96 under *in vitro* conditions, with reduced toxin production, secretion and motility observed at 37 compared to 25 °C. Using a two-way factorial experiment to investigate the role of temperature on the *in vivo* transcriptome of MH96, a major temperature- and host-dependent regulatory mechanism was revealed, linking a core set of putative VFs to

Yen-TC, including several co-secreted proteins. Other genes important VFs were found to be strictly under temperature-dependent control, including the flagella and T6SS. Finally, a lack of secretion systems was observed at 37 °C *in vivo*, but a number of putative VFs including an filamentous hemagglutinin N-terminal containing and LRR-containing proteins and *yenV*, all of which may have the potential to interact with host cells as well as increased expression of several CSPs and sulfur-acquisition systems, indicating increase response to host-induced stress at 37 °C may be encountered by MH96.

4.4.2 An improved method for RNA quality/quantity, dual rRNA depletion and alignment

This thesis describes a useful method for retrieval and enrichment of bacterial mRNA from within the hemolymph of *G. mellonella* and a very straightforward method with comparative reduced hands-on time compared to most other previously reported *in vivo* bacterial transcriptomes. The key step that has been developed (i.e., removal of host hemocytes by centrifugation prior to RNA extraction) is a technique that any molecular lab can use to enrich for *in vivo* bacteria in hemolymph (or plasma in mammals), independent of high-cost enrichment kits or complex protocols involving specialized equipment or training. Even at the lowest cell density sampled (10^7 CFU/g), a substantial amount of RNA (average 5.6 μg total RNA) was obtained by pooling hemolymph from only 12 individual *G. mellonella*. Operating time was also reduced by sending total dry-stabilized RNA to a third-party sequence centre that completed further quality control, dual rRNA depletion, library preparation and sequencing. This method provides an attractive alternative to previous *in vivo* RNA extraction methods, especially for researchers with little experience or expertise in molecular techniques required for working with highly-degradable RNA. At present, overly complex *in vivo* RNA extraction methods and enrichment of bacterial mRNA from mix total RNA samples present some of the major barriers for labs that cannot afford sophisticated equipment or do not routinely work with mixed RNA extraction. Furthermore, we envision wide applicability of this technique, especially since enriched host RNA could also be extracted from pelleted hemocytes and concurrently sequenced with pathogen

enriched RNA from the hemolymph, providing insights into the cross-talk between bacterial pathogen and host immune cells during infection (i.e., dual-seq approach). Dual-seq has already been widely applied in the study of human pathogens (Damron et al., 2016; Nuss et al., 2017; Thänert et al., 2017; Westermann et al., 2016), and perhaps dual-seq of *G. mellonella* hemocytes and pathogen-enriched hemolymph can provide similar insights for entomopathogenic bacteria in the future.

Overall, total RNA extracted from hemolymph was found to be of high quality and contained the anticipated ratios of host and bacterial rRNA depending on cell density. As expected, the lag phase *in vivo* RNA samples contained the highest amounts of contaminating host rRNA, even following depletion with an average of nearly a quarter of the reads aligning to *G. mellonella* rRNA sequence. The Ribo-zero human/mouse/rat rRNA depletion kit has been reported to remove > 98 % of insect (*D. ananassae*) rRNA from total RNA samples (Kumar et al., 2012). However, similar to the MH96 *in vivo* lag libraries, Smith et al. also found that insect RNA could only be partially depleted using the Ribo-Zero human/mouse/rat rRNA depletion kit, with 50 % of 28S rRNA in *D. melanogaster* remaining following depletion (2011). It is likely that the lag *in vivo* samples contained a much higher ratio of host rRNA compared to pathogen, such that the dual rRNA depletion kit was insufficient in depleting all contaminating host rRNA. It was found however, that in higher density *in vivo* samples (10^8 – 10^9 CFU/g) only trace amounts (0.3 – 1.1 %) of contaminating host rRNA remained, indicating sufficient enrichment of bacterial RNA in these samples. To potentially increase effectiveness of dual rRNA depletion in future studies (especially for low cell densities, or enrichment of pathogen associated with hemocytes, for example) library preparations may be modified by varying the ratio of host and pathogen rRNA-specific probes to quantitatively or qualitatively matching ratios of host/pathogen rRNA obtained from the Bioanalyzer or agarose gel, respectively.

Fortunately, by increasing the sequencing depth for the lag phase (~ 10^7 CFU/g) *in vivo* libraries it was possible to achieve an average alignment of 9 % of non-rRNA reads to the MH96 reference genome. Rarefaction analysis confirmed that these libraries contained enough aligned transcripts (~ 2 million) to sufficiently capture the entire dynamic range of expression by MH96, including lowly expressed transcripts. Recent

investigation of bacterial transcriptome sequencing depth have also reported that DE can be detected from numerous genes even when sequencing depth is reduced to 2-3 million non-rRNA fragments per sample, as long as biological replicates were available (Haas et al., 2012). Here the authors identified an actual risk of detecting spurious transcripts of little biological relevance, if sequencing depth were to increase beyond a certain point and suggested increasing the breadth of growth conditions or number of biological replicates, in favour of increasing sequencing depth. Further to this, McClure et al (2013) also found no significant degradation in the estimate of transcript abundances when using only 2 million reads per library.

The alignment rates for *in vitro* libraries were high (92 – 96 %), as expected for bacteria grown in pure culture. The alignment rates for exponential and stationary *in vivo* libraries were also high (73 – 84 %), which could be considered very good for RNA obtained from mixed eukaryotic/prokaryotic source; however, a sizeable fraction of contaminating MH96 rRNA (22.1 – 27.0 % of total reads) was also identified in these libraries. The microbial rRNA depletion seemed to be more effective for *in vitro* samples, which had on average less than 10 % of total reads attributed to MH96 rRNA among exponential *in vitro* libraries (25 °C). Subsequently, all rRNA sequences were removed from the count data prior to statistical analysis to avoid any confounding effects due to different efficiencies of rRNA depletion among different treatments. While, adequate depletion of host/bacterial rRNA remains a significant issue for the study of bacteria from within non-mammalian hosts, steadily decreasing high-throughput sequencing costs over time will eventually reduce the requirement to deplete rRNA, in favour of increasing sequencing depth enough to sufficiently sample signal from lower abundance bacterial mRNAs present in mixed RNA samples.

Not unexpectedly, most transcripts aligned to predicted protein coding regions of the genome (72 – 85 %). Interestingly, a markedly higher proportion of reads aligned to non-annotated regions of the genome (i.e., intergenic regions) from the *in vivo* libraries (24 – 26 %) compared to *in vitro* (13 – 15 %). It has been estimated that 10 – 15 % of bacterial genomes encode RNA (i.e., rRNA, tRNA and ncRNAs) (Westhof, 2010), so it would be quite unexpected to find 24 – 26 % of alignments from *in vivo* libraries representing ncRNA alone. To this end, the raw count data for all 207 predicted

ncRNAs was examined; however, none of the predicted ncRNAs were found to be exclusive to the *in vivo* libraries. A better explanation, is perhaps that greater sequencing depths (*in vivo* libraries were sequenced at a greater depth than *in vitro* libraries) result in greater capture of spurious transcriptional events or cDNA library contaminants as described by Haas and Défago (2005). Furthermore, since the *in vivo* samples came from a mixed source with host and pathogen total RNA, there is a greater risk of erroneous alignments of host-derived reads to non-annotated regions of the reference genome, despite pre-removal of host-derived sequences prior to alignment. Also, greater gene expression variability was identified among *in vivo* libraries compared to *in vitro* libraries as well; however this phenomenon has been previously described from the transcriptome of *E. coli* OP50 within the gut of *Ca. elegans* compared to culture plate (Chan et al., 2019). It is not surprising that *in vivo* transcriptome of a pathogenic bacteria would be inherently noisier than transcriptome from a laboratory culture, including low-level spurious alignments and increased expressional noise; however, these issues were not preventative for robust data analysis, which primarily focused on transcription of protein-coding genes and included adequate biological replication for each *in vivo* treatment.

4.4.3 *In vivo* putative VFs characterization

4.4.3.1 Putative toxins genes, *yenT*, *aidA*, *cdtAB*, *yenC3* and *vip2* respond to *in vivo* conditions at 25 ℃

Functional enrichment of putative VFs identified among *in vivo* clusters captured several genes for putative toxins that most likely interact with host hemocytes resulting in modulation of host immune responses to promote infection. Specifically, the gene for heat stable enterotoxin *yenT* (PL78_03785) was found to be highly and constitutively expressed *in vivo* compared to *in vitro* across all cell densities. YenT shares 23/51 (45.1 %) amino acid sequence similarity with Yersinia heat-stable enterotoxin type B (YstB) from *Y. enterocolitica* str. 845. Most biotype 1A strains of *Y. enterocolitica* encode *ystB* and the production of this enterotoxin is thought to be the major factor contributing to diarrhea in symptomatic cases caused by this otherwise non-pathogenic biotype (Grant et al., 1998). In the MH96 genome, *yenT* is located on

pathogenicity island, PAI_{YE96} that also encodes Yen-TC, but unlike most of the genes for the Yen-TC components, the expression of *yenT* was host-dependent, indicating that this toxin could play an important role in pathogenesis specifically during infection of the hemocoel by having targeted activity against hemocytes. The PAI_{YE96}-encoded putative virulence regulator, *yen1*, was found to share the same expression profile as *yenT,* establishing a potential *in vivo* regulatory linkage between *yen1* and *yenT*, which could be an interesting area for future research, especially since these data suggest YenT may have specificity against insects.

Aside from *yenT*, the remaining putative toxin genes identified from the *in vivo* VF functional enrichment analysis included *yenC3, cdtA, cdtB,* and *vip2* and all were identified from early *in vivo* clusters. Intrahemocoelic infection of *G. mellonella* with just a single MH96 cell was estimated to result in 100 % host mortality at both 25 °C (LD_{50} of ~ 0.5 cells, see Chapter 5, section 5.3.1.2, Table 5.2). So, for MH96 to achieve such an incredibly high rate of successful infection, the pathogen must be equipped with a plethora of strategies that act in synergy to overwhelm the *G. mellonella* immune response by multifaceted attack. Therefore, any toxin genes identified from early *in vivo* clusters, such as *yenC3*, *cdtA, cdtB* and *vip2*, represent (at least) some of the putative VFs that contribute to a highly effective overkill strategy.

The main virulence determinant required for successful infection of MH96 via oral challenge against insects was shown to be Yen-TC (Hurst et al., 2011a; Marshall et al., 2012). The YenC component of the holotoxin has been shown to contain the cytotoxin, which is auto-proteolytically cleaved from the C-terminus of the unfolded protein (Busby et al., 2013b; Landsberg et al., 2011; Meusch et al., 2014) and translocated into the host cell through a transmembrane pore formed by the B and N-terminal region of the C subunit (Gatsogiannis et al., 2013, 2016; Meusch et al., 2014) and A-subunit (Piper et al., 2019). MH96 encodes three homologous *yenC* component genes, *yenC1, yenC2* and the recently designated *yenC3* (previously known as *rhsA*) (Busby et al., 2019). *YenC1* and *yenC2*, are co-located with the main Yen-TC componentry within PAI_{YE96}, while *yenC3* is located within Rhs-associated region 2 and it has been suggested that YenC3 may likely replace YenC1 or YenC2 as the C component of the Yen-TC holotoxin (Hurst et al., 2016).

While almost all the Yen-TC component genes were found to be constitutively and highly expressed at 25 °C under both *in vivo* and *in vitro* growth conditions, on the contrary *yenC3* was identified as a putative VF with higher *in vivo* expression during early infection, which is an interesting finding, suggesting an important possible role for the third orphan *yenC3* homolog against hemocytes. The toxin containing C-terminus of each of the three *yenC* variants encoded by MH96 are expected to produce unique effectors which supports a multi-functional role for Yen-TC as a Rhs-containing cytotoxin translocator, contributing to virulence through infection by intoxication of host cells using divergent effector molecules.

Previous work has shown that insecticidal TCs effect the cytoskeleton of host cells. Notably, insect hemocytes were exposed to Yen-TC were observed to become irregular in shaped and ruffled in appearance (Hurst et al., 2015; Marshall et al., 2012). Also, exposure of HeLa cells and *G. mellonella* hemocytes to TC proteins of *P. luminescens* resulted in cytoskeleton modification and inhibition of phagocytosis, respectively (Lang et al., 2010). Furthermore, the TccC3 and TccC5 components of *P. luminescens* were identified as adenosine diphosphate (ADP)-ribosyltransferases, which can act on host-cell microtubules and Rho-GTPases (Lang et al., 2010). So, based on the host-dependent gene expression profile of *yenC3* and insecticidal TC abilities to effect cytoskeleton of insect cells in MH96 and other systems, it is tempting to speculate the YenC3 protein is processed in a similar manner as YenC1 or YenC2, and that the cytotoxic C-terminus is translocated specifically into hemocyte cells during early invasion of the hemocoel.

In addition to *yenC3*, the only other Yen-TC component found to have enhanced *in vivo* expression during insect infection at 25 °C was *chi1*, which, similar to *yenC3* was also identified as highly expressed *in vivo* during early infection. Chi1 has already been characterized as an important VF factor of MH96, especially as a critical component of Yen-TC structure (Busby et al., 2012; Landsberg et al., 2011) and now more recent work further supports a multifunctional role for Chi1 in host-ligand binding and pore formation by Yen-TC (Piper et al., 2019). Perhaps enhanced expression of *chi1* during insect infection could be interpreted as possible increased activities of Yen-TC during

early hemocoel infection, which may otherwise be limited by Chi1 component availability (one of the outer-most components of the holotoxin).

The genes for both components of the cytolethal distending toxin (CDT), *cdtA* and *cdtB* were also found to be more highly expressed *in vivo* during the lowest cell density but each gene was identified from separate early type clusters. CDTs are considered genotoxins that can interfere with the eukaryotic cell cycle, through the DNA-damaging activity of the CytB subunit that has deoxyribonuclease activity (Heywood et al., 2005; Lara-Tejero and Galán, 2000). In *Haemophilus ducreyi*, the CDT is a tripartite genotoxin, consisting of CdtA, CdtB and CdtC and based on the crystal structure CdtA and CdtC are lectin-type adhesins (like glycan-binding plant toxin ricin), while CdtB adopts a structure consistent with DNAse I family human endonucleases (Hu et al., 2006; Nešić et al., 2004).

Most CDTs have been described from human pathogenic bacteria (*E. coli, Shigella* (*Sh.*) *dysenteriae*, *Helicobacter* sp., *Campylobacter* sp., and *Sa. enterica*) and consist of a tripartite holotoxin (Guerra et al., 2011). With respect to insect-associated bacteria, the genome of protective endosymbiont *Hamiltonella defensa* (harbored by aphids) was found to encode a *cytB*-homolog within a toxin-encoding bacteriophage, that carries a gene for a shiga-like toxin (Degnan and Moran, 2008b, 2008a; Degnan et al., 2009) and this bacteriophage was later shown to confer the genes responsible for *H. defensa's* ability to protect its aphid host from attack of parasitoid wasps (Oliver et al., 2009). A CDT was also identified as potentially involved in a *Ps. entomophila-D. melanogaster* pathogenic interaction based on genomic-wide comparison study (Vodovar et al., 2006), but other than these two examples, the role of CDTs in virulence specific to entomopathogenic bacteria is fairly limited.

Like *cdtA, cdtB* and *yenC3* described above, the expression of another putative toxin encoding gene, *vip2* was identified among early type *in vivo* clusters. The predicted protein sequence of Vip2 contains a VIP conserved protein domain within its C-terminus and is currently annotated as a NAD:arginine ADP-ribosyltransferase toxin. Both *B. cereus* and *Bt* secrete VIP toxins during vegetative growth and these and similar binary toxins consisting of Vip1 and Vip2 components (Crickmore et al., 2014; de Maagd et al., 2003). Vip2 contains the cytotoxic A-domain and shares sequence

homology with other ADP-ribosyltransferase containing toxins including the enzymatic component of the iota-toxin from *Clostridium perfringens* (Tsuge et al., 2003, 2008; Tsurumura et al., 2013), CDT toxin from *C. difficile* (Gonçalves et al., 2004) or VIP toxin from *B. cereus* (Han et al., 1999), for example. All of the binary-family ADP-ribosyltransferase containing toxins are known to target globular (G)-actin of host cells, which may prevent phagocytosis, intracellular trafficking or induce cell death (Barth et al., 2004). Since a VIP binding component was not found within the genome of MH96 (Hurst et al., 2016) the translocation mechanism of Vip2 into target host cells is currently unknown.

4.4.3.2 Secretion systems

4.4.3.2.1 T2SS genes responded to *in vivo* conditions and possibly secrete chitinolytic exoenzymes

In addition to toxins, secretion systems also represent an important group of virulence factors that can contribute to defeating the immune response of the host by targeting hemocytes for effector delivery or secretion of exotoxins or degrading enzymes into extracellular milieu. Among the genes for secretion systems encoded by MH96, several related to T2SS and T3SSYE2 were found to be highly expressed *in vivo* at 25 °C. A large number of genes for the T2SS general secretion pathway (gsp) components were found to be highly expressed *in vivo* across all cell densities (e.g., *gspK*, *gspF*, *gspD*, *gspE* and putative lipoprotein *gspS*), while genes for other T2SS components (*gspJ*, *gspI*, *gspG* and a prepilin-like gene) showed a greater response to *in vivo* growth conditions only during late infection (10^9 CFU/g). *GspC* was not identified as a putative VF but demonstrated higher expression *in vivo* as well. Loci corresponding to possible remaining components (*gspL* and *gspM*) were not previously annotated in the genome (Hurst et al., 2016). The GspM component is supposed to assist in directing GspL toward appropriate sites in the cell well (Michel et al., 1998), but perhaps this process no longer required for active secretion by the T2SS in MH96.

Other T2SS have been identified from the genomes of entomopathogens such as *P. luminescens*, *X. bovienii* and *Ps. entomophila* (McQuade and Stock, 2018), but their functional role in secretion during pathogenesis has not been a focus of much research

to date. Given the higher expression of T2SS genes *in vivo* it is strongly suspected the T2SS contributes to virulence in MH96 by secreting folded exoproteins across the outer membrane. Recently, the Yts2 T2SS of a *Y. enterocolitica* biotype 1B strain was found to be significantly upregulated during intracellular conditions within murine macrophages; however another T2SS called Yts1 did not respond to either intra or extracellular growth conditions (Bent et al., 2015). Interestingly, secretion of putative chitin- and oligosaccharide-binding effectors (ChiY, EngY and YE3650) by Yst1 in *Y. enterocolitica* was shown to require low temperatures (17 °C) and high concentrations of $MgCl_2$ (Shutinoski et al., 2010). MH96 was found to highly express genes for two different putative lytic monooxygenases (LCMOs), otherwise knowns as chitin-binding proteins (CBPs), at 25 °C, one of which was identified among an *in vivo* early type cluster. The putative protein sequences of both of these LCMOs contain a signal sequence, implicating a role for T2SS in secretion of at least two important chitinolytic enzymes produced by MH96 during *in vivo* conditions, which would be similar to the T2SS Yst1 mode of action in *Y. enterocolitica* 1B strain (Shutinoski et al., 2010). Furthermore, the chitinase (PL78_11910) and the LCMO/CBP CbpA (PL78_05310) have been identified from the culture supernatant of MH96 cultures grown at 25 °C, so are known to be secreted (supplementary Figure S18).

4.4.3.2.2 T3SSYE2 and multiple effectors responds to intrahemocoelic conditions and likely contribute to MH96 virulence against hemocytes.

In addition to the T2SS, the genes for several different components and effectors associated with T3SSYE2 were also found to be among *in vivo* clusters. In total, eleven genes associated with the T3SSYE2 region of the genome were found to have a stronger response to *in vivo* growth conditions compared to *in vitro*, including genes for seven structural components: SoaS (PL78_14595), SpaP (PL78_14580), SpoA (PL78_14575), SpaL (PL78_14560), InvA (PL78_14550), InvE (PL78_14545) and prgH (PL78_14530) and two hypothetical proteins (PL78_14565 and PL78_145670). Genes for the T3SSYE2 putative effectors SicC (PL78_14610) and SicD (PL78_14615) both found to higher expression *in* vivo at 25 °C as well, with the former being more highly expressed across all cell densities and the latter more highly expressed during early infection only. Three other putative orphan T3SS effectors, the T3SSYE1-associated IpaB/EvcA (PL78_18085)

and SopD (PL78_14295) and IpgD/SopB (PL78_07345) were also identified as highly expressed under *in vivo* conditions at 25 °C and may be also secreted by the T3SSYE2. A final T3SSYE2 gene, for putative YscW (T3SS lipoprotein chaperone, PL78_14485) was identified as highly expressed *in vivo* but was clustered with the usher-chaperone fimbrial cluster also identified as very highly expressed *in vivo*, especially during early infection and these findings suggest a regulatory linkage between the a T3SSYE2 chaperone lipoprotein gene and the usher chaperone fimbriae gene cluster. Overall these results implicate T3SSYE2 as a critical secretion system of MH96 that is used during infection to translocate a variety of effectors into host cells. In alignment with this, the gene arrangement of T3SSYE2 shares the same gene order with the *Ysa* system of *Y. enterocolitica* biotype 1B (Hurst et al., 2016), which was demonstrated to facilitate intracellular replication in *D. melanogaster* S2 cells (Walker et al., 2013).

Some of the T3SS effectors identified from *in vivo* clusters share significant sequence similarities to other T3SS effectors that are known to be involved during cellular invasion by *Salmonella* (*Sa.*) *enterica* serovar Typhimurium and *Sh. flexneri*. The translated protein sequence of putative IpgD/SopB (PL78_07345) is significantly similar (E-value 1e-83) to SopB/SigD of *Sa. enterica* serovar Typhimurium (EAZ9709492.1) sharing 34 % identity and 73 % coverage. Likewise, the predicted protein coding sequence for putative effector SopD (PL78_14295) is significantly similar (E-value 2e-12) to SopD of *Sa. typhimurium* (TQS58761.1), sharing 30 % identity and 34 % coverage. In *Sa. enterica* serovar Typhimurium, SopB and SopD were shown to work cooperatively during invasion of non-phagocytic cells (Bakowski et al., 2007) and SopB can be translocated by the SPI-1 T3SS, while SopD can be translocated by both SPI-1 and SPI-2 T3SS (Boonyom et al., 2010). Since the genes for IpgD/SopB and SopD were highly expressed by MH96 under *in vivo* conditions and share similarities with known VFs from other pathogenic systems, it is hypothesized that these effectors may retain a similar cooperative mode of action as found in *Sa. enterica* serovar Typhimurium against *G. mellonella* hemocytes during infection.

Another host-specific putative effector (PL78_18085) that is predicted to have an IpaB/EvcA family domain was found to have higher *in vivo* expression as well. Further investigation revealed that the predicted secondary structure of this effector shares

significant amino acid similarities (100 % confidence and 58 % coverage) to the guanine nucleotide exchange factor (GEF)-mimicking IpgB2 from *Sh. flexneria* (c3lw8G), which is known to modify host actin filaments by interaction with Rho-GTPases (Klink, Björn et al., 2010; Kuehl et al., 2014).Taken together, MH96 is thought to be able to effectively modulate the immune response of the host by translocation of multiple T3SSYE2-secreted effectors because genes for these effectors and T3SSYE2 components were highly expressed *in vivo* at 25 °C.

The *in vivo* expression of the T3SSYE1 component genes at 25 °C were less clear compared to T3SSYE2 as genes for the T3SSYE1 structural components were more highly expressed at 10^8 and 10^9 CFU/ml cell densities *in vitro* compared to *in vivo*, suggesting that T3SSYE1 may be functional under different conditions that T3SSYE2. Among the T3SSYE1 operon, only the genes for putative effector IpaB/evcA and a T3SS chaperone (PL78_18170) genes were found to have higher expression *in vivo* at 10^7 CFU/g cell density and perhaps the chaperone is a required adaptor for IpaB/evcA secretion by T3SSYE2. A gene for putative orphan T3SS effector, SopD (PL78_14295) was also found to be more highly expressed at low cell density *in vivo*. It is not unexpected that the genes for multiple T3SS effectors can located separately within a genome but share coordinated regulation based on environmental variables. A similar scenario was found in *Y. enterocolitica*, where under *Ysa*-inducing conditions (e.g., 26 °C and high $NaCl_2$ concentration), 15 different effectors are secreted, including 3 Yop effectors that are located on the pYV plasmid and are usually secreted by the much better characterized Ysc T3SS) (Young and Young, 2002). Actually, only a single effector gene (*yspA*) is located within the *Ysa* gene cluster, and all the other genes for *Ysa*-secreted effectors are scattered throughout the genome (Matsumoto and Young, 2006). In this system the genes for the *Ysa* apparatus as well as the *trans*-located effectors are all co-ordinately regulated under the same environmental conditions (Venecia and Young, 2005; Walker and Miller, 2009).

In other pathogenic, T3SS effectors are primarily used during association with host cells and the invasion process or as cytotoxic agents, but not much has been studied among entomopathogenic bacteria. *P. luminescens* produces at least two T3SS effectors, the first is LopT (shares similarity to YopT from *Y. pestis*) that may contribute to

evasion of phagocytosis by hemocytes of the cutworm *Spodoptera* (*Sp.*) *littoralis* and the locust *Locusta migratoria*. (Brugirard-Ricaud et al., 2005) and a cycle-inhibiting factor (Cif) that shares similarities with Cif-type effectors pathogenic *E. coli* (Escoubas et al., 2009). While MH96 does encode a gene for a putative LopT-like effector (PL78_18760), it was not statistically identified among any *in vivo* clusters but did tend to be more highly expressed *in vivo* during early infection and is likely co-regulated with other T3SS effectors used during early infection and secreted by the T3SSYE2 as well.

Through this *in vivo* transcriptome investigation several putative T3SS effectors have been identified as likely important during invasion of insect hemocoel, contributing to the high virulence capabilities of MH96 against *G. mellonella*. A key next step to validating if the effectors identified in this work is visualization of effector translocation into host cells in a T3SS-depedent manner, which has been achieved in other systems using genetically encoded reporter systems like TEM-1 beta-lactamase (Charpentier and Oswald, 2004; Mills et al., 2008) or light-oxygen-voltage systems (Jayde et al., 2016), for example. Further, a mutant deficient in T3SSYE2 apparatus should be assessed for phenotype using either *G. mellonella* intrahemocoelic infection model or *Drosophila melanogaster* S2 macrophage-like cell line to further asses the role of this secretion systems role in host-pathogen interactions.

4.4.3.2.3 T6SS effectors respond to intrahemocoelic conditions and may have hemocoelic activities.

Generally, T6SS are regarded as one of the most widely distributed secretion systems among bacteria and are important for pathogenesis (Burtnick et al., 2011; Mougous et al., 2006; Mulder et al., 2012; Wang et al., 2018) and competition (Haapalainen et al., 2012; Hood et al., 2010; Koskiniemi et al., 2013; MacIntyre et al., 2010). The structure of T6SS resembles a bacteriophage-like contractile particle, with a needle-like complex enabling bacteria the ability to puncture target cells as well as translocate toxic effectors (Cianfanelli et al., 2016; Ho et al., 2014). The needle apparatus is formed from stacked hexametric ring components called hemolysin co-regulated protein (Hcp) topped by valine glycine repeat protein G (VgrG) spike component (Russell et al., 2014). Hcp and VgrG are considered to be both structural components and effectors of the T6SS

(Cianfanelli et al., 2016; Hood et al., 2010; Ma et al., 2017; Pukatzki et al., 2009) and the VgrG spike can also translocate C-terminal extension toxins, such as Rhs-containing proteins with N-terminal PAAR-repeat regions for example (Koskiniemi et al., 2013; Shneider et al., 2013). Specifically, Hcp has been attributed to pathogenesis of both *Ps. aeruginosa* in cystic fibrosis patients (Mougous et al., 2006) and *Burkholderia pseudomallei* in intracellular and Syrian hamster animal models (Burtnick et al., 2011).

During growth of MH96, the expression of genes for the T6SS structural components increased with cell density in both *in vivo* and *in vitro* treatments at 25 °C; however, some genes for putative effectors found within Rhs3-associated region (e.g., PL78_00900/Hcp and PL78_01040/Rhs), a putative T6SS orphan effector (PL78_02751/Hcp) and two duplicate putative effectors genes (PL78_03598/Hcp and PL78_03602/Hcp) were identified as putative VFs from *in vivo* clusters. Furthermore, the gene for putative Spt4 effector (PL78_00995) that is also located within Rhs-associated region 3 had constitutively higher expression *in vivo* throughout infection; however, the gene for the VgrG spike component was not found to have higher expression during *in vivo* growth compared to *in vitro.* The Spt4 effector shares sequence similarity with the eukaryotic transcription elongation factor and is considered to be an effector that targets host immune cell cycle, but the exact function remains to be elucidated. By using an *in vivo* transcriptomics approach additional putative VFs associated with T6SS could be identified as highly expressed *in vivo,* which implicates the T6SS as a potentially important VF during insect infection by MH96.

It is not uncommon for bacteria to encode genes for more one Hcp loci (Barret et al., 2011; Peng et al., 2016; Sha et al., 2013; Wang et al., 2018). Further investigations of different Hcp and VgrG paralogs produced by the emerging human pathogen, *Aeromonas hydrophila* SSU, indicated that for (at least) this system, the Hcp and VgrG proteins produced by genes located with the main T6SS cluster were more important to forming the T6SS structural components, while other Hcp and VgrG proteins produced by genes located outside of the T6SS cluster contributed more as effectors (Sha et al., 2013). In a different strain of *A. hydrophila* NH-35 (encodes one additional *hcp* locus compared to SSU), Wang et al. determined that each of the three Hcp paralogs took on

diverse roles related to environmental adaptation or virulence (Wang et al., 2018). MH96 encodes up to five putative *hcp* loci, four of which were found to be more highly expressed at various cell densities *in vivo* and based on these findings, it is hypothesized that each of the different *hcp* paralogs may play different roles involved in pathogenesis, environmental adaptation or competition. To further evaluate a potential role for the T6SS in pathogenesis, a MH96 mutant deficient for the T6SS sheath component gene, *vipB* (not to be confused with VIP binary toxin) , was assessed using the *G. mellonella* intrahemocoelic model, but was not found to have reduced virulence compared to wild-type (Chapter 5, section 5.3.1.2, Figure 5.8). Additional targeted mutagenesis of each putative *hcp* effector gene would help to further characterize the importance of the T6SS in MH96.

4.4.3.2.4 T5SS and holin-endolysin secretion systems also putative VFs of MH96 that responded to intrahemocoelic conditions.

Another gene for a major putative VF identified as highly expressed *in vivo* was for a T5SS autotransporter adhesin with a predicted protein sequence that contains both a pertactin-like passenger and outer-membrane autotransporter barrel domains. The predicted protein sequence for PL78_10240 is highly similar (E-value: 4e-91) to known VF, an AidA-I adhesin-like protein, from enteropathogenic of *E. coli* (EPEC) strains (WP_001386603), sharing 35 % amino acid similarity and 65 % sequence coverage. AidA is a autotransporter-type glycoprotein adhesin, that is present on the outer membrane of *E coli* cells (Benz and Schmidt, 1992; Suhr et al., 1996) and has shown to be versatile VF involved in cell aggregation and biofilm formation (Sherlock et al., 2004) as well as host cell adhesion and invasion (Benz and Schmidt, 1992, 2001). Based on the host-dependent response of AidA-like adhesion, PL78_10240, it is assumed that the adhesin would be associated with the outer membrane of MH96 during infection and contribute to pathogenesis by adhesion to hemocytes or possibly contribute to other important virulence processes, such as biofilm formation or aggregation of cells in a similar manner to AidA in enterohemorrhagic *E. coli* (EPEC).

Another secretion system encoded by MH96 that is potentially important for the secretion of important toxins or chitinolytic machinery is the holin-endolysin secretion

system, which is a phage-related lysis cassette located on an excisable holin-citrate island, HCUI$_{YE96}$ (Chapter 6, section 6.3.1). Indeed, many genes located on HCUI$_{YE96}$, including the putative holin-endolysin cassette (primary focus of another concurrent PhD project), as well as AraC-like regulator and several citrate fermentation related genes. The genes for holin-endolysin and many of the citrate fermentation cluster were not categorized as putative VFs by VFDB-annotation method. Despite the lack of virulence annotation, the holin-endolysin secretion system is thought to be important for toxin and/or chitinase/LCMO secretion by MH96 during insect infection because analogous holin-endolysin secretion systems in *Y. enterocolitica* W22703 and *S. marcescens* were shown to be required for insecticidal TC (Springer et al., 2018a) and chitinase/LCMO secretion (Hamilton et al., 2014), respectively.

4.4.3.3 Putative VFs related to stress response, defense and HGT also identified from MH96 *in vivo* transcriptome

Two other major groups of putative VFs were represented among *in vivo* clusters, which included defence and HGT and were considered part of an overall stress response by MH96 to the host environment. During infection pathogens must endure a variety of different stresses such as oxidative, pH, nutritional or temperature-induced stress and a number of genes related to general stress response that were upregulated by MH96 under *in vivo* conditions including a number of efflux pumps, acid shock and cold-shock proteins. Similarly, previous *in vivo* transcriptome of the plant pathogen *D. dadantii* also found a high number of genes related to stress response, including response to AMPs, efflux systems/transporters, acid shock, anaerobiosis, SoxR and Fe-S cluster within the body of aphids (Costechareyre et al., 2013). Efflux pumps especially seem to be important for detoxification of the cell during infection of insects by both MH96 and *D. dadantii* based on the similar findings between the two *in vivo* transcriptomes studies. Furthermore, a gene for a putative fumarate reductase (PL78_08240) (involved in anaerobic growth), was identified as highly expressed by MH96 during early infection implying the hemolymph is an anaerobic environment as well.

General stress response to *in vivo* growth conditions was also previously identified in *Y. pseudotuberculosis*, with genes related to anaerobiosis, defence against oxidative and

acid stress found to be more highly expressed during persistent infection (Avican et al., 2015). Recently the RNA-binding CSPs have been found to globally regulate mRNA secondary structures and rate of translation (Zhang et al., 2018b) and have been implicated as critical for regulating virulence in *Sa. enterica* serovar Typhimurium (Michaux et al., 2017), *Listeria monocytogenes* (Eshwar et al., 2017) and *Brucella melitensis* (Wang et al., 2016). Furthermore, among *Yersinia* spp., the expression of two CSP genes, *cspH* and *cspE*, responded to temperature shift from 26 to 37 °C in *Y. pestis* (Motin et al., 2004), while *cspE* and *cspI* were upregulated during macrophage infection (Fukuto et al., 2010) based on DNA microarray analysis. Leskinen et al. (2016) reported upregulation of four CSP genes in *Y. enterocolitica* during late infection of phage φR1-37. Recently, the main CSP in *Y. enterocolitica* (biotype O:3), *cspA,* was upregulated in a mutant deficient for *ybeY* (Leskinen et al., 2015), which is an RNAse involved in ribosome maturation and late-state 70S ribosome quality control (Jacob et al., 2013). Comparison of steady-state protein abundance and changes in mRNA expression of a cold shock-like protein gene, *cspI*, by *Y. pestis* KIM5 replicateing within macrophages found ~7-fold higher transcription of *cspI* between 1.5 to 8 h post-infection of macrophages (Keasey et al., 2016). Based on these emerging reports, and the prevalence of several CSP genes in the MH96 *in vivo* clusters, CSPs were identified as putative virulence regulators in MH96 and were the focus of more targeted investigations; A *cspA123* deficient strain was phenotyped by bioassay, protein visualization by SDS-PAGE and *in vivo* RNA-seq, reported in Chapter 5 section 5.3.1.2 - Table 5.2, Chapter 5 section 5.3.1.1 - Figure 5.4 and 5.6 and Chapter 6 section 6.3.1 - Tables 6.8 and Table 6.9, respectively.

Many genes related to mobilization of DNA from the genome and thought to be involved in HGT (i.e., excisionase, integrase, recombinases, transposase and phage) were identified as more highly expressed during *in vivo* conditions, but the majority of these HGT genes were not captured by VFDB annotations. While increased HGT during *in vivo* growth was not the primary focus of this study, these findings are very ineteresting, especially if MH96 has evolved host-dependent regulatory responses to increase the rate of HGT as a means of generating increased genomic flexibility required for survival as a pathogen (Dobrindt et al., 2002, 2004; Francino, 2012; Ochman and Moran, 2001). Indeed, conjugative transfer of mobile genetic elements has

been shown to occur *in planta*, among kiwifruit pathogen *Ps. syringae* pv. *actinidae* (Colombi et al., 2017) and potato pathogen *Pectobacterium atrosepticum* SCRI1043 (Vanga et al., 2015) and excision of the unstable pathogenicity island (ROD21) in *Sa. enterica* serovar Enteritidis was found to be induced during infection of murine phagocytic cells (Quiroz et al., 2011). While co-evolution of host and parasite has been shown to promote increased genome diversity, most studies have focused only on the effects of coevolution on host genome evolution. In one experiment, the genome of a *Ps. fluorescens* phage was shown to evolve with greater diversity when selected under co-adapted host, compared to genetically static host (Paterson et al., 2010). In another study, coevolution of *Bt* and *Ca. elegans* resulted in increased genetic diversity of *Bt* chromosome, specifically by acquisition of plasmid-born toxin genes (Schulte et al., 2013). The increased expression of HGT-related genes by MH96 during *in vivo* conditions could support a hypothesis of HGT driving increase genomic diversity during coevolutionary interaction with a host, but this remains an area for future research.

Finally, a surprising result was the high proportion of tRNA genes sharing increased expression during stationary stage under *in vivo* conditions, an expression profile that was shared with several bacteriophage- and HGT-related genes. While there are many types of mobile genetic elements (MGEs) known from prokaryotic genomes, including transposons, insertion sequences, integrons, conjugative plasmids, bacteriophages, integrated prophages, pathogenicity islands (PAIs)/genomic islands (GIs) and integrative and conjugative elements (ICEs) (Abebe-Akele et al., 2015; Frost et al., 2005; Oliveira Alvarenga et al., 2018; Wilkinson et al., 2009), some types of MGEs target tRNA loci for insertion into bacterial genomes, including PAIs/GIs, ICEs and prophages (Williams, 2002) which begs further investigations into potential linkages between lytic bacteriophage and PAIs/GIs/ICEs during high cell densities *in vivo* within the genome of MH96.

4.4.3.4 Usher chaperone fimbriae and other putative adhesions responded to intrahemocoelic conditions

In addition to the diverse array of putative toxins, T3SSYE2 secreted effectors, T6SS effectors and an autotransporter like AidA-like adhesin described above, several other

putative VFs were identified from *in vivo* clusters, especially within the category adhesion, notably an entire usher chaperone fimbrial gene cluster (PL78_12465 – PL78_12480) and genes for filamentous hemagglutinin N-terminal containing protein (PL78_11060, previously annotated as a hemolysin) and cognate two-partner secretion system (TPS) (PL78_11055). Usher chaperone fimbriae and filamentous hemagglutinins (also discussed below in section 4.4.4.2 below) are widely-known among pathogenic bacteria as important virulence actors involved in adhesion to host cells. The presence of putative VFs involved in adhesion among *in vivo* host clusters suggests that MH96 has evolved regulatory response mechanisms triggering production and secretion of an usher chaperon fimbriae as wells the filamentous hemagglutinin N-terminal containing and AidA-like adhesions to the surface of the cell where these adhesive structures can interact with hemocytes or even enhance intoxication by increased exposure to T3SSYE2, T6SS or other secreted toxins.

Due to the very unique and conserved *in vivo* expression pattern of the entire four-gene usher chaperone fimbrial gene cluster, this putative VF was targeted for mutagenesis; however, a reduction in virulence was not observed in the deficient mutant strain compared to the parental strain using the *G. mellonella* intrahemocoelic model at 25 °C (Chapter 5, section 5.3.1.2, Table 5.2), although not tested at 37 °C. Additional work to validate the usher chaperone fimbriae as a *bona fide* VF is required and may involve further studies of usher chaperone fimbriae attachment to host cells using *G. mellonella* hemocytes or *D. melanogaster* S2 macrophage-like cell lines. Some preliminary work using a transcriptional-GFP reporter under *in vivo* conditions was under taken; however, fluorescence was not observed.

4.4.3.5 Outer membrane proteins, lipopolysaccharide and flagella responded to *in vivo* conditions in MH96.

In addition to adhesions and secretion systems that are located on the outer surface of the cell, other important VFs include capsule, outer membrane proteins (OMPs) and lipo-polysaccharide (LPS), which are known to be involved in host-associated outer surface modifications, especially among pathogenic *Yersinia* spp. (Bengoechea et al., 2004; Eddy et al., 2014; Montminy et al., 2006; Rebeil et al., 2004; Suomalainen et al., 2010). Genes for two outer membrane proteins (PL78_12825, PL78_02925), an OMP-

assembly (PL78_11055), murein endopeptidase (PL78_04470) and LPS cluster (PL78_00680 – PL78_00755) were identified as putative VFs from *in vivo* clusters in MH96 and responded to the host environment constitutively through infection and these findings support an important role for LPS/OMPs during engagement with insect hemocytes during infection at 25 °C.

Another important extracellular structure, the flagella, was found to be more highly expressed during later infection stages implying motility is under density-dependent regulatory *in vivo* mechanisms in MH96. *In vivo* responses of flagella-related genes have been previously described from the transcriptome of the plant pathogen *D. dadantii*, during growth within aphids, where a 15- and 11-fold-change in expression occurred for the genes of two components of the flagellar motor MotA and MotB, respectively (Costechareyre et al., 2013). The increased transcription of flagella-associated genes during the highest cell density *in vivo* is an indicator that increased motility is important during later infection by MH96, potentially involved in deeper migration of bacteria into underlying tissue layers to access nutrients for bioconversion.

A complex interplay between the regulation of virulence and flagellar motility is widely recognized among entomopathogenic bacteria. For example, in *X. nematophila*, the flagellar master regulators, *flhD* and *fliZ* were both shown to be involved in virulence region (Givaudan and Lanois, 2000, 2017; Lanois et al., 2008). Whether the transcriptional regulators located among the MH96 flagellar operon have influence on the regulation of virulence in MH96 is an area of research that currently remains unexplored; however, significant temperature-dependent regulation of flagellar genes by MH96 was also reported and is discussed below (section4.3.8, Table 4.11). Furthermore, the loss of mobility at 37 °C in MH96 compared to 25° C is easily observable under wet mount on a compound light microscope and during infection of *G. mellonella* at 37 °C microaggregates of non-motile MH96 cells can be observed in the hemolymph (personal observation).

4.4.3.6 Iron acquisition system are upregulated in MH96 during late infection

Along with the significant *in vivo* transcriptional upregulation of genes for the flagella during later infection, genes for several putative iron acquisition systems were also identified from late *in vivo* clusters. Iron acquisition is widely known to be an important pathogenic strategy because hosts have evolved to sequester iron in response to infection. Genes related to iron-acquisition in MH96 that were upregulated during late infection of *G. mellonella* were primarily identified from two unique regions of the genome thought to be acquired by HGT (Hurst et al., 2016), especially genes related to iron-siderophore biosynthesis/transport (PL78_03905 – PL78_03955) and another iron acquisition island containing genes for enterobactin synthase, iron dicitrate/enterobactin transporters, chromophore lyase, enterochelin esterase, ferrichrysobactin receptor and transporters (PL78_09675 – PL78_09785) from unique regions 1 and 2, respectively.

Recent *in vivo* transcriptome investigations of *Y. enterocolitica* biovar 1B infecting murine macrophages, found that three different loci related to either zinc or iron uptake pathways were significantly downregulated in response to internalization by macrophages, suggesting that zinc and iron are scarcer in the *in vivo* extracellular environment compared to the intracellular environment within the macrophage. A similar finding was also identified by Fukoto et al. (2010). Considering these findings, the pattern of both the flagella and the iron acquisition related genes have increased *in vivo* expression during stationary phase of insect infection by MH96, likely indicates a metabolic transition into a high cell density scavenging mode requiring both motility and iron acquisition systems. During the initial invasion of the insect hemocoel, MH96 would acquire nutrients readily available in the hemolymph (Killiny, 2018) and made available by the activities of primary VFs that contribute to destruction and lysis of hemocytes to release the high nutrient intracellular content into the extracellular environment. Once these nutrient sources have all been bioconverted by MH96, at high cell density nutrient acquisition must involve migration deeper into host tissues and limited iron driving the increased expression of several iron scavenging strategies.

4.4.3.7 Two-component regulatory systems and other putative virulence regulators were upregulated during insect infection by MH96.

Host-specific nutrient availability, including iron have been shown to be important stimuli driving regulatory responses in entomopathogenic bacteria (Crawford et al., 2010; Fang et al., 2016; Tran et al., 2013). Among transcriptional regulators, TCRS, are probably one of the best studied regulatory sensing systems recognized among pathogenic bacteria (Capra and Laub, 2012; Zschiedrich et al., 2016), so it was not unexpected that a large proportion of the putative regulators with higher expression *in vivo* were identified as TCRS. Among entomopathogenic bacteria, the PhoP/PhoQ system in *P. luminescens* was shown to be involved in virulent against *Sp. littoralis* (Derzelle et al., 2004) and the analogous GacS/GacA TCRS in *Ps. entomophila* was also shown to be an important regulator of virulence (Vodovar et al., 2006). It is likely that the great diversity of TCRS that were found to be highly expressed during insect infection provide MH96 with extensive environmental sensing capabilities, which are important to adapting variable environmental conditions. Among genes for other putative regulators identified with higher *in vivo* expression, the AraC-like from $HCUI_{YE96}$ and IclR-like (PL78_06535), were both found to have significantly higher *in vivo* expression during early infection also represent potential virulence regulators due to their co-expression with Yen-TC and TC-associated factors, *in vivo* expression and high homology to virulence regulators from other pathogenic bacteria. Follow-up molecular phenotyping using *in vivo* RNA-seq of mutants deficient for regulators such as the AraC- or IclR-like regulators, would provide further insights into regulation of virulence.

4.4.3.8 Intrahemocoelic infection of *G. mellonella* by MH96 involves metabolic adaptation

While iron represents one nutrient that can be very important for growth and proliferation of entomopathogenic bacteria within the host environment, more generally metabolic adaptation during pathogenesis also becoming increasingly recognized as an important area of virulence research (Bücker et al., 2014; Heroven and Dersch, 2014; Rohmer et al., 2011). Indeed, functional enrichment of putative VFs related to metabolic adaption were broadly identified from nearly all *in vivo* clusters,

especially genes related to sugar, peptide, urea, nitrate, phosphate and phospholipid transporters. Recently, an untargeted metabolomic profiling of the hemolymph of *G. mellonella* using three methods of derivatization, including trimethylation (TMS), methyl chloroformate and boron trifluoride determined that *G. mellonella* hemolymph contains a rich source of essential nutrients consisting of wide range of metabolites (Killiny, 2018). The TMS derivative method identified high abundances of polyamine putrescine (56.67 % mM), amino acids proline (9.72 % mM), glycine (3.07 % mM) and serine (2.21 % mM), phosphoric acid (4.25 % mM) and citric acid (3.21 % mM).

With respect to the most abundant metabolite identified within the hemolymph, putrescine, the gene for a putative putrescine transporter, *potG* (PL78_02390) was identified as a putative VF with constitutively higher *in vivo* expression by MH96. Putrescine as also been identified in high abundance from the hemolymph of the corn earworm, *Heliothis zea* (Cheung et al., 1982) as well as non-diapausing European corn borer, *Ostrinia nubilalis* (Purać et al., 2015) and likely represents an important nutrient source available for entomopathogens that occupying the hemocoel. In addition, bacterial polyamines (such as putrescine and spermidine) are known to be important interkingdom signaling molecules that are important for survival of pathogenic bacteria within the host including and linked to acid resistance, iron scavenging and virulence (Espinel et al., 2016; Wortham et al., 2007). Also, a gene related to uptake of the most abundant proteinogenic amino acid, proline, a putative proline: sodium symporter, *putP* (PL78_17825) was identified as having higher expression *in vivo* during later infection but was not annotated as a putative VF. The acquisition of L-proline from insect hemolymph was shown to be an important regulating factor in *P. luminescens*, with linkages to metabolic shifts, bacterial proton motive force, virulence and antibiotic production (Crawford et al., 2010)

Furthermore, a cluster of genes related to citrate fermentation that are located on an excisable island (Chapter 7, section 7.3.1, Figure 7.1) were found to have higher *in vivo* expression as well, and one of these genes *citE* (PL78_17355) was classified as a putative VF. These results support the ability of MH96 to respond to the *in vivo* environment by increasing expression of genes related to uptake of both proline and putrescine, as well as citrate fermentation, all of which may contribute to acquisition of

some of the most abundant nutrient sources available from *G. mellonella* hemolymph (Killiny, 2018). A similar citrate fermentation operon was identified among the genomes of pathogenic strains of *Y. ruckeri* is thought to provide a metabolic trait associated with pathogenesis (Cascales et al., 2017). Further discussions specifically related to citrate fermentation and the holin-citrate unstable island are provided in Chapter 7 of this thesis and the same citrate fermentation gene cluster found only The MFC method used in the Killiny (2018) also identified compounds at low concentrations, including tryptophan, glutamic acid, threonine, aspartic acid, methionine and cysteine, which were not detected at all using the TMS method. In MH96, genes related to synthesis of tryptophan and cysteine (from serine) were identified as putative VFs from *in vivo* clusters, which is somewhat consistent with the low detection of tryptophan and cysteine from the hemolymph of *G. mellonella* by Killiny (2018).

Another important nutrient that must be acquired by MH96 during infection is carbon, which is available as sugars, especially the glucose-glucose disaccharide trehalose that was found to comprise nearly 80 % of the sugars found in *G. mellonella* hemolymph (Killiny, 2018). Trehalose is an important sugar for insects, an important form of carbon storage (Elbein et al., 2003) as well as a source of glucose energy (Reyes-DelaTorre et al., 2012) and cryo-protectant (Saeidi et al., 2013; Wang et al., 2010). During infection by MH96, many genes for sugar transporters where found broadly across all *in vivo* cluster types and a gene for putative *treC* (trehalose-6-phosphate, PL78_10570) was identified from a late *in vivo* cluster but not annotated as a putative VF. Other potential VFs that could contribute to trehalose uptake include sugar transporter-related loci PL78_11005, PL78_07145 and PL78_18960, which were all identified from either early or late *in vivo* clusters and share some similarities with components of the *Mycobacterium tuberculosis* trehalose-recycling ABC transporter that was shown to be essential for virulence (Kalscheuer et al., 2010; Rengarajan et al., 2005; Sassetti and Rubin, 2003) where trehalose-recycling was shown to be involved in formation of mycolate-containing cell wall and not really related to metabolism at all (Kalscheuer et al., 2010).

One key nutrient source for pathogenic bacteria within an insect host is chitin (Gooday, 1990; Souza et al., 2011). In contrast, the *in vivo* transcriptome of *Y. pestis* in the flea gut, identified that most sugar uptake systems were repressed except for the phosphotransferase system (PTS) uptake utilization system for chitobiase indicating available nutrient sources in the flea gut likely consisted of mostly proteins, lipids and chitin (Vadyvaloo et al., 2010). In MH96 the gene for the PTS uptake systems for the chitin monomer *N*-acetylglucosamine (GlcNAc) (PL78_01275/nagE PTS N-acetyl glucosamine transporter subunits IIABC) were not found to be highly expressed by MH96 *in vivo*; however, chitin is still thought to be a major carbon/nitrogen source for MH96 during insect infection and the gene for Yen-TC component *chi1* and a putative LCMO/CBP (PL78_08295) were identified as having increased *in vivo* expression, but a gene for second putative LCMO/CBP, *cbpA* (PL78_05310) did not have the same response to *in vivo* conditions as PL78_08295.

Crosstalk between metabolism and virulence have been previously identified among other pathogenic *Yersiniae* before, highlighting complex regulatory networks underpinning global regulation of metabolism and virulence at the post transcriptional level, which have been shown to involve conserved RNA-binding proteins, such as Hfq (Kakoschke et al., 2014; Lathem et al., 2014; Schmid et al., 2009) and carbon-storage regulator (Csr) (Heroven et al., 2012; Romeo et al., 2013). Among transcripts with the longest predicted 3′ and 5′ untranslated region (UTR) regions identified from the MH96 transcriptome were several genes related to carbon metabolism (supplementary Table S11) indicating an important role for post-transcriptional regulation in MH96 as well. So, while the *in vivo* transcriptome of MH96 does provide some insights into possible metabolic strategies occurring during insect infection, much further work is required to understand how post-transcriptional regulatory mechanisms effect metabolism and virulence in MH96 *in vivo* as well. Molecular phenotyping by *in vivo* RNA-seq of *hfq* or *csrA* deficient mutants would help to characterize important metabolic and virulence regulatory networks in MH96. As a follow-up, the temperature-dependent metabolic capabilities of MH96 were investigated using the Omnilog phenotype by microarray system to identify sole carbon sources that MH96 can utilize at 25 and 37 °C and is discussed in below in section 4.4.5.

4.4.4 Temperature- and host-dependent gene expression in MH96 during exponential growth

4.4.4.1 Yen-TC and other putative TC-associated VFs are co-regulated in a temperature- and host-dependent manner

In vivo transcriptomics approaches were also used to investigate the genome-wide temperature-dependent responses *in vivo* by MH96 using a two-way factorial experimental design. Previous work has shown that temperature is an important regulatory cue for MH96, with obvious phenotypic differences observed between 25 and 37 °C; briefly, at 37 °C *in vitro* MH96 has generally reduced global protein secretion and levels of YenA1 and YenA2 compared to 25 °C (Hurst et al., 2011a) (Chapter 2, 2.3.3, Figure 2.3) and during intrahemocoelic infection of *G. mellonella* at 37 °C the bacterial cell shape was found to be much smaller and more uniform compared to 25 °C where bulging, spherical and filamentous cell shapes occur at higher cell densities during infection (Hurst et al., 2015). Temperature also has a significant effect on MH96 virulence against *G. mellonella* by *per os* challenge, but not by intrahemocoelic injection (Hurst et al., 2015). Wide-spread transcriptional responses to temperature by MH96 confirmed temperature does play an important role in coordinating gene expression in MH96; however, the most critical VFs, Yen-TC and other TC-associated factors only became transcriptionally repressed at 37 °C when exposed to *in vivo* conditions, implying layered regulatory mechanisms involving both temperature and host-induced cues are involved in global repression of Yen-TC and TC-associated factors at the transcriptional level. Additional work to validate the temperature-dependent regulation of the *chi1-yenA1-yenA2-chi2* 4-gene operon using *lacZ*-translational reporter fused to the promoter region of *chi1* (proxy for Yen-TC activation) *in vitro* was undertaken and reported in Chapter 5, confirming that the level of Yen-TC production is reduced at 37 compared to 25 °C, implying post-transcriptional regulatory mechanisms are responsible for the temperature-dependent difference in toxin component levels under *in vitro* conditions.

Complex temperature-dependent regulation of important VFs in other species of *Yersinia* has been described before. The Ysc-Yop T3SS on the virulence plasmid pYV was found to be dramatically upregulated by *Y. pseudotuberculosis* pre-conditioned

host-cell culture medium (cell free) at 37 °C compared to cells grown in normal growth media at 26 °C, which suggested thermoregulation as the primary control mechanism for expression of the most critical VF produced by *Y. pseudotuberculosis* within the mammalian host (Bent et al., 2015). Thermoregulation of Ysc T3SS at 37 °C is known to involve two levels of transcriptional regulation in *Y. pseudotuberculosis*, including a thermo-sensitive riboswitch within the 5′ UTR region of the mRNA for the AraC-like regulator, *lcrF* (otherwise known as *virF*) and temperature-dependent repression of *lcrF* by YmoA (Böhme et al., 2012), a nucleoid-associated heat-labile global regulator that is known to interact with H-NS.

In *Yersinia* spp., YmoA is also an important temperature-dependent global regulator and in *Y. enterocolitca* YmoA partly controls expression of insecticidal TC genes (Starke and Fuchs, 2014). In *Y. enterocolitica* temperature-dependent regulation is also multi-layered, involving a thermolabile transcriptional activator, *tcaR2* located upstream from the *tc* genes as well (Starke et al., 2013).

While characterizing entire overlaying temperature-dependent regulatory mechanisms involved in coordinating this consortium of TC-associated factors by MH96 is well beyond the scope of this thesis, targeted work investigating putative temperature-dependent control of Yen-TC gene expression by PAI_{YE96}-encoded putative regulators *yen6* and *yen7*, represented a specific area of focus and is described in more detail in (see 0, 0section). Further work to validate the temperature-dependent regulation of *yen6* under both *in vitro* and *in vivo* conditions was undertaken by fusing the *yen6* promoter region to a *lacZ*-translational reporter (Chapter 5, section 5.3.3.6.2, Figure 5.26 and section 5.3.3.6.7, Figure 5.31). This finding determined that temperature-dependent regulation of *yen6* is affected by host-dependent conditions by post-transcriptional regulation at the 5′UTR region of *yen6*, but notably enzymatic levels produced by the reporter strain were estimated to be 39-fold higher at 37 °C compared to 25 °C during intrahemocoelic infection (Chapter 5, section 5.3.3.6.7, Figure 5.31) supporting previous *in vivo* transcriptome findings that *yen6* expression is also maximum under *in vivo* conditions at 37 °C as well.

Aside from *in vivo* temperature-dependent responses of Yen-TC and TC-associated factors, which is described above, the other most significant finding was wide-spread

temperature-dependent response by both the flagella and the T6SS, both of which were completed repressed at 37 compared to 25 °C. Reduced expression of flagellar-related genes at 37 °C has also been described in *Y. enterocolitca* (Bent et al., 2015; Horne and Prüss, 2006; Kapatral et al., 1996, 2004) and *Y. pseudotuberculosis* (Nuss et al., 2015). The flagella sigma factor, *fliA*, has been shown to be important to the regulation of not only the flagella apparatus but also several important VFs located on the virulence plasmid pYV (Horne and Prüss, 2006). It has been proposed, that in enteropathogenic *Yersinia*, that the temperature-dependent repression of the flagella at 37 °C is a strategy to avoid triggering the mammalian immune response by Tol-like receptor 5 activation (Minnich and Rohde, 2007), due to the promiscuous secretion of the flagellar components by the Ysc T3SS that is a critical VFs required for mammalian infection. This rationale however, is not consistent with the temperature-dependent response of the flagellar operons in MH96 because neither the T3SSYE1 or T3SSYE2 was found to be highly expressed at 37 °C *in vivo.* Like the flagella the entire T6SS of MH96 was repressed at mammalian host temperature as well, indicating that both motility the production of the T6SS apparatus may be detrimental in the mammalian host environment. A study characterizing the outer membrane proteome of *Y. pestis* cells grown at 26 and 37 °C also determined thirteen proteins associated with a function T6SS that were associated with cells grown at 26 but not at 37 °C (Pieper et al., 2009).

While the temperature-dependent expression of the pYV-encoded Ysc T3SS is a hallmark of other pathogenic *Yersinia* species, the T3SSYE1 and T3SSYE2 were not found to respond to temperature (during the exponential growth phase at least). The T2SS was identified as linked to the co-expression of Yen-TC and TC-associated factors, and since several components of the T2SS were found to have higher expression *in vivo* this can be taken as evidence that the T2SS is most likely involved in secretion of VFs during insect infection by MH96. As described above, the T2SS of *Y. enterocolitica* biotype 1B strains encode Yts1 (Iwobi et al., 2003; Thomson et al., 2006), which was found to secreted two chitin- and one oligosaccharide-binding exoenzymes (Shutinoski et al., 2010). Both of the putative LCMOs/CBPs encoded by MH96 contain signal sequence, *cbpA* (PL78_05310) is known to be 23 kDa and secreted into the culture medium by MH96 at 25 °C and co-expressed with Yen-TC and TC-associated factors, the second LCMO (PL78_08295) was identified as constitutively highly expressed *in*

vivo, and based on these transcriptome results the two LCMOs/CBPs thought to be important T2SS-secreted chitinolytic components of the insect-specific secreted machinery of MH96.

Additionally, the holin-endolysin secretion system was found to be co-expressed with Yen-TC and TC-associated genes and is also thought to be involved in coordinated secretion of Yen-TC components and co-secreted factors that likely target insect-specific tissue and cell-surface ligands. Considering genes for putative VFs associated with the T3SSYE2 and T6SS were also identified from *in vivo* clusters, and it is also thought that secretion systems could be important for pathogenesis of MH96 against insects but may more have targeted effects on hemocytes. While *in vivo* transcriptomics provided insights into a highly complex virulence strategy used by MH96, likely involving multiple toxins (including Yen-TC), translocated effectors, secretion systems and secreted exoenzymes, the cytotoxic effects and insect-specific targets and function roles of all but Yen-TC, remain to be validated.

4.4.4.2 During *in vivo* conditions at 37 °C MH96 upregulates genes for stress response, non-ribosomal peptide synthase, filamentous hemagglutinin adhesin and sulfur-acquisition

Like *in vivo* clusters containing genes with higher expression during early infection, multiple genes for several homologous CSPs were identified as having significantly higher expression *in vivo* at 37 °C compared to 25 °C. Overall, genes for six CSPs (three tandem – *cspA123* and duplicate *cspA4*; duplicate *cspB2* and *cspE2*) and were found to have significantly higher expression *in vivo* at 25 °C and 37 °C during early infection and middle infection, respectively compared to *in vitro* controls. The highly duplicative CSPs encoded by MH96 and significantly higher *in vivo* expression profile resulted in the *cspA123* locus becoming a focus of future work reported later in this thesis (Chapter 6, section 6.3.1).

In addition to the CSPs, another potential stress response regulator that was identified by this specific analysis was an IbpA-like heat shock protein that was co-expressed with genes for Yen-TC and TC-associated factors and may play a role in global virulence regulation by MH96. The primary role of heat-shock proteins is to assist in

protein folding during environmental stress, especially the rescue or degradation of denatured proteins or misfolded aggregates (Roncarati and Scarlato, 2017). While heat-shock proteins have not been investigated in entomopathogenic bacteria, given the regulatory linkage to Yen-TC and TC-associated factors, it is possible that the small heat-shock protein may be a chaperone that prevents aggregation of the large amount of secreted toxins and exoenzymes produced by MH96 at 25 °C. Two genes encoding the heat shock proteins, IbpB and IbpA in *Y. pestis* were both shown to have significantly higher expression during macrophage infection as well (Fukuto et al., 2010). Assessment of an *ibpA* deficient mutant for protein secretion would be an important next step to defining the role of heat-shock proteins in MH96.

While the genes for Yen-TC and co-secreted carbohydrate-binding/chitinolytic factors were highly repressed at 37 °C *in vivo*, the transcriptome identified at least three other putative toxins/effectors that had increased expression during growth at higher temperatures. The expression profile of filamentous hemagglutinin N-terminal containing protein, PAI_{YE96}-encoded *yenV* and an adenylate cyclase, *adyC* genes all had a strong positive response to 37 °C *in vivo* conditions. These findings indicate that these putative primary VFs represent toxins/effectors that could have specificity against mammalian cells. Furthermore, secretion of the filamentous hemagglutinin N-terminal containing protein in the culture supernatant from MH96 grown at 37 °C, but not 25 °C was confirmed by LCMS-ESI-MS/MS (Chapter 5, section 5.3.1.1, supplemental Figure S16 and supplemental Table S12), which confirms that this filamentous hemagglutinin N-terminal containing protein is highly expressed at 37 °C but also that it is secreted by MH96 in a temperature-dependent manner as well.

MH96 encodes two putative filamentous hemagglutinin N-terminal containing proteins (previously annotated as hemolysins: PL78_04365 and PL78_11060) (Hurst et al., 2016), but only PL78_04365 was found to have significantly greater expression at 37 °C *in vivo,* although PL78_11060 did share a similar expression profile, but the difference in expression were not determined to be statistically significant. Also, each of the filamentous hemagglutinin N-terminal containing proteins is associated with an upstream T5bSS autotransporter (PL78_04360 and PL78_11050) and these genes were also more highly expressed at 37 °C *in vivo*. These results show that the production of

filamentous hemagglutinin N-terminal containing proteins by MH96 is under temperature-dependent regulation, and because PL78_04360 and two putative TPSs were found to be more highly expressed at 37 compared to 25 °C, there is a very high likelihood that this putative VF plays an important role in MH96 virulence during mammalian infections.

In addition to high production of filamentous hemagglutinin N-terminal containing protein at 37 °C *in vivo*, MH96 also was found to highly express genes for two other interesting putative effectors: one a leucine-rich repeat (LRR) containing protein (PL78_01795, previously annotated as adenylate cyclase) and the other PAI_{YE96}-encoded *yenV* (PL78_03780), which is likely a calcium/calmodulin dependent protein kinase. Many pathogenic bacteria are known to express proteins containing LRR domains, which function in adhesion and host-cell invasion (Ireton, 2007) as well as effectors that modulate host signal transduction (Bernal-Bayard et al., 2010; Fernandez-Prada et al., 2000; Leung et al., 1990; Miao et al., 1999). YopM is a LRR containing effector from *Y. pestis* that was found to inhibit aggregation of platelets and required for virulence in the mouse infection model (Leung et al., 1990). The human pathogen, *Leptospira interrogans* was found to produce an extracellular LRR-containing effector with binding specificity against human E- and VE-cadherins and is thought to be an important adhesin (Eshghi et al., 2019). The MH96 encoded LRR-containing putative effector shares a predicted protein structural similarity to the crystal structure of the LRR protein lic10831 of *L. interrogans* (c4u06A) based on Phyre2 analysis (100 % confidence/35 % i.d.).

Another putative VF that was found to be under temperature-dependent regulation was NRPS (PL78_15621), which has higher expression at 37 compared to 25 °C and was also identified as having higher *in vivo* expression at 25 °C during early infection. In another transcriptome study, a significant increase in the expression of the NRPS gene *gxpS* in *P. asymbiotica* at 37 compared to 28 °C was also identified, the authors to suggest that the GameXPeptide synthase has a specific role in mammalian hosts compared to insects (Mulley et al., 2015). Other NRPS are known to respond to environmental different cues, for example in *X. nematophila*, a very complex co-regulatory network was identified, linking motility, exoenzyme and antibiotic

production, and the TCRS, consisting of membrane sensor kinase EnvZ and response regulator OmpR negatively regulated NRPS, *nrps1* (Park and Forst, 2006). Production of secondary metabolites by NRPS has also been described from entomopathogenic bacteria before such as those produced by the honey bee specific pathogen *Pa. larvae*, which uses NRPS and polyketide synthases (PKS) to produce secondary metabolites with antibacterial and antifungal activities thought to play a role in eliminating bacterial competitors (Garcia-Gonzalez et al., 2014a, 2014b; Müller et al., 2015; Sood et al., 2014) or possibly even have cytotoxic effects against host cells (Sood et al., 2014). MH96 encodes three putative NRPSs (PL78_15621, PL78_04165 and PL78_09730); however, PL78_15621 was shown to have the most interesting *in vivo* expression profile, notably linked to several CSPs discussed above and possibly contributing to the production of yet unidentified secondary metabolites that play an important role in pathogenesis by MH96. In general, the products of NRPS and PKS in entomopathogenic bacteria remains a largely untapped pharmacopeia of antibacterial, antifungal and insecticidal agents and much current research focus on secondary metabolites of *P. luminescens* (Nollmann et al., 2015; Orozco et al., 2016; Tobias et al., 2016).

Lastly, specifically under *in vivo* conditions at 37 °C, several gene clusters related to sulfur acquisition were upregulated by MH96, signifying sulfur acquisition is important during infection at 37 °C. Recently transposon mutagenesis study identified cysteine metabolism and sulfur assimilation as essential for survival of *Acinetobacter baumannii* in *G. mellonella* (Gebhardt et al., 2015) and sulfur assimilation was also found to important for virulence in the murine pneumonia model (Gebhardt et al., 2015; Wang et al., 2014a). In other pathogens, sulfate acquisition has been intrinsically linked to stress response and redox homeostasis (Álvarez et al., 2015; Kobayashi et al., 2014; Shatalin et al., 2011; Turnbull and Surette, 2010) and as an important part of the potent antioxidant glutathione (GSH) (Newton and Fahey, 2017). At 37 °C *in vivo*, MH96 upregulates genes related to acquisition of several sources of sulfur, including sulfate, thiosulfate, taurine and methionine. Perhaps in the absence of Yen-TC and TC-associated factors at 37 °C *in vivo*, MH96 has reduced capacity to prevent the immune system of *G. mellonella* from mounting an oxidative attack, so compared to 25 °C *in vivo*, sulfur uptake pathways are activated in response to increased oxidative stress at 37 °C

in vivo. Some genes related to cysteine metabolism were also identified as putative VFs with increased expression *in vivo* at 25 °C, and like *A. baumannii* sulfur acquisition maybe a critical requirement for growth of MH96 within the hemolymph of *G. mellonella*. Further experimentation would be required to determine whether the upregulation of numerous sulfur acquisition related pathways by MH96 *in vivo* at 37 °C is in response to temperature-induced nutrient availability within the host (Hanzal and Jegorov, 1991) or increased oxidative stress at 37 °C.

4.4.5 Temperature-dependent metabolic shifts in MH96 identified by phenotype microarray

The regulation of virulence and metabolism have been shown to be inextricably linked in pathogenic bacteria, which rely on sensory systems that regulate expression of both metabolic and virulence genes in response to available nutrient sources within the host (Kwaik and Bumann, 2013; Rohmer et al., 2011; Schmid et al., 2009). In many pathogens, temperature is an important cue that can induce specific signals according to whether a pathogen is in a mammalian (37 °C) or non-mammalian host, such as an insect. Furthermore, numerous *in vivo* transcriptome investigations have also implicated key regulatory themes related to metabolism and virulence during host infection (Jorth et al., 2013; Koczula et al., 2017; Wang et al., 2016). In one study of an emerging human pathogen, *P. asymbiotica* that can also infect insects and associated with nematodes, temperature was found to have limited effects on global gene transcription; however widespread reduction of sole-carbon and -nitrogen source utilization at 37 compared to 28 °C were identified (Mulley et al., 2015). Here also, sole-carbon utilization by MH96 was shown to be greatly impacted by temperature as well. Most notably, the utilization of sugar carbohydrates was significantly affected by temperature, such that MH96 was found to only utilize GlcNAc, D-trehalose, D-ribose, D-fructose, alpha-D-glucose, sucrose, D-mannose and D-mannitol at 25 °C and not 37 °C.

The hemolymph of *G. mellonella* was recently shown to be relatively high in sugars, especially trehalose (a glucose-glucose disaccharide), which was found to be the most abundant carbohydrate in the hemolymph (Killiny, 2018). Trehalose is a common sugar

known to be important in insects for carbon storage (Elbein et al., 2003) as well as a source of glucose energy (Reyes-DelaTorre et al., 2012) and cryo-protectant (Saeidi et al., 2013; Wang et al., 2010). Since MH96 cannot utilize trehalose at 37 °C based on phenotype microarray results, it is likely that trehalose represents a carbohydrate that present in greater abundance during insect infection compared to mammalian infections.

Similar to trehalose, GlcNAc is thought to also be a carbon and nitrogen source that would be readily available during infection of an insect host but not a mammalian host because GlcNAc is the building-block of the polymer chitin, a key component of the insect biomass (Gooday, 1990). MH96 is not able to utilize GlcNAc as a sole-carbon source at 37 °C but could grow on this compound at 25 °C, which suggests GlcNAc is also an abundant carbon and nitrogen source found within insect, but not in mammals. MH96 is known to secrete chitinolytic enzymes, including CbpA (PL78_15310) and three chitinases, two associated with Yen-TC (Chi1 and Chi2) and a third chitinase (PL78_11910) and these were found to be under temperature-dependent regulation, but only under *in vivo* conditions (Chapter 4, section 4.3.8, Table 4.13). The down regulation of the chitinolytic machinery by MH96 during *in vivo* conditions at 37 °C is coordinated with the metabolic down regulation of GlcNAc uptake and utilization systems in MH96 as observed by phenotype microarray.

Major differences in amino acid utilization were also observed as a result of temperature shifts in MH96; only at 37 °C could MH96 utilize L-glutamic acid, L-glutamine, Gly-glu and L-threonine, while only at 25 °C could MH96 utilize L-aspartic acid, L-asparagine and L-histidine. It is presently unclear what the biological purpose is for temperature-dependent shifts in amino acid utilization as sole-carbon source by MH96; however, it is assumed that these changes likely reflect true shifts in amino acid abundances encountered between infection of mammalian or insect hosts. Furthermore, at 37 °C MH96 was able to utilize the glycosylated pyrimidine-analog uridine and the low molecular-weight polymer dextrin, which is a non-digestible starch hydrolysis product (Chronakis, 1998; Sun et al., 2010), suggesting MH96 may have this ability at 37 °C to access dextrin present within the mammalian gut. While some of the phenotype microarray data were found to be somewhat unreliable (i.e.,

dextrin, sucrose, D-mannose, D-mannitol and L-lactic acid), for the most part the remaining findings were very consistent between the two replicate plates. Additional work to validate these temperature-dependent metabolic shifts observed in MH96 and also the role of important carbohydrate sources, like trehalose and GlcNAc in pathogenesis remains a fascinating area of future research.

4.4.6 Description of ncRNA identified in the MH96 transcriptome

A variety of non-coding transcripts were identified in the MH96 transcriptome. Whilst, not the key focus of this research, the below sections provide brief descriptions of notable patterns or findings related to the ncRNA species expressed by MH96 under both *in vivo* and *in vitro* conditions. Specific expression patterns of ncRNAs are described elsewhere in the thesis (i.e., Chapter 5, section 5.3.2.2, Figure 5.16 and 0 section 7.3.1, Figure 7.7Figure 7.8) and are considered within their local genetic context but a global analysis of ncRNAs was not considered here.

4.4.6.1 *Trans*-acting ncRNA prediction in MH96 transcriptome

Genome-wide and unbiased investigations of bacterial transcriptomes using RNA-seq have recently fostered a new appreciation for the complexity of the transcriptional and post-transcriptional landscapes that have evolved herein. One of the key findings of early bacterial RNA-seq studies was the discovery that bacteria widely produce *trans*-acting ncRNA species, which can act as global regulators of many important process, such as response to stress and pathogenesis (Caldelari et al., 2013; Chaloner et al., 2016; Citartan et al., 2016; Fröhlich and Papenfort, 2016; Quereda and Cossart, 2017; Westermann et al., 2016). Over 200 putative ncRNAs were predicted from the MH96 transcriptome during *in vitro* and *in vivo* growth conditions, across multiple cell-densities and two temperatures. Bacterial *trans*-acting ncRNAs (typically between 50 – 400 nt in length) have been widely reported as global regulators of human pathogenic bacteria (Caldelari et al., 2013; Chaloner et al., 2016; Fröhlich and Papenfort, 2016; Ortega et al., 2014; Quereda and Cossart, 2017; Westermann et al., 2016); however, there are very few (if any) reports of ncRNAs regulating virulence in insect pathogens to date. Typically, *trans*-acting ncRNA act by targeting mRNAs through base pairing of

short seed regions usually located in the 5′ UTR of the target messenger. The secondary structures formed by ncRNAs are important and binding of target often requires the presence of molecular chaperones, such as Hfq (Vogel and Luisi, 2011). While beyond the scope of this thesis, further characterization of the role of *trans*-acting ncRNAs in pathogenesis by MH96 represents an interesting area for future research. Such work could be facilitated by high-throughput identification of Hfq-dependent ncRNAs using UV-crosslinking and immunoprecipitation strategies as a means to validate/further characterize *trans*-acting ncRNAs in MH96 (Barquist and Vogel, 2015).

In addition to *trans*-acting ncRNA, it has only been recently found that (in at least some) bacteria produce functional microRNA-like ncRNAs (MncRNA), which are extremely small (< 35 nt) compared to *trans*-acting nc*RNA* (50 – 400 nt) (Choi et al., 2018; Kang et al., 2013; Lee and Hong, 2012; O'Neill et al., 2014). Here we report that greater than 50 % of the ncRNA predicted from the MH96 transcriptome were less than 50 nt in length. While the function of MncRNA by pathogenic bacteria remains largely unexplored, some evidence points to a potential role in host manipulation. For example, several different strains of the insect endosymbiont *Wolbachia* were shown to express two functional MncRNAs within *D. melanogaster*, *D. simulans*, and *Aedes aegypti* (O'Neill et al., 2014). This study also provided evidence that MncRNA produced by *Wolbachia* were exported into the host cells, and that one specific MncRNA, *W*snRNA-46, contains a seed region with perfect complementation to mRNA for the host actin heavy chain Dynein (*dhc*). When a *W*snRNA-46 mimic was injected into insect cells the amount of *dhc* increased and Wol^{+} mosquitos produced greater amounts of *dhc* compared to Wol^{-}. Based on these findings the authors suggested that modulation of host microtubules through regulation of d*hc* by *W*snRNA-46 likely played a role in bacterial placement within the oocyte facilitating vertical transmission. Secretion of both ncRNA and MncRNA in bacterial outer membrane vesicles has been recently characterized (Choi et al., 2017, 2018; Koeppen et al., 2016), also supporting a more direct role in host-manipulation by MncRNA in pathogenic bacteria.

While at least some of these predicted ncRNAs smaller than 50 nt in length may represent *bona fide* MncRNA, the MH96 transcriptome were derived from libraries with an average fragment length of 300 bp, making it very challenging to further

characterize these unusually small predicted ncRNAs. Future work to characterize MncRNA produced by MH96 may include deep sequencing cDNA libraries containing average fragments ~ 35 nt in length. It was found that some of these MncRNA-like species were detected directly adjacent to predicted transcriptional starts of annotated ORFs, or as intergenic components among operons. This latter phenomenon was also reported from the transcriptome of *Wolbachia* strain *w*Mel, where ncRNAs were detected as intergenic components of polycistronic mRNA (Woolfit et al., 2015). Subsequently, predicted ncRNAs less than 50 nt were generally not considered in this report, except for several predicted ncRNA (<50 nt) found within key genomic regions of interest (i.e., Yen-TC operon, for example). In this context, such ncRNAs were considered only within the local genetic landscape, in a presence/absence manner (i.e., differential expression was considered unreliable due to library preparation bias against small fragments).

When considering predicted ncRNAs from the MH96 transcriptome, it is also possible that some may be translated into peptides. For example, many annotation pipelines do not consider ORFs for proteins less than 100 residues; however, small RNAs have been shown to be common sources for small peptides among phylogenetically widespread bacteria (Friedman et al., 2017). For example, two putative ORFs identified as ncRNAs (ncRNA130 and ncRNA131) were found among the holin-endolysin system within the TC phage-associated region where identified in the MH96 transcriptome during targeted investigations of genomic regions of interest. Other similar holin-endolysin secretion systems used by pathogenic *Yersinia* spp. to secrete insecticidal TCs are comprised of four small ORFs without any intergenic space in between the coding regions of holin-endolysin cassette genes (Fuchs et al., 2008; Springer et al., 2018a); however further work is required to validate ncRNA130 and ncRNA131. Further work could be under-taken to improve ORF prediction in MH96 reference genome or further characterize the predicted ncRNAs using the Rfam database (Nawrocki et al., 2014).

4.4.6.2 *Cis*-acting ncRNA and regulatory overlapping UTRs predicted in MH96

By using strand-specific cDNA library preparations methods, asRNA can be identified in transcriptome data as reads originating from the DNA strand opposite to a protein

coding region and asRNAs were predicted to overlap approximately 1 - 2 % of protein coding regions of the MH96 genome. These so-called, *cis*-acting asRNAs share significant complementarity to the sense transcript, and are known to form dsRNA with targets, which often (but not always) result in increased rate of degradation by specific RNAses (Cossart et al., 2013; Georg and Hess, 2011; Saberi et al., 2016; Thomason and Storz, 2010). Of note, multiple asRNA were identified overlapping the genes of known VFs in MH96, including Yen-TC components *chi1* (two asRNA), *yenA1* (two asRNA) and *yenA2* (four asRNA), suggestive of potential post-transcriptional regulatory mechanisms here. AsRNAs were also found to overlap genes for other TC-associated factors, including peptidase M66 (one asRNA; PL78_05495), a hypothetical VF PL78_18785 (three asRNA), as well as known temperature-dependent secreted proteins *groEL*-like (PL78_08315) and two *ompA*-like homologs (PL78_01525 and PL78_02825). The expression of the asRNA overlapping any genes for Yen-TC components or TC-associated factors mirrored the expression profile of the sense transcripts (i.e., asRNA expression increased with sense transcript expression). AsRNA were also commonly predicted for many genes for ribosomal proteins, all of which were highly expressed across all libraries. While, the function of only a few bacteria asRNA have been described to date, it is also thought (at least by some) that pervasive asRNA transcription may be a result of transcriptional noise or sequencing artefact (Georg and Hess, 2011; Lloréns-Rico et al., 2016; Raghavan et al., 2012). So, further investigations would be required to determine if any of the asRNA overlapping genes for Yen-TC or TC-associated factors are important in the regulation of virulence in MH96, or just naturally occur for highly expressed genes, like the genes for the Yen-TC components.

Cis-acting asRNA originating from long overlapping 5′ and 3′ UTRs have also been identified in bacteria (Mraheil et al., 2011; Rasmussen et al., 2009; Sesto et al., 2013; Silby and Levy, 2008; Toledo-Arana et al., 2009) and cyanobacteria (Hernández et al., 2006; Hess et al., 2009). These types of asRNA are thought to provide a fine-tuning regulatory mechanism, such that the mRNA and asRNA can act on opposing operons simultaneously (Sesto et al., 2013). Many long 5′ and 3′ UTRs were identified in the MH96 transcriptome but were not examined specifically for potential overlaps in a genome-wide manner; however, one extremely long 3′UTR (646 nt) predicted on the

putative mRNA of *yen6* that is encoded on PAI$_{YE96}$. Both *yen6* and adjacent *yen7* were identified as a target for further study and are describe in more detail in Chapter 5 (section 5.3.2). Since, other interesting features can also be encoded in long 5′ and 3′UTR regions, such as RNA sensors or novel ncRNAs, future work to characterize such features could be carried by assigning possible annotations to predicted 3′ and 5′ UTR regions using the RNA family data Rfam (Nawrocki et al., 2014) or identification of secondary structures within these regions. Future studies may also focus on better defining transcriptional boundaries using dRNA-seq (differential RNA-seq) which allows better determination of transcriptional start sites by distinguishing between primary and processed transcripts in bacteria (Sharma and Vogel, 2014). A list of the 100 longest predicted 5′ and 3′ UTRs is provided in supplementary (Table S11).

4.4.6.3 Limitations and future directions with respect to putative VF identified by *in vivo* transcriptomics in MH96

The exploration of the temperature- and host-dependent *in vivo* responses of MH96 within the model host, *G. mellonella* provided a proof-in-principle that functional enrichment of putative VFs with higher *in vivo* expression is a useful approach to identify toxins and secreted effectors that likely have insecticidal activities against hemocytes of *G. mellonella.* Using these methods MH96 is predicted to exploit multiple synergistic modes of pathogenicity upon infection of the insect hemocoel, including secretion of toxins and LCMOs/CBPs, translocation of T3SS and T6SS effectors, adhesion to host cells by filamentous hemagglutinin and fimbrial adhesion or avoidance of detection by LPS and outer membrane protein modifications, which supports the observation that MH96 is extremely virulent during intrahemocoelic infection of *G. mellonella*, with only a single cell estimated to result in 100 % mortality. Other key findings from the *in vivo* transcriptome revealed a key shift for increased motility and iron acquisition at later stationary periods of *in vivo* growth at 25 °C and the unexpected observation of genes related to active mobilization by genetic elements as prevalent among *in vivo* clusters.

A prominent temperature- and host-dependent regulatory mechanism involving the global repression of Yen-TC and TC-associated factors, including several known co-

secreted chitinolytic and carbohydrate-binding exoenzymes, at 37 °C *in vivo* was identified by comparing genome-wide expression at 25 and 37 °C under both *in vitro* and *in vivo* growth conditions during exponential growth stage. This approach also revealed that transcriptional activation of both T6SS and flagella machineries are primarily driven by temperature in MH96, and greatly repressed at mammalian body temperature. While the expression of genes for all secretion systems, motility, Yen-TC and TC-associated factors are great reduced during *in vivo* growth at 37 °C, other putative VFs like filamentous hemagglutinin adhesin, multiple sulfur acquisition systems, multiple CSPs and NRPS were identified as more highly expressed at 37 °C *in vivo* compared to 25 °C. These results demonstrate that while MH96 is highly effective at killing *G. mellonella* at both temperatures by intrahemocoelic infection, there are significant changes occurring at the transcriptome level.

One of the major limitations of this approach was that many interesting putative VFs were not identified as sharing significant sequence homology to known VFs within the VFDB database. Like most annotations strategies, this approach is only able to capture known VFs, but is not able to capture novel or hypothetical VFs that may be important for virulence of MH96, especially since the VFDB database is likely biased towards VFs known from mammalian pathogens rather than entomopathogens. Future work may benefit from including additional annotation from other virulence databases or specific databases related to mobile genetic elements. Another recommendation that would improve the ability to assign robust clusters is to add at least one or two more additional collection timepoints which would greatly benefit the fuzzy clustering process and likely improve the ability to identify discrete clusters containing genes with tight regulatory linkages.

A major challenge with characterizing *bona fide* VFs in MH96 the difficulty demonstrating a reduction in virulence upon targeted gene disruption, given the diversity of putative VFs expressed during insect infection. In this sense, the results of the *in vivo* transcriptomic investigations were helpful to determine a list of putative VFs and regulators representing genes of interest (GOIs) for targeted mutagenesis and phenotyping, including LCMO/CBP encoded by TC-associated factor *cbpA* (PL78_05310), tandem triplicate CSPS encoded by *cspA123* (PL78_18365, PL78_17450

and PL78_18370), the UC fimbrial protein (PL78_12480), putative regulator *rovA* (PL78_05820), T6SS sheath protein encoded by *vipB* (PL78_00910) and putative PAI_{YE96}-encoded regulators *yen6* (PL78_03730) and *yen7* (PL78_03735), see Chapter 5, section 5.3.1.2, Table 5.2.).

Chapter 5. Mutant phenotyping by SDS-PAGE/intrahemocoelic infection of *G. mellonella* and molecular characterizations of Yen6 and Yen7 as potential thermoregulators of Yen-TC production and secretion by MH96.

5.1 Introduction

Previously reported, transcriptomic investigations of *Yersinia entomophaga* MH96, which aimed to discover important host- and temperature-dependent virulence factors (VFs), including virulence regulators, revealed many interesting genes of interest (GOIs) that were found to have explicit responses to temperature or in-host growth conditions (Chapter 4, sections 4.3.7 and 4.3.8, Figure 4.17, Table 4.13). Subsequently, several of these GOIs were targeted for mutagenesis (Table 5.1) and then deficient strains were investigated for phenotypic variation, which included mutation effects on global protein production and secretion, including components of the insecticidal toxin complex (TC) – Yen-TC, under *in vitro* conditions and pathogenesis using the intrahemocoelic *Galleria mellonella* infection model. Mutants with attenuated virulence compared to wild-type (i.e., Δ*cspa123*/ΔHCUI$_{YE96}$ and Δ*yen6*) were the subject of much more detailed phenotyping using molecular, transcriptome and targeted bioinformatics approaches reported in Chapter 6 and 7 of this thesis. The major focus of the remainder of this chapter was to determine whether hypothetical regulators Yen6 and Yen7 are directly involved in thermoregulation of the expression of Yen-TC component genes (see section 5.1.1 below). Applicable bioinformatics tools were used to further characterize putative regulators *yen6* and *yen7* and surrounding genomic regions. The role of Yen6 and Yen7 as temperature- and host-dependent regulators of Yen-TC were tested using molecular microbiology approaches, involving *lacZ*-reporter fusions, arabinose induction vectors and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Table 5.1: Summary of genes of interest targeted for mutagenesis and in this study.

Gene of Interest	Locus tag	Host-dependent factor? *	Temperature-dependent factor? †	Putative function	Other rationale
Chitin-binding protein, *cbpA*	PL78_05310	No.	Expression profile is shared with Yen-TC and TC-associated factors, with higher expression (> 2.5 fold-change log_2 CPM††) observed at 25 compared 37 °C *in vivo* (i.e., temperature-dependent response not observed *in vitro*).	Lytic chitin monooxygenase (also known as chitin-binding protein) that enhances virulence against insect host by oxidative cleavage of the insect-abundant biopolymer chitin.	Co-expressed with other TC-associated factors, including Yen-TC holotoxin, contributing to the degradation of the host peritrophic membrane allowing access to the mid-gut epithelium where Yen-TC and other VFs can bind to host cell ligands.
Cold-shock protein triplication: *cspA123*	PL78_18365, PL78_17450 and PL78_18370	More highly expressed *in vivo*, especially during early infection (> 4 fold-change log_2 CPM) compared to *in vitro* at 25 °C.	More highly expressed *in vivo* at 37 compared to 25 °C (> 2 fold-change log_2 CPM) *in vivo*.	RNA-interacting chaperone or ribosomal protein, involved in stress response and survival in the host.	Cold-shock proteins also identified as virulence regulators pathogenic bacteria, such as *Listeria monocytogenes* and *Salmonella* (*Sa.*) *enterica* for example.
Usher-chaperone fimbrial protein, *fim1*	PL78_12480	Extremely high *in vivo* expression, especially during early infection (> 6 log_2 fold-change) compared to *in vitro* at 25 °C.	More highly expressed (> 2.5 fold-change log_2 CPM) *in vivo* compared to *in vitro* at 37 °C.	Structural component of an usher-chaperone fimbriae involved in attachment to host.	Usher-chaperone fimbriae have been shown to be important virulence factors involved in attachment to host tissue.

Table 5.1: Summary of genes of interest targeted for mutagenesis and in this study.

Gene of Interest	Locus tag	Host-dependent factor? *	Temperature-dependent factor? †	Putative function	Other rationale
Putative regulator *rovA*	PL78_05820	More highly expressed *in vivo*, especially during late infection (~ 2 log_2 fold-change).	Moderately reduced expression (>1 fold-change log_2 CPM) at 37 compared to 25 °C in both *in vitro* and *in vivo* conditions).	MarR family, H-NS counter-silencer.	Well known for virulence regulation within pathogenic bacteria, including *Yersinia* spp. (*rovA)* and *Photorhabdus luminescens* (*slyA*) for example.
T6SS sheath protein *vipB*	PL78_00910	No.	T6SS is more highly expressed (> 2 fold-change log_2 CPM). at 25 compared to 37 °C in both *in vitro* and *in vivo* conditions.	Structural component of T6SS encoded on Rhs-associated region.	T6SS have been shown to be important virulence factors in other gram-negative bacteria.
Putative regulator *yen6*	PL78_03730	No.	More highly expressed (~ 1.6 fold-change log_2 CPM) at 37 compared to 25 °C *in vitro*. Even higher comparative expression (> 2.5 fold-change log_2 CPM) at 37 compared to 25 °C *in vivo*.	Putative temperature-repressor of Yen-TC *in vivo* at 37 °C.	This putative regulator is located 5′ of the Yen-TC components encoded on PAI_{YE96}.

Table 5.1: Summary of genes of interest targeted for mutagenesis and in this study.

Gene of Interest	Locus tag	Host-dependent factor? *	Temperature-dependent factor? †	Putative function	Other rationale
Putative regulator *yen7*	PL78_03735	More highly expressed (> 2 fold-change log_2 CPM) *in vivo* compared to *in vitro* at °25 C during infection.	Expression profile is shared with Yen-TC and other TC-associated factors, with higher expression (~ 1.7 fold-change log_2 CPM) observed only at 25 compared to 37 °C *in vivo* (i.e. temperature-dependent response not observed *in vitro*).	Putative transcriptional activator of Yen-TC.	This putative regulator is located upstream of Yen-TC components encoded on PAI_{YE96}.

*Host-dependent VFs were defined in Chapter 2, section 2.1, Table 2.1 as having host-specific expression based on RNA-seq and sharing homology with a known VF from the VFDB database and/or identified from a known pathogenicity island (i.e., PAI_{YE96}).

†Temperature-dependent VFs were defined in Chapter 4, section 4.3.8 using RNA-seq data comparing 25 and 37 °C both *in vitro* and *in vivo*.

††Log_2 counts-per-million (CPM) gene expression values were previously reported in Chapter4, using methods described in section 4.2.6.

5.1.1 Exploring *yen6* and *yen7* as putative temperature-dependent transcriptional regulators of Yen-TC in MH96

Interest in putative temperature-dependent regulators *yen6* and *yen7*, located on the PAI_{YE96} of MH96 directly 5′ of the operons encoding the Yen-TC components (described Chapter 2, section 2.3, Figure 2.1), was driven by results of the *in vivo* transcriptome analysis (Chapter 4, section 4.3.8, Table 4.13), which identified temperature- and host-dependent transcriptional responses in the expression profiles of *yen6* and *yen7*, as well as Yen-TC components *in vivo* (Figure 5.1). *Yen6* was found to be highly expressed at 37 compared to 25 °C and increased expression at 37 °C seemed to be enhanced further in the *in vivo* environment (Figure 5.2). A putative *cis*-encoded anti-sense RNA (asRNA) in the 3′ untranslated region (UTR) of *yen6* was also predicted from RNA-seq data. The 3′ UTR is predicted to be an asRNA that overlaps the *yen7* locus, which is encoded on the anti-strand in a tail-to-tail orientation asRNA (referred to as '*yen7as*'). In contrast, a temperature-dependent change in expression of *yen7* was only observed *in vivo*, with much higher expression at 25 compared to 37 °C, which linked *yen7* as co-expressed with the Yen-TC component genes and other TC-associated factors. Based on these findings, it was hypothesized that Yen6 may act as a repressor of *yen7*, which may then have down-stream effects on expression of Yen-TC components located directly 3′ to *yen7* (Figure 5.3). Much of the experimental investigations reported below focused on determining whether *yen6* and *yen7* contribute to the regulation of Yen-TC-mediated virulence in MH96.

The potential for *yen6* and *yen7* to regulate Yen-TC was also supported by recent characterizations of a temperature-dependent regulatory mechanism for a different insecticidal TC in *Y. enterocolitica* W22703, which involves two *cis*-encoded regulator genes *tcaR1* and *tcaR2* that are found 5′ of the insecticidal TC components and code for two LysR-type transcriptional regulators (LTTRs) (Figure 5.3). TcaR2 was shown to be an auto-regulating, heat-labile DNA-interacting protein that interacts with the promoter regions of insecticidal toxin components *tcaA* and *tcaB* in W22703 (Starke et al., 2013). The similar positioning of *yen6* and *yen7* directly 5′ of the Yen-TC components exactly matches the organization of *tcaR1* and *tcaR2*, with respect to the

insecticidal TC in W22703, supporting the hypothesis that *yen6* and *yen7* are likely involved in regulating Yen-TC. Furthermore, since the region of PAI_{YE96} containing the *yen6*, *yen7* and *chi1* genes has extremely high AT-content compared to the core genome (Figure 5.3), it was also considered likely that H-NS and YmoA also exert temperature-dependent regulatory control on this region through global nucleoid interactions with AT-rich regions of PAI_{YE96}. Along these lines, another important regulator of the insecticidal toxin complex in W22703 was identified as the modulator of yersinial virulence, YmoA (known as Hha in other genera), which contributes to the binding properties of global nucleoid-associated protein H-NS (Starke and Fuchs, 2014). Silencing of xenogenetic DNA by H-NS and other nucleoid- associating proteins, such as YmoA/Hha is emerging as a common theme in virulence regulation by pathogenic Enterobacteriaceae (Erhardt and Dersch, 2015; Navarre et al., 2007; Singh et al., 2016) and research has also revealed wide-spread evolution of H-NS counter-silencing mechanisms among pathogenic Enterobacteriaceae (Fang and Rimsky, 2008; Will et al., 2015, 2019), often involving PhoP/PhoQ two-component systems (Will et al., 2014), AraC-like activators GadX and GadW (Tramonti et al., 2006) and MarR family regulators such as SlyA (also known as RovA in *Yersinia* spp.) (Ellison and Miller, 2006; Heroven et al., 2004; Will et al., 2019). Therefore, it was postulated that a putative regulatory mechanism of Yen-TC component gene expression must involve temperature-dependent H-NS/YmoA silencing and counter-silencing (possibly by Yen7) or other commonly known counter-silencers from other pathogenic bacteria (i.e., AraC, PhoB, RovA, etc.).

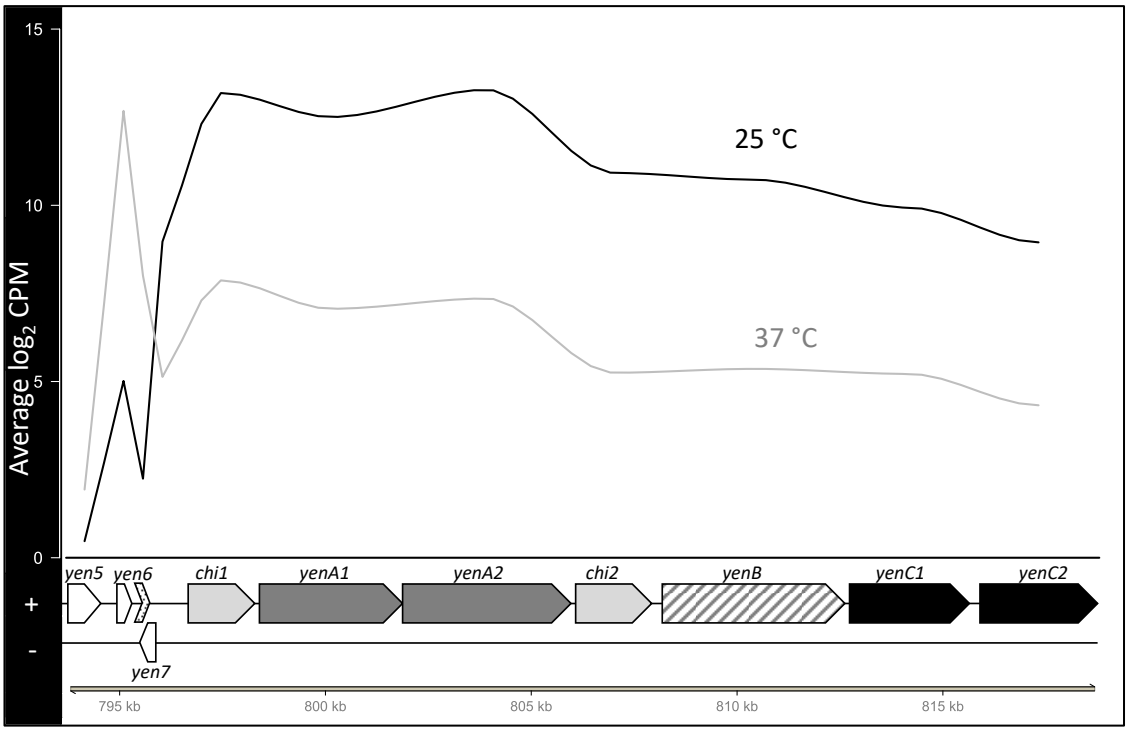


Figure 5.1: Mean Log_2 counts-per-million of the *yen5*, *yen6*, *yen7* and Yen-TC components expressed *in vivo* by *Y. entomophaga* during exponential growth at 25 and 37 °C . Data were previously determined in Chapter 4 using methods described in section 4.2.6. Mean expression were plotted using Gviz package (Hahne and Ivanek, 2016) smoothed using LOESS curve fitting for visualization purposes. Putative regulatory components are white, predicted asRNA is dotted, homolog exochitinases *chi1* and *chi2* are light grey, *yenA* components are dark grey, *yenB* component is diagonal lines and *yenC* components are black.

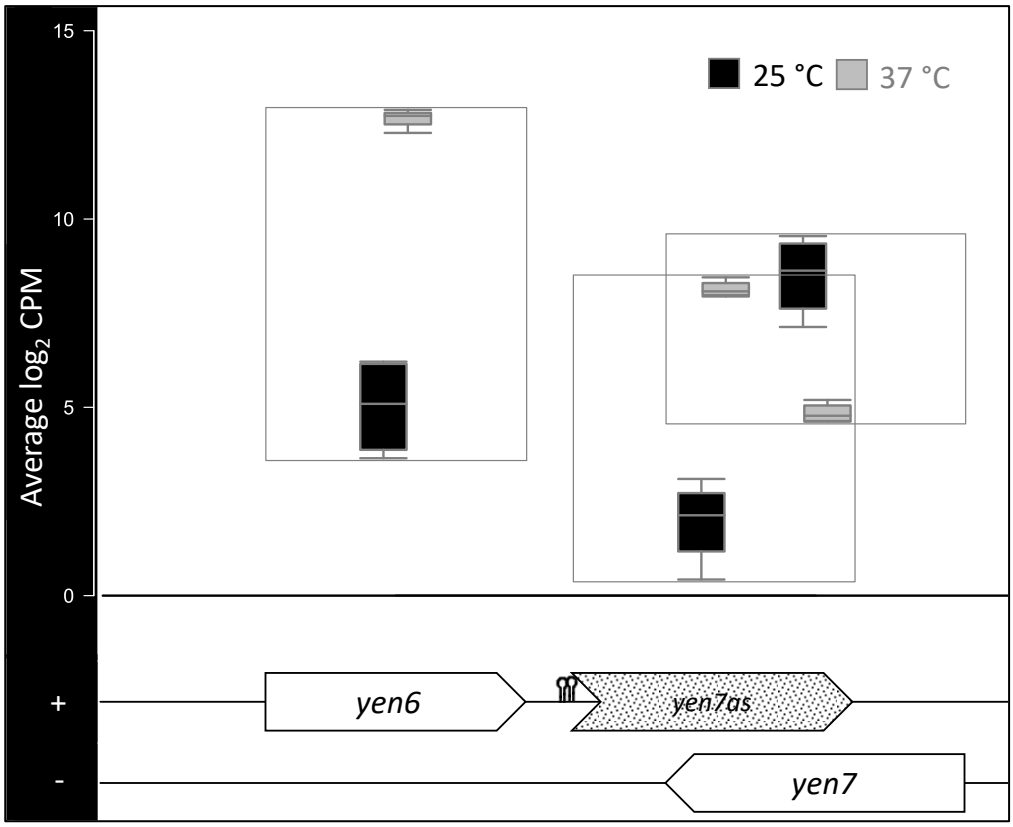


Figure 5.2: Mean $\log_2$ counts-per-million of putative temperature-dependent Yen-TC regulators *yen6, yen7as* and *yen7* expressed *in vivo* by *Yersinia entomophaga* MH96 during exponential growth at 25 and 37 °C . Data were previously determined in Chapter 4 using methods described in section 4.2.6. Boxes outline paired expression data for each of the three transcripts. Putative regulatory components are white and predicted asRNA is dotted. A predicted double-terminator structure is shown between *yen6* and *yen7as* on the positive strand.

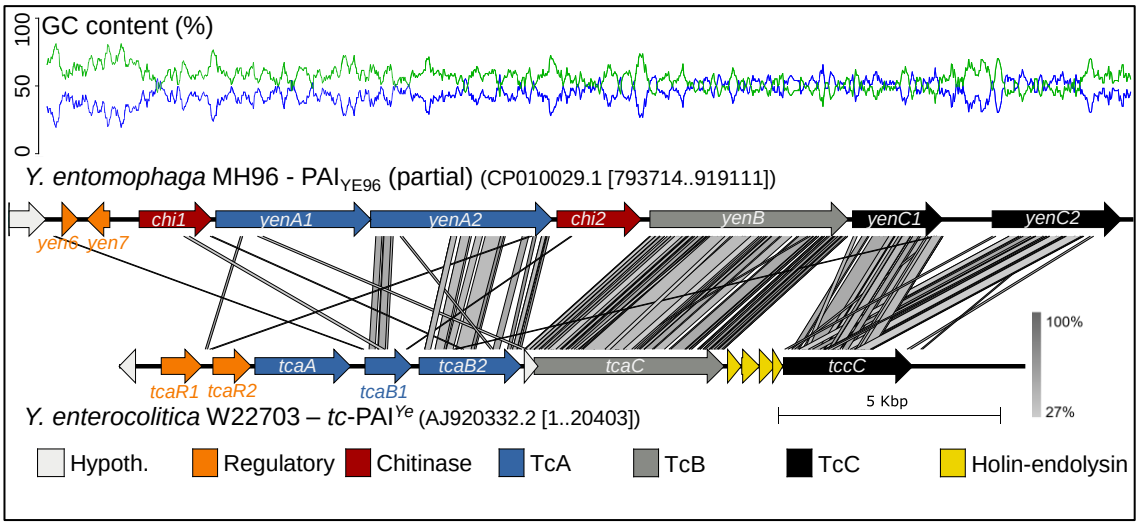


Figure 5.3: Partial PAI$_{YE96}$ in *Yersinia entomophaga* MH96 shares 14 high scoring pairs and 39 % coverage with the translated nucleotide sequence from a *tc*-PAI^{Ye} in *Y. enterocolitica* W22703. Grey lines indicate regions with 27 – 100 % identity between GIs based on tBLASTx searches with bacterial genetic code (translated nucleotide subjects using a translated nucleotide query, all three frames translated for DNA strand). Genomic organization of PAI$_{YE96}$ region encoding the putative temperature-dependent regulators *yen6*, *yen7* and Yen-TC components, *chi1*, *yenA1*, *yenA2*, *chi2*, *yenB*, *yenC1* and *yenC2*. Average AT- vs GC-content is provided (window size 20 bp), Green is AT-content and Blue is GC. Genomic organization of *tc*-PAI^{Ye} region encoding LTTR-like regulators *tcaR1* and *tcaR2*, insecticidal toxin components *tcaA*, *tcaB*, *tcaC*, *tccC1* and endolysin-holin cassette from *Y. enterocolitica* W22703.

Previously reported *in vitro* transcriptomics determined that Yen-TC component and TC-associated genes are readily transcribed at 37 °C *in vitro* (Chapter 4, section 4.3.8 and Figure 4.20), but in contrast the Yen-TC components are not secreted by MH96 when cultured at 37 °C (see Chapter 2, section 2.3.3, Figure 2.3) (Hurst et al., 2011). Also, previous findings from *in vivo* RNA-seq data that revealed genes for other TC-associated factors with the same expression pattern as Yen-TC, including chitin-binding protein, SrfA, SrfB and hypothetical protein (PL78_05310 – PL78_05325), peptidase M66 (PL78_05495), YenC3 and hypothetical (PL78_18780 – PL78_18785), Tc phage-associated (including putative endolysin-holin cassette P78_17400 – PL78_17390), heat-shock protein A, IbpA (PL78_12380), the entire type 2 secretion system (T2SS) (PL78_08975 – PL78_08930) and novel hydrolase cluster (PL78_06535-20) as being significantly repressed at 37 °C in the host compared to 25 °C (Chapter 4, section 4.14.3.8, Table 4.13). These interesting findings beg further investigation into temperature- and host-dependent regulation of Yen-TC at both transcriptional and post-transcriptional levels.

Here we combine targeted bioinformatics and molecular characterization of *yen6* and *yen7*, to gain further understanding of the function of these putative regulators in temperature-dependent control of Yen-TC production in MH96.

The main objective of this work was to:

1. Validate temperature-dependent expression of *yen6*, *yen7* and *chi1* (proxy for Yen-TC) at the translational level; and

2. Determine whether hypothetical regulators Yen6 and Yen7 act as transcriptional regulators with direct involvement in thermoregulation of Yen-TC.

Here the effects of over-expression of *yen6* and *yen7* on global protein secretion, Yen-TC production, general growth and plasmid stabilities were undertaken using crude protein visualization by SDS-PAGE. *LacZ* translational reporter fusions were used to validate temperature-dependent responses of *yen6* and *chi1* at the translational level as well as further interrogate the genetic functions of *yen6* and *yen7* on *chi1* translational reporter. These results contributed to further understanding of *in vitro* and *in vivo* temperature-dependent regulation of Yen-TC by *yen6* and *yen7*.

5.2 Summary methods

General molecular microbiology methods were used as described in Chapter 3, including bacterial strains, plasmids and growth conditions (section 3.1), bacterial enumerations (section 3.2), DNA manipulation (section 3.3), targeted mutagenesis (section 3.4), protein visualization by SDS-PAGE and identification (section 3.5) and *G. mellonella* intrahemocoelic bioassay (section 3.6) unless otherwise stated.

5.2.1 Targeted mutagenesis

For additional details related to targeted mutagenesis please see section targeted mutagenesis (Chapter 3, section 3.2). Targeted mutagenesis constructs were made either using naturally occurring restriction sites found within GOI or using overlap-extension polymerase chain reaction (PCR). The transcriptional regulator *rovA* (PL78_05820) and T6SS component *vipB* (PL78_00910) contained natural restriction sites, *ClaI* and *BgIII*, respectively. Amplicons of approximately 2 kb were generated containing these GOI, which were used as a base for the homologous recombination construct by gene interruption.

Homologous recombination constructs targeting the remaining GOI, *cbpA* (chitin-binding protein PL78_05310), *cspA123* (three tandem cold-shock proteins PL78_18365, PL78_17450 and PL78_18370), *fim1* (usher chaperone fimbrial protein PL78_12480), *yen6* (PL78_03730) and *yen7* (PL78_03735) were generated using overlap-extension PCR with primers CBP_P2/CBP_P5, CSP_P2/CSP_P5, UCF_P2/UCF_P5, Yen6_P2/Yen6_P5 or Yen7_P5 and Yen7_P2/Yen7_P5, respectively (supplementary Table S5).

5.2.2 Phenotyping targeted mutants by SDS-PAGE and protein identification by LC-ESI-MS/MS

For additional details related to protein visualization and identification see (section 3.5). To enable visualization of crude protein extracts, wild-type and mutant strains were grown in 3 ml broth for approximately eight hours with appropriate antibiotics. Flasks containing 50 ml nutrient broth were seeded with 2 % starter and grown at either 25 or 37 °C with 200 rpm shaking overnight (~ 18 hours) with antibiotic. The following morning, 1 ml culture were pelleted by centrifugation for 10 minutes at 8,000 x g and the cell supernatant was filter sterilized (0. 22 μm). Serial dilutions of the overnight culture were made in phosphate buffer solution (PBS) (Sigma) and colony forming units (CFU) were enumerated to confirm consistent cell density among the overnight cultures.

Protein visualization by SDS-PAGE was performed according to (Laemmli, 1970) on one-dimensional slab gels (1.0 mm thick) containing 0.1 % SDS. Mid-range proteins

were separated by size on hand-cast or Novex pre-cast tris-glycine with 10 % polyacrylamide under reducing conditions with 200 V for 55-60 minutes in SDS-PAGE running Buffer. Precision Plus™ Protein standard (Bio-Rad) was used as a markerUnstained marker. Smaller sized denatured proteins were mobilized on Novex pre-cast 10 % tricine (Life Technologies) under reducing conditions (by applying 125 V for 75 - 90 min in Novex 1 X tricine running buffer. Separated proteins were stained with silver as described by (Blum, H., Beier, H., Gross, 1987). Specific unknown protein samples for liquid chromatography electrospray ionization ion trap-tandem mass spectrometry (LC-ESI-MS/MS) analysis were separated by SDS-PAGE as outlined above in Chapter 3 (section 3.5.2).

5.2.3 Phenotyping targeted mutants by bioassay

For details on virulence testing of MH96 and mutatgenic strains using *Galleria mellonella* intrahemocoelic infection model (see section 3.6). Larvae were injected with 10 µl inoculum containing strain dilutions (ranging from ~ 0 – 10 cells) just below their third right leg with a 30-gauge needle on a 1 mL tuberculin syringe (BD) using a microinjector. Ten larvae were injected per dilution/strain and incubated at 25 or 37 °C. The bioassay was repeated three times per strain and temperature. To determine the LD_{50}, day five mortality rates were fit with binomial-type generalized linear model (includes the logit link function) and the LD_{50} and its standard error were estimated using the R package MASS (Crawley, 2013). Data were plotted using ggplot2 in R.

5.2.4 Yen6 and Yen7 protein sequence homology searches and protein model predictions.

The region of PAI_{YE96} encoding putative regulators *yen6* and *yen7* was interrogated with a variety of bioinformatics tools to assign additional annotations or genetic features to the region. The amino acid sequence of Yen6 and Yen7 were compared to the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein sequence database (May 3, 2019) using BLASTp, with default settings (Camacho et al., 2009). Amino acid sequence of hypothetical proteins sharing significant sequence homology

to Yen6 or Yen7 were aligned using clustalW (Thompson et al., 1994) with Mega X (Kumar et al., 2018). Evolutionary histories of Yen6 or Yen7 homologous hypothetical proteins were inferred by phylogenetic analysis using the Maximum Likelihood method and Le and Gascuel model (Le and Gascuel, 2008) with discrete Gamma distribution site substitution model and tested with 1,000 replicate bootstraps (Felsenstein, 1985) using Mega X. The amino acid sequences of Yen6 and Yen7 were investigated using the online database Protein Homology/Analogy Recognition Engine V 2.0 (Phyre^2) (Kelley, Lawrence et al., 2015)(accessed: May 11, 2018 and April 20, 2018) to predict protein structure and identify structural homologies to solved crystal protein structures using normal modelling mode. The predicted protein structures were visualized using the FirstGlance in Jmol viewer (v 2.74) (Hanson et al., 2013). Terminators were annotated by GAMOLA2 (Altermann et al., 2017) using TransTermHP (version 2.09) (Ermolaeva et al., 2000). The nucleotide sequence of the entire upstream region of *yen6* until the start codon of Chi1 (2,108 bp) was assessed for repeat motifs using the Repfind program (accessed: May 22, 2018) (Betley et al., 2002) with (P-value cut off = 0.001, minimum repeat length = 9, maximum repeat length = 14, with low complexity sequence filtered. Secondary RNA structures of *yen6* messenger RNA (mRNA) variants were modelled in Geneious (v10.0.9) using the tool Vienna RNAfold tool (Gruber et al., 2008) using the constraint generation RNA free energy parameter estimation method (Andronescu et al., 2007) and default settings.

5.2.5 Arabinose induction of *yen6*, *yen67as* and *yen7*

5.2.5.1 Cloning into arabinose induction vector

The arabinose induction pBAD18 derivative pAY2-4 vector (Shaw et al., 2003) (therein simply referred to as "pBAD") was used to characterize changes in growth and toxin production/secretion during over-production of *yen6*, *yen67as* and *yen7* in MH96 (supplementary Table S4). All pBAD vectors encode the regulatory gene *araC*, which under the presence of arabinose, will promote transcription of genes fused to the downstream P_{BAD} promoter (Guzman et al., 1995). The ORF regions of *yen6* (including 32 bp of 3'UTR), *yen67as* (including 477 bp of 3'UTR encoding a predicted terminator

and asRNA that overlaps with the ORF of *yen7* on the opposite strand) and *yen7* (including 31 bp of 3'UTR) were amplified using PCR with Platinum *Taq* DNA polymerase (Invitrogen) as described in Chapter 3 (section 3.3.2) using forward primer yen6_expF and reverse primers yen6_expR and yen67as_expR, respectively (supplemental Table S13, Figure S21). The restriction sites, *NdeI* and *EcoRI* were incorporated into the forward and reverse primers for restriction cloning of the PCR product, respectively. Two additional *yen6* expression variants were also generated using the aforementioned yen6_expF with Amb2 and Amb3 reverse primers, to generate a *yen6* variant with 3'UTR lengths of 67 and 148 bp, with the later variant containing the predicted terminator.

Amplicons were purified using the High Pure PCR Product Purification Kit (Roche). Digestion with *NdeI* and *EcoR1* (New England Biolabs (NEB)) and ligation using T4 DNA Ligase (Invitrogen) was carried out as previously described (section 3.3.7). The overnight ligation reaction was then precipitated in 100 % and further digested with *StuI* (NEB) to enrich for insertion of the PCR product into the vector. This step was determined as critical for successful recovery of correct clones (pBAD lacks a blue/white detection with *lacZ*). The enriched ligation was then transformed into DH10B by chemical transformation and clones were selected on ampicillin 100 μg/ml.

Insert validation was achieved by first screening ampicillin resistant colonies by PCR using DreamTaq Green PCR Master Mix (Thermo Scientific) and the original insert primers. Plasmids were isolated by minipreparation from PCR-positive clones and digested using *NdeI* and *EcoRI* to confirm insertion. Finally, plasmids showing a correct restriction digestion profile were isolated using the High Pure Plasmid Isolation Kit (Roche) and shipped to Macrogen Korea for sequencing with primers araF (5'-TCCATAAGATTAGCGGATCCTAC-3') and araR (5'-CATGGGGTCAGGTGGCAC-3') (supplementary Table S13). Sequence data were validated by visual inspection for PCR errors and correct in-frame fusion of GOI by alignment with the MH96 reference genome in Geneious.

Electrocompetent MH96 cells were generated as previously described for *E. coli* (Section 3.3.9). Validated pBAD-*yen6*, pBAD*yen67as* and pBAD-*yen7* induction vectors

were transformed into MH96 and derivative strains by electroporation and selected on ampicillin 400 μg/ml.

5.2.5.2 Over-expression of *yen6*, *yen67as* and *yen7* in MH96 and derivative strains

MH96 and derivative strains were grown in 3 ml Luria Broth Base (LB broth) (Invitrogen) at 30 °C with 250 rpm shaking for approximately eight hours with ampicillin μg/ml. Flasks containing 50 ml media were seeded with 1 % starter and grown at either 25 or 37 °C with 200 rpm shaking overnight (~ 16 hours) with carbenicillin 400 μg/ml (earlier work found that ampicillin did not produce enough selective pressure to maintain the plasmid in some growth conditions). For most arabinose induction experiments, arabinose was added immediately following the starter culture to a final concentration of 0.2 %, but sometimes final arabinose concentrations 0.02 or 0.002 % were used. In some specific *yen7* overexpresion experiments, the addition of 0.002 % arabinose was delayed for two hours after adding the inoculumn.

After 16 h, samples from overnight 50 ml cultures were serially diluted and plated for enumeration (see Chapter 3, section 3.2) on media supplemented with and without ampicillin 400 μg/ml. Plasmid containing cells were determed by their ability to form colonies on ampicillin supplement plates and plasmid stability was calculated as the percentage of plasmid-containing cells in the culture from non-replicated plates. Due to plating variability, estimated plasmid stabilities greater than 100 % were determined.

Crude protein samples were prepared from 1 ml of culture by centrifugation for 10 min at 8,000 x g to pellet the cells. Cell supernatant was filter sterilized (0. 22 μm) and the pellet resuspended in sterile H_2O. Protein extracts from cell supernatant (CS) and pellet (CP) were visualized by SDS-PAGE as described above (section 3.5.1) on either hand-cast or Novex Tris-Glycine pre-cast (Invitrogen) gel containing 10 % polyacrylamide with silver stain.

5.2.6 Translational P_{yen6}::, P_{yen7}:: and P_{chi1}::*lacZ* reporters

5.2.6.1 Transcriptional organization of *yen6*, *yen7* and *chi1*

Before construction of *lacZ*-promoter fusions for *yen6*, *yen7* and *chi1*, the transcriptional organization of the specific region was explored using reverse transcription PCR (RT-PCR). The RT-PCR reactions were performed on cDNA generated using SuperScript IV RT (Invitrogen) according to the manufacturer's guidelines with Random and Oligo (dT) Primers (Promega). The RNA used in the RT reactions was originally used for RNA-seq and collected from MH96 grown *in vitro* at 25 °C until mid-exponential growth (~ 10^8 CFU/ml) and was previously shown to be free of DNA (Chapter 4, section 4.3.2). A final volume of 13 µl, containing 1 µl random primers (0.5 µg/µl), 4 µl dNTPs (10 mM each), 5 µl RNA (DNAse treated) and 3 µl DEPC-treated H_2O was incubated at 65 °C for five minutes and then placed on ice for one minute. A RT reaction mix, containing 4 µl 5X SSIV buffer, 1 µl DTT (100mM), 1 µl RNAseOUT RNAse Inhibitor (40 U/µl) and 1 µl SuperScript IV Reverse Transcriptase (200 U/µl) were first mixed and then added to the reaction. The mixture was incubated at 23 °C, 55 °C and 80 °C for ten minutes each and then stored at -20 °C.

To assess the transcriptional organization of the *yen6*, *yen7* and *chi1* region, PCR reactions were carried out using the cDNA (1 µl), gDNA (0.5 µl) and RNA (0.5 µl) with primers provided in supplementary Table S14 and Platinum Taq DNA Polymerase as described before. These PCR reactions were performed using the following program: 95 °C for 3 min; then 95 °C for 25 s, 55 °C for 25 s, 72 °C for 1 min repeated 30 times; then 72 °C for 10 min. PCR products were visualized on 1 % agarose gel by electrophoresis on a UV transilluminator with red safe DNA dye.

5.2.6.2 *LacZ* reporter *cis*-merodiploid MH96 derivative strains

The promoter regions for *yen6* (1,006 bp), *yen7* (873 bp) and *chi1* (834 bp) were fused in-frame to promoterless *lacZY* cassette in the suicide plasmid pVIK107 (Kalogeraki and Winans, 1997). Primers, plasmids and strains used in this study are provided in supplementary Table S2, Table S4 and Table S13 and supplementary Figure S22). The pVIK107 plasmid contains a R6K vegetative origin but lacks the R6K *pir* gene (like the suicide plasmids used for targeted mutagenesis, see Chapter 3, section 3.3.8), so cannot

replicate in strains that do not have *pir*. The promoter regions were amplified using Phusion High-Fidelity DNA polymerase (NEB) following manufacturer recommendations. The PCR products were purified using the High Pure PCR Product Purification Kit (Roche) and then cloned into pGEM-T vector for sequence validation (see Chapter 3, section 3.3.5 above).

Once the cloned sequences were validated as correct, these promoter regions were restriction cloned into pVIK107 (kan^r) or pVIK107 ($chlor^r$) using *EcoR1* and *SalI* (NEB). The ligated DNA was purified by ethanol precipitation; briefly the DNA mixed with 2 volumes 100 % ethanol and 1:10 volume 3 M sodium acetate (pH 5.2) and incubated for 2 minutes at room temperature (~ 22 °C). Next the DNA was pelleted by centrifugation for 5 min at 15,700 x g, the supernatant was carefully removed, and the pellet was dried at room temperature 20 minutes and then resuspended in MilliQ water. The ligated DNA was further enriched by digestion with *XbaI* (NEB) and this enriched ligated DNA was transformed into EC100D (pir^+) by electroporation and clones were screened on either kanamycin 100 μg/ml or chloramphenicol 30 μg/ml (depending on whether the kan^r or $chlor^r$ pVIK107 plasmid variant was used) and X-gal 100 μg/ml for blue/white selection. White clones were then screened for sensitivity to ampicillin 100 μg/ml to ensure loss of the pGEM-T vector. Clones were validated by plasmid isolation, restriction enzyme digestion with EcoRI and *SalI*, and subsequent visualization of the correct sized inserts when by agarose gel by electrophoresis. Correct plasmids were then transformed into the conjugative *E. coli* strain, ST18 by chemical transformation as previously outlined (Chapter 3, Section 3.3.10).

By single-step homologous recombination, the reporter *lacZ*-fusions were integrated into the chromosomes of recipient strains by conjugation method as described for targeted mutagenesis (see Chapter 3, section 3.2). Putative *cis*-merodiploid MH96 derivative strains were screened on either kanamycin 100 μg/ml or chloramphenicol 90 μg/ml. Correct integration of the translational fusion within natural chromosomal locations was confirmed by PCR using validation primers (supplemental Table S13), extending beyond the point of recombination and into the *lacZ* gene. Mutant strain validation was carried out as described above in Chapter 3, section 3.4 and strains used in this study are provided in supplementary Table S2.

Preliminary attempts were taken to *trans*-complement the P_{chi1}::lacZ *cis*-merodiploid reporter Δ*yen6* and Δ*yen67* strains using the mid-copy vector pACYC184 Table S4. Briefly, the entire region containing the promoter and coding regions of *yen6* to *yen7* (2,761 bp) were amplified by PCR and ligated by restriction at the *BamHI and SphI* sites within the tetracycline cassette using primer pairs 184_Y67F/184_Y76R (Table S13 and Figure S22).

5.2.6.3 Calibration curve

MH96 was grown overnight in 3 ml LB broth at 30 °C with 250 rpm shaking. Next, three 50 ml flasks per temperature (25 and 37 °C) were each inoculated with 1 % starter and grown until optical density at 600 nm (OD_{600}) of approximately 1.0, from which a range of dilutions with OD_{600} ranging from 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9 were made with PBS from each culture flask and measured on the SPECTROstar^{Nano} (BMG Labtech). Absorbance data were normalized against blank using the MARS Data Analysis software (BMG Labtech). Subsequently, the cell density of each of the dilutions from each of the culture flasks were further serially diluted and plated for CFU enumeration. The experiment was repeated three times and the combined mean cell density (CFU/ml) was plotted against mean OD_{600} to determine the calibration curve for MH96 grown at 25 and 37 °C.

5.2.6.4 β-galactosidase assay *in vitro*

In vitro cultures for β-galactosidase (β-gal) assays were grown overnight (16 hours) in 50 ml LB at either 25 or 37 °C with 200 rpm shaking. Flasks were seeded with 1 % overnight culture and grown with appropriate antibiotics to maintain any plasmids. Every *lacZ*-fusion strain was grown in triplicate, and a negative control strain not expected to produce any lacZ was included with each experiment. Each experiment was repeated three times. To mitigate inherent variability associated with sampling from clumpy cell culture, 1 ml samples were collected into 1.5 mL microcentrifuge tubes and gently vortexed before 200 µl of culture were collected into a sterile 96-well plate (F-bottom) (Greiner Bio-One Cellstar). Also, 500 µl were diluted 1/10 in PBS,

vortexed gently and then 200 µl of the dilution was collected. The 96-well plate was frozen at -80 °C.

A single-step method for β-galactosidase assay using a 96-well microplate reader was carried out as described in Schaefer et al., 2016a & 2016b, with minor modifications.

Modified β-gal mix used in this study were made up in distilled water:

- 1x Z-buffer pH 7.0 (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$);
- β-mercaptoethanol (1.8 µl/ml);
- Lysozyme from chicken egg white (≥ 40,000 units/mg lyo.; Sigma) (0.2 mg/ml);
- Bacterial Protein Extraction Reagent (B-PER) (Thermo Fisher Scientific) (1/150 dilution); and
- 2-nitrophenyl-β—galactopyranoside (ONPG) (Sigma) (1 mg/ml).

Samples were thawed at 37 °C for 20 - 25 minutes (until no more ice crystals remained). Absorbance of the undiluted and 10^{-1} cell samples was measured on a SPECTROstar^{Nano} (BMG Labtech) at 600 nm. Next 80 µl of 10^{-1} cell sample was added to a Nunc F96 MicroWell plate (Thermo scientific) assay plate to which 120 µl of β-gal mix were added. Absorbance at 420 nm were immediately measured on a pre-warmed (30 °C) plate-reader. Readings were taken every 60 s for 1 h (20 flashes per well each cycle) and the plate was shaken at 500 rpm (double orbital shaking) between each reading.

Using the MARS Data Analysis software (BMG Labtech), the slope of blank-normalized absorbance values (OD_{420}/min) was determined using ranges of linear increase of β-gal only (measurements taken after maximum OD_{420} absorbance were not included). The blank corrected absorbance at OD_{600} of the 10^{-1} diluted cell sample, and the blank corrected slope of β-gal activities were expressed in Miller Units (MU) using the following equation (Schaefer et al., 2016b):

$$MU = \frac{1000 * (OD_{420}/min)}{10^{-1} \text{ culture } OD_{600} * \text{volume used (ml)}}$$

Statistical significance between average enzyme activities for P_{yen6}::lacZ and P_{chi1}::lacZ in wild-type MH96 at 25 and 37 °C were tested using Welch's t-test in R. Statistical significance between average enzyme activities for *P_{chi1}::lacZ* in various genetic backgrounds and under over-production of Yen6 and Yen67as were assessed, first by fitting an analysis of variance model to the data and then by applying post-hoc Tukey's Honest Significant Difference test in R.

5.2.7 β-galactosidase assay *in vivo*

Cultures for *in vivo* β-gal assays were grown in 3 ml LB broth for 7 – 8 hours at 30 °C with 250 rpm shaking. Cultures were serially diluted in PBS and 10 µl containing and between 260 – 740 cells were injected into the hemocoel of *G. mellonella* as described in section 3.6. Hosts were either incubated at 25 or 37 °C and hemolymph were collected at specific post-infection time points to try and normalize the cell density between the two temperatures (the cells grow faster at higher temperatures). Subsequently, hemolymph from hosts incubated at 25 and 37 °C were collected after 19 – 22 h and 16 – 17 h, respectively. Three individual hosts from each temperature/experiment were homogenized and used to make enumeration plates to estimate cell density (method previously described in Chapter 4, section 4.2.1.2) and the experiment was repeated three times. The first time was with a fresh batch of *G. mellonella* that were used within one week upon arrival and the second and third experiment were carried out with a less fresh batch of *G. mellonella*, which required at least three weeks on artificial diet before reaching the final L6-stage (this is the stage used for experiments).

Hemolymph used for β-gal assays were collected following the same protocol as the *in vivo* RNA samples (Chapter 4, section 4.2.1.2) with minor modifications. Briefly, approximately 50 µl of hemolymph was collected by pipette from a puncture of the dorsal cuticle and immediately placed into 200 µl pre-chilled PBS containing crystals of phenylthiocarbamide (1-Phenyl-2-thiourea; Sigma) to prevent melanization, which was

found to interfere with OD_{600} readings during preliminary experiments. The hemocytes were pelleted by centrifugation at 500 x g for five minutes at 4 °C. Next, 200 µl of the supernatant was collected into a pre-chilled 96-well plate and stored at -80 °C.

The *in vivo* β-gal assays were completed using the same single-step method used for *in vitro* samples described above in section 5.2.6.4, except the undiluted OD_{600} was used for the MU calculation, but the enzymatic assay was conducted on 1:10 samples diluted in PBS. The third experiment was omitted from the results due to highly variability of enzymatic activity observed between individual hemolymph samples at both temperatures.

5.3 Results

5.3.1 Targeted mutant phenotyping

Based on previously reported RNA-seq analysis (Chapter 4, sections 4.3.7.2 -Figure 4.17; section 4.3.8 - Figure 4.20) investigating global transcriptome responses by MH96 to *in vivo* conditions within the model host *G. mellonella* and different temperatures (25 and 37 °C), several GOIs representing putative VFs were identified and targeted for mutagenesis (see Table 5.1). These included: *cbpA* (PL78_05310) that encodes a secreted lytic chitin monooxygenase (LCMO)/chitin-binding protein (CBP); *cspA123* (PL78_18365, PL78_17450 and PL78_18370) encoding three triplicate cold-shock proteins (CSPs); *fim1* (PL78_12480) that encodes an usher chaperone fimbrial protein; *rovA* (PL78_05820) that encodes putative H-NS counter-silencer; *vipB* (PL78_00910) that encodes T6SS sheath protein; *yen6* (PL78_03730) that encodes putative LytTR-containing DNA-binding protein; and *yen7* (PL78_03735) that encodes putative winged-helix-turned-helix (wHTH) DNA-binding protein. Phenotypic investigations were carried out using protein visualization and virulence testing in the *G. mellonella* intrahemocoelic model to determine whether mutagenesis resulted in significant effects at the protein level or attenuated virulence of MH96.

5.3.1.1 Targeted mutant phenotyping by protein analysis

Denatured proteins from cell pellets (CP) and supernatant (CS) from MH96 and derivative mutant strains (Δ*cbpA*, Δ*cspA123* (subsequently renamed Δ*cspA123*/

ΔHCUI$_{YE96}$ due to the presence of 17.5 kb excision of HCUI$_{YE96}$), Δ*fim1*, Δ*rovA*, Δ*vipB*, Δ*yen6*, *Δyen7* and Δ*yen67*) were visualized by SDS-PAGE, targeting mid-large molecular weight proteins on 10 % glycine gel (Figure 5.4, Figure 5.5, Figure 5.6 and Figure 5.7). The protein profiles were examined for phenotypic variation in the mutant strains compared to wild-type, with an emphasis on production and secretion of large molecular weight Yen-TC components YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi1 (70kDa) and Chi2 (60 kDa) (MH96 is known to produce and secrete large amounts of Yen-TC components at temperatures of 25°C or lower). A phenotype was identified for Δ*cspA123*/ΔHCUI$_{YE96}$, which did not produce Yen-TC toxin components (Figure 5.4) and showed greatly reduced secretion (Figure 5.5) when grown at 25 °C. The protein profiles of the CP produced by Δ*cspA123*/ΔHCUI$_{YE96}$ at 25 °C was a near exact match to the CP profile of MH96 when cultured at 37 °C, which was found to lack Yen-TC components (YenB, YenA2, YenA1) as well as four other previously identified proteins: peptidase M66 (81 kDa), a hypothetical (44 kDa), chitinase (33 kDa) and CbpA (23 kDa) (Figure 5.4 and Figure 5.5). It was difficult to determine if less YenC1, YenC2, Chi2 and Chi1 were produced by Δ*cspA123*/ΔHCUI$_{YE96}$ at 25 °C due to the presence of similar sized proteins in the CP, however protein bands located at 70 and 60 kDa (corresponding to Chi2 and Chi1) were less prominent in Δ*cspA123*/ΔHCUI$_{YE96}$ compared to wild-type 25 °C.

Notably, the CS produced by Δ*cspA123*/ΔHCUI$_{YE96}$ when cultured at 37 °C only contained a single high molecular weight protein (~175 kDa) (this protein was also produced by wild-type at 37 °C). The LCMS-ESI-MS/MS analysis identified 51 peptides, covering 48 % of the filamentous hemagglutinin N-terminal containing protein (PL78_04365) with a predicted false positive rate of 8.6E-40 (i.e., chance that false identification would achieve the same or matching score) (Zhang et al., 2012) (supplemental Table S12 and Figure S16). This was the same filamentous hemagglutinin N-terminal containing protein identified as highly expressed *in vivo* at 37 °C compared to 25 °C and this high molecular weight protein is known to be secreted by wild-type at 37 °C as well.

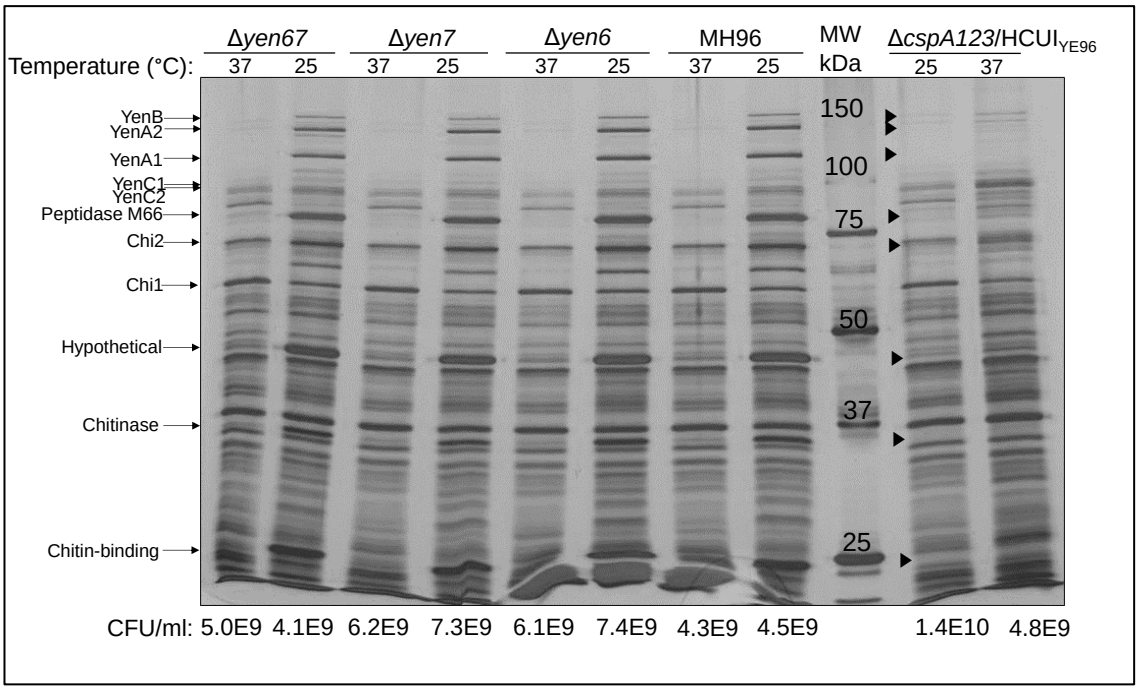


Figure 5.4: Crude protein extracts from cell pellet of overnight cultures of *Yersinia entomophaga* MH96 and Δ*yen67*, Δ*yen6* and Δ*yen7* strains grown for 18 hours with 200 rpm shaking at either 25 or 37 °C. Proteins are visualized on 10 % glycine gel under denaturing conditions and stained with silver. MW = Precision Plus Protein™ Unstained Protein Standard (Bio-Rad). Δ*yen67*= deletion mutant lacking *yen6 – yen7* region, Δ*yen7* = deletion mutant lacking *yen7*, Δ*yen6* = deletion mutant lacking *yen6*, Δ*cspA123*/ΔHCUI$_{YE96}$ = deletion mutant lacking cold-shock protein triplication and HCUI$_{YE96}$ excision. Arrows denote Yen-TC components: YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), Yen C1 (109 kDa), YenC3 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa) and peptidase M66 (81 kDa), hypothetical (44 kDa), chitinase (33 kDa) and chitin-binding proteins (23 kDa). Black triangles denote Yen-TC components and other secreted proteins missing from Δ*cspA123*/ΔHCUI$_{YE96}$.

The proteins present in the CS from the remaining derivative strains (i.e., Δ*cspA123*, Δ*vipB*, Δ*cbpA*, Δ*rovA* and Δ*fim1*) grown at 25 °C were compared to wild-type on denaturing 10 % glycine gel to identify potential secretion phenotypes (Figure 5.6). A distinct protein band of 23 kDa was absent from Δ*cbpA* and presumed to be the previously identified 23 kDa LCMO/CBP, CpbA. The only other phenotype identified using this method, was a slight variation in Yen-TC production by Δ*rovA*, which secreted a protein associated with the same size as YenA1 (130 kDa) but did not secrete as the predicted 156 kDa YenA2 (Figure 5.6).

To explore potential phenotypic variation of smaller molecular weight proteins, CP from MH96, Δ*cspA123*/ΔHCUI$_{YE96}$, Δ*vipB*, Δ*cbpA*, Δ*fim1* and Δ*rovA* were visualized on 10 % tricine gel under denaturing conditions (Figure 5.7). Like the glycine gels, the

profile of the CP from Δ*cspA123*/ΔHCUI$_{YE96}$ grown at 25 °C was found to lack Yen-TC components and other dominant bands that would normally be produced in wild-type. The tricine gel was found to be a better method to visualize the CbpA (estimated ~ 23 kDa), which was clearly missing from the CP produced by Δ*cbpA*. CpbA was also missing from the CP of MH96 when grown at 37 °C and Δ*cspA123*/ΔHCUI$_{YE96}$ when grown at both 25 and 37 °C.

One key result from the tricine gel was the appearance of a noticeable band around 8 kDa in size, which appeared only in the CP of Δ*cspA123*/ΔHCUI$_{YE96}$ when grown at 37 °C (Figure 5.7B). Using LCMS-ESI-MS/MS this band was found to contain five unique peptides, covering 31 % of the 50S ribosomal protein L6 (RplF, PL78_11255) with a predicted false positive rate of 1.4E-08 (i.e., chance that false identification would achieve the same or matching score) (Zhang et al., 2012) (supplementary Figure S17 and Table S12)

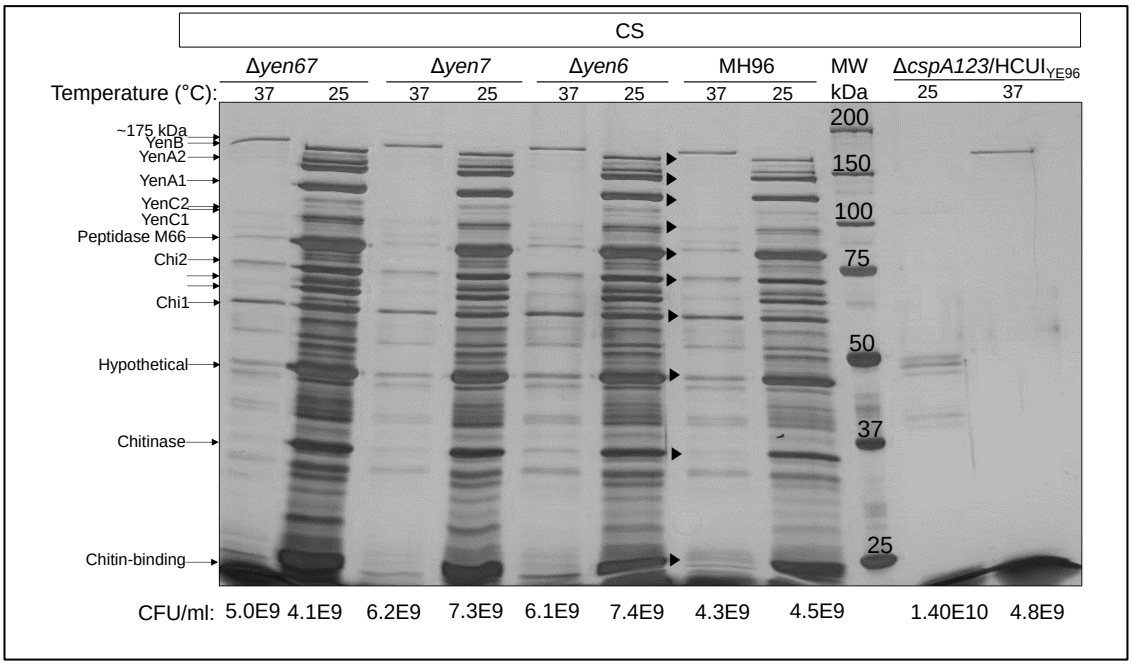


Figure 5.5: Crude protein extracts from cell supernatant of overnight cultures of *Yersinia entomophaga* MH96 and Δ*yen67,* Δ*yen6* and Δ*yen7* strains grown for 18 hours with 200 rpm shaking at either 25 or 37 °C. Proteins are visualized on 10 % glycine gel under denaturing conditions and stained with silver. MW = Precision Plus Protein™ Unstained Protein Standard (Bio-Rad). Δ*yen67*= deletion mutant lacking *yen6 – yen7* region, Δ*yen7* = deletion mutant lacking *yen7*, Δ*yen6* = deletion mutant lacking *yen6*, Δ*cspA123*/ΔHCUI$_{YE96}$ = deletion mutant lacking cold-shock protein triplication and HCUI$_{YE96}$ excision. Arrows denote Yen-TC components: YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), Yen C1 (109 kDa), YenC3 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa) and peptidase M66 (81 kDa), hypothetical (44 kDa), chitinase (33 kDa) and chitin-binding proteins (23 kDa). Black triangles denote Yen-TC components, peptidase M66, hypothetical protein, chitinase and chitin-binding missing from the CS of MH96 cultured at 37 °C.

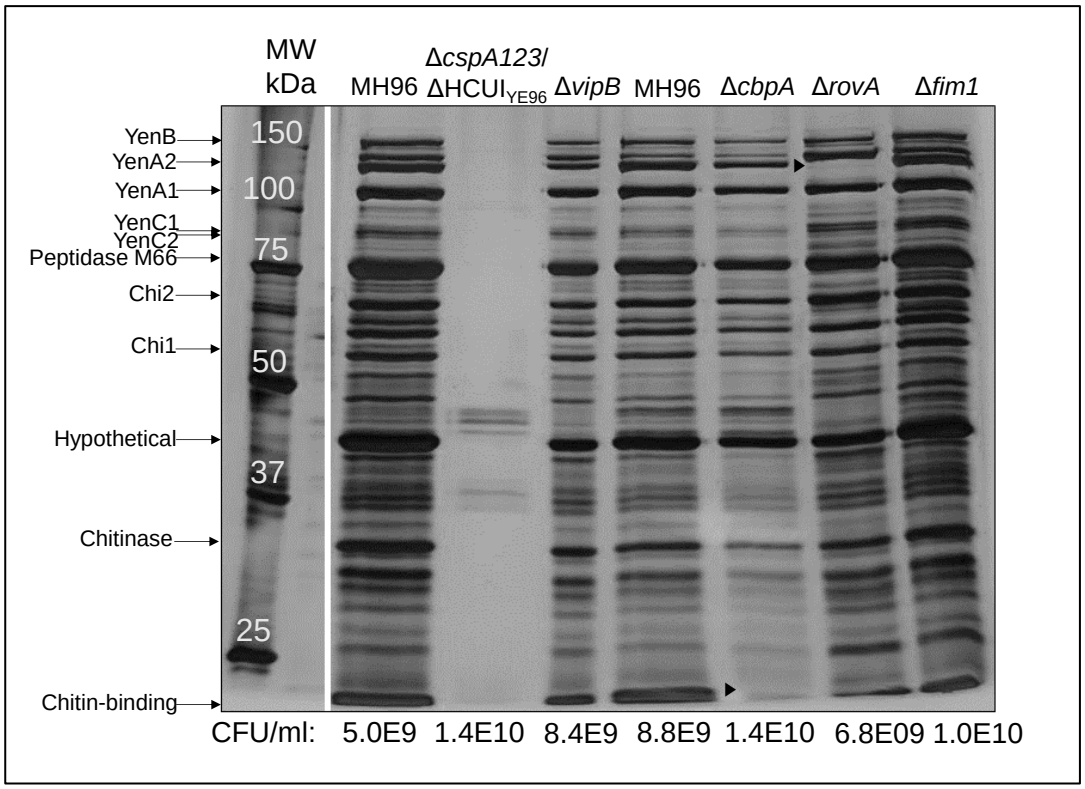


Figure 5.6: Crude protein extracts from cell supernatant of overnight cultures of *Yersinia entomophaga* MH96 and Δ*cspA123*/ΔHCUI$_{YE96}$, Δ*vipB*, Δ*cbpA*, Δ*rovA* and Δ*fim1* strains grown for 18 hours with 200 rpm shaking at 25 °C. Proteins are visualized on 10 % glycine gel under denaturing conditions and stained with silver. MW = Precision Plus Protein Unstained Protein Standard (Biorad). Δ*cspA123*/ΔHCUI$_{YE96}$ = deletion mutant lacking cold-shock protein triplication and HCUI$_{YE96}$ excision, Δ*vipB* = T6SS *vipB* sheath protein deletion mutant, Δ*cbpA*= deletion mutant lacking chitin-binding protein, Δ*fim1* = deletion mutant lacking usher/chaperone fimb1 protein, Δ*rovA* = deletion mutant lacking transcriptional regulator *rovA*. Arrows denote Yen-TC components: YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), Yen C1 (109 kDa), YenC3 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa) and peptidase M66 (81 kDa), hypothetical (44 kDa), chitinase (33 kDa) and chitin-binding proteins (23 kDa). Black triangles denote proteins missing in the Δ*cbpA* and Δ*rovA* mutants.

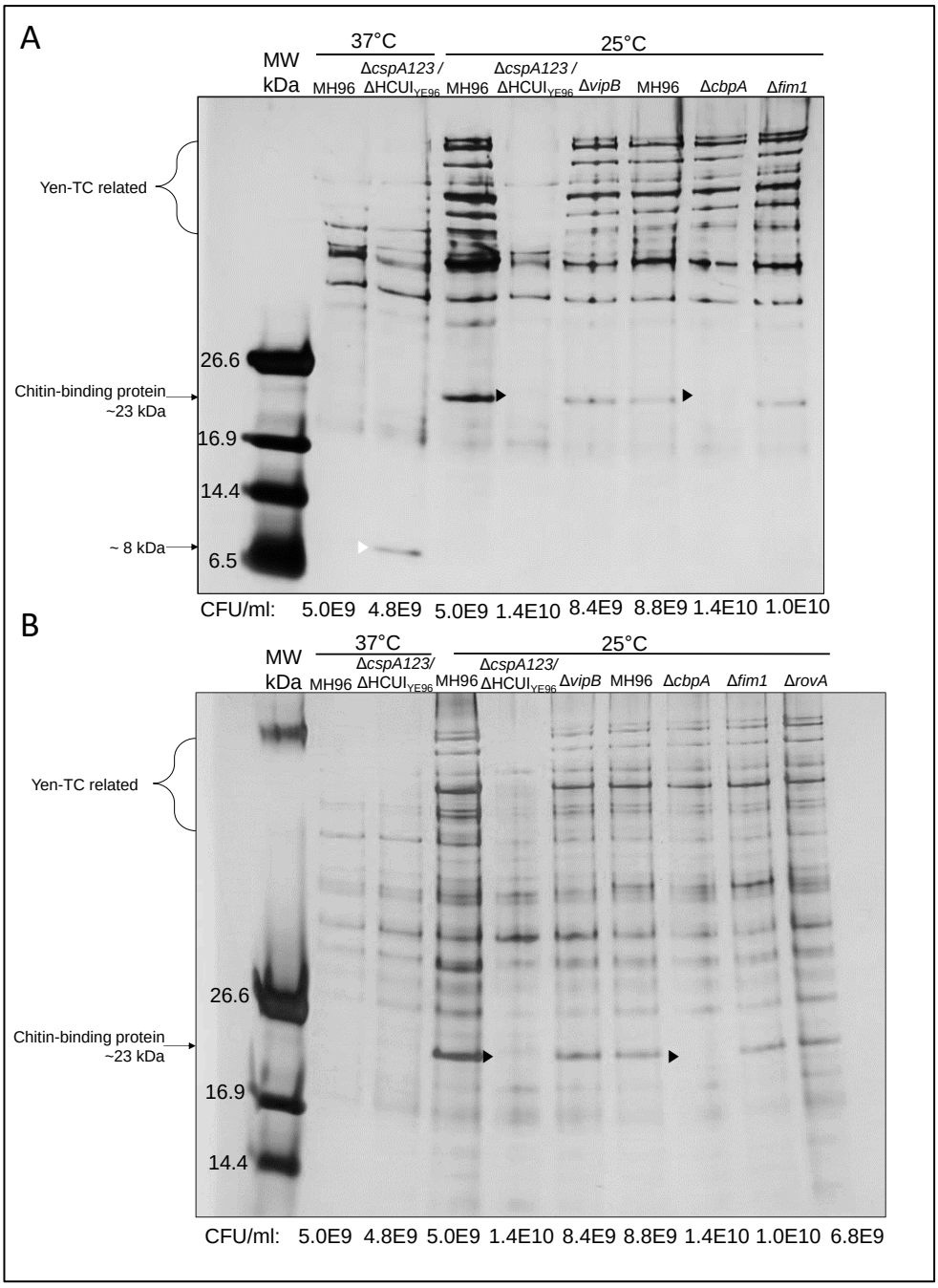


Figure 5.7: Crude protein extracts from cell pellet of overnight cultures of *Yersinia entomophaga* MH96 and Δ*cspA123*/ΔHCUI$_{YE96}$, Δ*vipB*, Δ*cbpA*, Δ*rovA* and Δ*fim1* strains grown for 18 hours with 200 rpm shaking at either 25 or 37 °C. Proteins are visualized on 10 % tricine gel under denaturing conditions and stained with silver. A. Gel run 125V for 90 minutes. B. Same samples as A, but 125 V applied for 75 minutes. MW = Ultra low-range molecular-weight marker (Sigma). Δ*cspA123*/ΔHCUI$_{YE96}$ = deletion mutant lacking cold-shock protein triplication and HCUI$_{YE96}$ excision, Δ*vipB* = T6SS *vipB* sheath protein deletion mutant, Δ*cbpA* = chitin-binding protein deletion mutant, Δ*fim1* = usher/chaperone fimb1 protein deletion mutant and Δ*rovA*= *rovA* deletion mutant (gel B only). Arrows denote chitin-binding protein (23 kDa). Black triangles denote chitin-binding protein (23 kDa) missing from Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*cbpA* mutant and white triangle denotes unknown protein of ~ 8 kDa present only in Δ*cspA123*/ΔHCUI$_{YE96}$ grown at 37 °C.

5.3.1.2 Virulence testing – intrahemocoelic bioassay *Galleria mellonella*

In addition to protein visualization by SDS-PAGE as a means of phenotyping, wild-type MH96 and each of the deletion mutant strains (except Δ*yen6*, Δ*yen7*, Δ*yen67*) were assessed for attenuated virulence using the model host *G. mellonella* at 25 °C. Since, previous reports demonstrated that a mutant deficient for Yen-TC (ΔYen-TC) showed no reduced virulence against *G. mellonella* when directly injected (Hurst et al., 2015), so deletion of *yen7* (a putative transcriptional regulator of Yen-TC) was thought to be unlikely to have reduced virulence in this model. Furthermore, we did not observe any obvious decreases in Yen-TC production or secretion within Δ*yen6*, Δ*yen67* or Δ*yen7* protein profiles visualized by SDS-PAGE (Figure 5.4 and Figure 5.5), which further reduced the chance of observing attenuated virulence in these strains in this model due to continued high levels of Yen-TC secretion at 25 °C. To this end, Δ*yen6* was only assessed in this model at 37 °C and not at 25 °C because *yen6* was found to be most highly expressed only at higher temperatures *in vivo* (Chapter 4, section 4.3.8, Figure 4.20), so it was assumed to be functional under these specific conditions. Unlike Δ*yen6*, Δ*cspA123*/ΔHCUI$_{YE96}$ was assessed at both temperatures because the expression of *cspA1*, *cspA2* and *cspA3* was shown to be significantly higher during early infection at 25 and also at 37 °C *in vivo* and may play a role in stress response, defense against host or virulence regulation during infection at either temperature.

Assessment of wild-type MH96 found that the number of cells required to kill 50 % of hosts in four days was estimated to be less than one cell at 25 and 37 °C, respectively (Table 5.2). Since it is not possible to have < 1 cell, these results were interpreted such that only a single cell is likely to kill 100 % of hosts within four days. None of the deletion mutants were found to have attenuated virulence compared to wild-type at 25 °C, based on LD_{50} estimates (all < 1 cell). However, at 37 °C two strains: Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* had a 3-fold reduction in virulence compared to wild-type, albeit not statistically significant (Table 5.2, Figure 5.8). Also, 100 % mortality rates were not observed with either Δ*cspA123*/ΔHCUI$_{YE96}$ or Δ*yen6* at 37 °C, even with the lowest dilution dose (Figure S19 and Figure S20).

Table 5.2: Logistic regression and estimation of median lethal dose (LD_{50}) from bioassay of *Yersinia entomophaga* MH96 and derivative strains with intrahemocoelic injection of larval *Galleria mellonella* after four days. Each bioassay included ten hosts and was repeated three times. SE = estimated standard error

Strain	Intercept	Intercept SE	Slope	Slope SE	LD_{50} (Cells)	LD_{50} SE
*Incubation temperature:*25 °C						
MH96	0.34	0.24	1.11	0.20	0.7	1.2
Δ*cbpA*	0.71	0.26	1.53	0.23	0.6	1.2
Δ*cspA123*/ΔHCUI$_{YE96}$	0.35	0.17	0.64	0.15	0.6	1.4
Δ*fim1*	1.05	0.22	0.81	0.19	0.3	1.3
Δ*rovA*	1.38	0.28	1.36	0.23	0.4	1.2
Δ*vipB*	2.62	0.50	1.34	0.32	0.1	1.3
Incubation temperature: 37 °C						
MH96	0.14	0.19	0.91	0.17	0.9	1.2
Δ*cspA123*/ΔHCUI$_{YE96}$	-0.49	0.13	0.48	0.10	2.8	1.2
Δ*yen6*	-0.83	0.21	0.84	0.20	2.7	1.4

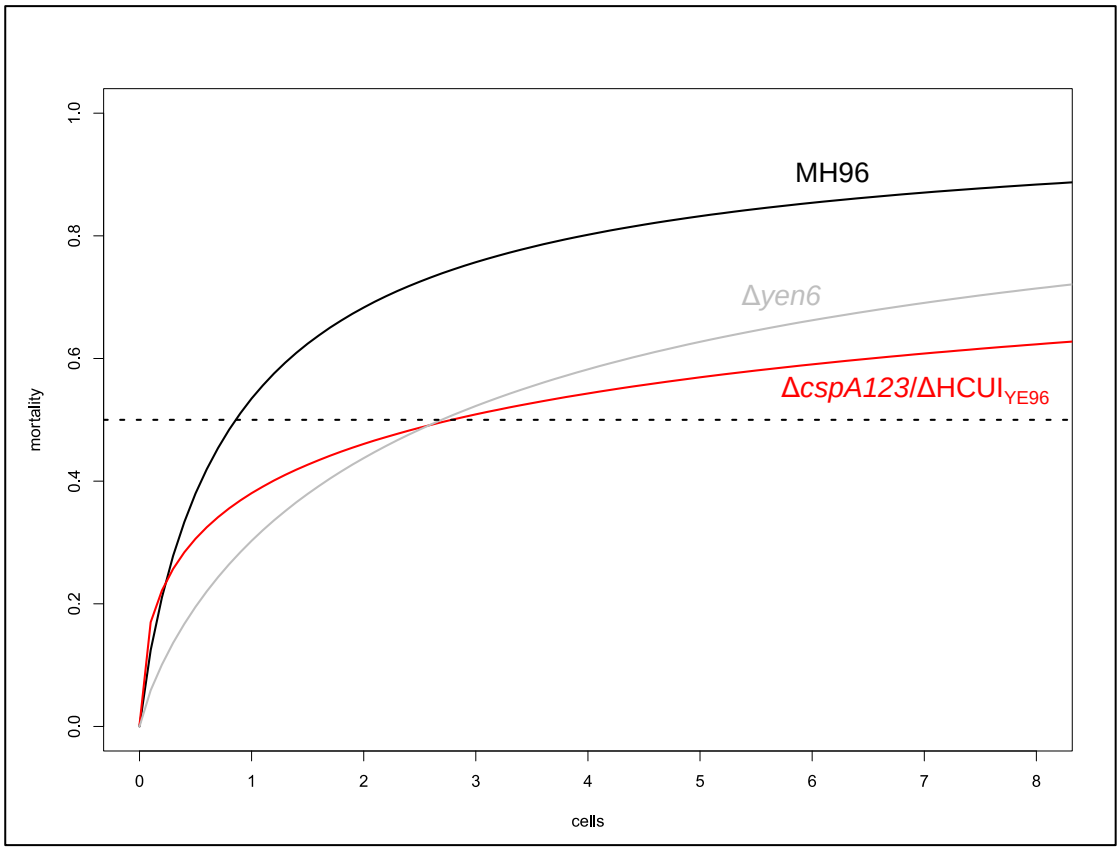


Figure 5.8: Dose response fitted generalized linear model for *Yersinia entomophaga* MH96 and Δ*yen6* and Δ*cspA123*/ΔHCUI$_{YE96}$ strain using intrahemocoelic infection of *Galleria mellonella* incubated at 37 °C for four days. The dashed line represents day five 50 % mortality rate used to estimate the median lethal dose (LD$_{50}$) value for each strain. Δ*cspA123*/ΔHCUI$_{YE96}$ = deletion mutant lacking cold-shock protein triplication and HCUI$_{YE96}$ excision, Δ*yen6* = deletion mutant lacking *yen6*. Each bioassay included 10 larva per strain and was repeated three times.

5.3.2 Focused investigations of putative regulators *yen6* and *yen7*

5.3.2.1 Secondary protein structure prediction and phylogenetics of Yen6 and Yen7

Protein modelling and homology searches using the Phyre^2 database identified high-confidence matches for the predicted protein structures of both Yen6 and Yen7, which suggested putative roles as DNA-interacting regulatory proteins. By re-constructing the evolutionary histories of Yen6 and Yen7 using phylogenetic analysis of homologous protein sequences from GenBank, the prevalence of these putative regulators within pathogenic lineages among Gammaproteobacteria was revealed. These focused investigations were helpful to further predict putative functions for

Yen6 and Yen7, especially the former which is currently annotated as a hypothetical protein.

The highest confidence match for predicted secondary structure of Yen6 using Phyre[2] was the DNA-binding domain of the toxin regulator AgrA from *Staphylococcus* (*St.*) *aureus* (c3bs1A, Sidote et al., 2008) (77 % of sequence modelled with 95.8 % confidence) (Figure 5.9). The predicted model for Yen6 was found to be very close in structure to the LytTR-domain AgrA model, consisting of eight β-strands (the template had ten β-strands) arranged into three β-sheets characteristic for LytTR containing proteins (Figure 5.9 and Figure 5.10). Compared to the AgrA model, the amino acid sequence of Yen6 included an insertion relative to the template (residues 71 – 79 of Yen6) potentially giving rise to a novel loop-turn region. A high degree of residue conservation was found among Yen6 and homologous proteins from GenBank within some of the predicted β-sheets regions (specifically β3-β8), but the predicted DNA-interacting residues were not found to be highly conserved among taxa (Figure 5.10).

Using the structure of the AgrA LytTR-binding domain as a guide (specifically DNA-interacting residues: H169, N201 and R233) (Sidote et al., 2008), potential corresponding DNA-interacting residues (N34, R67 and E108) from Yen6 were mapped onto predicted loop-turn structures in the 3D model (Note: E108 was selected over P109, which is comparatively more hydrophobic, Figure 5.9). Like the DNA-binding region of AgrA, all the predicted DNA-bind residues were found to both share the same face of the predicted protein and have a downward orientation. The predicted novel loop region unique to Yen6, was found to also face a similar orientation as the other loops containing DNA-interacting residues, suggesting Yen6 may possess and additional DNA-interacting loop region. A potential DNA-interacting residue on the novel predicted loop is asparagine (N81) which is 100 % conserved among the taxa (Figure 5.10).

Figure 5.9: Predicted secondary structure model for Yen6 from *Yersinia entomophaga* MH96 based on crystal structure of LytTR domain of AgrA from *S. aureus* (c3bs1A) with Phyre2 (confidence 95.8 %/coverage: 77 %). Visualized in JSmol interactive viewer in secondary structure view with putative DNA-interacting residues highlighted with yellow circles (N34, R67 and E108) and novel predicted loop region indicated by orange triangle. Yellow ribbon arrows = β-strands, pink helices = α-helices, purple helices = 3_{10} turn helices, blue coil = turns and while coil = unpredicted structure.

Hypothetical proteins with high sequence homology (E-value of less than or equal to 1e-25) to Yen6 were found among pathogenic species of Gammaproteobacteria, including diverse representatives among Enterobacteriales and Pasteurellales (Figure 5.11). All the other taxa, except for *Y. entomophaga* and closely related *Y. nurmii*, are known to be associated with or pathogens of humans and mammals. For example, *Pasteurella oralis* has been isolated from the oral cavities of animals including cats and dogs but is not known to be pathogenic (Christensen et al., 2012), but *Histophilus somni* (or *Haemophilus somnus*) is a pathogen of cattle, sheep and bison (Angen et al., 2003). Some well-known human pathogens were also found to encode a Yen6 homolog, including *Escherichia coli*, *Klebsiella pneumoniae* and *Sa. enterica*. The phylogenetic reconstruction was not able to robustly resolve the evolutionary history of Yen6 and Yen6-homologs because the percentage of trees in which the associated taxa clustered among boot-straps was low in some cases, including separation of *Sa. enterica* from *K. pneumoniae* and *E. coli* (only 48 % of trees) or separation of *Yersinia* spp. and *Actinobacillus* spp. (only 54 % of trees).

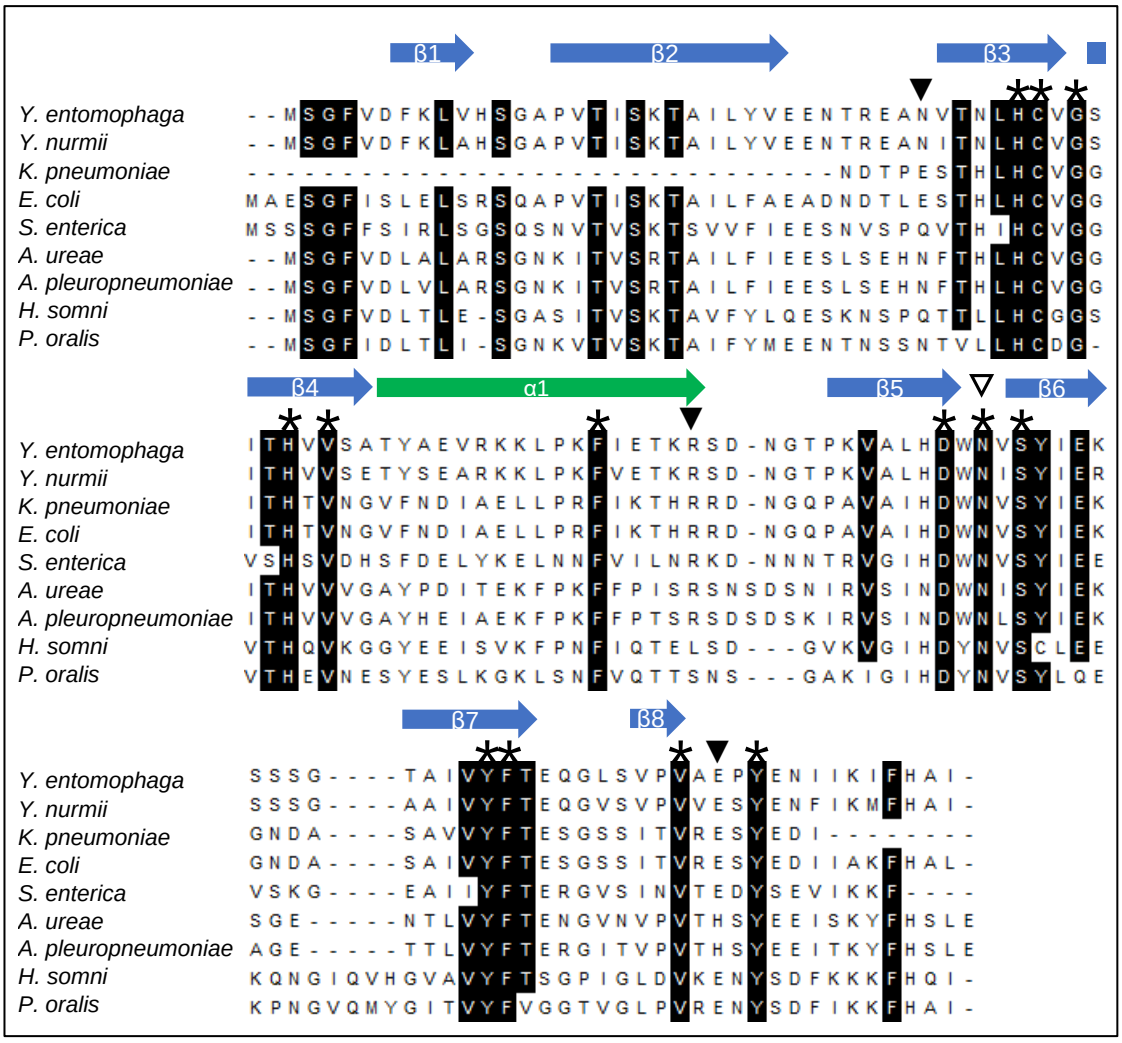


Figure 5.10: ClustalW amino acid alignment of putative LytTR-domain containing regulator, Yen6 of *Yersinia entomophaga* MH96 and homologous hypothetical protein sequences sharing significant similarity (E-value of less than or equal to 1e-25). Shaded loci share 80 % or greater homology. Asterix (*) indicate loci with 100 % conservation across taxa. Black triangle (▼) indicate putative DNA-interacting residues (N34, R67 and E108) based on protein secondary structure predictions from the structure of AgrA from *St. aureus*. White triangle (▽) indicate hypothetical DNA-interacting residue (N81) located on novel predicted loop region. Predicted β-sheets and α-helices are indicated by blue and green arrows, respectively. Alignments were conducted using MEGA X. Sequence accessions are provided in Figure 5.11, below.

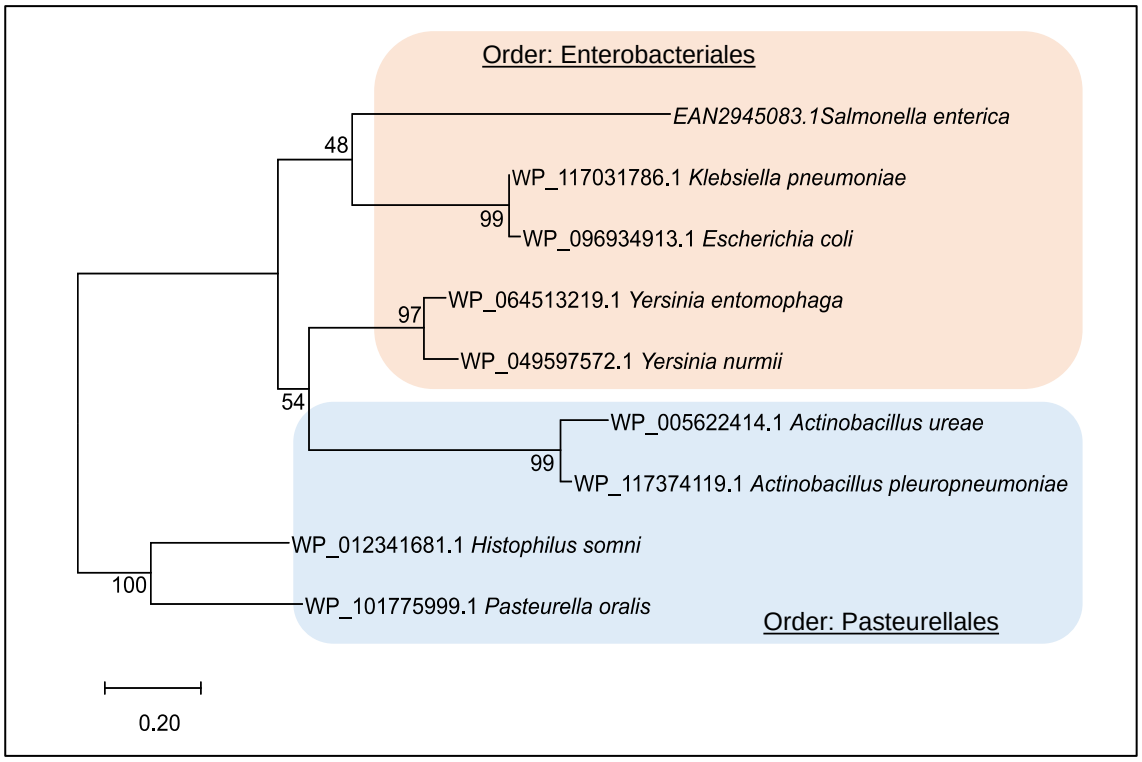


Figure 5.11: The evolutionary history of putative LytTR-containing regulator Yen6 of *Yersinia entomophaga* MH96 was inferred by using the Maximum Likelihood method and Le and Gascuel model. Taxa from GenBank nr protein database encoding hypothetical proteins with significant sequence similarity to Yen6 (E-value threshold 1e-25) were identified using BLASTp. The most homologous sequence from each species was aligned to Yen6 using ClustalW. The most appropriate phylogenetic model was determined as the model with the lowest Bayesian Information Criterion (best description of the substitution pattern) among 56 different amino acid substitution models. The bootstrap consensus tree inferred from 1,000 replicates is taken to represent the evolutionary history of the taxa analyzed. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.2861)). Sequence alignment and evolutionary analyses were conducted in MEGA X.

Using Phyre[2] the top protein model match for Yen7 was limited to the DNA-binding domain of PhoP response regulated from *Mycobacterium tuberculosis* (c2rv8A, Wang et al., 2007) (67 % of sequence modelled with 99.9 % confidence). The predicted structure of Yen7 consists of a typical wHTH motif, which is often found among DNA-interacting domain of two-component response regulators (Figure 5.12). The predicted structure of Yen7 is expected to begin with three stranded anti-parallel b-sheet, followed a bundle of three α-helices, and an C-terminal β-hairpin that forms a 3

stranded anti-parallel β-sheet with the short β-strand located between the first and second α-helices (Figure 5.12 and Figure 5.13).

Figure 5.12: Predicted secondary structure model for Yen7 from *Y. entomophaga* based on crystal structure of the DNA-binding domain of PhoP from *Mycobacterium tuberculosis* (c2rv8A) with Phyre2 (confidence 99.9 %/coverage: 67 %). Visualized in JSmol interactive viewer in secondary structure. Yellow ribbon arrows = β-strands, pink helices = α-helices, purple helices = 3_{10} turn helices, blue coil = turns and while coil = unpredicted structure.

Based on the crystal structure of the *M. tuberculosis* PhoP DNA-binding domain, DNA-interacting residues within wHTH motifs are primarily located within the recognition helix (α8) (Wang et al., 2007). The recognition helix of Yen7 is α3, which was found to contain several putative DNA-interacting residues (Figure 5.13 and Figure 5.14). Furthermore, a high degree of sequences conservation was noted within the recognition helix as well as the β3 and α2 regions (Figure 5.13).

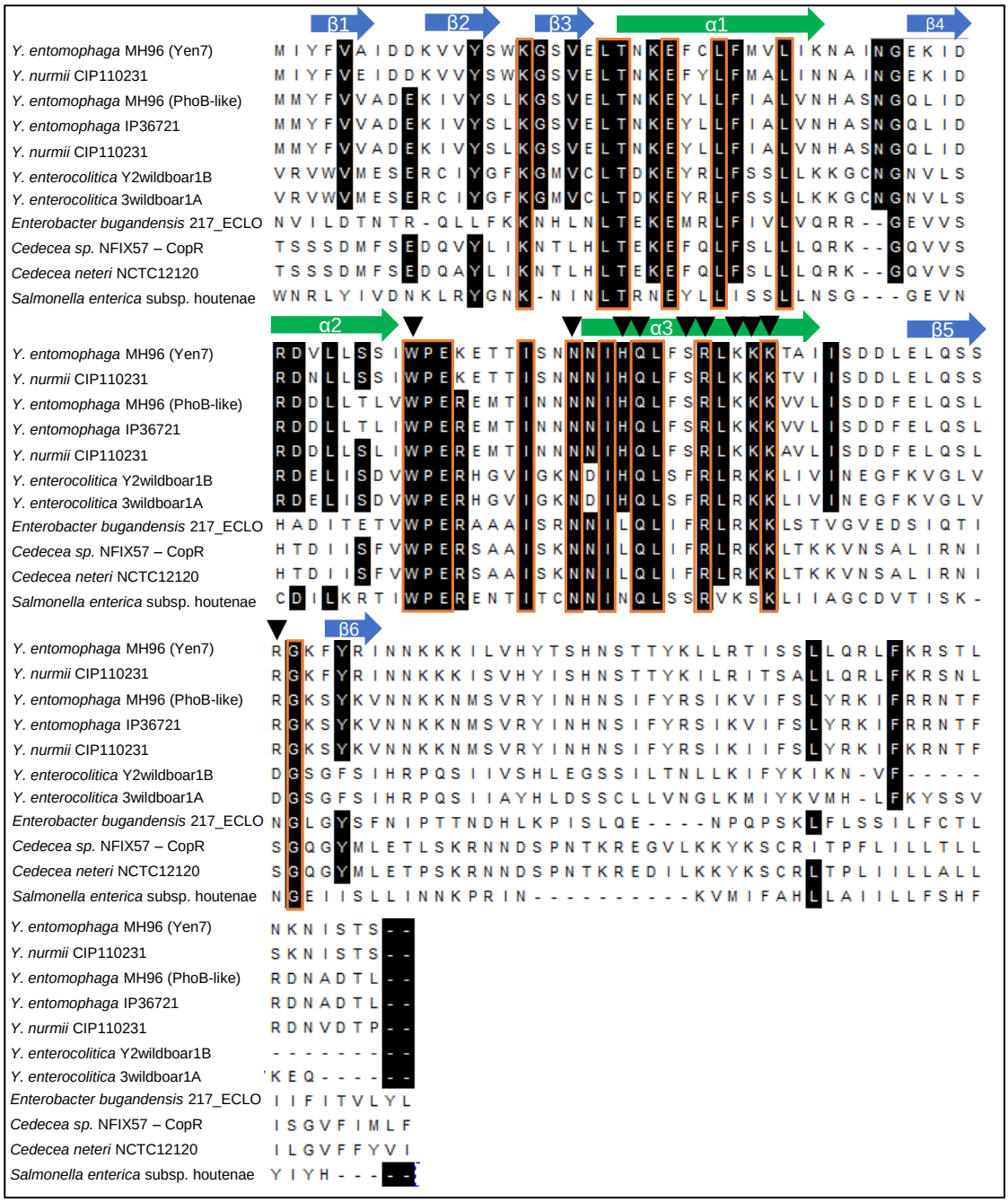


Figure 5.13: ClustalW amino acid alignment of putative wHTH-domain containing regulator, Yen7 of *Yersinia entomophaga* MH96 and homologous hypothetical protein sequences sharing significant similarity(E-value of less than or equal to 7e-7) with. Shaded loci share 60 % or greater homology. Orange boxes indicate loci with 100 % conservation across taxa. Black triangle (▼) indicate putative DNA-interacting residues (W52, N62, H65, Q66, S69, R70, K72, K73, K74, R88) based on protein secondary structure predictions from the structure of PhoP from M. tuberculosis. Predicted β-sheets and α-helices are indicated by blue and green arrows, respectively. Alignments were conducted using MEGA X. Sequence accessions are provided in phylogeny (Figure 5.15).

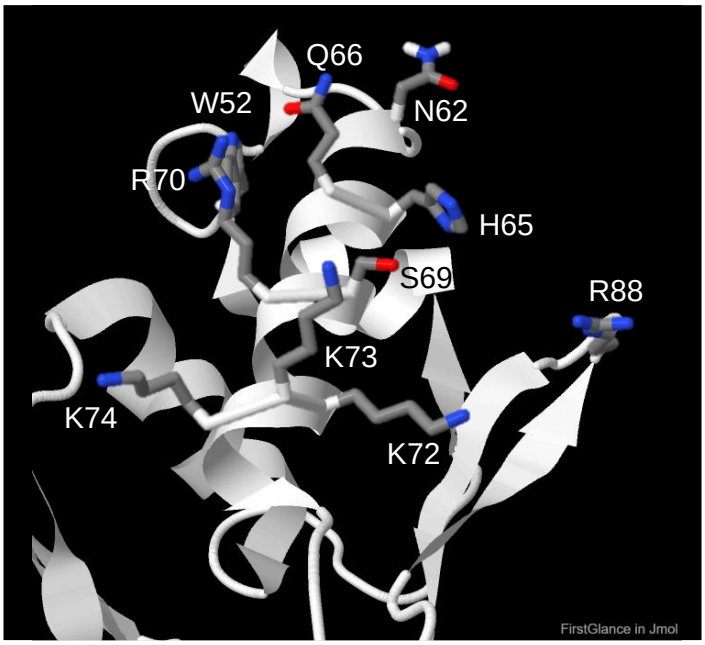


Figure 5.14: Predicted secondary structure of Yen7 from *Yersinia entomophaga* MH96 based on crystal structure of the DNA-binding domain of PhoP from *Mycobacterium tuberculosis* (c2rv8A) using Phyre[2] (confidence 99.9 %/coverage: 67 %). Visualized in JSmol interactive viewer in cartoon view with specific focus on the third or 'recognition' α-helix. Amino acid residue side chains arising from the Yen7 recognition helix (N62 – K75) that are predicted to interact with DNA are superimposed on the protein model.

Phylogenetic reconstruction of Yen7 and other homologous proteins with significant sequence similarities (E-value less than or equal to 7e-7) revealed the presence of a putative paralog, PhoB-like protein among *Y. entomophaga* and closely related *Y. nurmii* (Figure 5.15). The PhoB-like protein (PL78_17385) is encoded on the phage-related region and was shown to share a similar expression profile with Yen-TC and other Yen-TC-associated factors (Chapter 4, section 4.3.8, Table 4.13). Yen7 orthologues were also identified in *Y. enterocolitica* from wild boar and fallow deer carcasses (Macori et al., 2017), as well as other species known to be animal pathogens. There is some evidence that the Yen7 orthologue (WP_007175603.1) encoded by *Y. enterocolitca* strains isolated from fallow deer and wild boar carcasses may act as a insecticidal TC regulator because it is located directly upstream of three putative toxins (WP_077175604.1 - hypothetical protein, WP_077175605.1 - toxin, and WP_076706524.1 - virulence protein), which share significant homology and conserved domains with insecticidal TCs produced by other pathogenic bacteria including *Bacillus cereus*, *B. thuringiensis*, *Paenibacillus alvei*, *P. luminescens* and *Sa. enterica.*

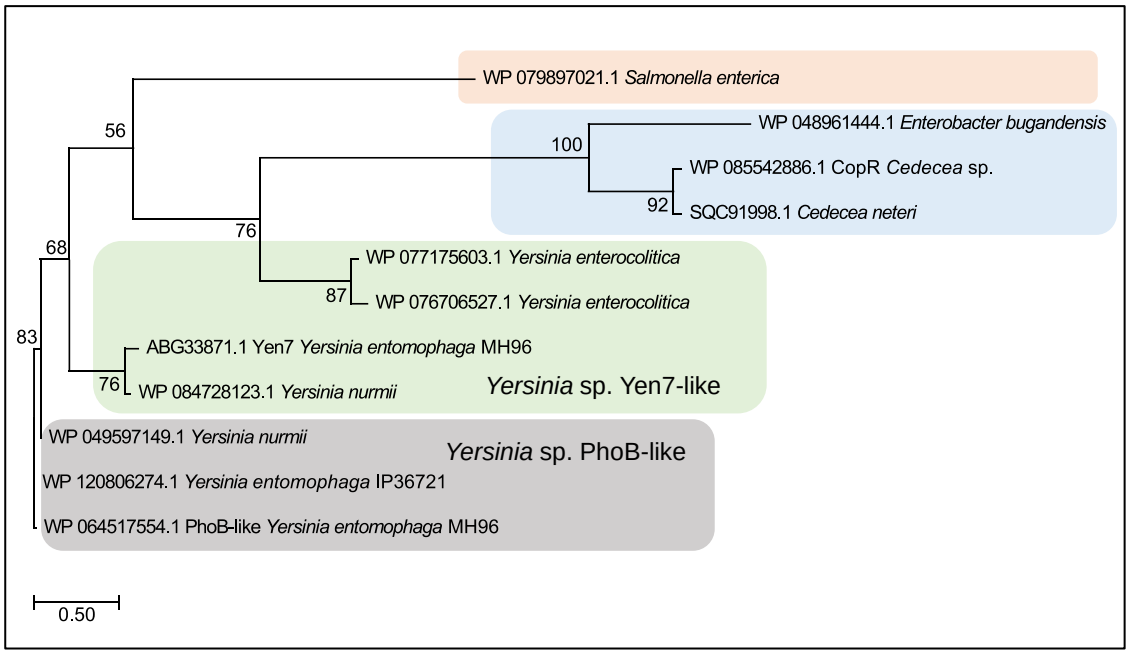


Figure 5.15: The evolutionary history of putative wHTH-containing regulator Yen7 of *Yersinia entomophaga* MH96 was inferred by using the Maximum Likelihood method and Le and Gascuel model. Taxa from GenBank nr protein database encoding hypothetical proteins with significant sequence similarity to Yen6 (E-value threshold 7e-7) were identified using BLASTp. Homologous sequences were aligned using ClustalW and trimmed as required. The most appropriate phylogenetic model was determined as the model with the lowest Bayesian Information Criterion (best description of the substitution pattern) among 56 different amino acid substitution models. The bootstrap consensus tree inferred from 1,000 replicates is taken to represent the evolutionary history of the taxa analyzed. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.2861)). Sequence alignment and evolutionary analyses were conducted in MEGA X.

5.3.2.2 Predicted non-coding RNA, terminators and repeat motifs within PAI_{YE96} regions encoding *yen6*, *yen7* and Yen-TC components.

The presence of 5′ and 3′ UTRs, non-coding RNA (ncRNA) and asRNA within the region of PAI_{YE96} encoding *yen6, yen7* and the Yen-TC components were predicted from transcriptome data using Rockhopper as described in Chapter 4 (section 4.2.4). To some extent, the transcription organization was validated using RT-PCR (described below in section 5.3.3.4, Figure 5.23). A double hairpin terminator structure was predicted within the long 3′UTR of Yen6, which also was predicted to partially overlap

the ORF of *yen7* found on the reverse strand. Many very small ncRNA and asRNA (size range 26 – 406 bp) were predicted among the Yen-TC components (Figure 5.16).

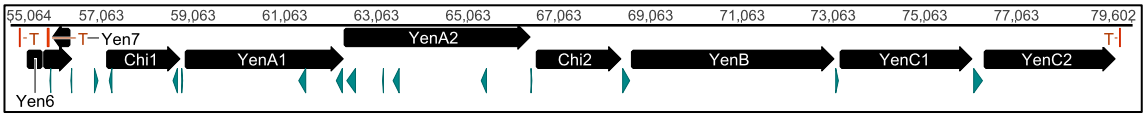


Figure 5.16: Terminators, non-coding and anti-sense RNA predicted within *Yersinia entomophaga* MH96 PAI$_{YE96}$ genomic region encoding putative regulators *yen6* and *yen7* and Yen-TC components . A long 3'UTR (646 bp) was predicted for *yen6* which overlaps the *yen7* ORF. A 5'UTR (166 bp) was predicted for *chi1*. "T"/orange rectangles = terminators and turquoise triangles represent non-coding and anti-sense RNAs.

The nucleotide sequence of the entire upstream region of *yen6* until the start codon of *chi1* (2,108 bp) was assessed for repeat motifs using the Repfind program (Betley et al., 2002) with (P-value cut off = 0.002, minimum repeat length = 9, maximum repeat length = 14, with low complexity sequence filtered). One perfect repeat sequence of 9 nt (5'-GATATATTT-3') separated by a 32 nt spacer was identified as overlapping the putative 35-box of *yen7* with P-value < 0.001(Figure 5.17). Another 9 nt (5'-AGCGACGAT-3') and 10 nt (5'-AAAAAGAAAA) perfect repeat pairs were also identified within the protein coding region of *yen7*, with P-value < 0.001 and 0.002, respectively (Figure 5.17).

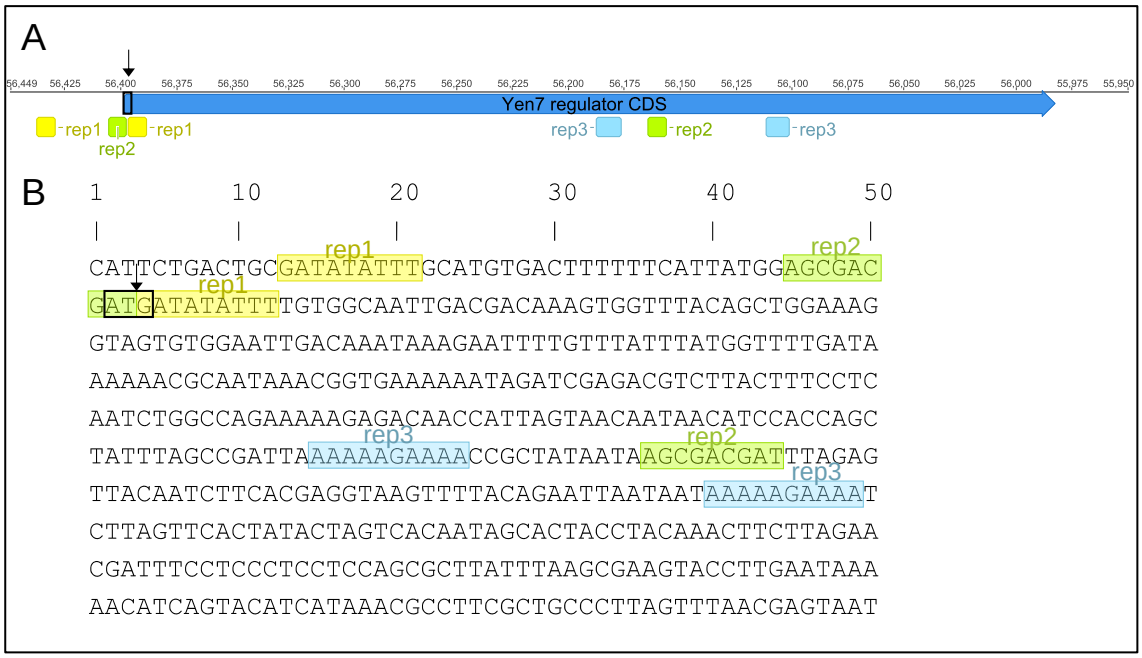


Figure 5.17: Overlapping perfect repeat motifs identified from *yen7* loci including upstream promoter region. A. Location of three perfect repeat pairs (rep1, rep2 and rep3) identified within PAI$_{YE96}$ of *Yersinia entomophaga* (MH96) using the Repfind program (Betley et al., 2002) (P-value cut off = 0.002, minimum repeat length = 9, maximum repeat length = 14, with low complexity sequence filtered). B. Genetic sequence with start codon of *yen7* (black box and down arrow) and putative transcriptional regulator binding sites identified as perfect repeat pairs highlighted in yellow (rep1), green (rep2) and blue (rep3).

5.3.2.3 *yen6* and *yen67as* predicted RNA structures

The predicted stabilities of RNA species of *yen6* messengers with a 32 and 478 bp 3'UTR region were modeled in Geneious using (v.10.0.9) DNA/RNA secondary structure fold viewer with the Vienna RNAfold tool (Gruber et al., 2008) using the constraint generation RNA free energy parameter estimation method (Andronescu et al., 2007) and default settings (Figure 5.18). While the *yen6* messenger species with the shorter, 32 bp 3'UTR was found to be somewhat stable (free energy of ensemble at 37 °C of -98.56 kcal/mol) , the predicted secondary structure of the *yen6* messenger containing the longer, 478bp 3' UTR, including the predicted double hairpin and a portion of the Yen7as (649 bp), was found to be relatively more stable (free energy of ensemble at 37 °C of -232.48 kcal/mol) (Table 5.3). Remarkably, the Yen7as region was predicted to have a high degree of complementarity to the 5' region of the *yen6* messenger, and the double hairpin also formed a very stable secondary structure in the predicted model.

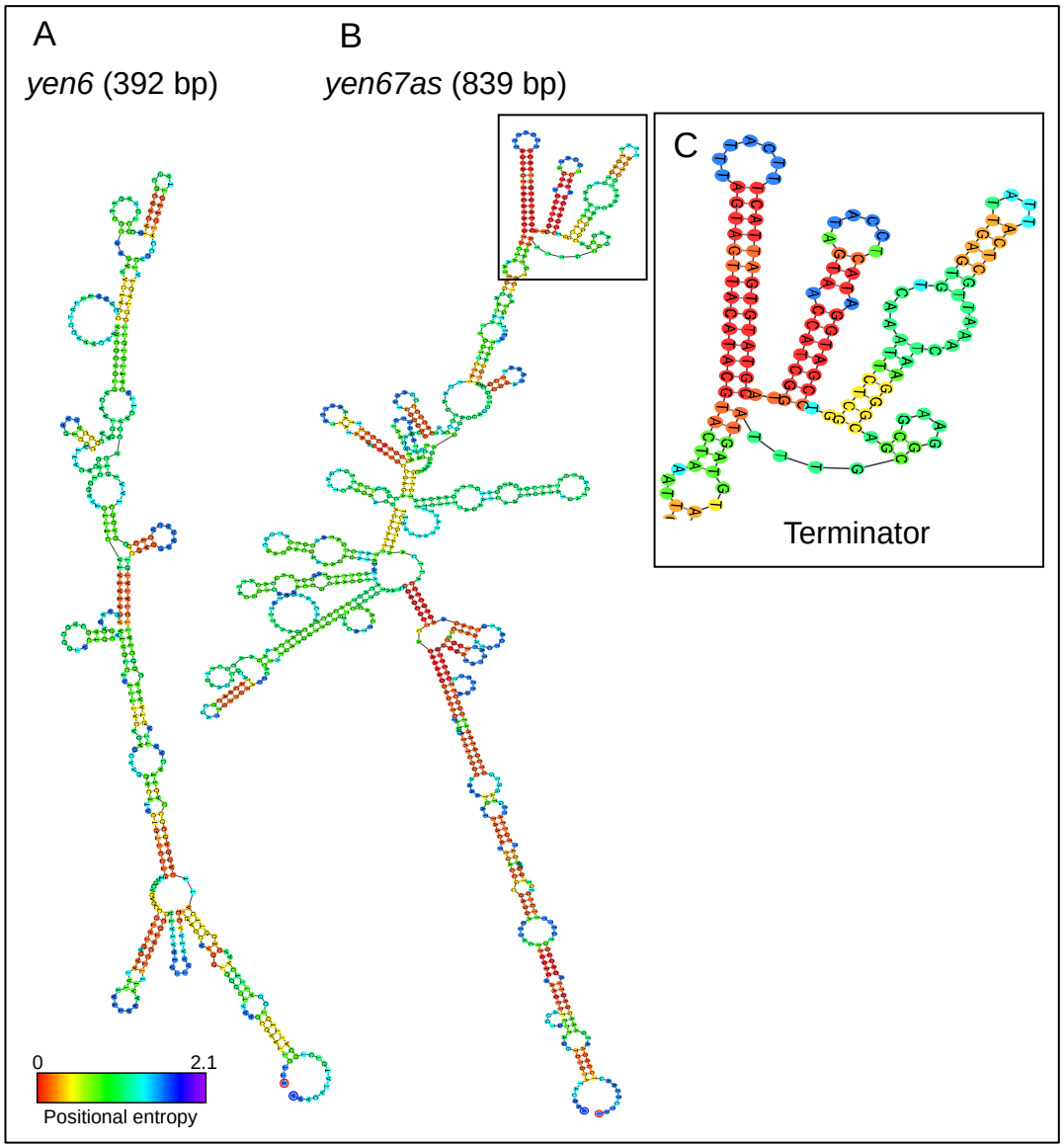


Figure 5.18: *yen6* variant mRNA secondary structure stabilities.short *yen6* mRNA variant including 32 bp 3'UTR. B. long *yen67as* mRNA variant including 477 bp 3'UTR. C. close-up of predicted terminator in 3'UTR region of *yen6*. The mRNA secondary structures were predicted using Geneious (v.10.0.9) DNA/RNA secondary structure fold viewer with the Vienna RNAfold tool (Gruber et al., 2008) using the constraint generation RNA free energy parameter estimation method (Andronescu et al., 2007) and default settings. Color scales is reliability/well-definedness measure, ranging from well-defined/low entropy (red) to ill-defined/high entropy (blue/violet) secondary structure predictions.

Table 5.3: *Yersinia entomophaga* MH96 *yen6* messenger RNA variant secondary structure stability predictions

yen6 mRNA variant	3' UTR length (bp)	Temperature (°C)	Free energy of ensemble (kcal/mol)
yen6	392	37	-98.56
yen6	392	25	-141.58
yen67as	839	37	-232.48
yen67as	839	25	-324.94

5.3.3 Characterization of *yen6* and *yen7* regulatory effects on Yen-TC

5.3.3.1 Optimization of pBAD, pBAD-*yen6*, pBAD-*yen67as* and pBAD-*yen7* in MH96

During preliminary experiments using MH96 strains carrying pBAD, pBAD-*yen6*, pBAD-*yen67as* and pBAD-*yen7,* variable effects on growth and plasmid retention were discovered under certain growth conditions. Also, ampicillin 400 μg/ml was found to be somewhat inefficient to maintain certain pBAD vectors, especially with arabinose 0.2 % (the highest concentration of arabinose used in this thesis). Since species of *Yersinia* are well known to produce β-lactamases conferring resistant to ampicillin, the alternative carbenicillin was used as a preferred antibiotic to maintain pBAD because it is considered more stable against β-lactamases. Even with high amounts of carbenicillin some plasmids were still found to be highly unstable with certain growth conditions.

When the empty arabinose expression vector, pBAD was maintained in MH96 as a control strain under strong inducing conditions (0.2 % arabinose) at 25 °C a reduction in cell density was consistently observed after 16 hours, with density only reaching ~ 5 x 10^8 CFU/ml (Table 5.4). Typically, MH96 reaches an early-stationary phase in 16 h and corresponding cell density of around ~ 1 x 10^{10} CFU/ml, so finding reduced cell density of MH96 carrying pBAD under arabinose 0.2 % was considered an interesting, albeit unintentional finding. A reduction in cell density was not observed when MH96, lacking any plasmids in the presence of 0.2 % arabinose (data not shown) attributing the reduction in growth to the interaction of the P_{BAD} and P_c promoters present on the empty vector with MH96 RNA polymerase holoenzyme and/or CAP-cAMP. Furthermore, when either *yen6* or *yen67as* was induced from pBAD the final growth

rates were consistently restored to almost wild-type levels, with these cultures reaching ~ 3 x 10^9 CFU/ml (Table 5.4). Induction of *yen7* under 0.2 % arabinose from the pBAD vector however had the greatest negative effect on growth, with average cell density only reaching ~ 6 x 10^6 CFU/ml after 16 hours.

Table 5.4: Mean cell density and pBAD/pBAD-*yen6*/pBAD-*yen67as*/pBAD-*yen7* stability in *Yersinia entomophaga* MH96 under 0.2 % arabinose inducing conditions at 25 °C. CFU = colony forming unit and SD = standard deviation.

Strain	N	plasmid stability (%) ± SD	cell density (CFU/ml) ± SD
MH96	1	-	1.0 x 10^{10}
MH96 + pBAD	3	35 ± 16	5.4 ± 1.1 x 10^8
MH96 + pBAD-*yen6*	2	95 ± 0	2.7 ± 1.0 x 10^9
MH96 + pBAD-*yen67as*	2	56 ± 15	2.9 ± 3.1 x 10^9
MH96 + pBAD-*yen7*	2	1 ± 1	5.9 ± 2.2 x 10^6

Noticeable and consistent differences in average plasmid stability were also observed among MH96 strains carrying pBAD, pBAD-*yen6*, pBAD-*yen67as* and pBAD-*yen7* under strong inducing conditions (Table 5.4). The mean stability of empty plasmid in MH96 with 0.2 % arabinose was only around 35 %, indicating the presence of selection pressures against the empty plasmid vector under these growth conditions (low pBAD stabilities were only observed under inducing conditions and plasmid was maintained at 100 % under ampicillin 400 μg/ml when not exposed to arabinose). On the contrary, the mean stability of pBAD-*yen6* was consistently higher at 95 %, indicating stronger selection for maintenance of the plasmid when yen6 was induced. Retention of pBAD-*yen67as* and pBAD-*yen7* was lower than pBAD-*yen6*, with only 56 and only 1 % of total cells retaining the plasmid, respectively.

Since Yen6, Yen67as and Yen7 over-producing strains were intended to be used for protein analysis, it was important to find appropriate inducing conditions to ensure consistent cell density and plasmid stability among strains. Specifically, further assessment into the effect of cell growth and plasmid stability of pBAD and pBAD-*yen7* were undertaken using three arabinose concentrations (0.2, 0.02 and 0.002 %) and the aforementioned growth conditions at 25 and/or 37 °C. By reducing the amount of arabinose to 0.002 %, higher cell densities and improved rates of plasmid retention

were achieved for MH96 strains carrying pBAD and pBAD-*yen7* (Table 5.5). Even at the lowest arabinose concentration (0.002 %) plasmid stability was still very low (17 %) and cell density only reached 1 x 10^7 CFU/ml when *yen7* was induced in MH96 at 25 °C.

Table 5.5: Cell density and pBAD/pBAD-*yen7* plasmid in *Yersinia entomophaga* MH96 under inducing conditions (0.2 %, 0.02 % and 0.002 % arabinose concentration) at 25 °C. CFU = colony forming unit.

Strain	Arabinose (%)	plasmid stability (%)	cell density (CFU/ml)
MH96 + pBAD	0.2	35	5.8 x 10^8
MH96 + pBAD	0.02	68	1.9 x 10^9
MH96 + pBAD	0.002	100	6.0 x 10^9
MH96 + pBAD-*yen7*	0.2	0	4.8 x 10^6
MH96 + pBAD-*yen7*	0.02	9	2.2 x 10^7
MH96 + pBAD-*yen7*	0.002	17	8.0 x 10^7

Preliminary trials using MH96 strains carrying pBAD-*yen6*, pBAD-*yen67as* and pBAD-*yen7* at 37 °C under 0.2 % arabinose were found to have variable effects on final cell density and plasmid stability after 16 h (Table 5.6). Unlike at 25 °C, pBAD-*yen6* and pBAD-*yen67as* were not found to be very stable at 37 °C with only 5 and 4 % estimated retention, respectively. The empty pBAD plasmid was also completely unstable at 37 °C, but unlike 25 °C, a deficiency in growth was not observed at this temperature, with cell densities reaching ~ 3 x 10^9 CFU/ml. In comparison, pBAD-*yen7* was highly stable at 37 °C under arabinose 0.2 %, which was in stark contrast to 25 °C where pBAD-*yen7* was estimated to be retained in less than 1 % of cells. Irrespective of temperature however, cell density of MH96 carrying pBAD-*yen7* were always much lower than MH96 with empty pBAD under these conditions.

Table 5.6: Cell density and pBAD/pBAD-*yen6*/pBAD-*yen67as*/pBAD-*yen7* stability in *Yersinia entomophaga* MH96 under inducing conditions (0.2 % arabinose) at 37 °C. CFU = colony forming units.

Strain	N	plasmid stability (%)	cell density (CFU/ml)
MH96 + pBAD	1	0	3.4 X 10^9
MH96 + pBAD-*yen6*	1	5	2.4 X 10^9
MH96 + pBAD-*yen67as*	1	4	1.3 X 10^9
MH96 + pBAD-*yen7*	1	97	4.7 X 10^7

Further work was undertaken to assess final cell density and plasmid stability of MH96 with pBAD and pBAD-*yen7* under 0.002 % arabinose at 25 and 37 °C. The addition of the arabinose was also delayed 2 h and the total culture time was 17 h. Under these conditions, control strains carrying empty pBAD were completely stable and reached normal cell densities at both temperature (Table 5.7). Strains carrying pBAD-*yen7* were found to have lower mean cell densities compared to other strains (only reached 10^8 CFUs/ml compared to 10^9 CFUs/ml) and mean plasmid stability was noticeably higher at 37 compared to 25 °C, with approximately 81 compared to 17 %, respectively. The stark temperature-dependent difference in plasmid stability of pBAD-*yen7* was considered a significant result, which eludes to the effect of *yen7* expression on the cell survival. While it is not ideal for inducing conditions (in this case 0.002 % arabinose) to result in reduced selection against the expression vector and deleterious effects on cell growth, these conditions were ultimately much better than previous attempts to induce with 0.02 and 0.002 %. As such, 0.002 % arabinose (added after 2 hours pre-incubation in 50 ml media) were considered the best possible inducing conditions available for pBAD-*yen7* in MH96 and was also used to prepare cultures for protein analysis (section 5.3.3.2) and β-gal assay for a *chi1* translational reporter strain (section 5.3.3.6.4), which allowed for a more thorough investigation of the role of *yen7* on Yen-TC production in MH96.

Table 5.7: Mean cell density and pBAD/pBAD-*yen7* stability in *Yersinia entomophaga* MH96 under inducing conditions (0.002 % arabinose added after 2 hours growth) at 25 and 37 °C . SD = standard deviation, CFU = colony forming units.

Strain	Temperature (°C)	N	Plasmid stability (% ± SD)	cell density (CFU/ml ± SD)
MH96	25	2	-	$9.1 \pm 1.3 \times 10^9$
MH96 + pBAD	25	3	115 ± 24	$7.0 \pm 0.9 \times 10^9$
MH96 + pBAD-*yen7*	25	3	17 ± 14	$2.4 \pm 2.6 \times 10^8$
MH96	37	2	-	$8.0 \pm 1.7 \times 10^9$
MH96 + pBAD	37	3	112 ± 31	$5.4 \pm 0.4 \times 10^9$
MH96 + pBAD-*yen7*	37	3	81 ± 20	$7.4 \pm 4.3 \times 10^8$

Based on these results further assessment of the effects of over-production of Yen6, Yen67as on protein secretion and Yen-TC production by SDS-PAGE and β-galactosidase assay using *lacZ*-reporter strains were undertaken using either 0.2 or 0.02 % arabinose. Following experiments related to over-production of Yen7 were subsequently undertaken with addition of arabinose to final concentration of 0.002 % delayed two hours post-inoculation.

5.3.3.2 Effect of over-production of Yen6, Yen67as and Yen7 on Yen-TC production and global protein secretion

Under strong inducing conditions (0.2 % arabinose) at 25 °C, the over-production of Yen67as from pBAD was initially found to increase secretion of Yen-TC toxin components compared to MH96 strains with pBAD or pBAD-*yen6* (Figure 5.19A). Much more prominent protein bands corresponding to Yen-TC components YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa) were observed in the CS from the from MH96carrying pBAD-*yen67as* under these inducing conditions. Other putative secreted VFs including M66 peptidase (81 kDa), hypothetical VF (44 kDa) and chitinase (33 kDa) were also observed when Yen67as was induced under 0.2 % arabinose in MH96 at 25 °C. On the contrary, increased production of Yen-TC components were not observed in the CP from MH96 with pBAD-*yen67as* compared to the other strains, but greater amounts of the M66 peptidase protein were observed in the CP (Figure 5.19B); however, reduced plasmid retention was also shown for MH96 carrying pBAD-*yen67as* compared to pBAD-*yen6* under these conditions (Table 5.4).

Though this experiment was repeated three times, visualizing increased secretion of Yen-TC components by the MH96 pBAD-*yen67as* strain were difficult to repeat, perhaps due to variation in SDS-PAGE hand-cast gels and silver stain development times (supplementary Figure S23, Figure S24 and Figure S25). We also observed that β gal production of a *chi1*::lacZ reporter strain (proxy for Yen-TC operon) with pBAD-*yen67as* under 0.2 % was highly variable, with only one out of three culture flasks reaching extremely high enzymatic activity per experiment (section 5.3.3.2, Figure 5.29).

Attempts to identify effects on Yen-TC secretion and production from induction of two other *yen6* variants carried by pBAD-Amb2 and pBAD-Amb3 (Table S13) with 0.2 % arabinose were also assessed. The mRNA of *yen6* is predicted to form a very stable secondary structure due to the presence of a long 3'UTR including a prominent terminator and putative *yen7as* (*cis*-encoded asRNA sharing extended overlap with the coding portion of *yen7*) (Figure 5.18). In order to assess whether truncations of the *yen6* 3'UTR resulted in observable effects on Yen-TC production and secretion (which was observed, albeit in-consistently with pBAD-*yen67as* as reported above) *yen6* 3'UTR variants Amb2 (3'UTR of 67 bp; terminator absent) and Amb3 (3' UTR of 148 bp; terminator present) were induced and then protein visualized by SDS- PAGE; however, this preliminary work did not find any difference with respect to Yen-TC secretion and production when MH96 carried pBAD-Amb2 and pBAD-Amb3 under 0.2 % arabinose compared to empty vector (Figure S24 and Figure S25).

One key finding of this work was evidence of restored secretion when Yen7 was overproduced at 37 °C, as indicated by the presence of prominent protein bands in the CS by SDS-PAGE (Figure 5.19A); however restored secretion of Yen-TC components were not observed under these growth conditions. Protein bands corresponding to Chi1 (60 kDa) and Chi2 (70 kDa) were much more prominent in the supernatant when *yen7* was induced (even with greatly reduced cell densities observed under these strong inducing conditions) compared to CS from MH96 carrying empty vector, but since other proteins are known to migrate at the same position as Chi1 and Chi2 (Hurst et al., 2011a) increased secretion of these specific Yen-TC components could be definitively demonstrated using this approach. Protein bands corresponding to Yen-TC components and other putative secreted VFs were not observed in the CP when *yen7* was induced at 37 °C either; however, a great reduction in cell density due to induction of *yen7* under inducing conditions (0.2 % arabinose) likely confounded these results. It should be noted, that later experiments using the *cis*-merodiploid P_{chi1}::lacZ reporter strain under the same conditions determined that over-expression of *yen7* at 37 °C resulted in significantly greater β-gal production by this strain compared to empty vector (section 5.3.3.6.5 and Figure 5.30) and these results are consistent with observations at the protein-level, that when *yen7* is over-expressed at 37 °C in MH96 there is a positive transcriptional effect on *chi1*.

Reduced secretion of Yen-TC components at 37 °C with arabinose 0.2 % was observed in MH96 strains carrying pBAD-*yen6* and pBAD-*yen67as* compared to MH96 with empty vector. These strains, along with wild-type also secreted a larger molecular weight protein (~ 175 kDa) only at 37 °C (not secreted at 25 °C), which was visualized as the highest protein band found on the CS SDS-PAGE (Figure 5.19A) and this high molecular weight band was not observed at 37 °C when y*en7* was induced from pBAD. This band was excised and assessed by LC-ESI-MS/MS (section 5.3.1.1, supplemental Figure S16 and supplemental Table S12).

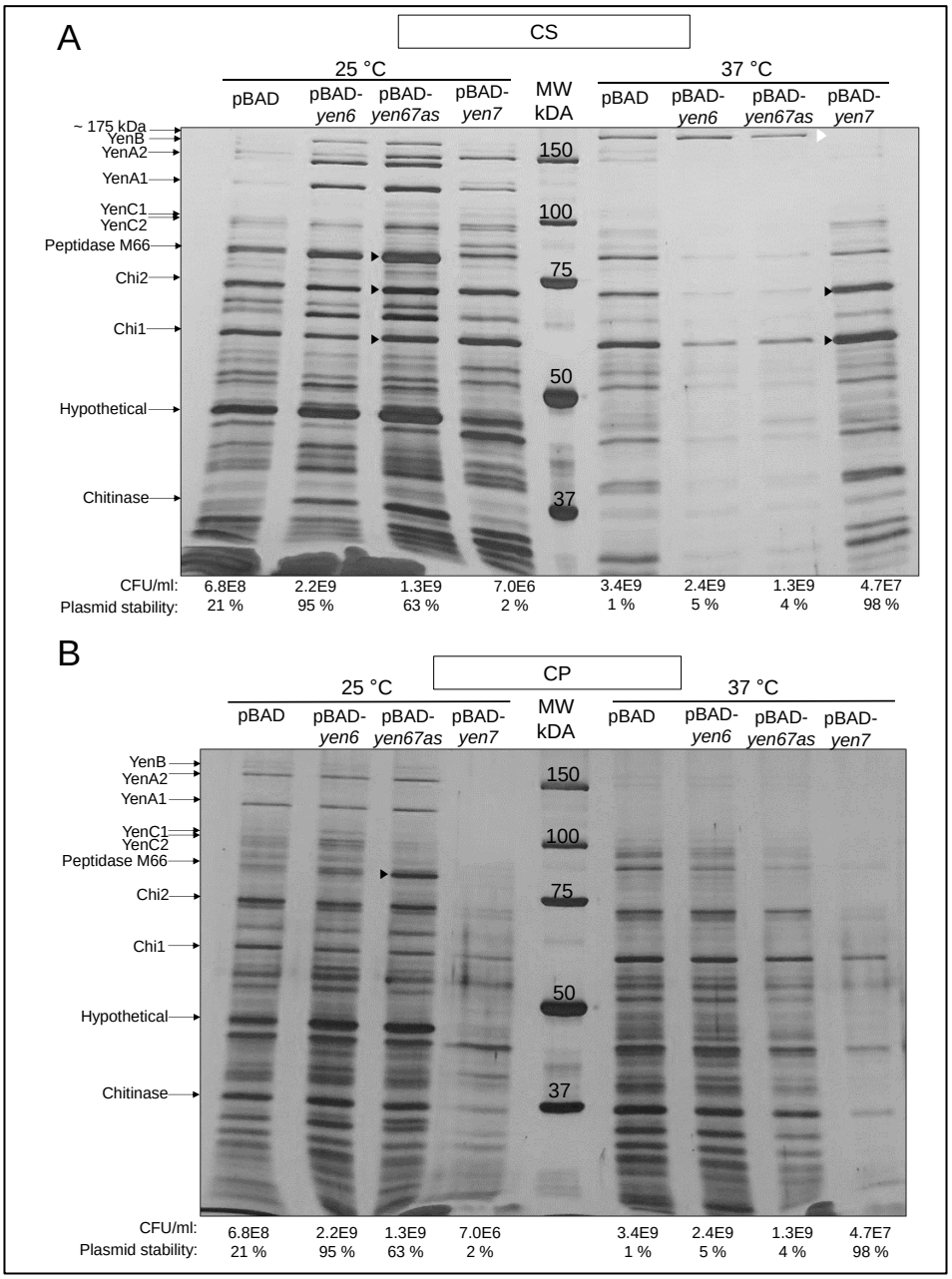


Figure 5.19: Silver-stained SDS-polyacrylamide gel of protein produced and secreted by *Yersinia entomophaga* MH96 carrying arabinose-inducible expression vector pBAD, pBAD-*yen6*, pBAD-*yen67as* and pBAD-*yen7* under arabinose 0.2 % at 25 and 37 °C grown for 16 h (early stationary). A) filtered cell supernatant (CS) and B) cell pellet (CP). The Marker (MW) lane contains Bio-Rad Precision Plus Protein unstained protein standard, with respective ladder sizes given. Arrows correspond to Yen-TC components, YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa) and peptidase M66 (81 kDa), hypothetical (44 kDa) and chitinase (33 kDa). Cell density is reported as CFU/ml as determined by enumeration plates and plasmid stability is reported as the proportion of CFUs enumerated from ampicillin 400 μg/ml compared to CFUs enumerated from media without antibiotic. Black triangles denote Yen-TC components and peptidase M66, which were found to be secreted and/or produced in greater amounts when *yen67as* or *yen7* was over-expressed.

Reducing the amount of arabinose in the culture from 0.2 to 0.002% had a positive effect on growth and plasmid stability in both MH96 with pBAD and pBAD-*yen7*. Decreased amounts of arabinose at 25 °C resulted in increased amounts of Yen-TC components being produced in both the cell pellet and the culture supernatant (Figure 5.20). Importantly nearly equal amounts of toxin components were visualized from in the CP of MH96 with pBAD and pBAD-*yen7* with 0.002 % arabinose, even though the final cell density of MH96 with pBAD-*yen7* was two-fold less than with vector only.

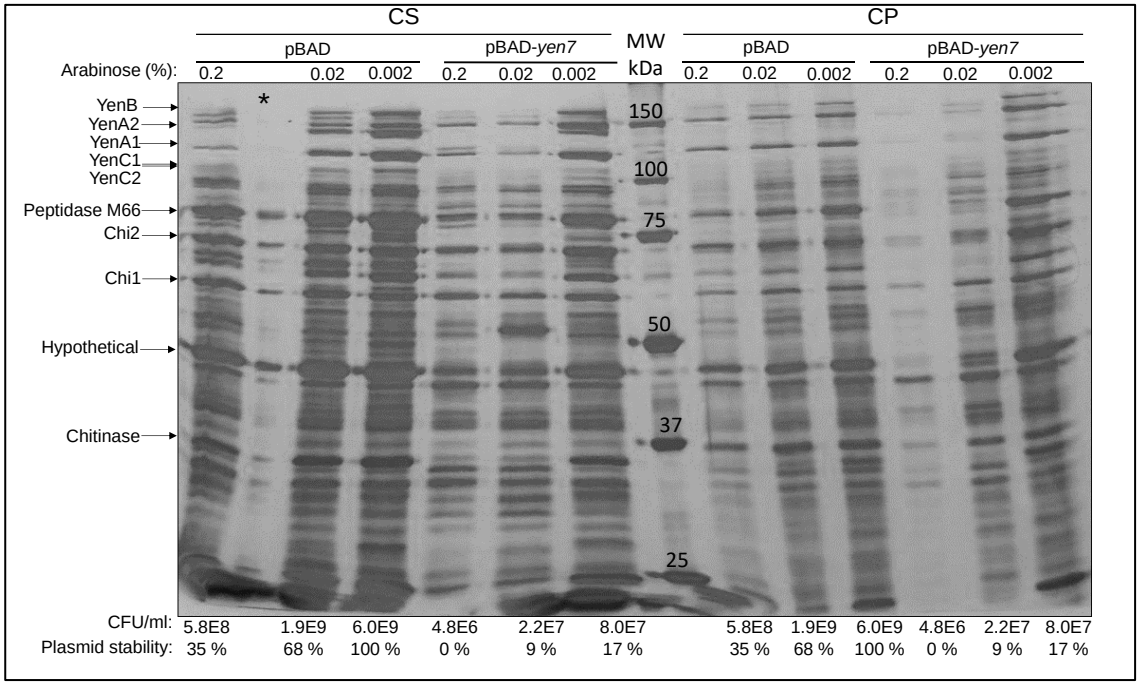


Figure 5.20: Silver-stained SDS-polyacrylamide gel of proteins produced and secreted by *Yersinia entomophaga* MH96 carrying arabinose-inducible expression vector pBAD and pBAD-*yen7* under arabinose 0.2, 0.02 and 0.002 % at 25 °C for 16 h. CS = cell supernatant, CP = cell pellet, MW = molecular weight over marker lane containing Bio-Rad Precision Plus Protein unstained protein standard, with respective ladder sizes given. Arrows correspond to Yen-TC components, YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa) and peptidase M66 (81 kDa), hypothetical (44 kDa) and chitinase (33 kDa). Cell density is reported as CFU/ml as determined by enumeration plates and plasmid stability is reported as the proportion of CFUs enumerated from ampicillin 400 μg/ml compared to CFUs enumerated from media without antibiotic.

At 37 °C, MH96 does not normally secrete proteins into the supernatant when cultured in LB broth (with or without pBAD), apart from a large molecular weight protein (~ 175 kDa) identified through LC-ESI-MS/MS (supplementary Table S12 and Figure S16) as a putative filamentous hemagglutinin N-terminal containing protein (PL78_04365)

and two other unknown proteins located at around 80 kDa and 60 kDa in size, which are observed as very faint bands on the SDS-PAGE of the 37 °C CS samples (Figure 5.21A). When *yen7* was induced with 0.002 % arabinose at 37 °C protein secretion was partially restored compared to wild-type and protein bands tentatively corresponding to Yen-TC components YenC1, YenC2, Chi2 and Chi1 were present in the supernatant (Figure 5.21A and supplemental Figure S26). Several other small bands were also found in large amounts in the CS along with YenC1, YenC2, Chi2 and Chi1 and could possibly represent putative virulence factors peptidase M66, hypothetical virulence factor and chitinase. A salient find was the comparative lack of Yen-TC and other putative secreted VFs in the CP when expression of *yen7* was induced with 0.002 % arabinose at 37 °C compared to the same strain cultured at 25 °C, where the Yen-TC bands were more evident compared to wild-type carrying empty pBAD vector.

As outlined above, induction of *yen7* at either 25 or 37 °C results in a reduction of total cell density compared to MH96 with or without pBAD under these conditions. It was observed, that even despite the reduction in total cell density after 16 hours, when *yen7* was induced at 25 °C, greater amounts Yen-TC components were secreted compared to MH96 with or without pBAD grown in the same conditions, which grew to 10 times greater cell density. Additionally, greater amounts of all Yen-TC toxin components, YenB, YenA2, YenA1, Chi2 and Chi1 (but not YenC1 and YenC2) were identified in the CP when Yen7 was over-produced at 25 °C (Figure 5.21B), although the YenC and Chi components are masked by other similar sized proteins making it difficult to quantify these specific proteins. Large amounts of other secreted putative VFs, including peptidase M66 (81 kDa) and hypothetical VF (44 kDa) were also found in the CP when Yen7 was over-produced at 25 °C. Conversely, the third secreted chitinase (PL78_05310) (33 kDa) was comparatively reduced in the CP when *yen7* was induced at 37 °C, compared to MH96 with or without pBAD under the same growth conditions.

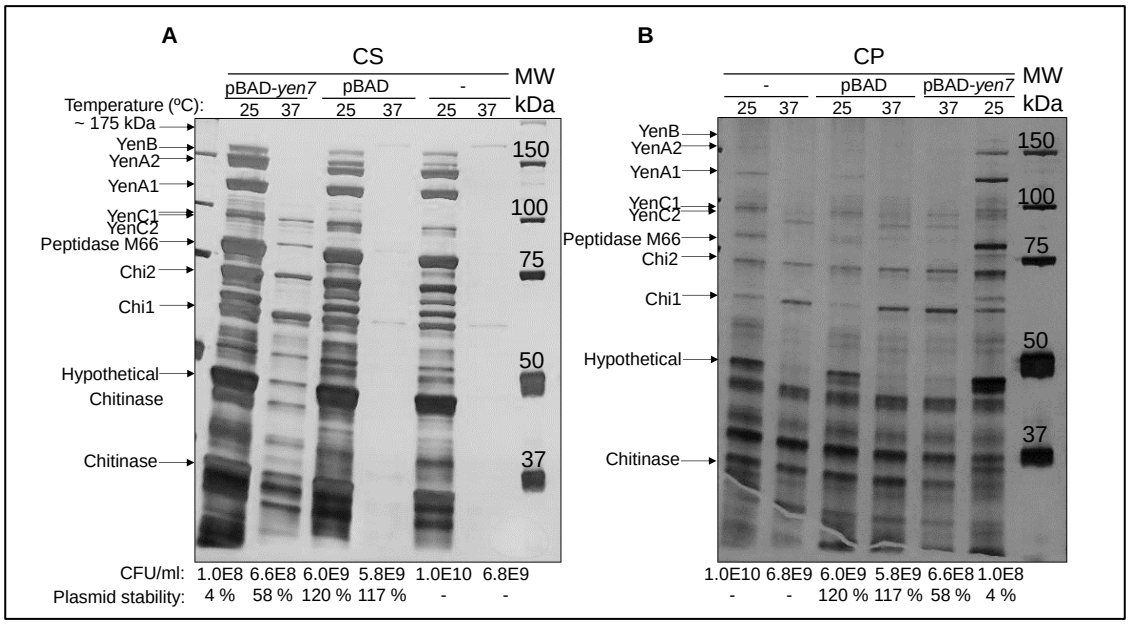


Figure 5.21: Silver-stained SDS-PAGE of proteins produced and secreted by *Yersinia entomophaga* MH96 carrying arabinose-inducible expression vector pBAD and pBAD-*yen7* under arabinose 0.002 % at 25 or 37 °C for 16 h. CS = cell supernatant, CP = cell pellet, MW = molecular weight over marker lane containing Bio-Rad Precision Plus Protein™ unstained protein standard, with respective ladder sizes given. Arrows correspond to Yen-TC components, YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa) and filamentous hemagglutinin (~175 kDa), peptidase M66 (81 kDa), hypothetical virulence factor (44 kDa) and chitinase (33 kDa). Cell density is reported as CFU/ml as determined by enumeration plates and plasmid stability is reported as the percentage of CFUs enumerated from ampicillin 400 μg/ml compared to CFUs enumerated from media without antibiotic.

5.3.3.3 Effect of over-production of *yen7* on Yen-TC production and global protein secretion in Δ*yen6* and Δ*yen6yen7*

To determine if deletion of *yen6* or *yen6yen7* resulted in greater global toxin secretion at 37 °C compared to wild-type when *yen7* was over-expressed from the pBAD arabinose induction vector additional overnight cultures were grown under 0.002 % arabinose at 25 and 37 °C. Protein from the culture supernatant were visualized by SDS-PAGE that showed no difference in the levels of Yen-TC components secreted between MH96, Δ*yen6* and Δ*yen6yen7* strains when *yen7* was over-expressed at 37 and 25 °C.

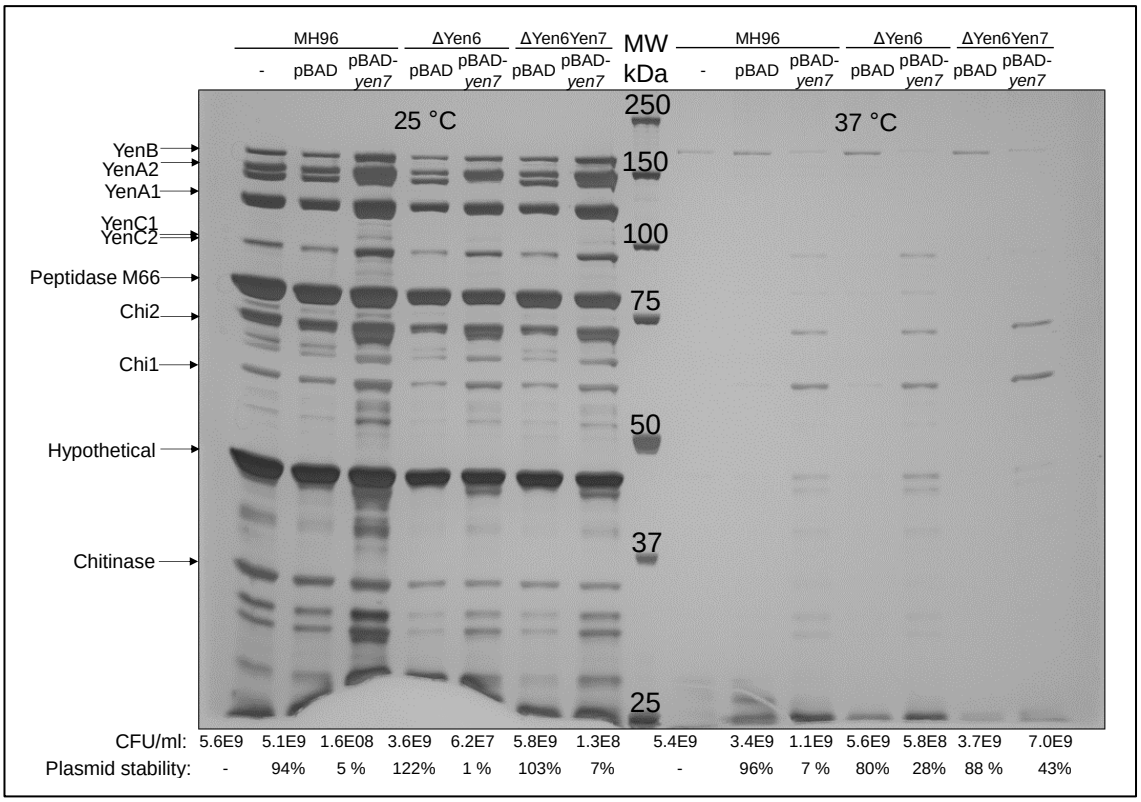


Figure 5.22: Silver-stained SDS-polyacrylamide gel of cell supernatant from *Yersinia entomophaga* MH96 strains carrying arabinose-inducible expression vector pBAD and pBAD-*yen7* under arabinose 0.002 % at 25 or 37 °C for 15 h. MW = molecular weight over marker lane containing Bio-Rad Precision Plus Protein unstained protein standard, with respective ladder sizes given. Arrows correspond to Yen-TC components, YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa) and peptidase M66 (81 kDa), hypothetical virulence factor (44 kDa) and chitinase (33 kDa). Cell density is reported as CFU/ml as determined by enumeration plates and plasmid stability is reported as the percentage of CFUs enumerated from ampicillin 400 μg/ml compared to CFUs enumerated from media without antibiotic.

5.3.3.4 Transcriptional organization of *yen6*, *yen7* and *chi1*

In order to validate the transcriptional organization of the *yen6*, *yen7* and *chi1* region predicted by the RNA-seq data and confirm operon structure of these genes, RT-PCR was performed on RNA isolated from MH96 grown at 25 °C *in vitro* during exponential growth (~ 1 x 10^8 CFU/ml) and nine regions ranging from 250 – 875 bp among *yen5*, *yen6*, *yen7* and *chi1* were investigated (Figure 5.23A). Seven out of nine RT-PCR reactions generated a product of similar size to the product generated from genomic DNA template (Figure 5.23B). RT-PCR across region 4 (spans from *yen6* across the predicted intergenic double hairpin into *yen7as*) was successful, providing evidence

that *yen6* is transcribed with a long 3'UTR extending into anti-sense region of *yen7* (as predicted in the transcriptome data).

RT-PCR also demonstrated that *yen6* and *yen7* are independently promoted, since RT-PCR across regions 1 and 7 did not generate a PCR product, but these regions could be amplified using gDNA as template. Since *yen6* and *yen7* independently promoted, these promoter regions were considered suitable for additional analysis using *lacZ* translational fusions. Successful amplification by RT-PCR from region 8 also determined that *chi1* has an even longer 5'UTR than previously estimated (i.e., 166 bp was estimated from RNA-seq data). Additional work is required to characterize the exact transcriptional start site of *chi1*; however, since *yen7* (the gene located directly 5' of *chi1*) is encoded on the anti-strand, it was also assumed that *chi1* is under independent promotion. Furthermore, based on previously reported operon prediction from RNA-seq data (Chapter 4, section 4.3.8, Table 4.13), *chi1* is the first gene in the 4 gene operon *chi1-yenA1-yenA2-chi2*, which also supports *chi1* as a good proxy for (at least) expression of YenA/Chi components of the Yen-TC holotoxin. The RNA-seq data also predicted that *yenB* and *yenC1* are expressed in a 2-gene operon but *yenC2* is transcribed separately, but these findings need to be further validated using RT-PCR.

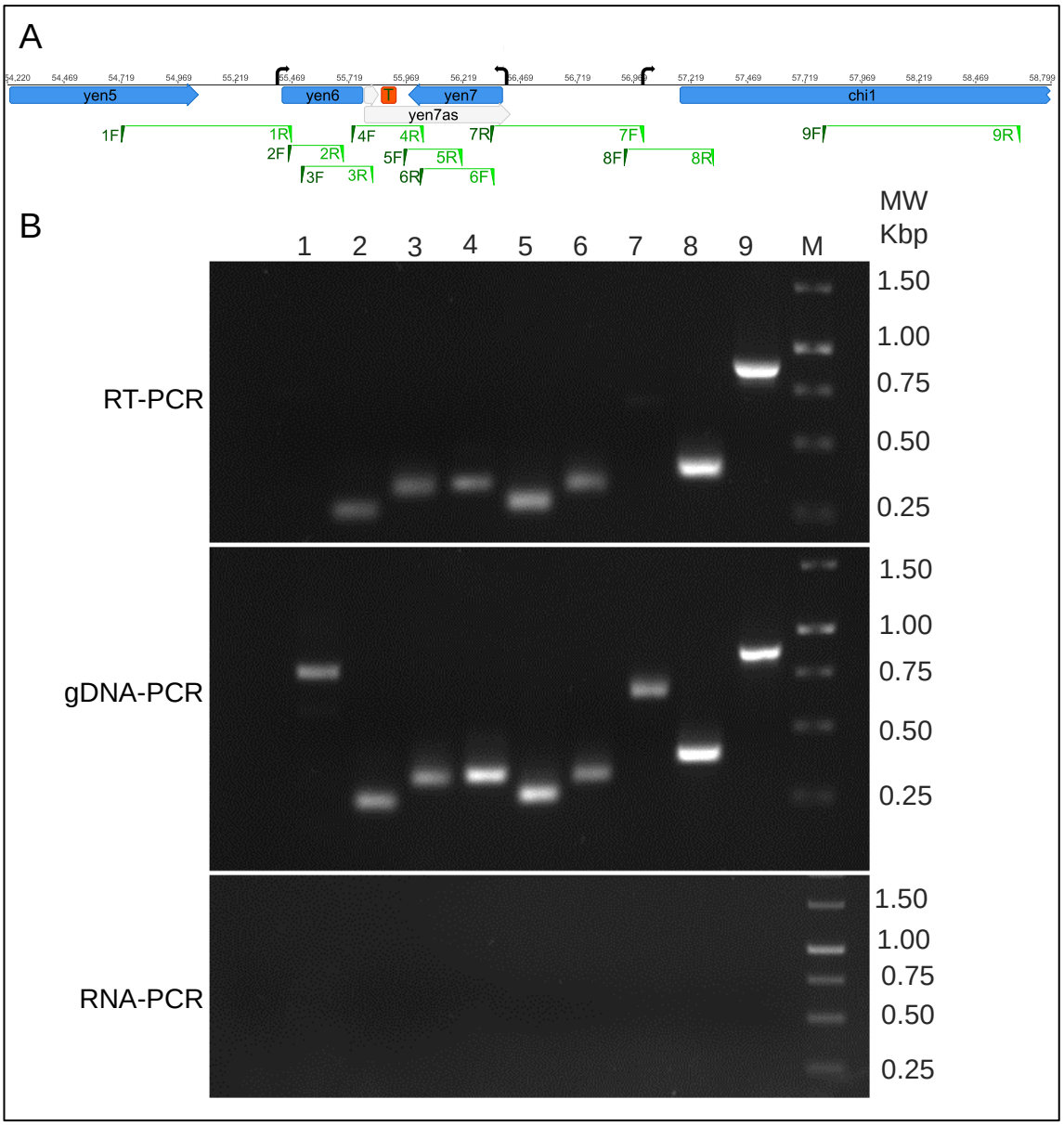


Figure 5.23: Transcriptional organization of *yen5*, *yen6*, *yen7* and *chi1* coding region of PAI_{ye96} of *Yersinia entomophaga* MH96. A) Transcriptional start sites (black curved arrows) were estimated using transcriptome data and independent promotion was confirmed by RT-PCR for *yen6*, *yen7* and *chi1*. Putative *cis*-acting anti-sense RNA (*yen7as*) was predicted from transcriptome data. A terminator (T) was predicted in the intergenic region between *yen6* and *yen7*. RT-PCR regions 6 and 7 were tested on the reverse strand. B) PCR amplicons generated using cDNA (RT-PCR), gDNA and RNA as template for nine different regions (lanes 1-9) using specific primers visualized on 1 % agarose gel using Red Safe fluorescent dye.

5.3.3.5 Calibration curve: 25 vs 37 °C

Prior to carrying out the β-gal assay on the plate-reader, calibration curves for MH96 grown at 25 and 37 °C were generated using absorbance OD_{600} on the SPECTROstar^{Nano}

plate-reader. These calibration curves were generated as a method to estimate actual cell densities based on absorbance readings from the plate-reader and determine the validity of making comparisons between cells grown at two different cultures. The calibration curves revealed a major difference between cells grown at 25 and 37 °C. The absorbance from culture grown at 37 °C was consistently correlated with higher CFUs compared to the absorbance from culture grown at 25 °C (Figure 5.24). The estimated slope of the linear model of the 37 °C was also found to be three times greater than that predicted for the 25 °C data.

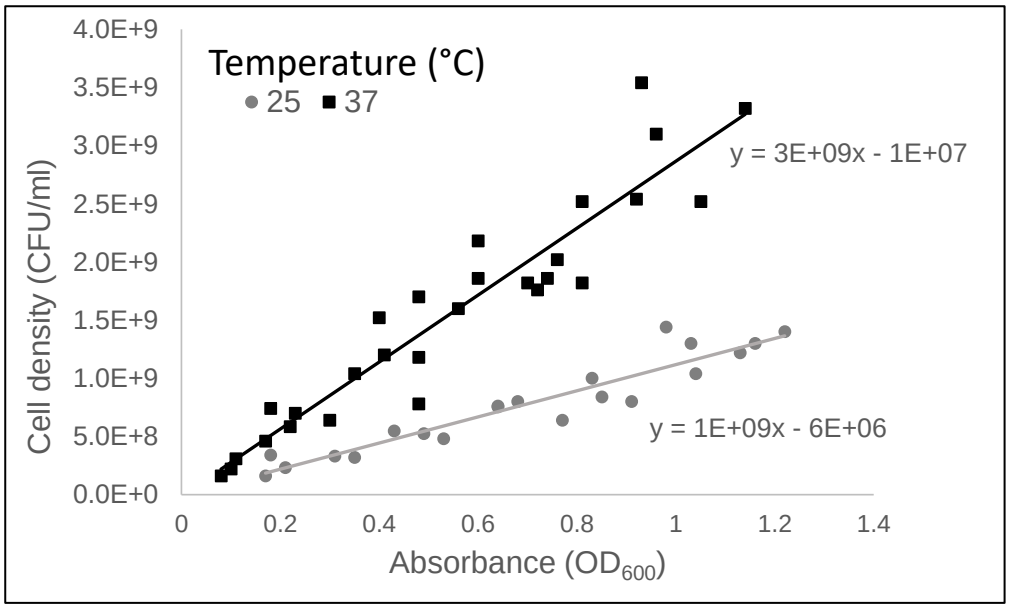


Figure 5.24: Calibration curves for *Yersinia entomophaga* MH96 grown at 25 °C and 37 °C in 50 ml LB broth shaking at 200 rpm until OD_{600} reached approximately 1.0. Absorbance were quantified on SPECTROstar^{Nano} plate-reader in 96-well plates. Cell density were determined from dilution plate counts. Data were combined from three experiments, each consisting of three replicate flasks per temperature.

5.3.3.6 B-galactosidase assay for *lacZ* translational reporter fusions

5.3.3.6.1 MH96 carrying P_{yen7}::*lacZ* reporter fusion did not produce β galactosidase

Cis-merodiploid MH96 strains carrying P_{yen6}::*lacZ*, P_{yen7}::*lacZ* and P_{chi1}::*lacZ* translational fusions (supplemental Table S2) were initially assessed at 25 °C using the β-gal assay but the P_{yen7}::*lacZ* reporter bearing a 873 bp region 5′ of *yen7* failed to produce any detectable β-gal and were not further assessed (supplementary Table S15). The correct fusion was confirmed in the P_{yen7}::*lacZ* by PCR using validation primers yen7_valF and

LacZ_valR (Table S13), so it was assumed that critical regulatory elements located internal to the *yen7* ORF were missing in P_{yen7}::*lacZ* reporter fusion and under these conditions translation of *lacZ* does not occur in this strain.

5.3.3.6.2 Temperature effects on P_{yen6}::*lacZ* and P_{chi1}::*lacZ* reporter fusions *in vitro*

The rate of β-gal production (measured in MU) from MH96 strains carrying P_{yen6}::*lacZ* and P_{chi1}::*lacZ* reporter fusions were assessed at 25 and 37 °C. The rate of β-gal production was significantly lower at 37 compared to 25 °C for both P_{yen6}::*lacZ* (Welch two sample t-test; P-value = 0.0026) and P_{chi1}::lacZ reporter fusions (P-value = 0.0003) (Figure 5.25 and Figure 5.26).

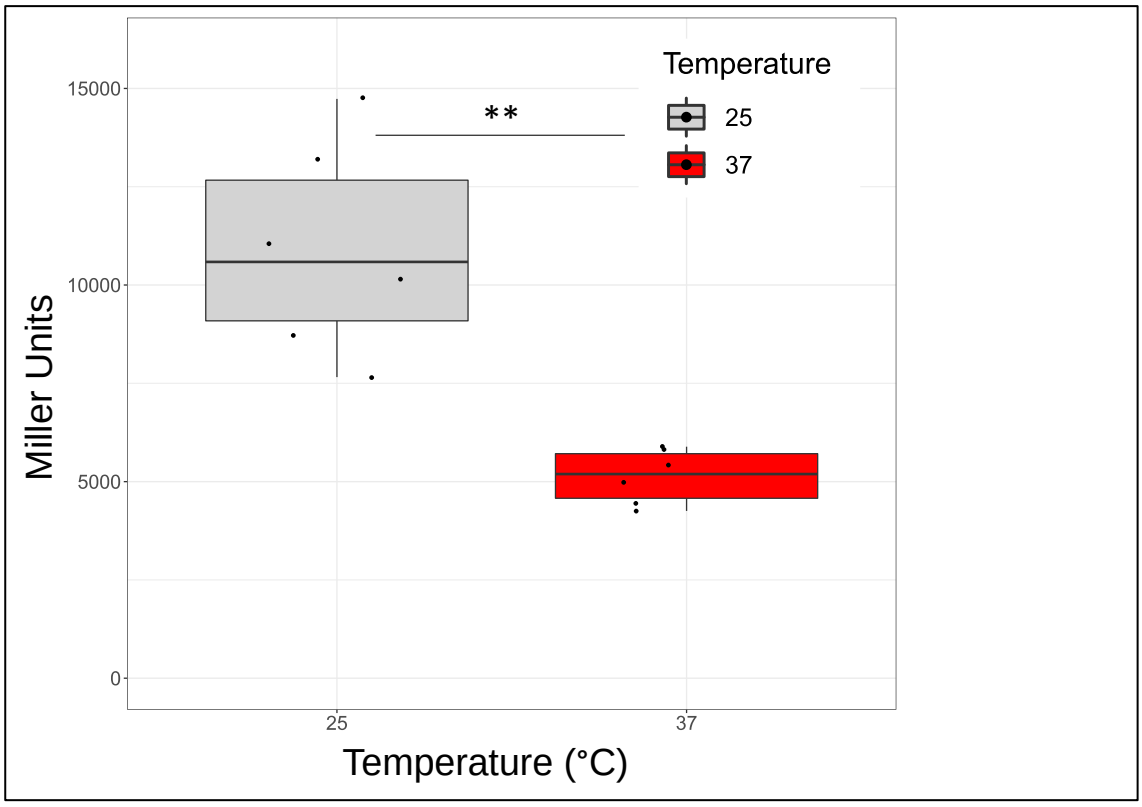


Figure 5.25: Effect of temperature on β-galactosidase production in *cis*-merodiploid *Yersinia entomophaga* MH96 strains harbouring P_{yen6}::*lacZ*. Rate of β-galactosidase was measured from 1:25 culture dilution in a 96-well plate at OD_{420} after 16 hours of growth and represented as Miller Units (normalized by OD_{600}). Data were combined from two separate experiments, each containing three biological replicates/temperature. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively. Horizontal bars indicate statistically significant differences: ** represents a P-value less than 0.01 but greater than 0.001 based on Welch two-sample t-test.

The decrease in β gal production by the P_{chi1}::*lacZ* reporter at 37 compared to 25 °C is consistent with previous SDS-PAGE results, showing reduced production and secretion of Yen-TC components at higher temperatures. These results however, are not consistent with the *in vitro* RNA-seq from 25 and 37 °C, where a significant transcriptional change for Yen-TC components was not identified between different temperatures.

The RNA-seq data did however identify a very strong temperature-dependent response in transcription of *yen6*, which had a much higher expression at 37 compared to 25 °C for both *in vivo* and *in vitro* growth conditions. Significantly lower enzymatic activity at 37 °C was detected in the P_{yen6}::*lacZ* reporter compared to 25 °C under *in vitro*

conditions, which in is in contrast to the previous findings from RNA-seq that showed transcription of *yen6* is significantly higher at 37 compared to 25 °C *in vitro*, suggesting the presence of some yet unknown culture-specific post-transcriptional repression mechanism for y*en6* at 37 °C *in vitro*.

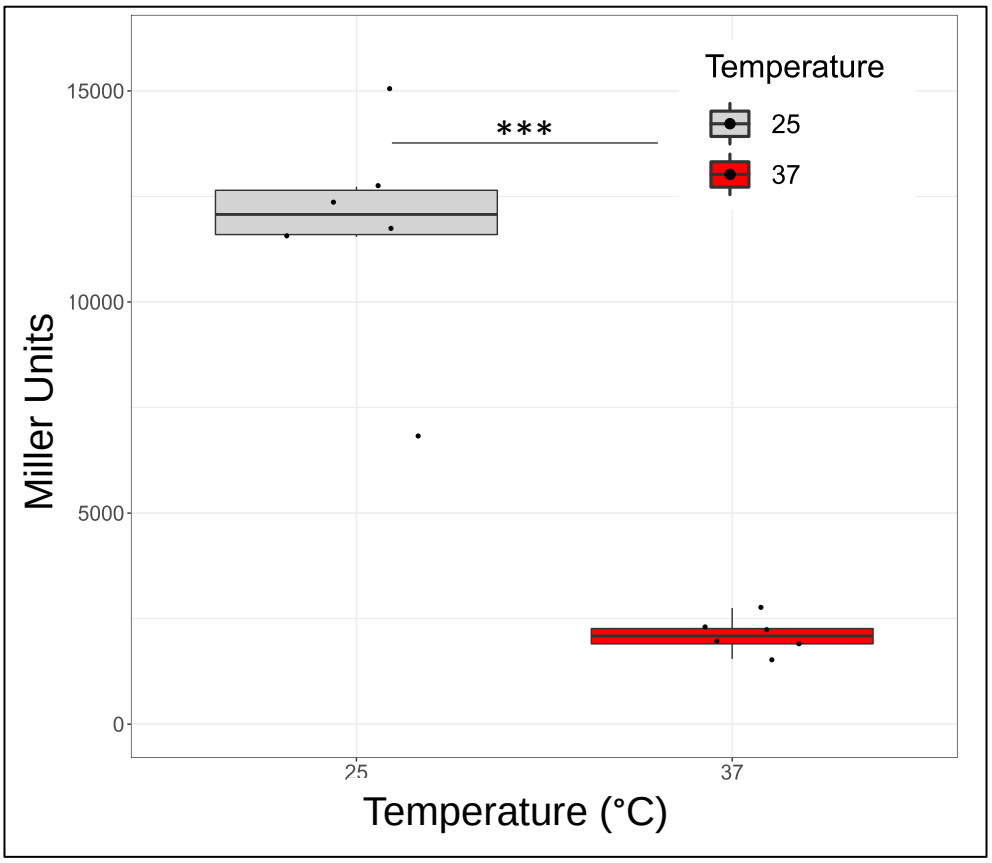


Figure 5.26: Effect of temperature on β-galactosidase production in *cis*-merodiploid *Yersinia entomophaga* MH96 strains harbouring P*chi1*::*lacZ*. Rate of β-galactosidase was measured from 1:25 culture dilution in a 96-well plate at OD_{420} after 16 hours of growth and represented as Miller Units (normalized by OD_{600}). Data were combined from two separate experiments, each containing three biological replicates/temperature. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively. Horizontal bars indicate statistically significant differences: *** represents a P-value less than 0.001 based on Welch two-sample t-test.

5.3.3.6.3 β galactosidase activity of MH96, Δ*yen6*, Δ*yen67* and Δ*yen7 cis*-merodiploid P*Chi1*::*lacZ* strains at 25 °C *in vitro*.

When the enzymatic activity of P*chi1*::*lacZ* in Δ*yen6*, Δ*yen67* and Δ*yen7* genetic backgrounds were compared to wild-type, MH96, significant differences were

identified (Figure 5.27). The mean rate of β-gal activity in wild-type was more than two times less compared to the Δ*yen6* and two and three times greater in Δ*yen67* and Δ*yen7*, respectively.

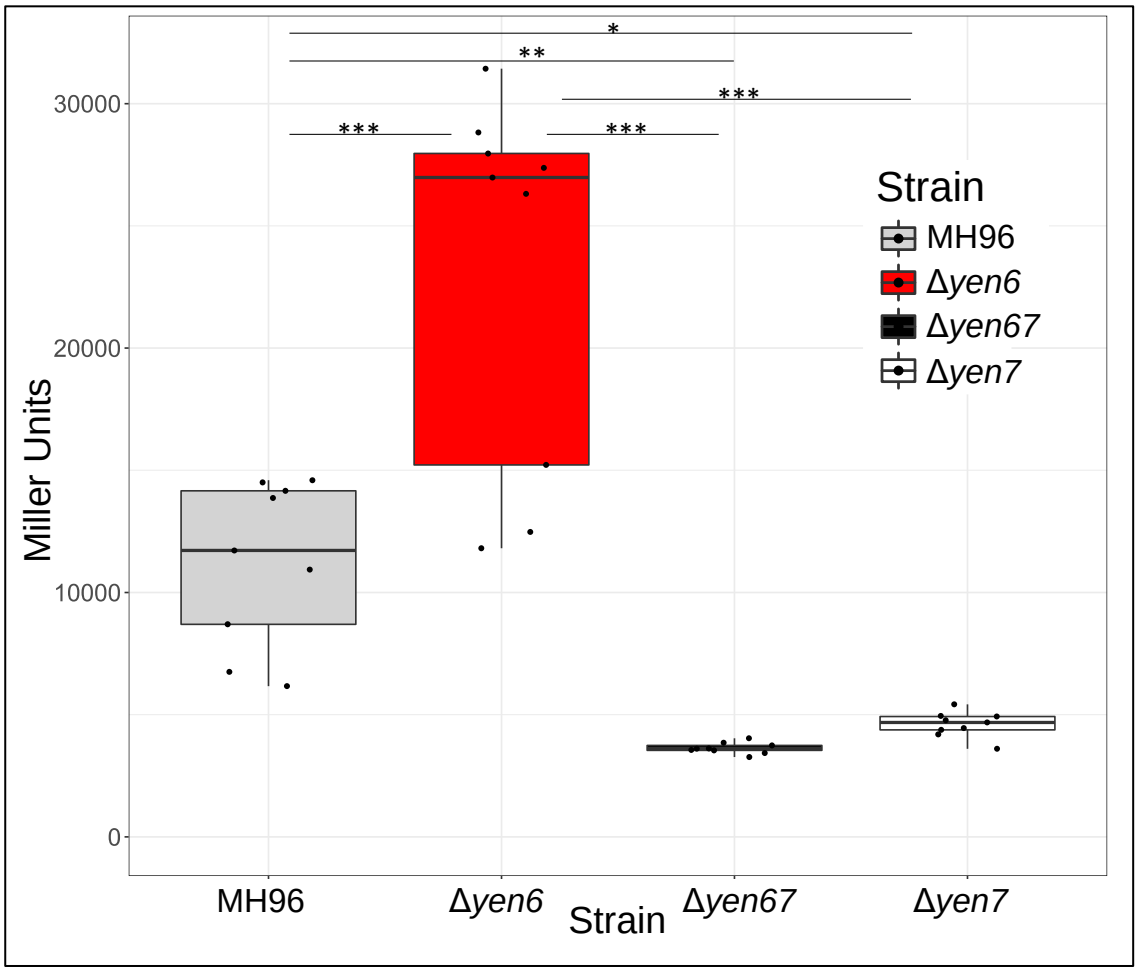


Figure 5.27: Increased β-galactosidase production in *cis*-merodiploid strains harbouring P_{chi1}::*lacZ* in Δ*yen6* and decreased β-galactosidase production in *cis*-merodiploid strains harbouring P_{chi1}::*lacZ* in Δ*yen67* and Δ*yen7* compared to wild-type *Yersinia entomophaga* MH96. Rate of β-galactosidase was measured from 1:25 culture dilution in a 96-well plate at OD_{420} after 16 hours of growth at 25 °C and represented as Miller Units (normalized by OD_{600}). Data were combined from three separate experiments, each containing three biological replicates/strain. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively. Horizontal bars indicate statistically significant differences: *** = adjusted P-value < 0.001, ** = 0.001 < adjusted P-value < 0.01 and * = adjusted P.value < 0.05 based on Tukey multiple comparison of means with 95 % family-wise confidence level.

To date, only preliminary attempts have been made to complement Δ*yen6* and Δ*yen67* *in trans* using mid-copy vector containing both *yen6* and *yen7* under their own natural

promoters (pACYC184-P_{yen6}-P_{yen7}; this region extended slightly beyond regions characterized using RT-PCR, ensuring full capture of both *yen6* and *yen7* independent promoter regions) (see Figure 5.23 above). When this complementation plasmid was carried in a Δyen*6* containing P_{chi1}::*lacZ* reporter fusion, β-gal activity was only marginally decreased compared to when the empty vector was carried in this strain (Table 5.8). Similar attempts were made to complement Δ*yen67*, but in this experiment the pACYC184-P_{yen6}-P_{yen7} plasmid further reduced the amount of enzyme activity in Δ*yen67* compared to the empty vector, and significantly less β-gal levels were detected in both Δ*yen67* strains compared to MH96 carrying the empty vector (Table 5.9). Lastly, preliminary complementation attempts using Δ*yen6* with P_{chi1}::lacZ and pBAD-*yen67as* under inducing conditions with 0.002 % arabinose were undertaken (Table 5.10). Both Δyen*6* P_{chi1}::lacZ strains (either carrying pBAD or pBAD-*yen67as*) were found to have markedly higher β-gal production compared to MH96 P_{chi1}::*lacZ* carrying pBAD-*yen67as* under arabinose 0.002 % at 25 °C. Also, as observed in wild-type, over-expression of Yen67as from pBAD resulted in greater enzymatic activities of the *chi1* reporter in both Δyen*6* and wild-type, compared to when these strains carried empty vector.

Table 5.8: Preliminary β galactosidase assay for pACYC184- P_{yen6}-P_{yen7} complementation of P_{chi1}::*lacZ* in *Yersinia entomophaga* MH96 and Δ*yen6* (single replicate).

Strain	N	Miller Unit
MH96 + P_{chi1}::lacZ + pACYC184sp	1	3,121
MH96 + P_{chi1}::lacZ + pACYC184sp	2	2,641
MH96 + P_{chi1}::lacZ + pACYC184sp	3	3,342
Δ*yen6* + P_{chi1}::lacZ + pACYC184sp	1	12,323
Δ*yen6* + P_{chi1}::lacZ + pACYC184sp	2	14,217
Δ*yen6* + P_{chi1}::lacZ + pACYC184sp	3	15,237
Δ*yen6* + P_{chi1}::lacZ + pACYC184sp-P_{yen6}-P_{yen7}	1	11,263
Δ*yen6* + P_{chi1}::lacZ + pACYC184sp-P_{yen6}-P_{yen7}	2	9,995
Δ*yen6* + P_{chi1}::lacZ + pACYC184sp-P_{yen6}-P_{yen7}	3	11,331

Table 5.9: Preliminary β galactosidase assay for pACYC184-P_{yen6}-P_{yen7} complementation of P_{chi1}::lacZ in *Yersinia entomophaga* MH96 and Δ*yen67* (single replicate).

strain	N	Miller Unit
MH96 + P_{chi1}::lacZ + pACYC184cm	1	4,589
MH96 + P_{chi1}::lacZ + pACYC184cm	2	8,108
MH96 + P_{chi1}::lacZ + pACYC184cm	3	7,186
Δ*yen67* + P_{chi1}::lacZ + pACYC184cm	1	3,053
Δ*yen67* + P_{chi1}::lacZ + pACYC184cm	2	4,040
Δ*yen67* + P_{chi1}::lacZ + pACYC184cm	3	3,593
Δ*yen67* + P_{chi1}::lacZ + pACYC184cm-P_{yen6}-P_{yen7}	1	2,873
Δ*yen67* + P_{chi1}::lacZ + pACYC184cm-P_{yen6}-P_{yen7}	2	2,258
Δ*yen67* + P_{chi1}::lacZ + pACYC184cm-P_{yen6}-P_{yen7}	3	2,256

Table 5.10: Preliminary β galactosidase assay for *trans* complementation of Δ*yen6* with P_{chi1}::*lacZ* translational fusion by induction of *yen67as* in pBAD under 0.02 % arabinose at 25 °C (single replicate).

strain	plasmid	N	Miller unit
MH96 + P_{chi1}::lacZ	pBAD	1	4,563
MH96 + P_{chi1}::lacZ	pBAD	2	3,641
MH96 + P_{chi1}::lacZ	pBAD	3	3,447
MH96 + P_{chi1}::lacZ	pBAD-*yen67as*	1	7,812
MH96 + P_{chi1}::lacZ	pBAD-*yen67as*	2	6,718
MH96 + P_{chi1}::lacZ	pBAD-*yen67as*	3	6,103
Δ*yen6* + P_{chi1}::lacZ	pBAD	1	14,703
Δ*yen6* + P_{chi1}::lacZ	pBAD	2	15,515
Δ*yen6* + P_{chi1}::lacZ	pBAD	3	11,721
Δ*yen6* + P_{chi1}::lacZ	pBAD-*yen67as*	1	36,343
Δ*yen6* + P_{chi1}::lacZ	pBAD-*yen67as*	2	37,634
Δ*yen6* + P_{chi1}::lacZ	pBAD-*yen67as*	3	28,877

To date, no attempts to complement the reduction in β-gal production for the P_{chi1}::lacZ fusion in Δ*yen7* have been undertaken, but there is potential to attempt complementation of this strains using pBAD-*yen7* under arabinose 0.002 %, although some reduction of growth would be expected compared to strains only carrying pBAD, potentially confounding results. *In trans* complementation of Δ*yen7* with pBAD-*yen7* remains a future area of work.

5.3.3.6.4 β galactosidase activity of MH96 *cis*-merodiploid P_{chi1}::lacZ strains over-producing Yen6 or Yen67as at 25 °C.

To validate increased Yen-TC component secretion that was observed in MH96 carrying pBAD-*yen67as* under arabinose 0.2 % at 25 °C from CS visualized by SDS-

PAGE, β-gal activity from MH96 P_{chi1}::*lacZ* reporter strain was assessed with either empty vector (i.e., pBAD), pBAD-*yen6* or pBAD-*yen67as* under the same inducing conditions. A significant increase was determined for MH96 carrying pBAD-*yen67as*, compared to strains carrying pBAD and pBAD-*yen6* (Figure 5.28), although the rate of β-gal production with pBAD-*yen67as* was highly variable among experimental replicates (Figure 5.29).

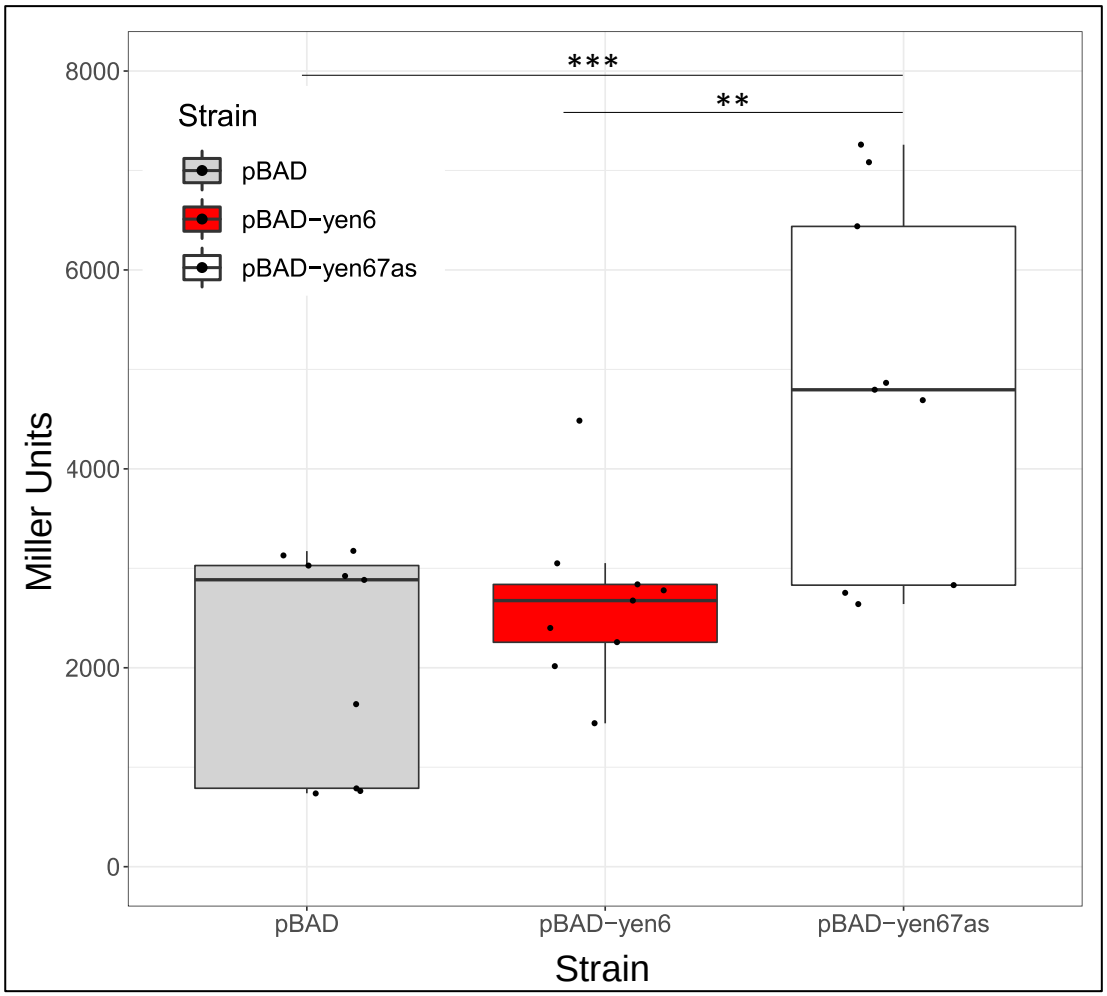


Figure 5.28: β-galactosidase production following induction of pBAD-*yen6* and pBAD-*yen67as* in *cis*-merodiploid *Yersinia entomophaga* MH96 strain harbouring P_{chi1}::*lacZ*. Rate of β-galactosidase was measured from 1:25 culture dilution in a 96-well plate at OD_{420} after 16 hours of growth at 25 °C with arabinose 0.2 % and represented as Miller Units (normalized by OD_{600}). Data were combined from three separate experiments, each containing three biological replicates/strain. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively. Horizontal bars indicate statistically significant differences: *** = adjusted P-value < 0.001, ** = 0.001 < adjusted P-value < 0.01 and * = adjusted P-value < 0.05 based on Tukey multiple comparison of means with 95 % family-wise confidence level.

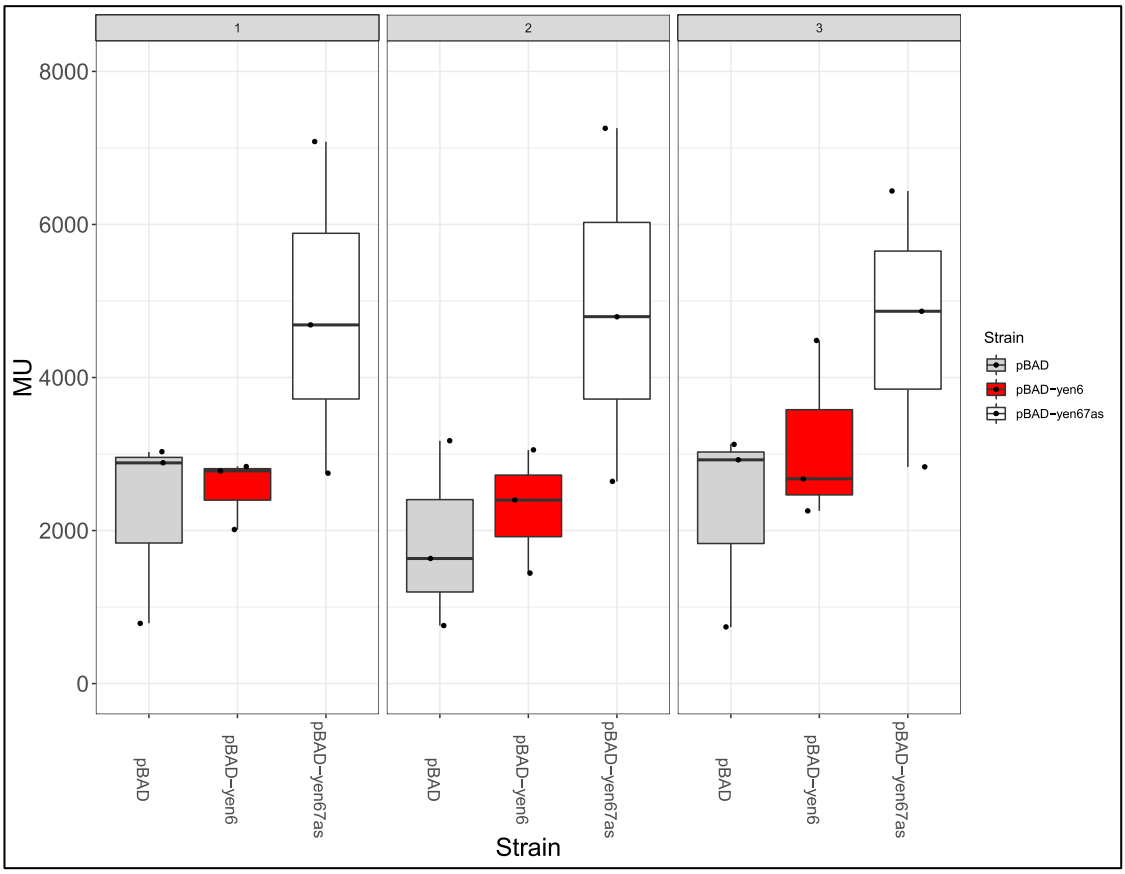


Figure 5.29: β-galactosidase production following induction of pBAD-*yen6* and pBAD-*yen67as* in *cis*-merodiploid *Yersinia entomophaga* MH96 strain harbouring P_{chi1}::*lacZ* showing results of separate experiments. Rate of β-galactosidase was measured from 1:25 culture dilution in a 96-well plate at OD_{420} after 16 hours of growth at 25 °C with arabinose 0.2 % and represented as Miller Units (normalized by OD_{600}). Data were combined from three separate experiments (facets 1, 2 and 3), each containing three biological replicates/strain. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively.

5.3.3.6.5 β galactosidase activity of MH96 *cis*-merodiploid P_{chi1}::lacZ strains over-producing Yen7 at 25 and 37 °C.

Induction of *yen7* with a low concentration of 0.002 % arabinose in MH96 was found to significantly increase enzymatic activity of the P_{chi1}::lacZ reporter strain at both 25 and 37 °C (p-values of 0.001 and 0.010, respectively) (Figure 5.30). These findings were consistent with the aforementioned SDS-PAGE analysis of the same strains grown under the same inducing conditions, that also showed the Yen-TC components and other highly secreted putative virulence factors (Peptidase M66 – PL78_05495,

Hypothetical virulence factor - PL78_18785, chitinase – PL78_11910 and chitin-binding protein – PL78_05310).

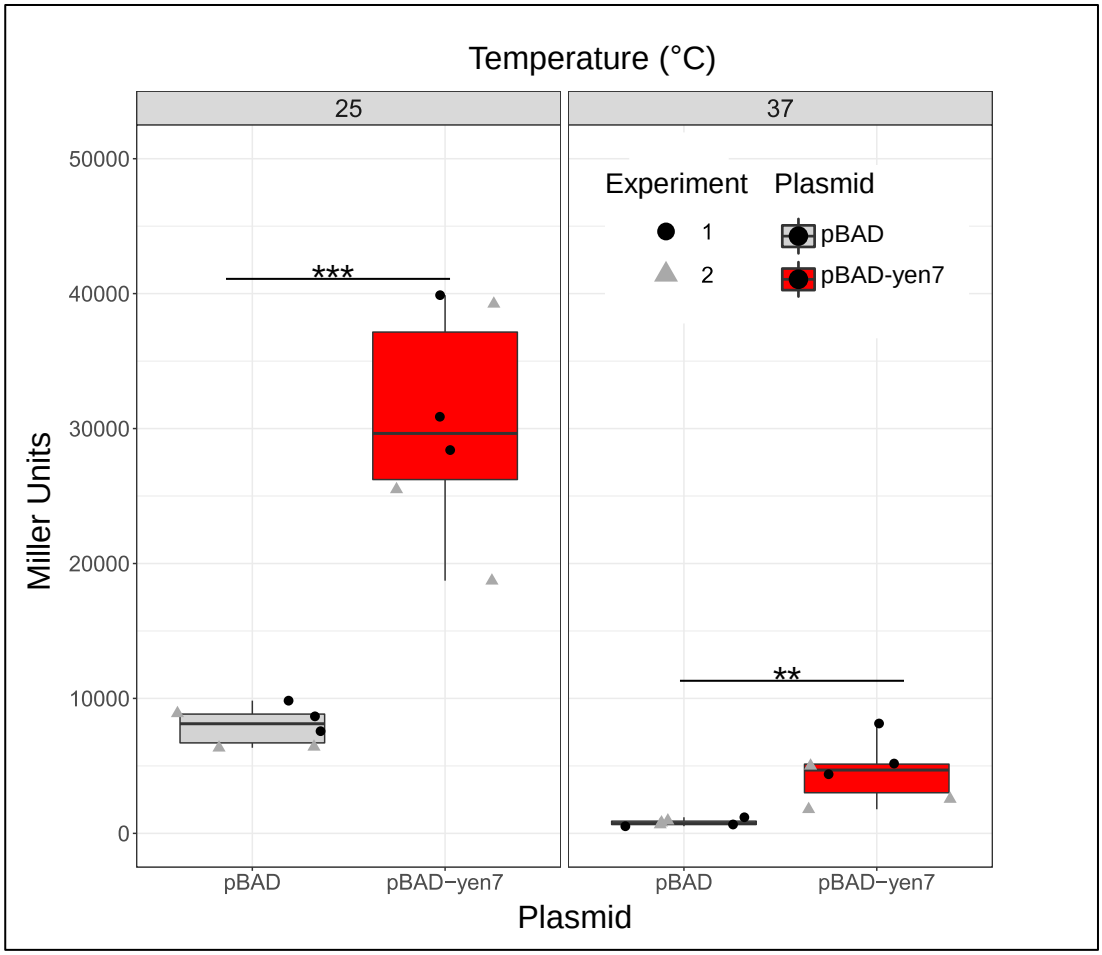


Figure 5.30: β-galactosidase production following induction of pBAD-*yen7* in *cis*-merodiploid *Yersinia entomophaga* MH96 strains harbouring P_{chi1}::*lacZ*. Rate of β-galactosidase was measured from 1:25 culture dilution in a 96-well plate at OD_{420} after 16 hours of growth at 25 °C with arabinose 0.002 % and represented as Miller Units (normalized by OD_{600}). Data were combined from two separate experiments carried out at 25 and 37 °C (facets 1 and 2), each containing three replicate flasks per strain. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively. Horizontal bars indicate statistically significant differences: *** = adjusted P-value < 0.001, ** = 0.001 < adjusted P-value < 0.01 and * = adjusted P-value < 0.05 based on Welch two-sample t-test.

5.3.3.6.6 Cumulative analysis of *yen7* over-expression on MH96 cell growth, plasmid stability, Yen-TC production/secretion and P_{chi1}::*lacZ* reporter activity at 25 and 37 °C.

The hypothetical protein Yen7 is predicted to act as a transcriptional activator of Yen-TC, so in order to test this hypothesis the effect of *yen7* over-expression on MH96 cell growth, plasmid stability, Yen-TC production/secretion and P_{chi1}::*lacZ* reporter activity at 25 and 37 °C were undertaken as part of this thesis. Over-expression of *yen7* was found to have significant impacts on MH96, observed as deleterious effects on cell growth, reduced plasmid stability, partial restoration of secretion at 37 °C, increased production of some Yen-TC/co-secreted TC-associated factors at 25 C and increased β gal activity in the P_{chi1}::lacZ reporter strain at both 25 and 37 °C. Consistent inducing conditions (0.002 % arabinose; delayed 2 hours after inoculation of 50 ml culture flasks) and sample collection time-point (~15 h post-induction) for plasmid stability, cell density, protein visualization by SDS-PAGE and β-gal assay of *cis*-merodiploid P_{chi1}::*lacZ* reporter strain permitted for cumulative analysis of the overall-effects of *yen7* over-expression on MH96 and toxin secretion at 25 and 37 °C (Table 5.11).

Table 5.11: Cumulative results from over-expression of Yen7 from pBAD under 0.002 % in *Yersinia entomophaga* (MH96)*

Experiment	Temperature (°C)	
	25	**37**
Plasmid stability (%)	17	81
Cell density (CFU/ml)	2.4×10^8	7.4×10^8
Secretion of Yen-TC and other putative VFs (SDS-PAGE of cell supernatant)	Increased secretion of Yen-TC components and other putative secreted VFs compared to wild-type with or without empty vector.	Secretion restored (wild-type with or without empty vector does not secrete).
Cellular levels of Yen-TC and other putative VFs (SDS-PAGE of cell pellet)	Increased levels of all Yen-TC components and other putative VFs compared to wild-type with or without empty vector.	Absence of Yen-TC or other co-secreted factors was observed in the cell pellet.
Enzymatic activity of *cis*-merodiploid P_{chi1}::lacZ strain	Over-expression of y*en7* resulted in significantly increased β-gal production	Over-expression of y*en7* resulted in significantly increased β-gal production strain.

*50 ml flask cultures were grown for 2 hours; 0.002 % arabinose was used to induce expression of yen7, and samples were collected after ~ 15 h.

Protein visualization and β-gal assay of the *chi1::lacZ* reporter strain both provided highly repeatable evidence supporting Yen7 as an activator of Yen-TC production at 25 °C. At lower temperatures greater levels of Yen-TC were produced as well as secreted under inducing conditions and under these same conditions significantly greater enzymatic activity was measured from the P_{chi1}*::lacZ* reporter strain, which is consistent with the hypothesis that Yen7 interacts with the promoter region of *chi1* to activate transcription of the downstream Yen-TC operon. The protein analysis also supports a more global function for Yen7 in activating the production of Yen-TC and perhaps other co-secreted exoenzymes, like peptidase M66, hypothetical VF (PL78_18785) and chitinase and possibly secretion as well. Visualization of cell supernatant at 37 °C under inducing conditions very clearly and repeatably shows Yen7 (except for Yen-TC components) can restore secretion in MH96 at this temperature. This is a key finding because MH96 does not usually secrete any proteins/toxins at 37 °C (apart from a large molecular weight filamentous hemagglutinin) but produces copious amounts of proteins/toxins at temperatures lower than 25 °C, a phenomenon that is very likely important to its entomopathogenic lifestyle. Using the β-gal assay of the *chi1::lacZ* reporter strain, increased enzymatic activity was also confirmed under inducing condition at 37 °C, albeit to a lesser degree than when the experiment was conducted at 25 °C, further supporting the finding that Yen7 can activate Yen-TC production and secretion at 37 °C as well as 25 °C. These findings also indicate complex regulatory mechanisms are likely controlling Yen-TC expression.

While induction of *yen7* under 0.002 % arabinose did result in slight reduction in cell density of the cultures, the final cell density was similar across both temperatures that were tested, but a clear difference in plasmid stability was observed, where at 25 °C there was a strong selection against pBAD-*yen7* in MH96, compared to 37 °C. Under these same conditions, the empty pBAD vector is 100 % maintained in MH96 at both 25 and 37 °C. Another salient difference between induction at different temperatures was observed from protein analysis of the CPs; at 25 °C abundant levels of Yen-TC and other putative VFs (especially peptidase M66 and hypothetical VF (PL78_18785) were clearly observed, however no obvious Yen-TC or other putative VFs protein bands were found in CPs from 37 °C. It is suspected that the major difference in plasmid

stability between 25 and 37 °C is linked to the presence of large amounts of toxins in the cellular fractions at 25 °C only, attributing to increased selection pressure against pBAD-*yen7* at 25 °C under inducing conditions compared to the higher temperature. Furthermore, the absence of Yen-TC components in the CP from 37 °C and the obvious presence of toxin components in the CS taken from the same culture flasks suggests possible density-dependent lysis mechanism may be present in MH96.

5.3.3.6.7 β galactosidase activity of MH96 *cis*-merodiploid P_{yen6}::lacZ strains *in vivo* at 25 and 37 °C.

Significantly higher transcription of *yen6* at 37 °C compared to 25 °C was observed from both *in vitro* and *in vivo* RNA-seq expression profiles (Chapter 4, section 4.3.8, Figure 4.20, Table 4.13); however, validation experiments using a P_{Yen6}::*lacZ* translational fusion strain found the opposite result for *in vitro* growth conditions, as there was significantly greater β gal activity produced by the P_{Yen6}::*lacZ* reporter strain at 25 compared to 37 °C *in vitro* (Figure 5.25). As a follow-up validation experiment, we assessed the P_{Yen6}::*lacZ* reporter strain in *G. mellonella* and found that during infection of *G. mellonella* the *cis*-merodiploid strain harbouring P_{Yen6}::lacZ has significantly higher enzymatic activity (> 39x fold-change) at 37 °C compared to 25 °C (Figure 5.31). These results provide evidence of host-dependent post-transcriptional regulation of *yen6*, in addition to the temperature-dependent transcriptional differences noted from the transcriptome data.

It should be noted that the results of the *in vivo* β-gal assays were found to be especially variable from experiment 2 and 3, which were collected from host *G. mellonella* that had been maintained in the laboratory for a prolonged period of greater than three weeks (typically *G. mellonella* would be used within one week for experimental purposes but access to hosts was limited at this time). The raw data, including OD_{600} values are provided in the supplementary section (Table S16). For the combined data analysis, a single individual from experiment two with unexpectedly low absorbance for OD_{600} was omitted as was the entire experiment 3, which suffered from numerous hosts reporting unexpectedly low OD_{600} measurements as wells as negative controls reporting positive MUs (negative controls should always generate negative MUs,

perhaps suggesting some pipette error). Though we note that biological insect systems are variable, it would be beneficial to repeat the experiment again using more standardized hosts, ensuring consistent age, time on lab-based artificial diet, size and weight as this assay was found to be highly variable. This is a logistical challenge associated with working pet food/bait suppliers, that often cannot guarantee size/age of hosts at time of delivery.

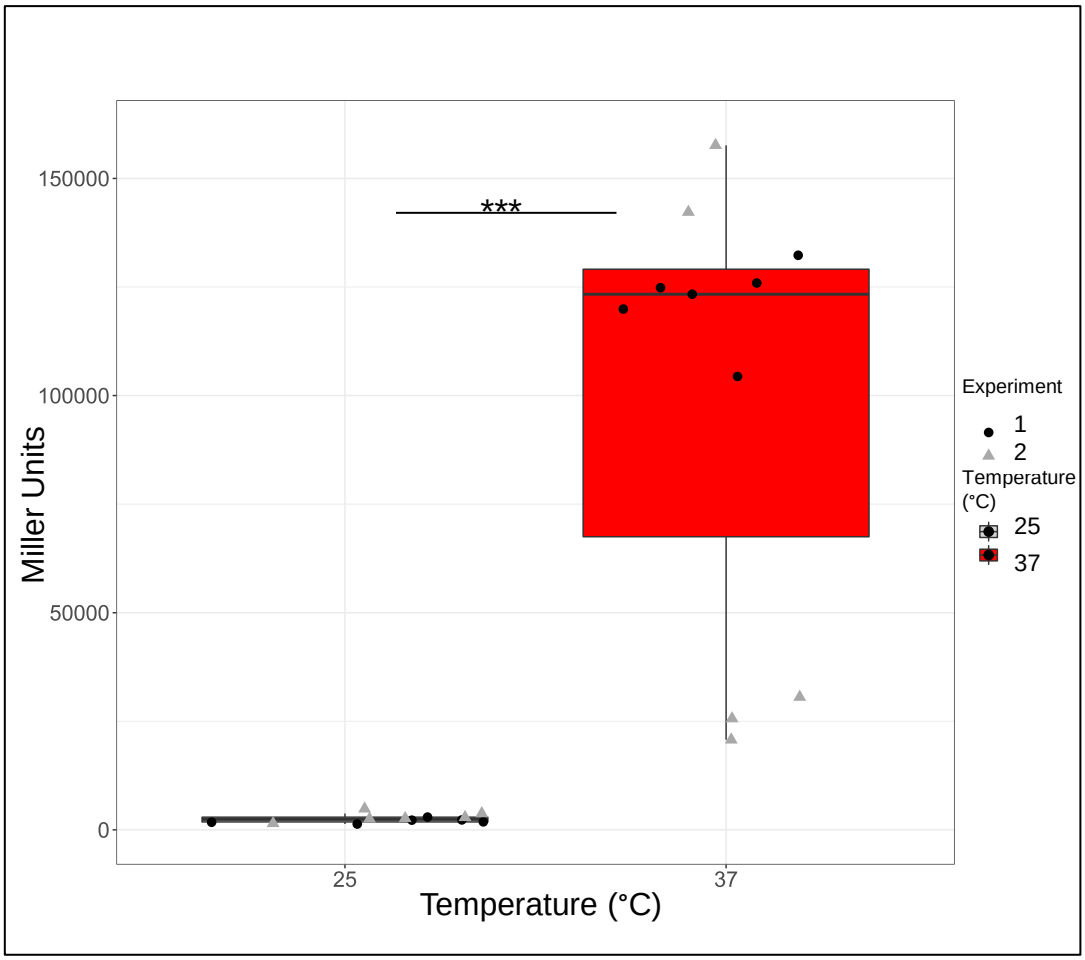


Figure 5.31: β-galactosidase production in *cis*-merodiploid *Yersinia entomophaga* MH96 strains harbouring P_{yen6}::*lacZ* during infection of *Galleria mellonella* at 25 or 37 °C. Rate of β-galactosidase was measured from 1:25 dilution of hemolymph sample in a 96-well plate at OD_{420} after 19-22 or 16-17 h post infection at 25 or 37 °C, respectively. Enzymatic activity is represented as Miller Units (normalized by OD_{600}). Data were combined from two separate experiments, each containing six biological replicates/temperature. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively. Horizontal bars indicate statistically significant differences: *** represents a P-value less than 0.001 based on Welch two-sample t-test.

5.4 Discussion

Initial phenotypic investigations of MH96 mutant strains by visualization of secreted protein and virulence testing in the intrahemocoelic *G. mellonella* infection model determined only few selected mutants showed a phenotype at the protein level (i.e., Δ*cspA123*/ΔHCUI$_{YE96}$, Δ*cbpA* and Δ*rovA*) and/or an attenuation in virulence (i.e., Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6*). Based on these results, and previously reported *in*

vivo RNA-seq findings determining interesting host- and temperature-dependent expression patterns for both PAI_{YE96}-encoded putative regulators, *yen6* and *yen7* (Chapter 4, section 4.3.8, Figure 4.20, Table 4.13) further investigations focused on determining whether *yen6* and *yen7* participate in the co-ordinated regulation of primary virulence determinant, Yen-TC were undertaken as a primary focus of this chapter. The outcome of these investigations expands our understanding the contribution of *yen6* and *yen7* in the regulation of virulence in MH96, but also importantly unexpected phenotypic findings associated with over-expression of *yen6* inspired new and interesting research initiatives, inspiring further characterization of the *yen6* regulon reported in Chapter 6 of this thesis.

5.4.1 Revealing the secreted weaponry of MH96 – considering SDS-PAGE and transcriptome findings

Previously unpublished work by S. Jones (AgResearch Ltd.) characterized LC-ESI-MS/MS many of the highly secreted proteins (in addition to the high molecular weight Yen-TC components) from the supernatant of MH96 grown in rich culture broth at 25 °C, which included 81 kDa peptidase M66 (PL78_04595), 33 kDa chitinase (PL78_11910), 23 kDa CbpA (LCMO/CBP - PL78_05310) and 44 kDa hypothetical VF (PL78_18785) (a modular-type protein containing putative cellulose-binding and intimin domains). Gene clusters encoding all of these secreted proteins were also previously identified as sharing identical expression profiles with the Yen-TC genes under *in vitro* and *in vivo* growth conditions at 25 and 37 °C (Chapter 4, section 4.3.8, Table 4.13). Notably, expression of the Yen-TC genes and the above-mentioned co-secreted factors were shown to be significantly repressed at 37 °C during *in vivo* conditions only. In this respect, comparison of protein secretion and the *in vivo* transcriptome of MH96 at 25 and 37 °C both identify the same core-set of putative VFs, which are all under dynamic temperature- and host-dependent regulation and this result was considered a key validating factor for the transcriptome.

5.4.1.1 Chitinolytic TC-associated factors are co-expressed and co-secreted with Yen-TC

Synergism of Yen-TC with chitinolytic factors would be considered especially effective for degradation of peritrophic membrane (PM) which is comprised of chitin fibrils attached to glycoproteins and proteoglycans and forms a physical barrier between the gut lumen and the midgut epithelial (Erlandson et al., 2019; Kelkenberg et al., 2015). In this model, chitinolytic factors act to degrade the PM, which exposes the underlying midgut cells to Yen-TC, then Chi1, Chi2 and YenA components bind specific receptors located on the surface of midgut epithelial cells (Landsberg et al., 2011; Piper et al., 2019), which initiates pore-formation (Gatsogiannis et al., 2016; Meusch et al., 2014) and intoxication of the cells (Busby et al., 2013b) ultimately resulting in dissociation of the midgut (Landsberg et al., 2011; Marshall et al., 2012). Other co-secreted factors, such as M66 peptidase and the hypothetical intimin/carbohydrate-binding domain contain protein would also be predicted to act synergistically with Yen-TC, chitinase and CbpA to perhaps aid by interacting with proteoglycan and glycoproteins associated with the PM?

Chitin also represents an abudnant nutrient resource within the insect host, as well as more generally in the biosphere, as a key component of arthropod exoskeleton and fungi cell wall, making it the second most abundance biopolymer in nature (Gooday, 1990; Souza et al., 2011) and diverse bacteria have evolved the ability to degrade this especially recalcitrant crystalline polysaccharides by syngergistic activity of chitinolytic enzymes (Beier and Bertilsson, 2013; Vaaje-Kolstad et al., 2013). Bacterial chitinases typically include glycosyl hydrolases (GHs) within the GH18 family based on the CAZy classification (Lombard et al., 2014) and can be broadly classified as either endochitinases that hydrolyse random internal β1-4-glycosidic bonds to release low-molecular weight multimeric units of β-1,4-*N*-acetylglucosamine (GlcNAc) or exochitinases that target the non-reducing end of the chitin polymer, either releasing mono or disaccharide GlcNAc sugar molecules (Adrangi and Faramarzi, 2013; Beier and Bertilsson, 2013; Berini et al., 2018). Furthermore, many chitinolytic bacteria produce chitobiase (or *N*-acetylhexosaminidase), which can hyrdolyse the β1-4-glycosidic bond in $(GlcNAc)_2$ and are classified within the GH20 family (Toratani et al., 2008; Uchiyama et al., 2003). Another important group of chitin degrading enzymes are

the LCMOs/CBPs that belong to the AA10 family (formerly the CBM33 family) (Aachmann et al., 2012; Lombard et al., 2014). Remarkably, LCMOs/CBPs have been shown to catalyse the depolymerization of chitin chains by an oxidative reaction (Vaaje-Kolstad et al., 2010) requiring copper (Aachmann et al., 2012) and can act synergistically with chitinases in the degredation of chitin is *Serratia* spp. (Manjeet et al., 2013; Purushotham et al., 2012) by increasing substrate accessibility for chitinases (Vaaje-Kolstad et al., 2005b).

Like MH96, most *Serratia* species are known to be proficient secretors, which are widely known to use potent lytic enzymes as VFs in immune resistance, invasiveness, and destruction of competitors (Petersen and Tisa, 2013). *S. marcescens* is known to be a proficient secretor as well, and was shown to effiently degrade chitin by secretion of a multi-component chitinolytic machinery (Suzuki et al., 1998; Vaaje-Kolstad et al., 2005a, 2013) consisting of three GH18-family chitinases (ChiA, B and C1) (Suzuki et al., 2002), a GH20-family chitobiase (Toratani et al., 2008; Uchiyama et al., 2003) and the LCMO/CBP, CBP21 (Suzuki et al., 1999). Similar to *S. marcescens*, MH96 also secretes an array of known and putative chitinolytic enzymes into the supernatant at temperatures of 25 °C and lower, including three GH18 family chitinases (i.e., Yen-TC components: Chi1 and Chi2 as well as PL78_11910) and CbpA, which is an LCMO/CBP that shares homology with CBP21 of *S. marcescens*.

Previous work functionally characterized Yen-TC-associated chitinases (Chi1 and Chi2) as important to the activity of Yen-TC holotoxin against insect mid-gut barrier, including involvedment in both host-cell adhesion and endochitinase activities (Busby et al., 2012; Landsberg et al., 2011; Piper et al., 2019). Furthermore, Chi1 was demonstrated to be especially critical to the functionality of Yen-TC secretion, with a structure consisting of a foot and leg that is involved in pore formation (Piper et al., 2019), a dynamic process by which Yen-TC holotoxin can translocate the YenC-Rhs toxic payload into host cells upon binding of host cell ligands (Busby et al., 2013b; Meusch et al., 2014). Screens using high-throughput glycan microarry for Chi1 and Chi2 identified galactose, glucose, *N*-acetylgalactosamine (GalNAc) and GlcNAc as significant binding motifs of these chitinases, further implicating Chi1 and Chi2 as multifunctional Yen-TC components with lectin-like properties (Piper et al., 2019). The

apparent divergence between the alkaline pH of the mid-gut of the host *C. giveni* (Biggs and McGregor, 1996) and optimum pH for Chi1 (~ 6) and Chi2 (~ 5) endochitinase activity (Busby et al., 2012) has been suggested by Piper et al. (2019) as another indicator that Chi1 and Chi2 may primarily function to bind glycophosphlipids, known to be prevalent on the surface of insect host cells (Aoki and Tiemeyer, 2010; Kim et al., 2003a; Walski et al., 2017a, 2017b), thus playing more of a primary role in host-cell binding and toxin translocation, rather than degrading β1-4-glycosidic bond to acquire GlcNAc for metabolism.

Given that MH96 is known to co-secrete a third chitinase, as well as CbpA (a LCMO/CBP), along with Yen-TC, perhaps endochitinase activity of Chi1 and Chi2 are not essential to MH96's ability to survive on GlcNAc as a sole carbon source at 25 °C (Chapter 4, section 4.3.9, Table 4.14). In order to test this idea, the optimal pH for the third chitinase and CbpA should be investigated, as these enzymes may have evolved optimum chitinolytic activity for alkaline conditions known from insect gut (Biggs and McGregor, 1996) and likely contribute to both virulence and metabolism of MH9 within the insect host.

With respect to biopesticide development, the Yen-TC chitinolytic co-secreted factors of MH96 represent very promising options for additives to pre-established products. Early work by Regev et al., (1996) showed that a chitinase from *S. marcescens* increased the insecticidal effects of *B. thuringiensis* delta-endotoxin CryIC against *Spodoptera* (*Sp.*) *littoralis*. In another study, a chitinase from tobacco, *Nicotiana tabacum*, was found to provide a synergistic effect when combined with recombinant *cry*1Ac in *B. thuringiensis* against *Helicoverpa armigera* (Ding et al., 2008). More recently, a chitinase from fungi *Penicillium orchrochloron* were shown to be an effective insecticidal agent against *H. armigera* (Patil and Jadhav, 2015). While the major focus of this work was not directly focused on the chitinolytic activities of MH96, the transcriptome analysis and subsequent comparisons with secreted proteins provided an effective method to identify a few select targets for furture research.

5.4.1.2 Other TC-associated factors are co-expressed and co-secreted with Yen-TC and chitinolytic enzymes

Two other secreted exoenzymes, M66 peptidase and putative intimin/carbohydrate binding protein, were also both found to be co-expressed/co-secreted with Yen-TC and TC-associated factors and are highly suspected to also act as key VFs within the insect host as well. Peptidase M66 contains a conserved zinc metalloprotease domain, sharing similar sequence homology to an important VF from *E. coli* O157:H7 called StcE (Lathem et al., 2002; Yu et al., 2012). Like the peritrophic membrane (PM) in insects, the mammalian gut is lined with a mucosal barrier consisting of heavily glycosylated mucin glycoproteins that protect the underlying epithelium from infection (Pelaseyed et al., 2014). To this end, enterohemorrhagic *E. coli* secretes StcE, which acts to hydrolyse the glycoproteins of the mucosal layer giving the pathogen access to penetrate the underlying epithelium (Lathem et al., 2002). *V. cholera* also secretes orthologous TagA which also is known to cleave mucin glycoproteins (Szabady et al., 2011). Based on the predicted protein structure, peptidase M66 is anticipated to perform a similar function against the PM in insects, and activity which may enhance Yen-TC intoxication of underlying epithelial cells, solidifying a possible motive driving co-expression/co-secretion of peptidase M66 with Yen-TC and TC-associated factors, but this hypothesis remains to be tested.

Finally, the putative intimin-containing adhesin, shares similar properities with the peptidase M66, in that it is likely involved into adhering to host cells through recognition of a carbohydrate-containing host ligand. Based on predicted secondary structure, this secreted protein can be expected to contain an intimin domain and a carbohydrate binding region but the precise role of this putative VFs remains currently unknown. Since both of these secreted exoenzymes have been clearly linked to the secretion of Yen-TC and other chitinolytic enzymes by MH96, both M66 peptidase and the putative intimin/carbohydrate-binding protein represent novel enzymes that likely enhance pathogenesis of MH96 against insects, possibly by interacting with glycoligands associated with the surface of the PM to increase access for co-secreted chitinolytic enzymes to access the chitin-rich fibril component of the PM.

5.4.2 Production and secretion of Yen-TC and TC-associated factors is temperature-dependent in MH96

Similar to Hurst *et al.* (2011), here MH96 was also shown to have greatly reduced protein secretion at 37 compared to 25 °C in LB broth, evidenced by notably fewer proteins in the CS at 37 °C, specifically including Yen-TC components and the co-secreted TC-associated factors. Important temperature-dependent effects on Yen-TC secretion and also virulence against *G. mellonella* have already been described before (Hurst et al., 2011a, 2015), indicative that Yen-TC and co-secreted TC-associated factors are the primary offensive artillery used by MH96 to attack insect hosts and thermoregulatory mechanisms have evolved to shut off this system at mammalian host temperatures.

Thermoregulation of important VFs from other pathogenic *Yersinia* spp. have also been demonstrated before. Expression of insecticidal TC genes *tcaA, tcaB, tcaC and tccC* in *Y. enterocolitica* strain W22703 is strictly under temperature-dependent regulation, such that all *tc* genes are silenced at 37 °C and maximally expressed at 10 °C *in vitro* (Bresolin et al., 2006). The tight thermoregulation of the *tc* genes in W22703 was shown to involve a thermal-labile regulator, TcaR2, which at lower temperatures can out-compete the H-NS/YmoA complex to activate the *tc* genes, but at higher temperatures was subjected to increased proteolysis resulting in suppression of the *tc* genes by H-NS/YmoA at 37 °C (Starke and Fuchs, 2014; Starke et al., 2013). Furthermore, optimal induction of the chromosomally encoded Ysa-T3SS in *Y. enterocolitica* biovar 1B was also found at temperatures of 26 but not 37 °C, and similar to MH96 secretion of Yen-TC components, the Ysa-T3SS effectors YspA, YopN, YopP and YopE could be easily visualized in the CS by SDS-PAGE when grown under Ysa-T3SS inducing conditions (i.e., 26 °C in nutrient broth with high sodium chloride concentrations) (Venecia and Young, 2005).

The effect of temperature on MH96 toxin secretion was found to be very consistent among experiments; however, this was not the case for the presence/absence of Yen-TC toxins and co-secreted factors in the cellular fraction at 37 °C, which was found to be variable between experiments (i.e., sometimes the toxin bands were visible but other times not). Earlier, Hurst et al. (2011) showed the CP protein profiles from MH96

cultured at 30 and 37 °C contained relatively reduced levels of Yen-TC components compared to 25 °C, but the sonicated filtrates of cells cultured at higher temperatures still remained toxic against *C. giveni per os* while the CS from the same cultures did not suggesting a temperature-mediated secretion mechanism. In contrast, the results presented in this thesis found that the characteristic toxin bands were usually not visible in the protein profiles of MH96 cells grown at 37 °C, especially when care was taken to promptly collect and store the CP samples without letting the culture rest at room temperature for extended periods of time. Considering that no significant difference in the expression of the Yen-TC and associated factors was identified from cultures grown at 25 or 37 °C using RNA-seq (i.e., data suggest Yen-TC components and associated factors are expressed equally at both temperatures *in vitro*), the variability in corresponding protein levels at 37 °C is speculated to involve some yet unknown post-transcriptional regulatory mechanism.

Indeed, post-transcriptional regulation of many VFs produced by *Yersinia* species has been reported, including regulation by *cis*- and *trans*-acting ncRNAs, thermoswitches/riboswitches, RNA-binding proteins and RNases (reviewed in Schiano and Lathem, 2012, Righetti et al., 2016). For example, an important AraC-family regulator LcrF, which is the activator of the plasmid encoded Ysc-Yop-T3SS (a key VF among in human-pathogenic *Yersinia* spp.) contains a thermoswitch in the 5′ region of its messenger RNA (Böhme et al., 2012; Hoe and Goguen, 1993). At low temperatures the 5′ region forms complex stem-loop structures that obscures the Shine-Dalgarno sequences preventing translation initiation, but at 37 °C the secondary structure denatures permitting translation of LcrF and subsequent Ysc-Yop-T3SS activation at mammalian body temperatures.

Compared to other regulatory mechanisms, post-transcriptional regulatory mechanisms may provide a more rapid way for pathogenic bacteria to adjust to environment changes and quickly fine tune the production of energetically expensive VFs, such as toxins and complex secretion machinery (Schiano and Lathem, 2012). A key finding of previously reported transcriptome and now proteomic investigations of MH96 at 25 and 37 °C suggests that temperature-dependent post-transcriptional regulatory mechanisms have evolved in MH96; however, the exact mechanism(s)

remain to be elucidated. Spontaneous point mutation in the RNA chaperone, *hfq*, results in a rough variant of MH96 that does not secrete or produce Yen-TC (Hurst et al. 2019a; in prep), which strongly implicates a global role for some yet uncharacterized ncRNA regulatory network in regulation of Yen-TC production and secretion, but the role of small ncRNAs networks is considered beyond the scope of this thesis.

5.4.3 MH96 secretes filamentous hemagglutinin N-terminal containing protein at 37 °C *in vitro*

MH96 is known to have greatly reduced secretion at 37 compared to lower temperatures *in vitro* (Hurst et al., 2011), (Chapter 2, section 2.3.3, Figure 2.3); however, at 37 °C a large molecular weight protein (~175 kDa) was found to be consistently secreted into the culture supernatant. LCMS-ESI-MS/MS methods were used to identify this protein as filamentous hemagglutinin N-terminal containing protein (also annotated as a hemolysin), which is encoded by PL78_04365 and this putative cell-surface adhesion is likely secreted by cognate two-partner secretion (TPS)/activation protein ShlB/FhaC/HecB encoded by PL78_04360 (Leyton et al., 2012). Filamentous hemagglutinins have been shown to be key VFs among diverse pathogenic bacteria, including *Bordetella* (*Bo.*) *pertussis*, *A. baumannii*, *Erwinia chrysanthemi* and *Ps. fluorescens* (Darvish Alipour Astaneh et al., 2014; Rojas et al., 2002; Scheller and Cotter, 2015; Sun et al., 2016). Filamentous hemagglutinins produced by *Bo. pertussis* are known to be secreted and contribute to adherence to host cells (Melvin et al., 2015; Serra et al., 2011), agglutination (Menozzi et al., 1994) and immune-modulation of host cells (Henderson et al., 2012; Inatsuka et al., 2005; Julio et al., 2009).

The filamentous hemagglutinin N-terminal containing protein that is secreted at 37 °C by MH96 has significant sequence similarity (E-value of 0.0) to ShlA of *S. marcescens* (ALE98162.1), with 55.75 % identity over 99 % of the sequence. ShlA is a large '*Serratia*-type' hemolysin that is secreted by means of TPS (Poole et al., 1988), but homologous toxin/translocator pairs have been identified from other pathogenic bacteria, including PhlA/PhlB from *P. luminescens* and YhlA/YhlB from *Y. ruckeri.* While, *phlA* was not found be a major virulence determinant of *P. luminescens* against *Sp. littoralis* (Brillard et al., 2002), both *yhlA* and *yhlB* were both found to be involved in virulence of *Y.*

ruckeri (Fernández et al., 2007) and numerous studies have associated *shlA* with pathogenicity in *S. marcescens* (González-Juarbe et al., 2015; Kurz et al., 2003; Lin et al., 2010; Di Venanzio et al., 2017)

The filamentous hemagglutinin N-terminal containing protein that is secreted at 37 °C by MH96 is considered a putative secreted toxin that is under thermoregulation at the transcriptional level. Previously reported *in vivo* RNA-seq results identified significantly higher expression of PL78_04365 at 37 compared to 25 °C, under both *in vitro* and *in vivo* conditions during exponential growth phases (Chapter 4, section 4.4.4.2, supplemental Table S9). These results support a strictly temperature-dependent transcriptional regulatory mechanisms for both secreted toxin and TPS which were predicted to be expressed in an operon. While MH96 encodes another filamentous hemagglutinin N-terminal containing protein (PL78_11060), the expression of this gene was not found to have a strong response to increased temperature but was found to have high expression at 25 °C during early infection (Chapter 4, section 4.3.7.3, Figure 4.17).

5.4.4 Phenotypes identified for Δ*cspA123*/ΔHCUI$_{YE96}$, Δ*cbpA* and Δ*rovA* mutant strains by SDS-PAGE

Since MH96 is known to secrete large amounts of Yen-TC components as well as Yen-TC-associated factors in a temperature-dependent manner (Hurst et al., 2011) (Chapter 2, section 2.3.3, Figure 2.3), visualization of the proteins in CS from mutant strains by SDS-PAGE provides a convenient method to screen for phenotypes related to secretion and Yen-TC production in MH96. Out of the eight mutant strains screened in this study, only three (Δ*cspA123*/ΔHCUI$_{YE96}$, Δ*cbpA* and Δ*rovA*) were identified to influence either global secretion or secretion of individual Yen-TC components or associated factors. The Δ*cspA123*/ΔHCUI$_{YE96}$ strain had the most conspicuous phenotype, with complete shut-down of global secretion at 25 °C compared to wild-type; however, transcriptome analysis of the Δ*cspA123*/ΔHCUI$_{YE96}$ strain revealed a 17.5 kb excision (Chapter 6, section 6.3.1.4, Table 6.7) of a unstable genomic element, HCUI$_{YE96}$ that became the focus of a different chapter of this thesis (i.e., Chapter 7) and later work showed the non-secreting phenotype of Δ*cspA123*/ΔHCUI$_{YE96}$ was solely attributed to

the excision of HCUI$_{YE96}$ and not actually a result of deletion of *cspA123* (Chapter 7, section 7.3.2, Figure 7.9).

In addition to absence of secretion at 25 °C, Δ*cspA123*/ΔHCUI$_{YE96}$ was observed to have two other phenotypic traits at the protein-level; First, the characteristic protein bands representing Yen-TC components as well as other co-secreted TC-associated factors (i.e., peptidase M66, chitinase, CbpA and hypothetical VF) were absent from the CP of Δ*cspA123*/ΔHCUI$_{YE96}$ grown at 25 °C, such that the protein profile of this strain more resembled MH96 grown at 37 °C, which is typically found to have reduced amounts of Yen-TC and TC-associated factors in the CP as well. Second, a small molecular weight protein was observed in the cell pellet of Δ*cspA123*/ΔHCUI$_{YE96}$ grown only at 37 °C, that was not present in CP from MH96. This protein was identified as ribosomal protein L6, which is known to be essential for the assembly of functional 50S rRNA subunits at the late state in *E. coli* (Shigeno et al., 2016). In *B. subtilis*, L6 was shown to be properly positioned by an essential GTPase on the 23S rRNA prior to incorporation of L16 and other late assembly ribosomal proteins (Gulati et al., 2014). It currently remains unclear why excessive amounts of L6 are produced in the Δ*cspA123*/ΔHCUI$_{YE96}$ grown at 37 °C and whether this phenotypic trait can be attributed to the *cspA123* deletion or excision of HCUI$_{YE96}$.

The only strains with phenotypes observed by SDS-PAGE, Δ*cbpA* and Δ*rovA,* were found to have comparatively more subtle phenotypes. As expected, the Δ*cbpA* mutant did not produce or secrete the previously-identified 23 kDa chitin-binding protein, confirming the successful deletion of the *cpbA* gene. *RovA* was found to influence the secretion of just one Yen-TC component, YenA2, which was found to be absent from the CS of Δ*rovA*. Compared to wild-type, Δ*cbpA* and Δ*rovA* were not found to have attenuated virulence in the *G. mellonella* intrahemocoelic model, despite the lack of secretion of CbpA and YenA2, respectively and subsequently *rovA* and *cbpA* were not carried further in this work despite a notable phenotype at the protein level from the targeted gene deletions in both these strains.

5.4.5 Exploring potential temperature- and host-dependent regulatory linkages between *yen6*, *yen7* and Yen-TC components.

5.4.5.1 Protein production and toxin secretion by Δ*yen6*, Δ*yen6yen7* and Δ*yen7* was consistent with wild-type MH96

In order to investigate potential linkages between putative regulators *yen6* and *yen7* and temperature-dependent production and secretion of Yen-TC, protein profiles produced by MH96-derivative mutants strains lacking *yen6* and *yen7* were compared to wild-type. By visualizing proteins from cell pellet and supernatant by SDS-PAGE, it was evident that deletion of *yen6*, *yen7* and the entire *yen6-yen7* region had nil effects on global protein production or secretion. These results suggest that both *yen6* and *yen7* do not have any major regulatory control over the expression of the Yen-TC by MH96 *in vitro* at 37 and 25 °C; however, later investigations using a translational *lacZ*-reporter fused to the promoter region of *chi1*, revealed possible subtle regulatory effects on β-gal production in the Δ*yen6*, Δ*yen6yen7* and Δ*yen7* strains were found to be significantly different compared to wild-type and will be discussed further below in section 5.4.8.6.

The notable lack of phenotype by SDS-PAGE observed for Δ*yen6*, Δ*yen6yen7* and Δ*yen7* may also be an indicator of additional interacting factors contributing to the regulation of toxin production/secretion in MH96. Previously reported transcriptome data supports the presence of both temperature- and host-dependent regulatory mechanisms governing the expression of genes for Yen-TC and TC-associated factors (Chapter 4, section 4.3.8, Table 4.13), as does the disparity between Yen-TC/TC-associated factor gene expression and Yen-TC component/co-secreted protein levels at 37 °C under *in vitro* conditions (discussed above, section 5.4.2). Furthermore, some additional MH96 mutants under investigation in the Hurst lab, including a spontaneous *hfq* deficient (Hurst et al., 2019a *in prep*) and Δ*ymoA* strains (Hurst, et al., 2019b *in prep*), are also known to have phenotypes linked to global regulation of Yen-TC production and secretion, which supports the possibility of multi-layered interacting processes governing the regulation of a consortia of primary secreted VFs in MH96.

In addition to the above mentioned global regulators, *hfq* and *ymoA*, another possible Yen-TC regulator is the HCUI$_{YE96}$-encoded PhoB-like regulator (PL78_17385), that is associated with the holin-endolysin secretion system module encoded within HCUI$_{YE96}$ (Chapter 7, section 7.3.1, Figure 7.1). While this putative regulator shares significant protein sequence homology to *yen7*, suggestive of functional redundancy in Yen-TC regulation, the specific roles of PL78_17385 in the temperature-dependent regulation of Yen-TC was considered out of scope of this thesis and research and was captured under the umbrella of another separate project. Further to this, a complete absence of secretion and Yen-TC/TC-associated factor production in the ΔHCUI$_{YE96}$ strain (reported in Chapter 7) provided even more support that HCUI$_{YE96}$ is connected to the regulation of Yen-TC and TC-associated factors.

It would not be considered unusual for complex temperature-dependent regulatory mechanisms to have evolved in MH96, especially since secretion and production of Yen-TC/TC-associated factors seems to comprise a core secreted virulence machinery. By comparison, arguably the most important VF in *Y. pseudotuberculosis*, the virulence plasmid pYV-encoded Ysc-Yop T3SS, is known to be transcriptionally regulated by two different temperature-sensitive mechanisms; First, mRNA of the Ysc-Yop T3SS transcriptional activator *lcrF* (an AraC-like regulator otherwise known as *virF*) includes a *cis*-acting thermo-sensitive riboswitch that permits ribosomal binding of ribosomal site (RBS) at 37 °C but not 25 °C and second, temperature-dependent repression of *lcrF* by heat-labile YmoA (nucleoid-associated regulator that is known to interact with H-NS) which is less stable at 37 compared to 25 °C (Böhme et al., 2012). YmoA was also shown to be involved as a heat-labile repressor of insecticidal TC expression in *Y. enterocolitica* W22703 (Starke and Fuchs, 2014), in addition to *trans*-acting TcaR1 and TcaR2 (Starke et al., 2013), which together form a complex temperature-dependent regulatory mechanism.

5.4.6 MH96 is extremely virulent against *G. mellonella* by intrahemocoelic injection at 25 °C

Overall the LD$_{50}$ estimated for MH96 was found to be extremely low, on average less than a single cell. These results, albeit highly variable, were taken to mean that only a

single cell of MH96 is required to kill 100 % of *G. mellonella* within 5 days after injection. The extreme virulence of MH96 in this infection model made it difficult to detect mutant phenotypes at 25 °C and not one deficient mutant strain that was found to have attenuated virulence compared to wild-type at this temperature. The mutants that were tested included strains deficient for *cspA123* and HCUI$_{YE96}$ region excised (Δ*cspA123*/ΔHCUI$_{YE96}$), T6SS sheath component (Δ*vipB*), Yen-TC/TC-associated LCMO/CBP (Δ*cbpA*), usher/chaperone fimbrial protein (Δ*fim1)* and putative regulator (Δ*rovA*).

Further virulence testing of mutant strains using *G. mellonella* hemocytes or the *Drosophila melanogaster* macrophage-like cell line called S2 may provide a more sensitive method to determine subtle phenotypes. For example, S2 cell line was shown to be an effective model system for characterizing the role of Ysa T3SS in *Y. enterocolitica*, which shares the same gene organization as T3SSYE2 in MH96 (Walker et al., 2013). Further experiments using insect cell lines would be particularly ideal for specifically Δ*vipB* and Δ*fim1*, which are deficient for the T6SS and usher-chaperone fimbriae, both putative VFs that are likely to play a role in close-range interactions with the host-cell during intrahemocoelic infection.

5.4.7 Significant reduction in virulence detected for Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* strains in the intrahemocoelic infection of *G. mellonella* at 37 °C

Bioassay results reported in this thesis found that wild-type MH96 is equally as virulent against *G. mellonella* by injection at both 25 and 37 °C, which is consistent with previously reported findings of Hurst et al., (2015). Despite this, MH96 is still thought to have reduced pathogenic capability within *G. mellonella* at 37 compared to 25 °C because previous *in vivo* RNA-seq data revealed significant repression of co-secreted VFs, including Yen-TC/TC-associated factors at 37 °C, but only during intrahemocoelic infection of *G. mellonella* (Chapter 4, section 4.3.8, Table 4.13). Subsequent bioassays conducted at 37 °C for two mutant strains (Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6*) found an estimate three-fold reduction in virulence by intrahemocoelic injection of *G. mellonella* compared to MH96 wild-type. As a result of the reduced virulence of Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* during intrahemocoelic infection of *G. mellonella* at 37

°C, *in vivo* transcriptome of both strains under these conditions were investigated and reported in Chapter 6 of this thesis, and from there experimental work supported the molecular characterization of Yen6 and HCUI$_{YE96}$, which can be found in Chapters 6 and 7 of this thesis, respectively.

5.4.8 Focused investigations of Yen6 and Yen7 as putative temperature-dependent regulators of Yen-TC

Targeted investigations of the PAI$_{YE96}$-encoded putative regulators *yen6* and *yen7*, including predicted secondary structure comparison, amino acid sequence alignment and phylogenetic analysis provided further support to the hypothesis that *yen6* and *yen7* likely function as DNA-binding transcriptional regulators in MH96. Also, some bioinformatic data also suggest the presence of additional regulatory processes involved in the temperature-dependent regulation of *yen6*, *yen7* and downstream toxin genes, including widespread detection of both ncRNA and asRNAs being transcribed in and around the *yen6, yen7* and toxin coding genes (previously identified from *in vivo* RNA-seq), specifically 3'UTR-encoded terminator and potential *cis*-acting asRNA associated with *yen6*. The discovery of multiple perfectly repetitive binding motifs within the promoter and coding regions of *yen7* also suggests there are multiple interacting DNA-binding proteins involved in the regulation of *yen7*.

5.4.8.1 *In silico* characterization of Yen6 – putative novel LytTR-containing regulator

Based on the current NCBI assigned-annotation, *yen6*, produces a small hypothetical protein; however, research presented in this thesis found that *yen6* functions as a virulence regulator in MH96. Due to the location of *yen6* on PAI$_{YE96}$, responsiveness of *yen6* expression to 37 compared to 25 °C (especially under *in vivo* conditions) and contribution of *yen6* to virulence of MH96 in the *G. mellonella* intrahemocoelic infection model at 37 °C, characterization of the functional role of Yen6 as thermoregulator of virulence in MH96 was the major focus of this thesis (especially Chapter 6). As such, additional targeted bioinformatics were undertaken to support further characterization of this novel regulator, revealing *yen6* to encoding small LytTR-containing binding protein sharing significant structural similarity to the DNA-binding domain of AgrA,

an important virulence regulator involved in density-dependent toxin secretion by *St. aureus* (Bronner et al., 2004; Dunman et al., 2001).

Proteins containing LytTR DNA-binding domains are most commonly found to be associated with bacterial two-component regulatory systems (TCRSs), in which the LytTR domain is found within the effector module of the response regulator (RR) component (Behr et al., 2016; Carter et al., 2007; Morici et al., 2007; Nicod et al., 2014; Sidote et al., 2008). Bacterial RRs are modular-type proteins consisting of two domains: a receiver and an effector, which have been mixed-and-matched over evolutionary timescales, giving rise to an incredible diversity of TCRSs used by bacteria to sense and adapt to changing environments (Capra and Laub, 2012; Galperin, 2006; Galperin et al., 2005). Structure classification of receiver and effector domains of a variety RRs encoded in prokaryotic genomes found the great majority encode DNA-binding HTH domains, including either NarL-type (33 % of total) or OmpR/PhoP-type wHTH (19 % of total) (Galperin, 2006). In this same analysis, only 3 % of all RR examined were found to contain a LytTR-containing effector domain, making these types comparatively rare.

Interestingly, Yen6 appears to be a small regulatory protein, that primarily contains the DNA-binding LytTR-domain and does not appear to be part of a TCRS. This contrasts with AgrA of *St. aureus,* which is part of a TCRS that includes cognate membrane-bound sensor, AgrC (Novick and Geisinger, 2008; Novick et al., 1993). While the majority of LytTR-domain containing proteins are associated with TCRS, some have been linked to other N-terminal domains, including CheY, LMO ('elmo', 40 bp with coiled-coil region), ATP-binding cassette of ABC-type transporters and PAS-domains (Galperin, 2008; Nikolskaya and Galperin, 2002); however, like Yen6, smaller standalone LytTR-containing proteins have also been identified as well, for example BlpS from *Streptococcus pneumoniae* (AAK74683) and the orf50 from the *Pseudomonas* phage D3 (AAF80809) both represent small LytTR-domain containing proteins (Nikolskaya and Galperin, 2002).

Most LytTR domains are ~105 residues and characterized by having a unique structure consisting of numerous stranded β-folds (Sidote et al., 2008) and compared to the secondary structure of AgrA which was found to contain 10 β-folds (Sidote et al., 2008), the predicted secondary structure of Yen6 is expected to contain only 8 β-folds, as well

as a novel loop structure between the fifth and sixth β-strands that may participate in DNA-binding interactions as it is located within the same plain as the other loops known to interact with DNA in the AgrA model. The most likely DNA-interacting residue located on this novel loop is an asparagine (N81), especially since asparagine (N201) was already demonstrated to be a DNA-interacting residue within a loop of AgrA, where side-chain water-mediated contact is made with A and T nucleic acid within the minor groove (Sidote et al., 2008).

It should be noted that the LytTR-containing domains of protein sequence are notoriously difficult to assess by sequence conservation, especially since the DNA-interacting residues (located within the loops) are highly variable and the repetitive structural elements (the many beta-sheets) are conserved. Considering this, Yen6 and other putative regulators sharing significant amino acid sequence similarity identified from pathogenic Enterobacteriales and Pasteurellales may comprise a novel family of standalone LytTR-containing regulators.

5.4.8.2 *In silico* characterization of Yen7 – putative wHTH-containing regulator

Like *yen6, yen7* is also predicted to code for a small hypothetical protein according to current NCBI annotation; however, the secondary structure prediction suggests that Yen7 contains a wHTH domain, sharing some similarities in structure to PhoP from *M. tuberculosis* (Wang et al., 2007). Proteins containing wHTH domains are often associated with the DNA-binding domains of RRs of TCRSs (Galperin, 2006) including virulence regulators GacS/GacA and PhoP/PhoQ, which were both implicated as virulence regulators in *Ps. entomophila* (Dieppois et al., 2015; Liehl et al., 2006; Opota et al., 2011; Vallet-Gely et al., 2010) and *P. luminescens* (Derzelle et al., 2004), respectively. In previously reported findings, TCRSs were found to be one of the most prevalent types of regulators identified as highly expressed under *in vivo* growth conditions at 25 °C that also shared significant sequence homology to previously characterized VFs from other pathogenic bacteria (Chapter 4, section 4.3.7.2, Figure 4.17); however, (like *yen6*) *yen7* also appears to encode a small stand-alone DNA-binding regulator and does not seem to be affiliated with any sort of phosphate receiver domain, which would be typical for a TCRS-associated RR component.

In addition to the predicted structure of Yen7 sharing a high degree of similarity to the wHTH binding-domain of PhoP from *M. tuberculosis*, the proximity of *yen7* directly 5′ of the genes for the Yen-TC components and the higher expression of *yen7* under *in vivo* growth conditions all point to *yen7* being a *trans*-acting transcriptional activator of downstream Yen-TC genes. Indeed, the α-3 recognition helix of the predicted protein structure of Yen7 was highly conserved among other orthologous proteins identified among the genomes of diverse pathogenic Enterobacteriaceae and phylogenetic reconstruction identified a possible gene duplication among *Y. entomophaga*/*Y. nurmii* strains. The putative *yen7* paralog, another small wHTH containing PhoB-like regulator (PL78_17385), encoded on the HCUI$_{YE96}$ unstable island (Chapter 7, section 7.3.1, Figure 7.1). The expression of *yen7* was found to be linked to the co-expression of Yen-TC and TC-associated factors, including the paralogous PhoB-like regulator and the holin-endolysin system (Chapter 4, section 4.3.8, Table 4.13). Cumulatively, this evidence supports a hypothesis that *yen7* and the PhoB-like regulator are functionally redundant transcriptional activators of Yen-TC production and secretion in MH96; however, research into the function of PhoB-like regulator was the focus of a separate PhD research project occurring concurrently with this one.

Three different potential binding motifs (ranging from 9 – 10 bp perfect repeat pairs) were identified in an overlapping pattern within the promoter and coding regions of the *yen7* loci and represent potential binding locations for regulators controlling the expression of *yen7*. One repeat motif (rep1) overlapped perfectly with the 35-box and the start codon of *yen7* and may be a particularly important regulatory motif, for which some, yet unknown, regulatory factor could bind. Since the LytTR-containing virulence regulator of *St. aureus* was found to bind to direct 9 bp repeats (often imperfect), and the predicted protein structure of Yen6 is shares similarities to the LytTR-binding domain of AgrA, there is potential that Yen6 may bind rep1, which would put Yen6 in a good position to act as transcriptional repressor of *yen7*. Additional work was carried out to characterize the genetic relationship between *yen6* and *yen7* including translational *lacZ* reporter fusions reported in this Chapter, as well as Δ*yen6* mutant RNA-seq (Chapter 6, section 6.3.1) and electrophoretic mobility shift assays (EMSA) testing recombinant Yen6 specific binding activity against the promoter region (including rep1) of *yen7* (Chapter 6, section 6.3.2).

5.4.8.3 Yen6 and Yen7 overproduction was found to affect growth, plasmid stability, Yen-TC production and secretion in MH96.

One of the first steps to characterizing the regulatory effects of *yen6* and *yen7* on toxin production and secretion by MH96 involved comparing cell growth, plasmid stability and protein production following over-expression of two *yen6* variants – a shorter *yen6* (contains 32 bp 3'UTR) and longer *yen67as* (contains 478 bp 3'UTR, and *yen7* (containing 31 bp 3'UTR) compared to empty vector. Under 0.2 % arabinose, strains containing pBAD-*yen6* and pBAD-*yen67as* plasmids consistently reached higher final cell densities and maintained significantly higher rates of plasmid stabilities compared to strains carrying an empty vector, indicating beneficial effects and strong selection favoring the over-expression of both *yen6* and *yen67as* under inducing conditions. On the contrary, strains containing pBAD-*yen7* under 0.2 % arabinose had greatly reduced growth rate and rates of plasmid stability compared to strains with empty vector, indicating high expression of *yen7* has deleterious effect on the cells.

Another interesting finding was drastic effect temperature had on plasmid stabilities; while pBAD-*yen6* and pBAD-*yen67as* had higher plasmid stability at 25 °C under arabinose 0.2 %, the stability completely dropped at 37 °C under the same conditions. Likewise, the plasmid stability for pBAD-*yen7* was found to be very low at 25 °C but the plasmid was perfectly maintained at 37 °C under inducing conditions. These findings were very consistent, and suggest temperature is an important factor in the activities of both putative regulators, *yen6* and *yen7* in MH96.

While over-expression of *yen6* was found to have little impact on toxin production/secretion, over-expression of *yen67as* under 0.2 % arabinose was (on occasion) found to result in increased toxin production and secretion visualized by SDS-PAGE compared to empty vector and pBAD-*yen6* based on SDS-PAGE. One significant limitation of SDS-PAGE is that protein levels can only be considered semi-quantifiable; however, the highly variable effect of over-expression of *yen67as* under arabinose 0.2% resulting in increased toxin translation was also identified using the P_{chi1}::lacZ translational reporter strain, but this may have been a coincidence and further validation is required.

In contrast, over-expression of *yen7* was shown to consistently result in partial restoration of secretion at 37 °C and tentatively showed greater production of some Yen-TC components as well, despite deleterious effects on growth and plasmid stability, even at much lower arabinose concentrations (0.002 %). It was, however, challenging to quantify differences in the larger Yen-TC components (YenB, YenA1 and YenA2) when *yen7* was overexpressed due to the semi-quantitative nature of SDS-PAGE, so whether *yen7* can also activate yenB, *yenA1* and *yenA2* remains to be definitively shown. Ultimately, these findings support *yen7* as a putative transcriptional activator of Yen-TC production and secretion in MH96, but the lack of phenotype by SDS-PAGE of the Δ*yen7* mutant suggests *yen7* may not be the only transcriptional regulator involved in Yen-TC production and secretion. These findings may also suggest higher order DNA structural complexities associated with this region of DNA and or potential over-regulation by the PhoB-like regulator.

5.4.8.4 Differences between transcription and translation of *yen6* and *chi1* provide evidence for *in vitro* temperature-dependent post-transcriptional mechanisms regulatory in MH96 virulence.

One important goal for experiments using the P_{yen6}, P_{yen7} and P_{chi1}::*lacZ* reporter strains was to validate temperature-dependent translational activity associated with the promoter region of each of these genes. Previous RNA-seq results determined that *yen7* and *chi1* had significantly higher expression at 25 compared to 37 °C, while *yen6* had significantly higher expression at 37 compared to 25 °C *in vitro* (Chapter 4, section 4.3.8, Table 4.13). Evidence for post-transcriptional thermoregulatory mechanisms was gained by experimental work with *lacZ* reporter strains measuring temperature-dependent effects on β-gal production in MH96 under *in vitro* (and *in vivo* for P_{yen6}::*lacZ* strain) conditions.

Preliminary assessment of the *lacZ* reporter strains determined that the P_{yen7}::lacZ *cis*-merodiploid strain did not produce any β-gal when grown at 25 °C *in vitro*. The 5′ *yen7* protomer region of 837 bp was fused in-frame to the promoterless *lacZ* reporter gene (Kalogeraki and Winans, 1997); however, the amount of β-gal produced was essentially equal to the MH96 wild-type control. These finding imply that additional promoter sequences must be present beyond the *yen7* start codon, located within the protein

coding region of the gene that represent putative binding motifs. These findings are consistent with the prediction of three different overlapping perfect repeat pairs are located with the protein coding region of *yen7*. Conjugative *E. coli* strains containing suicide plasmids for generating additional *yen7::lacZ* fusion strains containing the promoter and increasing portions of the coding region were constructed, but time restraints did not allow for recombination and follow-up assessment for β-gal production in these strains and this remains an i area of future work.

Unlike P_{yen7}::lacZ, the P_{yen6}::lacZ and P_{chi1}::lacZ reporter strains did produce β-gal and were used in subsequent β-gal assays. P_{chi1}::lacZ was found to produce significantly greater amounts of β-gal when cultured at 25 compared to 37 °C, which was consistent with previous RNA-seq findings (Chapter 4, section 4.3.8, Table 4.13). On the contrary, P_{yen6}::lacZ was found to also produce greater amounts of β-gal at 25 compared to 37 °C, which was the opposite trend identified in the RNA-seq data; based on the transcriptome, *yen6* was actually identified as the most significantly different expressed (DE) gene compared to 25 and 37 °C of any gene in the genome, with greater than 8 and 4 $\log_2$CPM fold-change under *in vivo* and *in vitro* conditions, respectively. These conflicting results indicate multiple regulatory interactions are controlling the temperature-dependent expression and translation of *yen6* in MH96.

In order to characterize thermoregulation of *yen6* under *in vivo* conditions, β-gal production by P_{yen6}::lacZ strain was measured after injection into the hemocoel of *G. mellonella* either incubated at 25 or 37 °C. While best efforts were taken to capture samples at consistent stages of the infection, a great variability in pathogen growth rates were encountered, especially when hosts were maintained for longer than adequate periods in the laboratory. Preliminary measurements of β-gal activities from hemolymph samples identified that the production of dark pigment melanin interfered with OD measurements at 600 nm, so a phenol oxidase inhibitor, phenylthiocarbamide/phenylthiourea, was added to subsequent samples. The production of melanin still remained an issue with the *in vivo* β-gal assay and it was noted that highly conserved measurements could only be obtained if adequately fresh hosts were used. Therefore, future applications of the *in vivo* β-gal assay must follow common recommendations for *G. mellonella* experiments, that include best practices to

improve host consistency, especially using hosts within one week of arrival, such as described in (Tsai et al., 2016).

Despite challenges associated with host-variability, using the *in vivo* β-gal method to investigate thermoregulation of the P_{yen6}::lacZ strain revealed a > 39x fold-change in enzymatic activity, with much higher β-gal production at 37 compared to 25 °C, after intrahemocoelic injection after 16-17 and 19-22 h, respectively. These results support a hypothesis that *yen6* is able to sense and respond to the host environment by some yet unknown post-transcriptional regulatory factors active within the *in vivo* environment at 37 °C. These unknown regulator factors must interact with the 5′ UTR region of the *yen6* messenger RNA to increase production of this regulator specifically under *in vivo* conditions. The *in vivo* β-gal assay represents a promising new tool for future work investigating the effects of *in vivo* cues on production of important putative VFs identified MH96 (with the caveat that hemolymph samples must only be collected from healthy and consistently-sized *G. mellonella* that have been maintained on diet for a week or less). Since *G. mellonella* is increasing as a wide-spread model host in the study of pathogenic microorganisms (Chapter 4, section 4.1.4), perhaps this approach for studying reporter-fusions within hemolymph samples can be more widely applied to other pathogens also.

5.4.8.5 Limitations of comparing β-gal activity of cells from different temperatures due to temperature-induced dimorphism in MH96

The calculation of MU, which is used for comparison of β-gal activity of MH96 *lacZ* reporter strains is normalized by OD, which is measured as absorbance at 600 nm (Schaefer et al., 2016a, 2016b); however, the calibration curve data demonstrated that the linear relationship between CFU and OD at 600 nm of MH96 was quite different at 25 compared to 37 °C, which may be related to differences in cell shape (MH96 at 37 °C has small round cells, but cells are more rod-shaped or variable at 25 °C). While, this natural temperature-dependent dimorphism of MH96 does not completely prevent comparison of β-gal activity using MUs, it is important to consider this confounding factor when extrapolating results from any β-gal assays using MH96 strains to comparing temperature effects on translational levels and is a recognized limitation of this method for quantifying translational levels in MH96. Further information,

including mitigation measures taken in order to improve temperature-wise comparison of MH96 using the β-gal assay are included in Chapter 8 (section 8.4).

5.4.8.6 Mutagenesis of *yen6* and *yen7* was shown to affect beta-gal production of P_{chi1}::lacZ strains, but this could not be complemented in *trans*.

One of the primary goals of this chapter was to conduct genetic experiments to test whether *yen6* is a *trans*-acting transcriptional repressor of *yen7* or the first gene in the Yen-TC operon, *chi1* and whether *yen7* acts *in trans* to activate transcription of *chi1*. Unfortunately, the P_{yen7}::*lacZ* reporter failed to produce any measurable β-gal which limited investigations somewhat; however, experimentally it could be shown that deletion of *yen6*, *yen67* and *yen7* significantly affected the amount of β-gal produced by the P_{chi1}::lacZ reporter strain under *in vitro* conditions at 25 °C. When *yen6* was deleted, the translational level of the *chi1* reporter strain increased > 2-fold compared to wild-type and these results suggest that Yen6 acts as a repressor of Yen-TC transcription. When either *yen67* or *yen7* were abolished, the translational level of the *chi1* reporter strain decreased > 3-fold and > 2-fold, respectively suggesting Yen7 is a transcriptional activator of *chi1*.

Initially these findings supported the original hypothesis: *yen6* acts as a temperature-dependent repressor of *yen7*, which is a transcriptional activator of *chi1*; however, attempts to *trans* complement Δ*yen6* and Δ*yen67* by introduction of a mid-copy cloning vector with *yen6* and *yen7* under their natural promoters or induction of *yen67as* under P_{BAD} where not successful. The Δ*yen6* strain carrying the mid-copy vector with the entire P_{yen6} – P_{yen7} region was found to be only partially complemented compared to the Δ*yen6* strain carrying empty vector. Over-expression of *yen67as* from pBAD under 0.02 % arabinose was also not a successful method to complement the Δ*yen6* strain, which actually resulted in increased β-gal production by the *chi1* rather than reducing the translational level as expected; however these results were in-line with previous findings that over-expression of *yen6* and *yen67as* from the arabinose vector has a positive effect on cell growth and plasmid stability, and in the case of *yen67as*, can (albeit variably) result in increased levels of Yen-TC components and co-secreted TC-associated factors as visualized by SDS-PAGE and quantified by β-gal assay of the *chi1* reporter strain.

Trans complementation of Δ*yen67* by mid-copy vector containing P_{yen6} – P_{yen7} was found to further reduce β-gal production by the *chi1* reporter strain compared to both MH96 and Δ*yen67* carrying empty vector; however, complementation would expect to increase β-gal production up to wild-type levels. Perhaps, confounding effects resulting from excessive binding of the mid-copy vector by DNA/RNA-binding protein with specificity to the P_{yen6} – P_{yen7} region may also contribute to the unexpected results, suggesting *trans* complementation using a mid-copy variable may not be suitable for this system and addition efforts to complement *in cis* may be required. Overall, these results suggest that there may more complex regulatory process occurring in MH96, which is consistent with results of transcriptome (Chapter 4, section 4.3.8, Table 4.13) and protein visualization reported in this chapter, which were found to be disparate with respect to Yen-TC/Co-secreted TC-associated factors expression and levels, at 37 °C under *in vitro* conditions, respectively.

5.4.8.7 Results summary table and consider if evidence supports the original model and propose revised model based on findings.

While initial targeted work focused on identifying phenotypes from several targeted mutant strain using protein visualization by SDS-PAGE and the bioassay using *G. mellonella* intrahemocoelic infection model, the bulk of experimental work reported in this chapter focused on investigating whether putative regulators *yen6* and *yen7* were involved in temperature- and host-dependent regulation of *chi1*, and essential component of Yen-TC and the first gene in the operon, and these results are summarized in Table 5.12, below. Over-expression experiments did not find *yen6* or *yen67as* exerted any repression on Yen-TC secretion or production at 25 °C, but possibly contributed to repression of Yen-TC production at 37 °C; however, the presence of Yen-TC and co-secreted TC-associated factors in CP is considered variable in MH96. Remarkably, over-expression of *yen6* and *yen67as* was found to have beneficial effects on cell growth, plasmid stability, and *yen67as* was shown to occasionally result in increased Yen-TC production/secretion. While β-gal assays did identify subtle, yet statistically significant differences in enzymatic activity of *chi1* reporter strain in Δ*yen6*, Δ*yen67* and Δ*yen7* strains these effects could not be *trans* complemented. While this work ultimately failed to generate definitive proof that *yen6*

acts as a transcriptional repressor of *yen7* or has a significant thermoregulatory effects on Yen-TC component gene transcription, further investigations reported in the next section of this thesis (i.e., Chapter 6) characterized a much different role for *yen6* in virulence regulation of MH96 at 37 °C *in vivo.*

Table 5.12: Summary of all molecular investigations into transcriptional activities of *yen6* and *yen7* on *chi1* . '+' = positive effect; '-' = negative effect.

Result	Gene of interest		
	yen6	*yen67as*	*yen7*
Over-expression experiments			
Cell growth effect	+	+	-
Plasmid stability effect at 25 °C	+	+	-
Plasmid stability effect at 37 °C	-	-	+
Secretion effects	None	Variable	Over-expression partially restored secretion at 37 °C and increased secretion at 25 °C
Cell pellet effects	None	Variable	None
	Region/gene of interest		
	yen6	*yen6yen7*	*yen7*
Genetic interactions using mutant/lacZ reporter stains			
Deletion mutant effect on *chi1* reporter beta gal activity	Increased > 2-fold	Decreased > 3-fold	Decreased > 2-fold
Ability to complement deletion *in trans* by *chi1* reporter bet gal assay	Partial with P_{yen6}-P_{yen7} in mid-copy vector, but over-expression of *yen67as* had opposite effect of further increasing *chi1* reporter activity	No, P_{yen6}-P_{yen7} in mid-copy vector resulted in opposite effect of further decreasing *chi1* reporter activity.	Untested
Temperature-dependent difference *in vitro*?	Significantly higher activity of P_{yen6}*::lacZ* reporter strain at 25 compared to 37 °C under these conditions.		Untested
Temperature-dependent difference *in vivo*?	Significantly higher activity of P_{yen6}*::lacZ* reporter strain at 37 compared to 25 °C under these conditions.		Untested
Contribution to virulence	3-fold reduction in virulence in the intrahemocoelic infection *G. mellonella* at 37 °C; Untested at 25 °C	Untested	Untested

Over-expression and *lacZ* assays were used to characterize the interaction of *yen7* on *chi1* and these experiments did provide evidence that *yen7* is likely an activator of *chi1* and secretion in MH96. One very compelling piece of evidence is the consistent partial restoration of secretion at 37 °C when *yen7* is over-expressed. The deleterious effects on cell growth during over-expression of *yen7* also support a model that *yen7* is an important regulator in MH96 and is likely under strict regulatory control in order to maintain proper growth. A recent preprint reports bimodal gene expression of the holin-endolysin secretion system in *S. marcescens* is orchestrated by *chiR*, such that only 1 % percent of cells are secreting the chitinolytic machinery via the holin-endolysin system within the population (de Assis Alcoforado Costa et al., 2019). Perhaps, a similar bimodal mechanism of *yen7* transcription is present in MH96 also, which allows cell population to reach stationary densities (~ 1 x 10^{10} CFU/ml), while only a portion of the population is responsible for the secretion large amounts of Yen-TC but further experimentation using *yen7* fluorescent reporters (similar to the methods used in the preprint) would be required to validate this theory.

Attempts to generate recombinant Yen7 protein for EMSA analysis against the promoter region of *chi1* were not successful. Briefly, using similar methods described in Chapter 6 (section 6.2.2.1) to express yen7$_{6x\text{-}His}$ in *E. coli* did not result in any visible recombinant protein production within induced whole cell lysate nor among the soluble lysate fraction (data not shown). Further characterization of the *yen7* binding interactions within the *chi1* promoter region and possible other promoter regions of other Yen-TC component genes would help to identify the putative binding motif of this novel toxin regulator. Other important next steps may also include comparing and contrasting of Yen7 and PhoB-like regulator binding dynamics and regulatory control over co-secretion of Yen-TC and TC-associated factors using complementary fluorescent reporter gene fusion strains.

Chapter 6. Characterization of the Yen6 regulon and molecular phenotyping of Δ*yen6* and Δ*cspA123*/HCUIYE96 by RNA-seq

6.1 Introduction

Further characterization of the deficient strains of *Yersinia entomophaga* (MH96), Δ*yen6* and Δ*cspA123*/ΔHCUI$_{YE96}$, was carried out because these deficient strains were found to have attenuated virulence in *G. mellonella* at 37 °C compared to wild-type (Chapter 5, section 5.3.1.2, Figure 5.8). The Δ*cspA123*/ΔHCUI$_{YE96}$ was also found to have a highly conspicuous non-secreting phenotype at 25 °C *in vitro,* however; *in vivo* RNA-seq data from Δ*cspA123*/ΔHCUI$_{YE96}$ identified a 17.5 kb excision of HCUI$_{YE96}$ and subsequent analysis determined the non-secreting phenotype could be attributed to the excision of the unstable region and not the targeted deletion of tandem cold-shock proteins, *cspA123*. This chapter reports the *in vivo* transcriptome of Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* strains at 37 °C in *G. mellonella*, from which further phenotypic assessment of Δ*yen6* and recombinant Yen6 DNA-binding dynamics were undertaken to show Yen6 is a DNA-binding regulator. Please note, Chapter 7 of this thesis also provides more details related to the Δ*cspA123*/ΔHCUI$_{YE96}$ strain, including phenotypic assessment of Δ*cspA123*, ΔHCUI$_{YE96}$ and Δ*cspA123*/ΔHCUI$_{YE96}$.

In addition to the significant attenuation in virulence observed in Δ*yen6* during intrahemocoelic infection of *G. mellonella* at 37 °C, prior evidence reported in previous chapters of this thesis also support Yen6 as small LytTR-containing transcriptional regulator that is responsive to environmental cues, such as *in vivo* growth conditions and temperature. Transcriptome data previously identified *yen6* as having the highest differential expression of any genes in the MH96 genome at 37 compared to 25 °C *in vivo* (Chapter 4, section 4.3.8, Figure 4.20, Table 4.13). Furthermore, the predicted protein structure of Yen6 was found to share significant similarities to the structure of the LytTR-domain of AgrA from *Staphylococcus* (*St.*) *aureus* (a well-known virulence response regulator) (Sidote et al., 2008) (Chapter 5, section 5.3.2.2, Figure 5.9); however unlike AgrA, Yen6 appears to be a standalone DNA-binding protein not affiliated with any two-component regulatory system (TCRS). The over-expression of *yen6* in MH96 with 0.2 % arabinose at 25 °C *in vitro* resulted in significantly higher growth rates as

well as positive selection on the pBAD-*yen6* induction vector compared to empty vector (Chapter 5, section 5.3.3.1, Figure 5.4). Additionally, targeted deletion of *yen6* was found to result in significantly greater β-galactosidase (β-gal) production by P_{chi1}::*lacZ* at 25 °C *in vitro* compared to empty pBAD (Chapter 5, section 5.3.3.6.3, Figure 5.27), but these results could not be easily complemented *in trans* (Chapter 5, section 5.3.3.6.3, Table 5.8 and Table 5.10). In a different reporter strain, β-gal production by P_{yen6}::*lacZ* strain was found to be significantly affected by temperature, and that the effect differed between to *in vitro* or *in vivo* conditions. The P_{yen6}::*lacZ* strain was found to have significantly higher β-gal activities at 25 compared to 37 °C under *in vitro* conditions (section 5.3.3.6.2, Figure 5.25) but enzymatic activity was significantly higher at 37 compared to 25 °C under *in vivo* conditions (section 5.3.3.6.7, Figure 5.31).

Taken together, these results demonstrate that regulation of *yen6* is complex, involving both temperature- and host-dependent mechanisms at transcriptional and post-transcriptional levels. While previous work did not determine definitive proof that Yen6 is a transcriptional repressor of *yen7* or *chi1*, here a deeper investigation of the transcriptional influence of *yen6* on both the transcriptome and metabolic phenotype of MH96 are undertaken. Furthermore, additional work to characterize the specific-binding affinities of recombinant $Yen6_{6x\text{-}His}$ against the promoter region of *yen7* and the promoter region of three other genes shown to be significantly affected by deletion of *yen6* based on *in vivo* RNA-seq. These findings, combined with additional evidence provided in previous chapters of this thesis, serve as definitive proof that Yen6 is a novel LytTR-containing DNA-binding protein that contributes to virulence during infection of *G. mellonella* at 37 °C.

6.2 Summary Methods

6.2.1 Molecular phenotyping of Δ*cspA123*/$\Delta HCUI_{YE96}$ and Δ*yen6* strains during infection of *G. mellonella* at 37 °C.

Unless otherwise stated *in vivo* RNA collection of MH96, Δ*cspA123*/$\Delta HCUI_{YE96}$ and Δ*yen6* from *G. mellonella* at 37 °C was carried out using methods previously described in Chapter 4, section 4.2.1.2. Total RNA samples (50 μl) were stabilized and shipped using RNA-stable tubes (Biomātrica) and then processed by Novogene in Beijing,

China. Total RNA quality and quantity were determined using 2100 Bioanalyzer Instrument (Agilent) and sample quality were also assessed using agarose gel electrophoresis. Samples were depleted using the Ribo-Zero Gold ribosomal RNA (rRNA) Removal Kit (Epidemiology)|Magnetic kit (Illumina, Singapore) according to the manufacturer's guidelines then strand-specific cDNA libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina. Libraries were selected for size 250 – 300 bp fragments and sequenced on the Novaseq 6000 PE150 (Illumina), which produces 150 bp paired-end sequence data.

The raw sequence data were trimmed as described in Chapter 4 (section 4.2.4), using adapter sequences listed in supplementary Table S17. The trimmed data were screened against the *G. mellonella* shot-gun genome assembly (GCA_003640425.1) and rRNA sequence (U65198.1, U65138.1, AF286298.1, AF423921.1 and X89491.1) as well as MH96 rRNA sequence (extracted from GCA_001656035.1) using the bbsplit function of BBMap (Bushnell, 2015). Rockhopper (v2.03) was used to align the screened paired-end reads to the MH96 reference genome (GCA_001656035.1/ASM165603v1) as well as predict 3' and 5' untranslated regions (UTRs), non-coding RNAs (ncRNAs), anti-sense RNAs (asRNAs) and operons, as described in Chapter 4, section 4.2.4.

Preliminary data exploration and differential expression (DE) analysis were carried out as previously described in Chapter 4, section 4.2.5, using the R packages: iNEXT (rarefaction analysis), RUVseq (PCA and RLE plots) and Limma (voom transformation and DE analysis). Due to limited biological replication (n = 2), the least conservative multiple-test correction method, "Benjamini & Hochberg" (Benjamini and Hochberg, 1995) was applied with an adjusted P-value cut-off of 0.1 (increased from standard cut-off of 0.05).

6.2.2 Electrophoresis mobility shift assay (EMSA)

6.2.2.1 Production and purification of recombinant Yen6

The coding region of *yen6* was amplified by polymerase chain reaction (PCR) from MH96 genomic DNA using high-fidelity Phusion DNA Polymerase (NEB) as previously described in Chapter 3 (section 3.3). Next, cloning, recombinant expression

and purification of Yen6$_{6x\text{-}His}$ protein was undertaken by Marion Schoof at Callaghan Innovation. The PCR produce was purified and cloned into plasmid pOPINF, which was previously linearized by digestion with KpnI and HindIII, using In-Fusion Cloning kit (Clontech) following manufacturer guidelines (supplemental Table S4). The vector pOPINF-*yen6* was transformed into *Escherichia coli* Turner cells (Novagen) and selected on ampicillin 100 μg/ml and sequence validated.

For induction of Yen6$_{6x\text{-}His}$, over-night cultures of *E. coli* containing pOPINF-*yen6* were diluted 1:250 of Terrific Broth (TB) and incubated at 25 °C until the optical density at 600 nm (OD_{600}) reached 0.5 – 1. Recombinant protein production was induced with 0.1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and cultures were incubated overnight at 16 °C. The following day cells were centrifuged at 5,000 x g for 20 min and the supernatant discarded. Cell pellets were resuspended in 25 ml phosphate buffer solution (PBS) and lysed using freeze thaw, lysozyme and sonication treatments and centrifuged at 25,000 x g for 20 min to separate soluble and insoluble fractions. Protein profiles from pre- and post-induced cellular samples and supernatant following cell lysis were assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to determine if Yen6$_{6x\text{-}His}$ was present as a soluble protein. To purify Yen6$_{6x\text{-}His}$, the soluble fractions were loaded onto a 5-ml Histrap FF nickel-affinity column (GE Healthcare Life Sciences) of the ÄKTA FPLC chromatography systems (GE Healthcare Life Sciences, Uppsala, Sweden) and equilibrated with 40mM imidazole (pH 7.4) and PBS solution. For protein dilution PBS with increasing imidazole concentration up to 500 mM (pH 7.4) was used.

6.2.2.2 Visualization of native Yen6 recombinant protein and buffer exchange

Once received from Callaghan Innovation, the purified Yen6$_{6x\text{-}His}$ was visualized under native conditions by using similar methods as SDS-PAGE described in Chapter 3 (section 3.5) with the following exceptions. Native PAGE was performed on ice using one-dimensional slab gels (1.0 mm thick) containing 12 % polyarcylamide without SDS using native running buffer (25 mM Tris/HCl, 200 mM glycine, pH 8.3). Undenatured samples (either 1 or 0.5 μl) were diluted into 20 μl 4x native loading buffer (125mM Tris/HCl (pH 6.8), 20 % glycerol (v/v), and 0.1 % (w/v) bromophenol blue) and 15 μl

was loaded into the wells following 120 V pre-run. Native Mark (Invitrogen) was used as a molecular marker and proteins were mobilized by charge with 120 V for 2 h and silver stained as previously outlined (Chapter 3, section 3.5.1).

For better long-term storage, purified Yen6$_{6x\text{-}His}$ was applied to VivaSpin 6, 10kDa MWCO (Satorius) and centrifuged at 9,500 x g for 15 minutes. The buffer was exchanged by washing the membrane with 1 ml EMSA buffer (20 mM NaH_2PO_4, 100 mM KCl, 5% glycerol, 2 mM 1,4-dithiothreitol, pH 8.0) to collect the protein. Yen6 in EMSA buffer was stored at -20 °C.

6.2.2.3 Design and preparation of biotin end-labelled DNA probes

Potential Yen6 DNA binding regions were selected primarily based on the RNA-seq data of the Δ*yen6* mutant, which found fructose and ribose gene clusters and a gene for putative RNA-binding protein, *yhbY* had DE compared to wild-type during *in vivo* conditions at 37 °C. These promoter regions were investigated for potential binding motifs using Geneious (v.R10) repeat finder and motif search (maximum 2 mismatches allowed). The potential for Yen6 to interact with a 9 bp perfect repeat putative binding motif called rep1, that overlaps the 35-box and the start codon of *yen7* was also investigated (see, Chapter 5, section 5.3.2.2, Figure 5.17).

The 5′ promoter regions of *yen7* (PL78_03735), *yhbY* (PL78_13050), D-ribose pyranase (PL78_12650) and PTS fructose transporter subunit IIBC (PL78_17165) were amplified by PCR using primers listed in Table 6.1, generating DNA probes ranging in size from 295 – 409 bp, referred to as P_{y7}, P_{yhbY}, P_{ribo} and P_{fruc}, respectively. Standard PCR primers and 5′-biotin labelled primers were sourced from Macrogen Inc.

Table 6.1: Primer sequences used to amplify probes for specific binding activity by Yen6 with EMSA

Primer	Sequence (5′ - 3′)	Amplicon size (bp)
EMSA_yen7F	CCTTCCTCCATTGGGCTACA	295
EMSA_yen7R	CCACACTACCTTTCCAGCTGT	
EMSA_riboseF	CGGAATCGGCAGCCCAG	409
EMSA_riboseR	GCGCCGGATCAGTTCGA	
EMSA_frucktoseF	ACGCCCACAACCACAGC	324
EMSA_fructoseR	CGCGTTGGCAGCCAAAT	
EMSA_yhbyF	TTAAACCAGGCGCGCGA	312
EMSA_yhbyR	CGGATGGGCCAGGCTTT	

Preparation of 5′-biotin-labelled DNA probes, preparation of EMSA reactions and electroblotting and detection were performed by Dr. Naren at School of Natural and Computational Sciences at Massey University, Albany Campus (New Zealand). Four DNA fragments containing putative Yen6 promoter regions were PCR amplified using 5′-biotinylated primers and purified by phenol/chloroform extraction and purified by cold ethanol precipitation and resuspended in distilled water.

6.2.2.4 Protein-DNA binding reactions, EMSA, electro blotting and detection

Specific binding activity of Yen6 against the probes was examined using EMSA by Dr. Naren at School of Natural and Computational Sciences at Massey University, Albany Campus (New Zealand). Binding reactions (20 µl total volume) including 20 nM biotin-labelled DNA and increasing concentrations of $Yen6_{6x\text{-}His}$ proteins in EMSA buffer and incubated for 30 min at 23 °C (room temperature). Excess unlabeled specific competitor DNA was added at a concentration of 200 nM to one reaction in order to test specific-binding activity. Excess non-specific competitor DNA (1 µg Salmon sperm DNA) and unrelated probe P_{crcZ} promoter DNA from *Ps. fluorescens* SBW25 as controls.

Samples were electrophoresed on 6 % native polyacrylamide gel in 0.5x TBE buffer at 120 V at 4 °C until the loading dye (60 % glycerol and 0.025 % bromophenol blue) reached the end of the plate.

DNA in native polyacrylamide gel was transferred to nylon membrane (Whatman Nytran SuPerCharge) using XCell SureLock mini-cell system (Invitrogen). Electroblotting was performed in 0.5x TBE buffer at 30 V for 1 hour. Next the DNA was

immobilized by incubating the membrane for 30 min at 80 °C. LightShift chemiluminescent EMSA kit (Thermo Fisher Scientific) was used to detect the biotin-labelled DNA probes following the manufacturer's instruction. The image was visualized by LAS-4000 Luminescent Imager equipped with ImageQuant LAS 4000 software (FujiFilm).

6.2.3 Phenotype microarray of MH96 and Δ*yen6* at 25 and 37 °C

Since *in vivo* transcriptome data from the Δ*yen6* revealed two gene clusters related to sugar uptake (in addition to a RNA-binding protein) were found to be DE compared to wild-type, so the ability of MH96 and Δ*yen6* to utilize a variety of sole-carbon sources under aerobic conditions was determined at 25 and 37 °C using the Omnilog phenotype microarray system (Biolog Inc.) as previously described (Chapter 4, section 4.3.9).

6.3 Results

Results presented here include *in vivo* transcriptome of MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* during intrahemocoelic infection of *G. mellonella* at 37 °C, DNA-binding specificity assessment of recombinant Yen6$_{6x\text{-}His}$ against DNA containing putative motifs within the promoter regions of *rbsD-xylG-rbsC-xylF-rbsK-ccpA*, *IIA/hpr-fruK-IIBC*, *yhbY* and *yen7* and phenotype microarray of sole-carbon source utilization by Δ*yen6* and MH96 at 25 and 37 °C.

6.3.1 Mutant transcriptome results

6.3.1.1 RNA collection

Inoculums and *in vivo* cell densities at time post-infection are provided below in Table 6.2 and Table 6.3, respectively

Table 6.2: *Yersinia entomophaga* MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* cell density pre-infection inoculum and *Galleria mellonella* incubation time . SD = standard deviation, CFU = colony forming units. Number of replicates = 3.

Strain	Mean inoculum cells (CFU ± SD)	Incubation time (h)
MH96	$1.27 \pm 0.38 \times 10^5$	9-10
Δ*cspA123*/ΔHCUI$_{YE96}$	$1.63 \pm 0.27 \times 10^5$	9-10
Δ*yen6*	$5.67 \pm 0.81 \times 10^4$	9-10

Table 6.3: *In vivo* cell densities at time of RNA collection for *Yersinia entomophaga* MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* at 37 °C in *G. mellonella* . SD = standard deviation, CFU = colony forming units. Number of replicates = 3.

Strain	Mean cell density per larvae (CFU / larvae ± SD)	Mean cell density per gram (CFU/g ± SD)
MH96	$5.73 \pm 3.31 \times 10^7$	$3.00 \pm 1.40 \times 10^8$
Δ*cspA123*/ΔHCUI$_{YE96}$	$1.75 \pm 0.70 \times 10^8$	$1.11 \pm 0.65 \times 10^9$
Δ*yen6*	$2.90 \pm 1.30 \times 10^7$	$1.29 \pm 0.48 \times 10^8$

6.3.1.2 RNA quantity/quality, library preparation and sequencing

Overall the samples were of high quality (mean RNA integrity number (RIN) values of ≥ 8.7 ± 0.7 SD) and reasonable quantities, except for one single low quantity sample (Table 6.4). Based on Agilent 2100 analysis, one Δ*yen6* RNA sample failed quality control due to low total amount of RNA (1.2 μg; minimum required is 3 μg); however, the sample were still sequenced with 70 – 80 % success rate as predicted by Novogene.

Table 6.4: Mean total RNA quantity and quality determined by Agilent 2100 analysis. SD = standard deviation, RIN = RNA integrity number. Number of replicates = 2.

Strain	Conc. ± SD (ng/μl)	RIN value ± SD	260/280 ± SD	260/230 ± SD
MH96	122 ± 8	9.0 ± 1.2	1.778 ± 0.314	0.688 ± 0.030
Δ*cspA123*/ΔHCUI$_{YE96}$	127 ± 16	9.8 ± 0.4	1.757 ± 0.061	1.007 ± 1.000
Δ*yen6*	82 ± 59	8.7 ± 0.7	1.907 ± 0.014	1.567 ± 0.054

6.3.1.3 Sequence data processing and quality control

Less than 0.05 % of sequences required removal during trimming. Host rRNA sequences were found to represent the largest proportion of sequence data contained

within the libraries (over 50 %) (Figure 6.1). Sizable fractions representing 15 – 25 % and 16 – 28 % of the sequence data aligned to MH96 and host genomes, respectively. Overall alignment rates for sequence data among trimmed libraries was 19 ± 8 % SD against the MH96 reference genome (data not shown). Due to the generally lower alignment rate of these data, rarefaction analysis was conducted to ensure sufficient capture of the dynamic range of expression (i.e., lowly expressed transcripts were captured) (Figure 6.2).

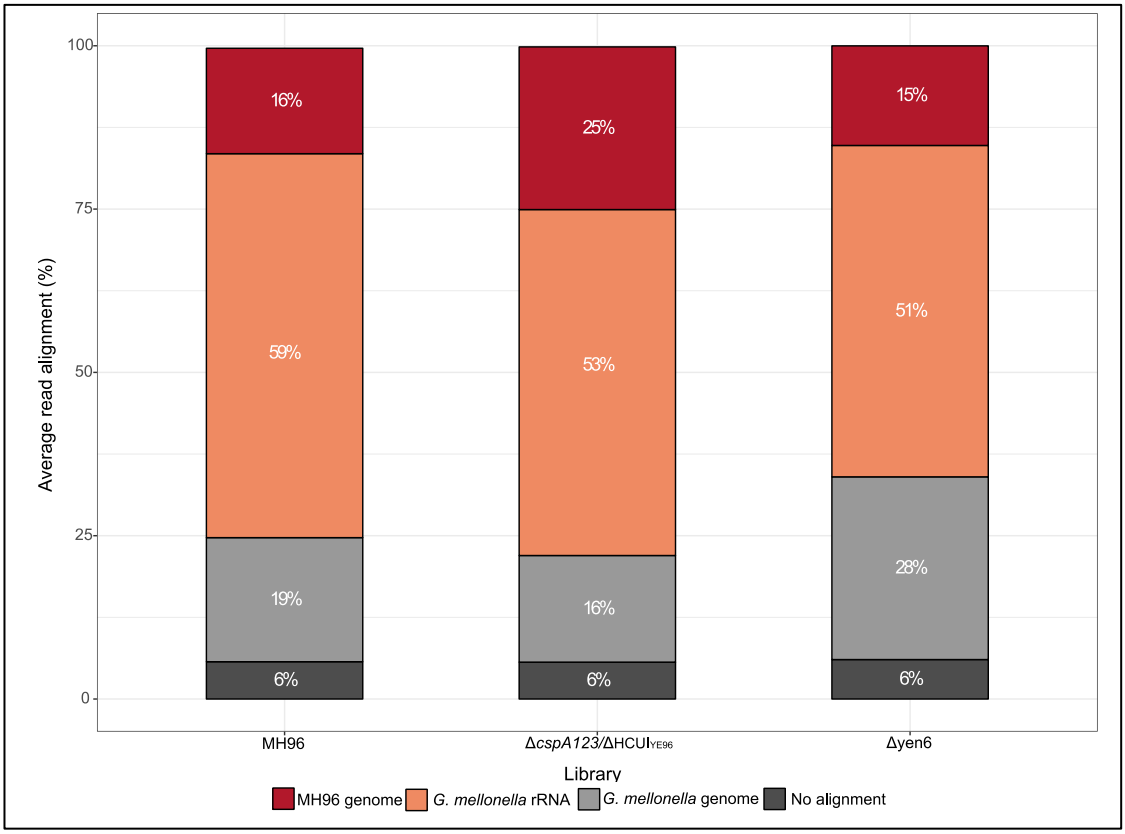


Figure 6.1: Proportion of trimmed paired-end reads aligning to *Yersinia entomophaga* MH96 genome or host, *Galleria mellonella* ribosomal RNA or genome sequence from *in vivo* RNA-seq libraries from MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* strains.

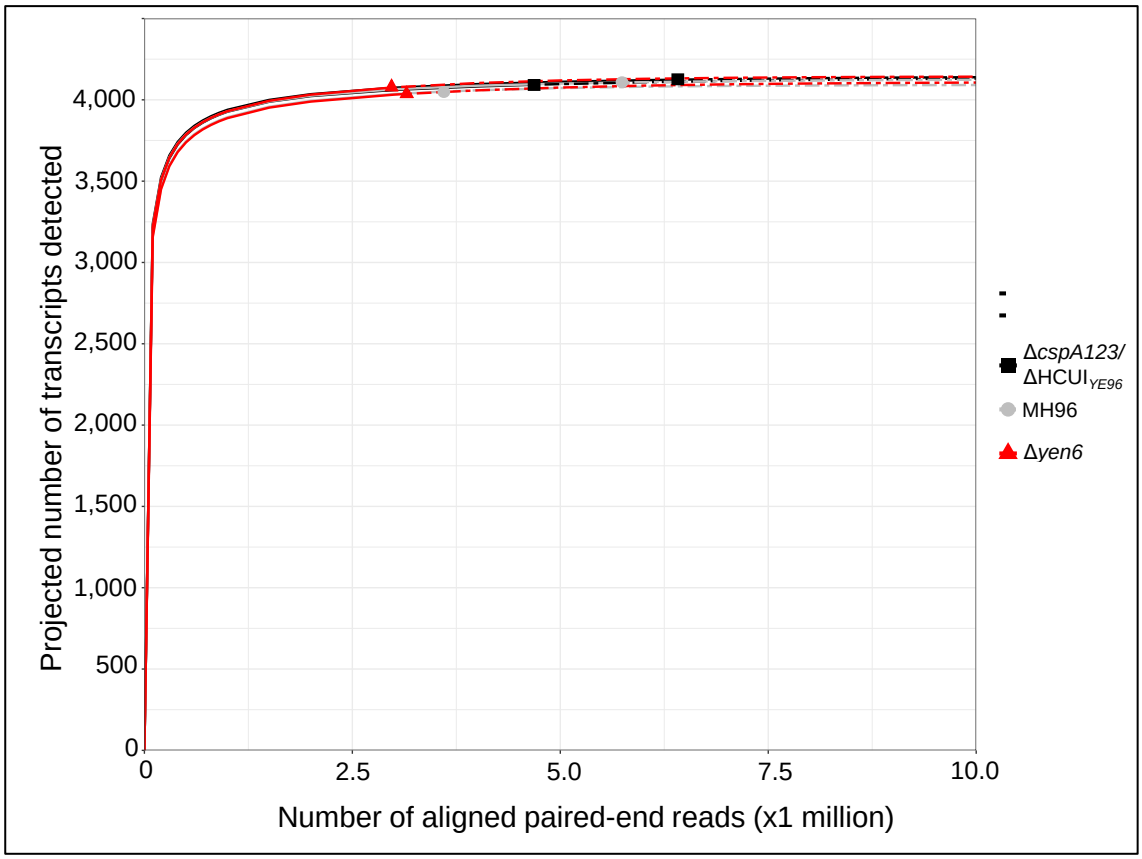


Figure 6.2: Predicted transcript detection based on the number of aligned paired-end reads for *Yersinia entomophaga* MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6 in vivo* libraries by rarefaction analysis. Solid and dashed lines are interpolated and extrapolated, respectively

Both unnormalized and upper quartile (UQ) normalized libraries clustered by strain based on principle coordinate analysis (PCA) (Figure 6.3). Based on inspection of relative log expression (RLE) plot, normalization by UQ method was found to be sufficient (i.e., the median RLE value for all libraries were normalized to approximately zero). Neither batch effects or excessive variation were identified among any of the normalized libraries. Regression analysis identified strong correlation between biological replicate samples with adjusted R_2-values ranging from 0.95 – 0.98; however, a large amount of variance was noted for asRNA among the two mutant libraries (Supplementary Figure S27).

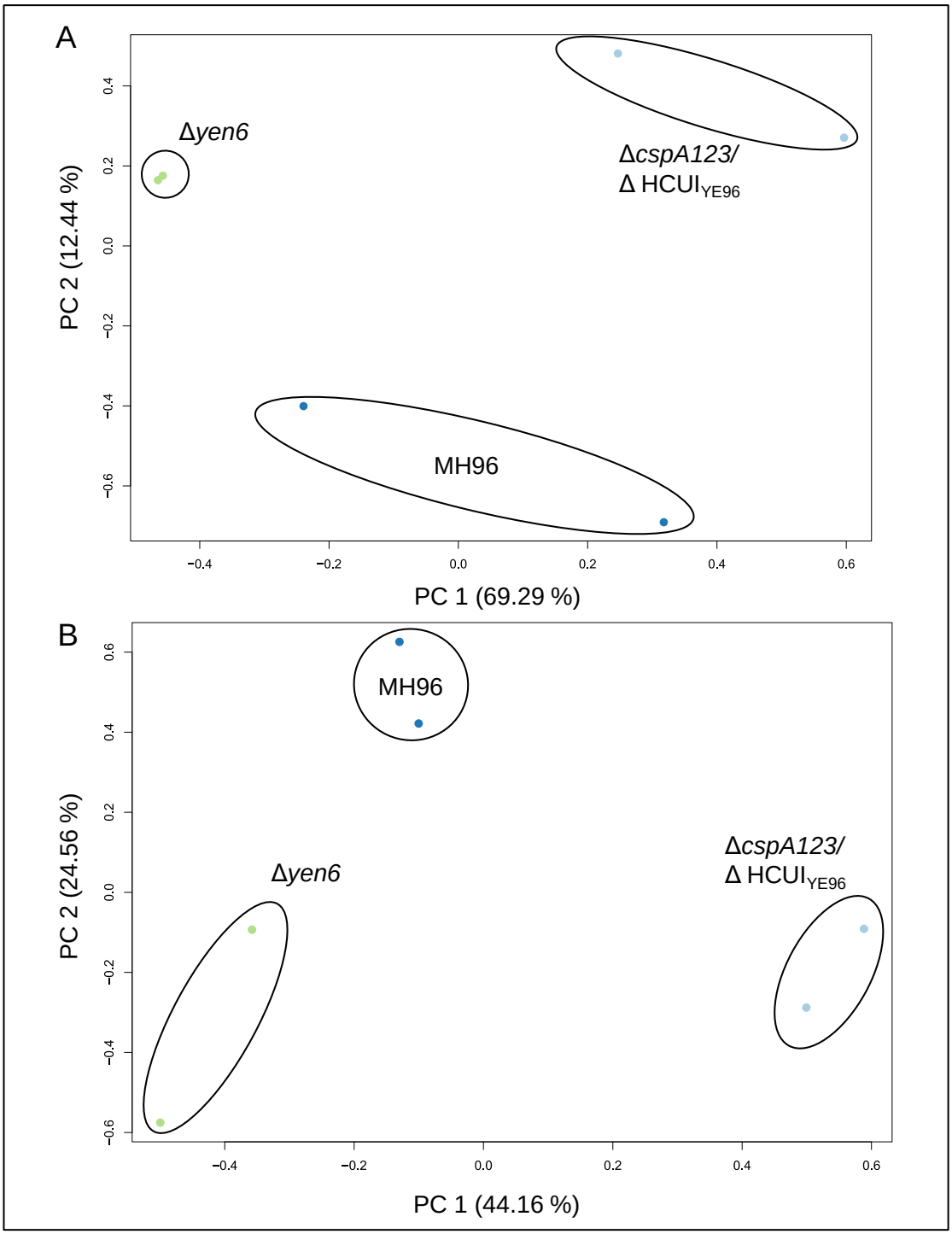


Figure 6.3: Principle component analysis of *Yersinia entomophaga* MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6 in vivo* libraries . A) unnormalized and B) upper quartile normalized count data.

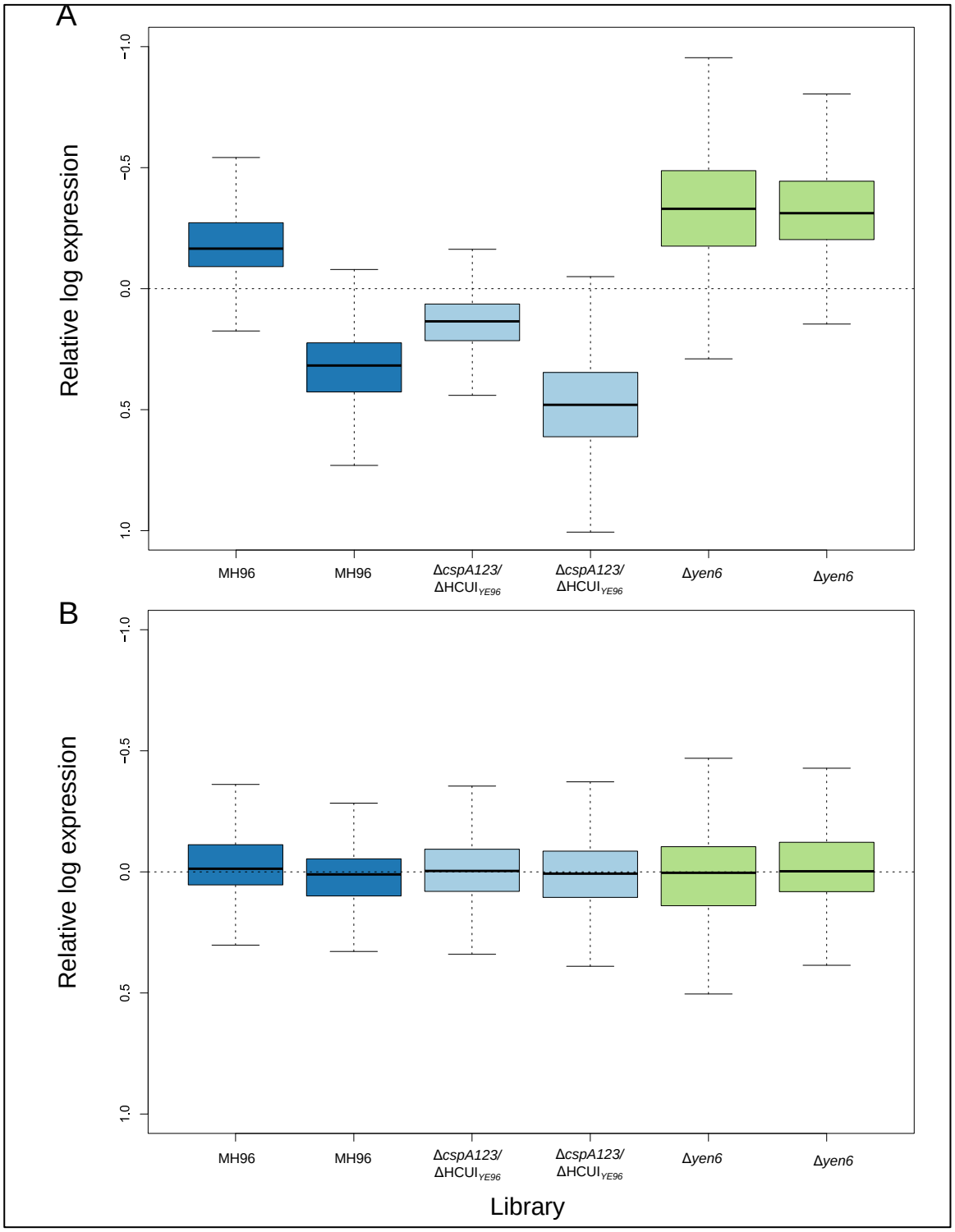


Figure 6.4: Relative log expression of *Yersinia entomophaga* MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6 in vivo* libraries represented as median boxplots . A) unnormalized and B) upper quartile normalized count data.

6.3.1.4 Differential expression analysis

Using the limma/voom transformation method (Supplementary Figure S28), very few DE genes identified between RNA collected *in vivo* at 37 °C from MH96 and Δ*yen6*, even using the least conservative method of multiple-test correction. The transcription of *yen6* and *yen7as* was significantly lower in Δ*yen6* (as expected because this gene was replaced with a kanamycin cassette), and not included in the analysis. Only six genes, were found to have significantly lower expression in Δ*yen6* compared to MH96 (Table 6.5). These genes are located in a single operon (*rbsD-xylG-rbsC-xylF-rbsK-ccpA*), and include *rbsD* (D-ribose pyranase, PL78_12650), *xylG* (D-xylose ABC transporter, PL78_12655), *rbsC* (ribose C transporter permease, PL78_12660), *xylF* (ABC-type sugar transport system, PL78_12665), *rbsK* (ribokinase, PL78_12670) and *ccpA* (catabolite control protein A, PL78_12675). All these genes were previously identified as more highly expressed *in vivo*, especially at 25 °C. Three genes located in a different putative operon related to fructose utilization, which includes PL78_17155 (bifunctional PTS fructose transporter subunit IIA/HPr protein), *fruK* (1-phosphofrucktokinase, PL78_17160) and PL78_17165 (PTS fructose transporter subunit IIBC) (*IIA-fruK-IIBC*), and as well as the gene encoding RNA-binding protein, *yhbY* (PL78_13050) were all identified to have significantly higher expression in Δ*yen6* compared to MH96 *in vivo* at 37 °C. The function of the *rbsD-xylG-rbsC-xylF-rbsK-ccpA* and *IIA-fruK-IIBC* were identified with respect to central carbon metabolism were identified (Figure 6.5) as well as other genes encoded by MH96 involved in glycolysis and the pentose phosphate pathway (Table 6.6).

Inspection of counts of raw alignments from the Δ*cspA123*/ΔHCUI$_{YE96}$ mutant libraries against the MH96 reference genome identified a spontaneously excised region of ~ 17.5 kb, which contained two integrase core domain proteins (PL78_17412 and PL7414) and several putative regulators, including AraC-family transcriptional regulator (PL78_17417), PhoB-like regulator (PL78_17385), *prsK* histidine kinase sensor (PL78_17340), response regulator of citrate/malate metabolism (PL78_17335) and NarL/FixJ family DNA-binding response regulator (PL78_17330) (Table 6.7). The excised region is flanked by two tRNA-Asn encoding genes (PL78_17415 and

PL78_1320) and contained the putative holin-endolysin system (PL78_17400-390) and a citrate lyase cluster (*citG*, *citX*, *citF*, *citE*, *citD* and *citC*; PL78_17370-45). Further validation of the excised region and characterization of the holin-endolysin secretion system/citrate fermentation unstable island, HCUI$_{YE96}$, is provide in Chapter 7 of this thesis.

Aside from the excised region, several genes were found to have significantly lower expression in Δ*cspA123*/ΔHCUI$_{YE96}$ compared to wild-type (Table 6.8). Notably, genes for Yen-TC components *chi1*, *yenA1*, *yenA2*and *yenB* had reduced expression in this mutant. The same xylose-transporter/ribose utilization cluster identified as DE in the Δ*yen6* was also found to have reduced expression in Δ*cspA123*/ΔHCUI$_{YE96}$ compared to wild-type. Other known host-specific putative virulence factors (VFs) with lower expression in Δ*cspA123*/ΔHCUI$_{YE96}$, included usher-chaperone fimbriae components (PL78_12480, PL78_12475 and PL78_12470), an RNA polymerase sigma factor (PL78_09636), downstream iron dicitrate transport regulator FecR (PL78_09630) and outer membrane cobalamin receptor protein (PL78_01710).

Several genes were also found to have higher expression in Δ*cspA123*/ΔHCUI$_{YE96}$ compared to MH96 *in vivo* at 37 °C (Table 6.9). Among these genes are clusters related to the acquisition of sulfur, including organic taurine (*tauA, tauB, tauD*; PL78_11445-11430) and inorganic sulfate and sulfite (*cysH, cysI, and cysJ*; PL78_14390-14380), which were previously found to be more highly expressed *in vivo* at 37 °C. Genes related to iron-sulphur cluster (Fe-S assembly, *fdx, hscA, hscB, iscA, iscU, iscS* and *iscR*; PL78_15480-15445) were also found to have higher expression in the cold-shock mutant as well as genes related to carbamoyl phosphate synthesis (*carB* and *carA*; PL78_13780-13775) two aspartate carbamoyltransferases subunits (*pyrI* and *pyrB*; PL78_10560-10555). Some genes found to have higher expression in Δ*cspA123*/ΔHCUI$_{YE96}$ compared to MH96 were previously identified as host specific putative VFs, including carbonic anhydrase (PL78_18360) and metallo-beta-lactamase fold hydrolase (PL78_08900).

Table 6.5. Transcripts identified as significantly differentially expressed *in vivo* at 37 °C in *Yersinia entomophaga* MH96 wild-type compared to Δ*yen6* . Log fold-change expression values and adjusted p-value determined from *in vivo* transcriptome using voom in limma R package.

Gene identifier	Transcript start/stop	Translation start/stop	Translation start/stop	Transcript start stop	Strand	Gene product	Log_2CPM fold-change	FC	Adj. P-value
PL78_17155	1,103,025	1,102,963	1,101,830	1,101,830	-	bifunctional PTS fructose transporter subunit IIA/HPr protein	-1.54	-	0.02
PL78_17160	1,101,833	1,101,833	1,100,895	1,100,881	-	FruK, hexose kinase,1-phosphofructokinase	-1.14	-	0.04
PL78_17165	1,100,880	1,100,880	1,099,192	1,099,182	-	PTS fructose transporter subunit IIBC	-1.21	-	0.08
PL78_13050	2,067,052	2,067,096	2,067,389	2,067,438	+	RNA-binding protein YhbY	-0.78	-	0.08
PL78_12650	2,523,603	2,523,586	2,523,167	2,523,167	-	RbsD, D-ribose pyranase	1.80	+	0.02
PL78_12655	2,523,166	2,523,159	2,521,627	2,521,627	-	XylG, D-xylose ABC transporter, ATP-binding protein	1.57	+	0.06
PL78_12660	2,521,630	2,521,630	2,520,665	2,520,665	-	RbsC, ribose ABC transporter permease	1.55	+	0.04
PL78_12665	2,520,664	2,520,601	2,519,711	2,519,711	-	XylF, ABC-type sugar transport system	1.59	+	0.02
PL78_12670	2,519,667	2,519,633	2,518,707	2,518,707	-	RbsK, ribokinase	1.43	+	0.01
PL78_12675	2,518,706	2,518,695	2,517,703	2,517,665	-	CcpA, DNA-binding transcriptional regulator, LacI/PurR family	1.29	+	0.02

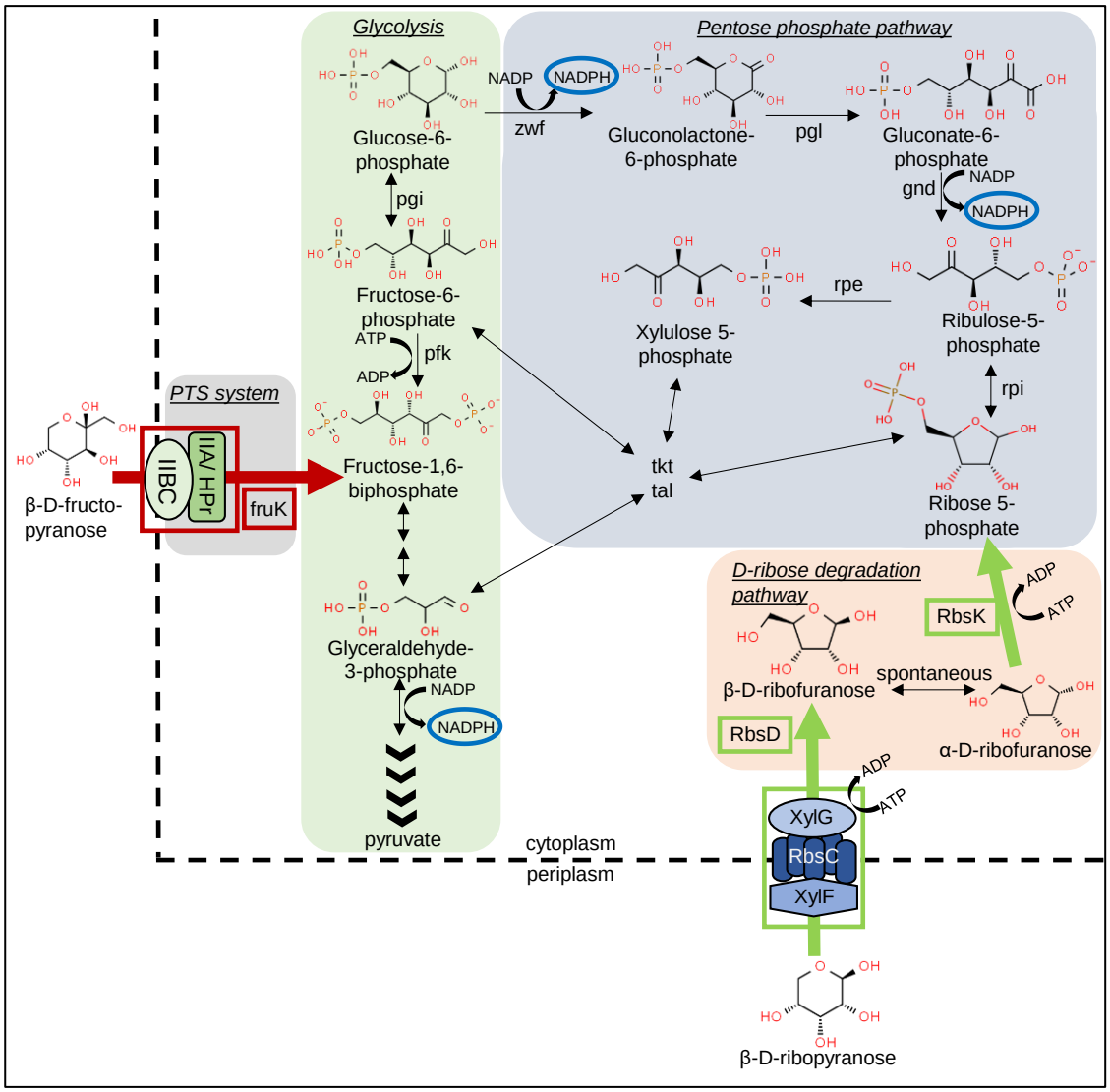


Figure 6.5: Metabolic pathways affected by *yen6* deletion in *Yersinia entomophaga* MH96 during infection of *G. mellonella* at 37 °C. Expression of *rbsD-xylG-rbsC-xylF-ccpA* were significantly decreased (indicated in green) and *IIA/hpr-fruK-IIBC* were significantly increased (indicated in red) in *yen6* compared to wild-type based on RNA-seq analysis. The locus tag and function of the metabolic enzymes included in this figure are provided in Table 6.6.

Table 6.6: Metabolic genes from *Yersinia entomophaga* MH96 that are depicted in Figure 6.5.

Gene	Function	MH96 locus tag
pgi	phosphoglucose isomerase	PL78_11035
pfk	6-phosphofructokinase	PL78_07890
zwf	glucose-6-phosphate dehydrogenase	PL78_04490
pgl	6-phosphogluconolactonase	PL78_01655
gnd	6-phosphogluconate dehydrogenase	PL78_02225
rpe	ribulose-5-phosphate epimerase	PL78_05440
tkt	transketolase	PL78_09365
tal	transealdolase	PL78_13695
rpi	ribose-5-phosphate isomerase	PL78_09120
rbsK	ribokinase	PL78_12670
rbsD	D-ribose pyranase	PL78_12650
xylG	ABC transporter, ATP-binding protein	PL78_12655
rbsC	ribose ABC transporter permease	PL78_12660
xylF	ABC transporter, periplasmic substrate binding protein	PL78_12665
fruK	1-phosophfructokinase	PL78_17160
IIBC	PTS, fructose transporter subunit IIBC	PL78_17165
IIA/hpr	PTS, fructose transporter subunit IIA/Hpr protein	PL78_17155

Table 6.7: Raw read alignments to a genomic region containing excision of putative holin-endolysin system and citrate fermentation cluster in Δ*cspA123*/ΔHCUI$_{YE96}$. Raw read aligned using Rockhopper package against reference *Yersinia entomophaga* MH96 genome.

Gene identifier	Gene Product	MH96	MH96	Δ*cspA123*/ ΔHCUI$_{YE96}$	Δ*cspA123*/ ΔHCUI$_{YE96}$	Δ*yen6*	Δ*yen6*
PL78_17410	tRNA-Asn	468	353	134	112	667	847
PL78_17412	integrase core domain protein	970	1,658	0	0	1,603	1,520
PL78_17414	integrase core domain protein	1,155	1,208	0	0	1,871	991
PL78_17417	AraC family transcriptional regulator	3,172	3,669	0	0	4,234	2,469
PL78_17400	hypothetical protein	214	525	0	0	161	0
PL78_17395	peptidase M15	342	194	0	0	99	109
PL78_17390	holin	730	307	0	0	353	320
PL78_17385	PhoB-like regulator	6,724	7,865	0	0	7,920	8,354
PL78_17386	hypothetical protein	472	403	0	0	738	1,574
PL78_17380	phosphopantetheinyl transferase	37,616	33,432	0	0	34,758	34,065
PL78_17375	anion permease	37,353	45,307	0	0	34,379	30,226
PL78_17370	CitG, Triphosphoribosyl-dephospho- CoA synthetase	27,890	31,128	0	0	26,424	24,674
PL78_17365	CitX, Apo-citrate lyase phosphoribosyl- dephospho-CoA transferase	18,092	16,079	0	0	11,272	12,387
PL78_17360	CitF, citrate lyase subunit alpha	43,030	50,281	0	0	32,525	42,535
PL78_17355	CitE, citrate lyase subunit beta	31,154	30,408	0	0	24,290	31,517
PL78_17350	CitD, citrate lyase subunit beta	12,894	12,674	0	0	10,429	11,848
PL78_17345	CitC, Citrate lyase synthetase	35,216	30,948	0	0	24,413	27,389
PL78_17340	PrsK, histidine kinase sensor	58,108	65,678	0	0	83,208	59,668
PL78_17335	Response regulator of citrate/malate metab.	37,723	42,923	0	0	42,125	33,852
PL78_17330	DNA-binding response regulator, NarL/Fax	7,319	4,909	0	0	7,418	2,309
PL78_17325	FbpC, ferric transporter ATP-binding subunit	55	1,582	0	0	194	622

Gene identifier	Gene Product	MH96	MH96	Δ*cspA123*/ ΔHCUI$_{YE96}$	Δ*cspA123*/ ΔHCUI$_{YE96}$	Δ*yen6*	Δ*yen6*
PL78_17326	hypothetical protein	0	52	0	0	79	112
PL78_17320	tRNA-Asn	421	448	373	94	585	343

Table 6.8: Transcripts with significantly higher expression in *Yersinia entomophaga* MH96 compared to Δ*cspA123*/ΔHCUI$_{YE96}$ during *in vivo* growth at 37 °C . Log fold-change expression values and adjusted p-value determined from *in vivo* transcriptome using voom in limma R package.

Gene identifier	Transcript start/stop	Translation start/stop	Translation start /stop	Transcript start/stop	Strand	Gene product	Log_2CPM fold-change	Adj. P-value
PL78_03740	57,123	57,172	58,800	58,800	+	Chi1, toxin complex component	2.24	0.02
PL78_03745	58,801	58,886	62,380	62,380	+	YenA1, toxin complex component	1.73	0.02
PL78_03750	62,373	62,373	66,467	66,467	+	YenA2, toxin complex component	0.99	0.08
PL78_03760	68,661	68,661	73,124	73,124	+	YenB, toxin complex component	0.98	0.09
PL78_05370	466,058	466,071	467,285	467,285	+	YcaO, ribosomal protein S12 methyltransferase accessory factor	1.89	0.01
PL78_05371	467,286	467,374	468,225	468,225	+	Putative class I SAM-dependent methyltransferase	1.26	0.07
PL78_18925	723,048	723,034	722,201	721,688	-	N-acetylmuramoyl-L-alanine amidase	0.72	0.09
PL78_17580	1,000,349	1,000,414	1,002,057	1,002,073	+	MycP, peptidase S8	0.94	0.02
PL78_17000	1,140,687	1,140,687	1,141,466	1,141,485	+	PapD	1.48	0.03
PL78_15020	1,610,998	1,610,985	1,610,302	1,610,053	-	ClpP, Clp protease	0.56	0.06
PL78_13715	1,907,875	1,907,957	1,908,523	1,908,572	+	Succinate-acetate transporter protein	0.72	0.07

Gene identifier	Transcript start/stop	Translation start/stop	Translation start /stop	Transcript start/stop	Strand	Gene product	Log_2CPM fold-change	Adj. P-value
PL78_08440	2,185,851	2,185,763	2,184,447	2,184,444	-	GltP, Glutamate/aspartate:proton symporter	0.86	0.10
PL78_12650	2,523,603	2,523,586	2,523,167	2,523,167	-	D-ribose pyranase	2.14	2.34E-03
PL78_12655	2,523,166	2,523,159	2,521,627	2,521,627	-	XylG, D-xylose ABC transporter, ATP-binding protein	2.70	1.12E-03
PL78_12660	2,521,630	2,521,630	2,520,665	2,520,665	-	RbsC, ribose ABC transporter permease	2.53	9.20E-04
PL78_12665	2,520,664	2,520,601	2,519,711	2,519,711	-	XylF, ABC-type sugar transport system	2.28	9.48E-04
PL78_12670	2,519,667	2,519,633	2,518,707	2,518,707	-	RbsK, ribokinase	1.99	3.07E-04
PL78_12675	2,518,706	2,518,695	2,517,703	2,517,665	-	CcpA, DNA-binding transcriptional regulator, LacI/PurR family	1.77	6.84E-04
PL78_12480	2,561,657	2,561,908	2,562,444	2,562,444	+	pilin type1 fimbria component protein	1.01	0.02
PL78_12475	2,562,445	2,562,540	2,563,250	2,563,288	+	fimbrial chaperone protein	1.55	0.02
PL78_12470	2,563,289	2,563,289	2,565,796	2,565,796	+	usher protein	1.38	0.02
PL78_12335	2,595,638	2,595,557	2,594,898	2,594,817	-	membrane protein	0.73	0.09
PL78_11890	2,713,331	2,713,351	2,714,094	2,714,114	+	16S rRNA (guanine(1516)-N(2))-methyltransferase	1.56	0.09
PL78_11340	2,840,585	2,840,561	2,839,140	2,839,140	-	cell wall-associated hydrolase, NlpC family	1.81	0.02

Gene identifier	Transcript start/stop	Translation start/stop	Translation start /stop	Transcript start/stop	Strand	Gene product	Log_2CPM fold-change	Adj. P-value
PL78_09636	3,200,043	3,200,058	3,200,585	3,200,585	+	RNA polymerase subunit sigma	1.09	0.03
PL78_09630	3,200,582	3,200,582	3,201,547	3,201,648	+	iron dicitrate transport regulator FecR	0.95	0.10
PL78_01415	3,811,943	3,811,943	3,812,962	3,812,970	+	membrane protein, putative aminobenzoyl-glutamate transporter	0.64	0.07
PL78_01710	3,875,523	3,875,523	3,877,568	3,877,568	+	Outer membrane cobalamin receptor protein	1.12	0.10
PL78_02435	4,036,173	4,036,173	4,035,394	4,035,394	-	EhuA, arginine ABC transporter ATP-binding protein	1.13	0.05

Table 6.9: Transcripts with significantly lower expression in *Yersinia entomophaga* MH96 compared to Δ*cspA123*/ΔHCUI$_{YE96}$ during *in vivo* growth at 37 °C . Log fold-change expression values and adjusted p-value determined from *in vivo* transcriptome using voom in limma R package.

Gene identifier	Transcript start/stop	Translation start/stop	Translation start/stop	Transcript start/stop	Strand	Gene product	Log_2CPM fold-change	Adj. P-value
PL78_05735	550,203	550,365	551,609	551,672	+	hypothetical protein, putative ahpD-like	-1.25	0.03
PL78_18360	840,050	840,050	839,316	839,316	-	carbonic anhydrase	-3.19	4.84E-04
PL78_17805	952,073	952,073	951,255	951,255	-	cysteine transporter subunit	-0.71	0.06
PL78_15970	1,384,229	1,384,203	1,383,166	1,383,166	-	thiosulfate transporter subunit	-1.25	0.02
PL78_15735	1,433,588	1,433,507	1,432,221	1,432,139	-	uracil/xanthine transporter	-1.24	0.06
PL78_15480	1,511,814	1,511,814	1,511,614	1,511,614	-	Fe-S, assembly, Iron-sulphur cluster assembly	-0.99	0.07
PL78_15475	1,512,204	1,512,202	1,511,867	1,511,815	-	Fdx, Ferredoxin	-0.92	0.04
PL78_15470	1,514,055	1,514,055	1,512,205	1,512,205	-	HscA, chaperone	-0.76	0.03
PL78_15465	1,514,831	1,514,823	1,514,299	1,514,146	-	HscB, co-chaperone	-0.99	0.02
ncRNA_19	1,514,919			1,514,832	-	ncRNA_19	-1.41	0.02
PL78_15460	1,515,281	1,515,281	1,514,958	1,514,920	-	IscA, iron-sulfur cluster assembly protein	-0.92	0.03
PL78_15455	1,515,761	1,515,731	1,515,345	1,515,282	-	IscU, Fe-S cluster formation	-1.02	0.02

Gene identifier	Transcript start/stop	Translation start/stop	Translation start/stop	Transcript start/stop	Strand	Gene product	Log_2CPM fold-change	Adj. P-value
PL78_15450	1,517,045	1,516,991	1,515,762	1,515,762	-	IscS, cysteine desulfurase	-0.86	0.02
PL78_15445	1,517,540	1,517,540	1,517,046	1,517,046	-	IscR ,transcriptional regulator	-0.59	0.10
PL78_14885	1,636,139	1,636,163	1,636,675	1,636,705	+	HpaR ,transcriptional repressor MprA	-0.86	0.03
PL78_14390	1,739,114	1,739,114	1,738,380	1,738,216	-	CysH, phosphoadenosine phosphosulfate reductase	-0.92	0.09
PL78_14385	1,740,841	1,740,841	1,739,111	1,739,111	-	CysI, sulfite reductase subunit beta	-0.90	0.07
PL78_14380	1,742,692	1,742,649		1,740,841	-	CysJ, sulfite reductase subunit alpha	-0.92	0.09
PL78_13780	1,892,512	1,892,503	1,889,270	1,889,264	-	CarB, carbamoyl phosphate synthase large subunit	-1.61	0.02
PL78_13775	1,893,750	1,893,685	1,892,519	1,892,518	-	CarA, carbamoyl phosphate synthase small subunit	-1.87	2.88E-03
PL78_13050	2,067,052	2,067,096	2,067,389	2,067,438	+	YhbY, RNA-binding protein	-0.64	0.05

Gene identifier	Transcript start/stop	Translation start/stop	Translation start/stop	Transcript start/stop	Strand	Gene product	Log_2CPM fold-change	Adj. P-value
PL78_07885	2,438,228	2,438,228	2,437,239	2,437,222	-	Sulfate transporter subunit	-1.37	0.01
PL78_12535	2,550,483	2,550,483	2,551,523	2,551,524	+	PstS, phosphate ABC transporter substrate-binding protein	-1.14	0.09
PL78_12510	2,555,327	2,555,345	2,556,181	2,556,195	+	amino acid ABC transporter substrate-binding protein	-1.45	0.02
PL78_12505	2,556,205	2,556,326	2,557,063	2,557,064	+	EhuC, amino acid ABC transporter permease	-1.14	0.09
PL78_11445	2,820,151	2,820,151	2,821,239	2,821,247	+	TauA ,taurine transporter ATP-binding subunit	-2.04	0.07
PL78_11440	2,821,248	2,821,248	2,822,015	2,822,015	+	TauB, taurine transporter ATP-binding subunit	-2.10	0.05
PL78_11430	2,822,872	2,822,872	2,823,720	2,823,928	+	TauD, taurine dioxygenase	-1.90	0.02
PL78_10560	3,000,358	3,000,348	2,999,884	2,999,829	-	PyrI, aspartate carbamoyltransferase regulatory subunit	-1.86	0.02

Gene identifier	Transcript start/stop	Translation start/stop	Translation start/stop	Transcript start/stop	Strand	Gene product	Log_2CPM fold-change	Adj. P-value
PL78_10555	3,001,329	3,001,294	3,000,359	3,000,359	-	PyrB, aspartate carbamoyltransferase catalytic subunit	-2.02	0.04
PL78_09275	3,270,077	3,270,163	3,270,849	3,270,928	+	hypothetical protein, trans-membrane helix predicted	-1.07	0.03
PL78_08900	3,341,509	3,341,494	3,340,607	3,340,557	-	MBL fold hydrolase	-1.89	0.06
PL78_02320	4,012,254	4,012,254	4,010,887	4,010,865	-	Cystathionine beta-synthase	-1.51	0.02

6.3.2 EMSA results

6.3.2.1 Recombinant Yen6 production

Recombinant Yen6 was found to be soluble and migrated at molecular weight of ~ 14.5 kDa under denaturing conditions, visualized from post-induced whole-cell and supernatant from sonicated lysate (Figure 6.6A). Oligomerization was observed when purified Yen6 was visualized under native conditions as five distinct bands ranging in size from approximately ~155 – 550 kDa and with varying abundance, suggesting multiple Yen6 subunits are combining to form a larger multi-monomeric structure (Figure 6.6B).

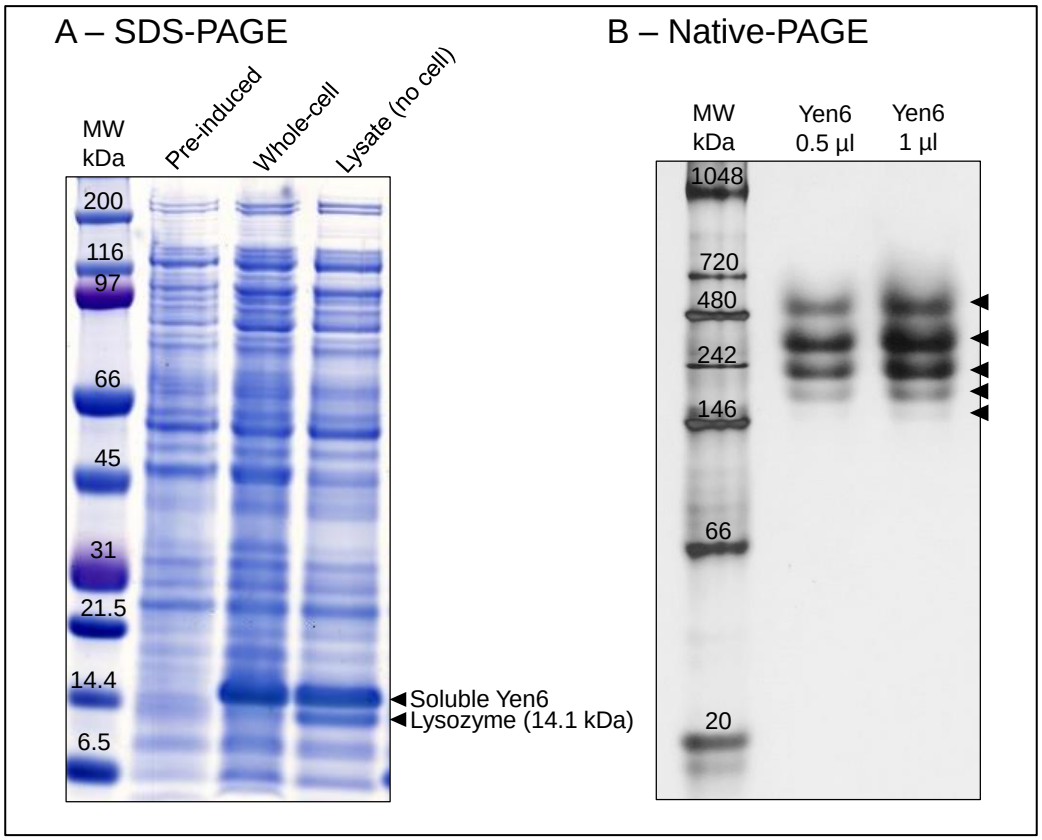


Figure 6.6: Visualization of recombinant Yen6$_{6x\text{-}His}$ by SDS- and native-PAGE .A) Coomassie stained 12 % SDS-PAGE of *E. coli* turner cells (Novagen) carrying pOPINF-*yen6*$_{6x\text{-}His}$ pre-induced, and post-induced whole-cell and lysate supernatant samples. MW = molecular weight over marker lane containing Bio-Rad unstained SDS-PAGE broad range standard with respective ladder sizes given. Black triangles correspond to recombinant Yen6 (~14.5 kDa) and lysozyme (14.1 kDa). This gel was prepared by Marion Schoof at Callaghan Innovation. B) Silver stained 12 % native-PAGE of purified Yen6. MW = molecular weight over marker lane containing Invitrogen NativeMark unstained protein standard with respective ladder sizes given. Black triangles correspond to five visible discrete bands.

6.3.2.2 Identification of palindromic motif within the promoter regions of putative *yen6* regulon genes

The promoter regions (P$_{fruc}$, P$_{ribo}$ and P$_{yhbY}$) of the gene/gene clusters that were found to respond significantly to the deletion of *yen6* (i.e., fructose cluster, ribose cluster and *yhbY*) were investigated for potential binding motifs and a shared palindromic motif was identified from multiple times within all three promoter regions (twice within the P$_{fruc}$ and P$_{ribo}$ promoter region and once within P$_{yhbY}$) (Figure 6.7). The promoter region of putative Yen-TC component activator, *yen7*, was also assessed using EMSA and the P$_{y7}$ probe contained a 9 nt perfect palindromic repeat (5'-GATATATTT-3') called rep1,

that was found to overlap with the 35-box and start codon of *yen7* (Chapter 5, section 5.3.2.2, Figure 5.17).

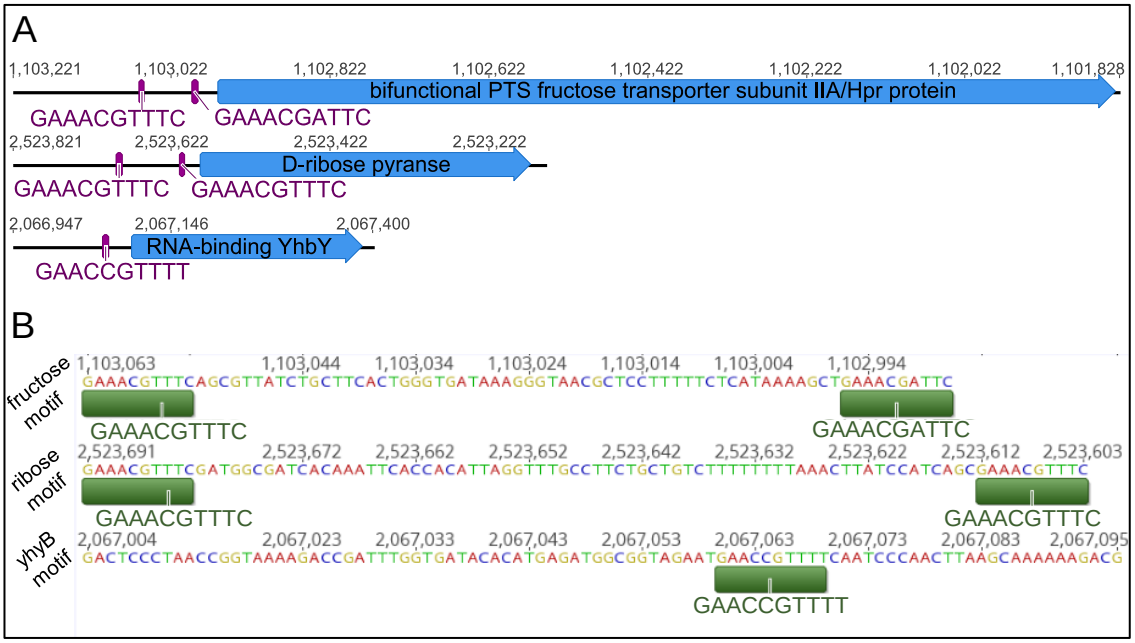


Figure 6.7: Palindromic motifs identified within the P_{fruc}, P_{ribo} and P_{yhbY} regions using Geneious repeat finder and motif search (maximum 2 mismatches allowed). A) Location of repeat motifs and protein coding region in the genome of MH96 and B) sequence-level view of repeat binding motifs identified.

6.3.2.3 EMSA

Recombinant $Yen6_{6x\text{-}His}$ was found to have specific binding activity against the P_{fruc}, P_{ribo} and P_{yhbY} based on EMSA analysis (Figure 6.8, Figure 6.9 and Figure 6.10). Binding of $Yen6_{6x\text{-}His}$ to the P_{fruc} probe showed clear specific binding and with increased Yen6 concentrations three DNA-protein complexes formed, possibly due to protein oligomerization/aggregation or the presence of multiple binding sites (Figure 6.8). Relatively weaker specific binding activity of $Yen6_{6x\text{-}His}$ was found for P_{ribo} and P_{yhbY} and only a small amount of DNA shifted, even when in reactions with the highest concentration of protein. A single DNA-protein complex formed between $Yen6_{6x\text{-}His}$ and P_{ribo}, which disappeared in the presence of unlabelled competitor probe, providing evidence of weak specific binding activity (Figure 6.9). As a control, $Yen6_{6x\text{-}His}$ was tested with an unrelated probe P_{crcZ} promoter DNA form *Ps. fluorescens* SBW25 as a negative control and no shifted band was observed (Figure 6.9). The probe P_{yhbY} formed two protein free bands ($Band_A$ and $Band_B$) on the EMSA, and at higher concentrations

of Yen6$_{6x\text{-}His}$ two DNA-protein complexes formed (Figure 6.10). In the presence of excess unlabelled competitor, the DNA-protein complexes became less visible, but one of the bands could not be completely reduced Figure 6.10).

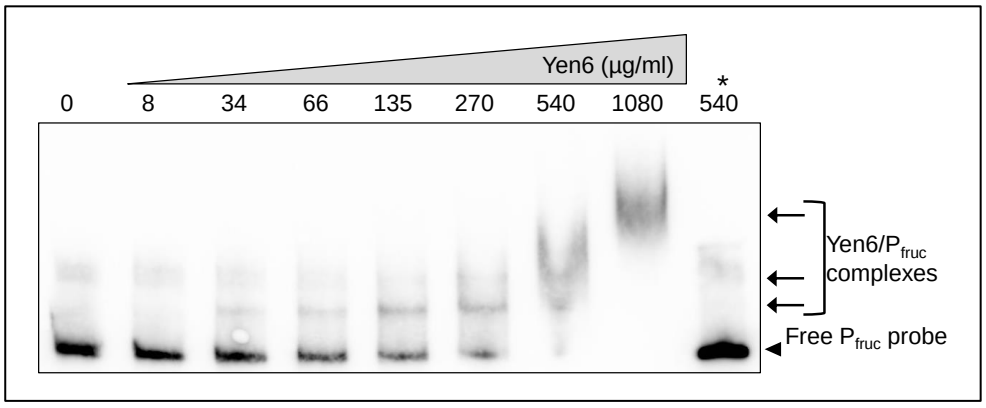


Figure 6.8: EMSA showing specific binding of Yen6 to the promoter region of gene for PTS fructose transporter subunit IIA/Hpr, P$_{fruc}$. EMSA was performed using 20 nM biotin-labeled 324 bp DNA probe corresponding to the promoter region containing two palindromic putative binding motifs. Purified Yen6$_{6x\text{-}His}$ was added with increasing concentration. A 200-fold molar excess of unlabelled probe (competitor DNA) was added and 1 µg salmon sperm DNA (non-specific competitor). * approximately 200 nM unlabelled specific competitor DNA added.

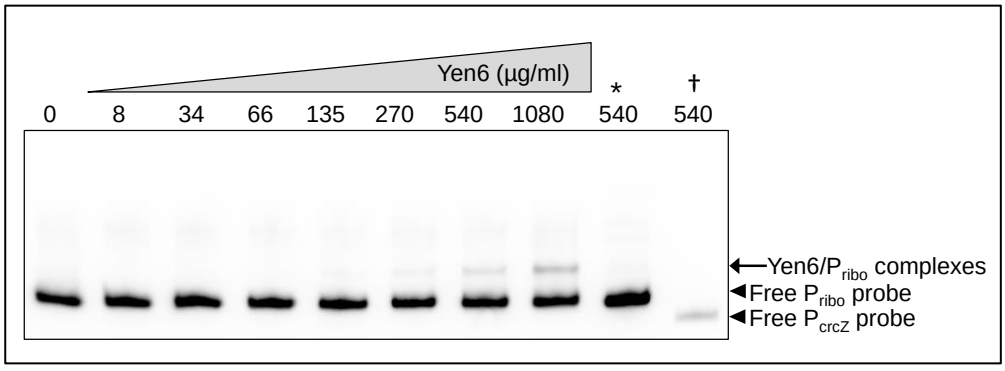


Figure 6.9: EMSA showing specific binding of Yen6 to the promoter region of *rbsD*, P$_{ribo}$. EMSA was performed using 20 nM biotin-labeled 409 bp DNA probe corresponding to the promoter region containing two palindromic putative binding motifs. Purified Yen6$_{6x\text{-}His}$ was added with increasing concentration. A 200-fold molar excess of unlabelled probe (competitor DNA) was added and 1 µg salmon sperm DNA (non-specific competitor). * approximately 200 nM unlabelled specific competitor DNA added. † 20nm unrelated probe P$_{crcZ}$ promoter DNA form *Ps. fluorescens* SBW25 as a negative control.

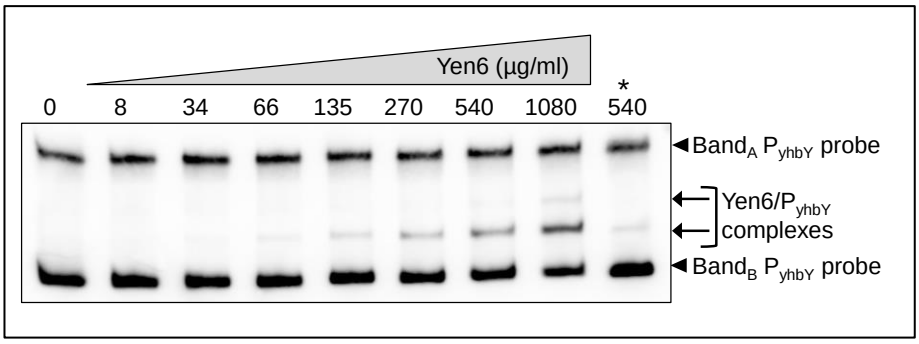


Figure 6.10: EMSA showing specific binding of Yen6 to the promoter region of *yhbY*, P_{yhbY}. EMSA was performed using 20 nM biotin-labeled 312 bp DNA probe corresponding to the promoter region containing a single putative binding motifs. Purified $Yen6_{6x\text{-}His}$ was added with increasing concentration. A 200-fold molar excess of unlabelled probe (competitor DNA) was added and 1 µg salmon sperm DNA (non-specific competitor). * approximately 200 nM unlabelled specific competitor DNA added.

Analysis by EMSA of $Yen6_{6x\text{-}His}$ affinity for the promoter region of *yen7* did not support specific binding activity (Figure 6.11). A weak shifted DNA-protein band was visible with higher concentrations of $Yen6_{6x\text{-}His}$ but this band remained (though difficult to see) in the presence of unlabelled competitor DNA. Also, P_{y7} produced two distinct protein-free bands and the large one ($Band_C$) became reduced in the presence of excess unlabelled DNA competitor, a result that cannot be explained. Higher order DNA structures were observed for P_{fruc}, P_{yhbY} and P_{y7} since these DNA probes do not show more than one band when visualized by agarose gel electrophoresis (supplemental Figure S29).

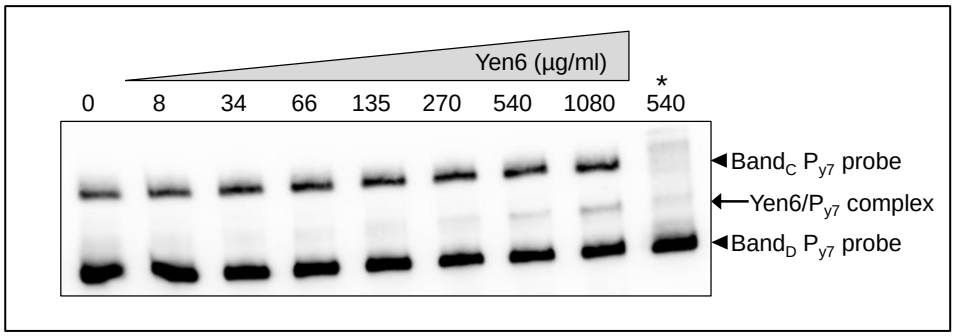


Figure 6.11: EMSA showing specific binding of Yen6 to the promoter region of yen7, P_{y7}. EMSA was performed using 20 nM biotin-labeled 295 bp DNA probe corresponding to the promoter region containing a repetitive binding motif rep1 (Chapter 5, section 5.3.2.2, Figure 5.17). Purified $Yen6_{6x\text{-}His}$ was added with increasing concentration. A 200-fold molar excess of unlabelled probe (competitor DNA) was added and 1 µg salmon sperm DNA (non-specific competitor). * approximately 200 nM unlabelled specific competitor DNA added.

6.3.3 Phenotype microarray of MH96 and *yen6* at 25 and 37 °C

Results of the Δ*yen6 in vivo* RNA-seq determined deletion of *yen6* results in significantly different expression of two different sugar uptake systems (i.e., fructose phosphotransferase system (PTS) and ribose/hexose ABC-type transporter system) suggesting *yen6* may play a role in in-host metabolic processes occurring at 37 °C *in vivo.* To further characterize effects of the *yen6* deletion on carbon utilization in MH96, phenotype microarray was carried out on MH96 and Δ*yen6* strains at 25 and 37 °C to identify any differences in ability to utilize 190 difference of carbon sources under aerobic conditions using the Omnilog phenotype microarray system (Biolog Inc.).

Based on the phenotype microarray very limited differences in sole-carbon utilization by MH96 and Δ*yen6* were identified (Table 6.10). Shifts in sole-carbon utilization by the Δ*yen6* strain was found to be somewhat spurious, but subtle reductions in the utilization of carbohydrate sources D-mannose and D-mannitol, amino acids L-asparagine and L-histidine and polymers laminarin and pectin were noted for 25 °C compared to wild-type. At 37 °C (the temperature that *yen6* is highly expressed) only the utilization of amino acid D-alanine was identified as possibly affected by the deletion. While, *yen6* was shown to have transcriptional effects on fructose and ribose gene clusters, there was no observable difference in utilization of fructose or ribose (or L-arabinose which is a hexose analogue ribose) as sole-carbon sources in the Δ*yen6*

strain compared to wild-type. Overall the findings were somewhat inconsistent among replicate batch plates and would need to be further validated.

Table 6.10: Differences in aerobic sole carbon source utilization by *Yersinia entomophaga* MH96 and Δ*yen6* strains at 25 and 37 °C by phenotype microarray. (+) indicates positive carbon-source utilization per replicate plate, green cells = maximum value/minimum value > 2 and yellow cells = maximum value/minimum value ≤ 2. Blank cells represent no growth curve detection based on threshold criteria.

		Strain	
Carbon source	Chemical	Δ*yen6*	MH96
Incubation temperature: 25 °C			
carboxylic acid	Succinic acid		+
carbohydrate	D-Mannose	+ +	+ +
carbohydrate	D-Mannitol	+ +	+ +
amino acid	L-Asparagine	+ +	+ +
amino acid	L-Histidine	+	+ +
polymer	Laminarin		+
polymer	Pectin		+
Incubation temperature: 37 °C			
amino acid	D-Alanine	+	+ +

6.4 Discussion

One previously reported finding of this thesis was that two mutants strains of MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6*, were shown to have significantly attenuated virulence during intrahemocoelic injection of *G. mellonella* at 37 °C compared to wild-type (Chatper 5, section 5.3.1.2). It was also found that Δ*cspA123*/ΔHCUI$_{YE96}$ did not have attenuated virulence by injection of *G. mellonella* at 25 °C compared to wild-type (Δ*yen6* was not tested at 25 °C). Bioassay results at 25 and 37 °C, also previously presented in this thesis (Chapter 5, section 5.3.1.2) and reported in literature (Hurst et al., 2015) have shown that shifting temperature from 25 to 37 °C does not affect the median lethal dose (LD_{50}) of MH96 during intrahemocoelic infection of *G. mellonella*; however, MH96 is still thought to have reduced pathogenic capability under these conditions at 37 compared to 25 °C because previous RNA-seq data revealed significant repression of co-secreted VFs, including Yen-TC/TC-associated factors at 37 °C, but only during intrahemocoelic infection of *G. mellonella* (Chapter 4, section 4.3.8, Table 4.13). So, as follow-up *in vivo* RNA-seq of Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6* were conducted at 37 °C

during intrahemocoelic infection of *G. mellonella* to further characterize the roles of *cspA123* and *yen6* as putative virulence regulators during infection of *G. mellonella* at mammalian host temperature. Subsequently, the transcriptome of Δ*cspA123*/ΔHCUI$_{YE96}$, revealed a 17.5 Kb excision in the genome, so while significant transcriptional effects on numerous genes were identified compared to wild-type in this strain, it was not possible to attribute transcriptional changes to the targeted *cspA123* deletion or excision HCUI$_{YE96}$, and is the focus of Chapter 7. Subsequently, follow-up phenotyping work reported in this chapter, including EMSA protein-DNA binding assays and carbon utilization testing, only focused on Yen6$_{6x\text{-}His}$, and the Δ*yen6* strain, respectively.

6.4.1 Deletion of *yen6* resulted in differential expression of fructose and ribose uptake/metabolic gene clusters during infection of *G. mellonella* at 37 °C

The *in vivo* transcriptome of Δ*yen6* at 37 °C revealed the regulon of Yen6 contains mostly genes related to central carbon metabolism as well as the gene for RNA-binding protein YhbY. These results also showed that deletion of *yen6* does not result in any significant transcriptional changes to *yen7* or Yen-TC component genes. Based on these findings Yen6 is revealed as a pleiotropic regulator that can simultaneously activate *rbsD-xylG-rbsC-xylF-rbsK-ccpA* and repress *IIA-fruK-IIBC* as well as *yhbY*. Additional work testing the specific binding activity of recombinant Yen6$_{6x\text{-}His}$ against DNA containing a putative binding motif within the promoter regions of *rbsD*, *IIA*/*Hpr* and *yhbY* were undertaken to determine the direct binding targets of recombinant Yen6 (section 6.4.4 below). Further to this, the binding affinity of the *yen7* promoter region was also tested with Yen6$_{6x\text{-}His}$.

The effects of the *yen6* deletion on the fructose and ribose uptake/acquisition clusters are of particular interest, especially since co-regulation of virulence and metabolism is known to be related to host-adaptation among other pathogenic *Yersiniae* spp. (Heroven and Dersch, 2014) and more broadly (Kwaik and Bumann, 2013) and metabolic shifts may represent a specific host-adaptation by MH96 for survival within a mammalian host, especially since Δ*yen6* has attenuated virulence against *G. mellonella* during intrahemocoelic infection at 37 °C (not tested at 25 °C) (Chapter 5, section

5.3.1.2, Table 5.2, Figure 5.8). *E. coli* can transport D-ribose (primarily as β-D-ribopyranose form) into the cytoplasm by ABC-type high-affinity transporter called RbsABC (Horazdovsky and Hogg, 1987; Park et al., 1999; Song and Park, 1998), although allose (Kim et al., 1997) and xylose ABC transporters (Song and Park, 1998) and an altered glucose PTS (Oh et al., 1999) have also been shown to function as low-affinity transporters of ribose as well. More recent investigations into the ABC transport mechanism suggest RbsABC share features of the both type I and II ABC importer systems, but likely represents a distinct superfamily of increasingly diverse transporters (Clifton et al., 2015).

Similar to MH96, *E. coli* encodes an inducible *rbsDACBK* operon that contains genes for ribose transport and initial metabolic enzymes required for ribose utilization (Laikova et al., 2001; Lopilato et al., 1984; Mauzy and Hermodson, 1992). In MH96, the *xlyG* and *xlyF* genes encode xylose ABC transporter ATP- and substrate-binding domains, respectively and should be considered analogous to (and possibly re-assigned to *rbsA* and *rbsB* of *E. coli*) (Horazdovsky and Hogg, 1987; Park et al., 1999; Song and Park, 1998), which combine with the permease RbsC to form the high-affinity transporter structure (Barroga et al., 1996; Clifton et al., 2015). MH96 encodes three other genes annotated as xylose ABC-transporter components, including PL78_17225 and PL78_02115 (both annotated as *xylG*) and PL78_02115 annotated as *xylF*. Unlike *E. coli*, where the gene for the LacI-type *rbs*DACBK repressor, *rbsR* is induced separately transcriptome data of MH96 identified that *ccpA* (orthologous to the *rbsR*) was expressed within the same operon as the ribose uptake/utilization genes based operon predictions generated by Rockhopper.

The β-pyranose form of D-ribose is the most abundant ribose available to bacteria in nature (Šišak et al., 2010) and represents an important sugar that is a component of nucleic acids and other biomolecules and also as an energy source. Once β-D-ribopyranose is transported from the periplasm by the RbsABC transporter, the sugar must undergo D-ribose degradation pathway before assimilation, which involves *rbsD* and rbsK, that encode for putative D-ribose pyranase and ribokinase, respectively and are also found in the *rbsD-xylG-rbsC-xylF-rbsK-ccpA* ribose operon of MH96. RbsD is a cytoplasmic sugar-binding protein that catalyse the conversion D-ribopyranose to D-

ribofuranose (Kim et al., 2003b) and RbsK phosphorylates D-ribofuranose into ribose 5-phophate (Sigrell et al., 1998) both reactions requiring ATP energy. Once in the ribose 5-phosphate form it can be further metabolized by the PPP (also called the hexose monophosphate shunt), which is an important component of cellular metabolism, both in production of key metabolites for nucleotide synthesis, biosynthesis of aromatic amino acids and vitamins (Lam and Winkler, 1990; Zhao and Winkler, 1994), as well as production of NADPH as a source of reducing power (Grant, 2008; Kovářová and Barrett, 2016; Spaans et al., 2015).

Genes related to ribose uptake and metabolism have been implicated in virulence in mammalian pathogens before. For example, an *E. coli* mutant deficient for *rbsK* was found to have minor defects in long term maintainance of the population within the mouse intestine but was not impaired with respect to colonization compared to wild-type (Chang et al., 2004). Genes for putative ribose uptake protein and ribose transporter RbsD were also found to be strongly induced in *Enterococcus faecalis* following three hours exposure to urine, suggesting these genes are induced during bladder infection (Hanin et al., 2010). In whole-genome microarray study of *Y. pestis* genes for ribose permeases and uptake of pentoses ribose, xylose and arabinose were found to be more highly expressed within the gut of blocked flea vector compared to *in vitro* growth conditions (Vadyvaloo et al., 2010) and an earlier microarray study of *Y. pestis* also found that *rbsK* and *rbsD* underwent early induction at 37 °C implying a potential role during mammalian infection (Motin et al., 2004). Even though ribose uptake/utilization have been found to be highly expressed by *Y. pestis* upon temperature transition or within the digestive tract of the flea vector, one of the key PPP enzymes, *zwf* (glucose-6-phosphate dehydrogenase) has recently acquired a missense mutation preventing the first step of PPP in epidemic strains of plague (Bearden et al., 2009). In *Y. pseudotuberculosis* transcriptome study of growth in human plasma compared to Luria Bertani broth and at 37 and 28 °C and found a gene for ribose-phosphate pyrophsphokinase, *prsA* had significantly greater expression in human plasma compared to broth culture, and both *rbsC* and *rpi* had significantly higher expression at 37 compared to 25 °C (Rosso et al., 2008).

In addition to activating the ribose transport and utilization cluster, deletion of *yen6* also results in significant repression of the fructose transport and utilization operon *IIA/hpr-fruK-IIBC* under *in vivo* conditions at 37 °C, which encode for the PTS components IIA/Hpr and IIBC and 1-phosphofructokinase, FruK, which are required for uptake of β-D-fructopyranose into the cytoplasm and production of fructose-1,6-biphosphate, an important substrate in glycolysis, respectively (Kornberg, 2001). Since, these result suggest that Yen6 may act transcriptionally to both activate ribose uptake/utilization and repress fructose uptake/ utilization, it is plausible that one of the primary functions of Yen6 is to control carbon fluxes within the central metabolic biosynthetic pathways of MH96. Further assessment of sole-carbon utilization of MH96 and Δ*yen6* by phenotype microarray were undertaken in order to reconcile potential metabolic effects of the deletion of *yen6* on carbon utilization by MH96 and discussed in section 6.4.5 below.

6.4.2 Deletion of *yen6* resulted in differential expression of the gene for an RNA-binding protein, YhbY during infection of *G. mellonella* at 37 °C

In addition to differential expression detected for the gene clusters *rbsD-xylG-rbsC-xylF-rbsK-ccpA* and *IIA/hpr-fruK-IIBC* in the *yen6* mutant compared to MH96 under *in vivo* conditions at 37 °C (Figure 6.5 above), the only other gene with significantly DE was *yhbY,* which encodes for an RNA-binding protein that belongs to a group hypothetical proteins widely distributed among archaea, eubacteria and plants (Pfam code UPF0044) (Barkan et al., 2007). The crystal structure of *E. coli* YhbY shares a lot of structural similarities (but little amino acid sequence similarity) with the C-terminal domain of the translation initiation factor 3 (IF3) (Ostheimer et al., 2002), which plays an important role in protein translation, especially fidelity of selection of $tRNA^{fMet}$ and start codon proof-reading (Hussain et al., 2016; Sharma et al., 2018). Also, in *E. coli*, YhbY can be co-sedimented with the 50S subunit (Jiang et al., 2006, 2007) and pre-50S particles (Barkan et al., 2007). In *Ralstonia solanacearum,* a *yhbY* homolog was found to play a role in virulence because a deficient mutant was found to have decreased virulence in tomato and *Arabidopsis* (Franks et al., 2008). In *Yersinia* spp., *yhbY* was found to have DE in a *rovA* deficient mutant strain of *Y. pseudotuberculosis* compared to

wild-type (Nuss et al., 2015), but that was not the case for *Y. enterocolitica* or *Y. pestis* (McNally et al., 2016). The function of *yhbY* in MH96 during infection at 37 °C remains to be determined.

6.4.3 Deletion of *yen6* has no significant effects on *yen7* or Yen-TC component genes at 37 °C *in vivo* and Yen6$_{6x\text{-}His}$ has no specific binding activity for the promoter region of *yen7*

The transcriptome of Δ*yen6* provided additional evidence that transcription of *yen7* and Yen-TC component genes are not influenced by *yen6* (at least during infection at 37 °C). Furthermore, EMSA analysis of Yen6$_{6x\text{-}His}$ affinity for P_{y7} did not provide evidence to support specific binding activity of Yen6 for the promoter region of Yen7 (although the reaction of Yen6$_{6x\text{-}His}$, biotin-labelled P_{y7} and excess unlabelled P_{y7} resulted the top-most protein-free DNA band disappearing, which could not be explained). Taken together, these results provide strong evidence that there is not likely a direct transcriptional interaction between Yen6 and *yen7*.

These findings are consistent with previous phenotyping by SDS-PAGE, that showed protein production and secretion by Δ*yen6* and Δ*yen6yen7* strains did not result in any major differences when grown at 25 and 37 °C *in vitro* compared to wild-type (Chapter 5, section 5.3.1.1, Figure 5.3 and Figure 5.4). Also, previous investigations using the transcriptional reporter P_{chi1}::*lacZ* strain (proxy for Yen-TC translational activity), determined that when *yen6* was deleted (i.e., Δ*yen6* strain) significantly higher β-gal production was observed at 25 °C under *in vitro* conditions compared to in the wild-type genetic background (Chapter 5, section 5.3.3.6.3, Figure 5.27); however, preliminary attempts at *trans* complementation of Δ*yen6* using either arabinose-inducible vectors, pBAD-*yen6* (protein coding region only) or pBAD-*yen67as* (protein coding region and portion of predicted 3'UTR-encoded anti-sense RNA) or the mid-copy cloning vector carrying *yen6* under its natural promoter (pACYC184-P_{yen6}-P_{yen7}) were not entirely successful and had opposite effects on the production of β-gal by the P_{chi1}::*lacZ* reporter strain, supporting a more complicated interaction (Chapater 5, section 5.3.3.6.3, Table 5.8, Table 5.10). So, to this end, the EMSA results and the *in vivo* transcriptome of Δ*yen6* at 37 °C both aligned with previous findings reported in

Chapter 5, that overall do not support Yen6 as a direct transcriptional regulator of *yen7* nor does Yen6 likely have any major downstream effects on Yen-TC component genes, as originally hypothesized.

6.4.4 EMSA validate Yen6 as transcriptional regulator that can interact with DNA from the promoter of the fructose and ribose uptake/acquisition clusters and RNA-binding protein, YhbY

Using EMSA analysis, the regulon of Yen6 was validated because $Yen6_{6x\text{-}His}$ was shown to have specific binding affinity for the promoter regions of *IIA/hpr-fruK-IIBC* (i.e., P_{fruc}), *rbsD-xylG-rbsC-xylF-rbsK-ccpA* (i.e., P_{ribo}), and *yhbY* (i.e., P_{yhbY}),. These results confirm that the palindromic 10 bp sequence (5′-GAA[A/C]CG[A/T]TT[C/T]-3′), which is located twice within P_{fruc} and P_{ribo} and once within P_{yhbY} is likely the binding motif of Yen6. These results validate the findings of the molecular phenotyping by *in vivo* RNA-seq of the Δ*yen6* strain at 37 °C, where significantly greater expression of *IIA/hpr-fruK-IIBC* operon and *yhbY* and significantly reduced expression of *rbsD-xylG-rbsC-xylF-rbsK-ccpA* compared to wild-type were identified. Combined with the EMSA results, Yen6 is now known to be a transcriptional regulator that can simultaneously repress the fructose uptake/acquisition cluster and *yhbY*, while also activating the ribose uptake/acquisition cluster during *in vivo* conditions at 37 °C.

The binding affinity of $Yen6_{6x\text{-}His}$ was strongest against P_{fruc} and comparatively weaker against P_{ribo} and P_{yhbY}, even though P_{fruc} and P_{ribo} share very similar imperfect repeats of the binding motif. Many of the biotin-labelled DNA probes used for the EMSA demonstrated higher order structures, which could stabilize DNA-protein reactions. Further EMSA analysis should be carried out with smaller DNA probes, which would help to reduce possible confounding effects associated higher order DNA structure.

6.4.5 Phenotype by microarray only identifies potentially subtle effects on utilization of sole carbon source by MH96 at both 37 and 25 °C

In vivo RNA-seq of the Δ*yen6* at 37 °C identified the regulon of Yen6 includes two sugar uptake/metabolism-related clusters, fructose and ribose and based on these additional investigations into the metabolic phenotype of Δ*yen6* were carried out.

Carbon utilization by MH96 and Δ*yen6* at 25 and 37 °C was assessed the Omnilog phenotype microarray system, which found that the deletion of *yen6* had limited effects on sole carbon utilization by MH96, especially at 37 °C. While some differences were noted among batches (i.e., replicate microarray plates), the results were overall inconsistent, and many of the subtle differences observed between the two strains requires further validation. Such limited differences in metabolic capabilities of MH96 compared to Δ*yen6* suggest that *yen6* may not have a primary role in regulating metabolism at all.

Transcriptome data support one of the key regulatory functions of Yen6 is the activation of the ribose uptake/utilization operon *rbsD-xylG-rbsC-xylF-rbsK-ccpA*, which is necessary to provide ribose for the PPP. The PPP is an important metabolic pathway in bacteria that generates carbohydrate precursors for biosynthesis and acts as a major carbon exchange center (Lam and Winkler, 1990; Zhao and Winkler, 1994). Another important function of PPP is the production of NADPH, which is an important reducing agent involved in reductive biosynthesis and resistance to oxidative stress (Grant, 2008; Kovářová and Barrett, 2016; Spaans et al., 2015). Based on the evidence form the microarray phenotyping suggesting *yen6* has limited effects on metabolism of MH96, it is now suspected that *yen6* plays a role in stress response during infection of *G. mellonella* at 37 °C, by activating ribose uptake systems and repressing fructose uptake systems to result in carbon flux to the PPP which is a survival strategy to increased NADPH in order to combat oxidative stress during infection. This response to oxidative stress hypothesis supports the findings of previously reported *in vivo* transcriptome (Chapter 4, section 4.3.8, Figure 4.20) of significantly higher expression of several other genes related to uptake of organic and inorganic sulfate, iron, nickel, cold-shock proteins, acid-resistance proteins, and multidrug transporters that were all found to be highly expressed at 37 °C during *in vivo* conditions by MH96 (Chapter 4, supplemental Table S9).

Bacterial pathogens are equipped with many strategies to deal with oxidative stress that is encountered within the host environment, including detoxification strategies involving reducing thiols, such as glutathione (GSH) or iron-sulfur (Fe-S) cluster, for example (Van Loi et al., 2015; Reniere, 2018; Turnbull and Surette, 2010; Zeller and

Klug, 2006). The polyamine GSH is a known potent antioxidant, containing a reactive Cys thiol core (Newton and Fahey, 2017; Reniere, 2018). In *E. coli*, synthesis of GSH is a two-step process involving glutamate cysteine-lyase and glutathione synthetase, encoded by *gshA* and *gshB*, respectively (Newton and Fahey, 2017). Similarly, Fe-S clusters are well known redox centers in the cell and have been shown to act as environmental sensors of oxidative and nitrosative stress (Kobayashi et al., 2014; Miller and Auerbuch, 2015). MH96 encodes homologous genes related to Fe-S (i.e., *iscR-iscS-iscU-iscA-hscB-hscA-fdx*-Fe-s assembly/PL78_15445 – PL78_15480) as well as glutamate cysteine-lyase (*gshA*/PL78_14695) and glutathione synthase (i.e., *gshB*/ PL78_09425) and several glutaredoxins (i.e., *grxA, grxC, grxD* and *nrdH*/PL78_02370, PL78_07935, PL78_05900 and PL78_14940) and it was noted that the entire Fe-S gene cluster was more highly expressed in the Δ*cspA123*/ΔHCUI$_{YE96}$ mutant compared to wild-type during infection of *G. mellonella* at 37 °C, albeit the transcriptional regulation of the Fe-S gene cluster cannot be attributed to the *cspA123* or the excision of the unstable region.

Another major clue that *yen6* is likely involved in a stress response is the co-expression of numerous genes related to sulfate, thiosulfate, taurine and methionine uptake that were identified previously as having higher expression at 37 °C *in vivo* (Chapter 4, section 4.3.8, Figure 4.20). As discussed above, sulfur is the key component of the reactive center of Fe-S as well as the Cys found in the glutathione and glutaredoxins, so upregulation of sulfur acquisition genes during the same conditions that Yen6 is expected to function to shift central carbon metabolism towards generating extra NADPH makes a feasible hypothesis for a role of Yen6 in response to oxidative stress during infection at 37 °C. It was also noted, that like the Fe-S cluster, an additional nine genes related to sulfur acquisition were identified in the transcriptome of the Δ*cspA123*/ΔHCUI$_{YE96}$ mutant as more highly expressed compared to wild-type, suggesting a possibility that *cspA123* and/or genes located on excisable region HCUI$_{YE96}$ are involved in the regulation of sulfur uptake and Fe-S cluster formation. Given the evidence that sulfur acquisition was an overriding theme of *in vivo* transcriptome of MH96 at 37 °C, these finding somewhat align with a functional role for *yen6* under these conditions to adapt to high levels of oxidative stress by fluxing more carbon into the PPP to generate necessary NADPH to oxidize the reactive cores of Fe-S and GSH.

While the role of Yen6 in oxidative stress response remains to be further tested, perhaps previous unexplained findings observed for *yen6* over-expression experiments also support this theory. When *yen6* was over-expressed using the arabinose induction vector pBAD, a consistent increase in plasmid retention and positive effects on cell growth were observed compared to strain carrying empty vector at 25 °C *in vitro*. In a 2007 study, a common mechanism of cellular death by antibiotics (including ampicillin/carbenicillin) was found to be a result of oxidative damage by stimulation of hydroxy radicals, regardless of drug-target (Kohanski et al., 2007, 2010). It is plausible that oxidative stress resistance explains why *yen6* over-expressed *in vitro* in the presence of carbenicillin or ampicillin consistently results in a positive effect on growth and plasmid retention compared to a strain with empty vector. Furthermore, *in vivo* β-gal analysis of the P_{yen6}::*lacZ* reporter stain also identified a much stronger induction of this strain during *in vivo* conditions at 37 compared to 25 °C (Chapter 5, section 5.3.3.6.7, Figure 5.31), which was consistent with expression of *yen6* under the same conditions as well based on transcriptome analysis (Chapter 4, section 4.3.8, Table 4.13). All of these findings indicate expression and translation of *yen6* is tightly regulated and induced specifically under *in vivo* conditions were MH96 is encountering oxidative stress. Further experimental work to identify potential signals of oxidative stress that trigger induction of *yen6* is an important next step for validating this new hypothesis as well as assessing oxidative resistance of Δ*yen6* compared to wild-type.

6.4.6 Transcriptome of Δ*cspA123*/ΔHCUI$_{YE96}$ identifies excision of unstable holin-endolysin secretion system and citrate fermentation cluster containing island

In vivo transcriptome of the Δ*cspA123*/ΔHCUI$_{YE96}$ strain identified a 17.5 Kbp excision, which became the focus of Chapter 7. In addition to the excised genes, 87 transcripts were found to have significant DE in the Δ*cspA123*/ΔHCUI$_{YE96}$ strain compared to wild-type under *in vivo* conditions at 37 °C. Among the DE transcripts identified included a portion of the Yen6 regulon (*rbsD-xylG-rbsC-XylF-rbsK-ccpA* and gene for RNA-binding protein *yhbY*), several Yen-TC component genes as well as a number of other potentially important putative VFs (i.e., usher-chaperone fimbriae, and sulfur

acquisition systems etc.) identified from previous *in vivo* RNA-seq investigations. These results suggest either HCUI$_{YE96}$ and/or *cspA123* are central to virulence regulation in MH96, but it is impossible to determine what transcriptional responses are associated with what deleted element from the genome in this strain, so additional *in vivo* RNA-seq should focus on characterizing the molecular phenotype of Δ*cspA123* and ΔHCUI$_{YE96}$, which would resolve much of the current confusion. Further details related to HCUI$_{YE96}$ are reported in Chapter 7.

Chapter 7. Characterization of an excisable holin-endolysin secretion system and citrate fermentation genomic island from *Yersinia entomophaga* MH96

7.1 Introduction

7.1.1 Horizontal gene transfer – a driver of bacterial genome evolution

Over twenty years of genome sequencing has revealed the mosaic-like and unstable structure of bacterial genomes, which is largely attributed to genome-wide rearrangement events and horizontal gene transfer (HGT) including the movement of mobile genetic elements (MGEs) (Boyd et al., 2009; Darmon and Leach, 2014; Frost et al., 2005; Land et al., 2015; Soucy et al., 2015). Despite the importance of MGEs in shaping modern prokaryotic genomes, these elements have remained somewhat understudied and currently lack a consistent classification system (Oliveira Alvarenga et al., 2018). A variety of different type of MGEs have been described and the most common types include transposons, insertion sequences, integrons, conjugative plasmids, bacteriophages, integrated prophages, genomic islands (GIs) - including pathogenicity islands (PAIs) and integrative and conjugative elements (ICEs) (Abebe-Akele et al., 2015; Frost et al., 2005; Oliveira Alvarenga et al., 2018; Wilkinson et al., 2009) but the focus here will be PAIs/GIs, which are considered a major driver of genome evolution among prokaryotes (Juhas et al., 2009; Rodriguez-valera et al., 2016), especially pathogens (Hacker and Carniel, 2001; Sui et al., 2009).

Usually occurring at repetitive sequences associated with tRNA genes, PAIs/GIs are associated with high diversity hotspots in the genome, evolving more rapidly than the core genome by frequent site-specific recombination events (Rodriguez-valera et al., 2016). The arrangement of genes within PAIs/GIs is modular, with blocks of genes sharing similar functions located together in clusters called modules and islands typically evolve by acquisition or loss of entire modules (Ogier et al., 2010; Rodriguez-valera et al., 2016). Usually (but not always) a mobilization module is associated with one end of the island and consists of remnant phage-related integrases or transposases required for mobilization of their cognate element (Bellanger et al., 2014; Wozniak and Waldor, 2010). Often, PAIs/GIs support gain-of-function phenotypes including,

symbiosis (Piel et al., 2004; Sullivan et al., 2002), metabolic adaptation (Gaillard et al., 2006), heavy metal resistance and iron acquisition (Larbig et al., 2002), antimicrobial resistance (Katayama et al., 2000; Luck et al., 2001), secondary metabolite synthesis (Egan et al., 2001; Metsä-Ketelä et al., 2002) and virulence (see below).

More specifically, PAIs are found within the genomes of pathogenic variants but are less frequently found within the genomes of avirulent relatives and represent some of the best studied GIs to date (Hacker and Carniel, 2001; Hacker et al., 1997; Kaper and Hacker, 2000). Usually PAIs encode genes related to virulence (i.e., exotoxins, adhesins, invasins, capsule synthesis, iron acquisition, secreted effectors and secretion systems) (Fischer et al., 2010; Kaper and Hacker, 2000) and like most GIs, are recognized by atypical G+C content and codon usage compared to the core genome, found within close proximity to tRNA genes and/or being flanked by direct repeats and may have evidence of multiple recombination events (i.e., multiple short repetitive elements), insertion sequences or remnant bacteriophage integrases or transposases (Boerlin, 2004; Kaper and Hacker, 2000; Schmidt and Hensel, 2004). With respect to mobility, PAIs can be highly variable - some remain mobile and contain a conjugation module (i.e., ICEs) (Johnson and Grossman, 2015) or viral machinery (integrated prophages) (Boyd, 2012), others are mobilized *in trans* by hijacking transmission machinery of other MGEs (Carpenter et al., 2016; Guédon et al., 2017; Moon et al., 2015, 2016) and still others demonstrate a more solid integration into the host genome, having lost the ability to mobilize all together (Kaper and Hacker, 2000). Although such elements are now recognized as ubiquitous among bacterial genomes and are major contributors to important evolutionary processes like pathogen emergence and antibiotic resistance, a clear understanding of the evolutionary relationships between integrated prophages, plasmids, ICEs and PAIs/GIs remains unclear at this time.

7.1.2 Pathogenicity islands in entomopathogenic bacteria

Among entomopathogenic bacteria, such as *Photorhabdus luminescens* and *Yersinia entomophaga* MH96, insecticidal toxin complex (TC) components are known to be present within PAIs (ffrench-Constant et al., 2003; Hurst et al., 2011a; Waterfield et al., 2004). *Photorhabdus luminescens* W14 is known to encode multiple tandem repeats of

insecticidal TC components among several PAIs located throughout the chromosome (Waterfield et al., 2001, 2002, 2004). Within PAI 1 of *P. luminescens* W14 are repetitive *tcdAB* genes interrupted by *tcc*-like sequences as well as a *tcd*A-like pseudogene (Duchaud et al., 2003b; ffrench-Constant et al., 2003). PAI 1 is flanked by tRNA-Asp loci that are associated with repetitive intergenic consensus sequences (ERIC)-like sequences, however PAI 1 is thought to be unstable because the genetic structure of this PAI among *P. luminescens* strains is highly variable (Duchaud et al., 2003b; ffrench-Constant et al., 2003; Waterfield et al., 2004). Within the genome of focal species, MH96 the main virulence determinant, Yen-TC, a insecticidal TC, is encoded on a virulence island called PAI_{YE96} that is flanked by a 76-bp direct repeat sequence (DRS), called DRL_{76} and DRR_{76}, that are associated with tRNA-Gly loci (Hurst et al., 2011a). A remnant tRNA is also located directly 5′ of DRR_{76} but it is not possible to identify historical residue specificity for this locus. PAI_{YE96} is thought to be non-mobile and permanently integrated into the genome of MH96 because the island does not encode any mobilization-related genes (Hurst et al., 2016). Furthermore, insecticidal TC components and a putative chitinase gene were found to be located on a tRNA-associated PAI within a *Xenorhabdus nematophila* cosmid library, that when expressed in *Escherichia coli* was found to increase virulence against cabbage white butterfly, *Pieris brassicae* by *per os* challenge (Morgan et al., 2001).

Interestingly, insecticidal TC genes are even more widely distributed among PAIs and on plasmids within pathogen genomes (e.g., *Serratia entomophila*, *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica*, *Y. mollaretii, Y. frederiksenii, Pseudomonas syringae* pv. tomato, *Fibrobacter succinogenes* and *Treponema denticola*) (Dobrindt et al., 2004, 2015; Dodd et al., 2006; Fuchs et al., 2008) and MH96 (Hurst et al., 2011a) making the evolution of TC islands an interesting system for investigating the influence of HGT in pathogen genome evolution. The widespread occurrence of insecticidal TCs among diverse bacterial lineages may in part be explained by high rates of HGT and frequent recombination of insecticidal TC modules associated with MGEs within bacterial communities over evolutionary timescales. In addition to providing an advantage for insect pathogenesis, TC containing MGEs may have also been responsible for increased rates of genome rearrangements, which is thought to give bacterial populations greater adaptive capabilities (Bellanger et al., 2014).

The genome of *P. luminescens* W14 contains other PAIs that encode toxins, for example PAI 2 is inserted into a tRNA-Phe site and carries the *mcf* gene encoding "makes caterpillar floppy" toxin as well as a filamentous hemagglutinin-like adhesion factor and PAI 3 encodes a CNF-like toxin gene *pnf* but this island (otherwise known as *Photorhabdus* virulence cassette) is not associated with any specific tRNA loci (ffrench-constant et al., 2000; ffrench-Constant et al., 2003; Waterfield et al., 2004). Within the genome of *Y. entomophaga* MH96, an additional fimbrial associated island FAI_{YE96} has been identified as flanked by 138-bp direct repeat sequence and tRNA-Asn genes but has yet to be functionally characterized with respect to virulence (Hurst et al., 2016). The genome of *Y. entomophaga* MH96 also contains five Rearrangement hotspot (Rhs) clusters (Hurst et al., 2016) and in addition to PAI_{YE96}, Rhs-associated region 3 and 4 are also associated with different tRNA genes, and represent putative PAIs. Rhs-associated region 3 is downstream from a tRNA-Arg and encodes a type VI secretion system apparatus, including a potential secreted effector with homology to eukaryotic gene Spt4 (PL_00995) and several short repetitive elements suggestive of orphan toxin/immunity pairs (Poole et al., 2011). Rhs4-associated region is flanked by tRNA-Pro and tRNA-Ser and encodes various putative type 3 and type 4 secretion system effectors.

7.1.3 Mobility and transmissibility of PAIs/GIs among pathogenic bacteria

Generally, PAIs/GIs are assumed be acquired by HGT but reports of active mobilization or transmission of such elements are somewhat limited. One widely-studied example of a mobile PAI is the 36 – 43 kb high pathogenicity island (HPAI) encoding an iron acquistion system called Yersiniabactin, that was initially identified as a mobile element present in the most pathogenic strains of *Yersinia*, including *Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica* phylotype 1B (Bach et al., 1999; Buchrieser et al., 1998; Hare et al., 1999) but was later shown to be highly conserved and widely-distributed among Enterobacteriaceae suggesting that HPAI was recently acquired by HGT among this group (Bach et al., 2000). In *Y. pseudotuberculosis* the excision process for HPAI requires a P4-like integrase (a tyrosine integrase) and a recombination directionality factor (RDF) call *hef*, both of which are encoded on HPAI

(Lesic et al., 2004). Under specific growth conditions (i.e., iron-deprivation at 4 °C) HPAI was demonstrated to be transmitted horizontally between strains of *Y. pseudotuberculosis* at a tRNA-Asn attachment site (Lesic and Carniel, 2005). A related mobile ICE was identified in *E. coli* ECOR31 that also encodes Yersiniabactin and is considered the possible progenitor of HPAI (Schubert et al., 2004). Subsequently, Flannery et al., (2009) identified an 94-kb ICE called ICE*Pm2*, prevelent among pathogens that commonly caused catheter-associated urinary tract infections, including *Proteus mirabilis*, *Providencia stuartii* and *Morganella morganii*. The ICE*Pm2* island consisted of three variable regions, one of which contains the Yersiniabactin, which further established this iron acquisition system on ICEs and mobile PAIs/GIs (Flannery et al., 2009).

Other linkages between ICEs and PAIs and GIs have also been demonstrated previously in the genomes of pathogenic bacteria, including both *Salmonella* (*Sa.*) *enterica* (Nieto et al., 2016; Quiroz et al., 2011) and *Legionella pneumophila* (Lautner et al., 2013; Trigui et al., 2013). A PAI present in several *Sa. enterica* serovars including Enteritidis, Gallinarum, Dublin and Typhi (McClelland et al., 2001; Thomson et al., 2008) known as ROD21 can excise from the genome and form an episomal genetic element when grown in laboratory culture medium and during infection of phagocytic cells (Quiroz et al., 2011) and was also shown to be transferrable between strains of *Salmonella* by conjugation (Salazar-Echegarai et al., 2014). A mutant strain of *Sa. enterica* serovar Enteritiditis that cannot excise ROD21 (deletion mutant of ROD21 cognate integrase) showed reduced ability to cause lethal disease in the mouse infection model as well as reduced colonization of mouse liver and spleen compared to wild-type, suggesting that excision was important for modulating virulence in this system (Tobar et al., 2013).

Other examples of mobile PAIs/GIs from pathogenic bacteria including uropathogenic *E. coli* (Blum et al., 1994; Hochhut et al., 2006; Middendorf et al., 2004), *Shigella* (*Sh.*) *flexoneri* (Sakellaris et al., 2004; Turner et al., 2001, 2004), *Sa. enterica* serovar Typhi (Bueno et al., 2004; Doublet et al., 2005) and *Vibrio cholerae* (Almagro-Moreno et al., 2010; Murphy and Boyd, 2008; Rajanna et al., 2003). In uropoathogenic *E. coli* 536, excision of multiple PAIs within the genome requires P4-like integrase activity, which

are generally restricted to their own cognate islands but some cross-talk between intregrases associated with different islands was also observed by Hochhut et al. (2006). The antibiotic restistance island, SRL in *Sh. flexneri* was also shown to require the SRL-encoded P4-like integrase for excision and intergration into two different tRNA *attB* sites (Turner et al., 2004), while excision of the *Sh. flexneri* virulence island SHE required the SHE-encoded P4-like integrase and RDF *rox* genes (Sakellaris et al., 2004). The very large (134 kb) virulence island SPI7 from *Sa. enterica* serovar Typhi is also a mobile element, which was shown to be prone to excision resulting in loss of capsular antigen Vi, resulting in changes to morphology and phage Vi-II resistance (Bueno et al., 2004; Nair et al., 2004). Excision of the VPI-2 metabolic/secretion island from *V. cholerae* was also shown to occur, but under specific conditions including low temeraptures following sublethal exposure to UV-light and excision of this island was also shown to require the VPI-2-encoded *intV2* (a tyrosine-recombinase) and the RDF gene, *vefA*, but does not require RDF gene *vefB*, which is also found on VPI-2 (Almagro-Moreno et al., 2010). Earlier studies on *V. cholerae* PAIs demonstrated that out of four PAIs located in the genomes of O1 and O139 serogroup isolates, three can excise (VPI-2, VPS-I and VPS-II) even though VSPI-1 does not carry any recombinase-like genes suggesting recombinase/RDF cross-talk between different *V. cholerae* mobile islands (Murphy and Boyd, 2008).

Similar to PAIs/GIs, mobilizable ICEs have also been identified from pathogenic bactieria as well. The transcriptome of *hfq* deletion mutant strain *L. pneumoniphila* revealed excision of a 100 kb copper resistance GI that also encodes an integrase and putative conjugative system. The island is able to excise from the chromosome and exist as episomal plasmid-like element called pLP100 and the rate of excision was effected by presence or absence of the *hfq* (Trigui et al., 2013). Two GIs within the genome in *L. pneumoniphila* Corby, designated Trb-1 and Trb-2, were both found to posses a conjugation/type 4a secretion system encoded by the genes *trb* and *tra*, both can excise and form episomal plasmid-like circles, but only Trb-1 could be transmitted to other *L. pneumoniphila* strains by conjugation (Glöckner et al., 2008). Furthermore, the congate integrases (*Int-1* and *Lpc1883*) were found to be responsible for the exision of Trb-1 and a separate secretion island ICE called LpcGI-2, respectivley (Lautner et al., 2013). In this same study, the activity of Int-1 was found to regulated by several

regualtory proteins (*lvrRABC*) also encoded on Trb-1 as well. Resistance to copper, arsenic and cadmium were also shown to be acquired *in plantae* by uptake of ICEs and plasmids by kiwi fruit pathogen *Ps. syringae* pv. *actinidiae* (Colombi et al., 2017).

With respect to entomopathogenic bacteria (the focus here), there have yet to date been any reports of mobile PAIs/GIs or ICEs. Recently, the methicillin resistance island McRI$_{mecD}$ of *Macrococcus caseolyticus* was shown compatible for site-specific integration in the genome of *B. thuringiensis* (Schwendener and Perreten, 2018). In this study, the integrase, *int*, and the attachment site on a circular element were found to be sufficient for integration at the *rpsl* gene, while two regulator genes *intR* and *xis* were also shown to influence recombination (Schwendener and Perreten, 2018). The importance of plasmids in the pathobiology and evolution of entomopathogenic bacteria has previously been recognized for several species, including *B. thuringiensis* and *S. entomophila* (Dodd et al., 2006; Gillis et al., 2018). It is now estimated that ICEs and PAIs/GIs are far more prevalent among bacterial genomes than plasmids and definetly have an important role in key evolutionary processes, such as pathogen evolution and antibiotic resistance, for example. So, while the importance of MGEs is reconginzed primarily from genome-wide comparison studies, very limited emperical data exist characterizing PAI/GI or ICE mobility mechanisms within entomopathogenic bacteria.

7.1.4 A mobile holin-endolysin secretion system/citrate fermentation genomic island is present in the genome of *Yersinia entomophaga* MH96

A ~17.5 Kb excision was observed in the transcriptome of the MH96 mutant Δ*cspA123*/ΔHCUI$_{YE96}$ strain (Chapter 6, section 6.3.1.4, Table 6.7). Since Δ*cspA123*/ΔHCUI$_{YE96}$ was also shown to have a very conspicuous non-secreting phenotype (Chapter 5, section 5.3.1.1, Table 5.6), further work was undertaken to determine whether this phenotype was attributed to the deletion of the *cspA123* operon or excision of the unstable island herein called holin-endolysin/citrate fermentation unstable island (HCUI$_{YE96}$). Previous linkages between the PAI$_{YE96}$-encoded Yen-TC and the holin-endolysin secretion system and have been identified and described in previous chapters of this thesis, notably as they share a co-expression profile with an exclusive group of secreted chitinolytic and carbohydrate-binding factors (Chapter 4,

section 4.3.8, Table 4.13) and both islands also encode a pair of homologous putative regulators, *yen7* (focus of previous work reported in Chapter 5, section 5.3.2) and, a PhoB-like regulator, respectively. Based on these, and additional work undertaken concurrently with this project by Marion Schoof (Bio-Protection/AgResearch, University of Lincoln) characterizing the role of the PhoB-like regulator and the holin-endolysin system in the secretion of Yen-TC in MH96, the likelihood that the non-secreting phenotype was linked to the excised region was considered very high. Targeted bioinformatics of HCUI$_{YE96}$ provided further insights into potential mobilization mechanisms involving two integrases found encoded within the island, that target DRSs associated with the flanking tRNA-Asn loci recombination.

Validation of the excision by targeted PCR and sequencing confirmed the transcriptome data were correct and that an excision event had occurred in Δ*cspA123*/ΔHCUI$_{YE96}$. Next, further experiments demonstrated that the excision event could be easily repeated during targeted mutagenesis of the *cspA123* operon and during over-expression of *yen7* within MH96 (wild-type). A ΔHCUI$_{YE96}$ mutant was generated allowing for further assessments by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to clarify that the non-secreting phenotype was attributed to the excised region confirming the role of the holin-endolysin system in secretion of Yen-TC and other TC-associated factors. In addition to the non-secreting phenotype, excision of HCUI$_{YE96}$ was also found to result in smaller cell shape and non-motility, eluding to the region having a more significant regulatory role governing global secretion, Yen-TC production, cell shape and motility in MH96.

7.2 Summary methods

Standard molecular microbiology and DNA manipulations were carried out as described in Chapter 3, sections 3.1 and 3.3, respectively. Reduced secretion at 25 °C (Chapter 5, section 5.3.1.1, Figure 5.6) and reduced virulence against *Galleria mellonella* by intrahemocoelic injection at 37 °C (Chapter 5, section 5.3.1.2, Figure 5.8) was previously reported for the Δ*cspA123* strain respectively; however subsequent *in vivo* transcriptome data from this strain identified excision of an unstable region called

HCUI$_{YE96}$ (Chapter 6, section 6.3.1.4, Table 6.7) and as more accurate reflection of the genetic status, this strain was subsequently renamed Δ*cspA12*/ΔHCUI$_{YE96}$. The excision of HCUI$_{YE96}$ in Δ*cspA123*/ΔHCUI$_{YE96}$ was validated by PCR using primers targeting regions internal or external to the HCUI$_{YE96}$ (Table 7.1). Strains used in this study are provided in Table 7.2.

Targeted mutagenesis by homologous recombination to replace the *cspA123* operon with a kanamycin cassette was repeated as previously described (Chapter 5, section 3.4) in order to test whether the excision of HCUI$_{YE96}$ was explicitly associated with this specific deletion. Eight potential recombinants were screened using validation primers (CSP_P1/KanR and KanF/CSP_P7) and HCUI$_{YE96}$ excision primers (EXC_F/EXC_R) (Table 7.1) and a single clean Δ*cspA123* strain was validated by sequencing.

Table 7.1: Primers used to validate the incision/excision of HCUI$_{YE96}$ in *Yersinia entomophaga* MH96 and mutant strains.

Internal or external to HCUI$_{YE96}$	Primer name	Sequence (5' → 3')
Internal	MS24_F	CTCCCCATCCTGTTCTTCAC
	MS23_R	CTACCCTTGGTTATTGTTTG
Internal	MS42_F	CCCTCCCCATCCTGTTCTTC
	MS43_R	AACTACTCCTCCTTAAAATTAG
External	EXC_F	GGTCGGCGTTGAGGGAG
	EXC_R	GCGGATAGCGACGCGAA

Table 7.2: *Yersinia entomophaga* MH96 strains used to determine the phenotypic attributes of HCUI$_{YE96}$.

Strain	Description	Reference
MH96	*Yersinia entomophaga* type strain isolated from diseased grass grub	(Hurst et al., 2011b).
Δ*cspA123*	Deletion mutant derivative MH96 lacking *cspA123* (tandem cold-shock proteins A1, A2 and A3; PL78_18365, PL78_17450 and PL78_18370, including intergenic regions); Kan^R	This study.
ΔHCUI$_{YE96}$	MH96 with holin-endolysin/citrate fermentation unstable island (HCUI$_{YE96}$) excised at the tRNA-Asn.	This study.
Δ*cspA123* / ΔHCUI$_{YE96}$ (previously designated Δ*cspA123*).	Deletion mutant derivative MH96 lacking *cspA123* (tandem cold-shock proteins A1, A2 and A3; PL78_18365, PL78_17450 and PL78_18370, including intergenic regions) and with holin-endolysin unstable island (HCUI$_{YE96}$) excised at the tRNA-Asn. Kan^R	This study.

Excision of HCUI$_{YE96}$ in wild-type MH96 was facilitated by over-expression of *yen7* (a suspected activator of secretion) to select for a non-secreting phenotype as previous experiments found that over-expression of *yen7* at 25 °C had both negative effects on growth and produced strong selection against the induction vector, pBAD-*yen7*, likely due to excessive lysis from activation of the holin-endolysin secretion system (Chapter 5, section 5.3.3.3). Briefly, MH96 with pBAD-*yen7* was grown in Luria Broth Base (LB broth) (Invitrogen) with ampicillin 400 μg/ml for eight hours at 30 °C with 250 rpm shaking, then diluted to 1 % in 50 ml media plus carbenicillin 400 μg/ml and grown for two hours at 25 °C with 200 rpm. Expression of *yen7* was induced by adding 0.002 % arabinose and the cultures were grown another 15 hours. The culture was diluted in phosphate buffer solution (PBS) (Sigma) and plated for single colonies on LB agar Miller (LB agar) (Merck) supplemented with ampicillin 400 μg/ml to select for maintenance of the expression plasmid (i.e., strains with HCUI$_{YE96}$ excised would altogether lack the holin-endolysin secretion system, alleviating the selective pressure against pBAD-*yen7* under inducing conditions). Six colonies were screened using PCR with HCUI$_{YE96}$ external primers, EXC_F and EXC_R (Table 7.1). An isolate testing positive for the excision of HCUI$_{YE96}$ was cured of pBAD-*yen7* by streaking for isolation on LB agar supplemented with 0.02 % arabinose (previous reports demonstrated that even the empty pBAD vector is selected against under higher concentrations of arabinose in MH96), grown for 24 h and transfer of single colonies onto a fresh LB plate

and screened for sensitivity against ampicillin (i.e., loss of plasmid). A cured ΔHCUI$_{YE96}$ strain was validated by carrying out plasmid isolation on overnight culture and then confirming absence of plasmids in the sample by visualization on 1 % agarose gel electrophoresis.

Sequence alignments were carried out using ClustalW and visualized using Geneious (v.R10). Translated nucleotide sequence from HCUI$_{YE96}$ (CP010029.1) and a similar suspected PAI from *Y. ruckeri* Big Creek 74 (CP011078.1) were compared using tBLASTx (Camacho et al., 2009) with online BLAST+ (v.2.9.0) and visualized using EasyFig (v.2.2.3) (Sullivan et al., 2011). Normalized $\log_2$ counts-per-million (CPM) gene expression values for *int1* and *int2* were determined from *in vivo* and *in vitro* RNA-seq as previously described in (Chapter 4, section 4.3.6).

Further phenotypic investigations were undertaken on MH96, Δ*cspA123*, ΔHCUI$_{YE96}$ and Δ*cspA123*/ΔHCUI$_{YE96}$. For visualization of protein extracts, wild-type and mutant strains were grown in 3 ml liquid media for approximately hours with appropriate antibiotics. Flasks containing 50 ml media were seeded with 1 % inoculum and grown at either 25 or 37 °C with 200 rpm shaking overnight (~ 17 hours) with appropriate antibiotic. In the morning, 1 ml culture were pelleted by centrifugation for 10 minutes at 8,000 x g and the cell supernatant was filter sterilized (0. 22 μm). Serial dilutions of the overnight culture were made in PBS and colony forming units (CFUs) were enumerated to confirm consistent cell density among the overnight cultures. SDS-PAGE was performed as previously described in Chapter 3, section 3.5) according to (Laemmli, 1970), using hand-cast 10 % tris-glycine under denaturing conditions, 200 V for 50 min and stained with silver (Blum, H., Beier, H., Gross, 1987). A sample of each strain was visualized by wet mount (1:2 dilution) under a Olympus BX50 light microscope at 400x magnification to observe cell shape, relative size and motility.

7.3 Results

7.3.1 Validation of $HCUI_{YE96}$ mobile genetic island in Δ*cspA123*/Δ$HCUI_{YE96}$

A excisable genomic region of 17.5 Kb, designated holin-endolysin/citrate fermentation unstable island ($HCUI_{YE96}$), was identified in the transcriptome of Δ*cspA123*/Δ$HCUI_{YE96}$ (Chapter 6, section 6.3.1.4, Table 6.7). A direct repeat sequences (DRSs) of 138 bp that contain tRNA-Asn loci (PL78_17410 and PL78_17315) was found to flank $HCUI_{YE96}$ (Figure 7.1). The 5′ DRS is referred to as left attachment site (*attL*) and the 3′ DRS is the right attachment site (*attR*). Two integrase genes, *int1* (PL78_17412) and *int2* (PL78_17414) are also found within $HCUI_{YE96}$ and are likely important for mobilization of the island and site-specific recombination at the DRS. The predicted proteins sizes for Int1 and Int2 are notably small, comprising 30 and 56 amino acid residues, respectively. In addition to the integrase-associated mobilization module, $HCUI_{YE96}$ contains two other modules, a holin-endolysin secretion system/PhoB-like regulator (PL78_17390 – PL78_17400/ PL78_17385) and citrate fermentation cluster, including *citC*, *citD*, *citE*, *citF*, *citX*, *citG* (PL78_17375 – PL78_17345), a putative anion permease (PL78_17375) and phosphopantetheinyl transferase (PPT) (PL78_17380).

Several putative regulators are found within the element, including a winged-helix-turn-helix (wHTH) containing PhoB-like regulator (PL78_17385), that shares significant sequence homology with *yen7*, which is encoded directly upstream of the Yen-TC genes on PAI_{YE96} and are a major focus of this thesis. Specifically, it was demonstrated in Chapter 5 (section 5.3.3.3), that over-expression of *yen7* partially restores protein secretion by MH96 at 37 °C (a temperature where MH96 does not normally secrete Yen-TC and other TC-associated factors). The AraC-like regulator is also found within the holin-endolysin secretion system module and was found to share the same co-expression pattern with Yen-TC and all other TC-association factors based on RNA-seq analysis (Chapter 4, section 4.3.8, Table 4.13)

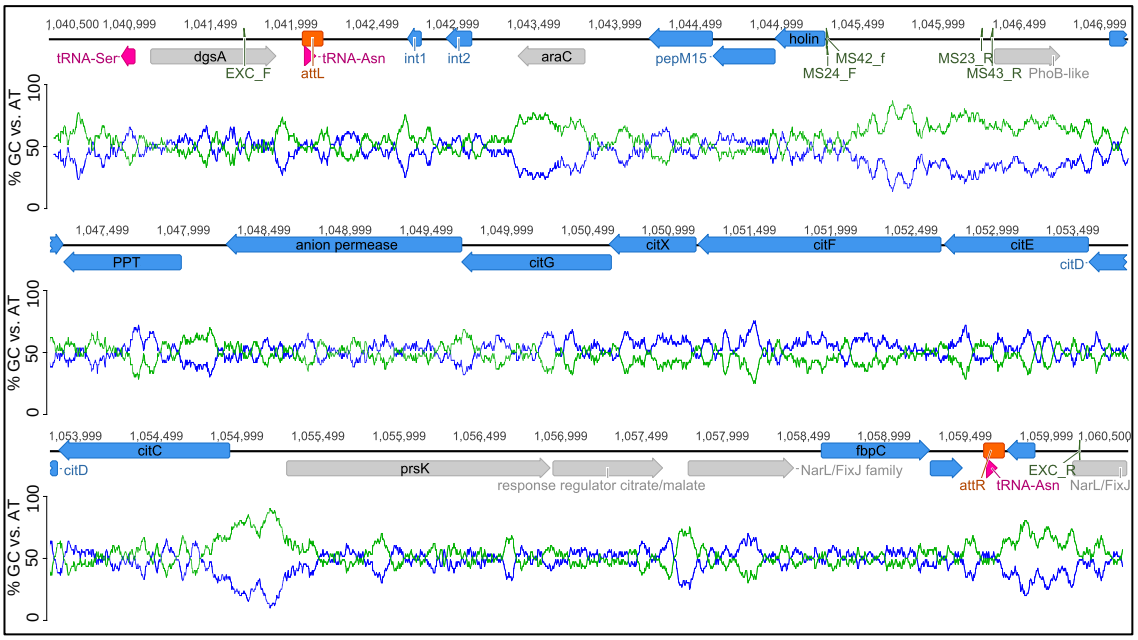


Figure 7.1: Genomic organization and GC content of HCUI$_{YE96}$ in *Yersinia entomophaga* MH96. Blue arrows represent ORFs encoding proteins thought to be non-regulatory, while grey arrows represent known and putative regulatory proteins. Pink arrows represent tRNA genes. Green markers represent primer binding locations used in this study. Orange boxes represent direct repeat sequences (DRS), also known as left and right attachment sites (i.e., *attL* and *attR*). PPT = phosphopantetheinyl transferase. The GC content was calculated using a window size of 60 bp, the green and blue lines represent AT- and GC-content, respectively.

The genome of MH96 contains one additional tRNA-Asn loci (PL78_17260), which is located directly downstream from HCUI$_{YE96}$, flanking the fimbrial associated island called FAI$_{YE96}$ (Hurst et al., 2016). This usher/chaperone fimbrial cluster represents a different putative adhesion and should not be confused with the usher/chaperone fimbrial cluster (PL78_12480, PL78_12475, PL78_12470 and PL78_12465) identified in the *in vivo* RNA-seq as a unique host-induced cluster (Chapter 4, section 4.3.7.2, Figure 4.15, supplemental Figure S10) and targeted for mutagenesis and phenotyping (section 5.3.1). The 138 bp sequence of the HCUI$_{YE96}$ *attL* and *attR* were compared to the third potential tRNA-Asn-containing DRS, which revealed the presence of several single-nucleotide polymorphisms (SNPs) within the 5′ and 3′ intergenic regions (Figure 7.2).

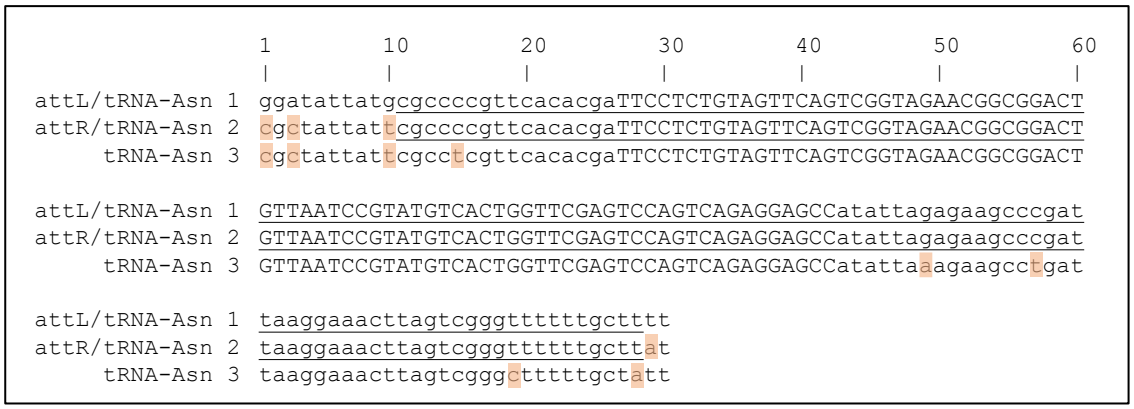


Figure 7.2: Nucleotide alignment of *Yersinia entomophaga* MH96 tRNA loci and flanking regions, containing the HCUI$_{YE96}$ direct repeat sequences (DRS). Uppercase nucleotides = tRNA coding sequences and underlined nucleotides = DRS *attL* and *attR*. Bases highlighted in red represent single nucleotide polymorphisms within intergenic regions.

Excision of HCUI$_{YE96}$ from the genome of Δ*cspA123*/ΔHCUI$_{YE96}$ was validated by targeted PCR, using HCUI$_{YE96}$ internal (MS24_F/MS23_R and MS42F/MS43_R) and external (EXC_F/EXC_R) primers (Figure 7.3). Genomic DNA template generated from Δ*cspA123*/ΔHCUI$_{YE96}$ failed to amplify a product with HCUI$_{YE96}$ internal primers, but did generate a 964 bp PCR product using external primers with one minute elongation, proving excision of HCUI$_{YE96}$ from the chromosome of this strain. The PCR product generated from the external primers was sequenced, allowing for delineation of the DRS/site of recombination (Figure 7.4).

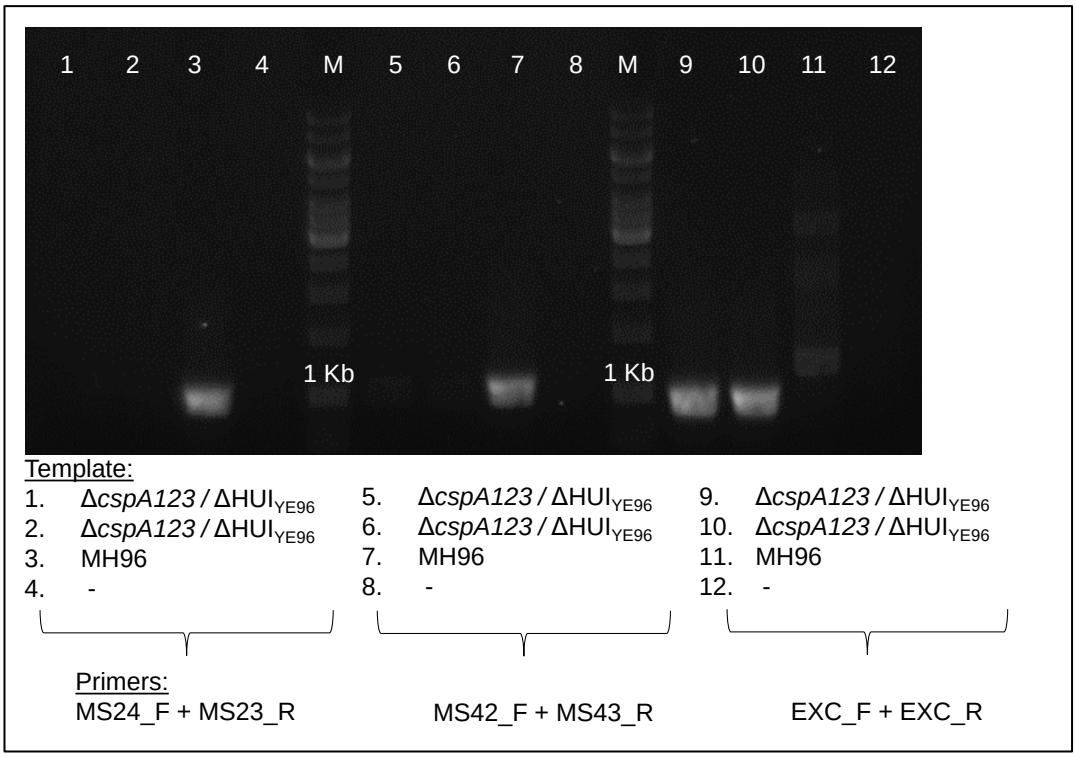


Figure 7.3: Targeted PCR validation of excision of HCUI$_{YE96}$ from *Yersinia entomophaga* MH96 Δ*cspA123*/ΔHCUI$_{YE96}$ mutant strain using internal (MS24_F + MS23_R and MS42_F + MS43_r) and external (EXC_F + EXC_R) primers. M = Marker; GeneRuler 1kb DNA ladder (Thermo Scientific).

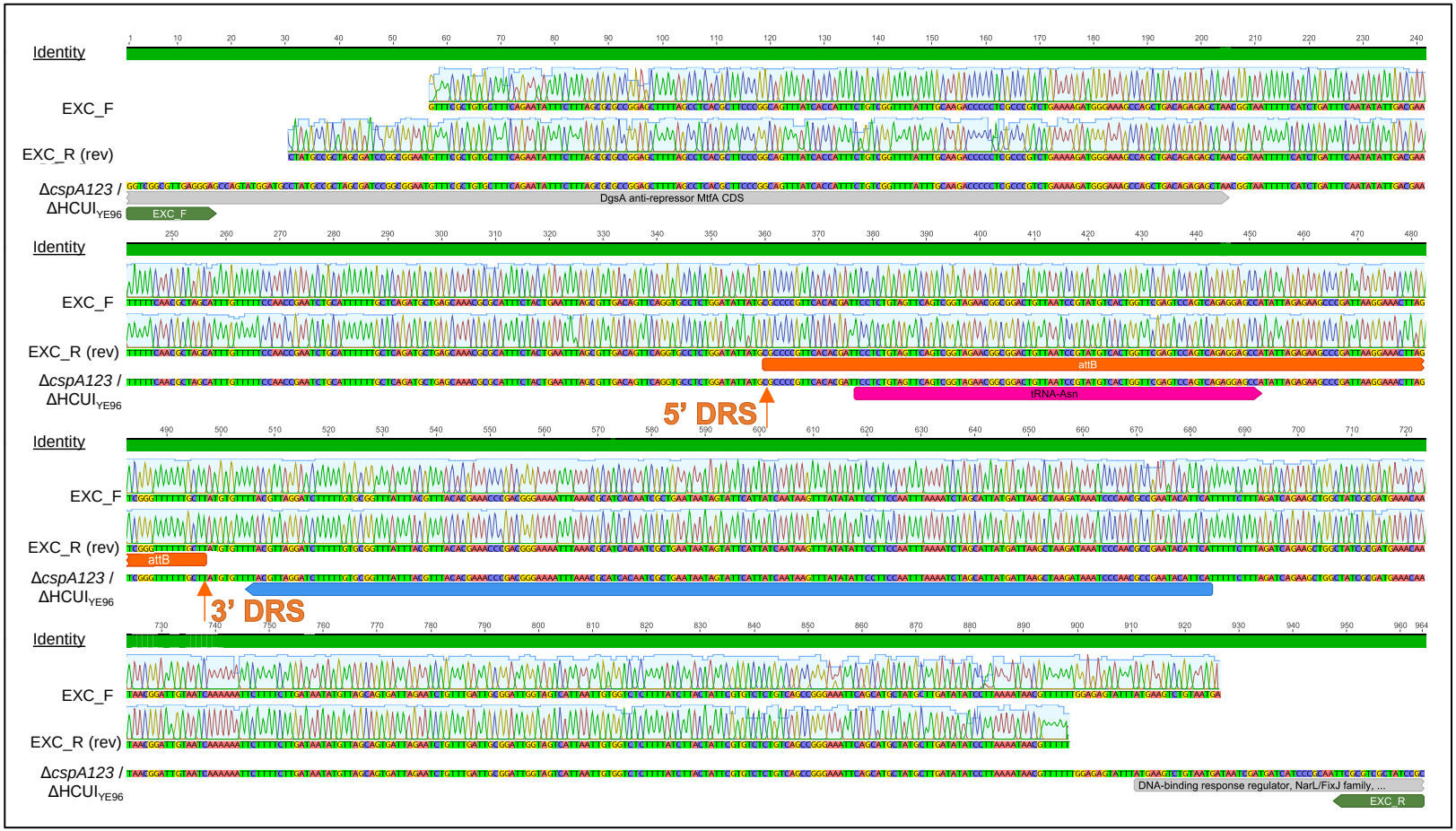


Figure 7.4: Alignment of quality trimmed DNA sequence electropherograms generated from HCUI$_{YE96}$ external primers EXC_F and EXC_R with Δ*cspA123*/ΔHCUI$_{YE96}$ genomic template delineating the point of recombination. Blue arrows represent ORFs encoding proteins thought to be non-regulatory, while grey arrows represent putative regulatory proteins. Pink arrows represent tRNA-Asn. Green arrows represent primer binding locations. Orange boxes represent direct repeat sequences, and the orange arrows indicate the point of excision.

Based on the sequencing results and identification of the DRS, the genomic organization of ΔHCUI$_{YE96}$, including the single remaining *attB* site, was determined (Figure 7.5A) and the extrachromosomal element that must have been excised from the chromosome containing the single *attP* site was predicted (Figure 7.5B).

During repeated mutagenesis of the the *cspA123* operon, in addition to several clean *ΔcspA123* deficient mutants, one out of eight potential recombinants that were checked was found to carry an excision of HCUI$_{YE96}$; however, further validation determined that this mutant was a product of single recombination (i.e., the suicide plasmid integrated into the chromosome) and was not used further in this study. Also, a single wild-type strain carrying an excision of HCUI$_{YE96}$ was identified from six islolates following over-expression of *yen7.* Perhaps coincidental, but worth noting that in each instance the strain carrying an excision was the first colony selected from the recovery plates.

HCUI$_{YE96}$ was found to share 11 high scoring pairs and 73% coverage with the translated nucleotide sequence from a suspected PAI identified from pathogenic strains of *Y. ruckeri* that encodes citrate fermentation, hexose phosphate acquistion and iron transport modules (Cascales et al., 2017) based on tBLASTx searches (E-value = 0.0) (Figure 7.6). Comparison of the two islands revealed significant areas of homology, including the entire citrate fermentation cluster plus the first gene in the hexose-phophate cluster (a regulatory factor) and the putative *fbpC*, a ferric transporter adenosine tri-phosphate (ATP)-binding subunit. Also, both HCUI$_{YE96}$ and the citrate/hexose phosphate/iron island from pathogenic *Y. ruckeri* share the same insertion site associated with tRNA-Asn loci; immediately 5′ of the *attL* site is a putative regulator, *dgsA* followed by a tRNA-Ser and the *attR* site contains a tRNA-Asn in both GIs. There are some differences between the two GIs, such as the presence of a second tRNA-Asn loci in the *Y. ruckeri* island that is not found in HCUI$_{YE96}$ and the presence of two putative integrases, an AraC-like and PhoB-like regulator and the holin-endolysin secretion system within HCUI$_{YE96}$ that are absent in the *Y. ruckeri* island. Genes found within regions external to the islands at the *attR* site were different in MH96 and pathogenic *Y. ruckeri* strains which encode an usher/chaperone fimbriae

and type 6 secretions system, respectively.

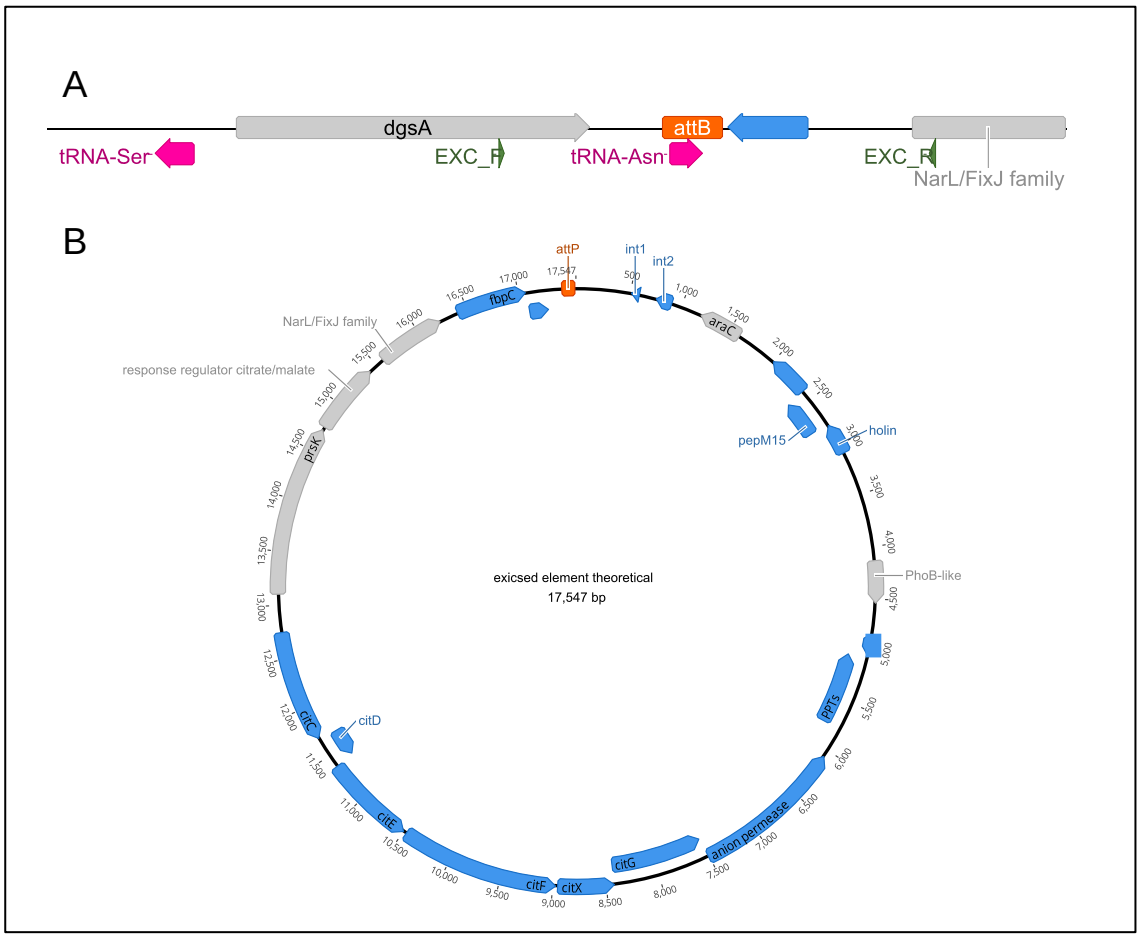


Figure 7.5: Genomic organization of HCUI$_{YE96}$ after excision from genome of *Yersinia entomophaga* MH96. A. Post-excision genomic organization of Δ*cspA123*/ΔHCUI$_{YE96}$ containing *attB* site. B. Theoretical HCUI$_{YE96}$ extrachromosomal element containing *attP* site. Blue arrows represent ORFs encoding proteins thought to be non-regulatory, while grey arrows represent known and putative regulatory proteins. Pink arrows represent tRNA genes. Green markers represent primer binding locations used in this study. Orange boxes represent direct repeat sequences (DRS), also known as left and right attachment sites (i.e., *attB* and *attP*). PPT = phosphopantetheinyl transferase.

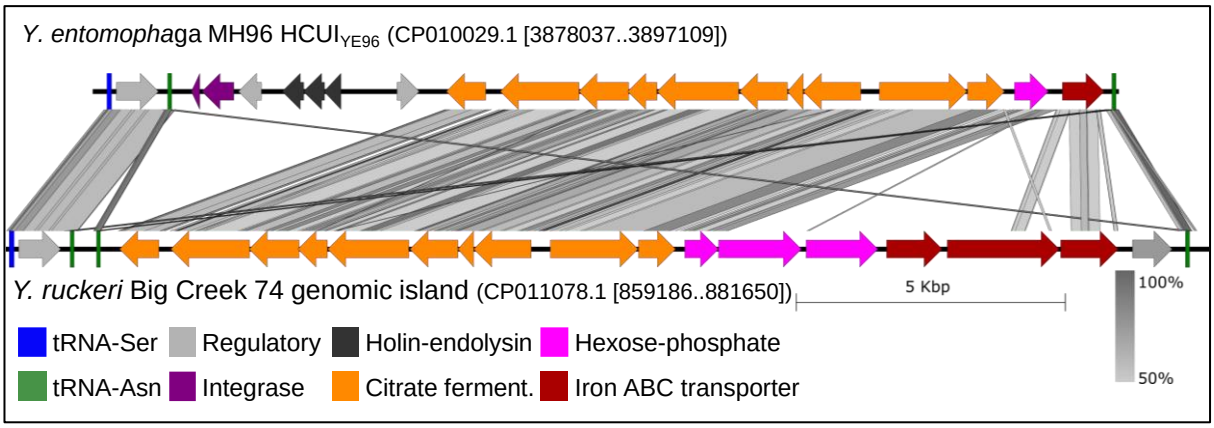


Figure 7.6: The mobile genetic island HCUI$_{YE96}$ in *Yersinia entomophaga* MH96 shares 11 high scoring pairs and 73 % coverage with the translated nucleotide sequence from a GI identified in *Y. ruckeri* Big Creek 74. Grey lines indicate regions with 50 – 100 % identity between GIs based on tBLASTx searches with bacterial genetic code (translated nucleotide subjects using a translated nucleotide query, all three frames translated for DNA strand).

Based on RNA-seq data, *int1* was found be more highly expressed *in vivo* across the three different cell densities tested (~ 10^7, 10^8 and 10^9 CFU/g or CFU/ml) (Figure 7.7A), while *int2* only had higher expression *in vivo* at the lowest cell density (Figure 7.8B). Two putative non-coding RNA (ncRNA_128 and ncRNA_129) were also identified from within the mobilization module of HCUI$_{YE96}$ and ncRNA_129 was found to be more highly expressed *in vivo* compared to *in vitro* across all cell densities tested (Figure 7.9).

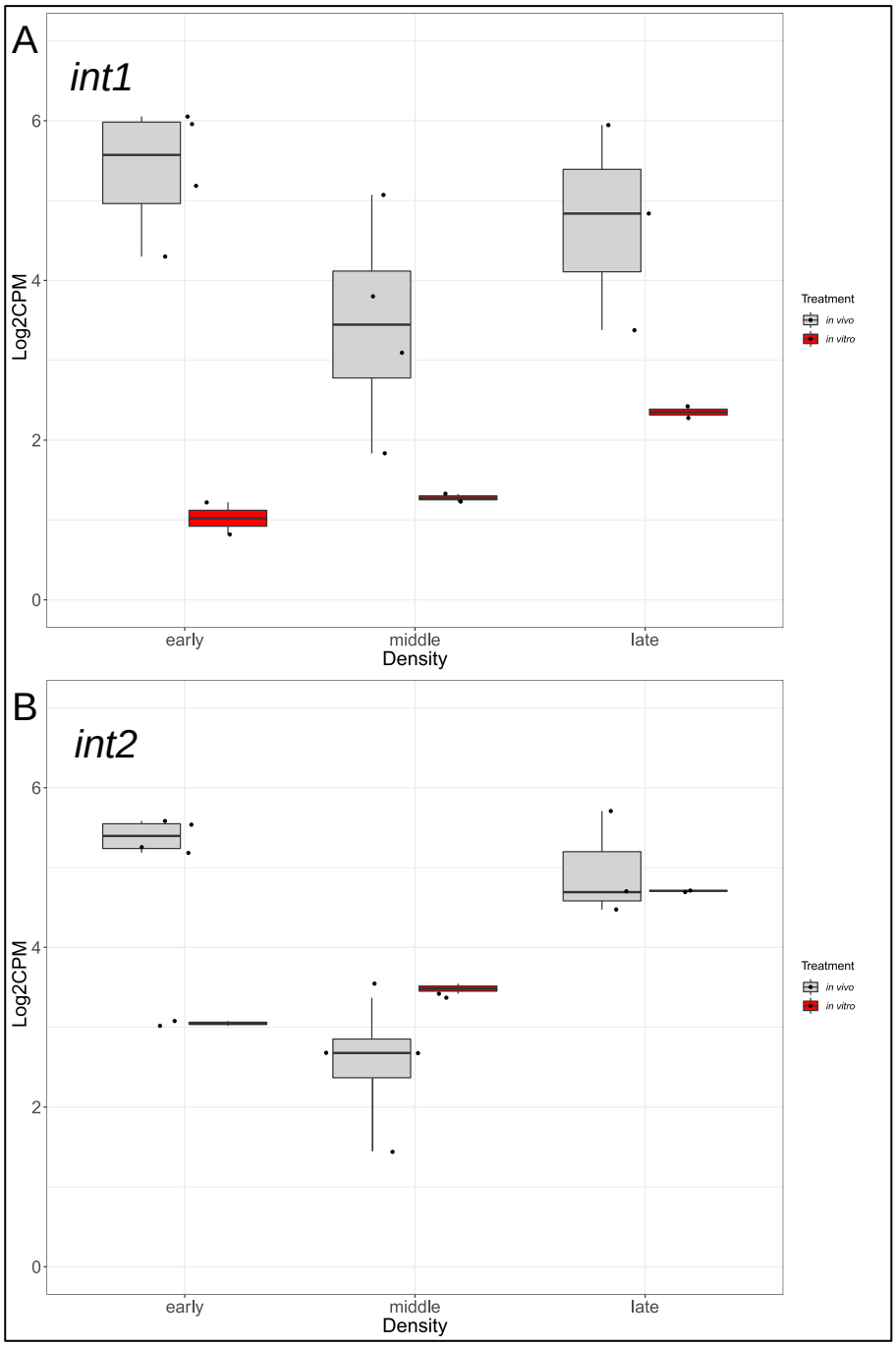


Figure 7.7: Normalized $\log_2$ counts-per-million gene expression of HCUI$_{YE96}$-associated integrases, *int1* and *int2* in *Yersinia entomophaga* MH96, from *in vivo* and *in vitro* treatments at 25 °C.RNA was collected from three different cell densities: early (~10^7), middle (~10^8) and late (~10^9) CFU/g or CFU/ml. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively.

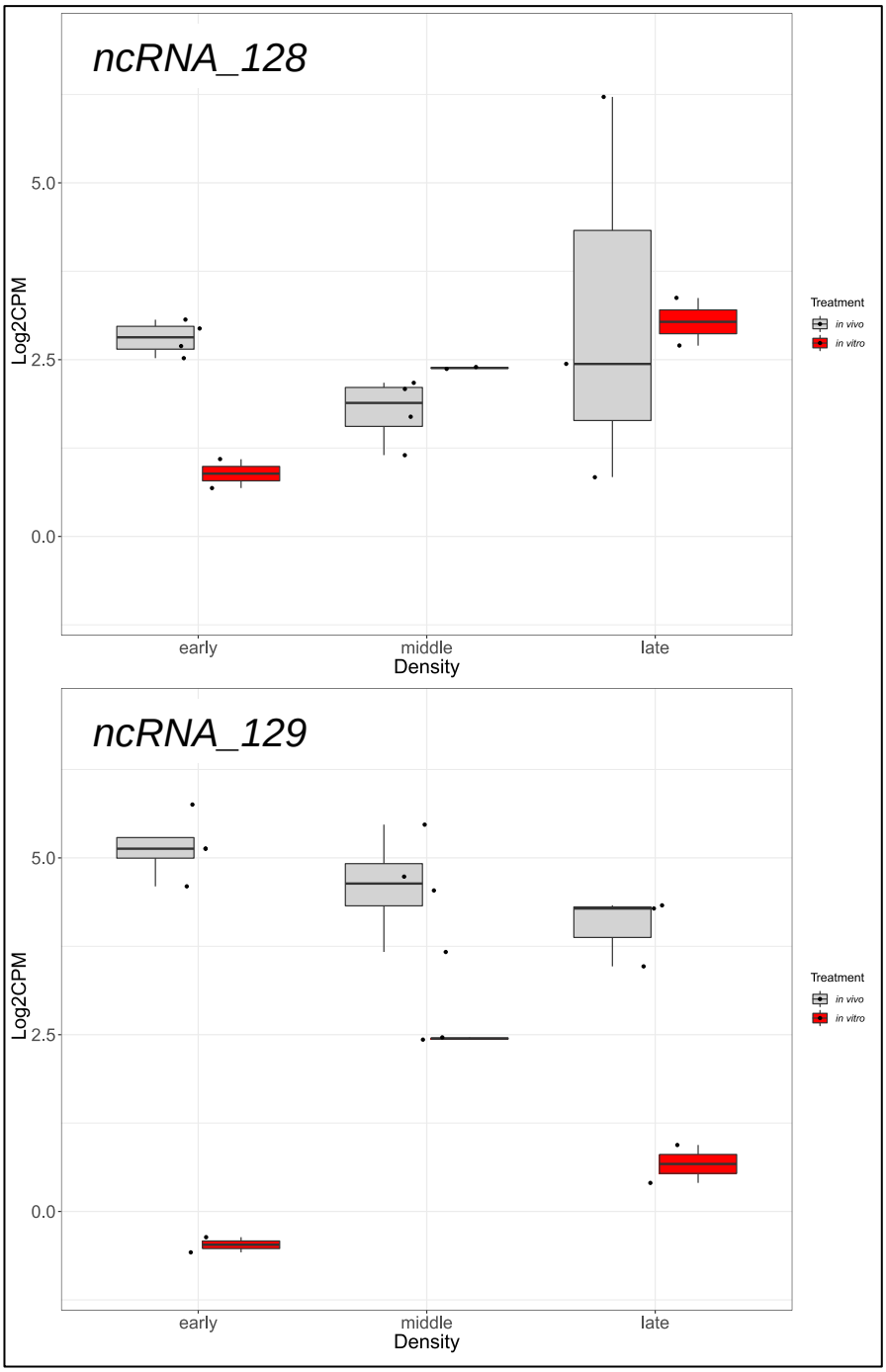


Figure 7.8: Normalized $\log_2$ counts-per-million gene expression of HCUI$_{YE96}$-associated non-coding RNAs (*ncRNA_128* and *ncRNA_129*) in *Yersinia entomophaga* MH96, from *in vivo* and *in vitro* treatments at 25 °C.RNA was collected from three different cell densities: early (~10^7), middle (~10^8) and late (~10^9) CFU/g or CFU/ml. Median boxplots displayed, with upper and lower hinges spanning the interquartile range (difference between the 25th to 75th percentiles) and whiskers representing the maximum and minimum values within 1.5 times the interquartile range over the 75th or under the 25th percentile, respectively.

7.3.2 Non-secreting phenotype was attributed to excision of $HCUI_{YE96}$, not deletion of the *cspA123* operon.

Examination of secreted proteins produced by MH96, Δ*cspA123*, $\Delta HCUI_{YE96}$ and Δ*cspA123*/$\Delta HCUI_{YE96}$ by SDS-PAGE determined that both strains lacking $HCUI_{YE96}$ (i.e., $\Delta HCUI_{YE96}$ and ΔcspA123/$\Delta HCUI_{YE96}$) did not secrete (Figure 7.9) or produce a wide range of proteins including the Yen-TC components (Figure 7.10) at 25 °C. Since the Δ*cspA123* strain still secreted and produced Yen-TC in the same manner as wild-type, the non-secreting phenotype was attributed to excision of $HCUI_{YE96}$. Visualization of the cell supernatant produced by Δ*cspA123* did reveal a large protein with molecular weight much greater than 250 kDa (top band of marker) these bands were not identified in the cell supernatant produced by any of the other strains. The largest gene in the MH96 genome (PL78_16910) is predicted to produce an repeats-in-toxin (RTX) autotransporter adhesin of an estimated 466.78 kDa in size and represents a potential identity for the large molecular weight protein secreted by Δ*cspA123.*

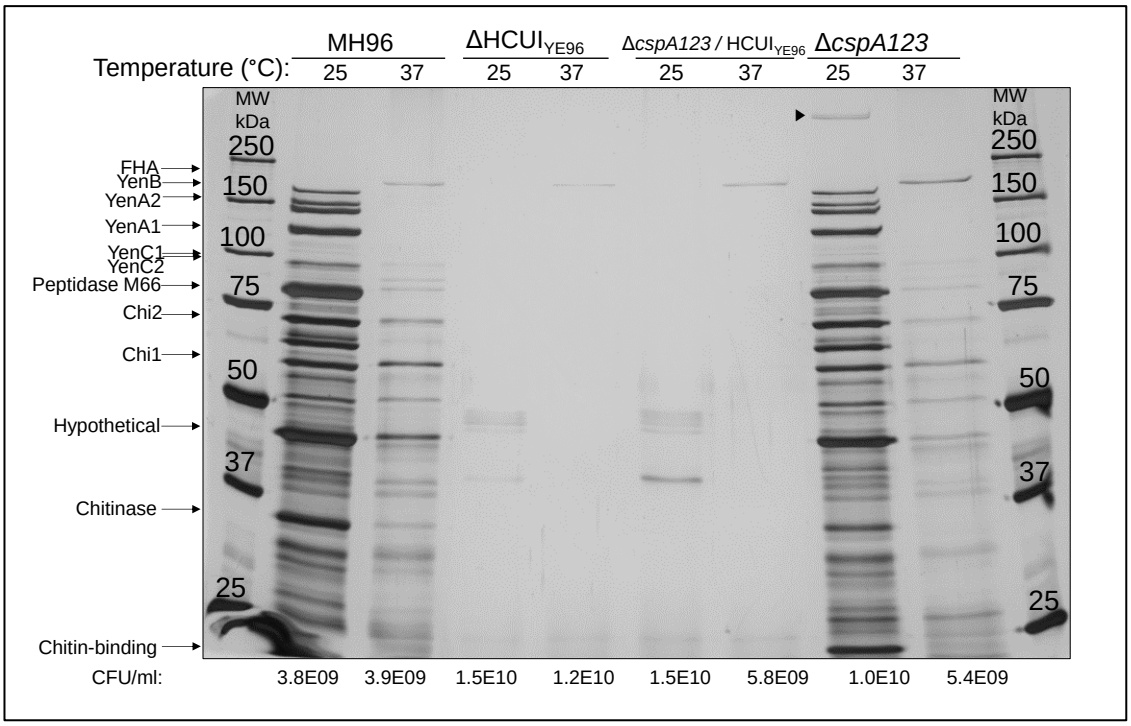


Figure 7.9: Silver-stained SDS-polyacrylamide gel of proteins secreted by *Yersinia entomophaga* MH96, ΔHCUI$_{YE96}$, Δ*cspA123*/HCUI$_{YE96}$ and Δ*cspA123* at 25 or 37 °C in LB broth for 17 hours. MW = molecular weight over marker lane containing Bio-Rad Precision Plus Protein unstained protein standard, with respective ladder sizes given. Arrows correspond to filamentous hemagglutinin (FHA) (~175 kDa), Yen-TC components: YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa), peptidase M66 (81 kDa), hypothetical virulence factor (44 kDa) and chitinase (33 kDa). Cell density is reported as CFU/ml as determined by enumeration plates.

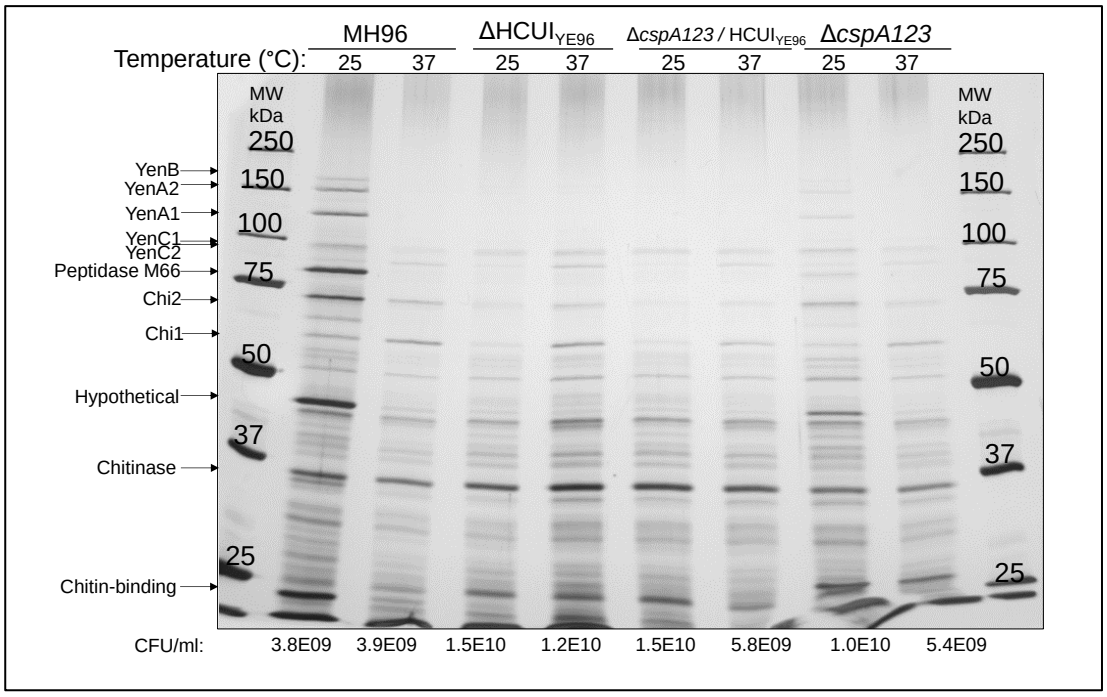


Figure 7.10: Silver-stained SDS-polyacrylamide gel of proteins from cell pellet of *Yersinia entomophaga* MH96, ΔHCUI$_{YE96}$, Δ*cspA123*/HCUI$_{YE96}$ and Δ*cspA123* at 25 or 37 °C in LB broth for 17 hours. MW = molecular weight over marker lane containing Bio-Rad Precision Plus Protein unstained protein standard, with respective ladder sizes given. Arrows correspond to filamentous hemagglutinin (FHA) (~175 kDa), Yen-TC components: YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa), peptidase M66 (81 kDa), hypothetical virulence factor (44 kDa) and chitinase (33 kDa). Cell density is reported as CFU/ml as determined by enumeration plates.

Visualisation of live cells under phase-contrast microscopy identified that a small/round cell shape is associated with the excision of HCUI$_{YE96}$ because both ΔHCUI$_{YE96}$ and Δ*cspA123*/ΔHCUI$_{YE96}$ strains produced cells with a much smaller and rounder appearance compared to MH96 and Δ*cspA123* that produced many more rod shaped cells of variable lengths (Figure 7.11). Under wet-mount, cells produced by both strains with HCUI$_{YE96}$ excised appeared non-motile compared to MH96 and Δ*cspA123* strains as well. Based on these results the excision of HCUI$_{YE96}$ could be attributed to a very conspicuous phenotype characterized by small, round, non-motile, non-secreting cells that do not produce as much Yen-TC and other TC-associated factors at 25 °C, compared to wild-type.

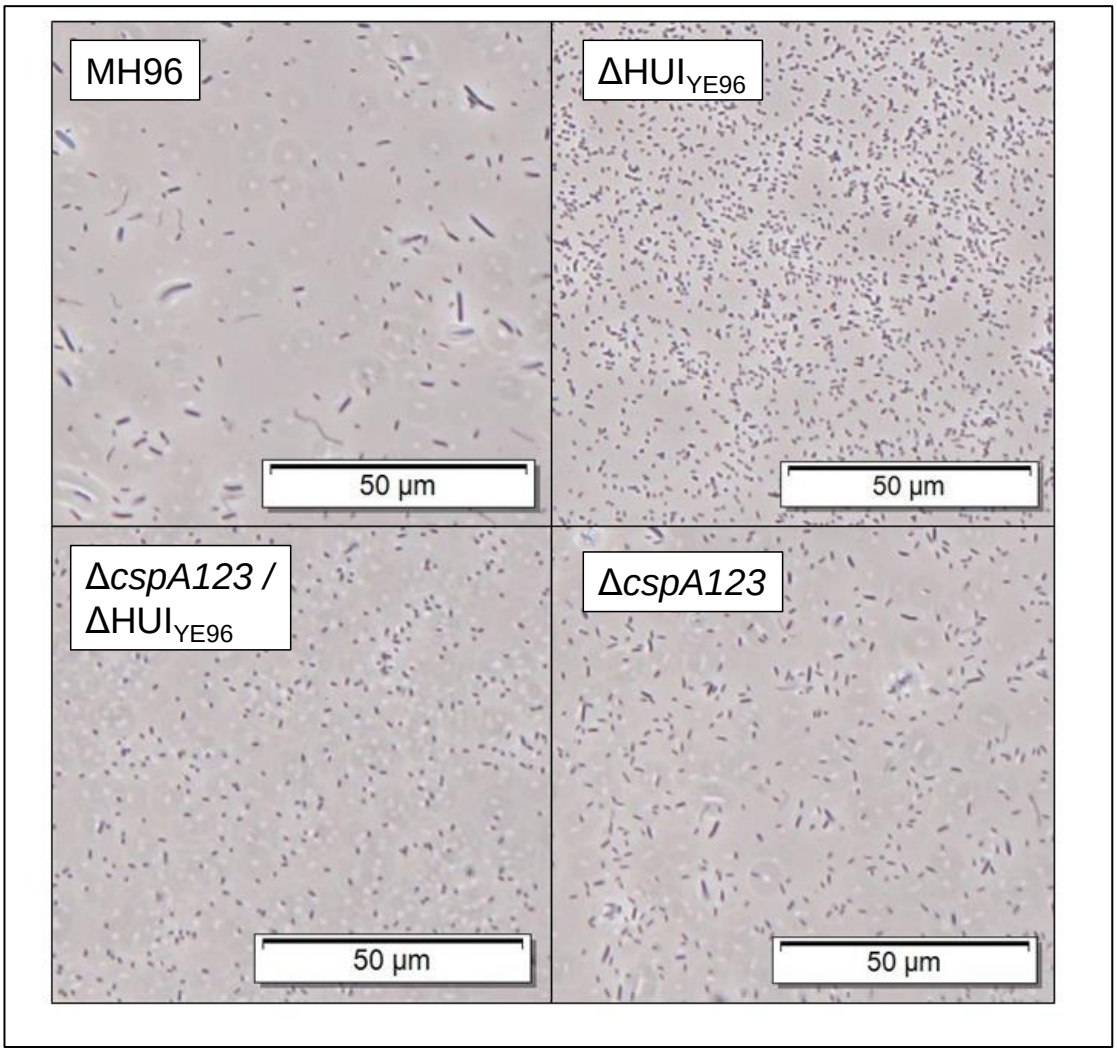


Figure 7.11: Phase-contrast microscopy of *Yersinia entomophaga* MH96, ΔHCUI$_{YE96}$, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*cspA123* strains grown in LB broth for 16 hours at 25 °C.

7.4 Discussion

7.4.1 HCUI$_{YE96}$ is an excisable island that inserts into direct repeat sequences located at tRNA-Asn attachment sites

This chapter provides preliminary characterization of an excisable secretion/metabolism island in MH96, that when excised from the genome produces a very distinctive phenotype, with altered cell-shape, motility, toxin production and secretion. This is the first report of a holin-endolysin secretion system located on an actively mobile island, which is likely a variant of a previously reported island containing citrate fermentation, hexose phosphate uptake and iron transport modules, found in the genomes of several *Yersinia* spp., including three virulent strains of *Y.*

ruckeri reported by Cascales et al. (2017). HCUI$_{YE96}$ and the related citrate/hexose-phosphate/iron islands are all inserted at the same location in the genomic backbone, which is associated with a tRNA-Asn locus. GIs are commonly associated with tRNAs as well as tmRNAS, and together these represent two of the more common targets for site-specific recombination by integrases encoded by GIs, bacteriophages and ICEs (Boyd, 2012; Boyd et al., 2009; Busby et al., 2013a; Lewis and Hatfull, 2001; Panis et al., 2012).

More specifically, tRNA-Asn attachment sites have previously been shown to be important in *Yersinia* because the iron-acquisition island, HPAI, also targets tRNA-Asn (Buchrieser et al., 1998; Hare et al., 1999). In *Y. pseudotuberculosis*, HPAI was shown to be transmittable and could integrate into any of three tRNA-Asn sites in the genome (Buchrieser et al., 1998) and more broadly among various Enterobacteriaceae (Bach et al., 2000; Karch et al., 1999; Schubert et al., 1998). The attachment site identified for HPAI was found to share sequence homology to the target sequence for a bacteriophage P4 integrase that normally targets *leuX* locus in *E. coli* (Buchrieser et al., 1998; Pierson III and Kahn, 1987). The 138 bp long DRS associated with excision of HCUI$_{YE96}$ is quite different than the 17 bp long DRS associated with HPAI. The tRNA-Asn seems to be a common target among *E. coli* too because HPAI was shown to be able to integrate in the tRNA-Asn of different strains of *E. coli* and tRNA-Asn is also a target for a colibactin synthesis island called *pks* also found among *E. coli* (Nougayrède et al., 2006). The observation that tRNA-Asn represents a common target for island-related integrases found among *Yersinia* spp. and *E. coli* strains, can be further supported by phylogenetic analysis of island-related integrases that showed that tRNA preference type was somewhat taxa specific (Boyd et al., 2009).

7.4.2 PAIs and GIs are associated with tRNA sites among other entomopathogenic bacteria

Considering what is known from the genomes of entomopathogens, insertion of GIs/PAIs at tRNA loci has been described before. A tRNA-Asn fimbrial island, FAI$_{YE96}$, in MH96 (Hurst et al. 2016) is located directly adjacent to HCUI$_{YE96}$, but this island is not known to be mobile, despite sharing a very similar attachment site with HCUI$_{YE96}$.

Since FAI$_{YE96}$ and HCUI$_{YE96}$ share an attachment site, these islands would be an example of tandem accretion, which has been known to be a driver of genome diversity in bacteria, such as *Streptococcus* (*Sc.*) *thermophilus* and *Sc. agalactiae* (Bellanger et al., 2011; Pavlovic et al., 2004; Puymège et al., 2013). In *Y. ruckeri*, at totally different island with a type VI secretion system (T6SS) occupies the flanking tRNA-Asn insertion site next to the citrate/hexose phosphate/iron island, which really demonstrates how tandem accretion can result in genome diversity among closely related pathogenic bacteria.

More generally, tRNA loci are also known to be important attachment sites for insecticidal TC encoding PAIs. The Yen-TC containing PAI$_{YE96}$ of MH96 is inserted into tRNA-Gly locus and in *P. luminescens* W14 the *tcd* containing PAI I is inserted at a tRNA-Asp locus. In contrast, another TC island, tc-PAI ^{Ye}, found among various strains of *Yersinia* spp. (i.e., *Y. enterocolitica*, *Y. pseudotuberculosis*, *Y. pestis* and *Y. mollaretii*) is located in a consistent location within the genomic backbone (Bresolin et al., 2006; Fuchs et al., 2008; Tennant et al., 2005), but there was no evidence of tRNAs, integrases or IS elements associated with these islands (Bresolin et al., 2006), with the exception of a single transposon-related sequence within tc-PAI^{Ye} only in the genome of *Y. pestis* Angola (Fuchs et al., 2008). Among the tc-PAI^{Ye} encoding *Yersinia* strains, several additional *tccC* duplications were identified externally from the primary TC island; however, protein sequence alignments of island- and non-island encoded TccCs formed separate clusters based on phylogenetic analysis (Fuchs et al., 2008). Since the tc-PAI^{Ye} lacks a mobilization module or attachment sites, and frameshift mutations have arisen within the genes for TC components in some strains (Fuchs et al., 2008), the island represents a relatively older acquisition since it has lost the ability to mobilize and is in the process of inactivated gene decay in some lineages (Kaper and Hacker, 2000). Like tc-PAI^{Ye}, PAI$_{YE96}$ has most likely lost its mobilization module too, but still maintains tRNA-Gly attachment sites suggesting PAI$_{YE96}$ was a more recent acquisition among a non-tc-PAI^{Ye}-encoding ancestral *Yersinia* lineage that would eventually give rise to MH96.

7.4.3 Comparison of HCUI$_{YE96}$ and other mobile islands described among pathogenic *Yersinia*

HCUI$_{YE96}$ is like HPAI in that both islands carry integrase genes, are inserted at tRNA-Asn and maintain the ability to excise from the genome. Mobilization of HPAI was shown to require P4-like tyrosine integrase, *int*, that targets a 17 bp-long DRS present at each end of the island for excision/incision (Buchrieser et al., 1998; Lesic et al., 2004; Schubert et al., 1999). Within the Enterobacteriaceae, the mobilization module, including *int* and the tRNA-Asn loci, of HPAI were found to be quite conserved among various strains of *E. coli*, *Citrobacter diversus* and *Klebsiella* spp., except for *K. oxytoca* and one strain of *E. coli* (4-85), which had an entire truncation of the *int*/tRNA-Asn region and a 300 bp deletion within the tRNA-Asn region, respectively (Bach et al., 2000). The mobilization module of HCUI$_{YE96}$ is maintained and (at least) transcriptionally active (i.e., *int1* and *int2*), while in contrast the *Y. ruckeri* citrate/hexose-phosphate/iron island is altogether missing any type of recognizable mobilization module, as well as the holin-endolysin module (including AraC- and PhoB-like regulators), but still maintains matching recombination sites *attL* and *attR*. Much of the 5′ end of HCUI$_{YE96}$ has decayed compared to the island in *Y. ruckeri*, lacking almost all of the original hexose phosphate uptake and iron transporter related genes, leaving only the original citrate fermentation cluster intact.

Based on this analysis the *Y. ruckeri* citrate/hexose phosphate/iron island may represent the ancestral version of HCUI$_{YE96}$, which has undergone subsequent modifications since becoming acquired into the genome of ancestral MH96. It is tempting to speculate that the holin-endolysin secretion system and associated PhoB-like regulator were acquired from PAI$_{YE96}$ during a subsequent recombination or genome rearrangement event during the evolutionary history of MH96, especially since insecticidal TC genes and holin-endolysin secretion system are co-located on tc-PAI^{Ye} found among different species of *Yersinia* (Fuchs et al., 2008).

7.4.4 Mobilization module of HCUI$_{YE96}$ consists of phage-related integrase genes, *int1 and int2*

Since HCUI$_{YE96}$ was found to excise in MH96, *int1* and *int2* are considered to be the mobilization module for the island. Further classification of *int1* and *int2* into a specific integrase family is hindered by their small sequences, so small that these genes are currently considered pseudogenized in the current GenBank reference genome. Despite this, transcriptional signal was clearly identified from *int1* and *int2* from RNA-seq analysis, suggesting the genes are transcriptionally active (especially *in vivo*), and likely contribute to excision of the island. Presently, it is unclear whether *int1* and *int2* belong to any of the most widely recognized PAI/GI-associated recombinases (i.e., tyrosine integrases, serine integrases or DDE transposases) or perhaps represent novel types of integrases. Thus, further classification of these very small, mobile island-encoded integrases and characterization of their role in HCUI$_{YE96}$ excision represent a priority for future research in MH96. Also, two non-coding RNA were found to be transcribed from the mobilization module and at least one showed increased expression *in vivo,* which suggests a potential role for ncRNAs in virulence and/or mobilization as well.

While *int1* and *int2* likely represent the mobilization module of HCU$_{YE96}$, there may be other *trans* factors involved in excision process. For example, PAIs encoding tyrosine integrases have been shown to require additional (often island-encoded) RDFs, which is the case for HPI, which requires *hef* (Lesic et al., 2004). Like island-encoded integrases, RDFs are also considered to originate from integrated prophages (Panis et al., 2012). Unlike, HPAI, HCUI$_{YE96}$ does not encoded any known RDFs; however, lack of an island-associated RDF does not necessarily rule out the possibility that *int1* and *int2* can excise the island independently. In *Clostridium difficile* for example, a transposon-encoded serine integrase was shown to effectively integrate and excise the island without the involvement of any helper RDFs (Lyras et al., 2004). Another possibility is that *int1* and *int2* may be assisted by the putative phage transcriptional regulator AlpA *in trans* (PL78_16240), which is a similar scenario shown found for PAIs found in both *St. aureus* (Lindsay et al., 1998; Ruzin et al., 2001) and *V. cholerae* (O'Shea and Boyd,

2002). As a next step, targeted deletion of *int1*/*int2* and *alpA* in MH96 would allow further characterization of the role these genes play in excision of HCUI$_{YE96}$.

In *Y. pseudotuberculosis*, the excision rates of HPAI varied with strain but a circular episomal form could be detected using PCR with primers targeting the *attP* site of the circular form (Lesic et al., 2004). The episomal element was found to be highly transient and rare, with less than one copy per cell, requiring two rounds of nested PCR to achieve a detectable PCR produce (Lesic et al., 2004). An important next step in characterizing the mobility of HCUI$_{YE96}$ in MH96 is to determine the natural rate of excision, and whether HCUI$_{YE96}$ can form a circular episomal form when excised using a similar targeted PCR approach as described in Lesic et al. (2004). Based on the phenotype conveyed by the island, it would not be unexpected to find excision of the island is highly regulated in MH96, especially when it was shown that excision occurred during targeted mutagenesis of *cspA123* or upon induction of *yen7* from the pBAD arabinose induction vector.

7.4.5 HCUI$_{YE96}$ is a secretion/metabolism island with mobile properties

Type 3, 4 and 6 secretion systems (T3SS, T4SS and T6SS) are widely known to be associated with MGEs, including PAIs/GIs among diverse bacteria. While not the focus here, the presence of a T3SS carried on the pYV plasmid (i.e., Yop virulon) is the hallmark of human-pathogenic *Yersinia* spp. and has been very widely studied (Cornelis, 2002; Rohde et al., 1999), but other T3SS are known from PAIs within the genomes of other significant human pathogens including, diarrheagenic *E. coli*, *Salmonella* spp., *Yersinia* spp., *Shigella* spp., and *Listeria* spp. as well as the plant pathogens *Ps. syringae* and *Erwinia* spp. (Kaper and Hacker, 2000). Furthermore, some T4SS found on GIs are known to delivery DNA and effectors into host cells among a variety of pathogenic and symbiotic bacteria, including *Agrobacterium tumefaciens*, *Sinorhizobium meliloti*, *Bradyrhizobium japonicum*, *Mesorhizobium loti*, *Photorhabdus* spp. and *Wolinella succinogenes* (Dobrindt et al., 2004). T6SS are also known to be associated with PAIs, for example *Photorhabdus* spp. has four predicted T6SS clusters, which are thought be located on pathogenicity islands (Rodou et al., 2010). Many T6SS-encoding island are known to contain highly repetitive elements as well, which frequently

consist of multiple orphan/toxin pairs arising from frequent recombination events (Alcoforado Diniz and Coulthurst, 2015; Jackson et al., 2009).

Holin-endolysin secretion systems are phage-derived lysis cassettes that have recently been recognized as essential for the secretion of insecticidal TC and chitinolytic machinery in *Y. enterocolitica* W22703 (Springer et al., 2018a) and *S. marcescens* (Hamilton et al., 2014; Owen et al., 2018), respectively. Here, it was shown that the ΔHCUI$_{YE96}$ mutant does not secrete at all, which directly implicates the holin-endolysin secretion system in the secretion of Yen-TC and possibly other chitinolytic and carbohydrate binding factors known to be co-secreted with Yen-TC. In light of these findings, HCUI$_{YE96}$ possibly represents a new type of excisable secretion island, one that carries a holin-endolysin secretion system that is associated with insecticidal TC secretion by MH96. Given that many GIs are thought to have arisen from phage origin, perhaps holin-endolysin secretion systems represent a very accessible module for incorporation into GIs due to the high likelihood of pre-existing phage-born regulators or elements that may be able to interact with the lysis cassette in an advantageous manner once acquired within an island.

7.4.6 A case for holin-endolysin secretion system acquisition from PAI$_{YE96}$ by HCUI$_{YE96}$ resulting in duplication of *yen7* and PhoB-like regulator

Direct linkages between the endolysin-holin secretion system and Yen-TC have been highlighted from transcriptome results reported in this thesis identifying co-regulation of Yen-TC and TC-associated factors (Chapter 4, section 4.3.8, Table 4.13). On a genetic level, both PAI$_{YE96}$ and HCUI$_{YE96}$ share paralogous putative regulators, containing wHTH DNA-binding motifs, *yen7* and PhoB-like, respectively. Evidence that *yen7* is an activator of downstream *chi1* (first gene in Yen-TC operon) was previously shown (Chapter 5, section 5.3.3.3, Figure 5.21 and section 5.3.3.6.5, Figure 5.30); however, a Δ*yen7* mutant still produced and secreted as much Yen-TC as wild-type when cultured at 25 °C (Chapter 5, 5.3.1.1, Figure 5.5) indicating, that like *yen7*, PhoB-like regulator may play a functionally redundant role as activator of Yen-TC transcription and secretion. The finding that ΔHCUI$_{YE96}$ does not produce nor secrete Yen-TC at 25 °C (even with an intact *yen7* gene) supports a slightly different hypothesis, where PhoB-

like and/or additional HCUI$_{YE96}$-encoded factors are the master transcriptional activator for both Yen-TC transcription and the holin-endolysin secretion system. In order to definitively show if PhoB-like regulator is the master activator of Yen-TC/secretion a PhoB-like regulator deficient mutant could be assessed for toxin production and secretion by SDS-PAGE.

Additional linkages between HCUI$_{YE96}$ and PAI$_{YE96}$ were identified from the expression profiles of the HCUI$_{YE96}$-associated holin-endolysin secretion system and PhoB- and AraC-like regulators, PAI$_{YE96}$-associated Yen-TC components and *yen7* and several other co-secreted chitinolytic and carbohydrate binding factors (Chapter 4, section 4.3.8, Table 4.13). One of the key findings of this transcriptome study was the consistent expression profiles of these so-called 'TC-associated' genes, which formed an explicit group of known and putative VFs all having significant down-regulation occurring at 37 °C under *in vivo* growth conditions. Clearly the many linkages between the two islands in MH96, including transcriptomic and phenotypic evidence, supports a recent acquisition of the holin-endolysin and important associated regulators by HCUI$_{YE96}$ from PAI$_{YE96}$ during a recombination event, also resulting in duplicated wHTH regulatory genes become split on either island. With the currently available information it is not possible to identify the source of the HCUI$_{YE96}$ mobilization module or whether the wHTH regulator gene duplication was a result of the recombination event. Regardless, genomic evidence strongly suggests the current placement of the holin-endolysin secretion system on HCUI$_{YE96}$ arose from an interaction between an insecticidal TC bearing PAI and a citrate fermentation GI during the evolutionary history of MH96.

7.4.7 Similar citrate fermentation clusters are associated with other genomic islands present in *Yersinia* and *Vibrio* species

In addition to the holin-endolysin secretion system, the other functional module found on HCUI$_{YE96}$ encodes a citrate fermentation cluster, which was also found to share high conservation with a putative GI found among *Y. ruckeri*, but a similar GI was identified as absent from avirulent strain ATCC29473, but present in three other virulent strains (150, Big Creek 74, CSF007-82 and SC09) (Cascales et al., 2017). This region contains the

entire citrate fermentation cluster and shares a few other homologous regions with HCUI$_{YE96}$ providing evidence of lateral acquisition of a closely related island in *Y. ruckeri* and MH96, and that the citrate fermentation genes may play a key role in pathogenicity in both a fish and insect pathogen. It is interesting that the *Y. ruckeri* island is lacking a mobilization module, indicating a more permanent integration of this ancestral variant into the chromosome of *Y. ruckeri* compared to MH96. Since the hexose phosphate acquisition and iron transport clusters remain intact, these modules must be important to the survival of *Y. ruckeri* compared to MH96, which has since lost these modules.

A citrate acquisition cluster has also been documented among variable metabolic islands identified from a genome-wide analysis of *V. parahaemolyticus* strains (Regmi and Boyd, 2019). The distribution of this metabolic island, was found to be limited to 39 different strains of *V. parahaemolyticus* (> 5 %), but a high proportion of isolates with the metabolic island (~ 87 %) were pathogenic and encoded a T3SS (Regmi and Boyd, 2019). Since this island was found among all 20 isolates from a 2009 outbreak in Peru (Gonzalez-Escalona et al., 2016), a causal linkage between citrate fermentation and emergence of the epidemic strain of the ST 120 clonal group was proposed (Regmi and Boyd, 2019). Other variants of this island were identified, including one associated with a CRISPR-Cas system in strain CDC_K563 and strain 3259 contained an even more expanded version also carrying genes for L-rhamnose and oligogalacturonide metabolism (Regmi and Boyd, 2019).

The variable citrate fermentation cluster found in *V. parahaemolyticus* includes more genes than the fermentation cluster found in MH96 on HCUI$_{YE96}$ and the related island found in *Y. ruckeri* virulent strains; The citrate fermentation module on HCUI$_{YE96}$ contains a single *cit* operon that encodes gene for the citrate lyase complex (CitD, CitE and CitF) and auxiliary proteins CitG, CitC and CitX. The *V. parahaemolyticus* island contains two diverging *cit* operons; one is like the citrate cluster on HCUI$_{YE96}$, encoding the same citrate lyase complex and auxiliary proteins, while the diverging operon encodes an extracellular citrate-sensing two-component regulatory system (TCRS) CitA/CitB, the oxaloacetate decarboxylase OadA/OadB/OadC and the citrate carrier CitS (Regmi and Boyd, 2019). The citrate cluster on HCUI$_{YE96}$ encodes a different

putative TCRS as well, consisting of a response regulator of citrate/malate metabolism and a histidine kinase that may be important in citrate sensing. With respect to transportation of citrate into the cell, the HCUI$_{YE96}$ lacks a CitS homolog, but does encode a putative anion permease and phosphopantetheinyl transferase (PPT) that may be involved in citrate acquisition. Previously, MH96 was shown to be able to utilize citric acid as a carbon source (Hurst et al., 2011b). Additional work investigating temperature-dependent effects on MH96 sole-carbon source utilization at 25 and 37 °C confirmed MH96 can grow on citrate at both temperatures (Chapter 4, section 4.3.9, data not shown).

7.4.8 Involvement of citrate uptake systems in virulence

Although, citrate uptake systems have been shown to contribute to virulence in plant pathogens, *Pectobacterium atrosepticum* (Urbany and Neuhaus, 2008) and *Xanthomonas campestris* (Tamir-Ariel et al., 2007, 2011) it is not exactly clear why acquisition of citrate and/or citrate metabolism may be important in pathogenesis. One suggestion is that citrate plays an important role in citrate-dependent iron acquisition, for example the plant pathogen *Erwinia chrysanthemi* produces a citrate-dependent siderophore called achromobactin (Münzinger et al., 2000) and the synthesis of achromobactin by closely related *Dickeya dadantii* was shown to be critical for plant infection (Franza et al., 2005).

Recently, intermediates of the tricarboxylic acid (TCA) cycle, including itaconate, succinate and acetyl-CoA have been implicated in macrophage signal transduction related to inflammatory responses (Infantino et al., 2014; Noe and Mitchell, 2019; Tannahill et al., 2013; Williams and O'Neill, 2018). Therefore citrate metabolism has been suggested as a pathogenic strategy for intracellular bacterial pathogen survival, for example, an itaconate degradation gene cluster found on HPAI, called *ripA*, in *Y. pestis* was shown to be essential for survival in macrophages (Pujol et al., 2005; Sasikaran et al., 2014). Similarly, a different itaconate degradation gene cluster was also found to be crucial for the survival of *Ps. aeruginosa* (Sasikaran et al., 2014). While Cascales et al. (2017) identified a citrate fermentation cluster on GI associated with virulent strains of *Y. ruckeri,* an actual pathogenic or metabolic function for this cluster has yet to be defined and remains a critical area for future research into the role of the

same cluster in entomopathogenic MH96. An unrelated citrate uptake operon, *yctCBA* in *Y. ruckeri* was found to be induced *in vivo* by Fernández et al. (2004) but was later found to be involved in citrate metabolism but not to be critical for virulence in the rainbow trout model (Navais et al., 2011).

In order to determine if the citrate fermentation cluster encoded on HCUI$_{YE96}$ in MH96 contributes to pathogenesis targeted deletion of the cluster would be required, followed by phenotyping by bioassay. Previous bioassay of the Δ*cspA123*/ΔHCUI$_{YE96}$ mutant by intrahemocoelic injection of *Galleria mellonella* found a 3-fold reduction of virulence in the mutant compared to wild-type MH96 at 37 °C, but not 25 °C, based on the estimated 5-day median lethal dose (Chapter 5, section 5.3.1.2, Table 5.2); however, the reduced virulence may also be attributed to the *cspA123* deletion and/or the excision of the entire HCUI$_{YE96}$ in this strain. Another important next step in defining the role of the citrate fermentation cluster in MH96 biology is to determine if the cluster contributes to MH96 ability to grow on citrate as a sole carbon source. Simmons citrate agar could be used to assess whether MH96 can grow aerobically (or possibly under microaerobic conditions) on citrate as a sole carbon source and whether the ΔHCUI$_{YE96}$ mutant suffers growth defects due to the loss of the citrate fermentation cluster when grown on the same media. Previous assessment of temperature effects on sole carbon source utilization found that MH96 can grow on citric acid at both 25 and 37 °C (Chapter 4, section 4.3.9, Table 4.14). Over the evolutionary history of HCUI$_{YE96}$, the mobilization, holin-endolysin secretion system and citrate fermentation modules have remained intact, but the hexose-phosphate acquisition and iron transport modules have been essentially lost. This pattern suggests that the citrate fermentation module is important in the biology of MH96, but whether it plays a direct role in virulence remains to be shown.

7.5 Summary

By investigating the transcriptome of mutant deficient for three tandem cold-shock proteins, excision of a holin-endolysin secretion system/citrate fermentation island was uncovered. When the island is excised, MH96 is no longer able to produce or secrete

Yen-TC, appears to have reduced motility, and the cells change from rods of variable length to small round cells. The lack of toxin secretion is attributed to the holin-endolysin secretion system, which has been identified as the secretion system for insecticidal toxin and chitinolytic machinery secretion in other pathogenic bacteria. The excisable region also shares a citrate fermentation cluster with pathogenic strains of the causative agent of red mouth disease in fish, *Y. ruckeri*, but whether the citrate cluster contributes to in-host metabolic adaptation or functions more directly in pathogenesis remains to be elucidated.

Based on the similarities and differences among modules on $HCUI_{YE96}$ and the suspected PAI found among pathogenic *Y. ruckeri,* it is thought that $HCUI_{YE96}$ acquired the holin-endolysin secretion system from PAI_{YE96} and the demonstrated mobility of $HCUI_{YE96}$ from the genome is evidence of recent horizontal transfer of the unstable island to the MH96 genome. Further work is required to characterize the motility of $HCUI_{YE96}$, especially whether it is transmittable and to determine if a coordinated excision processes has evolved that may support the complex and highly-regulated process of toxin secretion by holin-endolysin secretion system in MH96.

8.1 Key findings

8.1.1 MH96 secretes a highly effective insecticidal machinery that is co-regulated in a temperature- and host-dependent manner

The highly virulent nature of MH96 against a broad range of insect hosts has been attributed to effective midgut dissociation during *per os* challenge by Yen-TC (Hurst et al., 2011a; Marshall et al., 2012), an insecticidal toxin complex (TC) that is secreted by MH96 at temperatures of 25 °C or lower along with four other exoenzymes, two chitinolytic factors – a chitinase and a lytic chitin monooxygenase/chitin binding protein (LCMO/CBP) as well as mucinase-like M66 peptidase and putative intimin-domain containing protein. In this research, it was shown that these co-secreted toxins and exoenzymes are also co-regulated at the transcriptional level in response to host-dependent signals and thermally regulated at the post-transcriptional level. Along with Yen-TC and co-secreted exoenzymes, *in vivo* transcriptome also implicated several other putative virulence factors (VFs) as consortium of co-regulated insecticidal machinery, including holin-endolysin secretion system, type 2 secretion system (T2SS), a novel hydrolase/lyase cluster, a gene for heat shock protein, IbpA, and a number of putative transcriptional regulators (Chapter 4, section 4.3.8, Table 4.13).

8.1.2 Extreme pathogenicity of MH96 is due to a multifaceted secrete and destroy mode deployed during intrahemocoelic infection of *G. mellonella* at 25 °C.

In addition to the core repertoire co-regulated insecticidal machinery, *in vivo* transcriptomics also provided an unbiased view of the pathobiology of MH96 during intrahemocoelic infection of *G. mellonella*. These results found that MH96 deploys a multifaceted attack against the host during intrahemocoelic infection, which involves secretion of Yen-TC and co-secreted exoenzymes plus numerous accessory toxins. In addition, *in vivo* transcriptome combined with function virulence enrichment identified VFs that responded to the host environment, suggested MH96 virulence against insects also involves translocation of effectors by type 6 secretion system (T6SS) and type 3

secretion system(T3SS), T3SSYE2 directly into host cells, with concurrent expression of adhesion by usher chaperone fimbriae plus other host-induced transcriptional adaptations involving metabolism, motility, outer membrane, stress response, iron acquisition and horizontal gene transfer (HGT).

8.1.3 Transcriptional responses to *in vivo* conditions by MH96 at 37 °C provided insights pathogenic response to oxidative stress

MH96 was found to respond to *in vivo* conditions at 37 °C by upregulating multiple sulfur uptake gene clusters, cold-shock proteins, non-ribosomal protein synthetase (NRPS), genes for an YhlA-like filamentous hemagglutinin N-terminal containing protein and YhlB-like two-partner secretion (TPS) system as well as iron and nickel uptake genes. The secretion of the YhlA-like protein visualized by sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE) of a large ~175 kDa secreted protein, one of the few proteins secreted by MH96 during *in vitro* conditions at 37 °C and validated by liquid chromatography-mass spectrometry.

8.1.4 Yen6 is a virulence regulator that can activate ribose uptake/utilization genes and repress genes for fructose uptake/utilization and an RNA-binding protein, YhbY

Using transcriptomics and electrophoretic mobility shift assays (EMSA), the PAI_{YE96}-encoded *yen6* was shown to produce a small DNA-binding transcriptional regulator that is involved in virulence of MH96 during intrahemocoelic infection of *G. mellonella* at 37 °C. The findings of this work identify *yen6* as playing an important role in response to oxidative stress by regulating carbon fluctuations that drive nicotinamide dinucleotide phosphate oxidase (NADPH) production by the pentose phosphate pathway (PPP) through a process known as reserve carbon flux.

This work also provided evidence that Yen6 does not seem to directly interact with *yen7* or have significant influence on Yen-TC component genes because these genes did not have significant DE in the Δ*yen6 in vivo* transcriptome at 37 °C compared to wild-type and recombinant Yen6 protein did not show specific binding activities against the promoter region of *yen7* during *in vitro* binding assays. Furthermore, limited evidence

to support transcriptional interaction between Yen6 and *yen7* or Yen-TC component genes was provided from phenotypic assessment of Δ*yen6*, which secreted the same amount of protein as wild-type. The deletion of Δ*yen6* was shown to have an effect of increasing β-gal production by the *chi1::lacZ* reporter strain compared to wild-type but this phenotype was not easily complemented, and the positive effect on cell growth and plasmid stability by over-expression of *yen6* from the pBAD arabinose induction vector also suggest Yen6 plays a different role than regulating *yen7* and/or Yen-TC component expression in MH96.

8.1.5 An unstable region of the genome was identified in MH96 and excision was found to result in conspicuous phenotypic effects on secretion, Yen-TC production, motility and cell shape.

An unexpected, albeit important outcome of this work was the identification of a 17.5 kb unstable region, $HCUI_{YE96}$ that was found to actively excise from the genome at the tRNA-Asn site. Preliminary findings show that the region is easily mobilized through targeted mutagenesis of *cspA123* or induction of *yen7* from arabinose induction vector pBAD (Chapter 7, section 7.3.2). Previous *in vivo* transcriptomics work implicated many genes located on $HCUI_{YE96}$ as part of the Yen-TC/TC-associated factors co-regulated insecticidal machinery (Chapter 4, section 4.3.8, Table 4.13). Additional exploration of $HCUI_{YE96}$ and a similar pathogenicity island identified in *Y. ruckeri* informed a potential evolutionary history of $HCUI_{YE96}$ by acquisition of holin-endolysin cassette from PAI_{YE96} during the genomic evolution of MH96.

8.2 Updated model of intrahemocoelic infection of *G. mellonella* by MH96 at 25 and 37 °C

In light of new evidence presented in this thesis, considering both *in vivo* transcriptome of MH96 and Δ*yen6* and presence of secreted proteins in cell supernatant at 25 and 37 °C under *in vitro* conditions, updated virulence models for MH96 during infection at 25 and 37 °C were conceptualized These models represent the most current working model of MH96 pathogenesis but some aspects remain speculative and further molecular investigations are required to validate the models.

8.2.1 Insect infection model at 25 °C – secrete and destroy

The findings presented in this thesis support MH96 as one of the most entomopathogenic bacteria known to humanity. The minimum lethal dose of MH96 during intrahemocoelic infection of *G. mellonella* were shown to be ~ 1 cell at both 25 and 37 °C and *in vivo* transcriptomics revealed a plethora of diverse VFs are deployed by MH96 throughout infection. Both findings corroborate observations of extreme pathogenic capabilities of MH96 against insects. The *in vivo* transcriptome supports Yen-TC holotoxin, chitinolytic exoenzymes – ChiA and CbpA and auxiliary exoenzymes – a mucinase-like M66 peptidase and hypothetical VF containing intimin and carbohydrate-binding domains, as the co-secreted insecticidal machinery, ensuring MH96 is well-equipped to breach mid-gut epithelia during *per os* challenge. These co-secreted VFs (including Yen-TC components) are easily visualized by SDS-PAGE as dominant protein bands from cell supernatant of MH96 grown under *in vitro* conditions at 25 °C (Hurst et al., 2011a) (Chapter 5, section 5.3.1.1 Figure 5.5 and Figure 5.6) and filter-sterilized culture supernatant was also previously shown to have cytotoxic effects against diamond back moth *Plutella xylostella* and New Zealand grass grub *Costelytra giveni* (Hurst et al., 2011a). Further development of this secreted insecticidal machinery, including Yen-TC and co-secreted exoenzymes, represents a promising area for biopesticide development.

In addition to the co-secreted insecticidal machinery, MH96 was found to encode a number of other genes for putative toxins and effectors that showed significant response during infection of *G. mellonella* at 25 °C compared to *in vitro* conditions (Chapter 4, section 4.3.7.3, Figure 4.17). These findings support a more detailed view of MH96 offensive 'secrete and destroy' mode, which involves coordinated multifaceted attack using secretion of core insecticidal machinery plus other hemocoelically active toxins (i.e., CdtAB, VIP2, YenT, YenU and AidA) as well as other insecticidal secreted VFs (i.e., CbpB and YhlA-like) and translocation of numerous T3SS and T6SS effectors (Figure 8.1). In this model the co-secreted insecticidal machinery works synergistically to degrade the chitin-rich peritrophic membrane and mucin layer, enhancing intoxication efficiencies of Yen-TC holotoxin against the columnar cells of the mid-gut epithelia. The outcome of this concerted effort is degradation of the insect midgut,

providing access to the main body cavity of the insect host, which represents the primary focus of *in vivo* transcriptome investigations reported in this thesis.

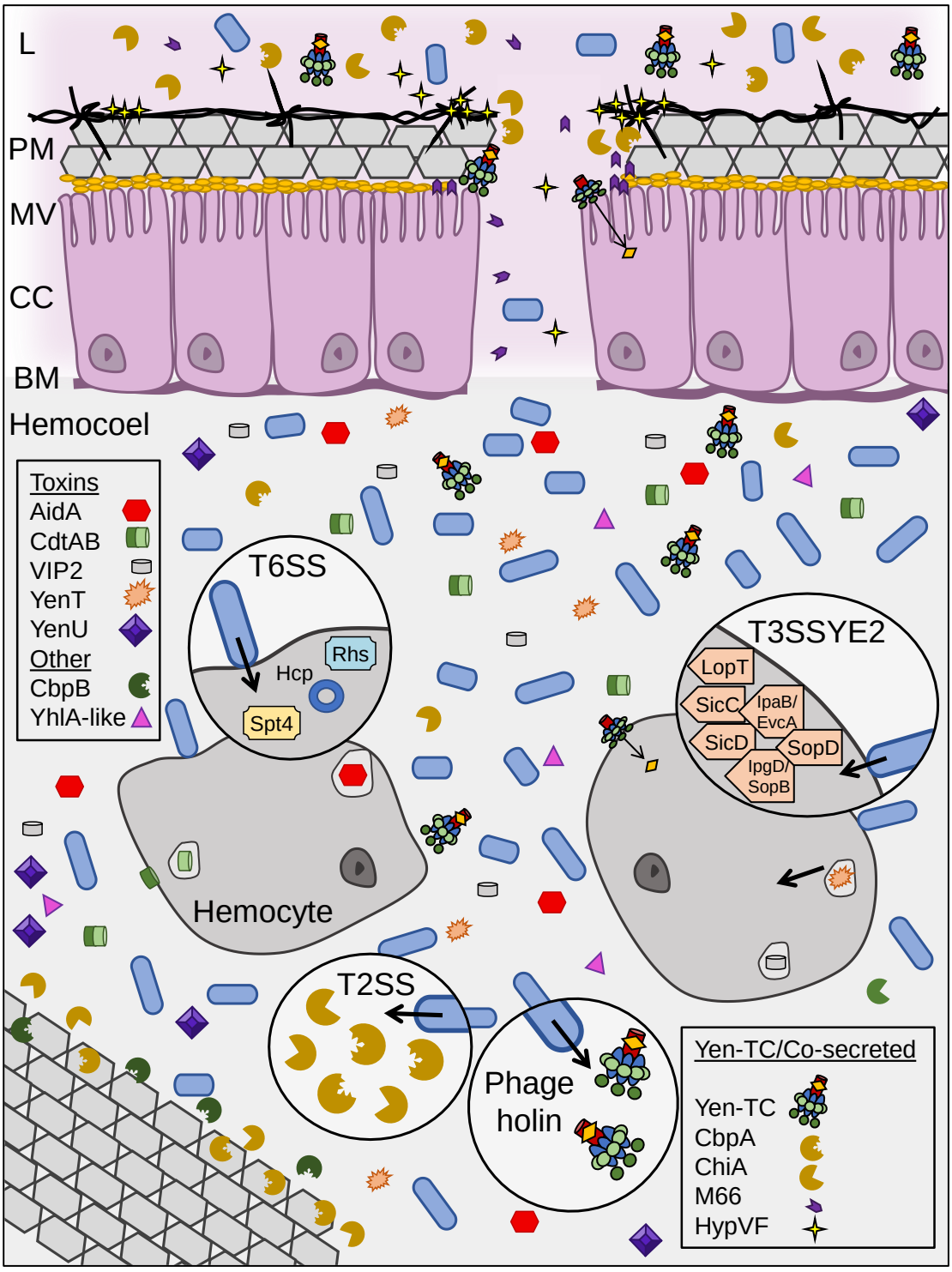


Figure 8.1: Conceptual model of *Yersinia entomophaga* MH96 secrete and destroy mode within insect host.Bacterial cells are blue rods. L = lumen, PM = peritrophic membrane, MV = microvilli, CC = columnar cells, BM = basement membrane. Fibrils associated PM are represented as black barbed-wire, chitin is represented as grey hexagons and mid-gut cell surface mucins are represented as yellow circles. Yen-TC/Co-secreted factors are secreted by MH96 and co-regulated by shared host-dependent and thermoregulatory mechanisms. Genes for toxins were found to have significantly higher expression during intrahemocoelic infection compared to *in vitro* conditions.

Once inside the hemocoel of the insect host, the expression of numerous genes for toxins and secreted effectors, as well as genes for the T3SSYE2 and T6SS are upregulated in MH96 and therefore considered hemocoelically-active. Within the hemocoel, the secrete and destroy mode involves coordinated activities of multiple toxins and effectors in order to effectively neutralize the host immune system allowing for growth, dissemination of bacterial cells throughout the body of the host and bioconversion of insect host tissues by MH96. Numerous other genes encoded by MH96 were shown to significantly respond to *in vivo* conditions and also shared sequence similarity to known VFs from other pathogens. These findings support pathogenesis of MH96 against insect hosts as a coordinated attack, not only involving toxins and translocated effectors, but also motility, host-cell attachment, metabolic adaptation, detoxification against host defense and outer membrane modifications.

Despite these findings, the T6SS, usher chaperone fimbriae and CbpA were not found to have major contributions to MH96 virulence against *G. mellonella* during intrahemocoelic infection as deficient mutants did not have attenuated virulence compared to wild-type; however, the ability to detect phenotypic differences using this approach is severely limited due to the extremely low LD_{50} of MH96 in this model combined with the prevalence of numerous toxins and effectors that likely share (at least) some degree of functionally redundancy.

8.2.2 Infection at 37 °C – redox and kill

While temperature is known to be a major regulatory cue for pathogenic bacteria, and much work has described thermoregulation of various VFs produced by pathogenic *Yersinia* spp., *in vivo* transcriptome of MH96 found that the interaction between temperature and host-dependent environmental conditions was critical for MH96 to undergo major transcriptional shifts. During intrahemocoelic infection at 37 °C, genes involved in production and secretion of the Yen-TC and TC-associated factors, including co-secreted insecticidal machinery became strongly repressed (Chapter 4, section 4.3.8, Table 4.13); however a totally different group of genes related to iron, nickel and sulfur acquisition, NRPS, YenV and filamentous hemagglutinin N-terminal containing hemolysin (YhlA-like) were found to respond to *in vivo* conditions at 37 °C,

and represent putative VFs involved in a different virulence strategy used by MH96 to survive within mammalian hosts (Chapter 4, section 4.3.8, Figure 4.20).

In vivo transcriptomics also identified *yen6* as the most differentially expressed gene in the MH96 genome at 37 compared to 25 °C and based on the proximity of *yen6* to the Yen-TC component genes within the pathogenicity island PAI_{YE96}, further characterization of *yen6* became a major focus of this thesis. Using a strain containing *lacZ* translational reporter fusion to the promoter region of *yen6,* significantly greater β-galactosidase (β-gal) production was determined at 37 °C *in vivo* compared to 25 °C, indicating host-dependent regulation of *yen6* translation have evolved in MH96 (Chapter 5, section 5.3.3.6.7, Figure 5.31). The opposite effect of temperature on β-gal production in the P_{yen6}:*lacZ* reporter strain was observed under *in vitro* conditions (Chapter 5, section 5.3.3.6.2, Figure 5.25) confirming the expression and translation of *yen6* is under complex regulation involving both temperature- and host-dependent factors.

Later assessment of the Δ*yen6 in vivo* transcriptome identified that Yen6 acts as a transcriptional regulator to activate ribose uptake/utilization genes (*rbsD-xylG-rbsC-xylF-rbsK-ccpA*) and repress genes for fructose uptake/utilization (*IIA/Hpr-fruK-IIBC*) and RNA-binding protein, *yhbY* (Chapter 6, section 6.3.1.4, Table 6.5). The specific binding activity of recombinant Yen6 for conserved 10 bp palindromic repeat motif located among the promoter regions of *rbsD*, *IIA/Hpr* and *yhbY* using EMSA, confirming Yen6 is a DNA-binding protein (Chapter 6, section 6.3.2.3 Figure 6.8, Figure 6.9 and Figure 6.10). Despite Yen6 involvement with sugar uptake and central carbon metabolism, limited effects on utilization of sole-carbon sources at 37 °C could be attributed to deletion of *yen6* compared to wild-type (Chapter 6,section 6.4.5, Table 6.10), implying that the regulatory activities of Yen6 may contribute to increased NADPH production by shifting carbon flux into PPP.

Considering at 37 °C under *in vivo* conditions, genes for multiple sulfur acquisitions (uptake sulfate, thiosulfate, methionine and taurine) are highly expressed compared to 25 °C and that one of the most highly expressed regulators at 37 °C, *yen6*, has likely evolved to increase intracellular reducing power by regulating core metabolic pathways during *in vivo* conditions, these results can be taken together as evidence that

MH96 employs a completely different strategy during intrahemocoelic infection of *G. mellonella* at 37 °C (Figure 8.2) compared to the secrete and destroy approach known from 25 °C. Under these conditions, the entire secreted insecticidal machinery, including Yen-TC and co-secreted exoenzymes, as well as many other putative VFs (i.e., T6SS, flagella, etc.) are repressed and it is expected MH96 will not be able to launch an effective multifaceted attack to prevent immune response by the host as observed for intrahemocoelic infections at 25 °C. Since the host immune system remains intact, MH96 would be exposed to far greater amounts of oxidative stress under these conditions and in response increases uptake of sulfur and production of NADPH in order to produce antioxidants including reactive thiols, such as glutathione, which would help MH96 to survive under such conditions. In line with this hypothesis, MH96 was found to have a minimum lethal dose of a single cell at both 25 and 37 °C (Chapter 5, section 5.3.1.2, Table 5.2), confirming MH96's alternative virulence strategy (including Yen6-induced increased redox potential) is equally as effective as the Yen-TC/co-secreted insecticidal machinery/T3SSYE2/T6SS/multiple toxin attack strategy, for killing *G. mellonella* by intrahemocoelic injection at 37 °C.

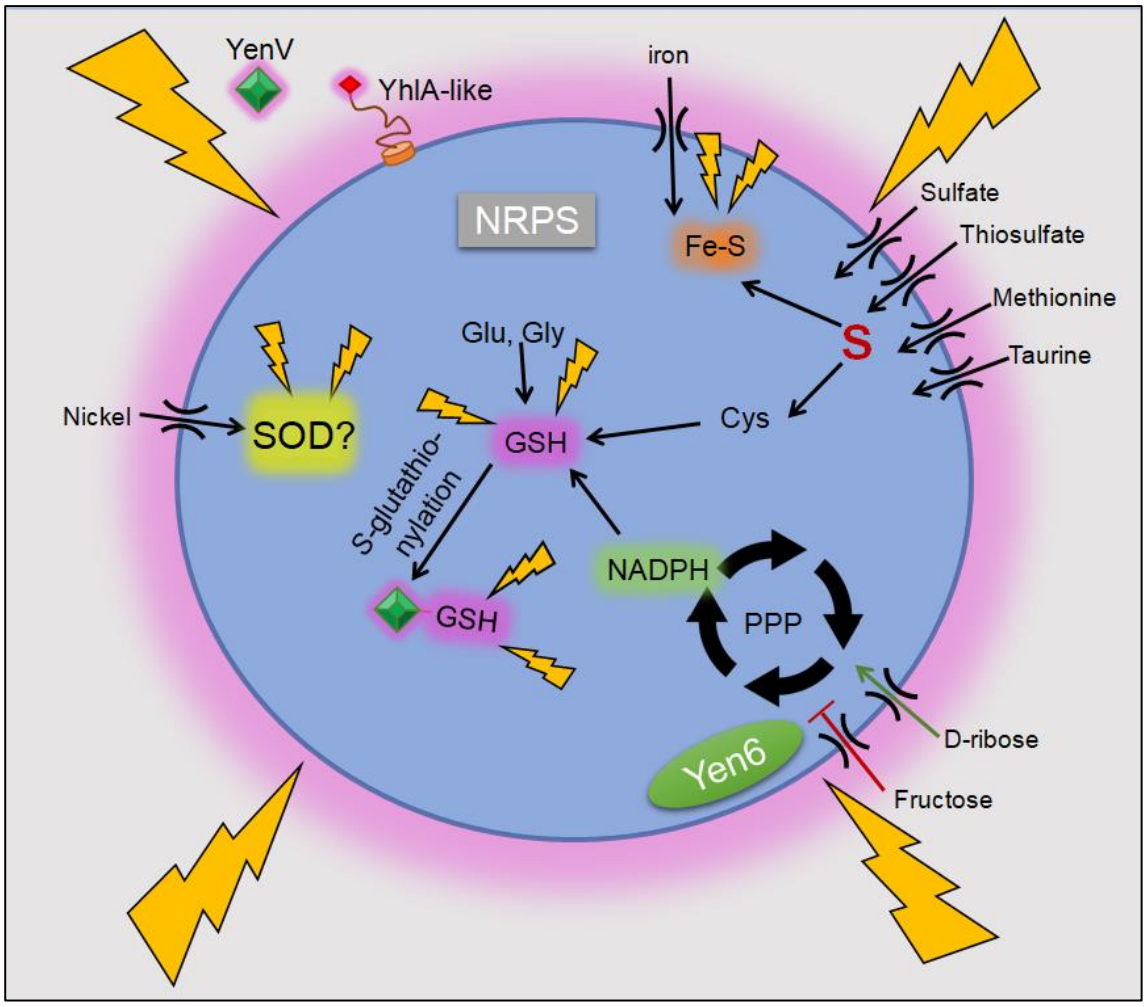


Figure 8.2: Conceptual model of *Yersinia entomophaga* MH96 cell undergoing redox and kill mode of pathogenesis during intrahemocoelic infection at 37 °C.The lightning bolts represent oxidative stress experienced by MH96 during infection at 37 °C. Fe-S = iron-sulfur cluster; Cys, Glu and Gly = cysteine, glutamic acid and glycine, respectively; PPP = pentose phosphate pathway; GSH = glutathione, NRPS = non-ribosomal protein synthetase, NADPH = nicotinamide dinucleotide phosphate oxidase and SOD = superoxide dismutase.

Redox stress is encountered by bacterial pathogens during assaults from the host immune system during infection, which include production of various reactive oxygen and nitrogen species (ROS and RNS) as well as other reactive chlorine and electrophilic species (Van Loi et al., 2015; Reniere, 2018). In order to survive within host environment, pathogenic bacteria have evolved a number of oxidative stress resistance strategies including the production of potent antioxidants, such as superoxide dismutase (SOD) (Broxton and Culotta, 2016), glutathione (GSH) (Song et al., 2013) and iron-sulfur (Fe-S) clusters (Amich et al., 2013; Miller and Auerbuch, 2015), for example. Since the operon for the nickel uptake system, *nikABCDE* was found to be highly

expressed *in vivo* at 37 °C, this suggests MH96 may harbor a SOD that utilizes nickel, in this instance SOD from other pathogens have been shown to use different metals such as zinc, copper, iron, manganese and nickel (Garcia et al., 2017; Pratt et al., 2015; Ryan et al., 2015).

Also, similar to eukaryotic cells, bacteria are also thought to post-transcriptionally modify proteins by the addition of reducing thiols under oxidative stress conditions in a process known as *S*-glutathionylation (Van Loi et al., 2015; Zhang et al., 2018a). Recently, *S*-glutathionylation was shown to be involved in virulence of *Y. pestis* because the tip protein, LcrV, of the T3SS that translocate Yop effectors is post-transcriptionally modified at Cys-273 by addition of glutathione when *Y. pestis* is inside the macrophage and this modification affected virulence by modulation of Yop secretion, macrophage killing and increased virulence in murine infection models (Mitchell et al., 2017). Taken together, it is tempting to speculate that RNA-binding protein YhbY, which was identified as part of the Yen6 regulon, may actually be involved in post-transcriptional modifications required to resist oxidative stress by MH96 during *in vivo* conditions at 37 °C. The primary toxin secreted by MH96 at 37 °C was shown to be a large molecular weight filamentous hemagglutinin N-terminal containing protein (Chapter 5, section 5.3.1.1, supplementary Figure S16 and Table S12) secreted by two-partner secretion (type 5b secretion system), but the PAI_{YE96}-encoded gene for YenV (putative calcium/calmodulin dependent protein kinase) was also found to significantly respond to 37 °C *in vivo* conditions (Chapter 4, section 4.3.8, supplemental Table S10) and may also be produced by MH96 during infection at 37 °C as well. Further discussions related to the potential role of Yen6 for oxidative stress response are provided in section 8.5.2 below.

8.3 Updated model of host- and temperature-dependent regulation of Yen-TC and TC-associated factors

A large proportion of work that was undertaken in this thesis involved dissecting the molecular mechanisms of temperature- and host-dependent regulation of Yen-TC component genes by two transcriptional regulators Yen6 and Yen7 (see Chapter 5).

Ultimately, phenotyping by protein visualization (Chapter 5, section 5.3.1.1, Figure 5.4 and Figure 5.5), transcriptome analysis of the Δ*yen6* strain (Chapter 6,section 6.3.1.4, Table 6.6) and EMSA (Chapter 6, section 6.3.2.3, Figure 6.11), as well as additional experimental involving *yen6* over-expression (Chapter 5, section 5.3.3.2, Figure 5.19) and P_{chi1}::*lacZ* translational reporter strain (Chapter 5, section 5.3.3.6.3, Table 5.10 and section 5.3.3.6.4, Figure 5.28), but all failed to provide evidence of direct transcriptional linkages between Yen6 and *yen7* and/or Yen-TC component genes.

While *yen6* (and possibly its 3'UTR-encoded asRNA *yen7as*) was originally hypothesized to act as a master transcriptional repressor of *yen7* and possibly Yen-TC component genes, in light of these investigations, *yen6* is no longer considered to be directly involved in transcriptional repression of *yen7* or Yen-TC components. Other well-known global virulence regulators such as heat-labile YmoA or H-NS counter-silencer RovA both represent potential candidates that may contribute the thermoregulation of virulence in MH96. Additional work to phenotype Δ*rovA* identified a phenotype at the protein-level, where the YenA2 component was not observed by SDS-PAGE at 25 °C (Chapter 5, section 5.3.1.1, Figure 5.6). The Δ*rovA* mutant was not found to have attenuated virulence during intrahemocoelic infection of *G. mellonella* at 25 °C. Additional work focusing on post-transcriptional regulation of virulence in MH96 is also warranted, especially since this work identified possible temperature-dependent post-transcriptional regulation of Yen-TC. In this sense, targeted deletion of the RNA-chaperone *hfq* would allow for further characterization of the role of ncRNA networks regulating virulence of MH96 (previously noted that work in Hurst lab identified a mutation in *hfq* abolishes secretion by MH96 at 25 °C).

While investigation to define the exact regulon of Yen7 was not the primary focus of this thesis, molecular experimental work did provide evidence that Yen7 can activate Yen-TC secretion at 37 °C (Chapter 5, section 5.3.3.2, Figure 5.21). Since the Δ*yen7* strain was still able to secrete Yen-TC components into the culture supernatant as wild-type (Chapter 5, section 5.3.1.1 and Figure 5.5), these results suggested other regulators, in addition to Yen7, also contribute to the transcriptional regulation of the Yen-TC components. The *in vivo* transcriptome helped to identify another potential regulator of Yen-TC that shares similarities with *yen7*; the PhoB-like regulator was identified

among the TC-associated factors and is co-expressed with Yen-TC components, *yen7* and all of the other co-secreted insecticidal machinery. The PhoB-like regulator also shares significant sequence homology to the amino acid sequence of Yen7 as well.

Later discovery of an unstable region, $HCUI_{YE96}$, which contains PhoB-like regulator, holin-endolysin secretion system and citrate fermentation cluster also support a functional redundancy between Yen7 and PhoB-like in the activation of Yen-TC component gene expression and holin-endolysin secretion system genes because $\Delta HCUI_{YE96}$ strain was found to lack production of Yen-TC and secretion at 25 °C. This thesis explored even further into the evolutionary history of $HCUI_{YE96}$ by comparison to a similar suspected PAI found only among pathogenic strains of *Y. ruckeri*, which suggested that $HCUI_{YE96}$ may have acquired the holin-endolysin secretion system from PAI_{YE96} during a historic recombination event. This theory could provide an explanation the duplication of the phage-holin transcriptional activator PhoB-like and a functionally redundant *yen7* paralogs located on different HGT-acquired elements within the genome of MH96 (Chapter 6, section 6.3.1.4 - Table 6.7; Chapter 7, section 7.4.1 - Figure 7.1). Additional work to generate a MH96 deficient for both *yen7* and PhoB-like regulator would help to determine if these putative paralogs are the primary activators of virulence in MH96.

8.4 Challenges and limitations of this work

Working with *in vivo* system presents a number of challenges as well as opportunities, so efforts were taken during *in vivo* experimentation with MH96 to improve consistency, reduce variability and improve statistical robustness whenever possible. Furthermore, due to the temperature-dependent characteristics of MH96, particularly increased clumping and flocking of cells when cultured at 37 °C, additional mitigation measures were required to improve consistencies and improve comparability between cultures grown at different temperatures. A summary of the major challenges, associated mitigation measures and remaining limitations are provided below in Table 8.1.

Table 8.1: Summary table of challenges encountered during this thesis, mitigations measures taken and outstanding limitations.

1.Highly variable mortality rates of *G. mellonella* by intrahemocoelic infection by MH96 due to extremely low median lethal dose of less than a single cell.	
Mitigations	• Increased sample size (n = 10); • repeated bioassay in triplicate; • enumerated inoculum from micro-injector; • inoculated with a range of five dilutions; and • used consistently sized *G. mellonella* within one week upon receipt maintained on standard artificial diet.
Outstanding limitations	– Highly variable phenotypes are a result of the interaction of host and pathogen genotype, therefore high levels of variability observed at the single-cell level making robust statistical analysis challenging using this model system.
Next steps	Alternative models of virulence testing should be identified for infection of MH96 at 25 °C, such as S2 hemocyte-like *Drosophila* or freshly harvested hemocytes from *G. mellonella*. Unlike 25 °C, the intrahemocoelic model of *G. mellonella* at 37 °C was found to be adequate for virulence testing as both Δ*yen6* and Δ*cspA123*/ΔHCUI$_{YE96}$ strains. This model would be recommended for future virulence testing for putative regulators or VFs involved in oxidative stress response, including components of the Yen6 regulon or the new Δ*cspA123* strain.

2.Expression data from the transcriptome were not validated by qPCR/Nanostring	
Mitigations/ Alternative validation	• Increased replicate samples for *in vivo* samples (n = 4); • confirmed high-degree of consistency among replicate RNA-seq libraries by regression analysis, with exception of *in vitro* stationary samples (see 1 above); • protein visualization at 25 and 37 °C confirmed co-expression of Yen-TC/TC-associated factors; • *lacZ* reporter fusions for *chi1* and *yen6* validated temperature-dependent regulation of these two genes (except P_{yen6}::*lacZ* under *in vitro* conditions, where temperature was found to have an opposite effect on translation); • complete lack of raw reads from Δ*cspA123*/$\Delta HCUI_{YE96}$ aligning to excised region in the genome; and • transcriptome-derived Yen6 regulon was validated by EMSA.
Outstanding limitation	– Increased variation related to *in vitro* stationary stage sample likely contributed increased noise affecting prediction of *in vivo* putative VFs from late infection stages and further validation would be recommended for these specific results; and – Expression of ncRNAs and asRNAs were found to have increased variability compared to protein coding transcripts, so further validation of ncRNA/asRNA would also be recommended.
Next steps	Temperature- and host-dependent changes in transcription of Yen-TC component genes and TC-associated factors could be further validated by including each of these genes in a Nanostring custom gene expression panel, possibly combined with other select putative VFs, such as the filamentous hemagglutinin N-terminal containing protein that is secreted at 37 °C for example.

3.**Difficult to collect consistent cell densities of MH96 between 25 and 37 °C.**	
Mitigations	• Bacterial enumeration by plate count was carried out when feasible (i.e., plasmid stabilities, culture samples used for protein visualization or *in vivo* cell density enumerations from *G. mellonella* homogenate); • Plate enumerations were not carried out for β-galactosidase (β-gal) assays; however, β-gal activity was reported in Miller Unit (MU), a relative enzymatic rate normalized by optical density at 600 nm (OD_{600}) to control for differences in cell density samples; • Inherent challenges with increased cell clumping/flocking of MH96 when cultured at 37 °C compared to 25 °C were mitigated by collection of larger volume samples, adequate mixing by vortex and all dilutions made into 5ml total volumes using consistent pipetting methods across all experiments; and • All experiments included adequate biological replicates and were repeated a minimum of two or three times to ensure consistency.
Outstanding limitation	– The calibration curve of MH96 grown at 25 and 37 °C as measured by OD_{600} from the plate reader identified inherent differences exist when MH96 is grown at 25 compared to 37 °C and this limitation must be considering when extrapolating β-gal assays results that compare MH96 *lacZ*-reporter strains grown at 25 and 37 °C as the rate of β-gal production is relative to total number of cells in the sample.
Next steps	Temperature-induced phenotypic dimorphism is a common property of pathogenic bacteria, including MH96. In this thesis, increasing sample collection and dilution volume, adequate mixing by vortex and consistent practices were key to producing consistent enumeration plates and β-gal assays results. In addition to continuing these practices for future experiments, a mathematical method for calculating MU equivalencies based on calibration curve data would possibly provide a more robust analysis assessing temperature-dependent effects using gene reporters in MH96.

4.Reduced arabinose-induction plasmid stability may have confounded molecular results	
Mitigations	• Selected carbenicillin over ampicillin for arabinose-induction experiments to improve plasmid stabilities and determined estimated plasmid stability following induction/complementation experiments; • determined pBAD-*yen6* and pBAD-*yen7* plasmid stabilities under a range of inducing conditions at 25 and 37 °C; and • Optimized arabinose induction conditions for pBAD-*yen6* and pBAD-*yen7* to achieve the highest possibly plasmid stabilities and most consistent cell numbers for protein visualization and β-gal assays.
Outstanding limitation	– While optimized conditions greatly improved cell growth and plasmid stability for pBAD and pBAD-*yen6*, this could not be achieved for pBAD-*yen7*. Therefore, experiments involving pBAD-*yen7* (i.e., protein visualization or β-gal assays results) must consider potential confounding effects due to inconsistent cell densities or plasmid stabilities among treatments.
Next steps	Plasmid stability was a major issue encountered during this thesis, and to avoid confounding effects a method of complementation or constitutive gene expression by chromosomal insertion would be advised. Along these lines, similar suicide plasmids, such as pVIK107 that was used to make the *lacZ*-reporter strains, which undergo single-recombination to form a *cis*-merodiploid strains are considered preferable for integrating genetic fragments into the genome of MH96 rather than plasmid-based systems.

8.5 Conclusions

8.5.1 Final outcomes

This thesis aimed to characterize the host- and temperature-dependent transcriptional responses of MH96 as a means to:

1. Provide a formative *in vivo* transcriptome of an entomopathogenic bacteria within model host *G. mellonella*, which fills a major gap in the literature.

a. The methodologies described in Chapter 4 of this thesis represent a guide on how to identify potential novel biopesticide agents from entomopathogenic bacteria as well as a means to reveal insights into host- and temperature-dependent virulence regulatory networks among pathogenic bacteria in general;

2. Provide a formative investigation into the molecular mechanisms controlling regulation of MH96's main insecticidal weapon: Yen-TC, a toxin delivery system found among diverse entomopathogenic and non-entomopathogenic bacteria;

   a. This thesis presents the first transcriptional investigation into thermoregulation of Yen-TC component genes, and as a result of this thesis an interaction between *in vivo* conditions and temperature was shown to result in significant downregulation of (not only) Yen-TC component genes, but a core consortia of putative VFs including co-secreted exoenzymes; and

   b. This thesis also determined the DNA-binding regulator Yen6 does not directly regulate putative regulator *yen7* or Yen-TC; however, *yen7* can activate Yen-TC production/secretion (even at 37 °C) but is not the only regulator involved.

3. Characterization of novel LytTR-containing DNA-binding protein Yen6 by combining mutant RNA-seq, mutant phenotype by microarray for sole-carbon utilization and specific DNA-binding EMSA;

   a. Characterization of Yen6 and its regulon revealed novel insights into a possible oxidative stress response strategy used by MH96 during intrahemocoelic infection of *G. mellonella* at 37 °C and these findings contribute to a greater functional understanding of Yen6 as a new type of

LytTR-containing regulator, which was previously known as a hypothetical protein.

4. First report of an unstable element that can actively excise from the genome of MH96, including preliminary phenotype investigations linking the region to Yen-TC production, secretion, cell shape and motility;

    a. An important, yet unanticipated, result of this thesis was discovery of unstable HCUI$_{YE96}$ in the genome of MH96, especially the linkages between the co-located PhoB-like regulator and holin-endolysin secretion system from HCUI$_{YE96}$ and *yen7* and the Yen-TC component genes from PAI$_{YE96}$. These finding also correlated with previous *in vivo* transcriptome results that also linked TC-associated factors (including genes for phage-related holin-endolysin secretion system, PhoB-like and AraC-like) as co-expressed with Yen-TC.

8.5.2 Future perspectives

Major efforts were undertaken as part of this thesis to determine whether Yen6 acts as a direct transcriptional regulator of *yen7* or Yen-TC component genes but ultimately several lines of inquiry converged to determine there were limited experimental evidence supporting a direct role of Yen6 as transcriptional repression of *yen7* or Yen-TC. These results suggest there must be other host- and temperature regulatory mechanisms controlling repression of Yen-TC component and TC-associated factor genes observed under *in vivo* conditions at 37 °C, yet these remain to be elucidated. Therefore, important future work should continue trying to identify these yet unknown regulatory mechanisms, including targeted mutagenesis and phenotyping of MH96-derivative strains deficient for putative global regulators YmoA and RovA. Along these lines, Δ*rovA* was found to have phenotypic differences in (at least) YenA2 production at the protein level (Chapter 5, section 5.3.1.1 and Figure 5.6) and mutant RNA-seq of Δ*rovA* may provide more clues into the role of this putative H-NS counter silencer in the virulence of MH96. Unfortunately, it is unlikely a H-NS deficient mutant

can be generated within the *Yersinia* spp. genetic background as it is considered an essential gene among the group.

Aside to the *in vivo* RNA-seq data, the most important finding of this thesis is the discovery of novel LytTR-containing regulator Yen6 and its unique regulon, which appears to play a role in response to oxidative stress by modulation of carbon flux through the PPP. These findings warrant more research, as this mechanism has not yet been proposed for pathogenic bacteria despite several previous reports of upregulation of ribose uptake/acquisition genes among pathogenic *Yersinia* spp. upon temperature shift to 37 °C (Motin et al., 2004; Rosso et al., 2008), exposure to blood plasma in *Y. pseudotuberculosis* (Rosso et al., 2008) or during occupation of the flea gut in *Y. pestis* (Vadyvaloo et al., 2010). Furthermore, genes for ribose uptake have been identified as induced under *in vivo* conditions using promoter traps for several mammalian pathogens including *Klebsiella pneumoniae* during mouse infection (Lai et al., 2001), *Haemophilus influenzae* infecting chinchilla (Mason et al., 2003) and *Salmonella enterica* serovar Typhi during infection of macrophage (Daigle et al., 2001). For the most part, when ribose uptake acquisition systems were recognized as responsive to temperature shifts or *in vivo* conditions it was assumed that this response must supported metabolic adaptation for uptake of host-derived sugars and not necessarily associated with oxidative stress resistance. Therefore, testing resistance of Δ*yen6* to oxidative stress represents the next critical experiment to prove that *yen6* functions to provide resistance to oxidative stress during infection at 37 °C, especially since the mutation was found to have limited effects on sole-carbon utilization by MH96.

Increased NADPH generation via the PPP by pathogenic bacteria in response to host-induced oxidative stress is a relatively novel idea (Christodoulou et al., 2018), but a similar strategy has already been recognized from blood-borne pathogenic trypanosomatids, such as medically important *Trypanosoma brucei, T. cruzi* and (Kovářová and Barrett, 2016). In *T. brucei* and *T. cruzi* glucose-6-phophate dehydrogenase (G6PDH, the rate limiting enzyme of PPP) was shown to be important for defense against oxidative stress (Gupta et al., 2011; Igoillo-Esteve and Cazzulo, 2006). Interestingly, a recently-acquired nonsense mutation in G6PDH of epidemic *Y. pestis* strains has also been recognized as a potentially key mutation supporting divergence of plague bacillus from ancestral *Y. pseudotuberculosis* as a specialized

zoonosis vectored by flea (Bearden and Brubaker, 2010; Bearden et al., 2009). The gene for G6PDH was still shown to be induced upon temperature shift to 37 °C despite the mutation (Motin et al., 2004). In humans, G6PDH deficiency is the most common enzymopathy, which gives the carrier resistance against severe malaria (Luzzatto, 2012; Mbanefo et al., 2017; Ruwende and Hill, 1998; Santana et al., 2013), but also makes the carrier susceptible to erythrocyte lysis if the antimalarial agent, primaquine, is taken (Recht et al., 2018).

The ability of cells to respond to oxidative stress by 'reserve flux capacity' involving shifting metabolism within the oxidative branch of PPP to generate NADPH was recently recognized as conserved across all kingdoms (Christodoulou et al., 2019), including *E. coli* (Christodoulou et al., 2018). Most organisms have evolved means to regulate either G6PDH, or in the case of yeast, glyceraldehyde 3-phosphate dehydrogenase (GADPH) (Ralser et al., 2007), which is another critical enzymatic step linking PPP and glycolysis. Based on the results of Chapter 6 of this thesis, including metabolic testing of the *yen6* strain, it is strongly believed that MH96 has evolved a novel mechanism to modify reserve flux capacity by using Yen6 to simultaneously activate and repress ribose and fructose uptake/utilization systems, respectively.

While outside the scope of this thesis, more broadly G6PDH and PPP represent novel targets under investigation in both anti-malarial (Alencar et al., 2015; Allen et al., 2015) and cancer drug research (Gorrini et al., 2013). Based on the emerging interest in oxidative stress response, reserve flux capacity and development of pharmaceutical agents, assessment of Yen6 ability to modulate carbon fluctuation/resistance to oxidative stress following heterologous expression in laboratory *E. coli* is warranted to determine if Yen6's regulatory function is conserved among Gram negative bacteria. The ability to readily produce recombinant soluble Yen6 and its tendency to oligomerize under native conditions also makes it an attractive target for protein structural analysis by x-ray crystallography or electron microscopy. The exact binding motif of Yen6 could also be further resolved by conducting additional EMSA analysis using smaller probes, a selection of mutated probes or nuclease protection assay.

One of the remaining key area of future research includes determining whether $HCUI_{YE96}$ is transmissible and excision is important for MH96 virulence against insects

as well as, determining the role of HCUI$_{YE96}$-encoded AraC- and PhoB-like regulators in host- and temperature-dependent transcriptional responses of MH96. While, the unexpected discovery of the excision from the genome of the Δ*cspA123*/ΔHCUI$_{YE96}$ inspired preliminary work reported in this thesis, this work is really a baseline to support future investigations into unusual quirk of the MH96 genome.

References

Aachmann, F.L., Sorlie, M., Skjak-Braek, G., Eijsink, V.G.H., and Vaaje-Kolstad, G. (2012). NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. Proc. Natl. Acad. Sci. *109*, 18779–18784.

Abebe-Akele, F., Tisa, L.S., Cooper, V.S., Hatcher, P.J., Abebe, E., and Thomas, W.K. (2015). Genome sequence and comparative analysis of a putative entomopathogenic *Serratia* isolated from *Caenorhabditis briggsae*. BMC Genomics *16*, 1–15.

Adrangi, S., and Faramarzi, M.A. (2013). From bacteria to human: A journey into the world of chitinases. Biotechnol. Adv. *31*, 1786–1795.

Agbonlahor, D.E. (1986). Characteristics of *Yersinia intermedia*-like bacteria isolated from patients with diarrhea in Nigeria. J. Clin. Microbiol. *23*, 891–896.

Ahantarig, A., Chantawat, N., Waterfield, N.R., ffrench-Constant, Richard Waterfield, N., and Kittayapong, P. (2009). PirAB toxin from *Photorhabdus asymbiotica* as a larvicide against Dengue vectors. Appl. Environ. Microbiol. *75*, 4627–4629.

Alcoforado Diniz, J., and Coulthurst, S.J. (2015). Intraspecies competition in *Serratia marcescens* is mediated by type VI-secreted Rhs effectors and a conserved effector-associated accessory protein. J. Bacteriol. *197*, 2350–2360.

Alencar, N.A.N. de, Juarez, J.J., Silva, J.L., and Garriga, J.L. (2015). Glucose-6-Phosphate Dehydrogenase as Potential and Selective Target Against Malaria. Rev. Process. Químicos *9*, 180–182.

Alenizi, D., Ringwood, T., Redhwan, A., Bouraha, B., Wren, B.W., Prentice, M., and McNally, A. (2016). All *Yersinia enterocolitica* are pathogenic: Virulence of phylogroup 1 *Y. enterocolitica* in a *Galleria mellonella* infection model. Microbiology *162*, 1379–1387.

Allen, S.M., Lim, E.E., Jortzik, E., Preuss, J., Chua, H.H., MacRae, J.I., Rahlfs, S., Haeussler, K., Downton, M.T., McConville, M.J., et al. (2015). *Plasmodium falciparum* glucose-6-phosphate dehydrogenase 6-phosphogluconolactonase is a potential drug target. FEBS J. *282*, 3808–3823.

Almagro-Moreno, S., Napolitano, M.G., and Boyd, E.F. (2010). Excision dynamics of *Vibrio* pathogenicity island-2 from *Vibrio cholerae*: role of a recombination directionality factor VefA. BMC Microbiol. *10*, 1–10.

Altermann, E., Lu, J., and McCulloch, A. (2017). GAMOLA2, a comprehensive software package for the annotation and curation of draft and complete microbial genomes. Front. Microbiol. *8*, 1–14.

Álvarez, R., Neumann, G., Frávega, J., Díaz, F., Tejías, C., Collao, B., Fuentes, J.A., Paredes-Sabja, D., Calderón, I.L., and Gil, F. (2015). CysB-dependent upregulation of the *Salmonella* Typhimurium *cysJIH* operon in response to antimicrobial compounds that induce oxidative stress. Biochem. Biophys. Res. Commun. *458*, 46–51.

Amich, J., Schafferer, L., Haas, H., and Krappmann, S. (2013). Regulation of sulphur assimilation is essential for virulence and affects iron homeostasis of the human-pathogenic mould *Aspergillus fumigatus*. PLoS Pathog. *9*.

Amorim-Vaz, S., Tran, V.D.T., Pradervand, S., Pagni, M., Coste, A.T., and Sanglarda, D. (2015). RNA enrichment method for quantitative transcriptional analysis of pathogens *in vivo* applied to the fungus *Candida albicans*. MBio *6*, 1–16.

An, R., Sreevatsan, S., and Grewal, P.S. (2009). Comparative *in vivo* gene expression of the closely related bacteria *Photorhabdus temperata* and *Xenorhabdus koppenhoeferi* upon infection of the same insect host, *Rhizotrogus majalis*. BMC Genomics *10*, 433.

Anderson, C.J., Clark, D.E., Adli, M., and Kendall, M.M. (2015). Ethanolamine signaling promotes *Salmonella* niche recognition and adaptation during infection. PLoS Pathog. *11*, 1–20.

Andrews, S.C., Robinson, A.K., and Rodríguez-Quiñones, F. (2003). Bacterial iron homeostasis. FEMS Microbiol. Rev. *27*, 215–237.

Andronescu, M., Condon, A., Hoos, H.H., Mathews, D.H., and Murphy, K.P. (2007). Efficient parameter estimation for RNA secondary structure prediction. Bioinformatics *23*, 19–28.

Angen, Ø., Ahrens, P., Kuhnert, P., Christensen, H., and Mutters, R. (2003). Proposal of *Histophilus somni* gen. nov., sp. nov. for the three species incertae sedis "*Haemophilus somnus*", "*Haemophilus agni*" and "*Histophilus ovis*." Int. J. Syst. Evol. Microbiol. *53*, 1449–1456.

Antunes, L.C.M., Ferreira, R.B.R., Buckner, M.M.C., and Finlay, B.B. (2010). Quorum sensing in bacterial virulence. Microbiology *156*, 2271–2282.

Aoki, K., and Tiemeyer, M. (2010). The glycomics of glycan glucuronylation in Drosophila melanogaster. In Methods in Enzymology: Glycobiology, M. Fukuda, ed. (San Diego, CA: Elsevier), pp. 297–321.

APRD (2019). Arthropod pesticide resistance database, https://www.pesticideresistance.org/.

Aranda, P.S., LaJoie, D.M., and Jorcyk, C.L. (2012). Bleach gel: A simple agarose gel for analyzing RNA quality. Electrophoresis *33*, 366–369.

de Assis Alcoforado Costa, M., Owen, R.A., Tammsalu, T., Buchanan, G., Palmer, T., Sargent, F., Costa, M. de A.A., Owen, R.A., Tammsalu, T., Buchanan, G., et al. (2019). Controlling and co-ordinating chitinase secretion in a *Serratia marcescens* population. BioRxiv 652685.

Autieri, S.M., Lins, J.J., Leatham, M.P., Laux, D.C., Conway, T., and Cohen, P.S. (2007). L-fucose stimulates utilization of D-ribose by *Escherichia coli* MG1655 Δ*fucAO* and *E. coli* Nissle 1917 Δ*fucAO* mutants in the mouse intestine and in M9 minimal medium. Infect. Immun. *75*, 5465–5475.

Avican, K., Fahlgren, A., Huss, M., Heroven, A.K., Beckstette, M., Dersch, P., and Fällman, M. (2015). Reprogramming of *Yersinia* from virulent to persistent

mode revealed by complex *in vivo* RNA-seq analysis. PLoS Pathog. *11*, 1–28.

Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H.B., Trombetta, J.J., Satija, R., Shalek, A.K., Xavier, R., Regev, A., et al. (2015). Pathogen cell-to-cell variability drives heterogeneity in host immune responses. Cell *118*, 6072–6078.

Bach, S., Buchrieser, C., Prentice, M., Guiyoule, A., Msadek, T., and Carniel, E. (1999). The high-pathogenicity island of *Yersinia enterocolitica* Ye8081 undergoes low-frequency deletion but not precise excision, suggesting recent stabilization in the genome. Infect. Immun. *67*, 5091–5099.

Bach, S., de Almeida, A., and Carniel, E. (2000). The *Yersinia* high-pathogenicity island is present in different members of the family Enterobacteriaceae. FEMS Microbiol. Lett. *183*, 289–294.

Baddal, B., Muzzi, A., Censini, S., Calogero, R.A., Torricelli, G., Guidotti, S., and Taddei, A.R. (2016). Dual RNA-seq of nontypeable *Haemophilus influenzae* and host cell transcriptomes reveals novel insights into host-pathogen cross talk. MBio *7*, 1–13.

Bader, M.W., Sanowar, S., Daley, M.E., Schneider, A.R., Cho, U., Xu, W., Klevit, R.E., Le Moual, H., and Miller, S.I. (2005). Recognition of antimicrobial peptides by a bacterial sensor kinase. Cell *122*, 461–472.

Bakowski, M.A., Cirulis, J.T., Brown, N.F., Finlay, B.B., and Brumell, J.H. (2007). SopD acts cooperatively with SopB during *Salmonella enterica* serovar Typhimurium invasion. Cell. Microbiol. *9*, 2839–2855.

Banville, N., Browne, N., and Kavanagh, K. (2012). Effect of nutrient deprivation on the susceptibility of *Galleria mellonella* larvae to infection. Virulence *3*, 497–503.

Barkan, A., Klipcan, L., Ostersetzer, O., Kawamura, T., Asakura, Y., and Watkins, K.P. (2007). The CRM domain: An RNA binding module derived from an ancient ribosome-associated protein. Rna *13*, 55–64.

Barnhart, M.M., and Chapman, M.R. (2006). Curli biogenesis and function. Annu. Rev. Microbiol. *60*, 131–147.

Barquist, L., and Vogel, J. (2015). Accelerating discovery and functional analysis of small RNAs with new technologies. Annu. Rev. Genet. *49*, 367–394.

Barquist, L., Westermann, A.J., and Vogel, J. (2016). Molecular phenotyping of infection- associated small non-coding RNAs. Philos. Trans. R. Soc. B Biol. Sci. *371*.

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.

Barroga, C.F., Zhang, H., Wajih, N., Bouyer, J.H., and Hermodson, M.A. (1996). The proteins encoded by the *rbs* operon of *Escherichia coli*: I. Overproduction, purification, characterization, and functional analysis of RbsA. Protein Sci. *5*, 1093–1099.

Barth, H., Aktories, K., Popoff, M.R., and Stiles, B.G. (2004). Binary bacterial toxins: Biochemistry, biology and applications of common *Clostridium* and *Bacillus* Proteins. Microbiol. Mol. Biol. Rev. *68*, 373–402.

Bearden, S.W., and Brubaker, R.R. (2010). Recent findings regarding maintenance of enzootic variants of *Yersinia pestis* in sylvatic reservoirs and their significance in the evolution of epidemic plague. Vector-Borne Zoonotic Dis. *10*, 85–92.

Bearden, S.W., Sexton, C., Pare, J., Fowler, J.M., Arvidson, C.G., Yerman, L., Viola, R.E., and Brubaker, R.R. (2009). Attenuated enzootic (pestoides) isolates of *Yersinia pestis* express active aspartase. Microbiology *155*, 198–209.

Behr, S., Heermann, R., and Jung, K. (2016). Insights into the DNA-binding mechanism of a LytTR-type transcription regulator. Biosci. Rep. *36*, e00326–e00326.

Beier, S., and Bertilsson, S. (2013). Bacterial chitin degradation-mechanisms and ecophysiological strategies. Front. Microbiol. *4*, 1–12.

Bellanger, X., Morel, C., Gonot, F., Puymege, A., Decaris, B., and Guédon, G. (2011). Site-specific accretion of an integrative conjugative element together with a related genomic island leads to *cis* mobilization and gene capture. Mol. Microbiol. *81*, 912–925.

Bellanger, X., Payot, S., Leblond-Bourget, N., and Guédon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol. Rev. *38*, 720–742.

Bengoechea, J.A., Najdenski, H., and Skurnik, M. (2004). Lipopolysaccharide O antigen status of *Yersinia enterocolitica* O:8 is essential for virulence and absence of O antigen affects the expression of other *Yersinia* virulence factors. Mol. Microbiol. *52*, 451–469.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Stat. *29*, 1165–1188.

Bennett, G.M., and Chong, R.A. (2017). Genome-wide transcriptional dynamics in the companion bacterial symbionts of the glassy-winged sharpshooter (Cicadellidae: *Homalodisca vitripennis*) reveal differential gene expression in bacteria occupying multiple host organs. G3 *7*, 3073–3082.

Bent, Z.W., Tran-Gyamfi, M.B., Langevin, S.A., Brazel, D.M., Hamblin, R.Y., Branda, S.S., Patel, K.D., Lane, T.W., and Vandernoot, V.A. (2013b). Enriching pathogen transcripts from infected samples: A capture-based approach to enhanced host-pathogen RNA sequencing. Anal. Biochem. *438*, 90–96.

Bent, Z.W., Brazel, D.M., Tran-Gyamfi, M.B., Hamblin, R.Y., VanderNoot, V.A., and Branda, S.S. (2013a). Use of a capture-based pathogen transcript enrichment strategy for RNA-Seq analysis of the *Francisella tularensis* LVS transcriptome during infection of murine macrophages. PLoS One *8*, 1–12.

Bent, Z.W., Poorey, K., Brazel, D.M., LaBauve, A.E., Sinha, A., Curtis, D.J., House, S.E.,

Tew, K.E., Hamblin, R.Y., Williams, K.P., et al. (2015). Transcriptomic analysis of *Yersinia enterocolitica* biovar 1B infecting murine macrophages reveals new mechanisms of extracellular and intracellular survival. Infect. Immun. *83*, 2672–2685.

Benz, I., and Schmidt, M.A. (1992). AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. Mol. Microbiol. *6*, 1539–1546.

Benz, I., and Schmidt, M.A. (2001). Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin. Mol. Microbiol. *40*, 1403–1413.

Bergin, D., Reeves, E.P., Renwick, J., Wientjes, F.B., and Kavanagh, K. (2005). Superoxide production in *Galleria mellonella* hemocytes: Identification of proteins homologous to the NADPH oxidase complex of human neutrophils. Infect. Immun. *73*, 4161–4170.

Berini, F., Katz, C., Gruzdev, N., Castertelli, M., and Tettamanti, G. (2018). Microbial and viral chitinases: Attractive biopesticides for integrated pest management. Biotechnol. Adv. *36*, 818–838.

Bernal-Bayard, J., Cardenal-Muñoz, E., and Ramos-Morales, F. (2010). The *Salmonella* type III secretion effector, *Salmonella* leucine-rich repeat protein (SlrP), targets the human chaperone ERdj3. J. Biol. Chem. *285*, 16360–16368.

Berrow, Nick, S., Alderton, D., and Owens, R.J. (2009). The precise engineering of expression vectors using high-throughput in-fusion PCR cloning. In Methods in Molecular Biology: High Throughput Protein Expression and Purification - Methods and Protocols, S.A. Doyle, ed. (New Jersey, USA: Humana Press), pp. 75–90.

Betley, J.N., Frith, M.C., Graber, J.H., Choo, S., and Deshler, J.O. (2002). A ubiquitous and conserved signal for RNA localization in chordates. Curr. Biol. *12*, 1756–1761.

Biggs, D.R., and McGregor, P.G. (1996). Gut pH and amylase and protease activity in larvae of the New Zealand grass grub (*Costelytra zealandica*; Coleoptera: Scarabaeidae) as a basis for selecting inhibitors. Insect Biochem. Mol. Biol. *26*, 69–75.

Blackburn, M.B., Martin, P.A.W., Kuhar, D., Farrar, R.R., and Gundersen-Rindal, D.E. (2011). The occurrence of *Photorhabdus*-like toxin complexes in *Bacillus thuringiensis*. PLoS One *6*.

Blum, H., Beier, H., Gross, H.J. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis *8*, 93–99.

Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschäpe, H., and Hacker, J. (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-Specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. *62*, 606–614.

Böck, D., Medeiros, J.M., Tsao, H., Penz, T., Weiss, G.L., Aistleitner, K., Horn, M., and Pilhofer, M. (2017). In situ architecture, function, and evolution of a contractile injection system. Science *357*, 713–717.

Boerlin, P. (2004). Evolution of bacterial virulence. In Pathogenesis of Bacterial Infections in Animals, C.L. Gyles, J.F. Prescott, J.G. Songer, and C.O. Thoen, eds. (Ames, IA: Blackwell Publishing), pp. 13–22.

Böhme, K., Steinmann, R., Kortmann, J., Seekircher, S., Heroven, A.K., Berger, E., Pisano, F., Thiermann, T., Wolf-Watz, H., Narberhaus, F., et al. (2012). Concerted actions of a thermo-labile regulator and a unique intergenic RNA thermosensor control *Yersinia* virulence. PLoS Pathog. *8*, e1002518.

Boman, H.G., and Hultmark, D. (1987). Cell-Free Immunity in Insects. Annu. Rev. Microbiol. *41*, 103–126.

Boonyom, R., Karavolos, M.H., Bulmer, D.M., and Khan, C.M.A. (2010). *Salmonella* pathogenicity island 1 (SPI-1) type III secretion of SopD involves N- and C-terminal signals and direct binding to the InvC ATPase. Microbiology *156*, 1805–1814.

Bottone, E.J. (1997). *Yersinia enterocolitica*: The charisma continues. Clin. Microbiol. Rev. *10*, 257–276.

Bottone, E.J. (1999). *Yersinia enterocolitica*: Overview and epidemiologic correlates. Microbes Infect. *1*, 323–333.

Bowen, D.J., and Ensign, J.C. (1998). Purification and characterization of a high-molecular-weight insecticidal protein complex produced by the entomopathogenic bacterium *Photorhabdus luminescens*. Appl. Environ. Microbiol. *64*, 3029–3035.

Bowen, D., Rocheleau, T.A., Blackburn, M., Andreev, O., Golubeva, E., Bhartia, R., and Ffrench-Constant, R.H. (1998). Insecticidal toxins from the bacterium *Photorhabdus luminescens*. Science *280*, 2129–2132.

Bowen, D.J., Rocheleau, T.A., Grutzmacher, C.K., Meslet, L., Valens, M., Marble, D., Dowling, A., Ffrench-Constant, R., and Blight, M.A. (2003). Genetic and biochemical characterization of PrtA, and RTX-like metalloprotease from *Photorhabdus*. Microbiology *149*, 1581–1591.

Boyd, E.F. (2012). Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity island interactions. In Advances in Virus Research, Łobocka M., and S. W.T., eds. (Amsterdam: Elsevier Inc.), pp. 91–118.

Boyd, E.F., Almagro-moreno, S., and Parent, M.A. (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. Trends Microbiol. *17*, 47–53.

Brameyer, S., Kresovic, D., Bode, H.B., and Heermann, R. (2014). LuxR solos in *Photorhabdus* species. Front. Cell. Infect. Microbiol. *4*, 1–11.

Bravo, A., Likitvivatanavong, S., Gill, S.S., and Soberón, M. (2011). *Bacillus thuringiensis*: A story of a successful bioinsecticide. Insect Biochem. Mol. Biol. *41*, 423–431.

Bresolin, G., Morgan, J.A.W., Ilgen, D., Scherer, S., and Fuchs, T.M. (2006). Low temperature-induced insecticidal activity of *Yersinia enterocolitica*. Mol. Microbiol. *59*, 503–512.

Brillard, J., Duchaud, E., Boemare, N., Kunst, F., and Givaudan, A. (2002). The PhlA hemolysin from the entomopathogenic bacterium *Photorhabdus luminescens* belongs to the two-partner secretion family of hemolysins. J. Bacteriol. *184*, 3871–3878.

Bronner, S., Monteil, H., and Prévost, G. (2004). Regulation of virulence determinants in *Staphylococcus aureus*: Complexity and applications. FEMS Microbiol. Rev. *28*, 183–200.

Broxton, C.N., and Culotta, V.C. (2016). SOD enzymes and microbial pathogens: Surviving the oxidative storm of infection. PLoS Pathog. *12*, 8–13.

Brugirard-Ricaud, K., Givaudan, A., Parkhill, J., Boemare, N., Kunst, F., Zumbihl, R., and Duchaud, E. (2004). Variation in the effectors of the type III secretion system among *Photorhabdus* species as revealed by genomic analysis. J. Bacteriol. *186*, 4376–4381.

Brugirard-Ricaud, K., Duchaud, E., Givaudan, A., Girard, P.A., Kunst, F., Boemare, N., Brehélin, M., and Zumbihl, R. (2005). Site-specific antiphagocytic function of the *Photorhabdus luminescens* type III secretion system during insect colonization. Cell. Microbiol. *7*, 363–371.

Buchrieser, C., Brosch, R., Bach, S., Guiyoule, A., and Carniel, E. (1998). The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn tRNA* genes. Mol. Microbiol. *30*, 965–978.

Bücker, R., Heroven, A.K., Becker, J., Dersch, P., and Wittmann, C. (2014). The pyruvate-tricarboxylic acid cycle node, a focal point of virulence control in the enteric pathogen *Yersinia pseudotuberculosis*. J. Biol. Chem. *289*, 30114–30132.

Bueno, S.M., Santiviago, C.A., Murillo, A.A., Fuentes, J.A., Trombert, A.N., Rodas, P.I., Youderian, P., and Mora, G.C. (2004). Precise excision of the large pathogenicity island, SPI7, in *Salmonella enterica* serovar Typhi. J. Bacteriol. *186*, 3202–3213.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics *11*, 94.

Burtnick, M.N., Brett, P.J., Harding, S. V., Ngugi, S.A., Ribot, W.J., Chantratita, N., Scorpio, A., Milne, T.S., Dean, R.E., Fritz, D.L., et al. (2011). The cluster 1 type VI secretion system is a major virulence determinant in *Burkholderia pseudomallei*. Infect. Immun. *79*, 1512–1525.

Busby, B., Kristensen, D.M., and Koonin, E. V. (2013a). Contribution of phage-derived genomic islands to the virulence of facultative bacterial pathogens. Environ. Microbiol. *15*, 307–312.

Busby, J.N., Landsberg, M.J., Simpson, R.M., Jones, S.A., Hankamer, B., Hurst, M.R.H.,

and Lott, J.S. (2012). Structural analysis of Chi1 chitinase from Yen-Tc: The multisubunit insecticidal ABC toxin complex of *Yersinia entomophaga*. J. Mol. Biol. *415*, 359–371.

Busby, J.N., Panjikar, S., Landsberg, M.J., Hurst, M.R.H., and Lott, J.S. (2013b). The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. Nature *501*, 547–550.

Busby et al., in prep. (2019). No Title.

Bushnell, B. (2015). BBMap short-read aligner, and other bioinformatics tools. http://sourceforge.net/projects/bbmap/.

Cafferkey, M.T., Sloane, A., McCrae, S., and O'Morain, C.A. (1993). *Yersinia frederiksenii* infection and colonization in hospital staff. J. Hosp. Infect. *24*, 109–115.

Caldas, C., and Cherqui, A. (2002). Purification and characterization of an extracellular protease from *Xenorhabdus nematophila* involved in insect immunosuppression. Appl. Environ. Microbiol. *68*, 1297–1304.

Caldelari, I., Chao, Y., Romby, P., and Vogel, J. (2013). RNA-mediated regulation in pathogenic bacteria. Cold Spring Harb. Perspect. Med. 1–22.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics *10*, 1–9.

Cao, Z., Casabona, M.G., Kneuper, H., Chalmers, J.D., and Palmer, T. (2016). The Ess/type VII secretion system of *Staphylococcus aureus* secretes a nuclease toxin that targets competitor bacteria. Nat. Microbiol. *2*.

Capra, E.J., and Laub, M.T. (2012). Evolution of two-component signal transduction systems. Annu. Rev. Microbiol. *66*, 325–347.

Carpenter, M.R., Rozovsky, S., and Fidelma, E. (2016). Pathogenicity island cross talk mediated by recombination directionality factors facilitates excision from the chromosome. J. Bacteriol. *198*, 766–776.

Carter, G.P., Lyras, D., Allen, D.L., Mackin, K.E., Howarth, P.M., O'Connor, J.R., and Rood, J.I. (2007). Binary toxin production in *Clostridium difficile* is regulated by CdtR, a LytTR family response regulator. *189*, 7290–7301.

Cascales, D., Guijarro, J.A., and Jessica, A.I.G. (2017). Comparative genome analysis reveals important genetic differences among serotype O1 and serotype O2 strains of *Y . ruckeri* and provides insights into host adaptation and virulence. Microbiologyopen *6*, 1–11.

Cathelyn, J.S., Crosby, S.D., Lathem, W.W., Goldman, W.E., and Miller, V.L. (2006). RovA, a global regulator of *Yersinia pestis*, specifically required for bubonic plague. Proc. Natl. Acad. Sci. U.S.A. *103*, 13514–13519.

Cathelyn, J.S., Ellison, D.W., Hinchliffe, S.J., Wren, B.W., and Miller, V.L. (2007). The RovA regulons of *Yersinia enterocolitica* and *Yersinia pestis* are distinct: Evidence that many RovA-regulated genes were acquired more recently than

the core genome. Mol. Microbiol. *66*, 189–205.

Cerenius, L., Lee, B.L., and Söderhäll, K. (2008). The proPO-system: pros and cons for its role in invertebrate immunity. Trends Immunol. *29*, 263–271.

Chaloner, T., van Kan, J.A.L., and Grant-Downton, R.T. (2016). RNA 'information warfare' in pathogenic and mutualistic interactions. Trends Plant Sci. *21*, 738–748.

Champion, O.L., Cooper, I.A.M., James, S.L., Ford, D., Karlyshev, A., Wren, B.W., Duffield, M., Oyston, P.C.F., and Titball, R.W. (2009). *Galleria mellonella* as an alternative infection model for *Yersinia pseudotuberculosis*. Microbiology *155*, 1516–1522.

Champion, O.L., Wagley, S., and Titball, R.W. (2016). *Galleria mellonella* as a model host for microbiological and toxin research. Virulence *7*, 840–845.

Chan, J.P., Wright, J.R., Wong, H.T., Ardasheva, A., Brumbaugh, J., McLimans, C., and Lamendella, R. (2019). Using bacterial transcriptomics to investigate targets of host-bacterial interactions in *Caenorhabditis elegans*. Sci. Rep. *9*, 1–12.

Chang, A.C.Y., and Cohen, S.N. (1978). Construction and characterization of amplifiable DNA cloning vectors derived from P15A cryptic plasmid. J. Bacteriol. *134*, 1141–1156.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., et al. (2004). Carbon nutrition of *Escherichia coli* in the mouse intestine. Proc. Natl. Acad. Sci. U.S.A. *101*, 7427–7432.

Charpentier, X., and Oswald, E. (2004). Identification of the secretion and translocation domain of the enteropathogenic and enterohemorrhagic *Escherichia coli* effector Cif, using TEM-1 β-lactamase as a new fluorescence-based reporter. J. Bacteriol. *186*, 5486–5495.

Chauhan, N., Hatlem, D., Orwick-Rydmark, M., Schneider, K., Floetenmeyer, M., van Rossum, B., Leo, J.C., and Linke, D. (2019). Insights into the autotransport process of a trimeric autotransporter, *Yersinia* Adhesin A (YadA). Mol. Microbiol. *111*, 844–862.

Chen, A., Wang, Y., Shao, Y., Zhou, Q., Chen, S., Wu, Y., and Chen, H. (2018). Genes involved in *Beauveria bassiana* infection to *Galleria mellonella*. Arch. Microbiol. *200*, 541–552.

Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: Hierarchical and refined dataset for big data analysis - 10 years on. Nucleic Acids Res. *44*, D694–D697.

Cheung, P.Y.K., Grula, E.A., Satayamurthy, N., and Berlin, K.D. (1982). Presence of free putrescine in the hemolymph of corn earworm (*Heliothis zea*) larvae. Insect Biochem. *12*, 41–48.

Choi, J.W., Um, J.H., Cho, J.H., and Lee, H.J. (2017). Tiny RNAs and their voyage via extracellular vesicles: Secretion of bacterial small RNA and eukaryotic

microRNA. Exp. Biol. Med. *242*, 1475–1481.

Choi, J.W., Kwon, T.Y., Hong, S.H., and Lee, H.J. (2018). Isolation and characterization of a microRNA-size secretable small RNA in *Streptococcus sanguinis*. Cell Biochem. Biophys. *76*, 293–301.

Choi, J.Y., Whitten, M.M.A., Cho, M.Y., Lee, K.Y., Kim, M.S., Ratcliffe, N.A., and Lee, B.L. (2002). Calreticulin enriched as an early-stage encapsulation protein in wax moth *Galleria mellonella* larvae. Dev. Comp. Immunol. *26*, 335–343.

Christensen, H., Bertelsen, M.F., Bojesen, A.M., and Bisgaard, M. (2012). Classification of *Pasteurella* species B as *Pasteurella oralis* sp. nov. Int. J. Syst. Evol. Microbiol. *62*, 1396–1401.

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., and Sauer, U. (2018). Reserve flux capacity in the pentose phosphate pathway enables *Escherichia coli's* rapid response to oxidative stress. Cell Syst. *6*, 569–578.e7.

Christodoulou, D., Kuehne, A., Estermann, A., Fuhrer, T., Lang, P., and Sauer, U. (2019). Reserve flux capacity in the pentose phosphate pathway by NADPH binding is conserved across kingdoms. IScience *19*, 1133–1144.

Chronakis, I.S. (1998). On the molecular characteristics, compositional properties, and structural-functional mechanisms of maltodextrins: A review. Crit. Rev. Food Sci. Nutr. *38*, 599–637.

Cianfanelli, F.R., Monlezun, L., and Coulthurst, S.J. (2016). Aim, load, fire: The type VI secretion system, a bacterial nanoweapon. Trends Microbiol. *24*, 51–62.

Citartan, M., Raabe, C.A., Hoe, C.H., Rozhdestvensky, T.S., and Tang, T.H. (2016). Bacterial sRNAs: Regulation in stress. In Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria, F.J. de Bruijin, ed. (New York, NY), pp. 108–114.

Clifton, M.C., Simon, M.J., Erramilli, S.K., Zhang, H., Zaitseva, J., Hermodson, M.A., and Stauffacher, C. V. (2015). *In vitro* reassembly of the ribose ATP-binding cassette transporter reveals a distinct set of transport complexes. J. Biol. Chem. *290*, 5555–5565.

Colombi, E., Straub, C., Künzel, S., Templeton, M.D., McCann, H.C., and Rainey, P.B. (2017). Evolution of copper resistance in the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* through acquisition of integrative conjugative elements and plasmids. Environ. Microbiol. *19*, 819–832.

Cornelis, G.R. (2002). The *Yersinia* YSC-YOP "type III" weaponry. Nat. Rev. Mol. Cell Biol. *3*, 742–752.

Cornells, G.R., Sluiters, C., Delor, I., Geib, D., Kaniga, K., de Rouvroit, C.L., Sory, M. -P, Vanooteghem, J. -C, and Michiels, T. (1991). *ymoA*, a *Yersinia enterocolitica* chromosomal gene modulating the expression of virulence functions. Mol. Microbiol. *5*, 1023–1034.

Cossart, P., Sesto, N., Wurtzel, O., Archambaud, C., and Sorek, R. (2013). The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat. Rev.

Microbiol. *11*, 75–82.

Costa, T.R.D., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., and Waksman, G. (2015). Secretion systems in Gram-negative bacteria: Structural and mechanistic insights. Nat. Rev. Microbiol. *13*, 343–359.

Costechareyre, D., Chich, J.F., Strub, J.M., Rahbé, Y., and Condemine, G. (2013). Transcriptome of *Dickeya dadantii* infecting *Acyrthosiphon pisum* reveals a strong defense against antimicrobial peptides. PLoS One *8*, e54118.

Cowles, K.N., and Goodrich-Blair, H. (2005). Expression and activity of a *Xenorhabdus nematophila* haemolysin required for full virulence towards *Manduca sexta* insects. Cell. Microbiol. *7*, 209–219.

Crawford, J.M., Kontnik, R., and Clardy, J. (2010). Regulating Alternative Lifestyles in Entomopathogenic Bacteria. Curr. Biol. *20*, 69–74.

Crawley, M.J. (2013). The R Book (Chichester, West Sussex, United Kingdom: Wiley).

Crennell, S.J., Tickler, P.M., Bowen, D.J., and Ffrench-Constant, R.H. (2000). The predicted structure of photopexin from *Photorhabdus* shows the first haemopexin-like motif in prokaryotes. FEMS Microbiol. Lett. *191*, 139–144.

Crickmore, N., Baum, J., Bravo, A., Lereclus, D., Narva, K., and Sampson, K. (2014). *Bacillus thuringiensis* toxin nomenclature.

Crosland, R.D., Fitch, R.W., and Hines, H.B. (2005). Characterization of β-leptinotarsin-h and the effects of calcium flux antagonists on its activity. Toxicon *45*, 829–841.

Culliney, Thomas, W. (2014). Crop losses to Arthropods. In Integrated Pest Management, D. Pimentel, and R. Peshin, eds. (Dordrecht: Springer), pp. 201–225.

Daborn, P.J., Waterfield, N., Blight, M.A., and Ffrench-Constant, R.H. (2001). Measuring virulence factor expression by the pathogenic bacterium *Photorhabdus luminescens* in culture and during insect infection. J. Bacteriol. *183*, 5834–5839.

Daborn, P.J., Waterfield, N., Silva, C.P., Au, C.P.Y., Sharma, S., and ffrench-Constant, R.H. (2002). A single *Photorhabdus* gene, makes caterpillars floppy (mcf), allows *Escherichia coli* to persist within and kill insects. Proc. Natl. Acad. Sci. U.S.A. *99*, 10742–10747.

Daigle, F., Graham, J.E., and Curtiss, R. (2001). Identification of <i>Salmonella typhi> genes expressed within macrophages by selective capture of transcribed sequences (SCOTS). Mol. Microbiol. *41*, 1211–1222.

Daleke, M.H., Ummels, R., Bawono, P., Heringa, J., Vandenbroucke-Grauls, C.M.J.E., Luirink, J., and Bitter, W. (2012). General secretion signal for the mycobacterial type VII secretion pathway. Proc. Natl. Acad. Sci. U.S.A. *109*, 11342–11347.

Damron, F.H., Oglesby-Sherrouse, A.G., Wilks, A., and Barbier, M. (2016). Dual-seq transcriptomics reveals the battle for iron during *Pseudomonas aeruginosa* acute

murine pneumonia. Sci. Rep. *6*, 39172.

Darmon, E., and Leach, D.R.F. (2014). Bacterial genome instability. Microbiol. Mol. Biol. Rev. *78*, 1–39.

Darvish Alipour Astaneh, S., Rasooli, I., and Mousavi Gargari, S.L. (2014). The role of filamentous hemagglutinin adhesin in adherence and biofilm formation in *Acinetobacter baumannii* ATCC19606T. Microb. Pathog. *74*, 42–49.

Degnan, P.H., and Moran, N.A. (2008b). Diverse phage-encoded toxins in a protective insect endosymbiont. Appl. Environ. Microbiol. *74*, 6782–6791.

Degnan, P.H., and Moran, N.A. (2008a). Evolutionary genetics of a defensive facultative symbiont of insects: Exchange of toxin-encoding bacteriophage. Mol. Ecol. *17*, 916–929.

Degnan, P.H., Yu, Y., Sisneros, N., Wing, R.A., and Moran, N.A. (2009). *Hamiltonella denfsa*, genome evoltuion of protective bacterial endosymbiont from pathogenic ancestors. Proc. Natl. Acadamy Sci. United States Am. *106*, 9063–9068.

Delepelaire, P. (2004). Type I secretion in gram-negative bacteria. Biochim. Biophys. Acta - Mol. Cell Res. *1694*, 149–161.

Delor, I., and Cornelis, G.R. (1992). Role of *Yersinia enterocolitica* Yst toxin in experimental infection of young rabbits. Infect. Immun. *60*, 4269–4277.

Derzelle, S., Turlin, E., Duchaud, E., Pages, S., Kunst, F., Givaudan, A., and Danchin, A. (2004). The PhoP-PhoQ two-component regulatory system of *Photorhabdus luminescens* is essential for virulence in insects. J. Bacteriol. *186*, 1270–1279.

Desbois, A.P., and McMillan, S. (2015). Paving the way to acceptance of *Galleria mellonella* as a new model insect. Virulence *6*, 410–411.

Dieppois, G., Opota, O., Lalucat, J., and Lamaitre, B. (2015). *Pseudomonas entomophila*: A versatile bacterium with entomopathogenic properties. In *Pseudomonas*: Volume 7: New Aspects of Pseudomonas Biology, J.L. Ramos, J.B. Goldberg, and A. Filloux, eds. (Springer), pp. 25–49.

Ding, L., Wang, Y., Hu, Y., Atkinson, S., Williams, P., and Chen, S. (2009). Functional characterization of FlgM in the regulation of flagellar synthesis and motility in *Yersinia pseudotuberculosis*. Microbiology *155*, 1890–1900.

Ding, X., Luo, Z., Xia, L., Gao, B., Sun, Y., and Zhang, Y. (2008). Improving the insecticidal activity by expression of a recombinant *cry*1Ac gene with chitinase-encoding gene in acrystalliferous *Bacillus thuringiensis*. Curr. Microbiol. *56*, 442–446.

Dobrindt, U., Hentschel, U., Kaper, J.B., and Hacker, J. (2002). Genome plasticity in pathogenic and nonpathogenic Enterobacteria. In Current Topics in Microbiology and Immunology, J. Hacker, and J.B. Kaper, eds. (Berlin, Heidelberg: Springer-Verlag), pp. 157–175.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic island in

pathogenic and environmental microorganisms. Nat. Rev. Microbiol. *2*, 414–425.

Dobrindt, U., Tjaden, S., Shah, S., and Hacker, J. (2015). Mobile genetic elements and pathogenicity islands encoding bacterial toxins. In The Comprehensive Sourcebook of Bacterial Protein Toxins, J. Alouf, D. Ladant, and M.R. Popoff, eds. (Amsterdam, Netherlands: Elsevier), pp. 40–76.

Dodd, S.J., Hurst, M.R.H., Glare, T.R., O'Callaghan, M., and Ronson, C.W. (2006). Occurrence of sep insecticidal toxin complex genes in *Serratia* spp. and *Yersinia frederiksenii*. Appl. Environ. Microbiol. *72*, 6584–6592.

Dodt, M., Roehr, J., Ahmed, R., and Dieterich, C. (2012). FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel). *1*, 895–905.

Doublet, B., Boyd, D., Mulvey, M.R., and Cloeckaert, A. (2005). The *Salmonella* genomic island 1 is an integrative mobilizable element. Mol. Microbiol. *55*, 1911–1924.

Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. *16*, 6127–6145.

Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M., Charles, J.F., et al. (2003a). The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nat. Biotechnol. *21*, 1307–1313.

Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M., Charles, J.-F., et al. (2003b). The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nat. Biotechnol. *21*, 1307–1313.

Dunman, P.M., Murphy, E., Haney, S., Palacios, D., Brown, E.L., Zagursky, R.J., Shlaes, D., and Projan, S.J. (2001). Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the *agr* and/or *sarA* loci. J. Bacteriol. *183*, 7341–7353.

Dunphy, G., and Halwani, A. (1997). Haemolymph proteins of larvae of *Galleria mellonella* detoxify endotoxins of the insect pathogenic bacteria *Xenorhabdus nematophilus* (Enterobacteriaceae). J. Insect Physiol. *43*, 1023–1029.

Dunphy, G.B., and Webster, J.M. (1991). Antihemocytic surface components of *Xenorhabdus nematophilus* var. *dutki* and their modification by serum of nonimmune larvae of *Galleria mellonella*. J. Invertebr. Pathol. *58*, 40–51.

Eddy, J.L., Gielda, L.M., Caulfield, A.J., Rangel, S.M., and Lathem, W.W. (2014). Production of outer membrane vesicles by the plague pathogen *Yersinia pestis*. PLoS One *9*, 1–11.

Egan, S., Wiener, P., Kallifidas, D., and Wellington, E.M.H. (2001). Phylogeny of *Streptomyces* species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. Antonie Van Leeuwenhoek *79*, 127–133.

Elbein, A.D., Pan, Y.T., Pastuszak, I., and Carroll, D. (2003). New insights on trehalose:

A multifunctional molecule. Glycobiology *13*, 17–27.

Eleftherianos, I., Boundy, S., Joyce, S. a, Aslam, S., Marshall, J.W., Cox, R.J., Simpson, T.J., Clarke, D.J., Ffrench-Constant, R.H., and Reynolds, S.E. (2007). An antibiotic produced by an insect-pathogenic bacterium suppresses host defenses through phenoloxidase inhibition. Proc. Natl. Acad. Sci. U.S.A. *104*, 2419–2424.

Ellison, D.W., and Miller, V.L. (2006). H-NS represses *inv* transcription in *Yersinia enterocolitica* through competition with RovA and interaction with YmoA. J. Bacteriol. *188*, 5101–5112.

Erhardt, M., and Dersch, P. (2015). Regulatory principles governing *Salmonella* and *Yersinia* virulence. Front. Microbiol. *6*, 1–20.

Erickson, D.L., Waterfield, N.R., Vadyvaloo, V., Long, D., Fischer, E.R., ffrench-constant, R., and Hinnebusch, B.J. (2007). Acute oral toxicity of *Yersinia pseudotuberculosis* to fleas: Implications for the evolution of vector-borne transmission of plague. Cell. Microbiol. *9*, 2658–2666.

Erlandson, M.A., Toprak, U., and Hegedus, D.D. (2019). Role of the peritrophic matrix in insect-pathogen interactions. J. Insect Physiol. *117*, 103894.

Ermolaeva, M.D., Khalak, H.G., White, O., Smith, H.O., and Salzberg, S.L. (2000). Prediction of transcription terminators in bacterial genomes. J. Mol. Biol. *301*, 27–33.

Escoubas, J.-M., Oswald, E., Givaudan, A., Nougayrède, J.-P., Jubelin, G., Banfield, M.J., Zumbihl, R., Nobe, R., Chavez, C.V., Taieb, F., et al. (2009). Cycle inhibiting factors (CIFs) are a growing family of functional cyclomodulins present in invertebrate and mammal bacterial pathogens. PLoS One *4*, e4855.

Eshghi, A., Gaultney, R.A., England, P., Brûlé, S., Miras, I., Sato, H., Coburn, J., Bellalou, J., Moriarty, T.J., Haouz, A., et al. (2019). An extracellular *Leptospira interrogans* leucine-rich repeat protein binds human E- and VE-cadherins. Cell. Microbiol. *21*, 1–14.

Eshwar, A.K., Guldimann, C., Oevermann, A., and Tasara, T. (2017). Cold-shock domain family proteins (Csps) are involved in regulation of virulence, cellular aggregation, and flagella-based motility in *Listeria monocytogenes*. Front. Cell. Infect. Microbiol. *7*, 1–15.

Espinel, I.C., Guerra, P.R., and Jelsbak, L. (2016). Multiple roles of putrescine and spermidine in stress resistance and virulence of *Salmonella enterica* serovar Typhimurium. Microb. Pathog. *95*, 117–123.

Ewing, W.H., Ross, A.J., Brenner, D.J., and Fanning, G.R. (1978). *Yersinia ruckeri* sp. nov., the redmouth (RM) bacterium. Int. J. Syst. Bacteriol. *28*, 37–44.

Fang, F.C., and Rimsky, S. (2008). New insights into transcriptional regulation by H-NS. Curr. Opin. Microbiol. *11*, 113–120.

Fang, F.C., Frawley, E.R., Tapscott, T., and Vázquez-Torres, A. (2016). Bacterial stress responses during host infection. Cell Host Microbe *20*, 133–143.

Farias, J.R., Andow, D.A., Horikoshi, R.J., Sorgatto, R.J., Fresia, P., dos Santos, A.C., and Omoto, C. (2014). Field-evolved resistance to Cry1F maize by *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Brazil. Crop Prot. *64*, 150–158.

Fauvarque, M.-O., Bergeret, E., Chabert, J., Dacheux, D., Satre, M., and Attree, I. (2002). Role and activation of type III secretion system genes in *Pseudomonas aeruginosa*-induced *Drosophila* killing. Microb. Pathog. *32*, 287–295.

Federici, B.A. (2005). Insecticidal bacteria: An overwhelming success for invertebrate pathology. J. Invertebr. Pathol. *89*, 30–38.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. EvolutionEvolution *39*, 783–791.

Fernandez-Prada, C.M., Hoover, D.L., Tall, B.D., Hartman, A.B., Kopelowitz, J., and Venkatesan, M.M. (2000). Shigella flexneri IpaH$_{7.8}$ facilitates escape of virulent bacteria from the endocytic vacuoles of mouse and human macrophages. Infect. Immun. *68*, 3608–3619.

Fernández, L., Marquez, I., and Guijarro, J.A. (2004). Identification of specific in vivo-induced (*ivi*) genes in *Yersinia ruckeri* and analysis of ruckerbactin, a catecholate siderophore iron acquisition system. Appl. Environ. Microbiol. *70*, 5199–5207.

Fernández, L., Prieto, M., and Guijaro, J.A. (2007). The iron- and temperature-regulated haemolysin YhlA is a virulence factor of *Yersinia ruckeri*. Microbiology *153*, 483–489.

ffrench-Constant, Richard Waterfield, N. (2005). An ABC guide to the bacterial toxin complexes. Adv. Appl. Microbiol. *58*, 169–183.

ffrench-constant, R.H., Waterfield, N., Burland, V., Perna, N.T., Daborn, P.J., Bowen, D., and Blattner, F.R. (2000). A genomic sample sequence of the entomopathogenic bacterium *Photorhabdus luminescens* W14: Potential implications for virulence. Appl. Environ. Microbiol. *66*, 3310–3329.

ffrench-Constant, R., Waterfield, N., Daborn, P., Susan, J., Bennett, H., Au, C., Dowling, A., Boundy, S., Reynolds, S., and Clarke, D. (2003). *Photorhabdus*: Towards a functional genomic analysis of a symbiont and pathogen. FEMS Microbiol. Rev. *26*, 433–456.

ffrench-Constant, R.H., Dowling, A., and Waterfield, N.R. (2007). Insecticidal toxins from *Photorhabdus* bacteria and their potential use in agriculture. Toxicon *49*, 436–451.

Fischer, W., Windhager, L., Rohrer, S., Zeiller, M., Karnholz, A., Hoffmann, R., Zimmer, R., and Haas, R. (2010). Strain-specific genes of *Helicobacter pylori*: Genome evolution driven by a novel type IV secretion system and genomic island transfer. Nucleic Acids Res. *38*, 6089–6101.

Flannery, E.L., Mody, L., and Mobley, H.L.T. (2009). Identification of a modular pathogenicity island that is widespread among urease-producing uropathogens and shares features with a diverse group of mobile elements.

Infect. Immun. *77*, 4887–4894.

Foster, J.W. (1999). When protons attack: microbial strategies of acid adaptation. Curr. Opin. Microbiol. *2*, 170–174.

Francino, M.P. (2012). The ecology of bacterial genes and the survival of the new. Int. J. Evol. Biol. *2012*, 1–14.

Franks, A., Mark-Byrne, G.L., Dow, J.M., and O'Gara, F. (2008). A putative RNA-binding protein has a role in virulence in *Ralstonia solanacearum* GMI1000. Mol. Plant Pathol. *9*, 67–72.

Franza, T., Mahé, B., and Expert, D. (2005). *Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection. Mol. Microbiol. *55*, 261–275.

Friedman, R.C., Kalkhof, S., Doppelt-Azeroual, O., Mueller, S.A., Chovancová, M., von Bergen, M., and Schwikowski, B. (2017). Common and phylogenetically widespread coding for peptides by bacterial small RNAs. BMC Genomics *18*, 1–21.

Fröhlich, K.S., and Papenfort, K. (2016). Interplay of regulatory RNAs and mobile genetic elements in enteric pathogens. Mol. Microbiol. *101*, 701–713.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005). Mobile genetic elements: The agents of open source evolution. Nat. Rev. Microbiol. *3*, 722–733.

Fu, Y., Yu, Z., Liu, S., Chen, B., Zhu, L., Li, Z., Chou, S.H., and He, J. (2018). c-di-GMP regulates various phenotypes and insecticidal activity of gram-positive *Bacillus thuringiensis*. Front. Microbiol. *9*, 1–16.

Fuchs, T.M., Bresolin, G., Marcinowski, L., Schachtner, J., and Scherer, S. (2008). Insecticidal genes of *Yersinia* spp.: Taxonomical distribution, contribution to toxicity towards *Manduca sexta* and *Galleria mellonella*, and evolution. BMC Microbiol. *8*, 1–11.

Fukuto, H.S., Svetlanov, A., Palmer, L.E., Karzai, A.W., and Bliska, J.B. (2010). Global gene expression profiling of *Yersinia pestis* replicating inside macrophages reveals the roles of a putative stress-induced operon in regulating type III secretion and intracellular cell division. Infect. Immun. *78*, 3700–3715.

Futschik, Matthias, E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. *3*, 965–988.

Gaillard, M., Vallaeys, T., Jörg, F., Minoia, M., Werlen, C., Sentchilo, V., Pühler, A., and Roelof van der Meer, J. (2006). The *clc* element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. J. Bacteriol. *188*, 1999–2013.

Galán, J.E., Lara-tejero, M., Marlovits, T.C., and Wagner, S. (2014). Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. Annu. Rev. Microbiol. *68*, 415–438.

Galperin, M.Y. (2006). Structural classification of bacterial response regulators:

Diversity of output domains and domain combinations. J. Bacteriol. *188*, 4169–4182.

Galperin, M.Y. (2008). Telling Bacteria: Do Not LytTR. Structure *16*, 657–659.

Galperin, M.Y., Gomelsky, M., and Galperin, M.Y. (2005). Bacterial signal transduction modules: from genomics to biology. ASM News *71*, 326–333.

Garcia-gonzalez, E., Poppinga, L., Fünfhaus, A., Hertlein, G., Hedtke, K., Jakubowska, A., and Genersch, E. (2014). *Paenibacillus larvae* chitin-degrading protein *Pl*CBP49 is a key virulence factor in American foulbrood of honey bees. PLoS Pathog. *10*, e1004284.

Garcia-Gonzalez, E., Müller, S., Hertlein, G., Heid, N., Süssmuth, R.D., and Genersch, E. (2014a). Biological effects of paenilamicin, a secondary metabolite antibiotic produced by the honey bee pathogenic bacterium *Paenibacillus larvae*. Microbiologyopen *3*, 642–656.

Garcia-Gonzalez, E., Müller, S., Ensle, P., Süssmuth, R.D., and Genersch, E. (2014b). Elucidation of sevadicin, a novel non-ribosomal peptide secondary metabolite produced by the honey bee pathogenic bacterium *Paenibacillus larvae*. Environ. Microbiol. *16*, 1297–1309.

Garcia, Y.M., Barwinska-Sendra, A., Tarrant, E., Skaar, E.P., Waldron, K.J., and Kehl-Fie, T.E. (2017). A superoxide dismutase capable of functioning with iron or manganese promotes the resistance of *Staphylococcus aureus* to calprotectin and nutritional immunity. PLoS Pathog. *13*, 1–19.

García, J., Ventura, María, I., Requena, M., Antonio, H.F., Parrón, T., and Alarcón, R. (2017). Association of reproductive disorders and male congenital anomalies with environmental exposure to endocrine active pesticides. Reprod. Toxicol. *71*, 95–100.

Gassmann, A.J. (2016). Resistance to Bt maize by western corn rootworm: Insights from the laboratory and the field. Curr. Opin. Insect Sci. *15*, 111–115.

Gatsogiannis, C., Lang, A.E., Meusch, D., Pfaumann, V., Hofnagel, O., Benz, R., Aktories, K., and Raunser, S. (2013). A syringe-like injection mechanism in *Photorhabdus luminescens* toxins. Nature *495*, 520–523.

Gatsogiannis, C., Merino, F., Prumbaum, D., Roderer, D., Leidreiter, F., Meusch, D., and Raunser, S. (2016). Membrane insertion of a Tc toxin in near-atomic detail. Nat. Struct. Mol. Biol. *23*, 884–890.

Gebhardt, M.J., Gallagher, L.A., Jacobson, R.K., Usacheva, E.A., Peterson, L.R., Zurawski, D. V, and Shuman, H.A. (2015). Joint transcriptional control of virulence and resistance to antibiotic and environmental stress in *Acinetobacter baumannii*. MBio *6*, 1–12.

Georg, J., and Hess, W.R. (2011). *cis*-antisense RNA, another level of gene regulation in bacteria. Microbiol. Mol. Biol. Rev. *75*, 286–300.

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

401–415.

Gillis, A., Fayad, N., Makart, L., Bolotin, A., Sorokin, A., Kallassy, M., and Mahillon, J. (2018). Role of plasmid plasticity and mobile genetic elements in the entomopathogen *Bacillus thuringiensis* serovar israelensis. FEMS Microbiol. Rev. *42*, 829–856.

Givaudan, A., and Lanois, A. (2000). *flhDC*, the flagellar master operon of *Xenorhabdus nematophilus*: Requirement for motility, lipolysis, extracellular hemolysis, and full virulence in insects. J. Bacteriol. *182*, 107–115.

Givaudan, A., and Lanois, A. (2017). Flagellar regulation and virulence in the entomopathogenic bacteria-*Xenorhabdus nematophila* and *Photorhabdus luminescens*. In The Molecular Biology of Photorhabus Bacteria, R.H. ffrenche-Constant, ed. (Cham: Springer), pp. 39–51.

Glare, T., Caradus, J., Gelernter, W., Jackson, T., Keyhani, N., Köhl, J., Marrone, P., Morin, L., and Stewart, A. (2012). Have biopesticides come of age? Trends Biotechnol. *30*, 250–258.

Glöckner, G., Albert-Weissenberger, C., Weinmann, E., Jacobi, S., Schunder, E., Steinert, M., Hacker, J., and Heuner, K. (2008). Identification and characterization of a new conjugation/type IVA secretion system (*trb*/*tra*) of *Legionella pneumophila* Corby localized on two mobile genomic islands. Int. J. Med. Microbiol. *298*, 411–428.

Gonçalves, C., Decré, D., Barbut, F., Burghoffer, B., and Petit, J.-C. (2004). Prevalence and characterization of a binary toxin (actin-specific ADP-ribosyltransferase) from *Clostridium difficile*. J. Clin. Microbiol. *42*, 1933–1939.

Gonyar, L.A., and Kendall, M.M. (2014). Ethanolamine and choline promote expression of putative and characterized fimbriae in enterohemorrhagic *Escherichia coli* O157: H7. Infect. Immun. *82*, 193–201.

Gonzalez-Escalona, N., Gavilan, R.G., Toro, M., Zamudio, M.L., and Martinez-Urtaza, J. (2016). Outbreak of *Vibrio parahaemolyticus* sequence type 120, Peru, 2009. Emerg. Infect. Dis. *22*, 1235–1237.

González-Juarbe, N., Mares, C.A., Hinojosa, C.A., Medina, J.L., Cantwell, A., Dube, P.H., Orihuela, C.J., and Bergman, M.A. (2015). Requirement for *Serratia marcescens* cytolysin in a murine model of hemorrhagic pneumonia. Infect. Immun. *83*, 614–624.

Gooday, G.W. (1990). The ecology of chitin degradation. In Advances in Microbial Ecology, K.C. Marshall, ed. (AMIE), pp. 387–430.

Gorrini, C., Harris, I.S., and Mak, T.W. (2013). Modulation of oxidative stress as an anticancer strategy. Nat. Rev. Drug Discov. *12*, 931–947.

Grant, C.M. (2008). Metabolic reconfiguration is a regulated response to oxidative stress. J. Biol. *7*, 6–9.

Grant, T., Bennett-Wood, V., and Robins-Browne, R.M. (1998). Identification of virulence-associated characteristics in clinical isolates of *Yersinia enterocolitica*

lacking classical virulence markers. Infect. Immun. *66*, 1113–1120.

Grenier, A., Duport, G., Page, S., Condemine, G., and Rahbe, Y. (2006). The phytopathogen *Dickeya dadantii* (*Erwinia chrysanthemi* 3937) is a pathogen of the pea aphid. Appl. Environ. Microbiol. *72*, 1956–1965.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36*, 70–74.

Guédon, G., Libante, V., Coluzzi, C., Payot, S., and Leblond-bourget, N. (2017). The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. Genes *8*, 1–28.

Le Guern, A.S., Savin, C., Brémont, S., Payro, G., Bon, D., Carniel, E., and Pizarro-Cerdá, J. (2018). First isolation of *Yersinia entomophaga* in human urinary tract. New Microbes New Infect. *26*, 3–7.

Guerra, L., Cortes-Bratti, X., Guidi, R., and Frisan, T. (2011). The biology of the cytolethal distending toxins. Toxins *3*, 172–190.

Gulati, M., Jain, N., Davis, J.H., Williamson, J.R., and Britton, R.A. (2014). Functional interaction between ribosomal protein L6 and RbgA during ribosome assembly. PLoS Genet. *10*.

Gupta, S., Igoillo-Esteve, M., Michels, P.A.M., and Cordeiro, A.T. (2011). Glucose-6-phosphate dehydrogenase of Trypanosomatids: Characterization, target validation, and drug discovery. Mol. Biol. Int. *2011*, 1–10.

Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. J. Bacteriol. *177*, 4121–4130.

Haapalainen, M., Mosorin, H., Dorati, F., Wu, R.F., Roine, E., Taira, S., Nissinen, R., Mattinen, L., Jackson, R., Pirhonen, M., et al. (2012). Hcp2, a secreted protein of the phytopathogen *Pseudomonas syringae* pv. tomato DC3000, is required for fitness for competition against bacteria and yeasts. J. Bacteriol. *194*, 4810–4822.

Haas, D., and Défago, G. (2005). Biological control of soil-borne pathogens by fluorescent pseudomonads. Nat. Rev. Microbiol. *3*, 307–319.

Haas, B.J., Chin, M., Nusbaum, C., Birren, B.W., and Livny, J. (2012). How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC Genomics *13*, 1.

Hacker, J., and Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity: A Darwinian view of the evolution of microbes. EMBO Rep. *2*, 376–381.

Hacker, J., Blum-Oehler, G., Mühldorfer, I., and Tschäpe, H. (1997). Pathogenicity islands of virulent bacteria, structure, function and impact on microbial evolution. Mol. Microbiol. *23*, 1089–1097.

Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and Bioconductor. In Statistical Genomics: Methods in Molecular Biology, M. Ewy,

and S. Davis, eds. (New York, NY: Springer Nature), pp. 444–455.

Halwani, A.E., Niven, D.F., and Dunphy, G.B. (2000). Apolipophorin-III and the interactions of lipoteichoic acids with the immediate immune responses of *Galleria mellonella*. J. Invertebr. Pathol. *76*, 233–241.

Hamilton, J.J., Marlow, V.L., Owen, R.A., Costa, M. de A.A., Guo, M., Buchanan, G., Chandra, G., Trost, M., Coulthurst, S.J., Palmer, T., et al. (2014). A holin and an endopeptidase are essential for chitinolytic protein secretion in *Serratia marcescens*. J. Cell Biol. *207*, 615–626.

Han, S., Craig, J.A., Putnam, C.D., Carozzi, N.B., and Tainer, J.A. (1999). Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. Nat. Struct. Biol. *6*, 932–936.

Han, Y., Zhou, D., Pang, X., Song, Y., Zhang, L., Bao, J., Tong, Z., Wang, J., Guo, Z., Zhai, J., et al. (2004). Microarray analysis of temperature-induced transcriptome of *Yersinia pestis*. Microbiol. Immunol. *48*, 791–805.

Hanin, A., Sava, I., Bao, Y., Huebner, J., Hartke, A., Auffray, Y., and Sauvageot, N. (2010). Screening of *in vivo* activated genes in *Enterococcus faecalis* during insect and mouse infections and growth in urine. PLoS One *5*, e11879.

Hanson, R.M., Prilusky, J., Renjian, Z., Nakane, T., and Sussman, J.L. (2013). JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia. Isr. J. Chem. *53*, 207–216.

Hanzal, R., and Jegorov, A. (1991). Changes in free amino acid composition in haemolymph of larvae of the wax moth, *Galleria mellonella* L., during cold acclimation. Comp. Biochem. Physiol. - A Physiol. *100*, 957–962.

Hare, J.M., Wagner, A.K., and McDonough, K.A. (1999). Independent acquisition and insertion into different chromosomal locations of the same pathogenicity island in *Yersinia pestis* and *Yersinia pseudotuberculosis*. Mol. Microbiol. *31*, 291–303.

Hares, M.C., Hinchliffe, S.J., Strong, P.C.R., Eleftherianos, I., Dowling, A.J., Ffrench-Constant, R.H., and Waterfield, N. (2008). The *Yersinia pseudotuberculosis* and *Yersinia pestis* toxin complex is active against cultured mammalian cells. Microbiology *154*, 3503–3517.

He, M., Ouyang, Z., Troxell, B., Xu, H., Moh, A., Piesman, J., Norgard, M. V., Gomelsky, M., and Yang, X.F. (2011). Cyclic di-gmp is essential for the survival of the lyme disease spirochete in ticks. PLoS Pathog. *7*.

Heermann, R., and Fuchs, T.M. (2008). Comparative analysis of the *Photorhabdus luminescens* and the *Yersinia enterocolitica* genomes: uncovering candidate genes involved in insect pathogenicity. BMC Genomics *9*, 40.

Held, K.G., LaRock, C.N., D'Argenio, D.A., Berg, C.A., and Collins, C.M. (2007). A metalloprotease secreted by the insect pathogen *Photorhabdus luminescens* induces melanization. Appl. Environ. Microbiol. *73*, 7622–7628.

Henderson, M.W., Inatsuka, C.S., Sheets, A.J., Williams, C.L., Benaron, D.J., Donato,

G.M., Gray, M.C., Hewlett, E.L., and Cotter, P.A. (2012). Contribution of *Bordetella filamentous* hemagglutinin and adenylate cyclase toxin to suppression and evasion of interleukin-17-mediated inflammation. Infect. Immun. *80*, 2061–2075.

Herbst, K., Bujara, M., Heroven, A.K., Opitz, W., Weichert, M., Zimmermann, A., and Dersch, P. (2009). Intrinsic thermal sensing controls proteolysis of *Yersinia* virulence regulator RovA. PLoS Pathog. *5*.

Hernández, J.A., Muro-Pastor, A.M., Flores, E., Bes, M.T., Peleato, M.L., and Fillat, M.F. (2006). Identification of a *furA cis* antisense RNA in the cyanobacterium *Anabaena* sp. PCC 7120. J. Mol. Biol. *355*, 325–334.

Heroven, A.K., and Dersch, P. (2014). Coregulation of host-adapted metabolism and virulence by pathogenic *yersiniae*. Front. Cell. Infect. Microbiol. *4*, 1–13.

Heroven, A.K., Nagel, G., Tran, H.J., Parr, S., and Dersch, P. (2004). RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and rovA expression in *Yersinia pseudotuberculosis*. Mol. Microbiol. *53*, 871–888.

Heroven, A.K., Dersch, P., Tran-Winkler, H., and Böhme, K. (2007). Regulatory elements implicated in the environmental control of invasin expression in enteropathogenic *Yersinia*. In The Genus *Yersinia*; Advances in Experimental Medicine and Biology, R.D. Perry, and J.D. Fetherston, eds. (New York, NY: Springer), pp. 156–166.

Heroven, A.K., Sest, M., Pisano, F., Scheb-Wetzel, M., Steinmann, R., Böhme, K., Klein, J., Münch, R., Schomburg, D., and Dersch, P. (2012). Crp induces switching of the CsrB and CsrC RNAs in *Yersinia pseudotuberculosis* and links nutritional status to virulence. Front. Cell. Infect. Microbiol. *2*, 158.

Hess, W.R., Scholz, I., Mitschke, J., Voß, B., Georg, J., and Wilde, A. (2009). Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. Mol. Syst. Biol. *5*.

Heywood, W., Henderson, B., and Nair, S.P. (2005). Cytolethal distending toxin: Creating a gap in the cell cycle. J. Med. Microbiol. *54*, 207–216.

Hill, C. (2012). Virulence or niche factors: What's in a name? J. Bacteriol. *194*, 5725–5727.

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.

Hochhut, B., Wilde, C., Balling, G., Middendorf, B., Dobrindt, U., Brzuszkiewicz, E., Gottschalk, G., Carniel, E., and Hacker, J. (2006). Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Echerichia coli* strain 536. Mol. Microbiol. *61*, 584–595.

Hoe, N.P., and Goguen, J.D. (1993). Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. J. Bacteriol. *175*, 7901–7909.

Van Der Hoeven, R., and Forst, S. (2009). OpnS, an outer membrane porin of

Xenorhabdus nematophila, confers a competitive advantage for growth in the insect host. J. Bacteriol. *191*, 5471–5479.

Hood, R.D., Singh, P., Hsu, F.S., Güvener, T., Carl, M.A., Trinidad, R.R.S., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Plemel, R.L., et al. (2010). A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. Cell Host Microbe *7*, 25–37.

Horazdovsky, B.F., and Hogg, R.W. (1987). High-affinity L-arabinose transport operon gene product expression and mRNAs. J. Mol. Biol. *197*, 27–35.

Horne, S.M., and Prüss, B.M. (2006). Global gene regulation in *Yersinia enterocolitica*: effect of FliA on the expression levels of flagellar and plasmid-encoded virulence genes. Arch. Microbiol. *185*, 115–126.

Hsieh, T.C., Ma, K.H., and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. *7*, 1451–1456.

Hu, X., Nesic, D., and Stebbins, C.E. (2006). Comparative structure-function analysis of cytolethal distending toxins. Proteins Struct. Funct. Genet. *62*, 421–434.

Humphrys, M.S., Creasy, T., Sun, Y., Shetty, A.C., Chibucos, M.C., Drabek, E.F., Fraser, C.M., Farooq, U., Sengamalay, N., Ott, S., et al. (2013). Simultaneous transcriptional profiling of bacteria and their host cells. PLoS One *8*.

Hurst, M.R.H., Glare, T.R., Jackson, T.A., and Ronson, C.W. (2000). Plasmid-located pathogenicity determinants of *Serratia entomophila*, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of *Photorhabdus luminescens*. J. Bacteriol. *182*, 5127–5138.

Hurst, M.R.H., Jones, S.A., Binglin, T., Harper, L.A., Jackson, T.A., and Glare, T.R. (2011a). The main virulence determinant of *Yersinia entomophaga* MH96 is a broad-host-range toxin complex active against insects. J. Bacteriol. *193*, 1966–1980.

Hurst, M.R.H., Becher, S.A., Young, S.D., Nelson, T.L., and Glare, T.R. (2011b). *Yersinia entomophaga* sp. nov., isolated from the New Zealand grass grub *Costelytra zealandica*. Int. J. Syst. Evol. Microbiol. *61*, 844–849.

Hurst, M.R.H., van Koten, C., and Jackson, T.A. (2014). Pathology of *Yersinia entomophaga* MH96 towards *Costelytra zealandica* (Coleoptera; Scarabaeidae) larvae. J. Invertebr. Pathol. *115*, 102–107.

Hurst, M.R.H., Beattie, A.K., Jones, S.A., Hsu, P.-C., Calder, J., and van Koten, C. (2015). Temperature-dependent *Galleria mellonella* mortality as a result of *Yersinia entomophaga* infection. Appl. Environ. Microbiol. *81*, 6404–6414.

Hurst, M.R.H., Beattie, A., Altermann, E., Moraga, R.M., Harper, L.A., Calder, J., and Laugraud, A. (2016). The draft genome sequence of the *Yersinia entomophaga* entomopathogenic type strain MH96T. Toxins *8*, 1–19.

Hurst, M.R.H., Jones, S.A., Beattie, A., van Koten, C., Shelton, A.M., Collins, H.L., and Brownbridge, M. (2019). Assessment of *Yersinia entomophaga* as a control agent

of the diamondback moth *Plutella xylostella*. J. Invertebr. Pathol. *162*, 19–25.

Hussain, T., Llácer, J.L., Wimberly, B.T., Kieft, J.S., and Ramakrishnan, V. (2016). Large-scale movements of IF3 and tRNA during bacterial translation initiation. Cell *167*, 133–144.e13.

Hüttener, M., Paytubi, S., and Juárez, A. (2015). Success in incorporating horizontally transferred genes: The H-NS protein. Trends Microbiol. *23*, 67.

Ibanez, F., Levy, J., and Tamborindeguy, C. (2014). Transcriptome analysis of "*Candidatus* Liberibacter solanacearum" in its psyllid vector, *Bactericera cockerelli*. PLoS One *9*, 1–12.

Iglauer, F., Lange, A., Parusel, R., Frick, J.-S., Beier, S., and Huson, D.H. (2018). Genome sequence of *Galleria mellonella* (Greater Wax Moth). Genome Announc. *6*, 1–2.

Igoillo-Esteve, M., and Cazzulo, J.J. (2006). The glucose-6-phosphate dehydrogenase from Trypanosoma cruzi: Its role in the defense of the parasite against oxidative stress. Mol. Biochem. Parasitol. *149*, 170–181.

Imori, P.F.M., Passaglia, J., Souza, R.A., Rocha, L.B., and Falcão, J.P. (2017). Virulence-related genes, adhesion and invasion of some *Yersinia enterocolitica*-like strains suggests its pathogenic potential. Microb. Pathog. *104*, 72–77.

Inatsuka, C.S., Julio, S.M., and Cotter, P.A. (2005). *Bordetella* filamentous hemagglutinin plays a critical role in immunomodulation, suggesting a mechanism for host specificity. Proc. Natl. Acad. Sci. *102*, 18578–18583.

Infantino, V., Lacobazzi, Vi., Menga, A., Avantaggiati, M.L., and Lmieri, F. (2014). A key role of the mitochondrial citrate carrier (*SLC25A1*) in TNFalpha- and INFgamma- triggered inflamation. Biochim. Biophys. Acta *1839*, 1217–1225.

Ireton, K. (2007). Entry of the bacterial pathogen *Listeria monocytogenes* into mammalian cells. Cell. Microbiol. *9*, 1365–1375.

Ishii, K., Adachi, T., Hamamoto, H., and Sekimizu, K. (2014). *Serratia marcescens* suppresses host cellular immunity via the production of an adhesion-inhibitory factor against immunosurveillance cells. J. Biol. Chem. *289*, 5876–5888.

Israel, J.W., Martik, M.L., Byrne, M., Raff, E.C., Raff, R.A., McClay, D.R., and Wray, G.A. (2016). Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea urchin genus *Heliocidaris*. PLoS Biol. *14*, 1–28.

Iwobi, A., Heesemann, J., Garcia, E., Igwe, E., Noelting, C., and Rakin, A. (2003). Novel virulence-associated type II secretion system unique to high-pathogenicity *Yersinia enterocolitica*. Infect. Immun. *71*, 1872–1879.

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*, 1–16.

Jacob, A.I., Köhrer, C., Davies, B.W., RajBhandary, U.L., and Walker, G.C. (2013).

Conserved bacterial RNase YbeY plays key roles in 70S ribosome quality control and 16S rRNA maturation. Mol. Cell *49*, 427–438.

Jayde, A., Laurent, A., Claire, M., Paul, D., John, M., Jost, E., and Andrew, J. (2016). Visualizing translocation and localization of bacterial type III effector proteins using a genetically encoded reporter system. Appl. Environ. Microbiol. *82*, 0.

Jiang, M., Datta, K., Walker, A., Strahler, J., Bagamasbad, P., Andrews, P.C., and Maddock, J.R. (2006). The *Escherichia coli* GTPase CgtAE is involved in late steps of large ribosome assembly. J. Bacteriol. *188*, 6757–6770.

Jiang, M., Sullivan, S.M., Walker, A.K., Strahler, J.R., Andrews, P.C., and Maddock, J.R. (2007). Identification of novel *Escherichia coli* ribosome-associated proteins using isobaric tags and multidimensional protein identification techniques. J. Bacteriol. *189*, 3434–3444.

Jimenez, P.N., Koch, G., Thompson, J.A., Xavier, K.B., Cool, R.H., and Quax, W.J. (2012). The multiple signaling systems regulating virulence in *Pseudomonas aeruginosa*. Microbiol. Mol. Biol. Rev. *76*, 46–65.

Johnson, C.M., and Grossman, A.D. (2015). Integrative and conjugative elements (ICEs): What they do an how they work. Annu. Rev. Genet. *49*, 577–601.

Jorth, P., Trivedi, U., Rumbaugh, K., and Whiteley, M. (2013). Probing bacterial metabolism during infection using high-resolution transcriptomics. J. Bacteriol. *195*, 4991–4998.

Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol. Rev. *33*, 376–393.

Julio, S.M., Inatsuka, C.S., Mazar, J., Dieterich, C., David, A., and Cotter, P.A. (2009). Natural-host animal models indicate functional interchagneability between the filamentous haemagglutinins of *Bordetlla pertussis* and *Bordetella bronchiseptica* and reveal a role for the mature C-terminal domain, but not the RGD motif, during . Mol. Microbiol. *71*, 1574–1590.

Junqueira, J.C. (2012). *Galleria mellonella* as a model host for human pathogens: recent studies and new perspectives. Virulence *3*, 474–476.

Jurat-Fuentes, J.L., and Jackson, T.A. (2012). Bacterial entomopathogens. In Insect Pathology, F.E. VEga, and H.K. Kaya, eds. (San Diego: Elsevier), pp. 265–349.

Kakoschke, T., Kakoschke, S., Magistro, G., Schubert, S., Borath, M., Heesemann, J., and Rossier, O. (2014). The RNA chaperone Hfq impacts growth, metabolism and production of virulence factors in *Yersinia enterocolitica*. PLoS One *9*.

Kalia, D., Merey, G., Nakayama, S., Zheng, Y., Zhou, J., Luo, Y., Guo, M., Roembke, B.T., and Sintim, H.O. (2013). Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis. Chem. Soc. Rev. *42*, 305–341.

Kalogeraki, V.S., and Winans, S.C. (1997). Suicide plasmids containing promoterless reporter genes can simultaneously disrupt and create fusions to target genes

of diverse bacteria. Gene *188*, 69–75.

Kalscheuer, R., Weinrick, B., Veeraraghavan, U., Besra, G.S., and Jacobs, W.R. (2010). Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U.S.A. *107*, 21761–21766.

Kamal, N.D.M., Jalil, N., Member, S., and Hashim, H. (2011). The analysis of shape-based, DWT and zernike moments feature extraction techniques for fasterner recognition using 10-fold cross validation multilayer perceptrons. *2011*.

Kang, S.M., Choi, J.W., Lee, Y., Hong, S.H., and Lee, H.J. (2013). Identification of microRNA-size, small RNAs in Escherichia coli. Curr. Microbiol. *67*, 609–613.

Kanonenberg, K., Schwarz, C.K.W., and Schmitt, L. (2013). Type I secretion systems - a story of appendices. Res. Microbiol. *164*, 596–604.

Kapatral, V., Olson, J.W., Pepe, J.C., Miller, V.L., and Minnich, S.A. (1996). Temperature-dependent regulation of *Yersinia enterocolitica* class III flagellar genes. Mol. Microbiol. *19*, 1061–1071.

Kapatral, V., Campbell, J.W., Minnich, S.A., Thomson, N.R., Matsumura, P., and Prüß, B.M. (2004). Gene arrays analysis of *Yersinia enterocolitica* FlhD and FlhC: Regulation of enzymes affecting synthesis and degradation of carbamoylphosphate. Microbiology *150*, 2289–2300.

Kaper, J.B., and Hacker, J. (2000). Pathogenecity islands and the evolution of microbes. Annu. Rev. Microbiol. *54*, 641–679.

Karavolos, M.H., Winzer, K., Williams, P., and Khan, C.M.A. (2013). Pathogen espionage: Multiple bacterial adrenergic sensors eavesdrop on host communication systems. Mol. Microbiol. *87*, 455–465.

Karch, H., Schubert, S., Zhang, D., Zhang, W., Schmidt, H., Olschläger, T., and Hacker, J. (1999). A genomic island, termed high-pathogenicity island, is present in certain non-O157 shiga toxin-producing *Escherichia coli* clonal lineages. Infect. Immun. *67*, 5994–6001.

Katayama, Y., Ito, T., and Hiramatsu, K. (2000). A new class of genetic element, *Staphylococcus* cassette chromosome mec, encodes methicillin resistance in *Staphylococcus aureus*. Antimicrob. Agents Chemother. *44*, 1549–1555.

Keasey, S.L., Natesan, M., Pugh, C., Kamata, T., Wuchty, S., and Ulrich, R.G. (2016). Cell-free determination of binary complexes that comprise extended protein-protein interaction networks of *Yersinia pestis*. Mol. Cell. Proteomics *15*, 3220–3232.

Kelkenberg, M., Odman-Naresh, J., Muthukrishnan, S., and Merzendorfer, H. (2015). Chitin is a necessary component to maintain the barrier function of the peritrophic matrix in the insect midgut. Insect Biochem. Mol. Biol. *56*, 21–28.

Kelley, Lawrence, A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, Michael, J.E. (2015). The Phyre2 web portal for protein modeling prediction and analysis. Nat. Protoc. *10*, 845–858.

Kendall, M.M., and Sperandio, V. (2016). What a dinner party! mechanisms and functions of interkingdom signaling in host-pathogen associations. MBio *7*, 1–14.

Kendall, M.M., Gruber, C.C., Parker, C.T., and Sperandio, V. (2012). Ethanolamine controls expression of genes encoding components involved in interkingdom signaling and virulence in enterohemorrhagic *Escherichia coli* O157:H7. MBio *3*, 1–10.

Kharade, S.S., and McBride, M.J. (2014). *Flavobacterium johnsoniae* chitinase ChiA is required for chitin utilization and is secreted by the type IX secretion system. J. Bacteriol. *196*, 961–970.

Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., and Tago, K. (2012). Symbiont-mediated insecticide resistance. Proc. Natl. Acad. Sci. United States Am. *109*, 8618–8622.

Killiny, N. (2018). Generous hosts: Why the larvae of greater wax moth, *Galleria mellonella* is a perfect infectious host model? Virulence *9*, 860–865.

Kim, B.T., Tsuchida, K., Lincecum, J., Kitagawa, H., Bernfield, M., and Sugahara, K. (2003a). Identification and characterization of three *Drosophila melanogaster* glucuronyltransferases responsible for the synthesis of the conserved glycosaminoglycan-protein linkage region of proteoglycans. J. Biol. Chem. *278*, 9116–9124.

Kim, C., Song, S., and Park, C. (1997). The D-allose operon of *Escherichia coli* K-12. J. Bacteriol. *179*, 7631–7637.

Kim, I.H., Aryal, S.K., Aghai, D.T., Casanova-Torres, Á.M., Hillman, K., Kozuch, M.P., Mans, E.J., Mauer, T.J., Ogier, J.C., Ensign, J.C., et al. (2017). The insect pathogenic bacterium *Xenorhabdus innexi* has attenuated virulence in multiple insect model hosts yet encodes a potent mosquitocidal toxin. BMC Genomics *18*, 1–25.

Kim, M.S., Shin, J., Lee, W., Lee, H.S., and Oh, B.H. (2003b). Crystal structures of RbsD leading to the identification of cytoplasmic sugar-binding proteins with a novel folding architecture. J. Biol. Chem. *278*, 28173–28180.

Klink, Björn, U., Barden, S., Heidler, T. V., Borchers, C., Ladwein, M., Stradal, Theresia, E.B., Rottner, K., and Heinz, D.W. (2010). Structure of *Shigella* IpgB2 in complex with human RhoA. J. Biol. Chem. *285*, 17197–17208.

Kobayashi, K., Fujikawa, M., and Kozawa, T. (2014). Oxidative stress sensing by the iron-sulfur cluster in the transcription factor, SoxR. J. Inorg. Biochem. *133*, 87–91.

Koczula, A., Jarek, M., Visscher, C., Valentin-Weigand, P., Goethe, R., and Willenborg, J. (2017). Transcriptomic analysis reveals selective metabolic adaptation of *Streptococcus suis* to porcine blood and cerebrospinal fluid. Pathogens *6*, 7.

Koeppen, K., Hampton, T.H., Jarek, M., Scharfe, M., Gerber, S.A., Mielcarz, D.W., Demers, E.G., Dolben, E.L., Hammond, J.H., Hogan, D.A., et al. (2016). A

novel mechanism of host-pathogen interaction through sRNA in bacterial outer membrane vesicles. PLoS Pathog. *12*, 1–22.

Kohanski, M.A., Dwyer, D.J., Hayete, B., Lawrence, C.A., and Collins, J.J. (2007). A common mechanism of cellular death induced by bactericidal antibiotics. Cell *130*, 797–810.

Kohanski, M.A., Dwyer, D.J., and Collins, J.J. (2010). How antibiotics kill bacteria: From targets to networks. Nat. Rev. Microbiol. *8*, 423–435.

Konkel, M.E., and Tilly, K. (2000). Temperature-regulated expression of bacterial virulence genes. Microbes Infect. *2*, 157–166.

Kornberg, H.L. (2001). Routes for fructose utilization by *Escherichia coli*. J. Mol. Microbiol. Biotechnol. *3*, 355–359.

Korves, T., and Colosimo, M.E. (2009). Controlled vocabularies for microbial virulence factors. Trends Microbiol. *17*, 279–285.

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

Kovářová, J., and Barrett, M.P. (2016). The pentose phosphate pathway in parasitic Trypanosomatids. Trends Parasitol. *32*, 622–634.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Kuehl, C.J., Dragoi, A.M., and Agaisse, H. (2014). The *Shigella flexneri* type 3 secretion system is required for tyrosine kinase-dependent protrusion resolution, and vacuole escape during bacterial dissemination. PLoS One *9*.

Kumar, L., and Futschik, M. (2007). Mfuzz: A software package for soft clustering of microarray data. Bioinformation *2*, 5–7.

Kumar, N., Creasy, T., Sun, Y., Flowers, M., Tallon, L.J., and Dunning Hotopp, J.C. (2012). Efficient subtraction of insect rRNA prior to transcriptome analysis of *Wolbachia-Drosophila* lateral gene transfer. BMC Res. Notes *5*, 1.

Kumar, N., Lin, M., Zhao, X., Ott, S., Santana-Cruz, I., Daugherty, S., Rikihisa, Y., Sadzewicz, L., Tallon, L.J., Fraser, C.M., et al. (2016). Efficient enrichment of bacterial mRNA from host-bacteria total RNA samples. Sci. Rep. *6*, 1–10.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *35*, 1547–1549.

Kurz, C.L., Chauvet, S., Andrès, E., Aurouze, M., Vallet, I., Michel, G.P.F., Uh, M., Celli, J., Filloux, A., De Bentzmann, S., et al. (2003). Virulence factors of the human opportunistic pathogen *Serratia marcescens* identified by *in vivo* screening. EMBO J. *22*, 1451–1460.

Kwaik, Y.A., and Bumann, D. (2013). Microbial quest for food in vivo: "Nutritional virulence" as an emerging paradigm. Cell. Microbiol. *15*, 882–890.

Lacey, L.A., Grzywacz, D., Shapiro-Ilan, D.I., Frutos, R., Brownbridge, M., and Goettel, M.S. (2015). Insect pathogens as biological control agents: Back to the future. J. Invertebr. Pathol. *132*, 1–41.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

Lai, Y.C., Peng, H.L., and Chang, H.Y. (2001). Identification of genes induced *in vivo* during *Klebsiella pneumoniae* CG43 infection. Infect. Immun. *69*, 7140–7145.

Laikova, O.N., Mironov, A.A., and Gelfand, M.S. (2001). Computational analysis of the transcriptional regulation of pentose utilization systems in the gamma subdivision of Proteobacteria. FEMS Microbiol. Lett. *205*, 315–322.

Lam, H.M., and Winkler, M.E. (1990). Metabolic relationships between pyridoxine (vitamin B_6) and serine biosynthesis in *Escherichia coli* K-12. J. Bacteriol. *172*, 6518–6528.

Lam, O., Wheeler, J., and Tang, C.M. (2014). Thermal control of virulence factors in bacteria: A hot topic. Virulence *5*, 852–862.

Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., et al. (2015). Insights from 20 years of bacterial genome sequencing. Funct. Integr. Genomics *15*, 141–161.

Landick, R., Wade, J.T., and Grainger, D.C. (2015). H-NS and RNA polymerase: A love-hate relationship? Curr. Opin. Microbiol. *24*, 53–59.

Landsberg, M.J., Jones, S.A., Rothnagel, R., Busby, J.N., Marshall, S.D.G., Simpson, R.M., Lott, J.S., Hankamer, B., and Hurst, M.R.H. (2011). 3D structure of the *Yersinia entomophaga* toxin complex and implications for insecticidal activity. Proc. Natl. Acad. Sci. *108*, 20544–20549.

Lang, A.E., Schmidt, G., Schlosser, A., Hey, T.D., Larrinua, I.M., Sheets, J.J., Mannherz, H.G., and Aktories, K. (2010). *Photorhabdus luminescens* toxins ADP-ribosylate actin and rhoA to force actin clustering. Science *327*, 1139–1143.

Lanois, A., Jubelin, G., and Givaudan, A. (2008). FliZ, a flagellar regulator, is at the crossroads between motility, haemolysin expression and virulence in the insect pathogenic bacterium *Xenorhabdus*. Mol. Microbiol. *68*, 516–533.

Lara-Tejero, M., and Galán, J.E. (2000). Cell cycle progression as a deoxyribonuclease I–like protein. Science *290*, 354–357.

Larbig, K.D., Christmann, A., Klockgether, J., Hartsch, T., Merkl, R., Wiehlmann, L., Fritz, H., and Tümmler, B. (2002). Gene islands integrated into $tRNA^{Gly}$ genes confer genome diversity on a *Pseudomonas aeruginosa* clone. J. Bacteriol. *184*, 6665–6680.

Lathem, W.W., Grys, T.E., Witowski, S.E., Torres, A.G., Kaper, J.B., Tarr, P.I., and Welch, R.A. (2002). StcE, a metalloprotease secreted by *Escherichia coli* O157:H7, specifically cleaves C1 esterase inhibitor. Mol. Microbiol. *45*, 277–288.

Lathem, W.W., Schroeder, J.A., Bellows, L.E., Ritzert, J.T., Koo, J.T., Price, P.A., Caulfield, A.J., and Goldman, W.E. (2014). Posttranscriptional regulation of the *Yersinia pestis* cyclic AMP receptor protein Crp and impact on virulence. MBio *5*, 1–12.

Lautner, M., Schunder, E., Herrmann, V., and Heuner, K. (2013). Regulation, integrase-dependent excision, and horizontal transfer of genomic islands in *Legionella pneumophila*. J. Bacteriol. *195*, 1583–1597.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Le, S.Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. Mol. Biol. Evol. *25*, 1307–1320.

Lee, H.J., and Hong, S.H. (2012). Analysis of microRNA-size, small RNAs in *Streptococcus mutans* by deep sequencing. FEMS Microbiol. Lett. *326*, 131–136.

Lee, S.A., Jang, S.H., Kim, B.H., Shibata, T., Yoo, J., Jung, Y., Kawabata, S., and Lee, B.L. (2018). Insecticidal activity of the metalloprotease AprA occurs through suppression of host cellular and humoral immunity. Dev. Comp. Immunol. *81*, 116–126.

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

Leo, J.C., Grin, I., and Linke, D. (2012). Type V secretion: Mechanism(s) of autotransport through the bacterial outer membrane. Philos. Trans. R. Soc. B Biol. Sci. *367*, 1088–1101.

Leon, L.J., Idangodage, H., Wan, C.P.L., and Weers, P.M.M. (2006). Apolipophorin III: Lipopolysaccharide binding requires helix bundle opening. Biochem. Biophys. Res. Commun. *348*, 1328–1333.

Lesic, B., and Carniel, E. (2005). Horizontal transfer of the high-pathogenicity island of *Yersinia pseudotuberculosis*. J. Microbiol. *187*, 3352–3358.

Lesic, B., Bach, S., Ghigo, J.M., Dobrindt, U., Hacker, J., and Carniel, E. (2004). Excision of the high-pathogenicity of *Yersinia pseudotuberculosis* requires the combined of actions of its cognate integrase and Hef, a new recombination directionality factor. Mol. Microbiol. *52*, 1337–1348.

Leskinen, K., Varjosalo, M., and Skurnik, M. (2015). Absence of YbeY RNase compromises the growth and enhances the virulence plasmid gene expression of *Yersinia enterocolitica* O:3. Microbiology *161*, 285–299.

Leskinen, K., Blasdel, B.G., Lavigne, R., and Skurnik, M. (2016). RNA-sequencing reveals the progression of phage-host interactions between φR1-37 and *Yersinia enterocolitica*. Viruses *8*.

Lesouhaitier, O., Clamens, T., Rosay, T., Desriac, F., Louis, M., Rodrigues, S., Gannesen, A., Plakunov, V.K., Bouffartigues, E., Tahrioui, A., et al. (2019). Host peptidic

hormones affecting bacterial biofilm formation and virulence. J. Innate Immun. *11*, 227–241.

Leung, K.Y., Reisner, B.S., and Straley, S.C. (1990). YopM inhibits platelet aggregation and is necessary for virulence of *Yersinia pestis* in mice. Infect. Immun. *58*, 3262–3271.

Lewis, J.A., and Hatfull, G.F. (2001). Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. Nucleic Acids Res. *29*, 2205–2216.

Leyton, D.L., Rossiter, A.E., and Henderson, I.R. (2012). From self sufficiency to dependence: Mechanisms and factors important for autotransporter biogenesis. Nat. Rev. Microbiol. *10*, 213–225.

Li, X., Gao, Y., Wang, J., Ji, G., Lu, Y., Yang, D., Shen, H., Dong, Q., Pan, L., Xiao, H., et al. (2016). Exposure to environmental endocrine disruptors and human health. Public Heal. Emerg. *1*, 47–47.

Liehl, P., Blight, M., Vodovar, N., Boccard, F., and Lemaitre, B. (2006). Prevalence of local immune response against oral infection in a *Drosophila*/*Pseudomonas* infection model. PLoS Pathog. *2*, 0551–0561.

Lin, C.S., Horng, J.T., Yang, C.H., Tsai, Y.H., Su, L.H., Wei, C.F., Chen, C.C., Hsieh, S.C., Lu, C.C., and Lai, H.C. (2010). RssAB-FlhDC-ShlBA as a major pathogenesis pathway in *Serratia marcescens*. Infect. Immun. *78*, 4870–4881.

Lindsay, J.A., Ruzin, A., Ross, H.F., Kurepina, N., and Novick, R.P. (1998). The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. Mol. Microbiol. *29*, 527–543.

Linhartová, I., Bumba, L., Mašn, J., Basler, M., Osička, R., Kamanová, J., Procházková, K., Adkins, I., HejnováHolubová, J., Sadílková, L., et al. (2010). RTX proteins: A highly diverse family secreted bya common mechanism. FEMS Microbiol. Rev. *34*, 1076–1112.

Litwin, C.M., and Calderwood, S.B. (1993). Role of iron in regulation of virulence genes. Clin. Microbiol. Rev. *6*, 137–149.

Lloréns-Rico, V., Cano, J., Kamminga, T., Gil, R., Latorre, A., Chen, W.H., Bork, P., Glass, J.I., Serrano, L., and Lluch-Senar, M. (2016). Bacterial antisense RNAs are mainly the product of transcriptional noise. Sci. Adv. *2*, 1–10.

Loftus, C.G., Harewood, G.C., Cockerill, F.R., and Murray, J.A. (2002). Clinical features of patients with novel *Yersinia* species. Dig. Dis. Sci. *47*, 2805–2810.

Van Loi, V., Rossius, M., and Antelmann, H. (2015). Redox regulation by reversible protein S-thiolation in bacteria. Front. Microbiol. *6*, 1–22.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. *42*, 490–495.

Lopilato, J.E., Garwin, J.L., Emr, S.D., Silhavy, T.J., and Beckwith, J.R. (1984). D-ribose

metabolism in *Escherichia coli* K-12: Genetics, regulation, and transport. J. Bacteriol. *158*, 665–673.

Lorow, D., and Jessee, J. (1990). Max efficiency DH10B: a host for cloning methylated DNA. Focus (Madison). *12*.

Luck, A.N., Slatko, B.E., and Foster, J.M. (2017). Removing the needle from the haystack: Enrichment of *Wolbachia* endosymbiont transcripts from host nematode RNA by Cappable-seq™. PLoS One *12*, 1–11.

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

Luo, G., Niesel, D.W., Shaban, R.A., Grimm, E.A., and Klimpel, G.R. (1993). Tumor necrosis factor alpha binding to bacteria: Evidence for a high- affinity receptor and alteration of bacterial virulence properties. Infect. Immun. *61*, 830–835.

Luzader, D.H., Clark, D.E., Gonyar, L.A., and Kendall, M.M. (2013). EutR is a direct regulator of genes that contribute to metabolism and virulence in enterohemorrhagic *Escherichia coli* O157: H7. J. Bacteriol. *195*, 4947–4953.

Luzzatto, L. (2012). G6PD deficiency and malaria selection. Heredity *108*, 456.

Lyras, D., Adams, V., Lucet, I., and Rood, J.I. (2004). The large resolvase TnpX is the only transposon-encoded protein required for transposition of the Tn4451/3 family of integrative mobilizable elements. Mol. Microbiol. *51*, 1787–1800.

Ma, J., Pan, Z., Huang, J., Sun, M., Lu, C., and Yao, H. (2017). The Hcp proteins fused with diverse extended-toxin domains represent a novel pattern of antibacterial effectors in type VI secretion systems. Virulence *8*, 1189–1202.

de Maagd, R.A., Bravo, A., Berry, C., Crickmore, N., and Schnepf, H.E. (2003). Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu. Rev. Genet. *37*, 409–433.

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.

Macori, G., Romano, A., Adriano, D., Razzuoli, E., Bianchi, D.M., Gallina, S., and Bellio, A. (2017). *Yersinia enterocolitica* strains, isolated from wild ungulate carcasses. Microbiol. Resour. Announc. *5*, 4–5.

Madrid, C., Nieto, J.M., and Juárez, A. (2001). Role of the Hha/YmoA family of proteins in the thermoregulation of the expression of virulence factors. Int. J. Med. Microbiol. *291*, 425–432.

Madrid, C., Balsalobre, C., García, J., and Juárez, A. (2007). The novel Hha/YmoA family of nucleoid-associated proteins: Use of structural mimicry to modulate the activity of the H-NS family of proteins. Mol. Microbiol. *63*, 7–14.

Maiques, E., Úbeda, C., Campoy, S., Salvador, N., Lasa, Í., Novick, R.P., Barbé, J., and

Penadés, J.R. (2006). β-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. J. Bacteriol. *188*, 2726–2729.

Manjeet, K., Purushotham, P., Neeraja, C., and Podile, A.R. (2013). Bacterial chitin binding proteins show differential substrate binding and synergy with chitinases. Microbiol. Res. *168*, 461–468.

Marceau, M. (2005). Transcriptional regulation in Yersinia: An update. Curr. Issues Mol. Biol. *7*, 151–178.

Marshall, S.D.G., Hares, M.C., Jones, S.A., Harper, L.A., Vernon, J.R., Harland, D.P., Jackson, T.A., and Hurst, M.R.H. (2012). Histopathological effects of the Yen-Tc toxin complex from *Yersinia entomophaga* MH96 (Enterobacteriaceae) on the *Costelytra zealandica* (Coleoptera: Scarabaeidae) larval midgut. Appl. Environ. Microbiol. *78*, 4835–4847.

Mason, K.M., Munson, R.S., and Bakaletz, L.O. (2003). Nontypeable Haemophilus influenzae gene expression induced in vivo in a chinchilla model of otitis media. Infect. Immun. *71*, 3454–3462.

Matsumoto, H., and Young, G.M. (2006). Proteomic and functional analysis of the suite of Ysp proteins exported by the Ysa type III secretion system of *Yersinia enterocolitica* Biovar 1B. Mol. Microbiol. *59*, 689–706.

Mauzy, C.A., and Hermodson, M.A. (1992). Structural and functional analyses of the repressor, RbsR, of the ribose operon of *Escherichia coli*. Protein Sci. *1*, 831–842.

Mavromatis, C. (Harris), Bokil, N.J., Totsika, M., Kakkanat, A., Schaale, K., Cannistraci, C. V., Ryu, T., Beatson, S.A., Ulett, G.C., Schembri, M.A., et al. (2015). The co-transcriptome of uropathogenic *Escherichia coli*-infected mouse macrophages reveals new insights into host-pathogen interactions. Cell. Microbiol. *17*, 730–746.

Mbanefo, E.C., Ahmed, A.M., Titouna, A., Elmaraezy, A., Trang, N.T.H., Phuoc Long, N., Hoang Anh, N., Diem Nghi, T., The Hung, B., Van Hieu, M., et al. (2017). Association of glucose-6-phosphate dehydrogenase deficiency and malaria: A systematic review and meta-analysis. Sci. Rep. *7*, 1–10.

McBride, M.J., and Zhu, Y. (2013). Gliding motility and por secretion system genes are widespread among members of the phylum *Bacteroidetes*. J. Bacteriol. *195*, 270–278.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.

McClelland, M., Sanderson, Kenneth, E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., et al. (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium. Nature *413*, 852–856.

McClure, R., Balasabramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, Caroline, A., Vanderpool, C.K., and Tjaden, B. (2013). Computational analysis

of bacterial RNA-seq data. Nucleic Aci *41*, e140.

McFeeters, R.L., Altieri, A.S., Cherry, S., Tropea, J.E., Waugh, D.S., and Byrd, R.A. (2007). The high-precision solution structure of *Yersinia* modulating protein YmoA provides insight into interaction with H-NS. Biochemistry *46*, 13975–13982.

McMullen, J.G., McQuade, R., Ogier, J.C., Pagès, S., Gaudriault, S., and Stock, S.P. (2017). Variable virulence phenotype of *Xenorhabdus bovienii* (γ - Proteobacteria: Enterobacteriaceae) in the absence of their vector hosts. Microbiology *163*, 510–522.

McNally, A., Thomson, N.R., Reuter, S., and Wren, B.W. (2016). "Add, stir and reduce": *Yersinia* spp. as model bacteria for pathogen evolution. Nat. Rev. Microbiol. *14*, 177–190.

McQuade, R., and Stock, S.P. (2018). Secretion systems and secreted proteins in gram-negative entomopathogenic bacteria: Their roles in insect virulence and beyond. Insects *9*.

Meier, R., Drepper, T., Svensson, V., Jaeger, K.E., and Baumann, U. (2007). A calcium-gated lid and a large β-roll sandwich are revealed by the crystal structure of extracellular lipase from *Serratia marcescens*. J. Biol. Chem. *282*, 31477–31483.

Melvin, J.A., Scheller, E. V., Noël, C.R., and Cotter, P.A. (2015). New insight into filamentous hemagglutinin secretion reveals a role for full-length FhaB in *Bordetella virulence*. MBio *6*, 12–15.

Menozzi, F.D., Boucher, P.E., Riveau, G., Gantiez, C., and Locht, C. (1994). Surface-associated filamentous hemagglutinin induces autoagglutination of *Bordetella pertussis*. Infect. Immun. *62*, 4261–4269.

Metcalf, W.W., Jiang, W., and Wanner, B.L. (1994). Use of the rep technique for allele replacement to construct new *Escherichia coli* hosts for maintenance of R6Kγ origin plasmids at different copy numbers. Gene *138*, 1–7.

Metsä-Ketelä, M., Halo, L., Munukka, E., Hakala, J., Pekka, M., and Ylihonko, K. (2002). Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various *Streptomyces* species. Appl. Environ. Microbiol. *68*, 4472–4479.

Meusch, D., Gatsogiannis, C., Efremov, R.G., Lang, A.E., Hofnagel, O., Vetter, I.R., Aktories, K., and Raunser, S. (2014). Mechanism of Tc toxin action revealed in molecular detail. Nature *508*, 61–65.

Miao, E.A., Scherer, C.A., Tsolis, R.M., Kingsley, R.A., Adams, L.G., Bäumler, A.J., and Miller, S.I. (1999). *Salmonella typhimurium* leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. Mol. Microbiol. *34*, 850–864.

Michaux, C., and Giard, J.C. (2016). New insight into cold shock proteins: RNA-binding proteins involved in stress response and virulence. In Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria, F.

de Bruijn, ed. (Hoboken, NJ: John Wiley & Sons), pp. 873–880.

Michaux, C., Holmqvist, E., Vasicek, E., Sharan, M., Barquist, L., Westermann, A.J., Gunn, J.S., and Vogel, J. (2017). RNA target profiles direct the discovery of virulence functions for the cold-shock proteins CspC and CspE. Proc. Natl. Acad. Sci. *114*, 201620772.

Michel, G., Bleves, S., Ball, G., Lazdunski, A., and Filloux, A. (1998). Mutual stabilization of the XcpZ and XcpY components of the secretory apparatus in *Pseudomonas aeruginosa*. Microbiology *144*, 3379–3386.

Middendorf, B., Hochhut, B., Leipold, K., Dobrindt, U., Blum-Oehler, G., and Hacker, J. (2004). Instability of pathogenicity islands in uropathogenic *Escherichia coli* 536. J. Bacteriol. *186*, 3086–3096.

Mielich-Süss, B., and Lopez, D. (2015). Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. Environ. Microbiol. *17*, 555–565.

Mikuliskis, A. V, Delor, I., Thi, V.H., and Cornelis, G.R. (1994). Regulation of the *Yersinia enterocolitica* enterotoxin Yst gene. Influence on growth phase, temperature, osmolarity, pH and bacterial host factors. Mol. Microbiol. *14*, 905–915.

Miller, H.K., and Auerbuch, V. (2015). Bacterial iron-sulfur cluster sensors in mammalian pathogens. Metallomics *7*, 943–956.

Miller, S.I., Hoffman, L.R., and Sanowar, S. (2007). Did bacterial sensing of host environments evolve from sensing within microbial communities? Cell Host Microbe *1*, 85–87.

Mills, E., Baruch, K., Charpentier, X., Kobi, S., and Rosenshine, I. (2008). Real-time analysis of effector translocation by the type III secretion system of enteropathogenic *Escherichia coli*. Cell Host Microbe *3*, 104–113.

Mills, E., Petersen, E., Kulasekara, B.R., and Miller, S.I. (2015). A direct screen for c-di-GMP modulators reveals a *Salmonella* Typhimurium periplasmic L-arginine-sensing pathway. Sci. Signal. *8*.

Minnich, S.A., and Rohde, H.N. (2007). A rationale for repression and/or loss of motility by pathogenic *Yersinia* in the mammalian host. In The Genus *Yersinia*, R.D. Perry, and D. Fetherston, Jacquieline, eds. (New York: Spinger-Verlag), pp. 298–311.

Mitchell, A., Tam, C., Elli, D., Charlton, T., Osei-Owusu, P., Fazlollahi, F., Faull, K.F., and Schneewind, O. (2017). Glutathionylation of *Yersinia pestis* LcrV and its effects on plague pathogenesis. MBio *8*, 1–15.

Mohanty, B.K., and Kushner, S.R. (2006). The majority of *Escherichia coli* mRNAs undergo post-transcriptional modification in exponentially growing cells. Nucleic Acids Res. *34*, 5695–5704.

Montminy, S.W., Khan, N., McGrath, S., Walkowicz, M.J., Sharp, F., Conlon, J.E., Fukase, K., Kusumoto, S., Sweet, C., Miyake, K., et al. (2006). Virulence factors of *Yersinia pestis* are overcome by a strong lipopolysaccharide response. Nat.

Immunol. *7*, 1066–1073.

Moon, B.Y., Park, J.Y., Hwang, S.Y., Robinson, D.A., Thomas, J.C., Fitzgerald, J.R., Park, Y.H., and Seo, K.S. (2015). Phage-mediated horizontal transfer of a *Staphylococcus aureus* virulence-associated genomic island. Sci. Rep. *5*, 1–6.

Moon, B.Y., Park, J.Y., Robinson, D.A., Thomas, J.C., Park, Y.H., Thornton, J.A., and Seo, K.S. (2016). Mobilization of genomic islands of *Staphylococcus aureus* by temperate bacteriophage. PLoS One *11*, 1–16.

Moreira, C.G., Russell, R., Mishra, A.A., Narayanan, S., Ritchie, J.M., Waldor, M.K., Curtis, M.M., Winter, S.E., Weinshenker, D., and Sperandio, V. (2016). Bacterial adrenergic sensors regulate virulence of enteric pathogens in the gut. MBio *7*, 1–14.

Morgan, J.A.W., Sergeant, M., Ellis, D., Ousley, M., and Jarrett, P. (2001). Sequence analysis of insecticidal genes from *Xenorhabdus nematophilus* PMFI296. Appl. Environ. Microbiol. *67*, 2062–2069.

Morici, L.A., Carterson, A.J., Wagner, V.E., Frisk, A., Schurr, J.R., Höner zu Bentrup, K., Hassett, D.J., Iglewski, B.H., Sauer, K., and Schurr, M.J. (2007). *Pseudomonas aeruginosa* AlgR represses the Rhl quorum-sensing system in a biofilm-specific manner. J. Bacteriol. *189*, 7752–7764.

Morrissey, J.A., Cockayne, A., Brummell, K., and Williams, P. (2004). The *Staphylococcal* ferritins are differentially regulated in response to iron and manganese and via PerR and Fur. Infect. Immun. *72*, 972–979.

Motin, V.L., Georgescu, A.M., Fitch, J.P., Gu, P.P., Nelson, D.O., Mabery, S.L., Garnham, J.B., Sokhansanj, B.A., Ott, L.L., Coleman, M.A., et al. (2004). Temporal global changes in gene expression during temperature transition in *Yersinia pestis*. J. Bacteriol. *186*, 6298–6305.

Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, A., Goodman, A.L., Joachimiak, G., Ordoñez, C.L., Walz, T., et al. (2006). A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. Science *312*, 1526–1530.

Mowlds, P., and Kavanagh, K. (2008). Effect of pre-incubation temperature on susceptibility of *Galleria mellonella* larvae to infection by *Candida albicans*. Mycopathologia *165*, 5–12.

Mraheil, M.A., Billion, A., Mohamed, W., Mukherjee, K., Kuenne, C., Pischimarov, J., Krawitz, C., Retey, J., Hartsch, T., Chakraborty, T., et al. (2011). The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages. Nucleic Acids Res. *39*, 4235–4248.

Mukherjee, K., Altincicek, B., Hain, T., Domann, E., Vilcinskas, A., and Chakraborty, T. (2010). *Galleria mellonella* as a model system for studying *Listeria* pathogenesis. Appl. Environ. Microbiol. *76*, 310–317.

Mulder, D.T., Cooper, C.A., and Coombes, B.K. (2012). Type VI secretion system-associated gene clusters contribute to pathogenesis of *Salmonella enterica*

serovar Typhimurium. Infect. Immun. *80*, 1996–2007.

Müller, S., Garcia-Gonzalez, E., Genersch, E., and Süssmuth, R.D. (2015). Involvement of secondary metabolites in the pathogenesis of the American foulbrood of honey bees caused by *Paenibacillus larvae*. Nat. Prod. Rep. *32*, 765–778.

Mulley, G., Beeton, M.L., Wilkinson, P., Vlisidou, I., Ockendon-Powell, N., Hapeshi, A., Tobias, N.J., Nollmann, F.I., Bode, H.B., Van Den Elsen, J., et al. (2015). From insect to man: *Photorhabdus* sheds light on the emergence of human pathogenicity. PLoS One *10*, 1–32.

Münzinger, M., Budzikiewicz, H., Expert, D., Enard, C., and Meyer, J.M. (2000). Achromobactin, a new citrate siderophore of *Erwinia chrysanthemi*. Zeitschrift Fur Naturforsch. - Sect. C J. Biosci. *55*, 328–332.

Murphy, R.A., and Boyd, E.F. (2008). Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. J. Bacteriol. *190*, 636–647.

Nagel, G., Lahrz, A., and Dersch, P. (2001). Environmental control of invasin expression in *Yersinia pseudotuberculosis* is mediated by regulation of RovA, a transcriptional activator of the SlyA/Hor family. Mol. Microbiol. *41*, 1249–1269.

Nair, S., Alokam, S., Kothapalli, S., Porwollik, S., Proctor, E., Choy, C., McClelland, M., Liu, S.L., and Sanderson, K.E. (2004). *Salmonella enterica* serovar Typhi strains from which SPI7, a 134-Kilobase island with genes for Vi exopolysaccharide and other functions, has been deleted. J. Bacteriol. *186*, 3214–3223.

Navais, R., Méndez, J., Reimundo, P., Pérez-Pascual, D., Gómez, E., and Guijarro, J.A. (2011). The *yctCBA* operon of *Yersinia ruckeri*, involved in in vivo citrate uptake, is not required for virulence. Appl. Environ. Microbiol. *77*, 1107–1110.

Navarre, W.W., McClelland, M., Libby, S.J., and Fang, F.C. (2007). Silencing of xenogeneic DNA by H-NS - Facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. Genes Dev. *21*, 1456–1471.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardern, P.P., Jones, T.A., Tate, J., et al. (2014). Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. *43*, D130–D137.

Nešić, D., Hsu, Y., and Stebbins, C.E. (2004). Assembly and function of a bacterial genotoxin. Nature *429*, 429–433.

Newton, G.L., and Fahey, R.C. (2017). Glutathione in prokaryotes. In Glutathione (1990), J. Vina, ed. (Boca Raton: CRC Press), pp. 69–78.

Nicod, S.S., Weinzierl, R.O.J., Burchell, L., Escalera-maurer, A., James, E.H., and Wigneshweraraj, S. (2014). Systematic mutational analysis of the LytTR DNA binding domain of *Staphylococcus aureus* virulence gene transcription factor AgrA. Nucleic Acids Res. *42*, 12523–12536.

Niere, M., Meißlitzer, C., Dettloff, M., Weise, C., Ziegler, M., and Wiesner, A. (1999). Insect immune activation by recombinant *Galleria mellonella* apolipophorin III.

Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. *1433*, 16–26.

Nieto, J.M., Madrid, C., Miquelay, E., Parra, J.L., Rodríguez, S., and Juárez, A. (2002). Evidence for direct protein-protein interaction between members of the enterobacterial Hha/YmoA and H-NS families of proteins. J. Bacteriol. *184*, 629–635.

Nieto, P.A., Pardo-Roa, C., Salazar-Echegarai, F.J., Tobar, H.E., Coronado-Arrázola, I., Riedel, C.A., Kalergis, A.M., and Bueno, S.M. (2016). New insights about excisable pathogenicity islands in *Salmonella* and their contribution to virulence. Microbes Infect. *18*, 302–309.

Nikolskaya, A.N., and Galperin, M.Y. (2002). A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. Nucleic Acids Res. *30*, 2453–2459.

Nivaskumar, M., and Francetic, O. (2014). Type II secretion system: A magic beanstalk or a protein escalator. Biochim. Biophys. Acta - Mol. Cell Res. *1843*, 1568–1577.

Noe, J.T., and Mitchell, R.A. (2019). Tricarboxylic acid cycle metabolites in the control of macrophage activation and effector phenotypes. J. Leukoc. Biol. 1–9.

Nollmann, F.I., Dauth, C., Mulley, G., Kegler, C., Kaiser, M., Waterfield, N.R., and Bode, H.B. (2015). Insect-specific production of new GameXPeptides in *Photorhabdus luminescens* TTO1, widespread natural products in entomopathogenic bacteria. ChemBioChem *16*, 205–208.

Nougayrède, J.-P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006). *Escherichia coli* induces DNA double-strand breaks in Eukaryotic cells. Science *313*, 848–851.

Novick, R.P., and Geisinger, E. (2008). Quorum Sensing in Staphylococci. Annu. Rev. Genet. *42*, 541–564.

Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B., and Moghazeh, S. (1993). Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. EMBO J. *12*, 3967–3975.

Nuss, A.M., Heroven, A.K., Waldmann, B., Reinkensmeier, J., Jarek, M., Beckstette, M., and Dersch, P. (2015). Transcriptomic profiling of *Yersinia pseudotuberculosis* reveals reprogramming of the Crp regulon by temperature and uncovers Crp as a master regulator of small RNAs. PLoS Genet. *11*.

Nuss, A.M., Beckstette, M., Pimenova, M., Schmühl, C., Opitz, W., Pisano, F., Heroven, A.K., and Dersch, P. (2017). Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host–pathogen transcriptomes. Proc. Natl. Acad. Sci. *114*, E791–E800.

O'Neill, S.L., Asgari, S., Mayoral, J.G., Iturbe-Ormaetxe, I., Joubert, D.A., and Hussain, M. (2014). *Wolbachia* small noncoding RNAs and their role in cross-kingdom communications. Proc. Natl. Acad. Sci. *111*, 18721–18726.

O'Shea, Y.A., and Boyd, E.F. (2002). Mobilization of the *Vibrio* pathogenicity island between *Vibrio cholerae* isolates mediated by CP-T1 generalized transduction.

FEMS Microbiol. Lett. *214*, 153–157.

Ochman, H., and Moran, N.A. (2001). Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis. Science *292*, 1096–1098.

Oerke, E.-C. (2006). Crop losses to pests. J. Agric. Sci. *144*, 31–34.

Ogier, J.C., Calteau, A., Forst, S., Goodrich-Blair, H., Roche, D., Rouy, Z., Suen, G., Zumbihl, R., Givaudan, A., Tailliez, P., et al. (2010). Units of plasticity in bacterial genomes: New insight from the comparative genomics of two bacteria interacting with invertebrates, *Photorhabdus* and *Xenorhabdus*. BMC Genomics *11*.

Oh, H., Park, Y., Park, C., Curtis, S.J., and Epstein, W. (1999). A mutated PtsG, the glucose transporter, allows uptake of D-ribose. J. Biol. Chem. *274*, 14006–14011.

Ohbayashi, T., Futahashi, R., Terashima, M., Barrière, Q., Lamouche, F., Takeshita, K., Meng, X.Y., Mitani, Y., Sone, T., Shigenobu, S., et al. (2019). Comparative cytology, physiology and transcriptomics of *Burkholderia insecticola* in symbiosis with the bean bug *Riptortus pedestris* and in culture. ISME J. 1469–1483.

Oliveira Alvarenga, D., Moreira, L.M., Chandler, M., and Varani, A.M. (2018). A practical guide for comparative genomics of mobile genetic elements in prokaryotic genomes. Methods Mol. Biol. *1704*, 213–242.

Oliver, K.M., Degnan, P.H., Hunter, Martha, S., and Moran, Nancy, A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. Science *325*, 992–994.

Olson, E.R. (1993). Influence of pH on bacterial gene expression. Mol. Microbiol. *8*, 5–14.

Ono, S., Goldberg, M.D., Olsson, T., Esposito, D., Hinton, J.C.D., and Ladbury, J.E. (2005). H-NS is a part of a thermally controlled mechanism for bacterial gene regulation. Biochem. J. *391*, 203–213.

Van Oosten, M., Rensen, P.C.N., Van Amersfoort, E.S., Van Eck, M., Van Dam, A.-M., Brevé, J.J.P., Vogel, T., Panet, A., Van Berkel, T.J.C., and Kuiper, J. (2002). Apolipoprotein E protects against bacterial lipopolysaccharide-induced lethality. J. Biol. Chem. *276*, 8820–8824.

Opota, O., Vallet-Gély, I., Vincentelli, R., Kellenberger, C., Iacovache, I., Gonzalez, M.R., Roussel, A., van der Goot, F.G., and Lemaitre, B. (2011). Monalysin, a novel ß-pore-forming toxin from the *Drosophila* pathogen *Pseudomonas entomophila*, contributes to host intestinal damage and lethality. PLoS Pathog. *7*.

Orozco, R.A., Molnár, I., Bode, H., and Stock, S.P. (2016). Bioprospecting for secondary metabolites in the entomopathogenic bacterium *Photorhabdus luminescens* subsp. *sonorensis*. J. Invertebr. Pathol. *141*, 45–52.

Ortega, Á.D., Quereda, J.J., Pucciarelli, M.G., and Garci-a-del Portillo, F. (2014). Non-

coding RNA regulation in pathogenic bacteria located inside eukaryotic cells. Front. Cell. Infect. Microbiol. *4*, 1–14.

Ostheimer, G.J., Barkan, A., and Matthews, B.W. (2002). Crystal structure of *E. coli* YhbY: A representative of a novel class of RNA binding proteins. Structure *10*, 1593–1601.

Otto, M. (2009). Bacterial sensing of antimicrobial peptides. Contrib. Microbiol. *16*, 136–149.

Owen, R.A., Fyfe, P.K., Lodge, A., Biboy, J., Vollmer, W., Hunter, W.N., and Sargent, F. (2018). Structure and activity of ChiX: a peptidoglycan hydrolase required for chitinase secretion by *Serratia marcescens*. Biochem. J. *475*, 415–428.

Oztug, M., Martinon, D., and Weers, Paul M. M.Manuscript, A. (2012). Characterization of the apoLp-III/LPS complex: insight in the mode of binding interaction. Biochemistry *51*, 6220–6227.

Pacheco, A.R., Munera, D., Waldor, M.K., Sperandio, V., and Ritchie, J.M. (2012). Fucose sensing regulates bacterial intestinal colonization. Nature *492*, 113–117.

Panis, G., Franche, N., Méjean, V., and Ansaldi, M. (2012). Insights into the functions of a prophage recombination directionality factor. Viruses *4*, 2417–2431.

Park, D., and Forst, S. (2006). Co-regulation of motility, exoenzyme and antibiotic production by the EnvZ-OmpR-FlhDC-FliA pathway in *Xenorhabdus nematophila*. Mol. Microbiol. *61*, 1397–1412.

Park, S.Y., Kim, C.H., Jeong, W.H., Lee, J.H., Seo, S.J., Han, Y.S., and Lee, I.H. (2005). Effects of two hemolymph proteins on humoral defense reactions in the wax moth, *Galleria mellonella*. Dev. Comp. Immunol. *29*, 43–51.

Park, Y., Cho, Y.J., Ahn, T., and Park, C. (1999). Molecular interactions in ribose transport: The binding protein module symmetrically associates with the homodimeric membrane transporter. EMBO J. *18*, 4149–4156.

Parker, M.W., and Feil, S.C. (2005). Pore-forming protein toxins: From structure to function. Prog. Biophys. Mol. Biol. *88*, 91–142.

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., et al. (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature *413*, 523–527.

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A.J., Thomson, N.R., Quail, M., Smith, F., Walker, D., Libberton, B., et al. (2010). Antagonistic coevolution accelerates molecular evolution. Nature *464*, 275–278.

Patil, N.S., and Jadhav, J.P. (2015). Significance of *Penicillium ochrochloron* chitinase as a biocontrol agent against pest *Helicoverpa armigera*. Chemosphere *128*, 231–235.

Patt, M. lw, Conte, L., Blaha, M., and J Plotkin, B. (2018). Steroid hormones as interkingdom signaling molecules: Innate immune function and microbial colonization modulation. AIMS Mol. Sci. *5*, 117–130.

Pavlovic, G., Burrus, V., Gintz, B., Decaris, B., and Guédon, G. (2004). Evolution of

genomic islands by deletion and tandem accretion by site-specific recombination: ICESt1-related elements from *Streptococcus thermophilus*. Microbiology *150*, 759–774.

Payelleville, A., Lanois, A., Gislard, M., Dubois, E., Roche, D., Cruveiller, S., Givaudan, A., and Brillard, J. (2017). DNA adenine methyltransferase (Dam) overexpression impairs *Photorhabdus luminescens* motility and virulence. Front. Microbiol. *8*.

Payelleville, A., Legrand, L., Ogier, J.C., Roques, C., Roulet, A., Bouchez, O., Mouammine, A., Givaudan, A., and Brillard, J. (2018). The complete methylome of an entomopathogenic bacterium reveals the existence of loci with unmethylated Adenines. Sci. Rep. *8*, 1–14.

Pech, L.L., and Strand, M.R. (1996). Granular cells are required for encapsulation of foreign targets by insect haemocytes. J. Cell Sci. *109 (Pt 8*, 2053–2060.

Pelaseyed, T., Bergström, Joakim, H., Gustafsson, J.K., Ermund, A., Birchenough, G.M.H., Schütte, A., van der Post, S., Svensson, F., Rodríguez-Piñeiro, Ana, M., Nyström, E.E.L., et al. (2014). The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. Immunol. Rev. *260*, 8–320.

Peleg, A.Y., Jara, S., Monga, D., Eliopoulos, G.M., Moellering, R.C., and Mylonakis, E. (2009). *Galleria mellonella* as a model system to study *Acinetobacter baumannii* pathogenesis and therapeutics. Antimicrob. Agents Chemother. *53*, 2605–2609.

Penadés, J.R., Chen, J., Quiles-Puchalt, N., Carpena, N., and Novick, R.P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. Curr. Opin. Microbiol. *23*, 171–178.

Penfold, R.J., and Pemberton, J.M. (1992). An improved suicide vector for construction of chromosomal insertion mutations in bacteria. Gene *118*, 145–146.

Peng, Y., Wang, X., Shou, J., Zong, B., Zhang, Y., Tan, J., Chen, J., Hu, L., Zhu, Y., Chen, H., et al. (2016). Roles of Hcp family proteins in the pathogenesis of the porcine extraintestinal pathogenic *Escherichia coli* type VI secretion system. Sci. Rep. *6*, 1–9.

Petersen, L.M., and Tisa, L.S. (2013). Friend or foe? A review of the mechanisms that drive *Serratia* towards diverse lifestyles. Can. J. Microbiol. *59*, 627–640.

Pickett, B.R., Gulzar, A., Ferré, J., and Wright, D.J. (2017). *Bacillus thuringiensis* Vip3Aa toxin resistance in *Heliothis virescens* (Lepidoptera: Noctuidae). Appl. Environ. Microbiol. *83*, e03506-16.

Piel, J., Höfer, I., and Hui, D. (2004). Evidence for a symbiosis island involved in horizontal acquisition of pederin biosynthetic capabilities by the bacterial symbiont of *Paederus fuscipes* beetles. J. Bacteriol. *186*, 1280–1286.

Pieper, R., Huang, S.T., Robinson, J.M., Clark, D.J., Alami, H., Parmar, P.P., Perry, R.D., Fleischmann, R.D., and Peterson, S.N. (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.

Pierson III, L.S., and Kahn, M.L. (1987). Integration of satellite bacteriophage P4 in *Escherichia coli*. J. Mol. Biol. *196*, 487–496.

Pinheiro, V.B., and Ellar, D.J. (2007). Expression and insecticidal activity of *Yersinia pseudotuberculosis* and *Photorhabdus luminescens* toxin complex proteins. Cell. Microbiol. *9*, 2372–2380.

Piper, S.J., Brillault, L., Rothnagel, R., Croll, T.I., Box, J.K., Chassagnon, I., Scherer, S., Goldie, K.N., Jones, S.A., Schepers, F., et al. (2019). Cryo-EM structures of the pore-forming A subunit from the *Yersinia entomophaga* ABC toxin. Nat. Commun. *10*, 1–12.

Pizarro-Cerdá, J., and Cossart, P. (2006). Bacterial adhesion and entry into host cells. Cell *124*, 715–727.

Poole, K., Schiebel, E., and Braun, V. (1988). Molecular characterization of the hemolysin determinant of *Serratia marcescens*. J. Bacteriol. *170*, 3177–3188.

Poole, S.J., Diner, E.J., Aoki, S.K., Braaten, B.A., t'Kint de Roodenbeke, C., Low, D.A., and Hayes, C.S. (2011). Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (cdi) and rearrangement hotspot (rhs) systems. PLoS Genet. *7*.

Porat, R., Clark, B.D., Wolff, S.M., and Dinarello, C.A. (1991). Enhancement of growth of virulent strains of *Escherichia coli* by interleukin-1. Science *254*, 430–432.

Pratt, A.J., Didonato, M., Shin, D.S., Cabelli, D.E., Bruns, C.K., Belzer, C.A., Gorringe, A.R., Langford, P.R., Tabatai, L.B., Kroll, S.J., et al. (2015). Structural, functional, and immunogenic insights on Cu, Zn superoxide dismutase pathogeni virulence factros from *Neisseria meningitidis* and *Brucella abortus*. J. Bacteriol. *197*, 3834–3847.

Pujol, C., Grabenstein, J.P., Perry, R.D., and Bliska, J.B. (2005). Replication of *Yersinia pestis* in interferon ɣ-activated macrophages requires ripA, a gene encoded in the pigmentation locus. Proc. Natl. Acad. Sci. *102*, 12909–12914.

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.

Purać, J., Kojić, D., Popović, Ž.D., Vukašinović, E., Tiziani, S., Günther, U.L., and Grubor-Lajšić, G. (2015). Metabolomic analysis of diapausing and non-diapausing larvae of the European corn borer *Ostrinia nubilalis* (Hbn.) (Lepidoptera: Crambidae). Acta Chim. Slov. *62*, 761–767.

Purushotham, P., Arun, P.V.P.S., Prakash, J.S.S., and Podile, A.R. (2012). Chitin binding proteins act synergistically with chitinases in *Serratia proteamaculans* 568. PLoS One *7*, 1–10.

Puymège, A., Bertin, S., Chuzeville, S., Guédon, G., and Payot, S. (2013). Conjugative transfer and cis-mobilization of a genomic island by an integrative and conjugative element of *Streptococcus agalactiae*. J. Bacteriol. *195*, 1142–1151.

Quade, N., Mendonca, C., Herbst, K., Heroven, A.K., Ritter, C., Heinz, D.W., and Dersch, P. (2012). Structural basis for intrinsic thermosensing by the master virulence regulator RovA of *Yersinia*. J. Biol. Chem. *287*, 35796–35803.

Quereda, J.J., and Cossart, P. (2017). Regulating bacterial virulence with RNA. Annu. Rev. Microbiol. *71*, 263–280.

Quiroz, T.S., Nieto, P.A., Tobar, H.E., Salazar-Echegarai, F.J., Lizana, R.J., Quezada, C.P., Santiviago, C.A., Araya, D. V., Riedel, C.A., Kalergis, A.M., et al. (2011). Excision of an unstable pathogenicity island in *Salmonella enterica* serovar Enteritidis is induced during infection of phagocytic cells. PLoS One *6*.

R Core Team (2017). R: A language and environment for statistical computing.

R10 Geneious.

Raghavan, R., Sloan, D.B., and Ochman, H. (2012). Pervasive transcription is widespread but rarely conserved in enteric bacteria. MBio *3*, 1–7.

Rajanna, C., Wang, J., Zhang, D., Xu, Z., Ali, A., Hou, Y.-M., and Karaolis, D.K.R. (2003). The *Vibrio* pathogenicity island of epidemic *Vibrio cholerae* forms precise extrachromosomal circular excision products. J. Bacteriol. *185*, 6893–6901.

Ralser, M., Wamelink, M.M., Kowald, A., Gerisch, B., Heeren, G., Struys, E.A., Klipp, E., Jakobs, C., Breitenbach, M., Lehrach, H., et al. (2007). Dynamic rerouting of the carbohydrate flux is key to counteracting oxidative stress. J. Biol. *6*.

Rasmussen, S., Nielsen, H.B., and Jarmer, H. (2009). The transcriptionally active regions in the genome of *Bacillus subtilis*. Mol. Microbiol. *73*, 1043–1057.

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

Rebeil, R., Ernst, R.K., Gowen, B.B., Miller, S.I., and Hinnebusch, B.J. (2004). Variation in lipid A structure in the pathogenic Yersiniae. Mol. Microbiol. *52*, 1363–1373.

Recht, J., Ashley, E.A., and White, N.J. (2018). Use of primaquine and glucose-6-phosphate dehydrogenase deficiency testing: Divergent policies and practices in malaria endemic countries. PLoS Negl. Trop. Dis. *12*, 1–27.

Regev, A., Keller, M., Strizhov, N., Sneh, B., Prudovsky, E., Chet, I., Ginzberg, I., Koncz-Kalman, Z., Koncz, C., Schell, J., et al. (1996). Synergistic activity of a *Bacillus thuringiensis* δ-endotoxin and a bacterial endochitinase against *Spodoptera littoralis* larvae. Appl. Environ. Microbiol. *62*, 3581–3586.

Regmi, A., and Boyd, E.F. (2019). Carbohydrate metabolic systems present on genomic islands are lost and gained in *Vibrio parahaemolyticus*. BMC Microbiol. *19*, 1–20.

Reisig, D.D., and Kurtz, R. (2018). Bt resistance implications for *Helicoverpa zea* (Lepidoptera: Noctuidae) insecticide resistance management in the United States. Environ. Entomol. *47*, 1357–1364.

Remaut, H., Tang, C., Henderson, N.S., Pinkner, J.S., Wang, T., Hultgren, S.J., Thanassi, D.G., Waksman, G., and Li, H. (2008). Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. Cell *133*, 640–652.

Rengarajan, J., Bloom, B.R., and Rubin, E.J. (2005). Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. Proc. Natl. Acad. Sci. U.S.A. *102*, 8327–8332.

Reniere, M.L. (2018). Reduce, induce, thrive: Bacterial redox sensing during pathogenesis. J. Bacteriol. *200*, 1–12.

Reniere, M.L., Whiteley, A.T., Hamilton, K.L., John, S.M., Lauer, P., Brennan, R.G., and Portnoy, D.A. (2015). Glutathione activates virulence gene expression of an intracellular pathogen. Nature *517*, 170–173.

Renwick, J., Reeves, E.P., Wientjes, F.B., and Kavanagh, K. (2007). Translocation of proteins homologous to human neutrophil $p47^{phox}$ and $p67^{phox}$ to the cell membrane in activated hemocytes of *Galleria mellonella*. Dev. Comp. Immunol. *31*, 347–359.

Reuter, S., Connor, T.R., Barquist, L., Walker, D., Feltwell, T., Harris, S.R., Fookes, M., Hall, M.E., Petty, N.K., Fuchs, T.M., et al. (2014). Parallel independent evolution of pathogenicity within the genus *Yersinia*. Proc. Natl. Acad. Sci. U.S.A. *111*, 6768–6773.

Reyes-DelaTorre, A., Peña-Rangel, M.T., and Riesgo-Escovar, J.R. (2012). Carbohydrate metabolism in *Drosophila*: Reliance on the disaccharide trehalose. In Carbohydrates: Comprehensive Studies on Glycobiology and Glycotechnology, C.-F. Chang, ed. (London: IntechOpen), pp. 317–338.

Ribeiro, C., Vignes, M., and Brehélin, M. (2003). *Xenorhabdus nematophila* (enterobacteriacea) secretes a cation-selective calcium-independent porin which causes vacuolation of the rough endoplasmic reticulum and cell lysis. J. Biol. Chem. *278*, 3030–3039.

Rienksma, R.A., Suarez-Diez, M., Mollenkopf, H.J., Dolganov, G.M., Dorhoi, A., Schoolnik, G.K., Dos Santos, V.A.P.M., Kaufmann, S.H.E., Schaap, P.J., and Gengenbacher, M. (2015). Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. BMC Genomics *16*, 1–15.

Righetti, F., Nuss, A.M., Twittenhoff, C., Beele, S., Urban, K., Will, S., Bernhart, S.H., Stadler, P.F., Dersch, P., and Narberhaus, F. (2016). Temperature-responsive in vitro RNA structurome of *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. *113*, 7237–7242.

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rodou, A., Ankrah, D.O., and Stathopoulos, C. (2010). Toxins and secretion systems of *Photorhabdus luminescens*. Toxins *2*, 1250–1264.

Rodríguez-Segura, Z., Chen, J., Villalobos, F.J., Gill, S., and Nuñez-Valdez, M.E. (2012). The lipopolysaccharide biosynthesis core of the Mexican pathogenic strain *Serratia entomophila* is associated with toxicity to larvae of *Phyllophaga blanchardi*. J. Invertebr. Pathol. *110*, 24–32.

Rodriguez-valera, F., Martin-Cuadrado, A.-B., and López-Pérez, M. (2016). Flexible genomic islands as drivers of genome evolution. Curr. Opin. Microbiol. *31*, 154–160.

Rohde, J.R., Luan, X.S., Rohde, H., Fox, J.M., and Minnich, S.A. (1999). The *Yersinia enterocolitica* pYV virulence plasmid contains multiple intrinsic DNA bends which melt at 37°C. J. Bacteriol. *181*, 4198–4204.

Rohmer, L., Hocquet, D., and Miller, S.I. (2011). Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. Trends Microbiol. *19*, 341–348.

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

Romeo, T., Vakulskas, C., and Babitzke, P. (2013). Posttranscriptional regulation on a global scale: From and function of Csr/Rsm systems. Environ. Microbiol. *15*, 313–324.

Roncarati, D., and Scarlato, V. (2017). Regulation of heat-shock genes in bacteria: from signal sensing to gene expression output. FEMS Microbiol. Rev. *41*, 549–574.

Rosso, M.L., Chauvaux, S., Dessein, R., Laurans, C., Frangeul, L., Lacroix, C., Schiavo, A., Dillies, M.A., Foulon, J., Coppée, J.Y., et al. (2008). Growth of *Yersinia pseudotuberculosis* in human plasma: Impacts on virulence and metabolic gene expression.

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.

Rutherford, S.T., and Bassler, B.L. (2012). Bacterial quorum sensing: Its role in virulence and possibilities for its control. Cold Spring Harb. Perspect. Med. *2*, 1–26.

Ruwende, C., and Hill, A. (1998). Glucose-6-Phosphate dehydrogenase deficiency and malaria. Am. J. Trop. Med. Hyg. *76*, 581–588.

Ruzin, A., Lindsay, J., and Novick, R.P. (2001). Molecular genetics of SaPI1 – a mobile pathogenicity island in *Staphylococcus aureus*. *41*, 365–377.

Ryan, K.C., Gruce, A.I., Johnson, O.E., Brunold, T.C., Cabelli, D.E., Garman, S.C., and Maroney, M.J. (2015). Nickel superoxidase dismutase, structural and functional roles of His1 and its H-bonding network. Biochemistry *54*, 1016–1027.

Saberi, F., Kamali, M., Najafi, A., Yazdanparast, A., and Moghaddam, M.M. (2016). Natural antisense RNAs as mRNA regulatory elements in bacteria: A review on function and applications. Cell. Mol. Biol. Lett. *21*, 1–17.

Saeidi, F., Moharramipour, S., and Barzegar, M. (2013). Seasonal patterns of cold hardiness and cryoprotectant profiles in *Brevicoryne brassicae* (Hemiptera: Aphididae). Environ. Entomol. *41*, 1638–1643.

Sakellaris, H., Luck, S.N., Al-hasani, K., Rajakumar, K., Turner, S.A., and Adler, B. (2004). Regulated site-specific recombination of the *she* pathogenicity island of *Shigella flexneri*. Mol. Micro *52*, 1329–1336.

Salazar-Echegarai, F.J., Tobar, H.E., Nieto, P.A., Riedel, C.A., and Bueno, S.M. (2014). Conjugal transfer of the pathogenicity island ROD21 in *Salmonella enterica* serovar Enteritidis depends on environmental conditions. PLoS One *9*.

Santana, M.S., Monteiroa, W.M., Siqueiraa, A.M., Costa, M.F., Sampaio, V., Lacerdaa, M. V., and Alecrima, M.G. (2013). Glucose-6-phosphate dehydrogenase deficient variants are associated with reduced susceptibility to malaria in the brazilian amazon. Trans. R. Soc. Trop. Med. Hyg. *107*, 301–306.

Santos, G., Schroeder, A.J., Goodman, J.L., Strelets, V.B., Crosby, M.A., Thurmond, J., Emmert, D.B., Gelbart, W.M., and Consortium, F. (2015). FlyBase: introduction of the *Drosophila melanogaster* release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. *43*, 690–697.

Sarkar, D. (2008). Lattice: multivariate data visualization with R (New York, NY: Springer).

Sarris, P.F., and Scoulica, E. V. (2011). *Pseudomonas entomophila* and *Pseudomonas mendocina*: Potential models for studying the bacterial type VI secretion system. Infect. Genet. Evol. *11*, 1352–1360.

Sasikaran, J., Zadora, P.K., Fleig, A., and Berg, I.A. (2014). bacterial itaconate degradation promotes pathogenicity. Nat. Chem. Biol. *10*, 371–379.

Sassetti, C.M., and Rubin, E.J. (2003). Genetic requirements for mycobacterial survival during infection. Proc. Natl. Acad. Sci. U.S.A. *100*, 12989–12994.

Schaefer, J., Jovanovic, G., Kotta-Loizou, I., and Buck, M. (2016a). A data comparison between a traditional and the single-step β-galactosidase assay. Data Br. *8*, 350–352.

Schaefer, J., Jovanovic, G., Kotta-Loizou, I., and Buck, M. (2016b). Single-step method for β-galactosidase assays in *Escherichia coli* using a 96-well microplate reader. Anal. Biochem. *503*, 56–57.

Scheller, E. V., and Cotter, P.A. (2015). *Bordetella* filamentous hemagglutinin and fimbriae: critical adhesins with unrealized vaccine potential. Pathog. Dis. *73*, ftv079.

Schiano, C.A., and Lathem, W.W. (2012). Post-transcriptional regulation of gene expression in *Yersinia* species. Front. Cell. Infect. Microbiol. *2*, 1–16.

Schmid, A., Neumayer, W., Trülzsch, K., Israel, L., Imhof, A., Roessle, M., Sauer, G., Richter, S., Lauw, S., Eylert, E., et al. (2009). Cross-talk between type three secretion system and metabolism in *Yersinia*. J. Biol. Chem. *284*, 12165–12177.

Schmidt, H., and Hensel, M. (2004). Pathogenicity islands in bacterial pathogensis. Clin. Microbiol. Rev. *17*, 14–56.

Schmit, A.R., Rowley, A.F., and Ratcliffe, N.A. (1977). The role of *Galleria mellonella* hemocytes in melanin formation. J. Invertebr. Pathol. *29*, 232–234.

Schubert, S., Rakin, A., Karch, H., Carniel, E., and Heesemann, J. (1998). Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. Infect. Immun. *66*, 480–485.

Schubert, S., Rakin, A., Fischer, D., Sorsa, J., and Heesemann, J. (1999). Characterization of the integration site of *Yersinia* high-pathogenicity island in *Escherichia coli*. FEMS Microbiol. Lett. *179*, 409–414.

Schubert, S., Dufke, S., Sorsa, J., and Heesemann, J. (2004). A novel integrative and conjugative element (ICE) of *Escherichia coli*: the putative progenitor of the *Yersinia* high-pathogenicity island. Mol. Microbiol. *51*, 837–848.

Schulte, R.D., Makus, C., and Schulenburg, H. (2013). Host-parasite coevolution favours parasite genetic diversity and horizontal gene transfer. J. Evol. Biol. *26*, 1836–1840.

Schuster, M., Joseph Sexton, D., Diggle, S.P., and Peter Greenberg, E. (2013). Acyl-homoserine lactone quorum sensing: From evolution to application. Annu. Rev. Microbiol. *67*, 43–63.

Schwämmle, V., and Jensen, O.N. (2010). A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. Bioinformatics *26*, 2841–2848.

Schwendener, S., and Perreten, V. (2018). The integrase of the *Macrococcus caseolyticus* resistance island mecD ($McRI_{mecD}$) inserts DNA site-specifically into *Staphylococcus* and *Bacillus* chromosomes. Mol. Microbiol. *110*, 455–468.

Senior, N.J., Bagnall, M.C., Champion, O.L., Reynolds, S.E., La Ragione, R.M., Woodward, M.J., Salguero, F.J., and Titball, R.W. (2011). *Galleria mellonella* as an infection model for *Campylobacter jejuni* virulence. J. Med. Microbiol. *60*, 661–669.

Sergeant, M., Jarrett, P., Ousley, M., and Morgan, J.A.W. (2003). Interactions of insecticidal toxin gene products from *Xenorhabdus nematophilus* PMFI296. Appl. Environ. Microbiol. *69*, 3344–3349.

Serra, D.O., Conover, M.S., Arnal, L., Sloan, G.P., Rodriguez, M.E., Yantorno, O.M., and Deora, R. (2011). FHA-mediated cell-substrate and cell-cell adhesions are critical for *Bordetella pertussis* biofilm formation on abiotic surfaces and in the mouse nose and the trachea. PLoS One *6*.

Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R., and Cossart, P. (2013). The excludon: A new concept in bacterial antisense RNA-mediated gene regulation. Nat. Rev. Microbiol. *11*, 75–82.

Sha, J., Rosenzweig, J.A., Kozlova, E. V., Wang, S., Erova, T.E., Kirtley, M.L., van Lier, C.J., and Chopra, A.K. (2013). Evaluation of the roles played by Hcp and VgrG type 6 secretion system effectors in *Aeromonas hydrophila* SSU pathogenesis. Microbiology *159*, 1120–1135.

Sharma, C.M., and Vogel, J. (2014). Differential RNA-seq: The approach behind and the biological insight gained. Curr. Opin. Microbiol. *19*, 97–105.

Sharma, S., H., H., Anand, A., and B., B. (2018). Assembly defects abrogate proofreading by initiation factors and license the entry of premature ribosomes into the translation cycle. BioRxiv Prepr.

Shatalin, K., Shatalina, E., Mironov, A., and Nudler, E. (2011). H_2S: A universal defense against antibiotics in bacteria. Science *334*, 986–990.

Shaw, R.J., Mcneill, M.M., Maass, D.R., Hein, W.R., Barber, T.K., Wheeler, M., Morris, C.A., and Shoemaker, C.B. (2003). Identification and characterisation of an aspartyl protease inhibitor homologue as a major allergen of *Trichostrongylus colubriformis*. Int. J. Parasitol. *33*, 1233–1243.

Sherlock, O., Schembri, M.A., Reisner, A., and Klemm, P. (2004). Novel roles for the AIDA adhesin from diarrheagenic *Escherichia coli*: Cell aggregation and biofilm formation. J. Bacteriol. *186*, 8058–8065.

Shigeno, Y., Uchiumi, T., and Nomura, T. (2016). Involvement of ribosomal protein L6 in assembly of functional 50S ribosomal subunit in *Escherichia coli* cells. Biochem. Biophys. Res. Commun. *473*, 237–242.

Shigenobu, S., and Stern, D.L. (2013). Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. Proc. R. Soc. B Biol. Sci. *280*.

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.

Shutinoski, B., Schmidt, M.A., and Heusipp, G. (2010). Transcriptional regulation of the Yts1 type II secretion system of *Yersinia enterocolitica* and identification of secretion substrates. Mol. Microbiol. *75*, 676–691.

Sidote, D.J., Barbieri, C.M., Wu, T., and Stock, A.M. (2008). Structure of the *Staphylococcus aureus* AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. Structure *16*, 727–735.

Sigrell, J.A., Cameron, A.D., Jones, T.A., and Mowbray, S.L. (1998). Structure of *Escherichia* coli ribokinase in complex with ribose and dinucleotide determined to 1.8 A resolution: insigts into a new family of kinase structures. Structure *6*, 183–193.

Silby, M.W., and Levy, S.B. (2008). Overlapping protein-encoding genes in *Pseudomonas fluorescens* Pf0-1. PLoS Genet. *4*, 1–8.

Singh, K., Milstein, J.N., and Navarre, W.W. (2016). Xenogeneic silencing and its impact on bacterial genomes. Annu. Rev. Microbiol. *70*, 199–213.

Šišak, D., McCusker, L.B., Zandomeneghi, G., Meier, B.H., Bläser, D., Boese, R., Bernd Schweizer, W., Gilmour, R., and Dunitz, J.D. (2010). The crystal structure of D-ribose - At last! Angew. Chemie - Int. Ed. *49*, 4503–4505.

Slamti, L., Lemy, C., Henry, C., Guillot, A., Huillet, E., and Lereclus, D. (2016). CodY regulates the activity of the virulence quorum sensor PlcR by controlling the import of the signaling peptide PapR in *Bacillus thuringiensis*. Front. Microbiol. *6*, 1–14.

Smith, E.R., Lin, C., Garrett, A.S., Thornton, J., Mohaghegh, N., Hu, D., Jackson, J., Saraf, A., Swanson, S.K., Seidel, C., et al. (2011). The little elongation complex (LEC) regulates small nuclear RNA transcription. Mol. Cell *44*, 954–965.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, 1–25.

Song, S., and Park, C. (1998). Utilization of D-ribose through D-xylose transporter. FEMS Microbiol. Lett. *163*, 255–261.

Song, M., Husain, M., Jones-Carson, J., Liu, L., Henard, C.A., and Vázquez-Torres, A. (2013). Low-molecular-weight thiol-dependent antioxidant and antinitrosative defences in *Salmonella* pathogenesis. Mol. Microbiol. *87*, 609–622.

Sood, S., Steinmetz, H., Beims, H., Mohr, K.I., Stadler, M., Djukic, M., von der Ohe, W., Steinert, M., Daniel, R., and Müller, R. (2014). Paenilarvins: Iturin family lipopeptides from the honey bee pathogen *Paenibacillus larvae*. Chembiochem *15*, 1947–1955.

Soucy, S.M., Huang, J., and Gogarten, J.P. (2015). Horizontal gene transfer: Building the web of life. Nat. Rev. Genet. *16*, 472–482.

Souza, C.P., Almeida, B.C., Colwell, R.R., and Rivera, I.N.G. (2011). The importance of chitin in the marine environment. Mar. Biotechnol. *13*, 823–830.

Spaans, S.K., Weusthuis, R.A., van der Oost, J., and Kengen, S.W.M. (2015). NADPH-generating systems in bacteria and archaea. Front. Microbiol. *6*, 1–27.

Spanier, B., Starke, M., Higel, F., Scherer, S., and Fuchs, T.M. (2010). *Yersinia enterocolitica* infection and tcaA-dependent killing of *Caenorhabditis elegans*. Appl. Environ. Microbiol. *76*, 6277–6285.

Springer, K., Reuter, S., Knüpfer, M., Schmauder, L., Sänger, P.A., Felsl, A., and Fuchs, T.M. (2018a). Activity of a holin-endolysin system in the insecticidal pathogenicity island of *Yersinia enterocolitica*. J. Bacteriol. *200*, 1–13.

Springer, K., Sänger, P.-A., Moritz, C., Felsl, A., Rattei, T., and Fuchs, T.M. (2018b). Insecticidal toxicity of *Yersinia frederiksenii* involves the novel enterotoxin YacT. Front. Cell. Infect. Microbiol. *8*, 1–13.

Srisucharitpanit, K., Yao, M., Chimnaronk, S., Promdonkoy, B., Tanaka, I., and Boonserm, P. (2013). Crystallization and preliminary X-ray crystallographic analysis of the functional form of BinB binary toxin from *Bacillus sphaericus*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. *69*, 170–173.

Srisucharitpanit, K., Yao, M., Promdonkoy, B., Chimnaronk, S., Tanaka, I., and Boonserm, P. (2014). Crystal structure of BinB: A receptor binding component of the binary toxin from *Lysinibacillus sphaericus*. Proteins Struct. Funct. Bioinforma. *82*, 2703–2712.

Stączek, S., Zdybicka-Barabas, A., Mak, P., Sowa-Jasiłek, A., Kedracka-Krok, S., Jankowska, U., Suder, P., Wydrych, J., Grygorczuk, K., Jakubowicz, T., et al. (2017). Studies on localization and protein ligands of *Galleria mellonella* apolipophorin III during immune response against different pathogens. J. Insect Physiol. *105*, 18–27.

Starke, M., and Fuchs, T.M. (2014). YmoA negatively controls the expression of insecticidal genes in *Yersinia enterocolitica*. Mol. Microbiol. *92*, 287–301.

Starke, M., Richter, M., and Fuchs, T.M. (2013). The insecticidal toxin genes of *Yersinia enterocolitica* are activated by the thermolabile LTTR-like regulator TcaR2 at low temperatures. Mol. Microbiol. *89*, 596–611.

Stewart, G.R., and Young, D.B. (2004). Heat-shock proteins and the host-pathogen interaction during bacterial infection. Curr. Opin. Immunol. *16*, 506–510.

Van Der Straaten, T., Zulianello, L., Van Diepen, A., Granger, D.L., Janssen, R., and Van Dissel, J.T. (2004). *Salmonella enterica* serovar Typhimurium RamA, intracellular oxidative stress response, and bacterial virulence. Infect. Immun. *72*, 996–1003.

Subramoni, S., Florez Salcedo, D.V., and Suarez-Moreno, Z.R. (2015). A bioinformatic survey of distribution, conservation, and probable functions of LuxR solo regulators in bacteria. Front. Cell. Infect. Microbiol. *5*, 1–17.

Suhr, M., Benz, I., and Schmidt, M.A. (1996). Processing of the AIDA-I precursor: Removal of AIDA^{c} and evidence for the outer membrane anchoring as a β-barrel structure. Mol. Microbiol. *22*, 31–42.

Sui, S.J.H., Fedynak, A., Hsiao, W.W.L., Langille, M.G.I., and Brinkman, F.S.L. (2009). The Association of virulence factors with genomic islands. PLoS One *4*, e8094.

Sullivan, J.T., Trzebiatowski, J.R., Cruickshank, R.W., Gouzy, J., Brown, S.D., Elliot, R.M., Fleetwood, D.J., Mccallum, N.G., Rossbach, U., Stuart, G.S., et al. (2002). Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. J. Bacteriol. *184*, 3086–3095.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011). Easyfig: A genome comparison visualizer. Bioinformatics *27*, 1009–1010.

Sun, J., Zhao, R., Zeng, J., Li, G., and Li, X. (2010). Characterization of dextrins with different dextrose equivalents. Molecules *15*, 5162–5173.

Sun, Y.Y., Chi, H., and Sun, L. (2016). *Pseudomonas fluorescens* filamentous hemagglutinin, an iron-regulated protein, is an important virulence factor that modulates bacterial pathogenicity. Front. Microbiol. *7*, 1–11.

Suomalainen, M., Lobo, L.A., Brandenburg, K., Lindner, B., Virkola, R., Knirel, Y.A., Anisimov, A.P., Holst, O., and Korhonen, T.K. (2010). Temperature-induced

changes in the lipopolysaccharide of *Yersinia pestis* affect plasminogen activation by the Pla surface protease. Infect. Immun. *78*, 2644–2652.

Suzuki, K., Suzuki, M., Taiyoji, M., Nikadou, N., and Watanabe, T. (1998). Chitin binding protein (CBP21) in the culture supernatant of *Serratia marcescens* 2170. Biosci. Biotechnol. Biochem. *62*, 128–135.

Suzuki, K., Taiyoji, M., Sugawara, N., Nikaidou, N., Henrissat, B., and Watanabe, T. (1999). The third chitinase gene (*chiC*) of *Serratia marcescens* 2170 and the relationship of its product to other bacterial chitinases. Biochem. J. *343*, 587–596.

Suzuki, K., Sugawara, N., Suzuki, M., Uchiyama, T., Katouno, F., Nikadou, N., and Watanabe, T. (2002). Chitinases A, B, and C1 of *Serratia marcescens* 2170 produced by recombinant *Escherichia coli*: Enzymatic properties and synergism on chitin degradation. Biosci. Biotechnol. Biochem. *66*, 1075–1083.

Szabady, R.L., Yanta, J.H., Halladin, D.K., Schofield, M.J., and Welch, R.A. (2011). TagA is a secreted protease of *Vibrio cholerae* that specifically cleaves mucin glycoproteins. Microbiology *157*, 516–525.

Tago, K., Kikuchi, Y., Nakaoka, S., Katsuyama, C., and Hayatsu, M. (2015). Insecticide applications to soil contribute to the development of *Burkholderia* mediating insecticide resistance in stinkbugs. Mol. Ecol. *24*, 3766–3778.

Tamayo, R. (2019). Cyclic diguanylate riboswitches control bacterial pathogenesis mechanisms. PLOS Pathog. *15*, e1007529.

Tamir-Ariel, D., Navon, N., and Burdman, S. (2007). Identification of genes in <i>Xanthomonas campestris</i pv. vesicatoria induced during its interaction with tomato. J. Bacteriol. *189*, 6359–6371.

Tamir-Ariel, D., Rosenberg, T., and Burdman, S. (2011). The *Xanthomonas campestris* pv. vesicatoria citH gene is expressed early in the infection process of tomato and is positively regulated by the TctDE two-component regulatory system. Mol. Plant Pathol. *12*, 57–71.

Tang, H. (2009). Regulation and function of the melanization reaction in *Drosophila*. Fly *3*, 105–111.

Tang, Q., Yin, K., Qian, H., Zhao, Y., Wang, W., Chou, S.H., Fu, Y., and He, J. (2016). Cyclic di-GMP contributes to adaption and virulence of *Bacillus thuringiensis* through a riboswitch-regulated collagen adhesion protein. Sci. Rep. *6*, 1–12.

Tangsongcharoen, C., Chomanee, N., Promdonkoy, B., and Boonserm, P. (2015). *Lysinibacillus sphaericus* binary toxin induces apoptosis in susceptible *Culex quinquefasciatus* larvae. J. Invertebr. Pathol. *128*, 57–63.

Tannahill, G.M., Curtis, A.M., Adamik, J., Palsson-McDermott, E.M., McGettrick, A.F., Goel, G., Frezza, C., Bernard, N.J., Kelly, B., Foley, N.H., et al. (2013). Succinate is a danger signal that induces IL-1beta via HIF-1alpha. *496*, 238–242.

Tennant, S.M., Skinner, N.A., Joe, A., and Robins-Browne, R.M. (2005). Homologues of insecticidal toxin complex genes in *Yersinia enterocolitica* biotype 1A and their

contribution to virulence. Infect. Immun. *73*, 6860–6867.

Thänert, R., Goldmann, O., Beineke, A., and Medina, E. (2017). Host-inherent variability influences the transcriptional response of *Staphylococcus aureus* during in vivo infection. Nat. Commun. *8*.

Thoma, S., and Schobert, M. (2009). An improved *Escherichia coli* donor strain for diparental mating. FEMS Microbiol. Lett. *294*, 127–132.

Thomason, M.K., and Storz, G. (2010). Bacterial antisense RNAs: How many are there, and what are they doing? Annu. Rev. Genet. *44*, 167–188.

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

Thomson, N.R., Crow, M.A., McGowan, S.J., Cox, A., and Salmond, G.P.C. (2000). Biosynthesis of carbapenem antibiotic and prodigiosin pigment in *Serratia* is under quorum sensing control. Mol. Microbiol. *36*, 539–556.

Thomson, N.R., Howard, S., Wren, B.W., Holden, M.T.G., Crossman, L., Challis, G.L., Churcher, C., Mungall, K., Brooks, K., Chillingworth, T., et al. (2006). The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. PLoS Genet. *2*, 2039–2051.

Thomson, N.R., Clayton, D.J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., Quail, M.A., Stevens, M., Jones, M.A., Watson, M., et al. (2008). Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res. *18*, 1624–1637.

Tischler, A.D., and Camilli, A. (2012). Cyclic diguanylate regulates *Vibrio cholerae* virulence gene expression. Infect. Immun. *73*, 5873–5882.

Tobar, H., Salazar-Echegarai, F., Nieto, P., Palavecino, C., Sebastian, V., Riedel, C., Kalergis, A., and Bueno, S. (2013). Chromosomal excision of a new pathogenicity island modulates *Salmonella* virulence *in vivo*. Curr. Gene Ther. *13*, 240–249.

Tobias, N.J., Mishra, B., Gupta, D.K., Sharma, R., Thines, M., Stinear, T.P., and Bode, H.B. (2016). Genome comparisons provide insights into the role of secondary metabolites in the pathogenic phase of the *Photorhabdus* life cycle. BMC Genomics *17*, 1–11.

Tojo, S., Naganuma, F., Arakawa, K., and Yokoo, S. (2000). Involvement of both granular cells and plasmatocytes in phagocytic reactions in the greater wax moth, *Galleria mellonella*. J. Insect Physiol. *46*, 1129–1135.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., et al. (2009). The *Listeria* transcriptional landscape from saprophytism to virulence. Nature *459*, 950–956.

Toratani, T., Shoji, T., Ikehara, T., Suzuki, K., and Watanabe, T. (2008). The importance of chitobiase and N-acetylglucosamine (GlcNAc) Uptake in N,N′-diacetylchitobiose [$(GlcNAc)_2$] utilization by *Serratia marcescens* 2170. Microbiology *154*, 1326–1332.

Tramonti, A., De Canio, M., Delany, I., Scarlato, V., and De Biase, D. (2006). Mechanisms of transcription activation exerted by GadX and GadW at the *gadA* and *gadBC* gene promoters of the glutamate-based acid resistance system in *Escherichia coli*. J. Bacteriol. *188*, 8118–8127.

Tran, S.-L., Guillemet, E., Lereclus, D., and Ramarao, N. (2013). Iron regulates *Bacillus thuringiensis* haemolysin *hlyII* gene expression during insect infection. J. Invertebr. Pathol. *113*, 205–208.

Trigui, H., Dudyk, P., Sum, J., Shuman, H.A., and Faucher, S.P. (2013). Analysis of the transcriptome of *Legionella pneumophila hfq* mutant reveals a new mobile genetic element. Microbiology *159*, 1649–1660.

Tsai, C.J.Y., Loh, J.M.S., and Proft, T. (2016). *Galleria mellonella* infection models for the study of bacterial diseases and for antimicrobial drug testing. Virulence *7*, 214–229.

Tseng, T.T., Tyler, B.M., and Setubal, J.C. (2009). Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol. *9*, 1–9.

Tsuge, H., Nagahama, M., Nishimura, H., Hisatsune, J., Sakaguchi, Y., Itogawa, Y., Katunuma, N., and Sakurai, J. (2003). Crystal structure and site-directed mutagenesis of enzymatic components from *Clostridium perfringens* Iota-toxin. J. Mol. Biol. *325*, 471–483.

Tsuge, H., Nagahama, M., Oda, M., Iwamoto, S., Utsunomiya, H., Marquez, V.E., Katunuma, N., Nishizawa, M., and Sakurai, J. (2008). Structural basis of actin recognition and ADP-ribosylation by *Clostridium perfringens* iota-toxin. Acta Crystallogr. Sect. A Found. Crystallogr. *64*, C356–C356.

Tsurumura, T., Tsumori, Y., Qiu, H., Oda, M., Sakurai, J., Nagahama, M., and Tsuge, H. (2013). Arginine ADP-ribosylation mechanism based on structural snapshots of iota-toxin and actin complex. Proc. Natl. Acad. Sci. *110*, 4267–4272.

Tu, Q., Yin, J., Fu, J., Herrmann, J., Li, Y., Yin, Y., Stewart, A.F., Müller, R., and Zhang, Y. (2016). Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency. Sci. Rep. *6*, 1–8.

Turnbull, A.L., and Surette, M.G. (2010). Cysteine biosynthesis, oxidative stress and antibiotic resistance in *Salmonella typhimurium*. Res. Microbiol. *161*, 643–650.

Turner, S.A., Luck, S.N., Sakellaris, H., Rajakumar, K., and Adler, B. (2001). Nested deletions of the SRL pathogenicity island of *Shigella flexneri* 2a. J. Bacteriol. *183*, 5535–5543.

Turner, S.A., Luck, S.N., Sakellaris, H., Rajakumar, K., and Adler, B. (2004). Role of *attP* in integrase-mediated integration of the *Shigella* resistance locus pathogenicity

island of *Shigella flexneri*. Antimicrob. Agents Chemother. *48*, 1028–1031.

Úbeda, C., Maiques, E., Knecht, E., Lasa, Í., Novick, R.P., and Penadés, J.R. (2005). Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. Mol. Microbiol. *56*, 836–844.

Uchiyama, T., Kaneko, R., Yamaguchi, J., Inoue, A., Yanagida, T., Nikaidou, N., Regue, M., and Watanabe, T. (2003). Uptake of N,N′-diacetylchitobiose [$(GlcNAc)_2$] via the phosphotransferase system is essential for chitinase production by *Serratia marcescens* 2170. J. Bacteriol. *185*, 1776–1782.

Urbany, C., and Neuhaus, H.E. (2008). Citrate uptake into *Pectobacterium atrosepticum* is critical for bacterial virulence. Mol. Plant-Microbe Interact. *21*, 547–554.

Vaaje-Kolstad, G., Horn, S.J., Van Aalten, D.M.F., Synstad, B., and Eijsink, V.G.H. (2005a). The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. J. Biol. Chem. *280*, 28492–28497.

Vaaje-Kolstad, G., Houston, D.R., Riemen, A.H.K., Eijsink, V.G.H., and Van Aalten, D.M.F. (2005b). Crystal structure and binding properties of the Serratia marcescens chitin-binding protein CBP21. J. Biol. Chem. *280*, 11313–11319.

Vaaje-Kolstad, G., WesterengBjørge, Horn, S.J., Liu, Z., Zhai, H., Sørlie, M., and Eijsink, V.G.H. (2010). An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. Science *330*, 219–222.

Vaaje-Kolstad, G., Horn, S.J., Sørlie, M., and Eijsink, V.G.H. (2013). The chitinolytic machinery of *Serratia marcescens* - A model system for enzymatic degradation of recalcitrant polysaccharides. FEBS J. *280*, 3028–3049.

Vadyvaloo, V., Jarrett, C., Sturdevant, D.E., Sebbane, F., and Hinnebusch, B.J. (2010). Transit through the flea vector induces a pretransmission innate immunity resistance phenotype in *Yersinia pestis*. PLoS Pathog. *6*.

Vallet-Gely, I., Lemaitre, B., and Boccard, F. (2008). Bacterial strategies to overcome insect defences. Nat. Rev. Microbiol. *6*, 302–313.

Vallet-Gely, I., Novikov, A., Augusto, L., Liehl, P., Bolbach, G., Péchy-Tarr, M., Cosson, P., Keel, C., Caroff, M., and Lemaitre, B. (2010). Association of hemolytic activity of *Pseudomonas entomophila*, a versatile soil bacterium, with cyclic lipopeptide production. Appl. Environ. Microbiol. *76*, 910–921.

Vanga, B.R., Ramakrishnan, P., Butler, R.C., Toth, I.K., Ronson, C.W., Jacobs, J.M.E., and Pitman, A.R. (2015). Mobilization of horizontally acquired island 2 is induced in planta in the phytopathogen Pectobacterium atrosepticum SCRI1043 and involves the putative relaxase ECA0613 and quorum sensing. Environ. Microbiol. *17*, 4730–4744.

Vannucci, F.A., Foster, D.N., and Gebhart, C.J. (2013). Laser microdissection coupled with RNA-seq analysis of porcine enterocytes infected with an obligate intracellular pathogen (*Lawsonia intracellularis*). BMC Genomics *14*, 1.

Di Venanzio, G., Lazzaro, M., Morales, E.S., Krapf, D., and García Véscovi, E. (2017). A

pore-forming toxin enables *Serratia* a nonlytic egress from host cells. Cell. Microbiol. *19*, 1–12.

Venecia, K., and Young, G.M. (2005). Environmental regulation and virulence attributes of the Ysa type III secretion system of *Yersinia enterocolitica* biovar 1B. Infect. Immun. *73*, 5961–5977.

Verma, S.C., and Miyashiro, T. (2013). Quorum sensing in the squid-*Vibrio* symbiosis. Int. J. Mol. Sci. *14*, 16386–16401.

Viegas, S.C., Mil-Homens, D., Fialho, A.M., and Arraiano, C.M. (2013). The virulence of *Salmonella enterica* Serovar Typhimurium in the insect model *Galleria mellonella* is impaired by mutations in RNase E and RNase III. Appl. Environ. Microbiol. *79*, 6124–6133.

Vigneron, A., Aksoy, E., Scolari, F., Weiss, B.L., Bing, X., Aksoy, S., Attardo, G.M., and Malacrida, A. (2017). Unravelling the relationship between the tsetse fly and its obligate symbiont *Wigglesworthia*: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. Proc. R. Soc. B Biol. Sci. *284*, 20170360.

Vodovar, N., Vallenet, D., Cruveiller, S., Rouy, Z., Barbe, V., Acosta, C., Cattolico, L., Jubin, C., Lajus, A., Segurens, B., et al. (2006). Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*. Nat. Biotechnol. *24*, 673–679.

Vogel, J., and Luisi, B.F. (2011). Hfq and its constellation of RNA. Nat. Rev. Microbiol. *9*, 578–589.

Vogel, H., Altincicek, B., Glöckner, G., and Vilcinskas, A. (2011). A comprehensive transcriptome and immune-gene repertoire of the lepidopteran model host *Galleria mellonella*. BMC Genomics *12*, 308.

Wagner, N.J., Lin, C.P., Borst, L.B., and Millera, V.L. (2013). YaxAB, a *Yersinia enterocolitica* pore-forming toxin regulated by RovA. Infect. Immun. *81*, 4208–4219.

Walker, K.A., and Miller, V.L. (2009). Synchronous gene expression of the *Yersinia enterocolitica* Ysa type III secretion system and its effectors. J. Bacteriol. *191*, 1816–1826.

Walker, K.A., Maltez, V.I., Hall, J.D., Vitko, N.P., and Miller, V.L. (2013). A phenotype at last: Essential role for the *Yersinia enterocolitica* Ysa: Type III secretion system in a *Drosophila melanogaster* S2 cell model. Infect. Immun. *81*, 2478–2487.

Walski, T., De Schutter, K., Cappelle, K., and Van Damme, E.J.M. (2017a). Distribution of glycan motifs at the surface of midgut cells in the cotton leafworm (*Spodoptera littoralis*) demonstrated by lectin binding. Front. Physiol. *8*, 1–11.

Walski, T., De Schutter, K., Van Damme, E.J.M., and Smagghe, G. (2017b). Diversity and functions of protein glycosylation in insects. Insect Biochem. Mol. Biol. *83*, 21–34.

Wang, N., Ozer, E.A., Mandel, M.J., and Hauser, A.R. (2014a). Genome-wide identification of *Acinetobacter baumannii* genes necessary for persistence in the lung. MBio *5*, 1–8.

Wang, N., Liu, J., Pang, M., Wu, Y., Awan, F., Liles, M.R., Lu, C., and Liu, Y. (2018). Diverse roles of Hcp family proteins in the environmental fitness and pathogenicity of *Aeromonas hydrophila* Chinese epidemic strain NJ-35. Appl. Microbiol. Biotechnol. *102*, 7083–7095.

Wang, S., Engahang-Ndong, J., and Smith, I. (2007). Structure of the DNA-binding domain of the response regulator PhoP from *Mycobacterium tuberculosis*. Biochemistry *46*, 14751–14761.

Wang, X.H., Qi, X.L., and Kang, L. (2010). Geographic differences on accumulation of sugars and polyols in locust eggs in response to cold acclimation. J. Insect Physiol. *56*, 966–970.

Wang, Y., Yi, L., Wang, S., Lu, C., and Ding, C. (2014b). Selective capture of transcribed sequences in the functional gene analysis of microbial pathogens. Appl. Microbiol. Biotechnol. *98*, 9983–9992.

Wang, Z., Liu, W., Wu, T., Bie, P., and Wu, Q. (2016). RNA-seq reveals the critical role of CspA in regulating *Brucella melitensis* metabolism and virulence. Sci. China Life Sci. *59*, 417–424.

Wassenaar, T.M., and Gaastra, W. (2001). Bacterial virulence: Can we draw the line? FEMS Microbiol. Lett. *201*, 1–7.

Waterfield, N.R., Bowen, D.J., Fetherston, J.D., Perry, R.D., and ffrench-Constant, R.H. (2001). The tc genes of *Photorhabdus*: a growing family. Trends Microbiol. *9*, 185–191.

Waterfield, N.R., Daborn, P.J., and Ffrench-Constant, R.H. (2002). Genomic islands in *Photorhabdus*. Trends Microbiol. *10*, 541–545.

Waterfield, N.R., Daborn, P.J., and Ffrench-Constant, R.H. (2004). Insect pathogenicity islands in the insect pathogenic bacterium Photorhabdus. Physiol. Entomol. *29*, 240–250.

Westermann, A.J., Förstner, K.U., Amman, F., Barquist, L., Chao, Y., Schulte, L.N., Müller, L., Reinhardt, R., Stadler, P.F., and Vogel, J. (2016). Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature *529*, 496–501.

Westermann, A.J., Barquist, L., and Vogel, J. (2017). Resolving host–pathogen interactions by dual RNA-seq. PLoS Pathog. *13*, 1–19.

Westhof, E. (2010). The amazing world of bacterial structured RNAs. Genome Biol. *11*.

White-Ziegler, C.A., and Davis, T.R. (2009). Genome-wide identification of H-NS-controlled, temperature-regulated genes in *Escherichia coli* K-12. J. Bacteriol. *191*, 1106–1110.

Whitten, M.M.A., Tew, I.F., Lee, B.L., and Ratcliffe, N.A. (2014). A novel role for an

insect apolipoprotein (apolipophorin III) in β-1,3-glucan pattern recognition and cellular encapsulation reactions. J. Immunol. *172*, 2177–2185.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (New York: Springer-Verlag).

Wilkinson, P., Waterfield, N.R., Crossman, L., Corton, C., Sanchez-contreras, M., Vlisidou, I., Barron, A., Bignell, A., Clark, L., Ormond, D., et al. (2009). Comparative genomics of the emerging human pathogen *Photorhabdus asymbiotica* with the insect pathogen *Photorhabdus luminescens*. BMC Genomics *10*, 302.

Will, R.W., Navarre, W.W., and Fang, F.C. (2015). Integrated circuits: How translational silencing and counter-silencing facilitate bacterial evolution. Curr. Opin. Microbiol. *0*, 8–13.

Will, W.R., Bale, D.H., Reid, P.J., Libby, S.J., and Fang, F.C. (2014). Evolutionary expansion of a regulatory network by counter-silencing. Nat. Commun. *5*, 1–12.

Will, W.R., Brzovic, P., Le Trong, I., Stenkamp, R.E., Lawrenz, M.B., Karlinsey, J.E., Navarre, W.W., Main-Hester, K., Miller, V.L., Libby, S.J., et al. (2019). The evolution of SlyA/RovA transcription factors from repressors to countersilencers in Enterobacteriaceae. MBio *10*, e00009-19.

Williams, K.P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res. *30*, 866–875.

Williams, N.C., and O'Neill, L.A.J. (2018). A role for the Krebs cycle intermediate citrate in metabolic reprogramming in innate immunity and inflammation. Front. Immunol. *9*, 1–11.

Wojda, I. (2017). Immunity of the greater wax moth *Galleria mellonella*. Insect Sci. *24*, 342–357.

Wong, J., Chen, Y., and Gan, Y.H. (2015). Host cytosolic glutathione sensing by a membrane histidine kinase activates the type VI secretion system in an intracellular bacterium. Cell Host Microbe *18*, 38–48.

Woolfit, M., Algama, M., Keith, J.M., McGraw, E.A., and Popovici, J. (2015). Discovery of putative small non-coding RNAs from the obligate intracellular bacterium *Wolbachia pipientis*. PLoS One *10*, 1–19.

Wortham, B.W., Patel, C.N., and Oliveira, M.A. (2007). Polyamines in bacteria: Pleiotropic effects yet specific mechanisms. In The Genus *Yersinia*, R.D. Perry, and J.D. Fetherston, eds. (New York, NY: Springer), pp. 106–115.

Wozniak, R.A.F., and Waldor, M.K. (2010). Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat. Rev. Microbiol. *8*, 552–563.

Wu, H.J., Wang, A.H.J., and Jennings, M.P. (2008). Discovery of virulence factors of pathogenic bacteria. Curr. Opin. Chem. Biol. *12*, 93–101.

Wu, L., Estrada, O., Zaborina, O., Bains, M., Shen, L., Kohler, J.E., Patel, N., Musch, M.W., Chang, E.B., Fu, Y.X., et al. (2005). Microbiology: Recognition of host immune activation by *Pseudomonas aeruginosa*. Science *309*, 774–777.

Yan, Y., Su, S., Meng, X., Ji, X., Qu, Y., Liu, Z., Wang, X., Cui, Y., Deng, Z., Zhou, D., et al. (2013). Determination of sRNA expressions by RNA-seq in *Yersinia pestis* grown in vitro and during infection. PLoS One *8*, 1–13.

Yook, K., Harris, T.W., Bieri, T., Cabunoc, A., Chan, J., Chen, W.J., Davis, P., de la Cruz, N., Duong, A., Fang, R., et al. (2012). WormBase 2012: more genomes, more data, new website. Nucleic Acids Res. *40*, 735–741.

Young, B.M., and Young, G.M. (2002). Evidence for targeting of Yop effectors by the chromosomally encoded Ysa type III secretion system of *Yersinia enterocolitica*. J. Bacteriol. *184*, 5563–5571.

Yu, A.C.Y., Worrall, L.J., and Strynadka, N.C.J. (2012). Structural insight into the bacterial mucinase StcE essential to adhesion and immune evasion during enterohemorrhagic E. coli infection. Structure *20*, 707–717.

Zdybicka-Barabas, A., Stączek, S., Mak, P., Skrzypiec, K., Mendyk, E., and Cytryńska, M. (2013). Synergistic action of *Galleria mellonella* apolipophorin III and lysozyme against Gram-negative bacteria. Biochim. Biophys. Acta - Biomembr. *1828*, 1449–1456.

Zdybicka-Barabas, A., Mak, P., Jakubowicz, T., and Cytryńska, M. (2014). Lysozyme and defense peptides as suppressors of phenoloxidase activity in *Galleria mellonella*. Arch. Insect Biochem. Physiol. *87*, 1–12.

Zeller, T., and Klug, G. (2006). Thioredoxins in bacteria: Functions in oxidative stress response and regulation of thioredoxin genes. Naturwissenschaften *93*, 259–266.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., and Ma, B. (2012). PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell. Proteomics *11*, 1–8.

Zhang, J., Ye, Z., Singh, S., Townsend, D.M., and Tew, K.D. (2018a). An evolving understanding of the s-glutathionylation cycle in pathways of redox regulation. Free Radic. Biol. Med. *120*, 204–216.

Zhang, X.X., Ritchie, S.R., and Rainey, P.B. (2014). Urocanate as a potential signaling molecule for bacterial recognition of eukaryotic hosts. Cell. Mol. Life Sci. *71*, 541–547.

Zhang, Y., Burkhardt, D.H., Rouskin, S., Li, G.W., Weissman, J.S., and Gross, C.A. (2018b). A stress response that monitors and regulates mRNA structure is central to cold shock adaptation. Mol. Cell *70*, 274–286.e7.

Zhao, G., and Winkler, M.E. (1994). An *Escherichia coli* K-12 *tktA tktB* mutant deficient in transketolase activity requires pyridoxine (vitamin B_6) as well as the aromatic amino acids and vitamins for growth. J. Bacteriol. *176*, 6134–6138.

Zhu, C., Ruan, L., Peng, D., Yu, Z., and Sun, M. (2006). Vegetative insecticidal protein enhancing the toxicity of *Bacillus thuringiensis* subsp *kurstaki* against *Spodoptera exigua*. Lett. Appl. Microbiol. *42*, 109–114.

Zoued, A., Brunet, Y.R., Durand, E., Aschtgen, M.S., Logger, L., Douzi, B., Journet, L., Cambillau, C., and Cascales, E. (2014). Architecture and assembly of the Type VI secretion system. Biochim. Biophys. Acta - Mol. Cell Res. *1843*, 1664–1673.

Zschiedrich, C.P., Keidel, V., and Szurmant, H. (2016). Molecular mechanisms of two-component signal transduction. J. Mol. Biol. *428*, 3752–3775.

Supplemental

Supplementary section for Chapter 3

Table S1: *Yersinia entomophaga* MH96 and targeted mutagenesis strains used in this study*

Strain	Description	Reference
Yersinia entomophaga		
MH96(T)	*Yersinia entomophaga* type strain isolated from diseased grass grub	(Hurst et al., 2011a)
Δ*cbpA*	Deletion mutant derivative MH96 lacking *cpbA* (chitin-binding protein, PL78_05310); Kan^R	This study.
Δ*cspA123* / $\Delta HCUI_{YE96}$	Deletion mutant derivative MH96 lacking *cspA123* (tandem cold-shock proteins A1, A2 and A3; PL78_18365, PL78_17450 and PL78_18370, including intergenic regions) and containing a naturally occurring excision $HCUI_{YE96}$ at tRNA-Asn of 17.5 kb length ; Kan^R	This study.
Δ*fim1*	Deletion mutant derivative MH96 lacking *fim1* (usher chaperone fimbria component protein, PL78_12480); Kan^R	This study.
Δ*rovA*	Deletion mutant derivative MH96 lacking *rovA* (putative transcriptional regulator PL78_05820); $Spec^R$	This study.
Δ*vipB*	Deletion mutant derivative MH96 lacking *vipB* (T6SS contractile sheath component *vipB*, PL78_00910); $Spec^R$	This study.
Δ*yen6*	Deletion mutant derivative MH96 lacking *yen6* (putative regulator, PL78_03730); Kan^R	This study.
Δ*yen7*	Deletion mutant derivative MH96 lacking *yen7* (putative regulator, PL78_03735); $Spec^R$	This study.
Δ*yen67*	Deletion mutant derivative MH96 lacking *yen6* and *yen7* (including intergenic region); $Spec^R$	This study.
Escherichia coli		
DH10B	F^- *mcrA* Δmrr-hsdRMS-mcrBCΔ80d lacZΔM15 ΔlacX74 *endA1 recA1 deoR* Δ(ara,leu)7697 *araD139* galU *galK nupG rpsL* λ^-	(Lorow and Jessee, 1990)
EC100D	F^- *mcrA* Δ(mrr-hsdRMS-mcrBC)ϕ80d *lacZ*ΔM15 Δ*lacX74 recA1 endA1 araD139* Δ(ara,leu)7697 *galU galK* λ^- rpsL *nupG* pir^+ (DHFR)	(Metcalf et al., 1994)
ST18	*pro thi hsd*R^+ Tp^r Sm^r; chromosome::RP4-2 Tc::Mu-Kan::Tn7/λpir ΔhemA	(Thoma and Schobert, 2009)

* This table excludes reporter strains, which are listed in supplementary Table S2.

Table S2: *LacZ*-reporter fusion strains used in this study.

Bacterial strain or plasmid	Description	Reference
Yersinia entomophaga		
MH96 P_{yen6}::*lacZ*	MH96 *cis*-merodiploid yen6-*lacZ* translational reporter strain; Kan^R	This study
MH96 P_{yen7}::*lacZ*	MH96 *cis*-merodiploid *yen7-lacZ* translational reporter strain; Kan^R	This study
MH96 P_{chi1}::*lacZ*	MH96 *cis*-merodiploid *chi1-lacZ* translational reporter strain; Kan^R	This study
Δ*yen6* P_{chi1}::*lacZ*	Δ*yen6 cis*-merodiploid *chi1-lacZ* translational reporter strain; Kan^R $Chlor^R$	This study
Δ*yen7* P_{chi1}::*lacZ*	Δ*yen7 cis*-merodiploid *chi1-lacZ* translational reporter strain; $Spec^R$ Kan^R	This study
Δ*yen67* P_{chi1}::*lacZ*	Δ*yen67 cis*-merodiploid *chi1-lacZ* translational reporter strain; $Spec^R$ Kan^R	This study

Table S3: Cloning and suicided plasmids used for targeted mutagenesis

Plasmid	Description	Source
pGEM	pGEM-T Vector, Amp^R, *lacZ* multicloning site (MCS)	Promega Ltd.
pGEM-*cbpA*_KO	Homologous recombination construct targeting *cbpA* in MCS of pGEM, Amp^R, Kan^R	This study.
pGEM-*cspA123*_KO	Homologous recombination construct targeting *cspA123* in MCS of pGEM, Amp^R, Kan^R	This study.
pGEM-*fim1*_KO	Homologous recombination construct targeting *fim1* in MCS of pGEM, Amp^R, Kan^R	This study.
pGEM-*vipB*	Chromosomal region from MH96 containing *vipB* in MCS of pGEM, Amp^R, $Spec^R$	This study.
pGEM-*vipB*_KO	Homologous recombination construct targeting *vipB* in MCS of pGEM, Amp^R, $Spec^R$	This study.
pGEM-*rovA*	Chromosomal region from MH96 containing *rovA* in MCS of pGEM, Amp^R, $Spec^R$	This study.
pGEM-*rovA_KO*	Homologous recombination construct targeting *rovA* in MCS of pGEM, Amp^R, $Spec^R$	This study.
pGEM-*yen6*_KO	Homologous recombination construct targeting *yen6* in MCS of pGEM, Amp^R, Kan^R	This study.
pGEM-*yen7*_KO	Homologous recombination construct targeting *yen7* in MCS of pGEM, Amp^R, $Spec^R$	This study.
pGEM-*yen67*_KO	Homologous recombination construct targeting *yen67* in MCS of pGEM, Amp^R, $Spec^R$	This study.
pJP5608	Suicide plasmid; mob(RP4), *lacZ* MCS, Tet^R	(Penfold and Pemberton, 1992)
pJP5603	Suicide plasmid; mob(RP4), *lacZ* MCS, Kan^R	(Penfold and Pemberton, 1992)
pJP5608-*cbpA*_KO	Homologous recombination construct targeting *cbpA* in MCS of pJP5608, Kan^R, Tet^R	This study.
pJP5608-*cspA123*-_KO	Homologous recombination construct targeting *cspA123* in MCS of pJP5608, Kan^R, Tet^R	This study.
pJP5608-*fim1*_KO	Homologous recombination construct targeting *fim1* in MCS of pJP5608, Kan^R, Tet^R	This study.
pJP5608-*vipB*_KO	Homologous recombination construct targeting *vipB* in MCS of pJP5603, $Spec^R$, Kan^R	This study.
pJP5608-*rovA*_KO	Homologous recombination construct targeting *rovA* in MCS of pJP5603, $Spec^R$, Kan^R	This study.
pJP5608-*yen6*_KO	Homologous recombination construct targeting *yen6* in MCS of pJP5608, Kan^R, Tet^R	This study.

Plasmid	Description	Source
pJP5608-*yen7*_KO	Homologous recombination construct targeting *yen7* in MCS of pJP5603, $Spec^R$, Kan^R	This study.
pJP5608-*yen67*_KO	Homologous recombination construct targeting *yen67* pJP5603, $Spec^R$, Kan^R	This study.

Table S4: Arabinose/IPGT induction, *lacZ*-reporter fusion and complementation plasmids used in this study*

Plasmid	Description	Source
pBAD	pAY2-4; a pBAD18 derivative ; arabinose inducible expression vector with modified poly-linker, Amp^R	(Shaw et al., 2003)
pBAD-*yen6*	Expression vector for arabinose induction of *yen6* (protein-coding region and 32 bp 3'UTR region). Amp^R	This study.
pBAD-*yen67as*	Expression vector for arabinose induction of *yen6*, including a 478 bp 3'UTR containing putative terminator and *cis*-acting anti-sense RNA. Amp^R	This study.
pBAD-Amb2	Expression vector for arabinose induction of *yen6*, including a 67 bp 3'UTR. Amp^R	This study.
pBAD-Amb3	Expression vector for arabinose induction of *yen6*, including a 148 bp 3'UTR containing putative terminator. Amp^R	This study.
pBAD-*yen7*	Expression vector for arabinose induction of *yen7* (protein-coding region only). Amp^R	This study.
pACYC184	Mid-copy cloning vector; $Chlor^R$, Tet^R.	(Chang and Cohen, 1978)
pACYC184-P_{yen6}-P_{yen7}	Mid-copy cloning vector containing entire *yen6* and *yen7* promoter and coding regions interrupting the tetracycline cassette; $Chlor^R$.	This study.
pVIK107kn	Integrative vector for *cis*-merodiploid mutants; Kan^R	(Kalogeraki and Winans, 1997)
pVIK107cm	pVIK107kn derivative with Kan^R interrupted with ChlorR at the *AgeI* site; $Chlor^R$	This study.
pVIK107kn-P_{yen6}::*lacZ*	*yen6* promoter region (1,006 bp) cloned into MCS of pVIK107kn for chromosomal integration of *yen6-lacZ* translational reporter; Kan^R	This study.
pVIK107kn-P_{yen7}::*lacZ*	*Yen7* promoter region cloned into MCS of pVIK107kn for chromosomal integration of *yen7-lacZ* translational reporter; Kan^R	This study.
pVIK107kn-P_{chi1}::*lacZ*	*Chi1* promoter region cloned into MCS of pVIK107kn for chromosomal integration of *chi1-lacZ* translational reporter; Kan^R	This study.

Plasmid	Description	Source
pVIK107cm-P*chi1*::*lacZ*	*Chi1* promoter region cloned into MCS of pVIK107cm, for chromosomal integration of *chi1-lacZ* translational reporter; Chlor^R	This study.
pOPINF	In-Fusion™ PCR cloning vector for 6X-His tagged constructs.	(Berrow, Nick et al., 2009)
pOPINF-*yen6*$_{6x\text{-}His}$	pOPINF with coding region of *yen6* cloned into *kpnI* and *hindIII*, 3′ and 5′ infusion sites, respectively. Amp^R	Marion Schoof (Callaghan Innovation)

Table S5: Primers used for targeted mutagenesis

Target	Primer name	Sequence (5' → 3')*	Cut site
(Selectable marker)			
pACYC177	KanF	TCTGCTGACGCACCGGTGCAG	
pACYC177	KanR	CCGTCCCGTCAAGTCAGCGT	
pHP45	SpecF	ACCCTCACTGATCCGCATGCC	*ClaI* & *BgIII*
pHP45	SpecR	GTGCTTAGTGCATCTAACGCT	
(Mutagenesis target – restriction)			
rovA	rovAF	GCCATACTTTTATCTATACCCGAAAT	*ClaI*
	rovAR	ATTTCATCTCAATTTACCAACTGACT	
	rovA_valF†	GCTATCGCTGCGGTTACTCT	
	rovA_valR†	AACAAGCCATGAAAACCGCC	
vipB			*BglII*
	vipB_F	GTGCGTCGCATTGACTCATC	
	vipB_R	TACCGAACTCTGGCAACCAC	
	vipB_valF†	TCATGTGGAAATAACCTGCCTGA	
	vipB_valR†	TTGAACTCCACCAGATCGCC	
(Mutagenesis target – overlap-extension)			
cbpA	CBP_P1†	CCAAGCTGACTACTGCCGAA	
	CBP_P3_kan	CATCAACCGTGGCTCCCTCACCGGC CATGTCTCAGTTGTGA	
	CBP_P4_kan	ACTGGCAGAGCATTACGCTGACCCA TGGGTGATGCTTGACT	
	CBP_P6†	CGTCATCTTATTTGCCGGCG	
	CBP_P2	TCCGAGGATTTTGAGCGCAA	*EcoR1*‡
	CBP_P5	GGGTTTTTCGACTTCTGCCG	
cspA123			
	CSP_P1†	TGGACAAAACCAGTCACCGA	
	CSP_P3_kan	CATCAACCGTGGCTCCCTCACAAAC CAAAGCCTTTACCCGC	
	CSP_P4_kan	ACTGGCAGAGCATTACGCTGACGGC CAATGTTGTTGTCCTG	
	CSP_P6	CTCAGATAGCGCCGTCTGTT	
	CSP_P2	AAGAAGTAACCTGGCACCGC	*EcoR1*‡
	CSP_P5	TGCCCTGGAAAAACGCCTTA	
	CSP_valR†	TTGCCGGATGTTCAAGCTCT	
fim1			
	UCF_P1†	TCTTCAGTGCCGATGATGCA	

Target	Primer name	Sequence (5' → 3')*	Cut site
	UCF_P3_kan	CATCAACCGTGGCTCCCTCACAATA ACGACACTGGGCCACA	
	UCF_P4_kan	ACTGGCAGAGCATTACGCTGATTAA GGGCGCAATAGCAGCA	
	UCF_P6†	CGACTATTTGCCTTTCCGCG	
	UCF_P2	AAAGAGCTCGACGTTAGCTGGCTAT CGCT	*SacI*
	UCF_P5	AAAGAGCTCTTGCTTAAACGCGTCT TGGC	
yen6			
	Yen6_P1†	CATCCAAGAGTCATCCCGCA	
	Yen6_P3_kan	CATCAACCGTGGCTCCCTCACACCT CCTTATCAAATTTGCCGA	
	Yen6_P4_kan	ACTGGCAGAGCATTACGCTGATGAG CAAGGTGGGGTAAAGC	
	Yen6_P6†	CGTCATCTTATTTGCCGGCG	
	Yen6_P2	CGTCGTTTCACATCGGATGC	*EcoR1*‡
	Yen6_P5	ATCAAACTGCTCACTACTGATACT	
yen7			
	Yen7_P1†	CCGGCAGTGGATTAGGCTTA	
	Yen7_P3_spec	CGGTTTTCTGGAAGGCGAGCACCCT CCTCCAGCGCTTATTT	
	Yen7_P4_spec	TCAAGCCGACGCCGCTTCGCGCCAC TTTGTCGTCAATTGCCA	
	Yen7_P6†	GCGCTTCTCCCACCAGTATC	
	Yen7_P2	TGACCACGAGTTTTAGCGCT	*EcoR1*‡
	Yen7_P5	TCCTTGTTGATTATTCTGGGGTCA	
yen67			
	Yen6_P1†	CATCCAAGAGTCATCCCGCA	
	Yen6_P3_spec	CGGTTTTCTGGAAGGCGAGCAACCT CCTTATCAAATTTGCCGA	
	Yen7_P4_spec	TCAAGCCGACGCCGCTTCGCGCCAC TTTGTCGTCAATTGCCA	
	Yen7_P6†	GCGCTTCTCCCACCAGTATC	
	Yen6_P2	CGTCGTTTCACATCGGATGC	*EcoR1*‡
	Yen7_P5	TCCTTGTTGATTATTCTGGGGTCA	

* underscore denotes the overlap sequence for the selectable marker. Double-underscore represents restriction site. †used for validation sequencing. ‡*EcoR1* site from pGEM-T multiple cloning site.

Supplementary section for Chapter 4

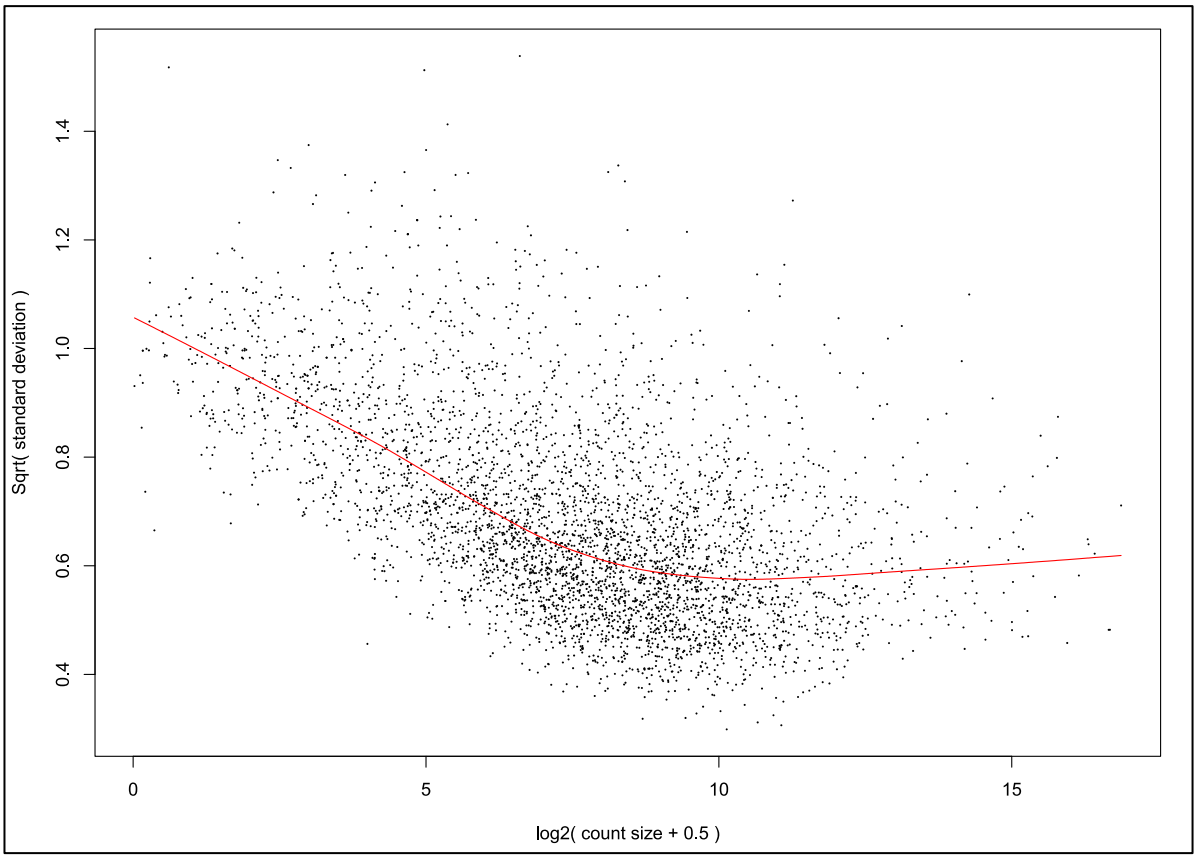


Figure S1. Normalized count data from *Yersinia entomophaga* MH96 *in vivo* transcriptome with transcript-wise mean-variance relationship modelled using voom . Red line represents lowess fit.

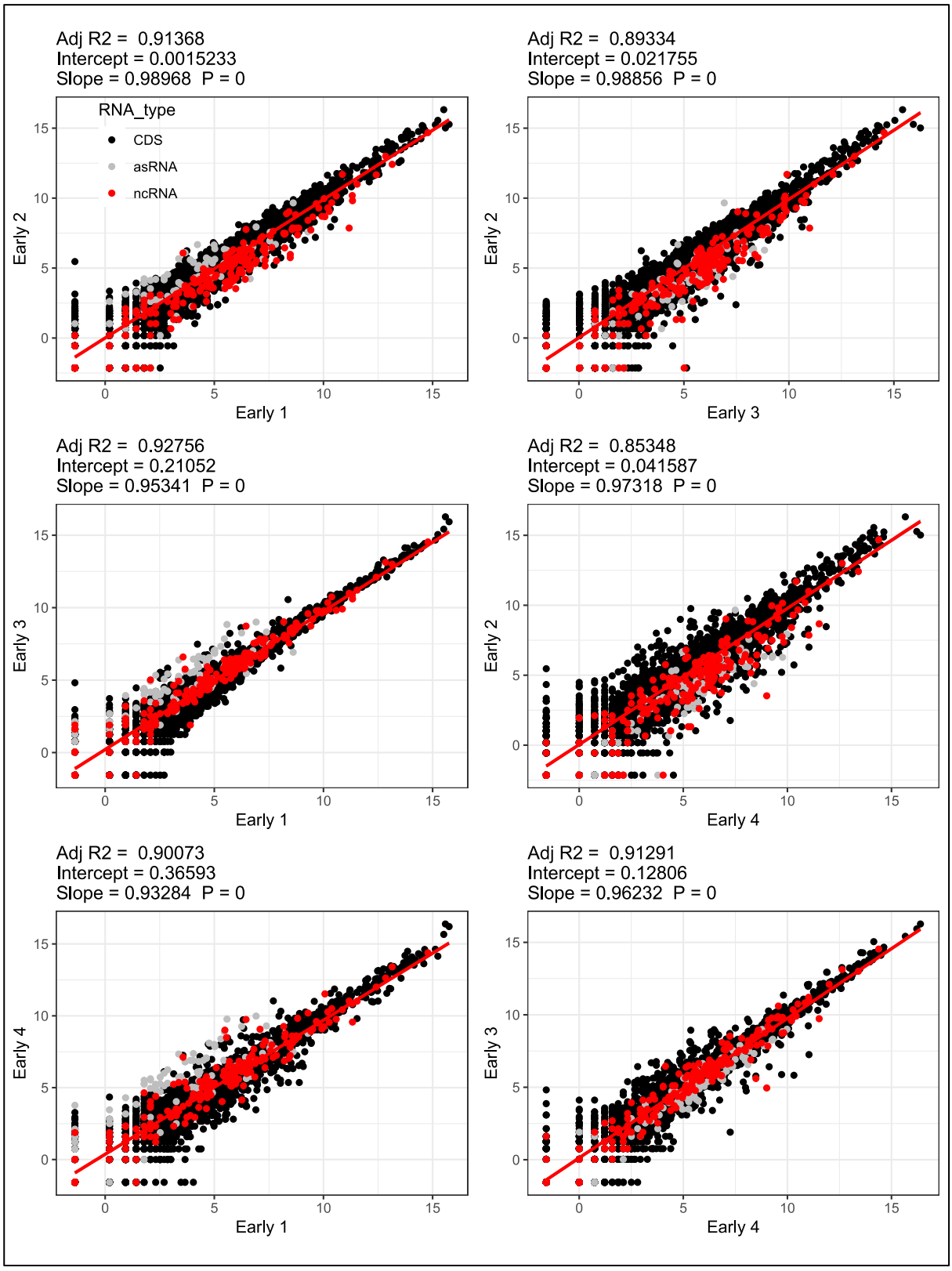


Figure S2. Regression of *in vivo* RNA-seq libraries from *Yersinia entomophaga* MH96 during intrahemocoelic infection of *Galleria mellonella* at 25 °C during lag phase (~1 x 10^7 CFU/g).

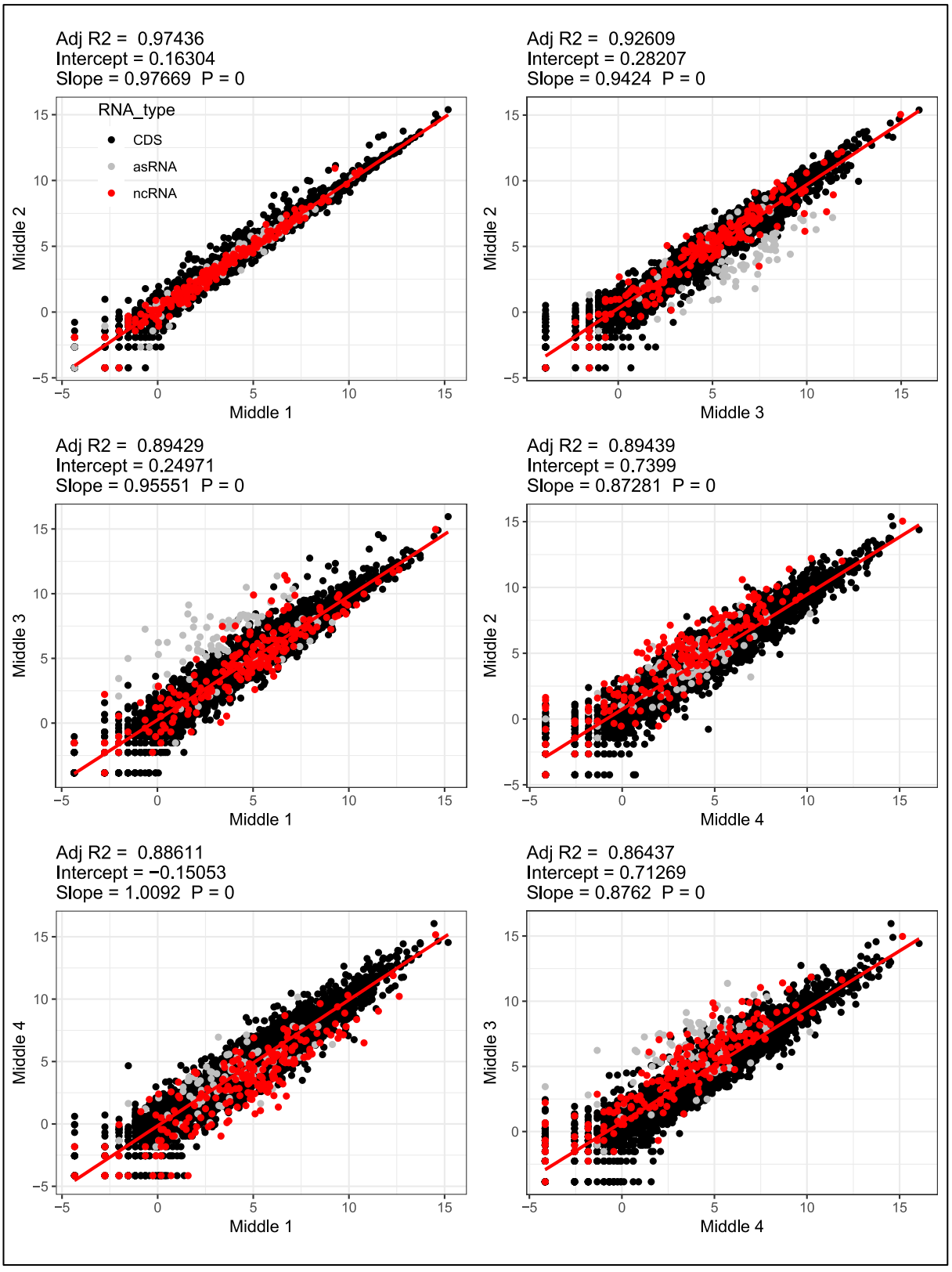


Figure S3. Regression of *in vivo* RNA-seq libraries from *Yersinia entomophaga* MH96 during intrahemocoelic infection of *Galleria mellonella* at 25 °C during exponential phase (~1 x 10^8 CFU/g).

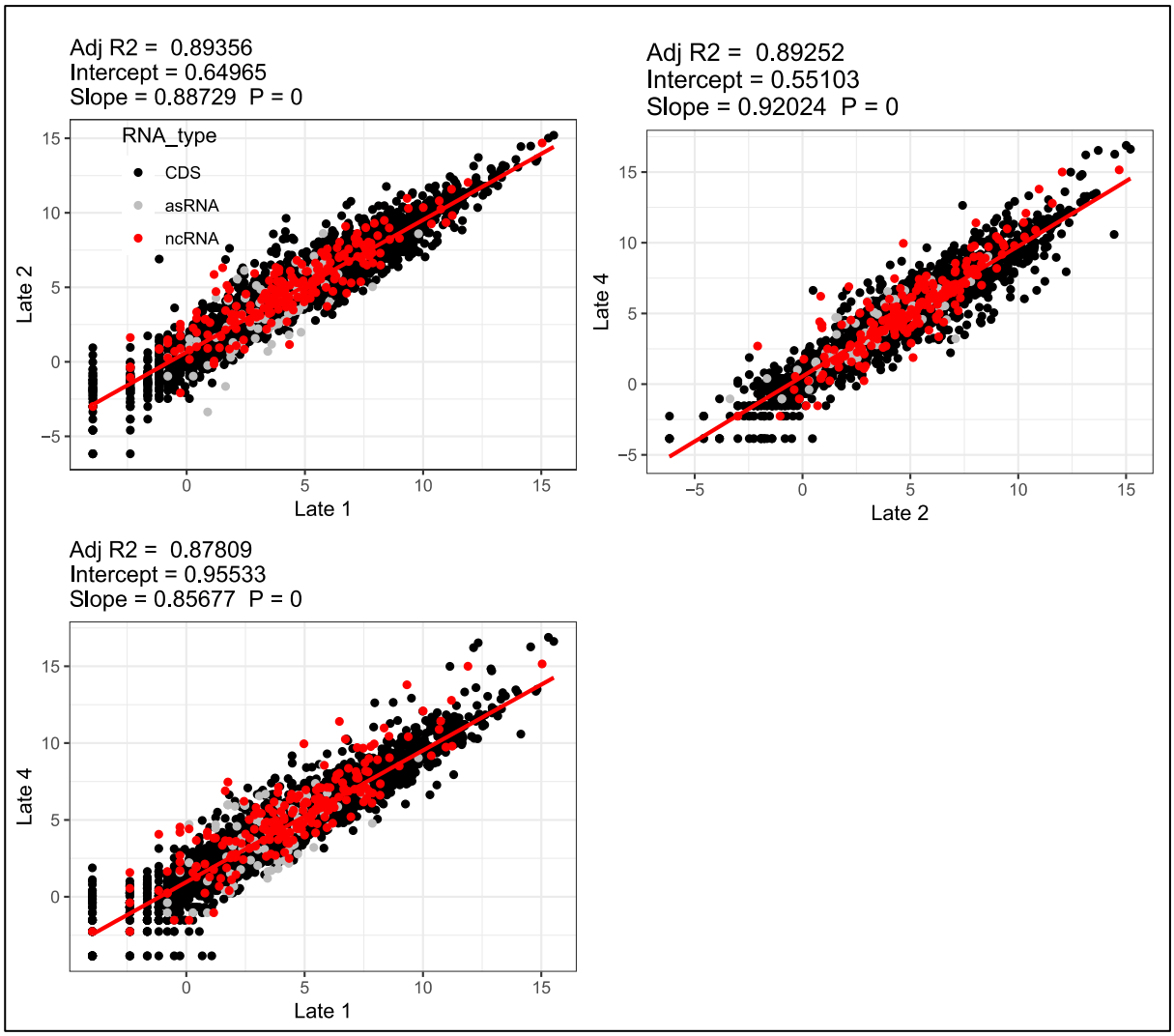


Figure S4. Regression of *in vivo* RNA-seq libraries from *Yersinia entomophaga* MH96 during intrahemocoelic infection of *Galleria mellonella* at 25 °C during stationary phase (~1 x 10^9 CFU/g).

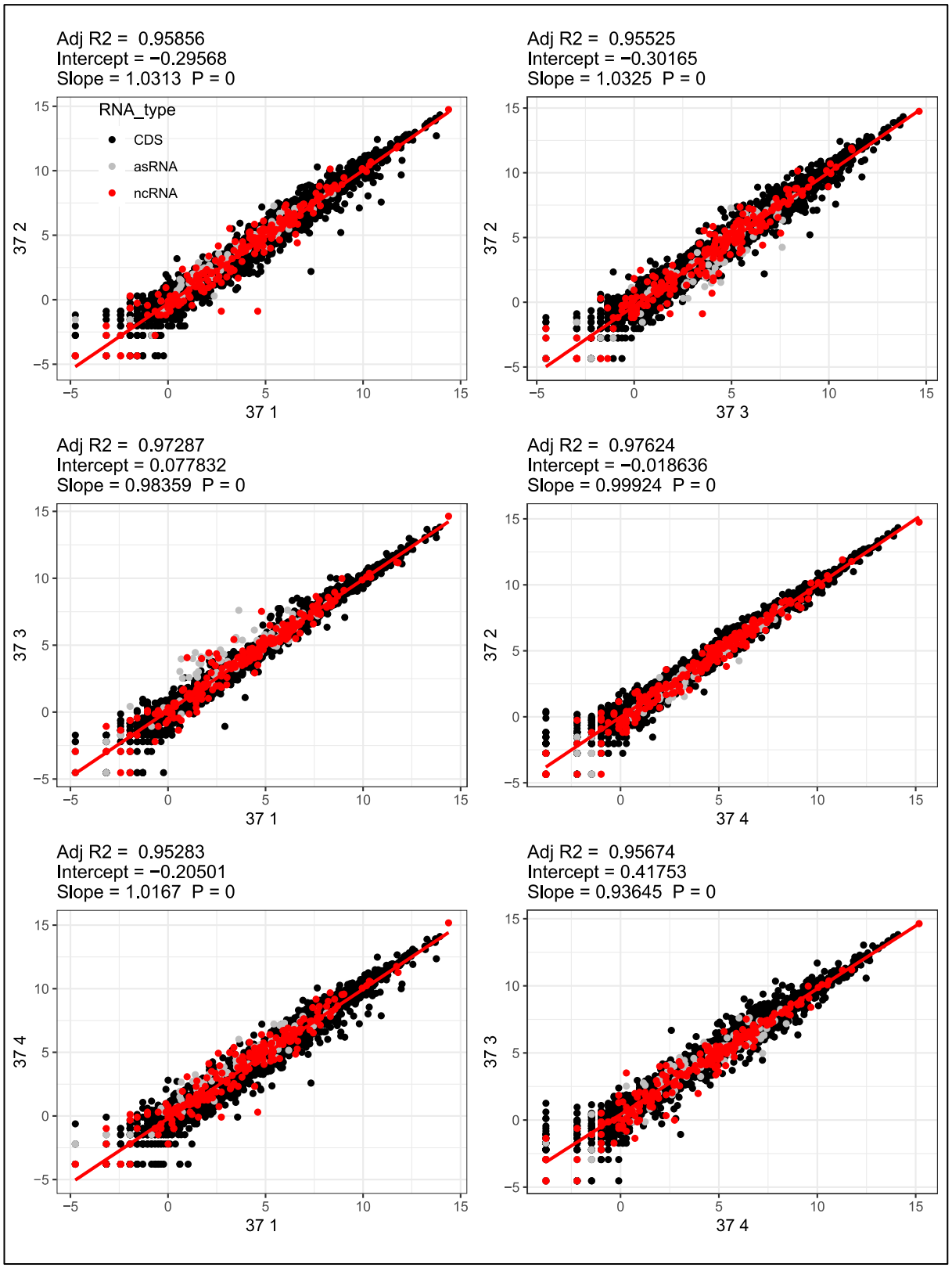


Figure S5. Regression of *in vivo* RNA-seq libraries from *Yersinia entomophaga* MH96 during intrahemocoelic infection of *Galleria mellonella* at 37 °C during exponential phase (~1 x 10^8 CFU/g).

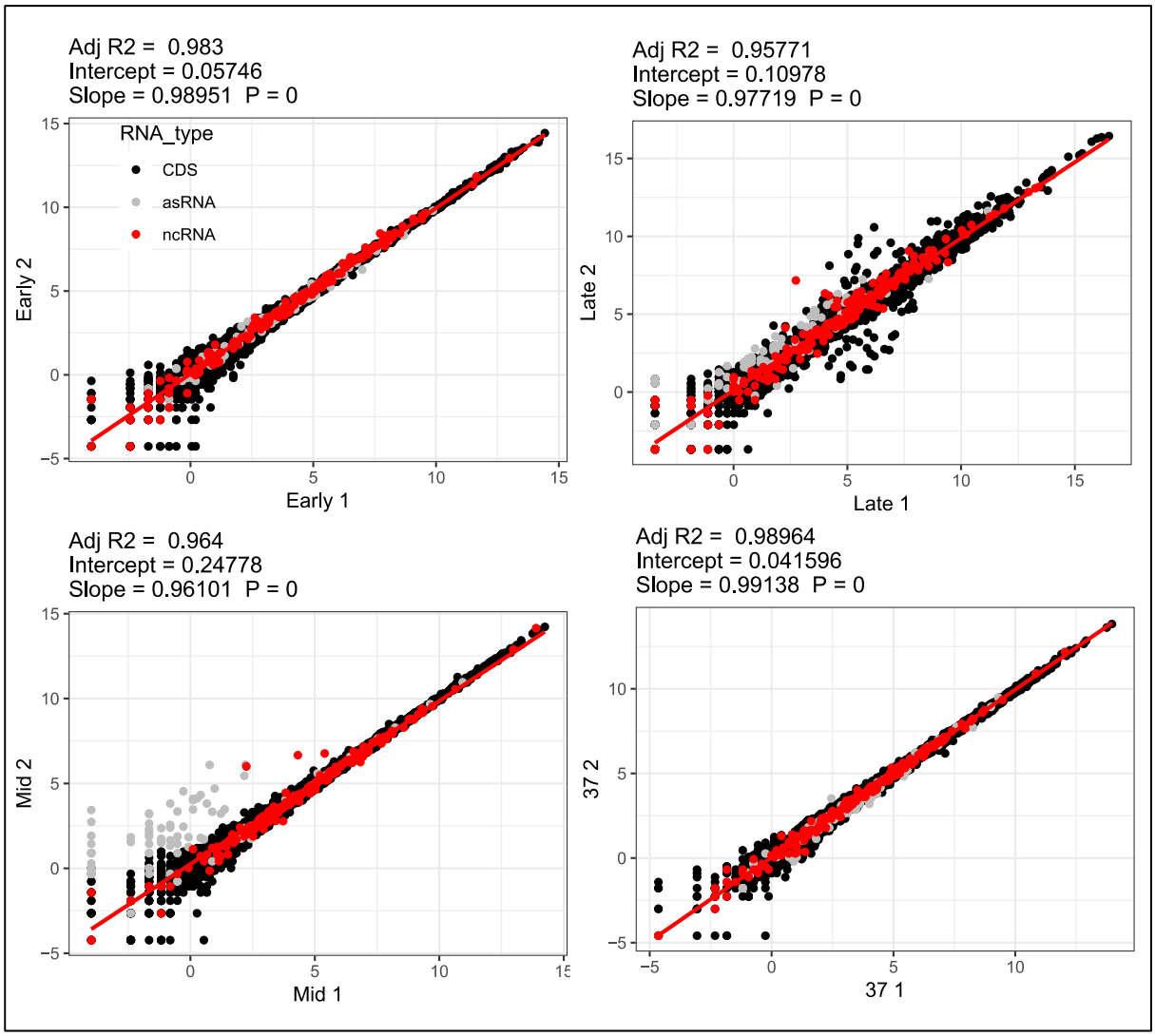


Figure S6. Regression of *in vitro* RNA-seq libraries from *Yersinia entomophaga* MH96 during culture in 50 ml LB broth at 25 or 37 °C during lag, exponential and stationary phases (i.e., 1 x 10^7, 10^8, 10^9 CFU/ml, respectively). Early = lag, Mid = exponential, Late = stationary and 37 = exponential at 37 °C.

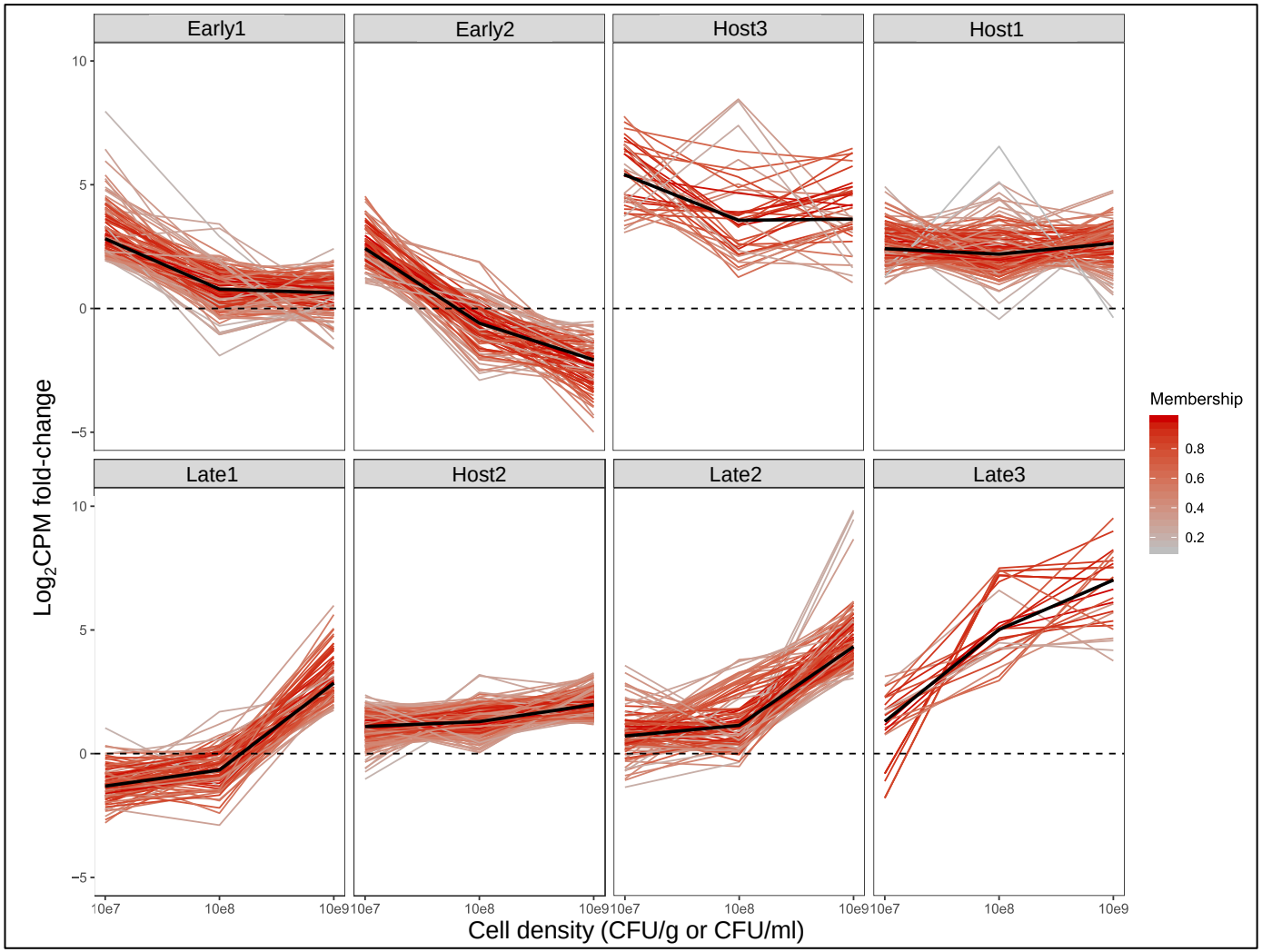


Figure S7: Host-specific transcripts from *Yersinia entomophaga* MH96 *in vivo* RNA-seq identified using c-means fuzzy clustering algorithm (k = 20, m = 1.5). Log_2 counts-per-million (CPM) fold-change between *in vivo* and *in vitro* samples for each cell density. Black line = median log_2 CPM fold-change. CFU = colony forming units.

Table S6: Adapter and barcode sequences used to trim all *in vivo* samples and *in vitro* samples "F22-1" and "F22-2" (exponential growth phase 25 °C).

Adapters_1.fasta

Table S7: Adapter and barcode sequences used to trim all *in vitro* samples, except "F22-1" and "F22-2" (exponential growth phase 25 °C).

Adapters_2.fasta

Table S8. In-host putative virulence factors identified in *Yersinia entomophaga* MH96 using fuzzy clustering and virulence annotations from VFDB.

Supplementary_table_VFDB_annotation

Table S9. Top one-hundred most highly expressed genes by *Yersinia entomophaga* MH96 at 37 °C compared to 25 C *in vivo*.

Supplementary_table_DE_37v25invivo.xlsx

Table S10. Top one-hundred most highly expressed genes by *Yersinia entomophaga* MH96 at 37 °C *in vivo* compared to *in vitro*

Supplementary_tables_DE_invivo_vs_invi

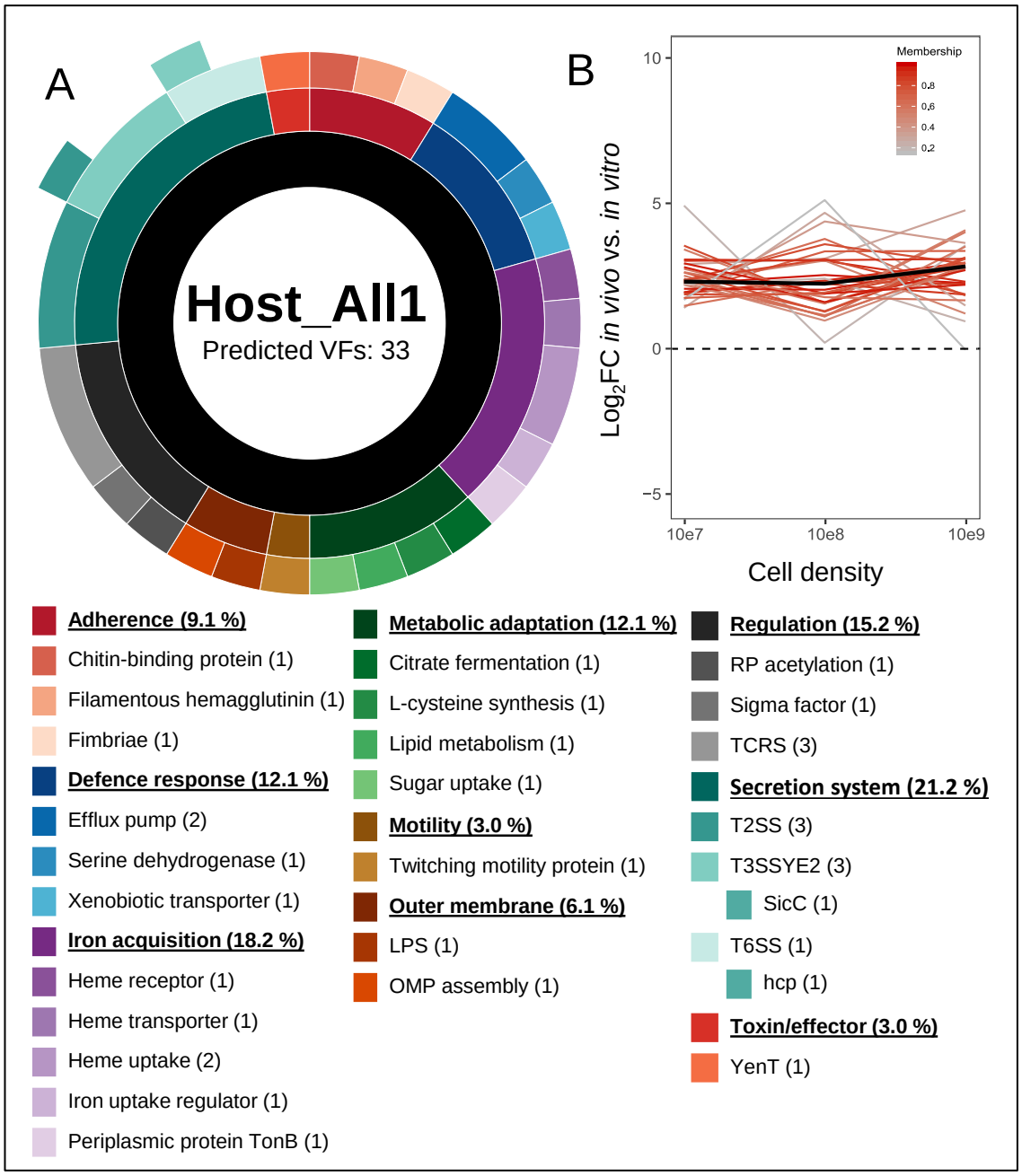


Figure S8: Functional enrichment of Host_All1 cluster of putative virulence factors. (A) Proportion of genes organized by functional enrichment categories. LPS = Lipopolysaccharide, OMP = outer-membrane protein, RP = ribosomal protein, TCRS = two-component regulatory system, T2SS = type II secretion system, T3SSYE2 = type 3 secretion system YE2 from MH96. Numbers in brackets indicate the number of VFs within each sub-category. (B) Log_2 fold-change of normalized counts-per-million expression between *in vivo* and *in vitro* samples for the functionally annotated genes in this cluster. Black line = median log_2 CPM fold-change.

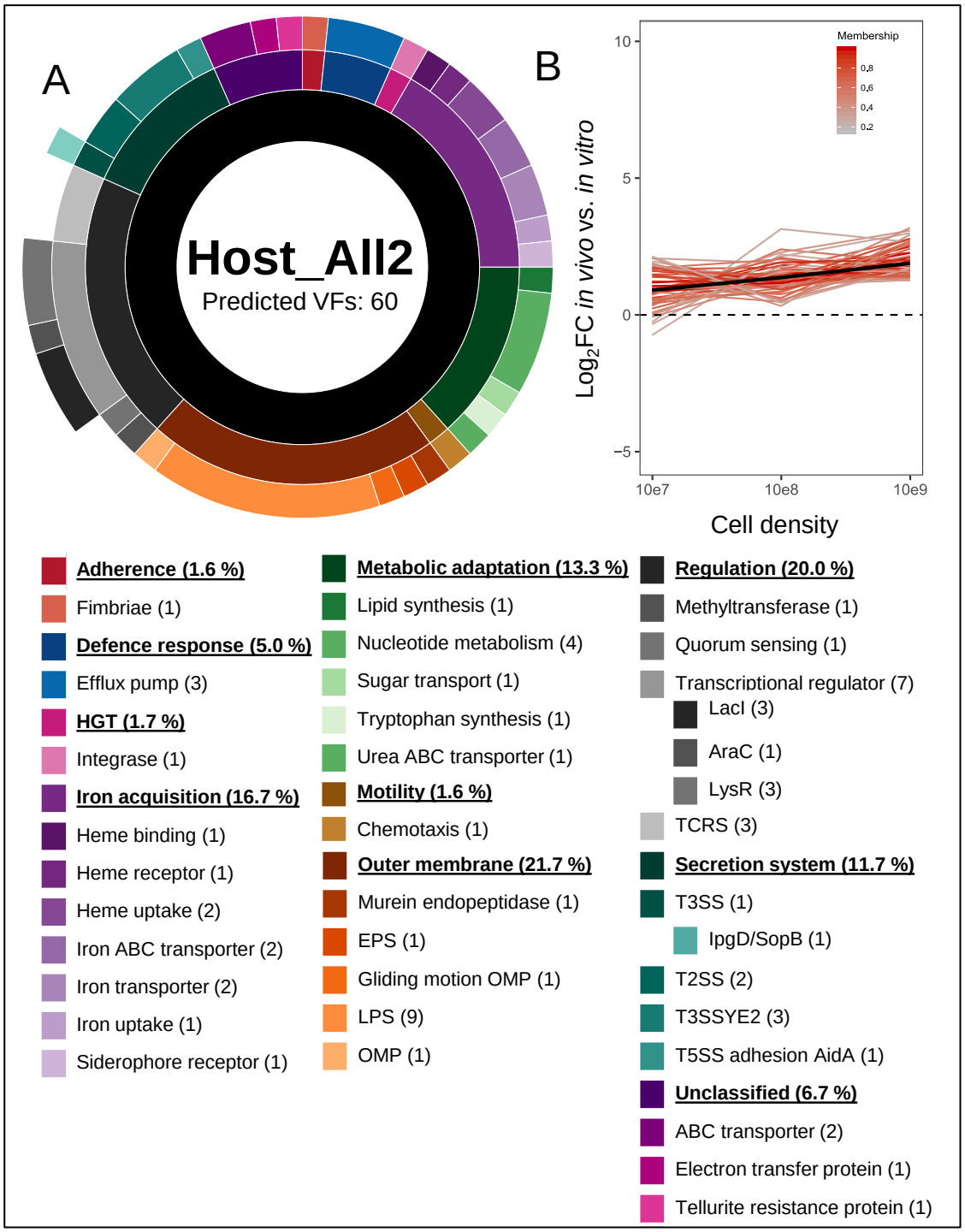


Figure S9: Functional enrichment of Host_All2 cluster of putative virulence factors. (A) Proportion of genes organized by functional enrichment categories. ABC = ATP-binding cassette, EPS = extracellular polysaccharide, OMP = outer-membrane protein, LPS = lipopolysaccharide, TCRS = two-component regulatory system, T3SS = type III secretion system, T2SS = type II secretion system, T3SSYE2 = type III secretion systems encoded by MH96, T5SS = type V secretion system. Numbers in brackets indicate the number of VFs within each sub-category. (B) Log_2 fold-change of normalized counts-per-million expression between *in vivo* and *in vitro* samples for the functionally annotated genes in this cluster. Black line = median log_2 CPM fold-change.

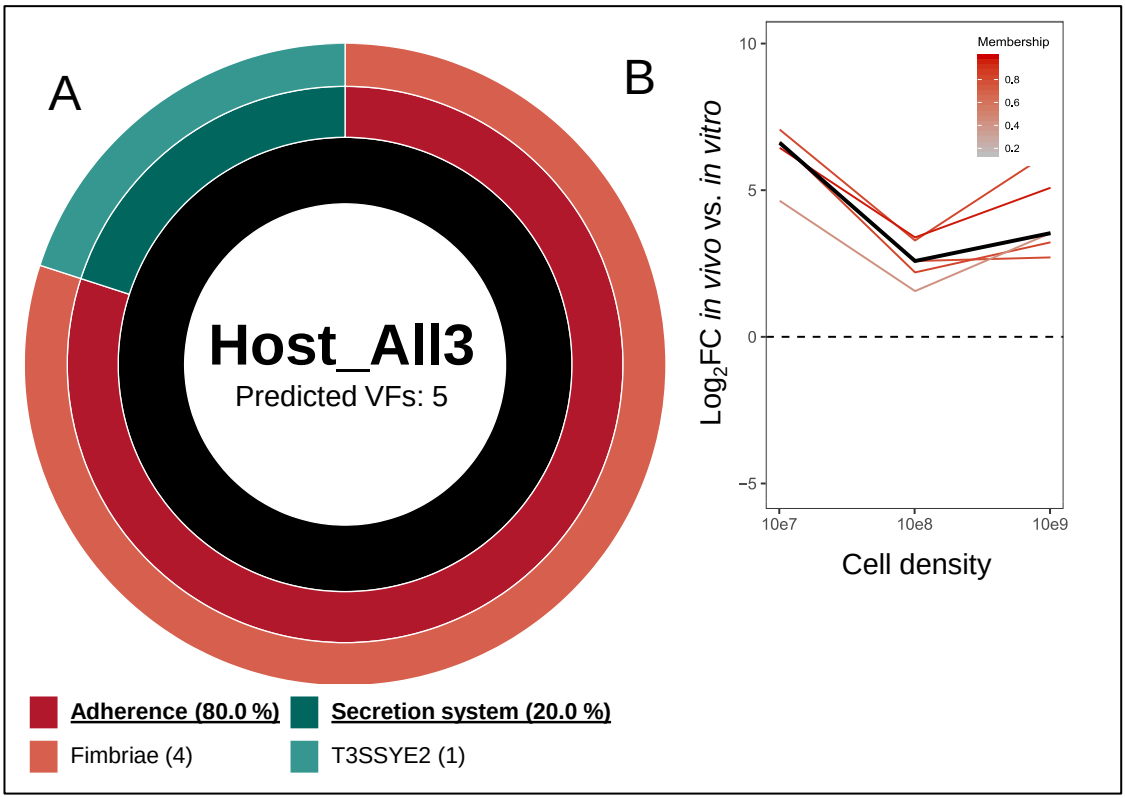


Figure S10: Functional enrichment of Host_All3 cluster of putative virulence factors. (A) Proportion of genes organized by functional enrichment categories. T3SSYE2 = type III secretion system encoded by MH96. Numbers in brackets indicate the number of VFs within each sub-category. (B) Log_2 fold-change of normalized counts-per-million expression between *in vivo* and *in vitro* samples for the functionally annotated genes in this cluster. Black line = median log_2 CPM fold-change.

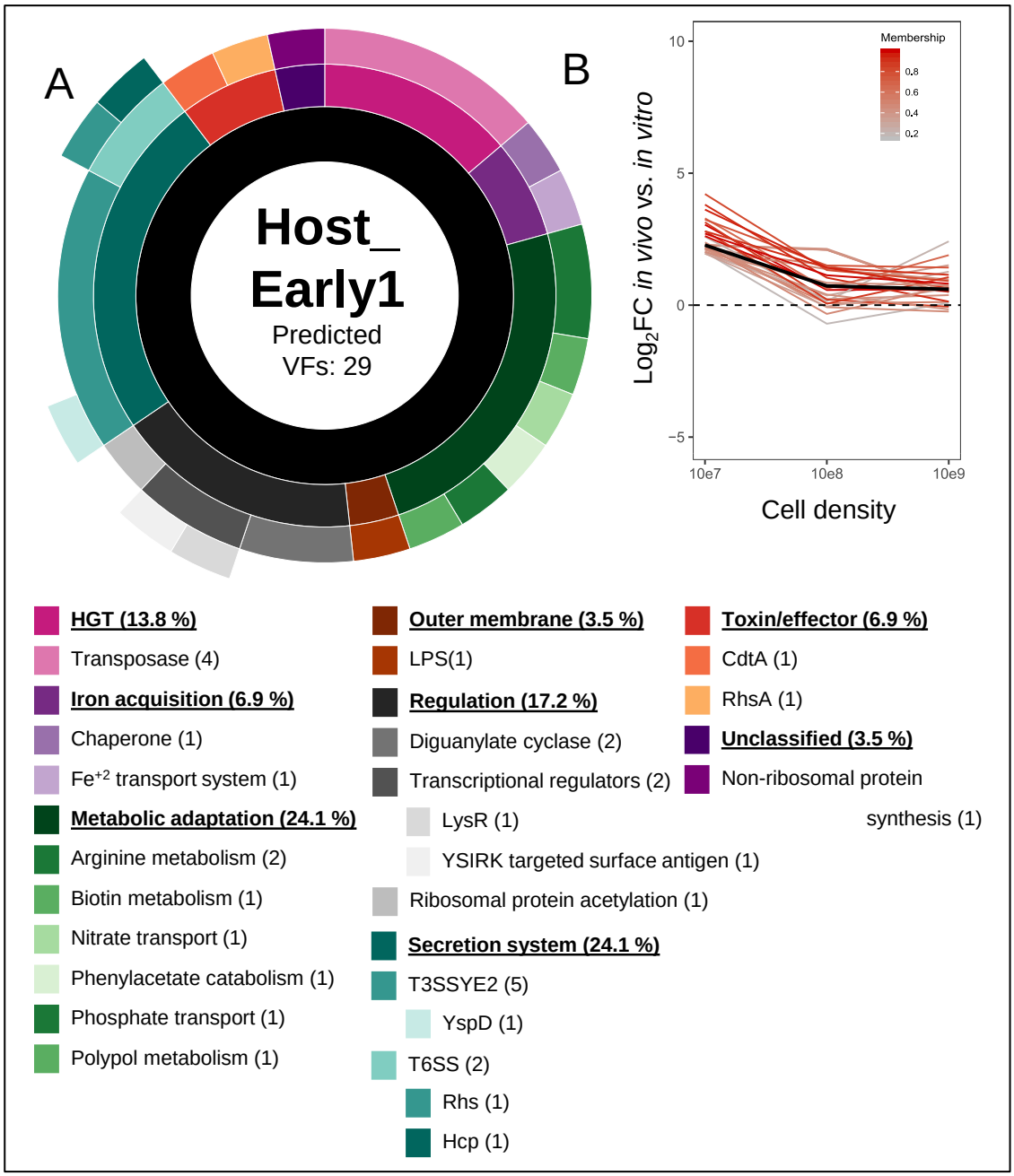


Figure S11: Functional enrichment of Host_Early1 cluster of putative virulence factors. (A) Proportion of genes organized by functional enrichment categories. LPS = lipopolysaccharide, T3SSYE2 = type III secretion system encoding by MH96, T6SS = Type VI secretion system. Numbers in brackets indicate the number of VFs within each sub-category. (B) Log_2 fold-change of normalized counts-per-million expression between *in vivo* and *in vitro* samples for the functionally annotated genes in this cluster. Black line = median log_2 CPM fold-change.

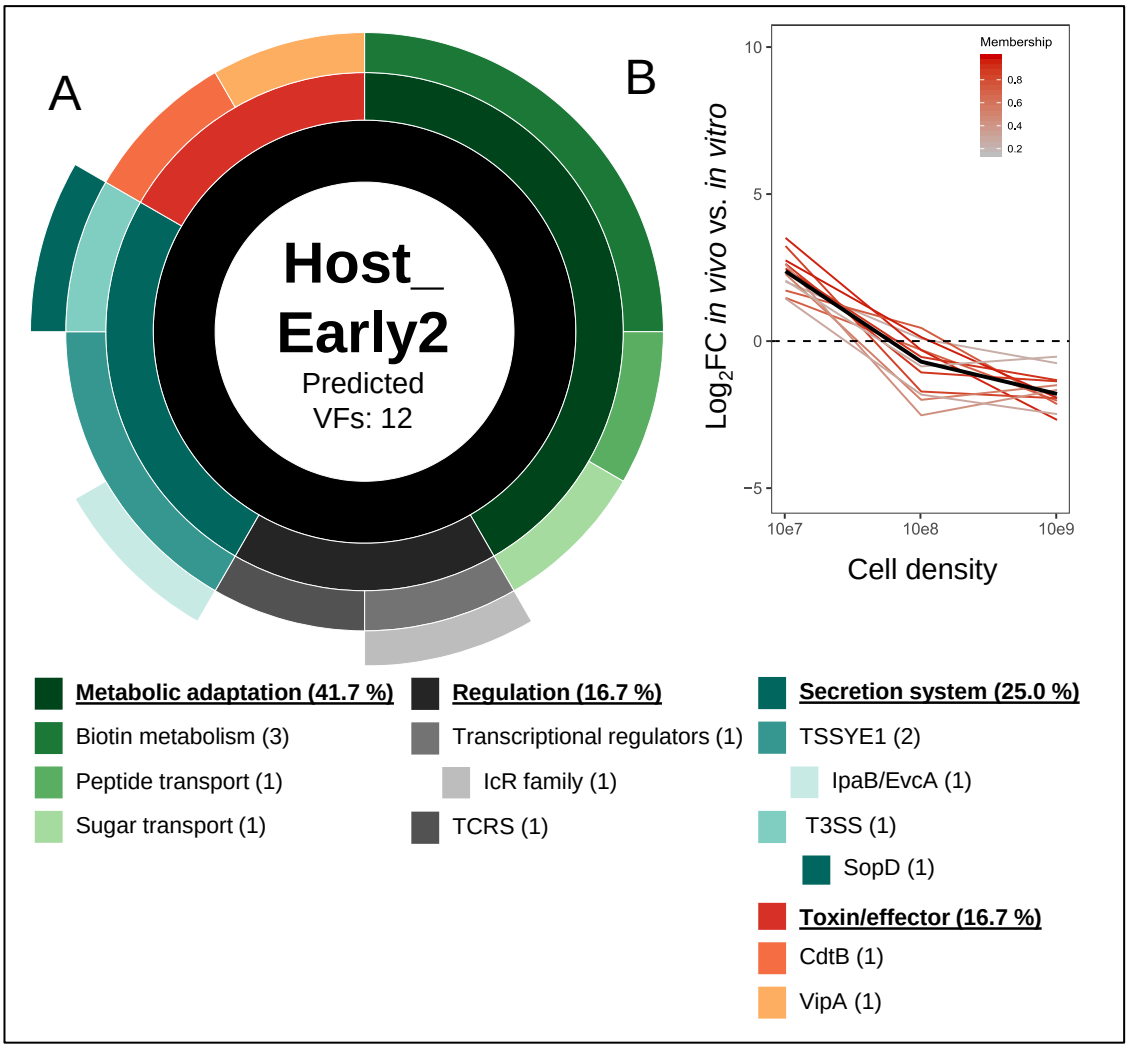


Figure S12: Functional enrichment of Host_Early2 cluster of putative virulence factors. (A) Proportion of genes organized by functional enrichment categories. TCRS = two-component regulatory system, T3SSYE1 = type III secretion system encoded by MH96, T3SS = Type III secretion system. Numbers in brackets indicate the number of VFs within each sub-category. (B) Log_2 fold-change of normalized counts-per-million expression between *in vivo* and *in vitro* samples for the functionally annotated genes in this cluster. Black line = median log_2 CPM fold-change.

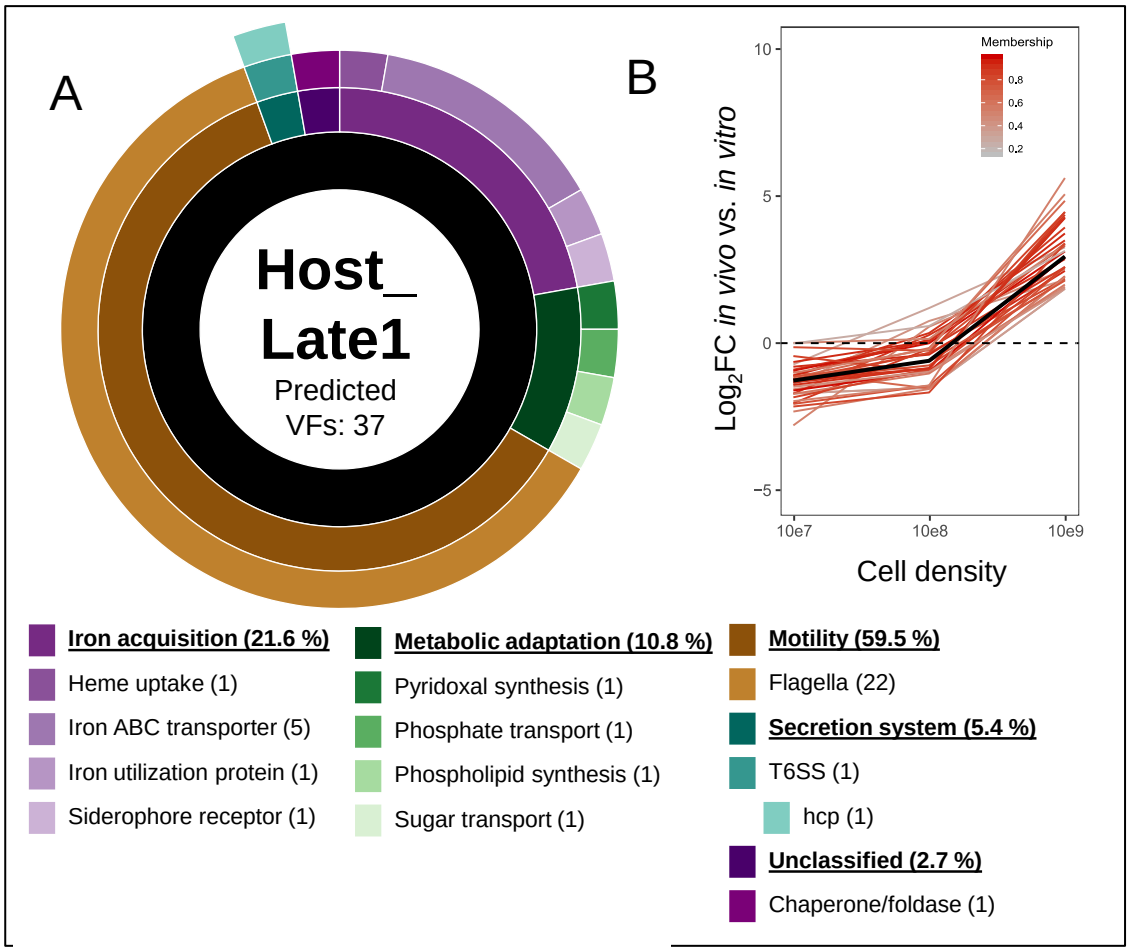


Figure S13: Functional enrichment of Host_Late1 cluster of putative virulence factors. (A) Proportion of genes organized by functional enrichment categories. T6SS = type VI secretion system. Numbers in brackets indicate the number of VFs within each sub-category. (B) Log_2 fold-change of normalized counts-per-million expression between *in vivo* and *in vitro* samples for the functionally annotated genes in this cluster. Black line = median log_2 CPM fold-change.

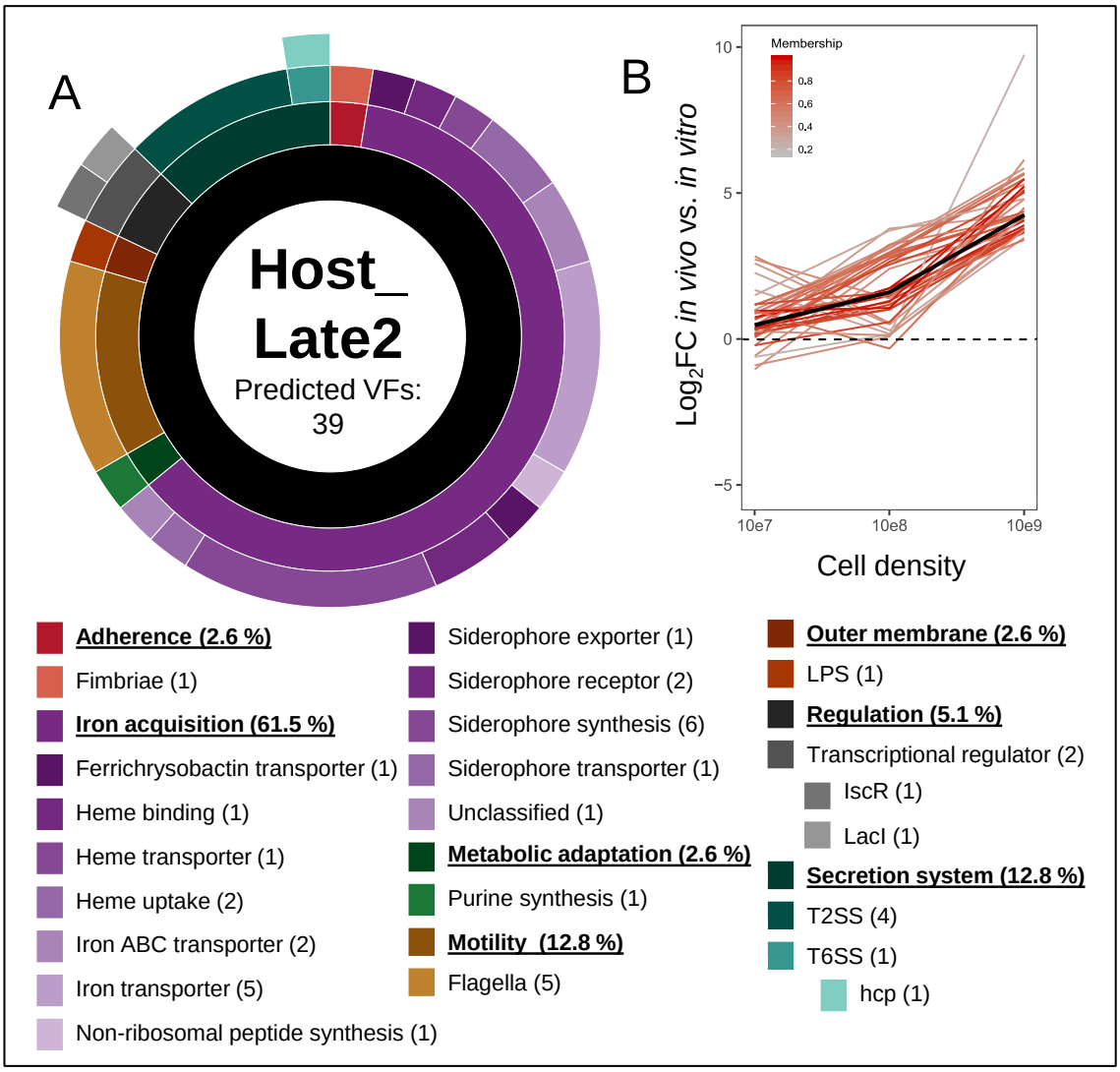


Figure S14: Functional enrichment of Host_Late2 putative virulence factors. (A) Proportion of genes organized by functional enrichment categories. ABC = ATP-binding cassette, LPS = lipopolysaccharide, T2SS = type II secretion system, T6SS = type VI secretion system. Numbers in brackets indicate the number of VFs within each sub-category. (B) Log_2 fold-change of normalized counts-per-million expression between *in vivo* and *in vitro* samples for the functionally annotated genes in this cluster. Black line = median log_2 CPM fold-change.

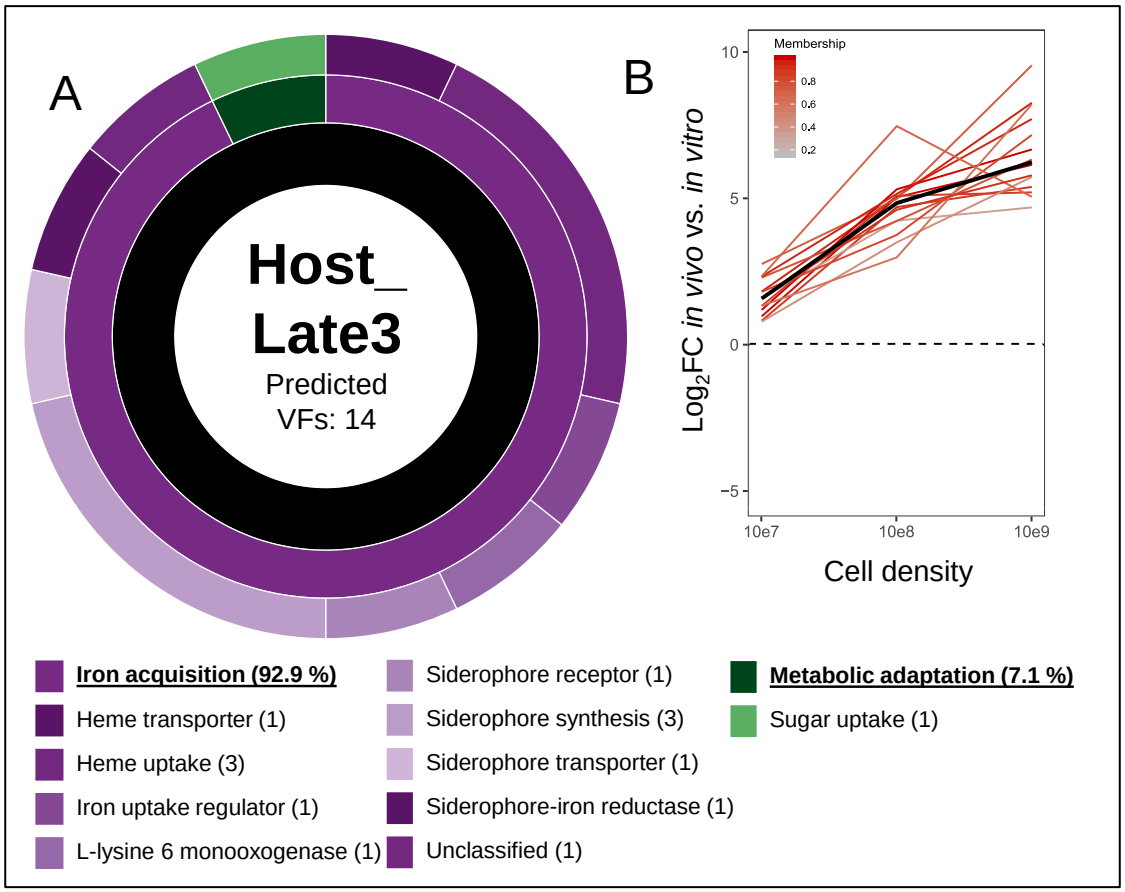


Figure S15: Functional enrichment of Host_Late3 putative virulence factors. (A) Proportion of genes organized by functional enrichment categories. Numbers in brackets indicate the number of VFs within each sub-category. (B) Log_2 fold-change of normalized counts-per-million expression between *in vivo* and *in vitro* samples for the functionally annotated genes in this cluster. Black line = median log_2 CPM fold-change.

Table S11. Top one-hundred transcripts with longest predicted 5′ and 3′ UTRs in the transcriptome of *Yersinia entomophaga* MH96.

Supplementary Table _long 5'_3'_UT

Source Code 1 – R source code for *in vivo* RNA-seq analysis

```
##load libraries
library("GenomicFeatures")
library("stringi")
library("EDASeq")
library("RUVSeq")
library("lattice")
library("edgeR")
library("limma")
library("statmod")
library("EDASeq")

##assign color from RColorBrewer
library(RColorBrewer)
colors <- brewer.pal(12, "Paired")

##read in raw count data, row names are locus tags
seqdata <- read.delim("count_data.txt", header=TRUE, row.names=1)

##filter non-expressed genes (requires at least 5 reads in at least two samples to be included)
filter <- apply(seqdata, 1, function(x) length(x[x>5])>=2)
seqdata_filtered <- seqdata[filter, ]

##select all filtered rows
sel.rows <- row.names(seqdata_filtered)

##assign treatment/collection batch as factors
myx <- as.factor(c("10^7 LB","10^7 LB","10^8 LB","10^8 LB","10^9 LB","10^9 LB","37 LB","37 LB","10^7 Galleria.A","10^7 Galleria.A","10^7 Galleria.A","10^7 Galleria.B","10^8 Galleria","10^8 Galleria","10^8 Galleria","10^8 Galleria","10^9 Galleria.A","10^9 Galleria.A","10^9 Galleria.A","10^9 Galleria.B","37 Galleria","37 Galleria","37 Galleria","37 Galleria"))
myset <- newSeqExpressionSet(as.matrix(seqdata_filtered),phenoData=data.frame(myx,row.names=colnames(seqdata_filtered)))

##apply upper quartile normalization
UQmyset <-betweenLaneNormalization(myset, which="upper")

##explore RLE plots of unnormalized and normalized data
plotRLE(myset, outline=FALSE,ylim=c(4,-4),col=colors[myx])
plotRLE(UQmyset, outline=FALSE,ylim=c(4,-4),col=colors[myx])

##explore PCA plots to look for outliers
plotPCA(myset, k=2, labels=FALSE, col=colors[myx])
plotPCA(UQmyset, k=2, labels=FALSE, col=colors[myx])

##subset to remove Late.3 outlier sample
myfilteredx <- seqdata_filtered[ , -which(names(seqdata_filtered) %in% c("Late_Galleria3"))]

##read in gene annotation information, locus tag is row names
genes <- read.delim("annotations.txt",header=TRUE, row.names=1)
```

```
genes <- genes[which(row.names(genes) %in% sel.rows), ]  ##Selects the filtered row names only

##assign treatment groups as factors
groups <- factor(c("early_LB","early_LB","mid_LB","mid_LB","late_LB","late_LB","X37_LB","X37_LB","early_host","early_host","early_host","early_host","mid_host","mid_host","mid_host","mid_host","late_host","late_host","late_host","X37_host","X37_host","X37_host","X37_host"),levels=c("early_host","mid_host","late_host","early_LB","mid_LB","late_LB","X37_host","X37_LB"))

##generate DGEList object and set design matrix based on treatment groups
dge <- DGEList(counts=myfilteredx, group=groups,genes=genes)
dge <- calcNormFactors(dge, method="upperquartile")
groups <- dge$samples$group
design <- model.matrix(~0+groups)

##check the BCV dispersion using EdgeR package
dispersion_check <- estimateDisp(dge,design,robust=TRUE)
plotBCV(dispersion_check)
dispersion_check$common.dispersion

##square root of common dispersion is the coefficient of biological variation, see: https://academic.oup.com/nar/article/40/10/4288/2411520 - paragraph 4
sqrt(dispersion_check$common.dispersion)

##apply voom mean-variance model
myvm <- voom(dge, design=design, plot=TRUE)

##print voom transformed LogCPM count data
write.table(myvm$E,file="LogCPM_count_data.txt")


#convert matrix to dataframe
myvm.df <- as.data.frame(myvm$E)

#function to pull the data out of a linear regression and return important values from https://susanejohnston.wordpress.com/2012/08/09/a-quick-and-easy-function-to-plot-lm-results-in-r/

ggplotRegression <- function (fit) {

 require(ggplot2)

 ggplot(fit$model, aes_string(x = names(fit$model)[2], y = names(fit$model)[1])) +
  geom_point() +
  stat_smooth(method = "lm", col = "red") +
  labs(title = paste("Adj R2 = ",signif(summary(fit)$adj.r.squared, 5),
                "\nIntercept =",signif(fit$coef[[1]],5 ),
                "\nSlope =",signif(fit$coef[[2]], 5),
                "\nP =",signif(summary(fit)$coef[2,4], 5)))
}
```

```
plot <-lm(Early.2 ~ Early.1, data  = log2Exprnorm.df)
regression_plot = ggplotRegression(plot)

##fit linear models
fit <- lmFit(myvm,design)

##smooth standard errors with Empirical Bayes
fit.eBayes <- eBayes(fit)

##specify host-specific contrast matrix
host_specific.contrast.matrix=makeContrasts(groupsearly_host-
groupsearly_LB,groupsmid_host-groupsmid_LB, groupslate_host-groupslate_LB,levels=design)
colnames(host_specific.contrast.matrix)=c("Early_hostvsLB","Middle_hostvsLB","Late_hostvsLB
")

##fit the host-specific contrast matrix
fit_host <- contrasts.fit(fit,host_specific.contrast.matrix)
fit_host <- eBayes(fit_host)

##identify DE genes
results_host <- decideTests(fit_host,adjust.method = "BY",p.value=0.05)

##print results
write.fit(fit_host, results_host,"results.txt")
vennDiagram(results_host)

##print all DE genes
AllDE.table <- topTable(fit_host, sort="non", n=Inf)
write.table(AllDE.table, "allDE_table.txt")

##fit contrasts for temperature/host factorial analysis (see section 9.5 of limma manual:
http://www.bioconductor.org/packages/devel/bioc/vignettes/limma/inst/doc/usersguide.pdf)
factorial.matrix <- makeContrasts(X37vs25inLB=groupsX37_LB-groupsmid_LB,
X37vs25inhost=groupsX37_host-groupsmid_host, Diff=(groupsX37_host-groupsmid_host)-
(groupsX37_LB-groupsmid_LB),levels=design)
fit <- lmFit(vm, design=design)
fit.factorial <- contrasts.fit(fit, factorial.matrix)
fit.factorial <- eBayes(fit.factorial)

##determine DE genes and print results
results <-decideTests(fit.factorial, adjust.method = "fdr",p.value=0.05)
summary(results)
write.fit(fit.factorial, results, "factorial.matrix.results.txt")
vennDiagram(results)

##order significant results by F value
modFpvalue=fit_host$F.p.value
indx = p.adjust(modFpvalue, method="BY")
sig = modFpvalue[indx]
nsiggenes = length(sig)
modF = fit_host$F
```

```
modFordered = order(modF, decreasing = TRUE)

##order the genes in the dge object based on F value
ranked_genes = dge$genes$GeneID[modFordered[1:nsiggenes]]
clust.table <- topTable(fit_host, sort="non",n=Inf)
ranked <- clust.table[which(row.names(clust.table) %in% ranked_genes), ]

##select only Log2CPM value columns and re-name
keeps.clust <- c("Early_hostvsLB","Middle_hostvsLB","Late_hostvsLB")
clust.log2FC <- clust[ , keeps.clust, drop=FALSE]
colnames(clust.log2FC) = c("Early","Middle","Late")

##convert object to matrix
clust.log2FC <- data.matrix(clust.log2FC, rownames.force=NA)

##convert matrix to data frame
clust.log2FC.df <- as.data.frame(clust.log2FC, rownames.force=NA)
clust.log2FC.df.ordered <- clust.log2FC.df[order(row.names(clust.log2FC.df)),]

##combine annotation information
sel.rows.b <- row.names(clust.log2FC)
genes.b <- genes[which(row.names(genes) %in% sel.rows.b), ]
genes.b <- genes.b[order(row.names(genes.b)),]
genes_annotated_df <- new("AnnotatedDataFrame",data=genes.b)
clust.log2FC.matrix.ordered <- data.matrix(clust.log2FC.df.ordered,rownames.force = NA)

##generate an ExpressionSet using Biobase
library(Biobase)
DEGesetlog2FC<-
ExpressionSet(assayData=clust.log2FC.matrix.ordered,featureData=genes_annotated_df)

##cluster the data using fuzzy clustering
library(Mfuzz)
fuzz <- mfuzz(DEGesetlog2FC, c = 20, m = 1.5)
mfuzz.plot2(DEGesetlog2FC,centre=TRUE,cl=fuzz)
overlap <- overlap(fuzz)
overlap.plot(fuzz, over=overlap,thres=0.3)
```

Supplementary section for Chapter 5

Table S12: LCMS-ESI-MS/MS results

geneID	protein name	accession	-10lgP	p-value	Coverage (%)	#Peptides	#Unique
PL78_04365	Hemolysin	tr|A0A3S6EX96|A0A3S6EX96_YERET	390.66	8.59E-40	48	51	51
PL78_11255	rplF, 50S ribosomal protein L6	tr|A0A210TK65|A0A210TK65_9GAMM	78.67	1.36E-08	31	5	5

Figure S16: Peaks DB peptide coverage of PL78_04365, filamentous hemagglutinin N-terminal containing protein from ~175 kDa protein LCMS-ESI-MS/MS. Blue bars represent covered regions of the 51 peptides identified from the protein.

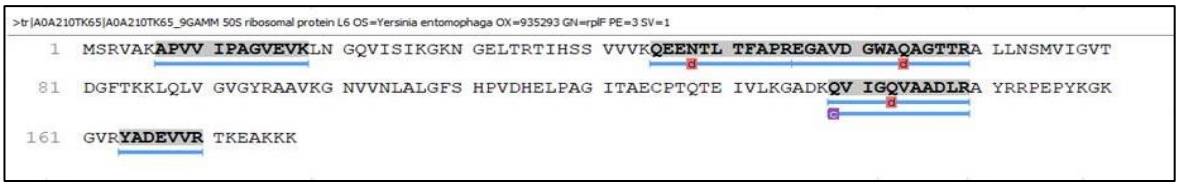


Figure S17: Peaks DB peptide coverage of PL78_11255, rplF 50S ribosomal protein L6 identified from ~8 kDa protein LCMS-ESI-MS/MS . Blue bars represent covered regions, red "d" represents deamidation, purple "c" represents carbamidomethylation.

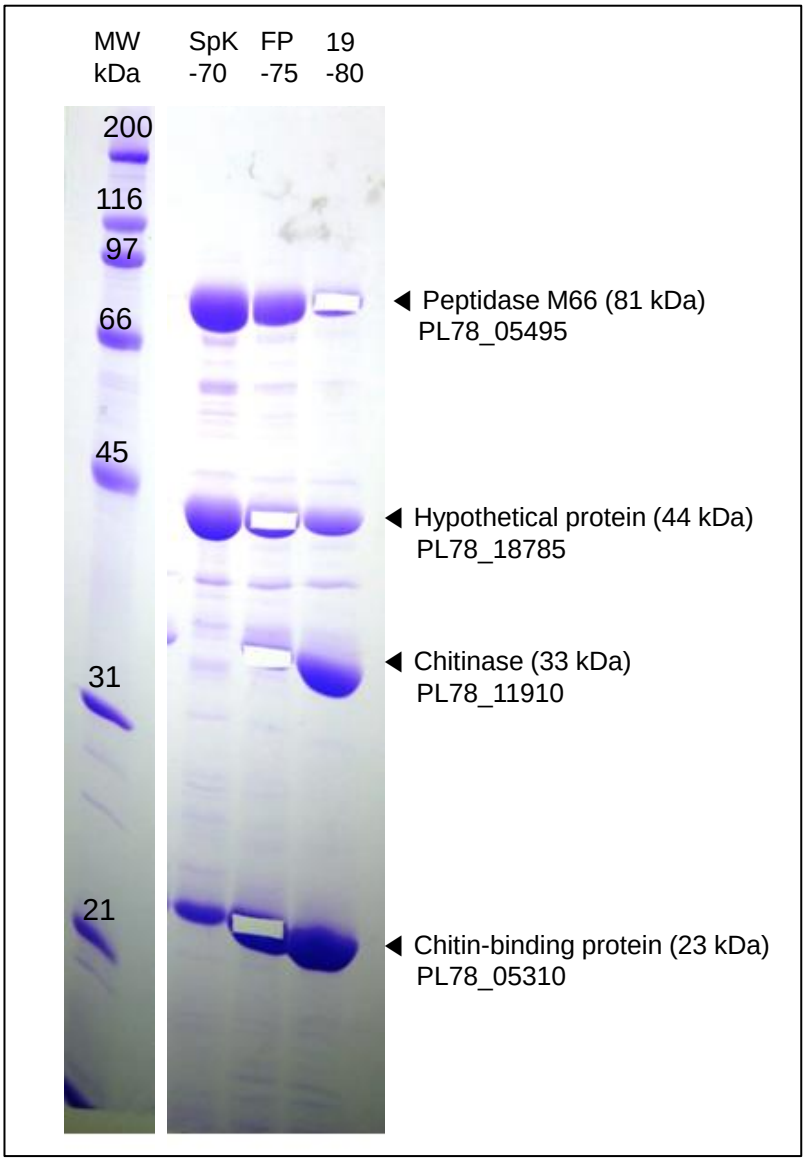


Figure S18: Size-exclusion chromatography fragment pools from cell supernatant of *Y. entomophaga* MH96 visualized on 12 % polyacrylamide gel by SDS-PAGE and stained with Coomassie brilliant blue. Excised bands are indicated by black arrow heads and identity determined by LC-ESI-MS/MS are given. This work was completed by Sandra Jones (AgResearch Ltd.).

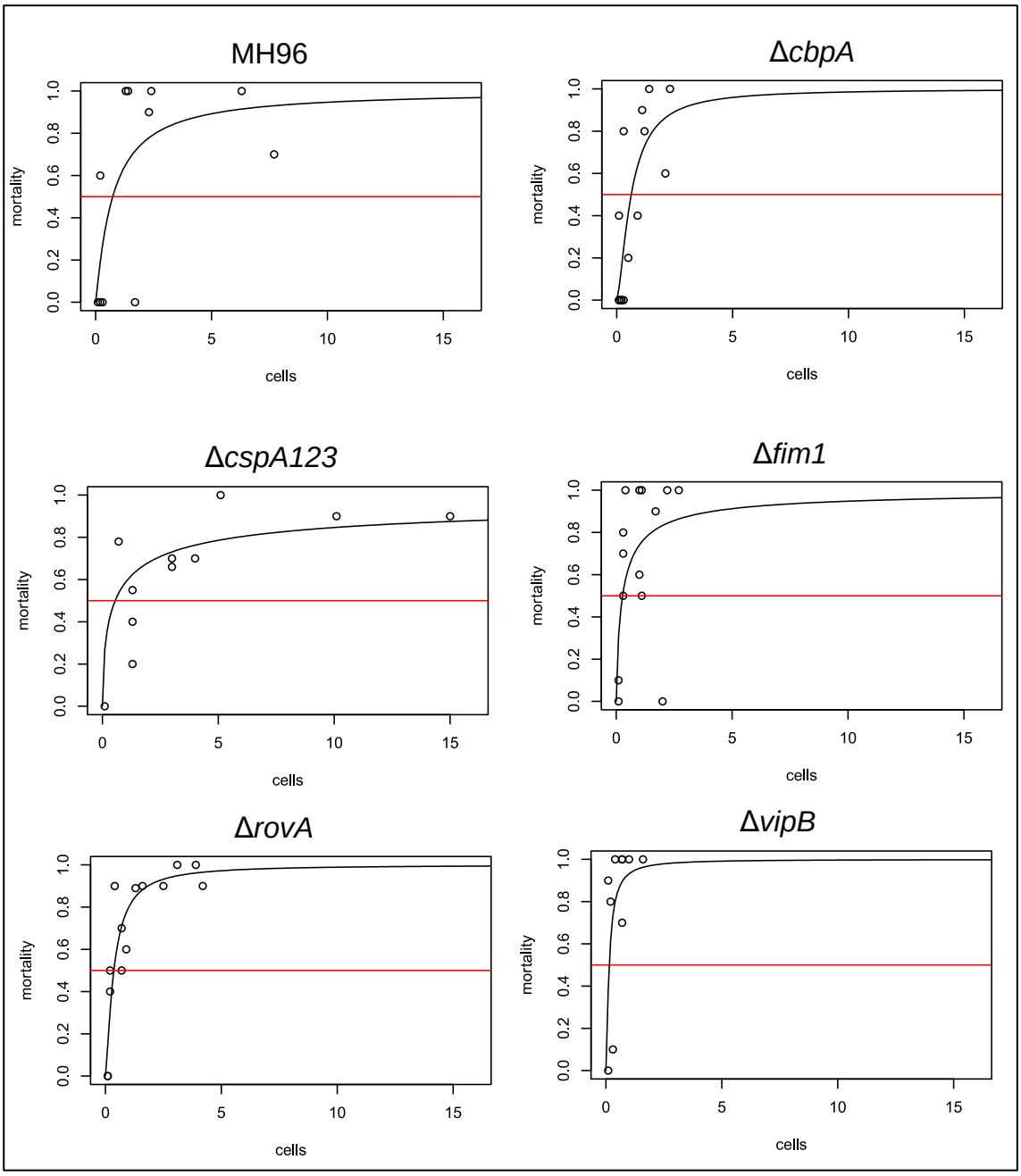


Figure S19: Fitted relationship using binomial logistic regression between dose (cells) of *Yersinia entomophaga* MH96 and Δ*cbpA*, ΔcspA123/ΔHCUI$_{YE96}$, Δ*fim1*, Δ*rovA* and Δ*vipB* and day five mortality rates of *Galleria mellonella* following intrahemocoelic injection and incubation at 25 °C. Each experiment included 10 larvae/dilution and five dilutions were tested each experiment as well as 10 larvae injected with PBS as a control. Each experiment was repeated three times and data were combined. Median lethal dose and standard error were estimated with the R package MASS.

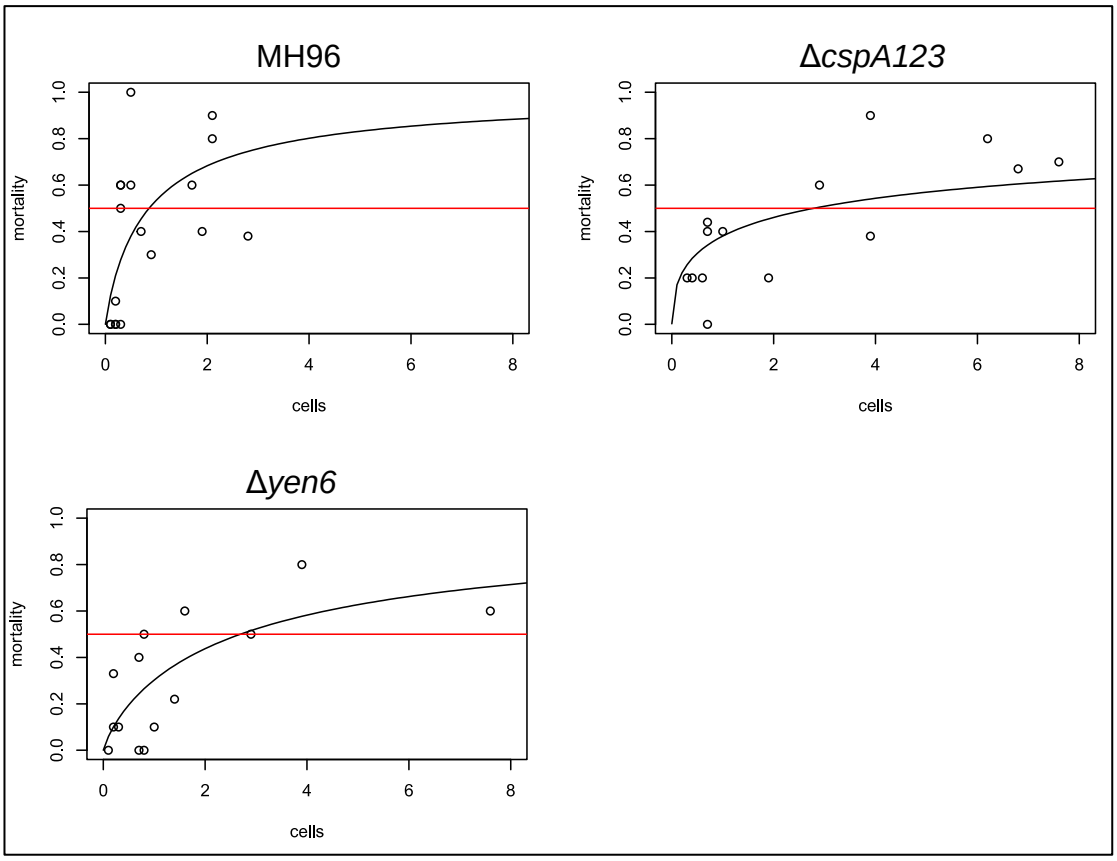


Figure S20: Fitted relationship using binomial logistic regression between dose (cells) of *Yersinia entomophaga* MH96 and Δ*cspA123*/ΔHCUI$_{YE96}$, and Δ*yen6* and day five mortality rates of *Galleria mellonella* following intrahemocoelic injection and incubation at 37 °C. Each experiment included 10 larvae/dilution and five dilutions were tested each experiment as well as 10 larvae injected with PBS as a control. Each experiment was repeated three times and data were combined. Median lethal dose and standard error were estimated with the R package MASS.

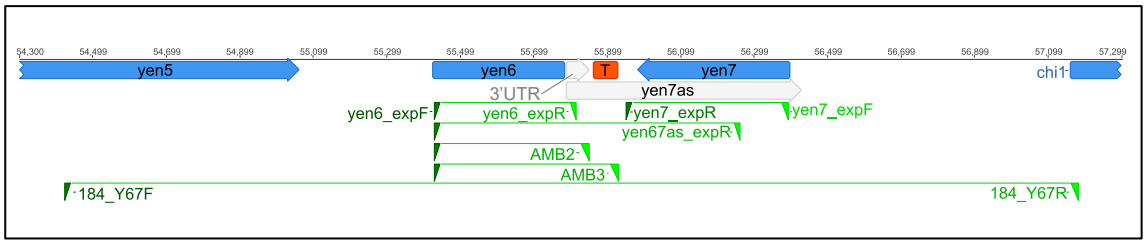


Figure S21: Portion of PAI$_{YE96}$ containing *yen5*, *yen6*, *yen7* and *chi1* of *Yersinia entomophaga* MH96 showing primer locations used for arabinose induction and complementation experiments. Orange T box = predicted terminator, grey arrow = predicted 3'UTR of *yen6* in including *yen7as* and blue arrows are protein coding regions.

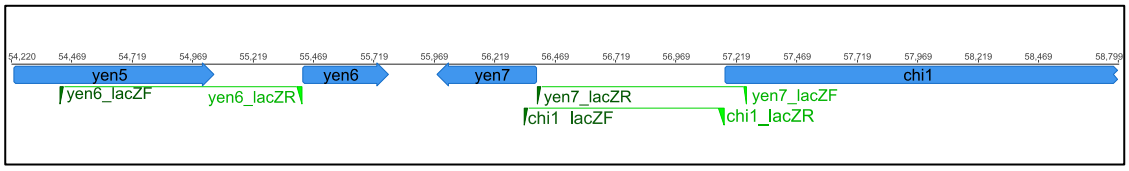


Figure S22: Portion of PAI$_{YE96}$ containing *yen5*, *yen6*, *yen7* and *chi1* of *Yersinia entomophaga* MH96 showing primer locations used to generate P$_{yen6}$::lacZ, P$_{yen7}$::lacZ and P$_{chi1}$::lacZ *cis*-merodiploid MH96-derivative strains.Blue arrows are protein coding regions.

Table S13: Primers used for *yen6*, *yen67as*, *yen7* arabinose induction, complementation and *lacZ* reporter cloning used in this study

Target	Primer name	Sequence (5' → 3')*	Cut site
Arabinose induction vector cloning			
Yen6	yen6_expF	AAACATATGTCTGGATTTGTGGATTTCAAA	*NdeI*
Yen6	yen6_expR	AAAGAATTCGCTTTACCCCACCTTGCTCA	*EcoRI*
Yen6	Amb2_expR	AAAGAATTCGTTATTAACCAAGATAAGTTCAGGG	
Yen6	Amb3_expR	AAAGAATTCGCCAGGCTACCTATGAGGTATCATT	
Yen67as	yen67as_expR	AAAGAATTCACGTCTTACTTTCCTCAATC	*EcoRI*
Yen7	yen7_expF	AAACATATGATATATTTTGTGGCAATTGACG	*NdeI*
Yen7	yen7_expR	AAAGAATTCTTACTCGTTAAACTAAGGGC	*EcoRI*
pAY2-4/*pBAD validation primers*			
pAY2-4/ pBAD	araF†	CATGGGGTCAGGTGGGAC	
pAY2-4/ pBAD	araR†	TCCATAAGATTAGCGGATCCTAC	
lacZ-reporter fusion cloning			
P_{yen6}	yen6_lacZF	AAAGAATTCCGTCGTTTCACATCGGATGC	*EcoRI*
P_{yen6}	yen6_lacZR	AAAGTCGACCATAATAACCTCCTTATCAAATTTGCCG	*SalI*
P_{yen6}	yen6_valF†	CTACGGTCATGGTGCCAGTT	
P_{yen7}	yen7_lacZF	AAAGAATTCAGGTCTCTTCTGCTGCTCCT	*EcoRI*
P_{yen7}	yen7_lacZR	AAAGTCGACCATCGTCGCTCCATAATGAAAA	*SalI*
P_{yen7}	yen7_valF†	ATCGGGATCAGTTGCTACGC	
P_{chi1}	chi1_lacZF	AAAGAATTCCCACACTACCTTTCCAGCTGT	*EcoRI*
P_{chi1}	chi1_lacZR	AAAGTCGACCATTTCAGACTCCATTTTTTAAATTAATGG	*SalI*
P_{chi1}	chi1_valF†	ATCGGGATCAGTTGCTACGC	
pVIK107	lacZ_valR†	GGGCCTCTTCGCTATTAC	
pACYC184	chlorF	AAAATCGATCCGCTAGCGCTGATGTCCGGC	*ClaI*
pACYC184	chlorR	AAAATCGATAAACCAGCAATAGACA	*ClaI*
Complementation			
P_{yen6}-P_{yen7}	184_Y67F	AAAGGATCCTCTTCTTTTTCCATTTCAGACTCCA	*Bam HI*
P_{yen6}-P_{yen7}	184_Y67R	AAAGCATGCCGTCGTTTCACATCGGATGC	*SphI*
pACYC184	TetF_val†	TTTCTATGCGCACCCGTTCT	
pACYC184	TetR_val†	ACCAGTGACGAAGGCTTGAG	

* Double-underscore represents restriction site. †indicates primer used for validation sequencing.

Table S14: RT-PCR primers used to explore the transcriptional organization of *yen6, yen7* and *chi1*

Target	Primer name	Sequence (5' → 3')	size (bp)
yen5 – yen6	YenTC_1_F	TGGATATTCGTCAGCCACCG	
yen5 – yen6	YenTC_1_R	AACTGGCGCTCCTGAATGAA	753
yen6	YenTC_2F	TTCATTCAGGAGCGCCAGTT	
yen6	YenTC_2R	AGTACCGCTGCTGCTTTTCT	248
yen6-3'UTR-*yen6*	YenTC_3F	ACTCGCGAGGCTAATGTCAC	
yen6-3'UTR-*yen6*	YenTC_3R	GGGTTAGTGGGGGCTTTACC	320
yen6-yen7as	YenTC_4F	AGTGTTCCTGTAGCAGAGCC	
yen6-yen7as	YenTC_4R	ACGATTTCCTCCCTCCTCCA	320
yen7as-yen7	YenTC_5F	AGTGGCATTACACGACTGGA	
yen7as-yen7	YenTC_5R	GAGCCAGGCTACCTATGAGG	262
yen7	YenTC_6F	AAACTAAGGGCAGCGAAGGC	
yen7	YenTC_6R	AGTAACAATAACATCCACCAGCT	331
yen7-5'UTR-*chi1*	YenTC_7F	ACAGCTGGAAAGGTAGTGTGG	
yen7-5'UTR-*chi1*	YenTC_7R	TGGAGGAGGGAGGAAATCGT	679
5'UTR-*chi1-chi1*	YenTC_8F	TCAGGAACCAAGAAAATTGACACA	
5'UTR-*chi1-chi1*	YenTC_8R	CCACACTACCTTTCCAGCTGT	399
chi1	YenTC_9F	GCGCCGGCAAATAAGATGAC	
chi1	YenTC_9R	GTTGGGGTCCAGGTGAGTTT	872

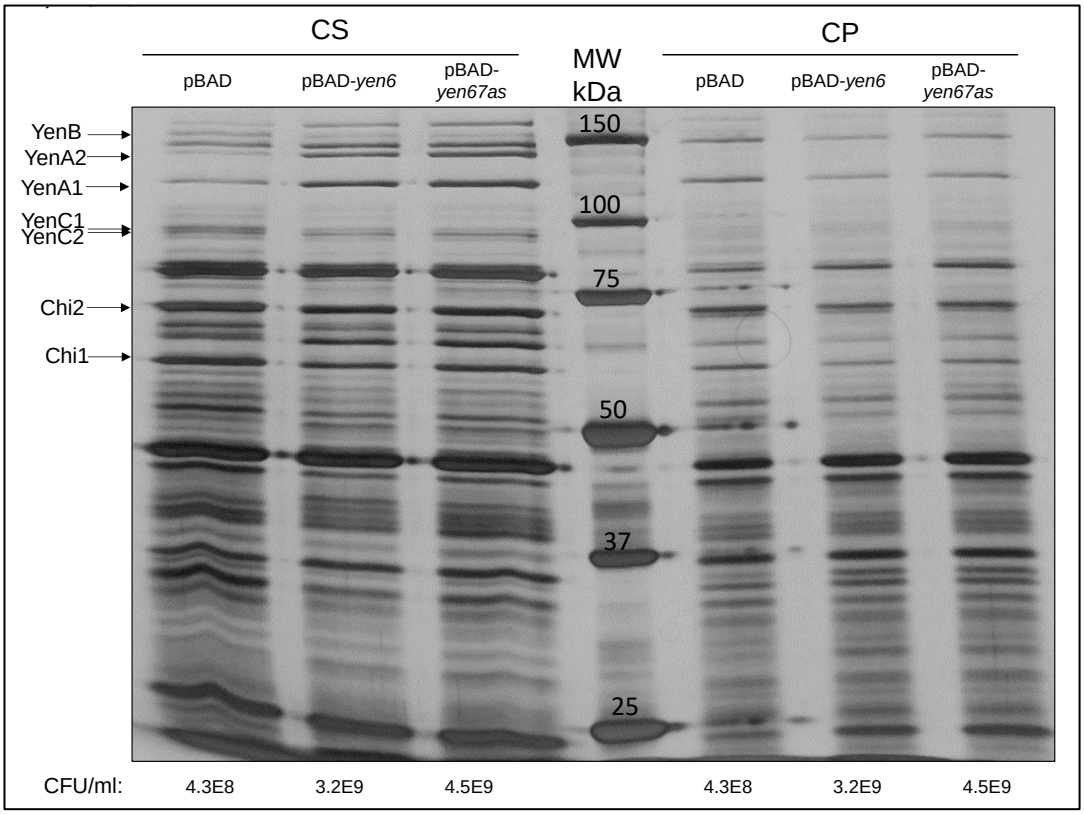


Figure S23: Silver-stained SDS-polyacrylamide gel of proteins produced and secreted by *Yersinia entomophaga* MH96 strains carrying arabinose-inducible expression vector pBAD, pBAD-*yen6* and pBAD-*yen67as* under arabinose 0.2 % at 25 °C and grown for 16 h (early stationary). CS = filtered cell supernatant and CP = cell pellet. The Marker (M) lane contains Bio-Rad Precision Plus Protein™ unstained protein standard, with respective ladder sizes given. Arrows correspond to Yen-TC components, YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa). Cell density is reported as CFU/ml as determined by enumeration plates. In this example the over-production of Yen-TC components in the pBAD-*yen67as* strain is difficult to observe based on band intensity.

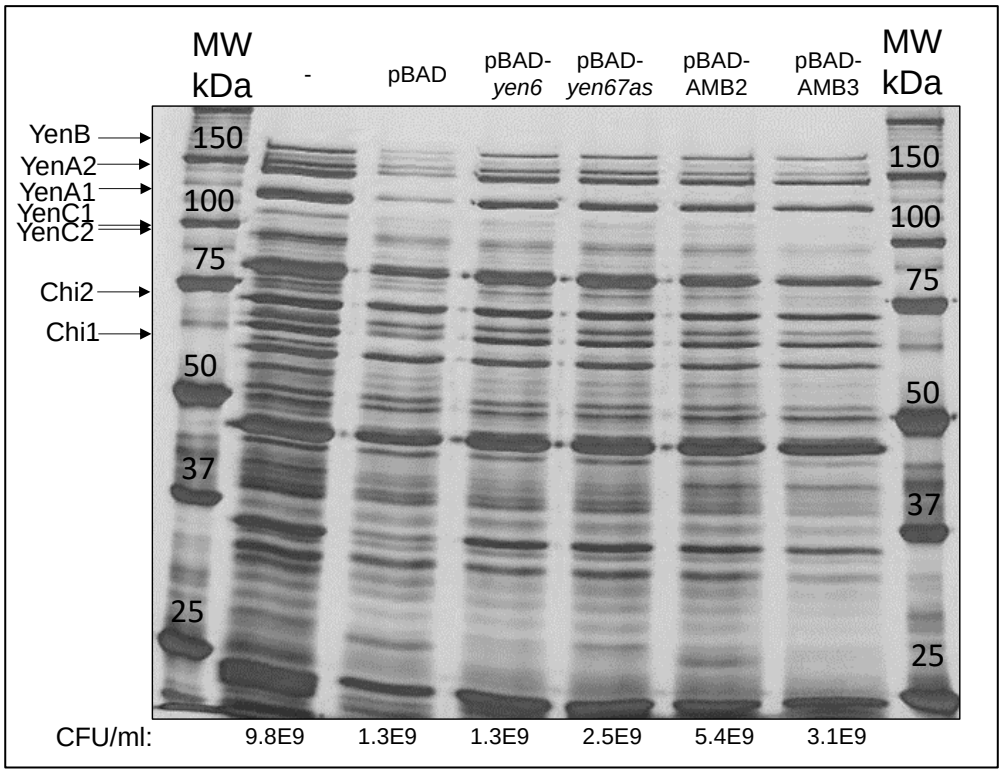


Figure S24: Silver-stained SDS-polyacrylamide gel of filtered cell supernatant from *Yersinia entomophaga* MH96 strains carrying arabinose-inducible expression vector pBAD, pBAD-*yen6* and pBAD-*yen67as* under arabinose 0.2 % at 25 °C and grown for 16 h 40 min (early stationary). The Marker (M) lane contains Bio-Rad Precision Plus Protein™ unstained protein standard, with respective ladder sizes given. In this example the over-production of Yen-TC components in the pBAD-*yen67as* strain is difficult to observe based on band intensity.

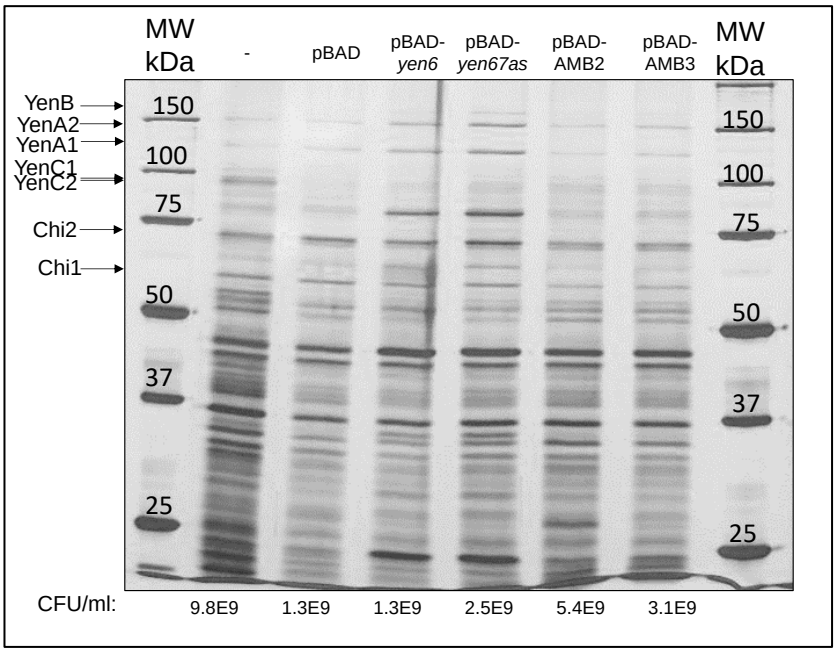


Figure S25: Silver-stained SDS-polyacrylamide gel of cell pellet from *Yersinia entomophaga* MH96 strains carrying arabinose-inducible expression vector pBAD, pBAD-*yen6* and pBAD-*yen67as* under arabinose 0.2 % at 25 °C and grown for 16 h 40 min (early stationary). The Marker (M) lane contains Bio-Rad Precision Plus Protein™ unstained protein standard, with respective ladder sizes given. In this example the over-production of Chi2 (~70 kDa) can be observed in the cell pellet of pARA-*yen67as* strain compared to pBAD and pBAD-*yen6* strain.

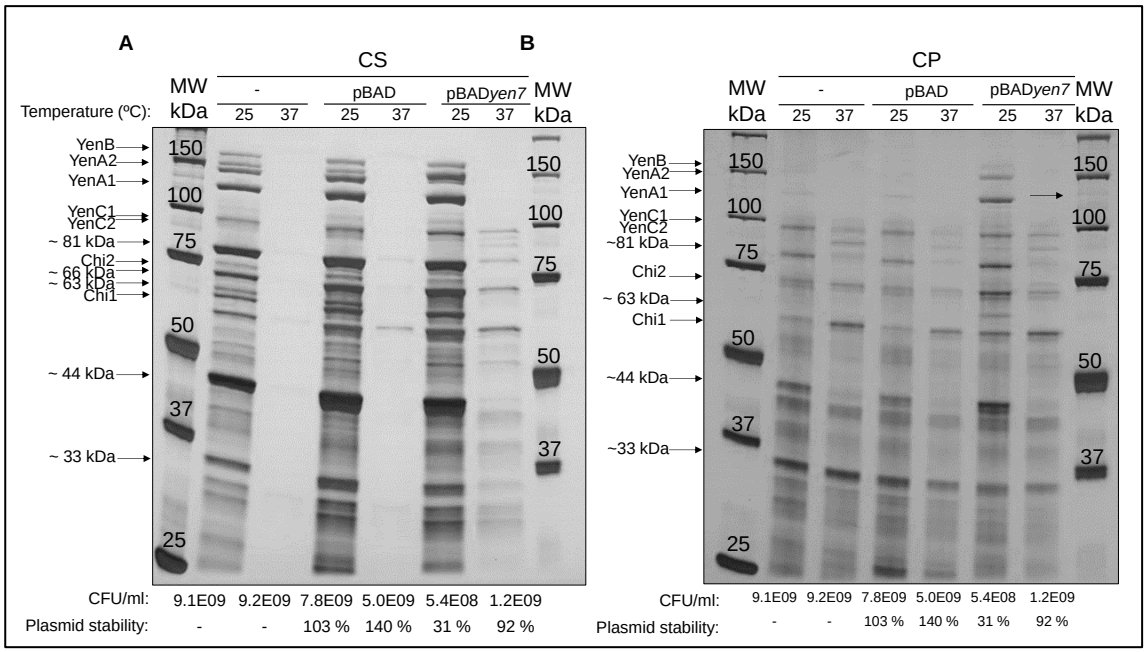


Figure S26: Replicate, silver-stained SDS-polyacrylamide gel of proteins produced and secreted by *Yersinia entomophaga* MH96 strains carrying arabinose-inducible expression vector pBAD and pBAD-*yen7* under arabinose 0.002 % at 25 or 37 °C for 16 h. A) CS = cell supernatant and B) CP = cell pellet. MW = molecular weight over marker lane containing Bio-Rad Precision Plus Protein™ unstained protein standard, with respective ladder sizes given. Arrows correspond to Yen-TC components, YenB (167 kDa), YenA2 (156 kDa), YenA1 (130 kDa), YenC1 (109 kDa), YenC2 (107 kDa), Chi2 (70 kDa) and Chi1 (60 kDa). Other marked proteins include Peptidase M66 (PL78_05495), Hypothetical virulence factor (PL78_18785) and Chitinase (PL78_11910) corresponding to approximate sizes of 81, 44 and 33 kDa, respectively. Cell density is reported as CFU/ml as determined by enumeration plates and plasmid stability is reported as the proportion of CFUs enumerated from ampicillin 400 μg/ml compared to CFUs enumerated from media without antibiotic.

Table S15: Preliminary β galactosidase assay for *Yersinia entomophaga* MH96 wild-type and *cis*-merodiploid strains P_{yen6}::*lacZ*, P_{yen7}::*lacZ* and P_{chi1}::*lacZ* at 25 and 37 °C.

Strain	OD_{600}	Temperature (°C)	Miller Unit
MH96	0.3	25	-8
MH96	0.6	25	-8
MH96	0.9	25	-2
P_{yen6}::lacZ	0.3	25	95
P_{yen6}::lacZ	0.6	25	167
P_{yen6}::lacZ	0.9	25	233
P_{yen7}::lacZ	0.3	25	-5
P_{yen7}::lacZ	0.6	25	-8
P_{yen7}::lacZ	0.9	25	-4
P_{chi1}::lacZ	0.3	25	28
P_{chi1}::lacZ	0.6	25	79
P_{chi1}::lacZ	0.9	25	106
MH96	0.3	37	-2
MH96	0.6	37	-4
MH96	0.9	37	-5
P_{yen6}::lacZ	0.3	37	74
P_{yen6}::lacZ	0.6	37	171
P_{yen6}::lacZ	0.9	37	196
P_{yen7}::lacZ	0.3	37	-3
P_{yen7}::lacZ	0.6	37	-2
P_{yen7}::lacZ	0.9	37	-3
P_{chi1}::lacZ	0.3	37	52
P_{chi1}::lacZ	0.6	37	81
P_{chi1}::lacZ	0.9	37	145

Table S16: Results of *in vivo* β galactosidase assay during *Galleria mellonella* infection at 25 and 37 °C using *cis*-merodiploid *Yersinia entomophaga* MH96 with P_{yen6}::*lacZ* (3 replicate experiments, 1 is independent, 2 and 3 used same hosts from the same cohort). HPI = hours-post infection, CFU = colony forming unit and MU = Miller Unit.

Temp-erature (°C)	Replicate	Batch	HPI	Mean cell density (CFU/g)	Optical density at 600 nm	MU	Note
MH96 with P_{yen6}::lacZ translational reporter							
37	1	1	16-17	1.4E10	0.083	125,904	
	2				0.022	123,337	
	3				0.029	104,405	
	4				0.081	132,302	
	5				0.044	124,837	
	6				0.158	119,916	
25	1	1	20-22	9.0E8	0.734	1,346	
	2				0.577	1,754	
	3				0.542	2,951	
	4				0.568	2,282	
	5				1.121	2,235	
	6				0.886	1,840	
37	1	2	17	2.9E9	0.031	30,590	
	2				0.156	157,633	
	3				0.022	20,757	
	4				0.016	25,628	
	5				0.001	-28,411	omitted
	6				0.111	142,243	
25	1	2	19	1.6E9	0.084	2,529	
	2				0.298	1,521	
	3				0.067	4,856	
	4				0.289	3,775	
	5				0.160	2,912	
	6				0.128	2,646	
37	1	3	17	1.6E8*	0.008	170,751	omitted
	2				0.015	18,446	omitted
	3				0.106	248,055	omitted
	4				0.023	96,065	omitted
	5				-0.003	-3,518	omitted
	6				0.016	-2,181	omitted
25	1	3	19	1.1E9*	0.089	1,463	omitted
	2				-0.008	1,452	omitted
	3				0.002	52,788	omitted
	4				-0.003	6,419	omitted
	5				0.001	6,028	omitted
	6				0.022	-2,239	omitted

Temp-erature (°C)	Replicate	Batch	HPI	Mean cell density (CFU/g)	Optical density at 600 nm	MU	Note
Negative control samples (wild-type MH96)							
37	NC	1			0.029	-2,752	
37	NC	2			0.470	290	
37	NC	3			0.067	-723	omitted
25	NC	1			0.290	-744	
25	NC	2			0.071	-337	
25	NC	3			0.020	2,929	omitted

*enumeration plate counts highly variable, with individuals very near or clearing infection

Supplementary section for Chapter 6

Table S17: Adapter and barcode fasta sequence used to trim raw RNA-seq libraries from *Yersinia entomophaga* MH96, Δ*cspA123*/ΔHCUI$_{YE96}$ and Δ*yen6*.

>CSP_2
AGTACAAGCGTACTAG
>CSP_2_reversed
AGTACAAGCGTACTAG
>CSP_3
AACAACCAAACGTGAT
>CSP_3_reversed
ATCACGTTTGGTTGTT
>K37_1
ACAAGCTACGCTCATT
>K37_1_reversed
AATGAGCGTAGCTTGT
>K37_2
CTGTAGCCTCTCGCGC
>K37_2_reversed
GCGCGAGAGGCTACAG
>Y6_1
AACCGAGAAAACATCG
>Y6_1_reversed
CGATGTTTTCTCGGTT
>Y6_2
AACGCTTACATCAAGT
>Y6_2_reversed
ACTTGATGTAAGCGTT
>P7
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGAAGAGAATCTCGTATGCC
GTCTTCTGCTTG
>P7_reverse
CAAGCAGAAGACGGCATACGAGATTCTCTTCAGTGACTGGAGTTCAGACGTGT
GCTCTTCCGATC
>P5
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTA
TCATT
>P5_reverse
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTA
TC

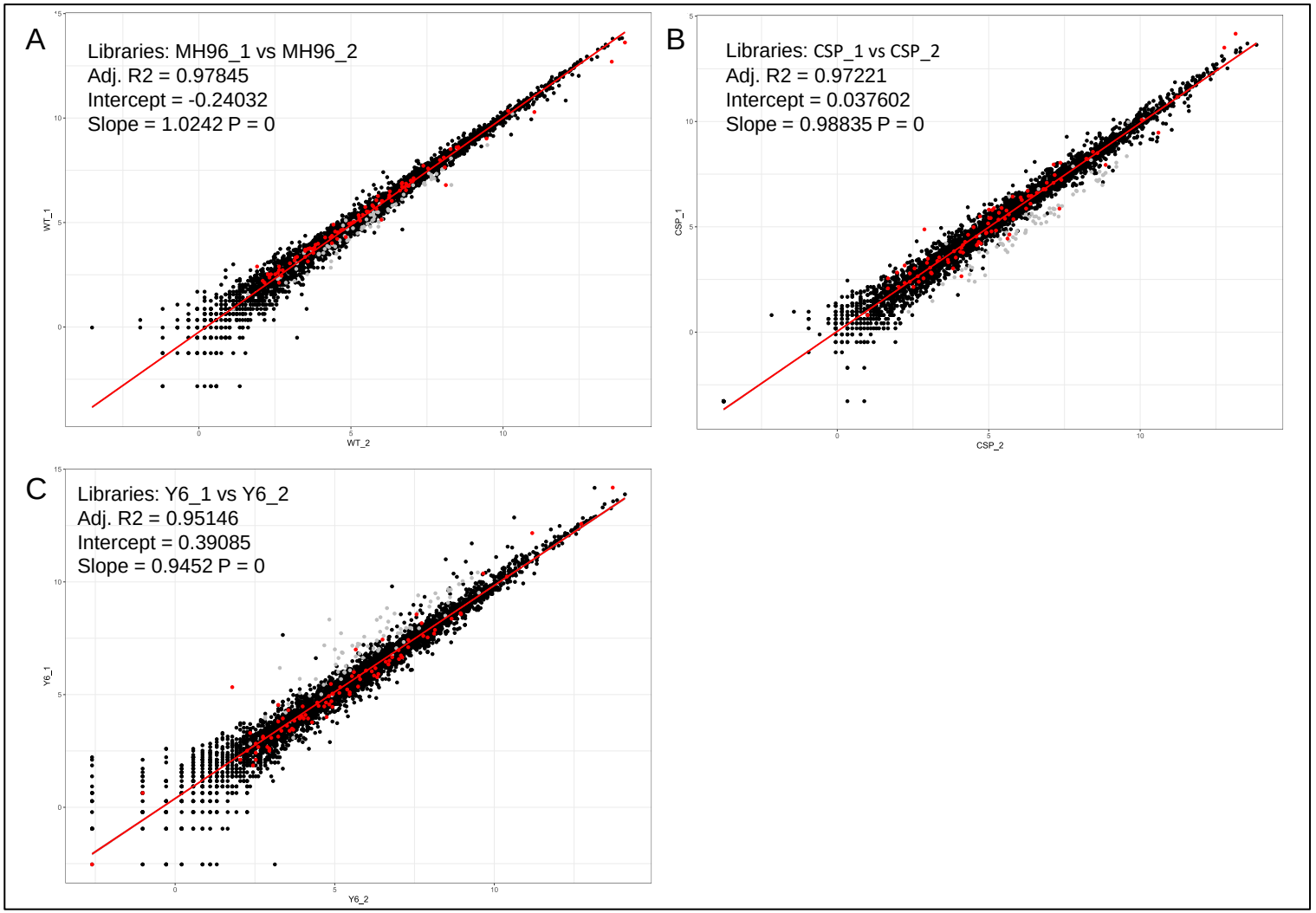


Figure S27: Regression of *Yersinia entomophaga* MH96, Δc*spA123*/ΔHCUI$_{YE96}$ and Δ*yen6 in vivo* count data comparing biological replicate samples among strains. A) Wild-type (WT) sample 1 vs. sample 2, B) ΔcspA123/ΔHCUI$_{YE96}$ (CSP) sample 1 vs. sample 2 and C) Δ*yen6* (Y6) sample 1 vs. sample 2. Black = CDS, Red = non-coding RNA, Grey = anti-sense RNA.

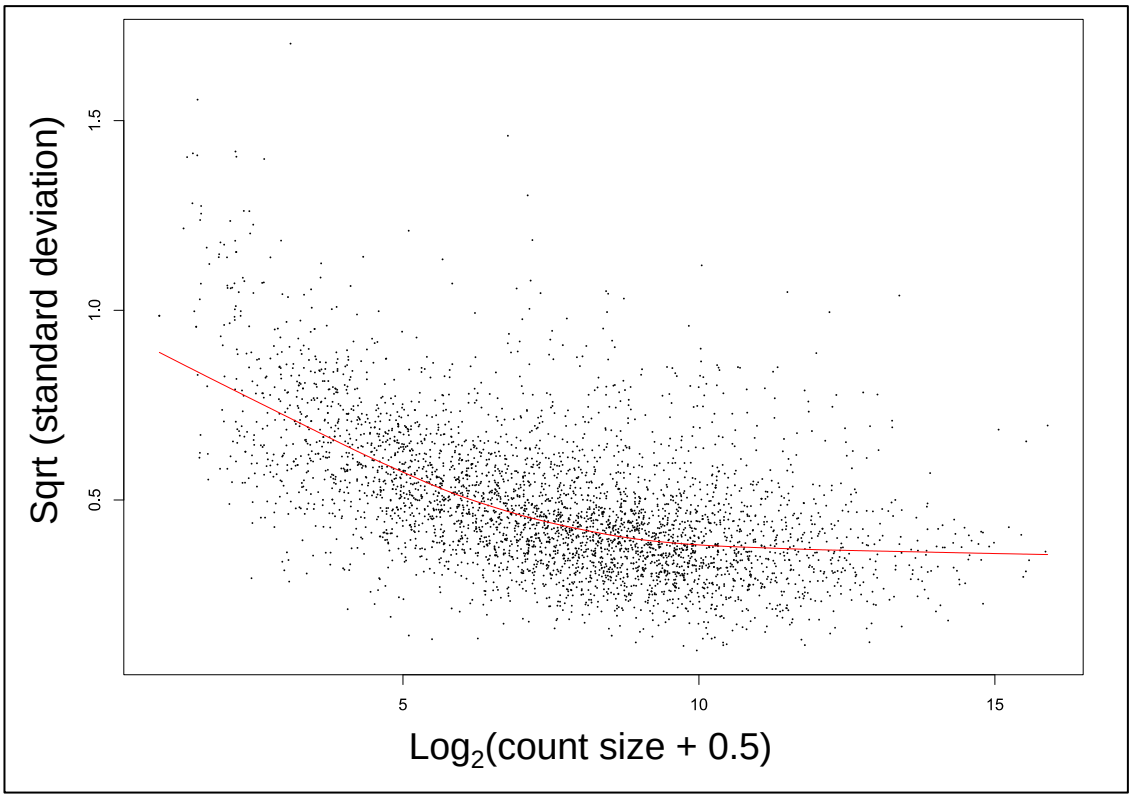


Figure S28: Transcript-wise mean-variance relationship of MH96, ΔcspA123/ΔHCUI$_{YE96}$ and Δ*yen6 in vivo* libraries modelled using voom. Red line represents lowess fit.

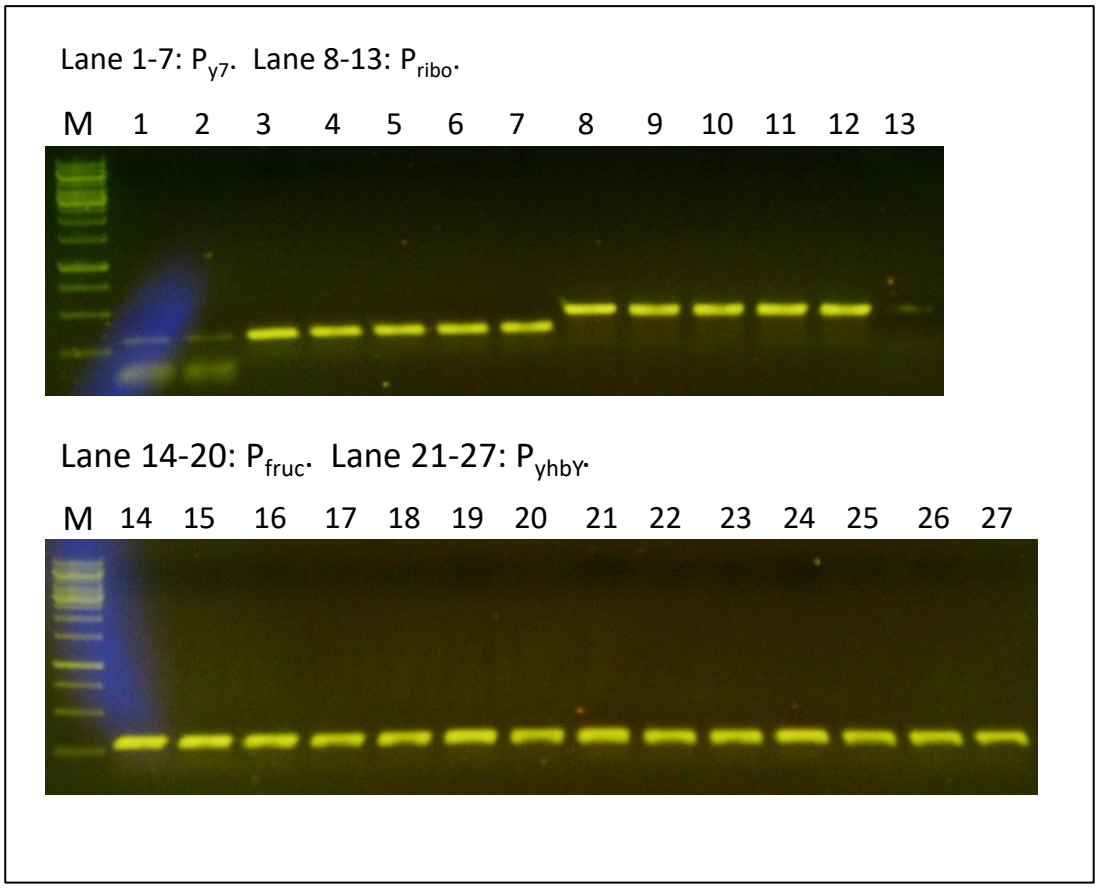


Figure S29: Visualization of DNA-probes used for EMSA assessment of recombinant Yen6 by agarose gel electrophoresis. This gel was provided by Dr. Naren at Massey University. P_{y7} = *yen7* promoter region, P_{ribo} = *rbsD* promoter region, P_{fruc} = *IIA/hpr* promoter region and P_{yhbY} = *yhbY* promoter region.