Investigation of *Pasteurella multocida* pathogenesis using transposon-directed insertion site sequencing and comparative genomics

Thomas Rodney Smallman

Bachelor of Science (Honours)

A thesis submitted for the degree of Doctor of Philosophy at Monash University in 2021

Department of Microbiology, Monash Biomedical Discovery Institute

Table of contents

Thesis including submitted works declaration

I hereby declare that this thesis contains no material that has been accepted for the award of any other degree at any university or equivalent institution. It is also to the best of my knowledge and belief that this thesis contains no material previously published or written by any other person, excluding where referenced throughout the text.

This thesis includes one original paper submitted for publication in a peer reviewed journal. The core theme of this thesis is comprehensive characterisation of *P. multocida* virulence factors and pathogenic mechanisms using Transposon-directed insertion site sequencing and comparative genomics. The ideas, development, and preparation of all manuscripts in the thesis were the principal responsibility of myself, the student, working within the Department of Microbiology under the supervision of Professor John Boyce and Doctor Marina Harper.

The inclusion of the following co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

For the manuscript in chapter two:

Thesis Chapter	**Publication Title**	**Status** *(published, in press, accepted or returned for revision, submitted)*	**Nature and % of student contribution**	**Co-author name(s) Nature and % of Co-author's contribution***	**Co-author(s), Monash student Y/N***
2	Whole genome investigation of essential genes and capsule production genes in *Pasteurella multocida*	Submitted for publication	75% including development of the concept of the work, performing experimental work, data analysis and manuscript writing	1) John Boyce, concept, and manuscript preparation 10% 2) Marina Harper, concept, and manuscript preparation 10% 3) Galain Williams, scanning electron microscopy 5%	1) No 2) No 3) Yes

I have not renumbered sections of the submitted paper in order to generate a consistent presentation within the thesis

Student name: Thomas Smallman

Student signature: **Date:**

The undersigned hereby certify that the above declaration above correctly reflects the nature and extent of the student's and co-author's contributions to this work.

Main supervisor name: John Boyce

Main supervisor signature: **Date:**

Publications and conference proceedings during enrolment

Publications under review

Thomas R. Smallman, Galain Williams, Marina Harper[#] and John D. Boyce. 2021. Whole genome investigation of essential genes and capsule production genes in *Pasteurella multocida*. Submitted to mBio on the 4th of August 2021.

Conference proceedings

2017	Poster	"Using transposon-directed insertion site sequencing to investigate *Pasteurella multocida* pathogenesis" BacPath14, Adelaide, Australia
2019	Poster	"Using transposon-directed insertion site sequencing to investigate *Pasteurella multocida* pathogenesis" BacPath15, Perth, Australia
2021	Talk	"Investigating *Pasteurella multocida* pathogenesis through comparative genomics and TraDIS" ASM 2021, Melbourne, Australia

Acknowledgements

I would like to wholeheartedly thank my supervisors John and Marina, you've both been amazing supervisors and I greatly appreciate everything you've done for me. The unending guidance, willingness to help, support and encouragement you've both given me (and everyone else in the lab) makes working with you both incredibly enjoyable and rewarding. I have no doubt if I was supervised by anyone else, I wouldn't have gotten to the end. Thank you both so much. A big thank you as well to my unofficial lab supervisor Amy for all the help you've given me throughout the years. Thank you to everyone I've worked with in the Boyce lab and the extended Boyce lab family. You're all truly wonderful people, working with you made work not feel like work, and your friendship and support has kept me mostly sane throughout this endeavour.

Thank you to everyone that has helped me in the lab or on a server. A big thank you to Despina Kotsanas, Tony Korman, Conny Turni and Justine Gibson for sharing human, cat and dog *Pasteurella* isolates. Thank you to Fran Short for sharing the TraDISort protocol with me and all the advice you've given me along the way. Thank you to Galain Williams for your assistance and time with electron microscopy. Thank you to Margareta Go, Scott Coutts and Jacob Amy for their advice and help with Illumina sequencing, this was invaluable when navigating a custom Illumina sequencing protocol. Thank you to Sarah Williams, Adele Bargahare and David Powell for their help with bioinformatic analyses. I want to also thank all the microbiology teaching team, especially Kylie Wilson, Jen Moffatt and Meredith Hughes; teaching was one of the highlights of the PhD and this was no doubt due to your enthusiasm and support.

I would also like to thank my family and friends for their support throughout my PhD. To Laura, thank you for everything you've done for me, if not for your help and encouragement I would not have finished this. Thank you so much for all the bioinformatics help you've given me, and sorry for being such an annoying student. To Mum and Dad, thank you for all the sacrifices and hard work for getting me through up until now, I owe you both so much.

Although this thesis will be submitted for an individual award, a PhD is always a team effort. I will have definitely missed people that have helped me get through this, so apologies if I've left you out.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship

Abbreviations

%	Percent
(p)ppGpp	guanosine 5’-diphosphate-3’-diphosphate and guanosine 5’-triphosphate-3’-diphosphate
~	Approximately
-RT	Minus reverse transcriptase
+RT	Plus reverse transcriptase
<	Less than
>	Greater than
°C	Degrees Celsius
3’	Three prime
5’	Five prime
ABC transporter	ATP-binding cassette transporter
ACT domain	Aspartokinase, choismate mutase
ANI	Average nucleotide identity
BLASTn	Nucleotide Basic Local Alignment Search Tool
BLASTp	Protein Basic Local Alignment Search Tool
bp	Base pairs
cAMP	Cyclic adenosine monophosphate
CARD	Comprehensive antibiotic resistance database
CC domain	Coiled coil
CFU	Colony forming units
COGs	Cluster of orthologous groups
DEG	Database of essential genes
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
Flp-pili	Fimbrial-low molecular weight pili
g	Gravitational force
GalNAc	N-acetyl galactosamine
GlcA	N-acetyl glucuronic acid
GlcNAc	N-acetylglucosamine
h	Hour(s)
HI	Heart Infusion
HITS	High-throughput insertion tracking by deep sequencing
ICE	Integrative conjugative element
ID_{50}	50% infectious dose
INSeq	Insertion sequencing
kb	Kilobase pairs
Kdo	β 3-deoxy-D-manno-octulsonic acid
L50	Number of contigs required for generating half of the genome
LD_{100}	100% lethal dose

LD_{50}	50% lethal dose
LPS	Lipopolysaccharide
LRT	Lower respiratory tract
M	Molar
MALDI-TOF	Matrix assisted laser deabsorption ionization-time of flight
ManNAcA	N-acetylmannosaminuronic
min	Minute(s)
mL	Millilitre
mm	Millimetre
N50	Smallest contig required for assembly of half of the genome
nm	Nanometer(s)
ORF	Open reading frame
p	P-value
PastyVRDB	*Pasteurella multocida* virulence and resistance database
PBS	Phosphate buffered saline
PCho	Phosphocholine
PCR	Polymerase chain reaction
PG	phosphatidylglycerol
PMT	*Pasteurella multocida* toxin
polyP	polyphosphate
PTS system	phosphoenolpyruvate: sugar phosphotransferase system
qPCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative reverse transcriptase polymerase chain reaction
RMP	Revolutions per minute
RNA	Ribonucleic acid
RRM domain	TyrA/RNA recognition motif
RSH	RelA/SpoT homologs
SAS	Small alarmone synthases
sec	Second(s)
SEM	Scanning electron microscopy
SNP	Single nucleotide polymorphism
SNVs	Single nucleotide variants
SR	Stringent response
STM	Signature tagged mutagenesis
T6SS	Type VI secretion system
Tad-locus	Tight adherence locus
TdT	Terminal deoxynucleotide transferase
TEM	Transmission electron microscopy
TGS domain	ThrRS GTPase, SpoT
Tn	Transposon
Tn-Seq	Transposon sequencing

TraDIS	Transposon-directed insertion site sequencing
UIS	Unique insertion sites
URT	Upper respiratory tract
v/v	Volume/volume
VFDB	Virulence factor database
w/v	Weight/volume
ZFD	Zinc finger domain
α	Alpha
β	Beta
µL	Microlitre(s)
µm	Micrometer(s)
µM	Micromolar(s)
Ω	Ohm

Summary

Pasteurella multocida is the causative agent of a wide range of distinct animal diseases in different mammal and bird species, including fowl cholera in birds, haemorrhagic septicaemia in ungulates, atrophic rhinitis in pigs, and bovine respiratory disease in livestock animals. *P. multocida* also causes a range of zoonotic infections in humans. The pathogenic mechanisms that underly most of these diseases are still incompletely understood, with almost nothing known about *P. multocida* strains that cause disease in humans. Several important *P. multocida* virulence factors have been identified and shown to be critical for disease; however, most of these remain to be fully characterised. This study aimed to further characterise known virulence factors and pathogenic mechanisms of *P. multocida*, as well as identify novel pathogenic mechanisms and virulence factors required for different *P. multocida* diseases. Specifically, we fully characterised genes involved in the production of hyaluronic acid (HA) and genes important for growth in chicken serum with active complement activity in the *P. multocida* fowl cholera isolate VP161 using transposon-directed insertion site sequencing (TraDIS). Secondly, we sequenced a number of characterised human *P. multocida* isolates and identified putative virulence factors and pathogenic mechanisms in *P. multocida* using a range of comparative genomic and bioinformatic analyses.

For TraDIS analysis of *P. multocida* strain VP161, a transposon mutant library with over 81,000 unique mutant strains was generated using the *Himar1* transposon. TraDIS analysis of the VP161 *Himar1* mutant library after growth in rich media identified 509 genes required for growth in rich media, the majority of which encoded proteins involved in housekeeping processes. To identify genes important for HA capsule production, the VP161 *Himar1* mutant library was separated into acapsular and capsulated populations by discontinuous Percoll gradient centrifugation. TraDIS analysis comparing the acapsular and capsulated populations identified 69 genes important for HA capsule production in *P. multocida*, including all 10 known capsule biosynthesis genes and positive regulators of capsule biosynthesis in *P. multocida*. Directed mutants were generated in five novel *P. multocida* capsule production genes, with HA capsule absorbance assays confirming these genes were required for capsule production in *P. multocida*. Several genes identified as important for *P. multocida* capsule production in this analysis were involved in the stringent response, including *spoT*. The stringent response is a bacterial stress response controlled by guanosine alarmones, with *spoT* encoding a bifunctional guanosine alarmone synthase/hydrolase. Mutagenesis of the 3' end of *spoT*, which disrupted putative regulatory domains within SpoT, resulted in reduced capsule gene expression, suggesting capsule gene expression is downregulated in *P. multocida* when the stringent response is activated.

TraDIS analysis was also performed following growth of the VP161 *Himar1* mutant library in 90% chicken serum. TraDIS analysis identified 52 genes essential for growth in serum that were not essential for growth in rich media. In addition, analysis comparing mutants recovered following growth in 90% chicken serum and growth in rich media identified 44 genes important for fitness in chicken serum. Genes essential or important for fitness encoded capsule biosynthesis and export proteins, proteins involved in aerobic ATP synthesis and components of the Tol-Pal export system. Comparison with the genes identified by the TraDIS analysis of acapsular mutants identified 26 genes important for HA capsule production and important for serum resistance, including genes encoding capsule biosynthesis and export proteins, the cAMP phosphodiesterase CpdA and a putative sodium proton antiporter. Further work is required to elucidate the role of these proteins in capsule production and serum resistance in *P. multocida* strain VP161.

The genomes of 59 *Pasteurella* spp. isolates recovered from human, dog and cat infections were sequenced and compared. Genome analysis of these 59 isolates identified 36 as *P. multocida*, 13 as *P. canis*, 3 as *P. dagmatis*, 4 as *P. stomatis* and 3 as *Frederiksenia canicola*. Virulence factor and antibiotic resistance genes were identified in these isolates by comparing genome sequences using BLASTn to publicly available virulence factor and antibiotic resistance databases, as well as to a database of *P. multocida* virulence factor and antibiotic resistance genes compiled in this study. The *P. multocida* isolates sequenced in this study encoded many of the virulence factors identified in other *P. multocida* strains; however, several isolates did not contain a capsule biosynthesis and/or did not contain a complete Flp-pili biosynthesis locus. Pan-genome analysis of the 36 *P. multocida* isolates sequenced in this study and 260 *P. multocida* whole-genome reference sequences identified 10 genes significantly associated with capsule type F strains, including genes encoding heavy metal resistance proteins, and 27 genes significantly associated with capsule genotype B and LPS genotype L2 strains, including a putative glycosyltransferase and lipoprotein. A core-genome phylogeny of the 296 *P. multocida* isolates showed that the majority of the human *P. multocida* isolates clustered with cat, dog, and mouse isolates as a well separated clade (clade A), while most animal isolates clustered into a second large clade (clade B). Genes encoding a putative carbohydrate-specific ABC-transport system and L-fucose catabolism enzymes were associated with strains from clade A, whereas genes encoding proteins involved in anaerobic respiration and carbohydrate uptake were associated with strains from clade B. These genes represent candidates that contribute to the differences in host predilection and types of diseases caused by different strains of *P. multocida*.

Chapter 1 - Introduction

1.1 *Pasteurella multocida*

Pasteurella multocida is a small, encapsulated, non-motile, Gram-negative coccobacilli (1). *P. multocida* has been isolated as a commensal from the upper respiratory tract of many mammal and bird species, including chickens, turkeys, water fowl, cattle, swine, rabbits, and domesticated dogs and cats (2-6). There are three *P. multocida* subspecies, subsp. *multocida, gallicida* and *septica* with strains differentiated based on fermentation of sorbitol and utilisation of trehalose (7). Strains of *P. multocida* can also be differentiated serologically by the presence of capsule and lipopolysaccharide (LPS) surface antigens. There are five capsule types (namely A, B, D, E and F) with strains from each type producing structurally different capsular polysaccharides (8). Capsule type can be identified by serology, or more commonly by a multiplex polymerase chain reaction (PCR) specific for type specific capsule polysaccharide biosynthesis genes (8, 9). Originally, serological methods were also used to differentiate *P. multocida* into 16 Heddleston types based on LPS antigens (10). Recently, Heddleston serotyping has been shown to be inaccurate and poorly reproducible and has been replaced by LPS genotyping via a multiplex PCR that targets type specific LPS outer core polysaccharide biosynthesis genes (11). There are eight LPS genotypes (namely L1 through L8), with strains from each of the eight LPS genotypes producing a structurally different LPS outer core polysaccharide structure (11, 12). Additionally, for some genotypes, strains can produce a shorter LPS structure due to point mutations within specific LPS biosynthesis genes (11, 12). Strains of *P. multocida* are often reported together with their capsule and LPS type (such a A:L1 or B:L2), and there is a correlation between specific capsule type, and host disease predilection (13).

1.2 *P. multocida* diseases

P. multocida is the causative agent of several different animal disease syndromes in different animal hosts (5). Animal diseases include fowl cholera (FC) in birds, haemorrhagic septicaemia (HS) in ungulates, atrophic rhinitis (AR) in pigs and rabbits, and bovine respiratory disease in livestock animals (13). These animal diseases can be primarily caused by *P. multocida* as is the case for FC and HS, or may also involve co-infection with another viral or bacterial respiratory pathogen as is the case for AR and bovine respiratory disease. These diseases range from self-limiting upper respiratory tract infections (URT) with little effect on the host, to acute or peracute infections that spread from the URT to the lower respiratory tract (LRT) or systemic infections. Peracute cases of disease can be fatal within 48 h of infection, with death often occurring before other clinical signs. Due to their high mortality, *P. multocida* diseases have a major economic impact on several agricultural industries (14, 15). As well as

high mortality, the bacteria can be easy transmitted between animals, as seen in large outbreaks of pasteurellosis in both livestock and wild animals (16, 17). Disease syndromes are often capsule type-specific, with strains from a subset of capsule genotypes responsible for each disease. *P. multocida* can also cause a range of serious zoonotic disease in humans, commonly following an animal bite.

1.2.1 Fowl cholera

FC is an peracute, acute, or chronic disease that can affect most bird species. *P. multocida* capsule type A and F, and LPS type L1 and L3 are most commonly isolated from cases of FC, however all capsule types excluding E and all LPS types excluding L7 have been isolated from cases of FC (6, 18-20). *P. multocida* initially colonises the upper respiratory tract (URT) epithelia before spreading to the lower respiratory tract (LRT) (21). The bacteria then enter the blood stream, likely through tissue that has been damaged by the immune response of the host (22, 23). Entry into the bloodstream via leg abrasions has also been reported indicating *P. multocida* may also enter the bloodstream via skin wounds (13). After entering the blood stream, the bacteria colonise and replicate in the liver and spleen (21). *P. multocida* cells may also be directly transported from the URT mucosa into the spleen and other tissues by macrophages (21). In late-stage FC, the bacterial numbers in the blood increase significantly. This severe bacteraemia and systemic infection stimulates a strong host immune response, with inflammation leading to widespread lesions and haemorrhaging that usually leads to death of the host (Figure 1) (20, 22). Lesions and haemorrhaging are primarily caused by the immune system of the host and are likely due to recruitment of heterophils. FC infections in chickens with depleted heterophil populations had reduced FC severity compared to untreated chickens (22, 24). Peracute cases of FC caused by highly virulent strains of *P. multocida* can be fatal within 48 hours (25, 26). In chronic cases of FC, *P. multocida* also causes a systemic infection, however a weak host immune response results in reduced clinical symptoms (27). *P. multocida* is cleared from privileged sites in the host but persists in the URT and LRT, maintaining *P. multocida* within wild and livestock bird populations (27, 28). Less pathogenic strains of *P. multocida* are more readily cleared by phagocytic uptake, likely contributing to clearance of less pathogenic strains from privileged sites, reducing the severity of the immune response (29)

P. multocida has been found as a commensal in several bird species, with carriage rates ranging between 6.2% and 25.6% for duck and chicken populations (4, 24, 30). *P. multocida* is readily transmitted between birds suffering FC and healthy birds, with healthy birds developing FC following transmission (24). Epidemiological data for fowl cholera is limited, however large outbreaks have been reported in several countries in both wild and domesticated birds, suggesting *P. multocida* strains capable of causing fowl cholera are

endemic to most countries (17, 31, 32). The mechanisms that determine whether *P. multocida* causes fowl cholera or not following transmission, and how *P. multocida* causes large FC outbreaks remains unclear.

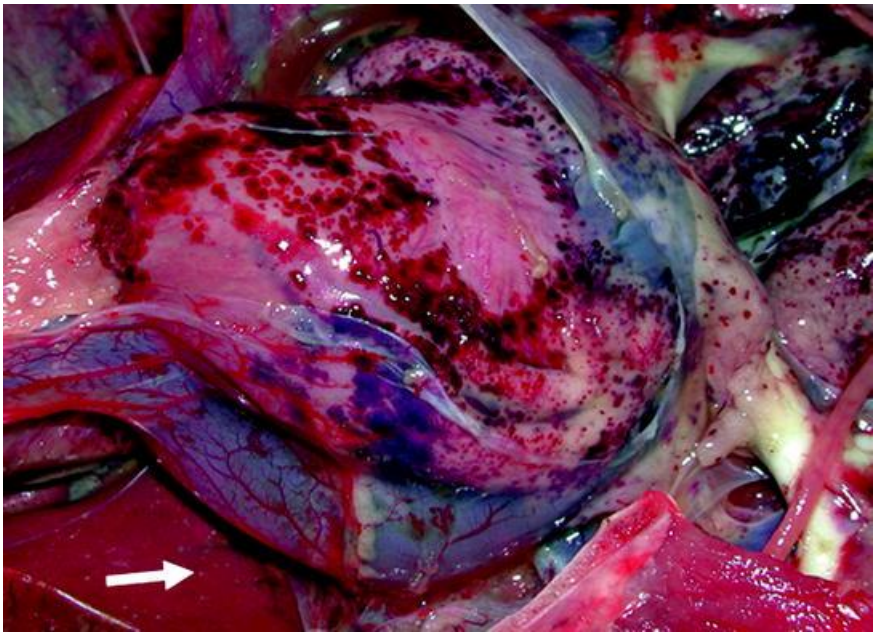


Figure 1.1: Field case of fowl cholera in a duck. Note multiple haemorrhages on serosal surfaces, particularly on the epicardium. There are also numerous small white foci of necrosis in the liver (arrow). Figure and legend taken from Wilkie *et al* 2012.

1.2.3 Haemorrhagic septicaemia

HS is an peracute and acute disease of ungulates, with buffalo and antelopes highly susceptible (15). Strains of *P. multocida* that produce a type B or E capsule, and LPS type L2, are the causative agents of HS (13, 33-35). *P. multocida* colonises and invades the URT mucosa at the tonsils, before being transported to the lymphatic system via macrophages (13, 36). The bacteria spread through the body via the lymphatic and circulatory systems, colonizing lymph nodes and tissues, resulting in widespread tissue necrosis and haemorrhaging on mucosal surfaces in the respiratory, gastrointestinal and urinary tracts, and in subcutaneous tissue (36). Endotoxin (lipid A of LPS) has been shown to cause septic shock during later stages of the disease (37). HS is endemic to several countries in Asia, Africa, and the Americas, causing large buffalo and cattle losses each year (15). Several HS outbreaks have been reported in wild ungulate populations, including a 2015 outbreak in Kazakhstan that caused 200,000 deaths in saiga antelopes (16, 38). HS is nearly always fatal as clinical signs of HS are only observed relatively late in this disease (39). In buffalo that have survived acute HS, *P. multocida* has been isolated from the URT, gastrointestinal tract and ureter (36).

Animals surviving HS harbouring *P. multocida* can become a reservoir for further HS outbreaks, with disruption of the immune system inducing secretion of *P. multocida* (36).

1.2.2 Atrophic rhinitis and snuffles

AR is a chronic disease that predominantly affects pigs and rabbits, but has also been observed in goats (13, 40, 41). Atrophic rhinitis is caused by the *Pasteurella multocida* toxin (PMT). PMT is produced by capsule type D and A strains of *P. multocida* (42, 43). Colonisation of *P. multocida* strains that cause AR generally requires a pre-existing infection or co-infection with another respiratory pathogen, such as *Bordetella bronchiseptica* or *Mannheimia haemolytica* (44). Following colonisation of the URT, *P. multocida* passively releases the PMT. PMT disrupts several host signalling pathways via modification of host heterotrimeric G proteins (45, 46). PMT dysregulation of osteoclast activity in the nose or snout stimulates bone reabsorption that leads to reduced bone density (47, 48). As a result, nasal bone atrophy occurs, and in some cases this leads to a complete loss of bone within the nose (49). If atrophic rhinitis is not treated, a secondary infection can occur, such as snuffles in rabbits. *P. multocida* cells are inhaled into the LRT where they can colonize and persist, causing severe pneumonia (50). Epidemiological data from several countries suggests toxigenic *P. multocida* strains are endemic in most countries.

1.2.4 Bovine respiratory disease

Bovine respiratory disease, or shipping fever, is a disease that affects livestock, including sheep, pigs and goats and is particularly prevalent in calves (51). Bovine respiratory disease ranges from a mild, self-limiting disease in sheep, pigs, and goats, to an acute and often fatal disease in cattle. *P. multocida* is one of several different bacterial species that has been identified as the cause of bovine respiratory disease; other common causative agents include *M. haemolytica* and *Histophilus somni* (52). *P. multocida* serogroup A, B and D strains are most commonly isolated from cases of *P. multocida* bovine respiratory disease (53). Bovine respiratory disease likely occurs when *P. multocida* from the URT is aspirated into the LRT, where clearance of bacteria by pulmonary macrophages is disrupted (13). Pre-existing or co-infection with other bacterial species such as *M. haemolytica, B. bronchiseptica, Mycoplasm* spp. or respiratory viruses such as bovine respiratory syncytial virus assist *P. multocida* colonisation and persistence in the LRT (54-57). Colonisation of the LRT by *P. multocida* generally results in pneumonia and widespread tissue necrosis in the lungs (13). As the signs of bovine respiratory disease caused by different pathogens are quite similar, it is likely that tissue damage is due primarly to the immune response of the host (13).

1.2.5 Human disease

P. multocida can cause a range of zoonotic infections. Isolates from human infections are typically capsule type A and F (58). The most common *P. multocida* zoonotic infections are skin infections following animal bite wounds (59-61). In severe skin infections, *P. multocida* can spread through fascial planes of the skin, causing necrotising fasciitis (62). *P. multocida* can also cause serious disease in immunocompromised individuals, including pneumonia, meningitis and septicaemia (61, 63-65). *P. multocida* isolates from human disease rarely show antibiotic resistance and are readily treated (58, 66). Very little is known about the *P. multocida* pathogenic mechanisms in human disease.

1.3 Virulence factors

Many *P. multocida* virulence factors have been identified, including capsule, lipopolysaccharide, PMT, filamentous haemagglutinin, fimbriae, and iron and nutrient scavenging systems (5). Of these virulence factors, capsule, LPS and PMT are the most well characterised. Other virulence factors have been identified via molecular methods such as signature tagged mutagenesis (STM), transcriptomics and microarray studies, or by whole-genome analysis that identifies genes with homology to virulence factors from closely related species such as *Haemophilus influenzae*, *Bordetella pertussis* and *B. bronchiseptica*. Apart from genes involved in capsule, LPS, PMT, few other virulence associated genes have been fully characterised. Given the wide range of hosts and disease syndromes caused by *P. multocida*, it is likely that there are many virulence factors yet to be identified. A more complete understanding of *P. multocida* pathogenic mechanisms and virulence factors would allow for novel therapeutic approaches to combat pasteurelloses.

1.3.1 Capsule

Bacterial capsule is an external polysaccharide layer that surrounds the cell wall and is an important virulence factor produced by several pathogens. Capsule helps pathogenic bacteria avoid recognition by the immune system, engulfment by phagocytes, and complement-mediated killing. *P. multocida* has five capsule serogroups (A, B, D, E and F), that are each comprised of different glycosaminoglycan polymers (Figure 1.2) (67-70). Glycosaminoglycans are polymers produced from repeating uronic sugar and amino sugar disaccharides, and are commonly found in animal extracellular tissue (71, 72). Mimicry of host polysaccharide likely helps *P. multocida* avoid immune recognition. Type A, D, and F capsule polymers are hyaluronic acid (HA), heparin and chondroitin, respectively (67-69). HA and heparin polymers are repeating N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA) with different linkages between monomers (69). Chondroitin is comprised of repeating N-acetylgalactosamine (GalNAc) and GlcA disaccharides (69). Capsule type B and E polymers are repeating GlcNAc

and N-acetylmannosaminuronic acid (ManNAc) disaccharides, with a fructose side chain attached to GlcNAc (70). Type B and E capsules have a different linkage between disaccharides and also differ in the position of the fructose side chain on GlcNAc (70). In addition, a glycine sidechain is attached to GlcNAc in type B capsule but this is not present in type E capsule (70).

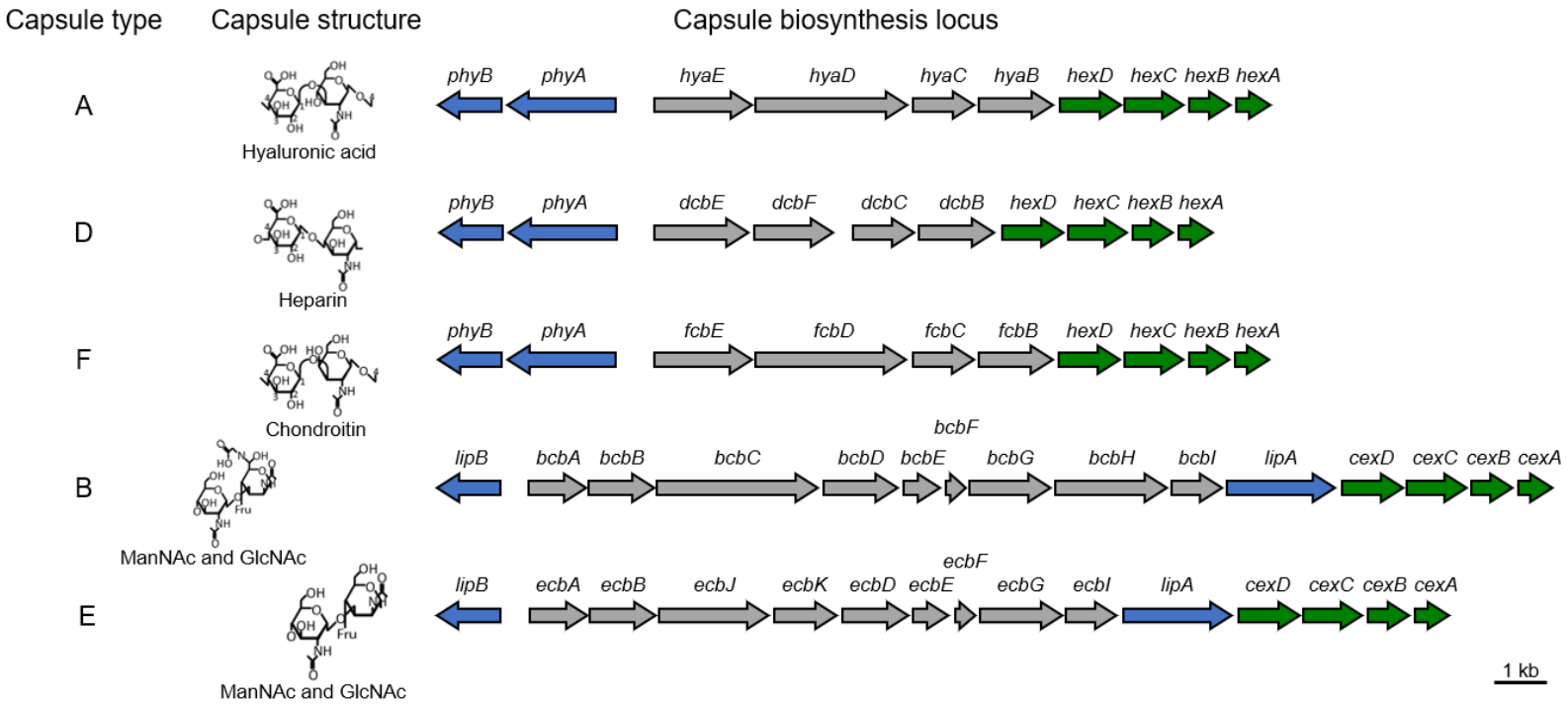


Figure 1.2: Capsule type, capsule repeating structure and capsule biosynthesis genes for the five capsule genotypes of *P. multocida.* Capsule structures represent repeating disaccharides for each capsule type. Genes involved in capsule export, attachment to phospholipids or capsule polysaccharide biosynthesis are coloured green, blue and grey, respectively. Scale bar represents 1 kb. Adapted from Townsend *et al* 2001 and Boyce pers. comm.

Capsule biosynthesis mechanisms are well conserved in Gram-negative bacteria despite the wide range of capsule polymers produced by different species. There are two main capsule biosynthesis mechanisms, ABC-transport-dependant and Wzy-polymerase-dependant systems, which are differentiated by the method used to export capsule to the cell surface (73, 74). *P. multocida* produces capsule using an ABC-dependant-transport system (75, 76). The capsule biosynthesis genes in all *P. multocida* strains are found in a single locus that can be split into three regions (Figure 1.2) (77, 78). Region one encodes the capsule ABC-transport system (HexABCD or CexABCD), region two encodes polysaccharide-specific biosynthesis proteins and region three encodes proteins predicted to be involved in anchoring the polysaccharide to phospholipids (PhyAB or LipAB). Homologous proteins encoded by genes in regions one and three have high amino acid identity across all serogroups and are predicted to have identical functions (Figure 1.2). HexABCD are homologs of capsule ABC-transport systems from other Gram-negative bacteria, including KpsTMED from *E. coli* (74). HexA contains ATP-binding domains and is likely the ATP-binding protein of the capsule ABC-transport system (76). HexB contains transmembrane domains and is predicted to form a channel across the inner membrane (76). HexC is a homolog of CtrB from *Neisseria meningitidis* that forms an oligomeric pore in the inner membrane that protrudes into the periplasm (79). HexD is a homolog of KpsD from *E. coli*, which has been shown to insert as a dimer into the *E. coli* outer membrane and also protrudes into the periplasm (80). Together HexC and HexD likely form the pore required for export of the polysaccharide through the periplasm to the surface of the cell. PhyA and PhyB are homologs of *E. coli* β 3-deoxy-D-manno-octulsonic acid (Kdo) transferases KpsC and KpsS, respectively, that synthesise a short Kdo polymer on a phosphatidylglycerol (PG) (81). KpsS adds the first Kdo residue to PG before KpsC adds the remaining Kdo residues to the polymer (81). Capsule is then synthesised from the Kdo polymer that is embedded in the outer membrane. KpsC and KpsS both have membrane association domains (81). In *E. coli*, interactions between KpsTMED and KpsS, as well as KpsC and capsule polymerase NeuS have been shown by *in vivo*-crosslinking and two hybrid screens respectively, showing capsule biosynthesis proteins form a complex at the inner membrane (80, 82). Thus, formation of a capsule biosynthesis complex would couple capsule biosynthesis with export.

Region two in the *P. multocida* capsule biosynthesis locus is variable. However, the proteins involved in biosynthesis of the A, D and F polysaccharides are highly conserved, as would be predicted given HA, heparin and chondroitin are highly similar in structure (Figure 1.2). Similarly, the proteins involved in biosynthesis of the structurally related B and E polysaccharides are also highly conserved (Figure 1.2). Type A, D and F each contain a capsule polymerase gene encoding HyaD (PmHAS), DcbD (HssA, PmHS) and FcbD (or

PmCS) that are HA, heparin, and chondroitin polymerases, respectively. *P. multocida* capsule polymerases are progressive glycosyltransferases that can add both monomers to a growing capsule polymer (83-85). These polymerases contain two separate glycosyltransferase domains that individually recognise and then add each monomer to the growing capsule polymer (83, 86, 87). In comparison, all other Gram-negative bacteria that have an ABC-transport dependant capsule biosynthesis system have two or more glycosyltransferases that each add an individual monomer to a growing capsule polymer (74). PmHAS contains a membrane association domain in the C-terminus, suggesting PmHAS interacts with other capsule biosynthesis proteins at the inner membrane (83). Recombinant *P. multocida* capsule polymerases have a faster rate of polymer synthesis when polymer acceptors are present *in vitro* and have almost no initiation activity, suggesting a short capsule polymer would be required *in vivo* for efficient capsule production (84, 85, 88). Proteins HyaB, DcbB and FcbB share 98% amino acid identity, and proteins HyaE and FcbE share 97% amino acid identity, indicating these homolog groups likely have identical function (9). These proteins have weak identity to glycosyltransferases, and could add capsule monomers onto the Kdo polymers, providing capsule synthases with a polymer acceptor *in vivo* (81). DcbC and FcbC share 92% and 98% amino acid identity with HyaC, respectively. HyaC is a UDP-glucose dehydrogenase that synthesises UDP-GlcA from UDP-glucose (89). Several *P. multocida* type A, D and F strains have a functional alternative heparin synthase, HssB, that is encoded outside of the capsule biosynthesis locus (90). In a type F strain, the *hssB* gene was shown to have 10-fold less expression compared to *fcbD* (synthase gene in the capsule locus) during mid-exponential phase growth, indicating *hssB* is not transcriptionally linked to the capsule biosynthesis locus (90). Production of heparin in type A and F strains in stad of HA and chondroitin may allow for capsule variation in response to different growth conditions; however, heparin has not yet been identified in capsule isolated from *P. multocida* type A and F strains. Capsule type B and E strains contain nine genes in region two that to date have not been functionally characterised. BcbA/EcbA and BcbB/EcbB are homologs of WecB and WecC from *E. coli* that synthesise UDP-ManNAc from UDP-GlcNAc (77). BcbC contains a glycosyltransferase domain, suggesting it is a synthase for type B capsule (77).

Two positive regulators of capsule biosynthesis have been identified in the *P. multocida* strain VP161 (A:L1). Capsule biosynthesis genes in type A strains are predicted to be under the control of two divergent promoters found between *phyA* and *hyaE* (91, 92). Fis is a nucleoid-associated protein that has been shown in *E. coli* to regulate a large number of genes during early exponential stage growth (93, 94). Fis has been shown to regulate capsule in *P. multocida* strain VP161 (A:L1), with three putative Fis binding sites identified between *phyA* and *hyaE* based on similarity to the *E. coli* Fis binding consensus sequence (92). *P. multocida*

strain VP161 *fis* mutants had reduced capsule biosynthesis gene expression and reduced HA capsule production, with complementation restoring capsule biosynthesis expression and HA production to that of the wild-type parent strain (92). The RNA-binding chaperone Hfq is also involved in capsule gene expression in *P. multocida*. Hfq mediates the interaction between small regulatory RNAs and their mRNA targets, resulting in either increased or decreased transcription, or degradation of the mRNA target (95). A directed VP161 *hfq* mutant had reduced capsule gene expression and capsule production compared to the wild-type parent strain (96). Given Hfq mediates regulation by sRNAs, it is likely that one or more Hfq-dependant sRNAs target capsule biosynthesis gene transcripts but these have yet to be identified.

P. multocida capsule has been shown to be essential for virulence in both capsule type A and B strains. Targeted mutagenesis of *cexA* in the *P. multocida* haemorrhagic septicaemia isolate M1404 (B:L2) abrogated capsule production (75). The virulence of the acapsular M1404 *cexA* mutant was highly attenuated; the 50% infectious dose (ID_{50}) of the mutant strain in mice was 10^7 colony forming units (CFU) compared to an ID_{50} of < 10 CFU for the wild-type parent strain (75). The M1404 *cexA* mutant was also four-fold more susceptible to uptake by mouse macrophages *in vitro*, compared to the wild-type parent strain (75). A targeted *hexA* mutant in the *P. multocida* fowl cholera isolate X73 (A:L1) was produced no HA capsule and caused no disease in chickens at a high infectious dose of 2.6×10^8 CFU, compared the wild-type parent strain that caused severe disease at an infectious dose of 6.5×10^4 CFU (76). The X73 *hexA* mutant was also more susceptible to killing by chicken serum compared to the wild-type parent strain (76). A global virulence study in mice using a *P. multocida* bovine pneumonia isolate UC6731 (A:L3) identified a highly attenuated *hyaD* (*has*) mutant (97). The UC6731 *hyaD* mutant was unable to cause disease in mice at an infectious dose of 6.2×10^5 CFU, whereas the wild-type parent strain had a 50% lethal dose (LD_{50}) of < 10 CFU (97).

1.3.2 Lipopolysaccharide

LPS is the main component of the outer leaflet of the outer membrane and is present in almost all Gram-negative bacteria. LPS has up to three regions, a hydrophobic lipid A region, a non-repeating core polysaccharide consisting of an inner core and outer core region, and a repeating O-antigen (98). Like many other Gram-negative species, *P. multocida* LPS does not contain an O-antigen component, with variability in LPS types due primarily to different outer core polysaccharide structures. *P. multocida* strains are currently typed using an LPS multiplex PCR that identifies type-specific regions present in the outer core LPS biosynthesis loci (11). Prior to genotyping, stains could only be typed by serology into 16 Heddleston serovars; however, Heddleston typing has been shown to be highly inaccurate and poorly reproducible, often identifying the incorrect serovar or assigning multiple serovars for the same

strain (10, 11). Identification of *P. multocida* LPS core polysaccharide structures and biosynthesis genes showed there were only eight conserved LPS outer core genotypes, but more than eight outer core structures were possible due to mutations within these loci (Figure 1.3) (99-104). For example, investigation of the outer core polysaccharide structure in 23 *P. multocida* strains belonging to the LPS genotype L3 identified a full-length structure of Lipid A-Hep-[Glc]-Gal-Gal-GalNAc-GalNAc (where Hep: heptose, Glc: glucose, Gal: galactose) (101). Five variants were identified that had the same LPS outer core locus but had inactivating mutations in different outer core glycosyltransferase genes (101). The role of non-functional glycosyltransferases in LPS outer core assembly determined the length of the variant LPS outer core structures (101). In addition to different outer core polysaccharide structures, all *P. multocida* strains produce two inner core polysaccharide glycoforms, A and B (105). Glycoform A and B differ by the number of Kdo residues; only one Kdo is present in glycoform A and two are present in glycoform B (105). Glycoform A also has additional phosphoethanolamine (PEtn) and glucose restudies compared to glycoform B (105).

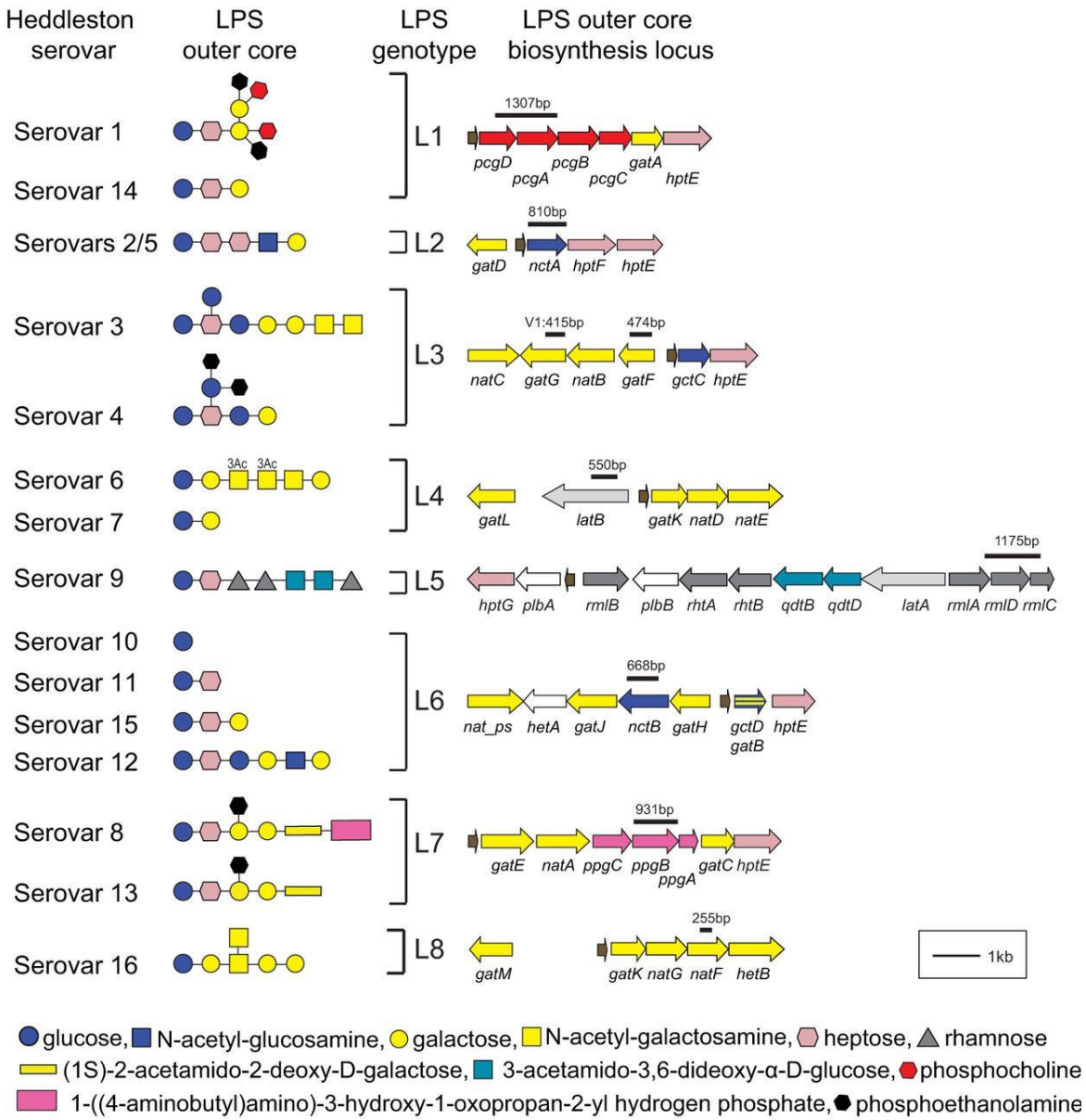


Figure 1.3: LPS outer core structure produced by each of the Heddleston serovar type strains and the genes responsible for LPS outer core biosynthesis in each strain. (Left) Schematic representation of the outer core LPS structures produced by each of the Heddleston serovar type strains. The last residue (glucose) of the conserved LPS inner core is shown on the far left as a reference point. Specific linkages between each of the residues are not shown. (Right) LPS genotype and genetic organization of each LPS outer core biosynthesis locus. The relative position and size of each genotype-specific PCR amplicon are shown above each LPS outer core biosynthesis locus. Each gene is color coded according to its known/predicted role in LPS biosynthesis; *gctD* and *gatB* (yellow and blue striped) in locus L6 differ by only a single nucleotide and are involved in the addition of glucose or galactose, respectively, to the outer core heptose. The *rpL31_2* gene, encoding ribosomal protein L31, is not involved in LPS biosynthesis and is colored brown. Figure and legend taken from Harper *et al* 2015.

Several studies have shown that truncation of the LPS structure reduces virulence. STM of strain UC6731 in mice and VP161 in chickens identified a *gatG* and *hptD* mutants has having reduced virulence, respectively (97, 106, 107). The genes *gatG* and *hptD* encode for an LPS outer core glycosyltransferase an inner core heptosyltransferase, respectively, indicating that truncating either the inner or outer core polysaccharide can result in reduced virulence. In addition, a VP161 *pcgC* mutant, which lacked phosphocholine (PCho) at the distal end of the LPS outer core structure displayed reduced growth in mice and had increased sensitivity to the chicken antimicrobial peptide fowlicidin-1 compared to the wild-type parent strain (101). In comparison, the expression of the glycoform B inner core is not essential for virulence. Mutagenesis the glycoform B specific glycosyltransferase *hptB* in *P. multocida* strain VP161 resulted in strains that only produced LPS with only glycoform A, but still produced a full-length outer core structure (105). The *hptB* mutant was fully virulent in chickens, showing that the expression of both LPS glycoforms was not required for virulence (105).

1.3.3 *Pasteurella multocida* toxin

PMT is produced by toxigenic capsule type A and D strains is the causative agent responsible for the atrophy of nasal bones seen in atrophic rhinitis (42, 108). The toxin is encoded by the *toxA* gene that is found within a 53 kb region that contains a λ-like prophage belonging to the Siphoviridae family (109, 110). PMT is not actively secreted by toxigenic *P. multocida* strains but is passively released upon cell lysis (111). The prophage containing PMT is excised from the genome following mitomycin C treatment of toxigenic cells, and bacteriophage are released following growth of toxigenic *P. multocida* strains, suggesting PMT is likely released from cells following the prophage entering the lytic cycle (109). Several non-toxigenic strains of *P. multocida* have genomic regions that match regions of the 53 kb *toxA* containing prophage, but lack the *toxA* gene (110). This indicates *toxA* may have been widespread in *P. multocida* before being lost by some strains.

PMT is a 145 kDa protein that forms a typical A-B type toxin. The receptor binding and translocation domains of PMT are found in the N-terminus (112, 113). Green fluorescent protein fused to the N-terminus was internalised by endocytosis into lysosomes within target cells, before being delivered into the cytosol (112). The C-terminal region of PMT contains three different domains, denoted C1, C2, and C3 (113). The C1 domain contains four helices essential for targeting the toxin to the plasma membrane (113, 114). There is no reported function for the C2 domain of PMT. The C3 domain is the biologically active domain of the toxin and is responsible for the modification of host heterotrimeric G proteins (45). G proteins are involved in signal transduction and are normally activated and inactivated by GTPases (115). PMT deamidates the α-subunit of G proteins (116), stopping GTPase interaction with G proteins, leaving the modified G proteins in a constitutively active conformation (116).

Several G family proteins are activated by PMT, resulting in the activation of mitogenic, calcium, cytoskeletal and osteogenic signalling pathways (46). One key process that is altered is the differentiation of preosteoclasts into osteoclasts (49). The increase in the number of host osteoclasts at the site of infection results in an increase in bone resorption, leading to atrophy of nasal bones observed in severe cases of atrophic rhinitis (49).

1.3.4 Adhesins and outer membrane proteins

P. multocida produces several putative adhesins, including filamentous haemagglutinin, a type IVa fimbriae and a fimbrial low-molecular weight (Flp) pili. Two putative filamentous haemagglutinin genes *fhaB1* and *fhaB2* (also known as *pfhB1*/*pfhaB1* and *pfhB2*/*pfhaB2*) were identified as important for virulence. Transposon mutants of *fhaB1* and *fhaB2*, and *fhaC2* (also known as *pfhaC2*) that encodes the cognate secretion partner for PfhB2 all displayed reduced virulence in mice compared to the wild-type parent strain (97). In addition, a targeted fhaB2 mutant in *P. multocida* strain P1059 (A:L3) was significantly attenuated in turkeys, with a 100% lethal dose (LD_{100}) of 6.6 x 10^6 CFU compared to the wild-type parent strain that had an LD_{100} of 5.5 x 10^2 CFU (117). *P. multocida* FhaB1 and FhaB2 are homologs of filamentous haemagglutinins (FhaB superfamily proteins) found in *Bordetella* spp. and *Haemophilus ducreyi* that have been shown to be important for adhesion to the upper respiratory mucosa and for avoiding phagocytosis, respectively (118-121). Both Fis and Hfq have been shown to regulate the expression of filamentous haemagglutinin genes in *P. multocida*. Independent VP161 *fis* and *hfq* mutants had reduced expression of *fhaB2* and *fhaC2*, with proteomics also showing reduced FhaB2 and FhaC2 production in the *hfq* mutant (92, 96). Two additional genes encoding putative filamentous haemagglutinins have been identified from analysis of the whole genome sequence of *P. multocida* strain P1059 (A:L3) (122).

Type four fimbriae are differentiated into either type IVa or type IVb based on the pilin subunit structure. Fimbriae were identified on the surface of *P. multocida* of several capsule type A, B and D strains by electron microscopy and N-terminal sequencing of isolated fimbriae peptides showed identity to other type IVa fimbriae pilins (123). The *P. multocida* type IVa pilin gene is homologous to other type IVa pilins including *pilA* from *Pseudomonas aeruginosa* and *pilE* from *Neisseria gonorrhoeae* (124). The role of type IVa fimbriae in *P. multocida* pathogenesis is yet to be determined, however microarray analysis of *P. multocida* strain X73 showed the pilin subunit gene *ptfA* had higher expression in chickens compared to growth in rich media (125). However, a strain VP161 *ptfA* mutant was not attenuated for virulence in chickens (Boyce and Harper pers. comm.). *P. multocida* has a locus with strong similarity to the tight adherence (Tad) locus from *Aggregatibacter actinomycetemcomitans*, which encodes a type IVb Flp-pilus (126). The Tad-locus contains genes *flp1-flp2-tadV-rcpCAB-tadZABCDEFG*, with all genes required for production of Flp-pili in *A. actinomycetemcomitans*, except *flp2*

(127-130). Genomic analysis of several *P. multocida* strains showed the Tad-locus is present in most strains, however some strains did not have a *flp2* homolog (110). Flp-pili are required for *A. actinomycetemcomitans* non-specific adherence and biofilm formation, and in *H. ducreyi* Flp-pili are required for virulence in humans (128, 129, 131). The Flp-pili biosynthesis genes *flp1* and *tadD* were identified as important for virulence by STM analysis of VP161 and UC6731, respectively, suggesting that Flp-pili are likely important for virulence in *P. multocida* (97, 106).

Several outer membrane proteins, porins and lipoproteins have been identified as important in *P. multocida* (132). Most *P. multocida* outer membrane proteins and lipoproteins have been identified via homology to characterised proteins from other bacterial species and many are yet to be functionally characterised. The membrane stabilisation protein OmpA functions as an adhesin in *P. multocida*, binding to fibronectin and heparin (133). *P. multocida* ComE1 has also been shown to bind to fibronectin (134). Both fibronectin and heparin are found within mammalian extracellular matrix (71). The ability of OmpA and ComE1 to bind extracellular matrix components likely aids *P. multocida* persistence in the host. In addition, ComE1 was able to bind double-stranded DNA suggesting a role in DNA uptake, however natural competence has not been shown in *P. multocida* (135). *P. multocida* PlpB is part of a methionine ABC-transport system, with a VP161 *plpB* mutant displaying reduced virulence in chickens (132). Several *P. multocida* outer membrane and lipoproteins are immunogenic and have been investigated as potential vaccine targets (136, 137). Most outer membrane-based vaccines only provide partial protection; however, a recombinant PlpE vaccine has been shown to provide strong homologous and heterologous protection against disease in both mice and chickens (137, 138).

1.3.5 Iron and nutrient scavenging

P. multocida has several different iron receptors and iron-specific ABC-transport systems. Sequencing of the *P. multocida* strain Pm70 genome identified 53 genes encoding proteins with high identity to iron acquisition proteins from other bacterial species including the iron ABC-transporters AfuABC, FbpABC, YfeABCD, FecBCDE, the TonB-ExbBD energy transduction system and iron receptors HasR, HemR, HbpA and PfhR (118). Several *P. multocida* outer membrane proteins have been shown to bind haemoglobin and/or haemin, including HgbA, HgbB, Pm0741, Pm1081, Pm1282, Pm1428, and the receptors identified in the Pm70 genome (HasR, HemR, HbpA and PfhR) (139-143). Generally, single mutants of genes encoding iron receptors do not display reduced virulence, indicating a level of functional redundancy (140, 143). However, a *hgbA* mutant in strain UC6731 displayed reduced virulence (97). The transferrin receptor TbpA has also been identified in *P. multocida*, but only in bovine respiratory isolates (144). *P. multocida* TbpA was able to bind bovine transferrin

without a secondary receptor; most other bacterial species have a dual TbpAB system (144). The TonB-ExbBE energy transduction system is required for virulence in *P. multocida*, as an *exbB* mutant in strain UC6731 displayed reduced virulence (97). The TonB-ExbBE energy transduction system utilises ATP at the inner membrane and provides energy for outer membrane proteins to import iron into the periplasm (145). *P. multocida* encodes a predicted homolog of the global iron-regulatory protein Fur, which is found in several bacterial species (146). A *fur* mutant in *P. multocida* strain 0818 had an LD_{50} of 1.65 x 10^8 CFU compared to the wild-type strain LD_{50} of 1.13 x 10^6 CFU and the mutant showed increased sensitivity to killing by duck serum (147). Taken together, these data show that Fur is important for *P. multocida* virulence. DNA microarrays have been used to identify changes to global gene expression following growth of strain Pm70 in iron-limited media or in media with defined iron sources (148, 149). Iron limitation resulted in increased expression of iron-acquisition genes including iron receptors, ABC-transport systems and the TonB-ExbBD system, as well as increased expression of stress response (*recX, hslV* and *htrA*) and cell membrane biosynthesis genes (*gmhA*, *skp*, *hexC*, *ponB*) (148). Increased expression of these genes is likely due to Fur responding to reduced iron availability and alleviating repression of these genes. Regulatory changes were different depending on the iron source available, showing there are likely iron source-specific regulators in *P. multocida* that have yet to be characterised (149). A large proportion of the genes identified as showing altered expression during growth in media with low iron or defined iron sources had unknown functions, showing *P. multocida* may have additional iron acquisition genes (148, 149). ABC-transport systems involved in iron uptake are yet to be further characterised in *P. multocida*.

P. multocida can scavenge sialic acid from the host. *P. multocida* strains produce up to two sialidases, NanH, a 2-3' sialyl lactase and NanB, a 2-3' and 2-6' sialyl lactase (150). Proteins involved in sialic acid uptake and metabolism were identified in *P. multocida* via homology to known sialic acid utilisation proteins from other bacterial species (151). This analysis identified the sialic acid TRAP-transporter NanU, the secondary binding protein NanP, the sialic acid metabolism proteins NanAKE-NagAB, and NeuAS1S2-Pm1710 that are involved transfer of sialic acid onto surface structures (151). NeuP was shown to be required for transport of sialic acid, and NeuA and Pm1710 were shown to be sialic acid cytidylyltransferases (151). A *P. multocida* strain P1059 *nanPU* double mutant was avirulent when injected intramuscularly into turkeys at a dose of 2.9 x 10^4 CFU, compared to the wild-type that caused disease at 2.8 x 10^2 CFU, showing sialic acid uptake is required for *P. multocida* virulence (152). Despite the importance of sialic acid uptake for *P. multocida* virulence, the specific mechanism of utilisation is unclear. In some other bacterial species, including *H. influenzae*, sialic acid is attached to LPS (153). In *P. multocida*, sialic acid was shown to be transferred onto a low

molecular weight product similar to LPS in strain P1059, however sialic acid has not been identified in any *P. multocida* LPS core polysaccharide structures to date (12, 153). It is unclear if sialic acid metabolism is required for pathogenesis. A *nanE* mutant in strain UC6731 had an LD_{50} of 3 x 10^5 CFU compared to the wild-type LD_{50} of < 10 CFU, suggesting *P. multocida* may utilise sialic acid as an essential carbon source *in vivo* (151). However, a *nanA* mutant, which encodes a protein acting in the same catabolic pathway was not attenuated.

1.3.6 Antibiotic resistance

Antibiotic susceptibility testing of *P. multocida* isolates shows that antibiotic resistance is not widespread in *P. multocida* (154, 155). To date, antibiotic resistance genes in *P. multocida* have only been identified on mobile-genetic elements, with no resistance observed due to mutations of antibiotic targets on the chromosome. A number ColE1-family plasmids have been identified in *P. multocida* (156). These ColE1 plasmids contain antibiotic resistance genes that confer resistance to β-lactams, sulfonamides, tetracycline, chloramphenicol, florfenicol, streptomycin and spectinomycin (157-161). Four different tetracycline resistance genes, *tetH*, *tetO*, *tetB* and *tetG*, as well as four different spectinomycin/streptomycin resistance genes, *strA*, *strB*, *aadA1* and *aadA14* have been identified on *P. multocida* plasmids (157, 160, 161). Antibiotic resistance genes in *P. multocida* have also been identified on integrative conjugative elements (ICE). Genome sequencing of the multidrug resistant bovine isolate 36950 identified an 82 kb ICE designated ICE*Pmu1* (162). ICE*Pmu1* had several homologs of genes on ICEs from *Histophilus somni* and *H. influenzae*, with many of these genes conferring antibiotic resistance. The *erm(42)*, *msr(E)* and *mph(E)* genes found on ICE*Pmu1* were identified as novel antibiotic resistance genes conferring resistance to macrolides, triamilides and lincosamides (163). Comparative analysis of whole-genome sequences from 12 haemorrhagic septicaemia and four avian and porcine *P. multocida* isolates identified a putative ICE designated ICE*Pmu2* (164). ICE*Pmu2* shared 35 homologs with ICE*Pmu1* and contained several antibiotic resistance genes, including chloramphenicol resistance gene *catA2* and a β-lactam resistance gene $bla_{TEM\text{-}1B}$, that were not present in ICE*Pmu1* (164).

1.4 *P. multocida* genomics and phylogenetics of *Pasteurella* spp.

1.4.1 Genomics

The first *P. multocida* genome sequenced was of strain Pm70 (118). Since the Pm70 sequence was published in 2001, the number of publicly available whole genome sequences for *P. multocida* has increased but has generally lagged behind other bacterial species. Despite the relatively low number of available genome sequences, several comparative genomic analyses have been conducted to try and identify genes specific for particular

diseases and/or hosts. Comparative genomics of 12 haemorrhagic septicaemia and four avian and porcine *P. multocida* isolates described above also identified two large regions of 15 kb and 34 kb, as well as 96 genes that were present only in haemorrhagic septicaemia isolates (164). Most of the haemorrhagic septicaemia-specific genes encoded putative bacteriophage products that may contribute to pathogenesis and/or bovine predilection of these strains (164). In addition, intact 44 kb and 50 kb bacteriophages were identified, but these were only found in *P. multocida* genomes of Asian or Pakistani origin (164).

Comparison of strain Pm70 and a closely related strain HN07 (F:L3) identified an ICE designated ICE*pmcn07* in HN07, which encodes a putative type IV secretion system and several antibiotic resistance genes (165). Comparison of highly virulent avian isolates X73 and P1059 and avian isolate Pm70 identified 336 genes unique to X73 and P1059, including L-fucose utilisation and ribose specific ABC-transport systems, a putative TonB-dependant iron receptor, genes encoding a trimetylamine-N-oxide reductase system, and two putative filamentous haemagglutinins (122). Genes identified in these studies were thought to be possible virulence genes as some isolates of Pm70 have been shown to be avirulent in chickens. However, it is now apparent that Pm70 can cause disease in chickens, although some laboratories received Pm70 with a single point mutation within an LPS biosynthesis gene, resulting in loss of virulence (Harper pers. comm.). Since these studies, an additional 254 *P. multocida* strains have been whole-genome sequenced. Investigation of 109 *P. multocida* genome sequences identified a core genome of 1,806 genes (166). Comparison of strains from different hosts identified no genes associated with predilection for a single host, suggesting there may be no host-specific virulence factors, or that multiple factors interact to determine host-specificity (166). Single nucleotide variants (SNVs) in 653 *P. multocida* strains compared to *P. multocida* strain ATCC 43137 identified 237,670 SNVs and allowed generation of a *P. multocida* SNV phylogeny, showing human *P. multocida* isolates clustered away from most animal isolates (167). Pan genome analysis identified several proteins over-represented in strains belonging to particular capsule types, including PtfA that was primarily found in capsule type B and F strains, HgbA which was primarily found in capsule type F strains, and OmpH1, OmpH3, PlpE and FhaB1 that were strongly associated with capsule type A strains (167).

1.4.2 Phylogenetics

Several *P. multocida* phylogenies have been produced using different methods. Whole 16s rRNA and MLST phylogenetic analyses have identified two clear clades, separating *P. multocida* subsp. septica isolates and several avian isolates away from other *P. multocida* strains (34, 168). Recent core-genome *P. multocida* phylogenies separated *P. multocida* strains into multiple clades containing several closely related strains (166, 169). Strains from

these clades were isolated from just one or two hosts, suggesting that these strains have a narrow host range (166). *P. multocida* is one of 13 recognised *Pasteurella* species that includes *aerogenes*, *bettyae*, *caballi*, *caecimuris*, *canis*, *dagmatis*, *langaaensis*, *mairii*, *multocida*, *oralis*, *skynesis*, *stomatis*, *testudinis*, with the genus belonging to the Pasteurellaceae family (https://lpsn.dsmz.de/genus/pasteurella) (170). *Pasteurella* sp. *aerogenes*, *bettyae*, *caballi*, *canis*, *dagmatis* and *stomatis* cause a range of human diseases, including skin infections following animal bite wounds, urethritis, peritonitis and bacteraemia (171-175). *P. caballi* has been shown to cause systemic infections in horses, resulting in endocarditis (176). Phylogenetic analysis of the Pasteurellaceae family using 16s rRNA genes and housekeeping genes show *Pasteurella* sp. *multocida*, *canis*, *dagmatis*, *stomatis* cluster together into what is known as *Pasteurella sensu stricto*, away from spp. *mairii*, *aerogenes*, *testudinis*, *bettyae*, *caballi* and *langaaensis* suggesting these are unlikely to be true *Pasteurella* species (177, 178).

1.5 Molecular studies performed in *P. multocida*

P. multocida is genetically amenable, but directed mutagenesis is more time consuming than in many other bacterial species. Other global molecular methods for probing gene function have been important for understanding *P. multocida* pathogenesis, including DNA microarrays, transcriptomics, proteomics, and STM. The first global studies in *P. multocida* used DNA microarrays to compare total bacterial gene expression (the transcriptome) under different growth conditions or between a mutant and wild-type parent strain. Genes with large changes to expression are likely important for growth in the tested condition or have altered regulation due mutagenesis of a target gene. DNA microarrays have been used to investigate *P. multocida* genes differentially expressed following growth in low iron or defined iron sources, as well as under nutrient limitation. In strain Pm70, 669 genes showed altered expression during growth in minimal media. Genes involved in energy metabolism and protein synthesis showed reduced expression in minimal media while genes involved in amino acid biosynthesis, transport and heat shock systems showed increased expression (179). Microarrays have also been used to identify changes in *P. multocida* expression following growth in a natural host. During growth of strain X73 in chickens, genes involved in amino acid biosynthesis, energy metabolism, type IVa fimbrial biosynthesis and several hypothetical genes showed increased expression compared to X73 cells grown in rich medium (125). However, it should be noted that DNA microarrays are inherently biased as expression is only measured for fragments present on the microarray.

RNA-sequencing overcomes microarray bias as all bacterial RNAs isolated from the bacteria are sequenced, allowing identification of total bacterial gene expression. Transcriptomics is often coupled with proteomic analysis, where protein levels are identified by mass

spectrometry. Similar to microarrays, transcriptomics and proteomics are used to identify changes to gene expression or protein production between two growth conditions, or between two strains. Transcriptomic and proteomic analysis have been used to identify genes that are regulated by the small RNA chaperone Hfq, and the small RNA GcvB in *P. multocida* strain VP161 (96, 180). In the *hfq* mutant, 109 genes and 49 proteins showed increased production, and 19 genes and 29 proteins showed decreased production compared to the wild-type parent strain at mid-exponential phase growth. As such, the Hfq chaperone primarily acts with sRNAs to repress expression/production of their targets. Expression of several virulence factor genes was altered in the *hfq* mutant, including capsule biosynthesis genes, filamentous haemagglutinin genes and genes encoding the LPS-specific PCho and PEtn transferases (96). Proteomic analysis of a VP161 sRNA *gcvB* mutant strain showed 36 proteins with increased production and 10 proteins with decreased production compared to wild-type VP161 (180). Of the proteins with increased production, 27 were involved in amino acid scavenging or biosynthesis, showing that the sRNA GcvB negatively regulates production of proteins involved in amino acid uptake and biosynthesis (180). Transcriptomics has also been used to investigate *P. multocida* gene expression following growth *in vivo*. Gene expression was compared between closely related virulent strain PmCQ2 and attenuated strain PmCQ6 (both capsule type A) following growth in mice (181). PmCQ2 had higher-fold change increase to expression of virulence related genes *in vivo* compared to PmCQ6 (181). Genes with increased expression in PmCQ2 included iron acquisition and haem utilisation genes, capsule and LPS biosynthesis genes, and *ompP5* (181). Although gene and protein expression analyses give valuable information about putative niche-essential genes, these methods do not directly identify essential genes for a given growth condition. Any genes with altered expression or proteins with altered production need to be be further characterised using mutagenesis to confirm they are important for under that growth condition.

STM is an unbiased, whole-genome approach for identifying genes essential for growth under particular conditions (182). A large pool of transposon mutants is generated, with each mutant having a uniquely tagged transposon. The mutant library is then grown in a condition of interest, with mutants that have insertions in genes essential for that condition being lost during growth. Mutant strains that cannot grow are identified by loss of their corresponding transposon tag, allowing essential genes for that condition to be identified (182). Two STM studies have been performed in *P. multocida* (97, 106). STM analysis of *P. multocida* strain UC6731 identified 34 genes essential for growth in mice (97). Several of these genes have been shown to encode virulence factors (see section 1.3). In addition, several housekeeping and hypothetical genes were identified (97). STM analysis of *P. multocida* strain VP161 identified 15 genes essential for growth in chickens, and five genes essential for growth in

mice, including several encoding confirmed virulence factors (106). Other genes identified as essential for VP161 growth in animals included a homolog of OMP P1 from *H. influenzae*, a peptidoglycan biosynthesis gene *ponC* and genes involved in nucleoside biosynthesis (106). Although STM allows for investigation of gene function by mutagenesis, the method suffers from low throughput. In the studies described above, approximately 2000 different mutant strains were screened (97). Given the low coverage, many essential genes were likely not identified in these studies.

1.6 Transposon insertion sequencing

Transposon insertion sequencing (TIS) methods allow for all genomic regions important for a particular bacterial phenotype to be identified (Figure 1.4) (183, 184). TIS methods utilise random transposon mutagenesis to produce a very large number of single transposon mutants, ideally achieving saturation of the genome across the whole library of mutants. The transposon mutant library is then grown in a condition of interest, or the mutant library is screened to identify mutants with a particular phenotype. Genomic DNA is recovered from all of the surviving or phenotype-positive transposon mutants and used to produce library of genomic DNA-transposon junctions that is then sequenced to identify all transposon insertion sites in recovered mutants. Analysis of the full set of insertion sites identifies genomic regions that are important for growth under the specific condition, or for the selected phenotype. TIS encompasses several related methods, including transposon-directed-insertion site sequencing (TraDIS), transposon sequencing (Tn-Seq), insertion sequencing (INSeq) and high-throughput insertion tracking by deep sequencing (HITS) (185-188). All of these methods have the same principal of coupling random mutagenesis with high-throughput sequencing with only minor differences in library preparation methods. TIS experiments overcome many of the weaknesses of STM described above as they do not require barcoding of individual mutants as in STM experiments, greatly increasing the number of mutant strains that can be tested. Additionally, TIS methods investigate gene importance by direct mutagenesis, rather than changes to gene expression or protein production.

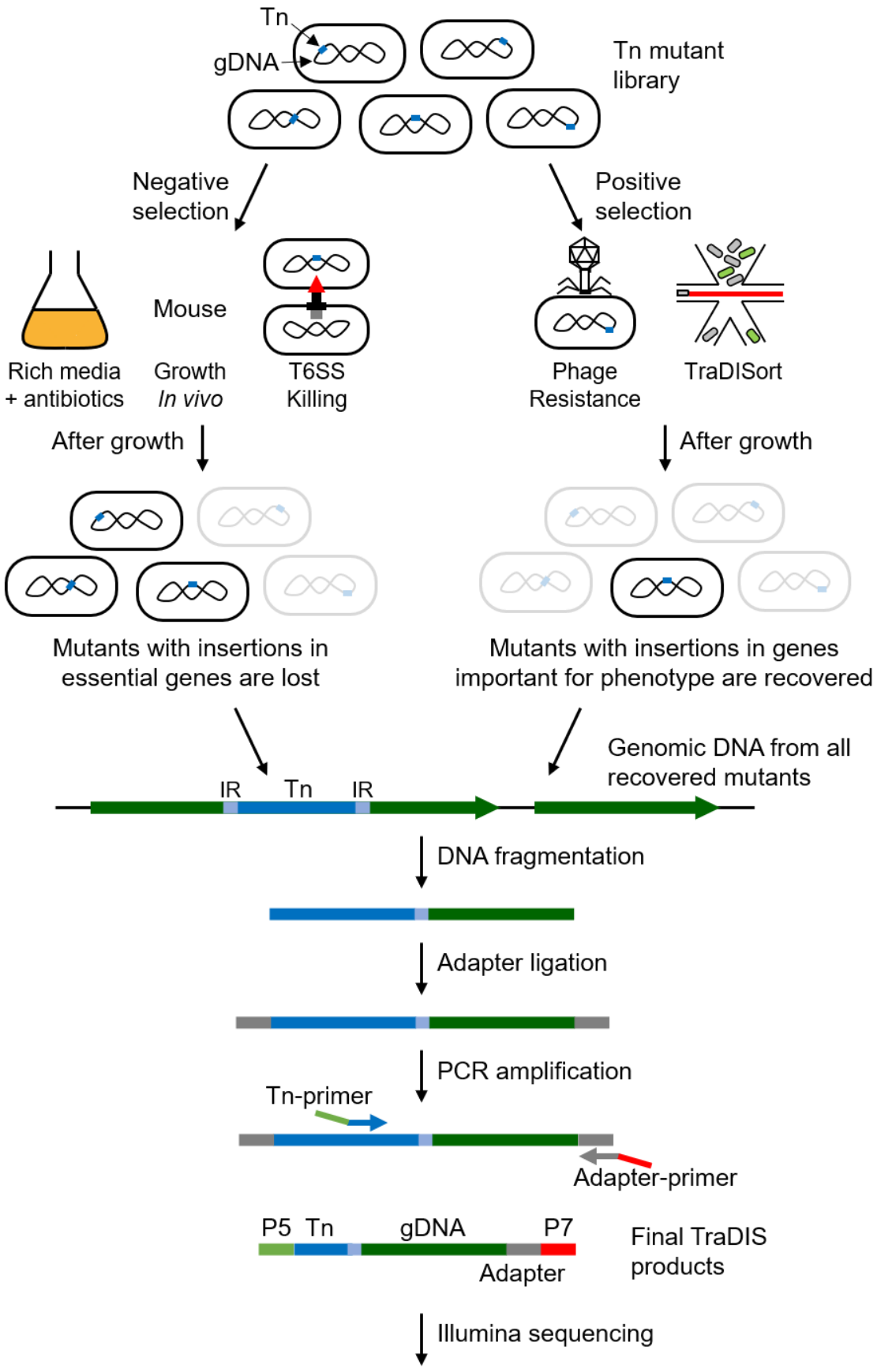


Figure 1.4: Figure continues over page.

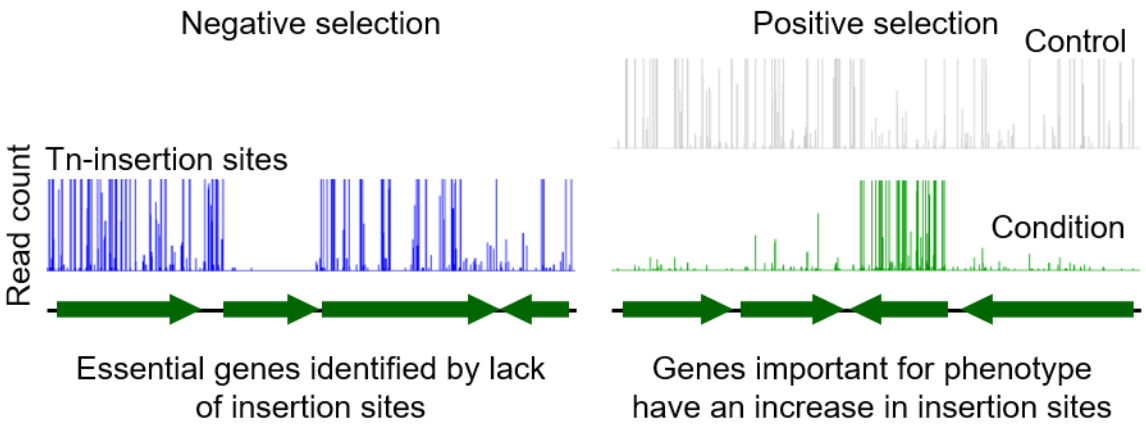


Figure 1.5: Overview of Transposon insertion sequencing (TIS) experiments. A large, random mutant library is generated by transposon (Tn) mutagenesis with a transposon that inserts quasi-randomly in the bacterial genome of interest. The Tn-mutant library is subjected to negative or positive selection. For negative selection, Tn-mutants are grown in a condition of interest. Any Tn-mutants with genes essential for growth in the condition are lost. Examples of negative-selection growth conditions include growth in media with antibiotics, growth *in vivo* and co-growth with a competing bacterial species (Type VI secretion system killing). For positive selection, Tn-mutants with a particular mutant phenotype are recovered, and as such these recovered Tn-mutants will have insertions in genes required for the wild-type phenotype. Examples of positive selection include physical sorting by fluorescent-activated cell sorting selection of Tn-mutants resistant to bacteriophage killing. Following selection, genomic DNA is extracted from all recovered Tn-mutants. TIS libraries are produced to allow for Tn-chromosome junctions to be sequenced via Illumina sequencing. Sequencing data is used to identify Tn-insertion sites in recovered mutant strains. For negative selection, genes with highly reduced unique insertion sites are identified as essential for growth in that condition. For positive selection, genes that have a fold-change increase in read count for the condition compared to the control are identified as important for that phenotype. Together, TIS experiments allow all genomic regions required for a phenotype to be identified. Tn: transposon, gDNA: genomic DNA, P5: Illumina P5 sequence, P7: Illumina P7 sequence.

1.6.1 TIS methodology

TIS methods require the production of a very large transposon mutant library that achieves high coverage across the entire genome. Most Tn-Seq-based studies have utilised *Himar1*, Tn*5* or Tn*5*-derivative transposons to produce a high-density mutant library (183, 189). Tn*5* and Tn*5*-derivatives have a weak recognition sequence, and the *Himar1* recognition site is only a TA dinucleotide, allowing both transposons to insert quasi-randomly into the genome of a wide range of bacterial species (190, 191). Transposition of both Tn*5* and *Himar1* does not require any host factors, allowing transposition in almost all bacterial species (189, 192). Transposons are delivered into the bacteria of interest via transformation or conjugation. Usually, the transposon is delivered on a suicide plasmid with the transposase gene on the plasmid backbone, or as purified transposome (a complex of transposon DNA and transposase). Delivery via a suicide vector or transposome is efficient, and results in a stable transposon insertion.

The prepared mutant library is then used in positive or negative phenotypic selection experiments. Genomic DNA is extracted from all recovered cells and used to produce TIS libraries to allow sequencing of all recovered transposon sites. Several similar TIS library production methods have been developed, each with advantages and disadvantages (183, 193). Genomic DNA is cleaved into smaller fragments by sonication or restriction digestion (186, 187). Methods that fragment DNA by restriction digestion use the type II restriction endonuclease *Mme*I that cuts 20 bp downstream of the target sequence (194). In methods that utilise *Mme*I, the *Himar1* inverted repeat is modified to include an *Mme*I recognition site. Genomic DNA digested with *Mme*I has transposon-genomic junction fragments of identical size, reducing PCR bias during library amplification compared to DNA that is fragmented by sonication where the DNA length adjacent to the transposon varies. Following fragmentation by any method, adapters are then added to the ends of all DNA fragments, most commonly by ligation. Some TIS methods use a terminal deoxynucleotide transferase (TdT) enzyme to add a polynucleotide tail onto all fragments, reducing the chance of adapter dimers or fragments not having adapters (195). However, TdT methods require two rounds of PCR to produce the final library for Illumina sequencing, increasing the chance of PCR bias (195). The adapter-ligated products are used in a PCR with primers specific to the transposon and the adapter or polynucleotide tail. This PCR is specific for fragments containing a transposon-chromosome junction, as only the transposon-specific primer can bind in the first round of the PCR. Some TIS methods utilise splinkerette adapters that form a hairpin on the strand containing the adapter-specific primer binding site (196). The hairpin further precludes adapter primer binding, increasing PCR specificity for fragments containing a transposon-chromosome junction (196). Other TIS methods include additional steps between adapter ligation and library

amplification to improve PCR specificity. The Tn-Circle library production method digests and ligates fragments containing a transposon-chromosome junction into circular products (197). The library is then exonuclease treated, removing linear products, and leaving only fragments containing a transposon-chromosome junction as templates for library amplification (197). The final TIS libraries are sequenced via Illumina sequencing. Reads are aligned to the reference genome, with the first base identifying where the transposon had inserted.

During mutant library construction, transposon insertions will occur randomly throughout the genome and should occur equally in both essential and non-essential genomic regions. During negative selection, any mutant with an insertion in a gene essential for that condition will be non-viable and lost from the population. As such, large regions without transposon insertions are identified as essential for growth in the condition of interest (Figure 1.4, Negative selection). Essential genes are identified by annotation-dependant or annotation-independent methods (183). Annotation-dependant analysis calculates an insertion index by normalising the number of unique insertion sites (UIS) for gene length. Insertion indexes are compared across the genome, and any gene with an insertion index below a predetermined cut-off are identified as essential (183). Annotation-independent methods compare the number of UIS in kmers across the genome in a sliding window approach, or adjacent to different points across the genome using a hidden Markov model (183). Annotation-independent analyses require a dense transposon mutant library to avoid falsely identifying essential regions (183). For positive selection, mutant strains with a specific phenotype are isolated from the mutant library. Mutants can be selected for by particular growth conditions or physical selection (198). Positively selected mutants will have transposon insertions in genomic regions that confer the phenotype of interest. As such, genomic regions that have a high number of UIS and increased fold-change of transposon insertions compared to the control are predicted to be important for the phenotype of interest (Figure 1.4, Positive selection).

Genes identified by TIS methods are only putatively essential or important for the phenotype of interest. Essential genes identified using TraDIS were recently compared with those identified using single gene deletion libraries in *E. coli* K-12 strain BW25113 (199). This comparison found that TIS methods can in some cases incorrectly identify non-essential genes as essential, and vice versa, due to the analysis method used to identify essential genes. Genes with a low insertion index (low unique insertion sites in that gene) may be required for wild-type levels of fitness rather than being essential, or may have a low number of transposon insertions by chance (199). Non-essential genes can be in an operon with an essential gene, with inactivating insertions in the non-essential gene affecting the expression of the essential gene. Essential genes can be called non-essential if only part of the gene or encoded protein is essential. For example, proteins may have an essential N-terminus domain

but mutants with transposon insertions downstream of the essential domain may be recovered, or genes may contain regulatory sequences required for proper expression of adjacent essential genes, with insertions tolerated around this regulatory region (199). In both cases, these genes would have high insertion indexes and be called non-essential, but removal of the entire gene would result in loss of viability. Therefore, it is important that genes identified by TIS methods as essential or important for a phenotype are confirmed in directed mutants.

1.6.2 TIS-based studies

TIS methods have been used to identify essential genes in many different bacterial species across a wide range of growth conditions. Many studies have initially focused on growth of the transposon mutant libraries *in vitro* in rich media, as it is predicted that essential genes identified under this growth condition identifies most genes required for all growth conditions. TIS analyses have identified 356 genes in *Salmonella enterica* Typhi Ty2 derivative strain, 827 genes in *Neisseria gonorrhoeae* strain MS11 and 352 genes in *Pseudomonas aeruginosa* strain MPAO1 as essential for growth in rich media (187, 200, 201). Genes found to be essential during growth in rich media are typically involved in housekeeping processes such as cell division, DNA replication, transcription, translation, protein folding and export, and peptidoglycan, LPS, capsule, lipid and amino acid biosynthesis (187, 200, 201). TIS methods can also identify essential non-coding regions and essential domains within a protein if the transposon mutant library has very high resolution (< 15 bp between each UIS). A *Caulobacter crescentus* mutant library was generated with a UIS every eight bp (202). TraDIS analysis of this mutant library following growth in rich media identified 480 open reading frames, 402 regulatory sequences and 130 non-coding sequences including 29 tRNAs and eight small RNAs as essential (202). Similarly, an *E. coli* K-12 strain BW25113 mutant library that had an insertion every 6 bp identified several regions within genes that encoded essential protein domains or essential regulatory regions (199). These studies show the importance of having a saturated mutant library for TIS studies. Several studies have identified genes essential for growth in the natural host, or in *in vitro* conditions mimicking growth in the host. Growth in the host or conditions mimicking the host allows for identification of putative primary and secondary virulence factors, some of which may become targets for novel therapeutic development. A *H. Influenzae* TIS study identified 136 genes as essential for persistence in the mouse lung, including genes required for LPS, cell membrane, and amino acid biosynthesis and transport, nutrient acquisition and transport as well as several required for general housekeeping (185). Tn-Seq analysis of an *E. coli* mutant library grown in human serum identified 56 genes essential for survival in human serum, the majority of which were involved in LPS biosynthesis (203). Serum sensitivity was assessed in 54 directed mutants

generated in genes identified by TraDIS, confirming 46 of the genes were essential for serum resistance (203).

TIS methods can also be used to identify genes involved in a specific biological process or phenotype. Tn-Seq was used to identify type VI secretion system (T6SS) bactericidal effectors and their cognate immunity proteins (that protect against self-toxicity) produced by *Vibrio cholerae* strain V52 (204). Immunity proteins by *tsiV1*, *tsiV2* and *tsiV3* were identified by comparing genes essential for growth in the wild-type V52 strain and a V52 double *hcp* mutant that had a non-functional T6SS (204). TsiV1, TsiV2 and TsiV3 were antitoxins against effectors VasX, VgrG3 and TseL, respectively (204). In another study, TraDIS analysis of *Clostridium difficile* strain R20291 mutant libraries following growth on either rich media or sporulation media identified 404 genes essential for growth in rich media and 798 genes involved in sporulation (205). Both of the studies above compared essential genes following negative growth conditions to identify genes important for a particular phenotype. Several recent TIS studies have utilised a positive selection process to identify genes required for a particular phenotype. An *E. coli* strain EC958 transposon mutant library was grown on motility LB agar plates, allowing for the isolation of hypermotile transposon mutant strains with TraDIS analysis identifying 30 genes that repressed motility in strain EC958 (206). In another study, an *E. coli* strain 9000 transposon mutant library was grown in the presence of either T4 or T7 bacteriophages (207). TraDIS analysis identified several novel genes that facilitated phage susceptibility, including genes in the *sap* ABC-transporter operon and a gene encoding a putative polynucleotide phosphorylase involved in DNA repair mechanisms (207). In addition to positive and negative selection processes, physical methods can be used to isolate mutants with a phenotype of interest. An *Acinetobacter baumannii* strain BAL062 mutant library was subjected to fluorescence-activated cell sorting following treatment with ethidium bromide (208). Ethidium bromide is exported by multidrug efflux pumps, therefore mutants with insertions in genes required to efflux the drug remained fluorescent compared to strains with intact drug efflux pumps (208). Comparison of fluorescent and non-fluorescent mutant strains identified AdeABC, AdeIJK and AmvA as the main efflux pumps for ethidium bromide, as well as novel regulators of these efflux pumps (208). Collectively, these studies show how TIS methods can be used to probe gene function.

1.7 Aims of this study

The aim of this study was to investigate *P. multocida* virulence using TraDIS and comparative genomics. Firstly, chapter two describes how the TraDIS methods was adapted for use in the highly virulent *P. multocida* fowl cholera isolate VP161, and reports on the initial TraDIS analysis that allowed for the identification of over 500 genes that were essential for VP161 growth in heart infusion broth. Following this, genes important for HA capsule production in

strain VP161 and for normal growth in complement-active chicken serum were investigated. To identify capsule-associated genes, the VP161 transposon mutant library was subjected to buoyancy density gradients and mutants with a lower density, and therefore produced little or no capsule, were isolated and investigated using TraDIS. In total, 69 genes were identified as important for HA capsule production. Mutants in five of the capsule associated genes, not previously known to be involved in capsule production, were confirmed to be required for capsule production by HA capsule absorbance assays and scanning electron microscopy. This analysis identified novel regulators of capsule biosynthesis in *P. multocida*, including genes involved in the stringent response. To identify genes required for growth in chicken serum, the VP161 transposon mutant library was grown in 90% chicken serum, with TraDIS analysis identifying 94 genes that were essential or important for VP161 survival in chicken serum, but were not essential for growth in rich media. Capsule biosynthesis genes within the main capsule locus were identified as important for VP161 survival in serum but not the known regulators of capsule biosynthesis genes.

Chapter three examines the genomes of *Pasteurella* spp. recently isolated from humans, cats and dogs, with comparison to all publicly available *P. multocida* whole-genome reference sequences using a range of bioinformatic analyses. In-depth analysis of *P. multocida*-specific virulence factors showed the recent human, cat and dog *P. multocida* isolates had virulence factor profiles similar to previously sequenced *P. multocida* strains. However, several isolates did not have a capsule locus, had a divergent Flp-pili biosynthesis locus or were missing genes encoding iron-receptors (compared to reference *P. multocida* strain Pm70). Pan-genome analysis of 296 *P. multocida* isolates (36 sequenced in this study and 260 reference genomes) identified a number of trait-associated genes, including genes encoding a putative lipoprotein and glycosyltransferase that were over-represented in type B:L2 strains that were associated with haemorrhagic septicaemia and several genes encoding heavy metal resistance proteins and stress response regulators over-represented in capsule type F strains. Also identified in human, cat and dog isolates were genes predicted to be involved in L-fucose uptake and utilisation, and most animal disease isolates contained genes involved in anaerobic ATP-synthesis and nutrient uptake. These trait-associated genes are likely to contribute to *P. multocida* host predilection and disease presentation.

Chapter Two – Investigation of *P. multocida* VP161 hyaluronic acid capsule production and resistance to chicken complement-mediated killing

The following manuscript and associated supplementary material were submitted verbatim to Mbio on the 8th of August 2021 as:

Thomas R. Smallman, Galain Williams, Marina Harper[#] and John D. Boyce. 2021. Whole genome investigation of essential genes and capsule production genes in *Pasteurella multocida*.

Whole genome investigation of essential genes and capsule production genes in *Pasteurella multocida*

Thomas R. Smallman, Galain Williams, Marina Harper[#] and John D. Boyce[#]

Infection and Immunity Program, Monash Biomedicine Discovery Institute and Department of Microbiology, Monash University, Victoria

Running title: *Pasteurella multocida* genes involved in capsule production

Keywords: Pasteurella multocida, capsule, hyaluronic acid

[#]Corresponding author: John Boyce, john.boyce@monash.edu

ABSTRACT *P. multocida* is the causative agent of a range of animal diseases resulting in severe morbidity and mortality in livestock animals. The *P. multocida* hyaluronic acid (HA) capsule is a critical virulence factor but the mechanism behind *P. multocida* capsule biosynthesis is not fully understood. We utilised transposon-directed insertion site sequencing (TraDIS) to first identify *P. multocida* genes essential for *in vitro* growth in rich media, and secondly to investigate HA capsule production in the highly virulent fowl cholera isolate VP161. By coupling discontinuous Percoll density gradients with TraDIS analysis, 69 genes were identified as important for capsule production, including all genes previously know to be involved in capsule biosynthesis and regulation. Five of the novel capsule biosynthesis genes were confirmed to be required for *P. multocida* capsule production by assessing capsule production in mutants and complemented mutants. Several other capsule-associated genes were associated with regulation or activation of the stringent response, a bacterial stress response. Fis is known positive regulator of capsule biosynthesis genes in *P. multocida*, with *fis* downregulated during the stringent response in other bacterial species. However, qRT-PCR analysis indicated *fis* expression was unchanged when the stringent response was activated, whereas capsule gene expression was reduced. Overall, this study has identified genes required for normal growth of *P. multocida in vitro* and those required for the production of HA capsule, and revealed that the activation of the stringent response leads to a downregulation of capsule biosynthesis genes in *P. multocida*.

IMPORTANCE The bacterium *P. multocida* can cause serious animal disease, including fowl cholera in birds, haemorrhagic septicaemia in cattle and buffalo, atrophic rhinitis in pigs and respiratory diseases in a range of livestock. *P. multocida* disease can be fatal within 24 h and can spread rapidly through wild and domestic animal populations, resulting in a large economic impact on agriculture worldwide. As such, a deep understanding of the pathogenic mechanisms used by *P. multocida* is required. *P. multocida* produces a capsule that is essential for disease, with the complete mechanism and regulation of capsule production yet

to be fully elucidated. We utilised TraDIS to identify all genes essential for *P. multocida* growth in rich media, and genes important for hyaluronic acid capsule production. These analyses have identified novel *P. multocida* essential genes, and novel capsule biosynthesis and regulatory genes in this important animal pathogen.

INTRODUCTION

P. multocida is a Gram-negative coccobacillus the causative organism of serious animal diseases, including avian cholera, haemorrhagic septicaemia, bovine respiratory disease, atrophic rhinitis and snuffles (1). These diseases range from mild, self-limiting infections to acute and peracute lower respiratory tract or systemic infections that can be fatal within 24 h (1). *P. multocida* is highly transmissible, as shown by large outbreaks of fowl cholera and haemorrhagic septicaemia in wild and domesticated animal populations (2, 3). In addition, *P. multocida* can also cause zoonotic disease in humans, commonly causing skin infections following animal bite wounds, and less frequently meningitis, peritonitis, and bacteraemia (4).

Virulent *P. multocida* strains express several crucial virulence factors including polysaccharide capsule (5, 6), lipopolysaccharide (7, 8), filamentous haemagglutinin (9) and several iron and nutrient scavenging systems (10, 11). Capsule is a primary *P. multocida* virulence factor. Acapsular *P. multocida* X73 (serogroup A) and *P. multocida* M1404 (serogroup B) mutants are severely attenuated for virulence in chickens and mice, respectively (5, 6). Five different *P. multocida* capsule serogroups have been identified (denoted A, B, D, E and F), with each group representing different glycosaminoglycan polymers (12). Capsule types A, D and F are comprised of hyaluronic acid (HA), heparosan and chondroitin respectively (13-15). Capsule types B and E are comprised of repeating N-acetylmannosaminuronic acid and N-acetylglucosamine (GlcNAc) disaccharide subunits with a fructose sidechain attached to GlcNAc (16). B and E capsules have different linkages between disaccharide repeats, and type B capsules have a glycine sidechain attached to GlcNAc that is not present in the type E capsule (16). Particular *P. multocida* diseases are associated with a subset of capsule serogroups, indicating capsule type contributes to the ability of different *P. multocida* strains to colonise and cause disease in different hosts (1).

The capsule biosynthesis loci from all of the *P. multocida* capsular serogroups have been identified (17-19). Strains representing each serogroup contain a single locus that can be split into three regions; genes in region one encode an ABC transport system required for capsule polymer export, and genes in region three encode putative phospholipid substitution proteins predicted to be required for anchoring of the exported capsule into the lipid A layer of LPS (17-19). Genes in regions one and three are highly conserved between strains from

different serogroups (17, 18). Genes in region two differ between serogroups as they encode the proteins required for specific sugar monomer synthesis and activation, as well as the glycosyltransferases required for capsule polymer synthesis (20-23). As well as the genes encoding synthases located within the primary capsule biosynthesis locus, *P. multocida* serogroup A, D and F strains also encode a functional heparosan synthase on a gene located elsewhere on the genome (24). This additional heparosan synthase may allow capsular antigen variation in serogroup A and F strains, although heparosan has not been identified as a component in strains belonging to these serogroups. The expression of genes in the primary *P. multocida* capsule biosynthesis locus is positively regulated by both Fis and Hfq (25, 26). In *P. multocida* strain VP161 (serogroup A), independent *fis* and *hfq* insertional mutants displayed reduced capsule biosynthesis gene expression and produced almost no detectable HA capsule compared to the wild-type parent strain (25, 26). The amount of HA capsule produced by *P. multocida* varies at different growth phases, with capsule production highest at mid-exponential stage growth and lowest at low at early- and late-exponential stage growth (25).

In this study, we used transposon-directed insertion-site sequencing (TraDIS) to identify genes in *P. multocida* strain VP161 essential for growth in rich media and in a separate experiment, identified genes required for capsule production. *P. multocida* strain VP161 is a highly virulent serogroup A strain recovered from a case of fowl cholera (27). TraDIS and related methods are whole-genome forward genetic approaches that can be used to identify genomic regions essential for growth in a condition of interest, or important for a phenotype of interest (28, 29). To perform TraDIS with *P. multocida*, a dense *Himar1* mutant library was produced in strain VP161. The VP161 *Himar1* mutant library was grown in rich media, allowing identification of genes required for general growth (Fig. 1). To investigate capsule production, the VP161 *Himar1* mutant library was separated through a discontinuous Percoll density gradient. Percoll gradients have been used in several other bacterial species to separate cells based on differences in capsule production (30-32), including in *Klebsiella pneumoniae* where gradients were coupled with TraDIS analysis to identify genes required for capsule production (33). Acapsular VP161 *Himar1* mutants had reduced buoyancy compared to capsulated cells, allowing fractionation of the mutant library into acapsular and capsulated populations, with TraDIS analysis allowing identification of genes crucial for HA capsule production (Fig 1).

RESULTS

Closure of the *P. multocida* strain VP161 genome using combined Illumina and Nanopore sequencing. TraDIS sequencing data analysis requires an accurate reference genome to ensure that the position of each transposon insertion site is precisely identified. However, the available *P. multocida* strain VP161 genome sequence was incomplete. To close the VP161 genome, Oxford Nanopore MinION™ sequencing of VP161 genomic DNA was performed, generating more than 10^6 reads with an average read length of 1.5 kb. Sequencing reads generated using Nanopore were *de novo* assembled using SPAdes, together with reads from previously performed Illumina sequencing of VP161, into a single 2,246,076 bp circular genome with 40.36% GC content. Annotation of the closed genome with Prokka identified a total of 2,104 open reading frames, with 2,025 encoding proteins, 59 encoding tRNAs, 19 encoding rRNAs and one encoding a tmRNA. The fully annotated VP161 genome was submitted to NCBI (accession no. CP048792.1).

***Himar1* mutant library production.** A modified *Himar1* suicide vector, pAL614, was used for mutagenesis of *P. multocida* strain VP161 (34). Given the low competency of *P. multocida*, conjugative transfer was used to deliver pAL614. To allow for selection of *P. multocida* following conjugation, the Tn*7* transposon (containing a kanamycin resistance cassette) was introduced into VP161 via triparental mating with donor *E. coli* strains AL2017 and AL2621 (Table S1). Tn*7* inserts at a single conserved intergenic region approximately 23 bp downstream of *glmS* and has been shown not to disrupt virulence in a wide range of bacterial species (35). A single Tn*7* insertion at the expected site was confirmed by direct Sanger sequencing of VP161-Tn*7* genomic DNA using the Tn*7*-specific primer BAP8265. The kanamycin resistant *P. multocida* strain VP161-Tn*7* was then used as a recipient in filter mating conjugations with donor *E. coli* strain AL2972 that harboured the *Himar1* delivery vector pAL614. Each conjugation generated more than 10^6 *P. multocida Himar1* transconjugants. A pool containing ~6.45 x 10^6 *P. multocida Himar1* transconjugants was stored in 15% glycerol broth at -80 ºC for subsequent TraDIS experiments.

Genes essential for *P. multocida* VP161-Tn*7* growth in rich media. Two separate replicates of the VP161-Tn*7* *Himar1* mutant library were grown to mid-exponential growth phase in HI broth. Genomic DNA was extracted from surviving mutants and two TraDIS libraries were generated from each replicate (full TraDIS methodology in supplementary information). The rich media TraDIS libraries were sequenced on an Illumina MiSeq generating 7,940,656 reads in total for the four libraries, with reads analysed using Bio-TraDIS scripts. Briefly, reads containing the *Himar1* inverted repeat were aligned to the VP161 genome, with the first sequenced VP161 base of each read assigned as a transposon insertion site. This analysis identified 7,208,425 reads containing a *Himar1* inverted repeat (90.7% of

total reads) and mapped 5,018,138 reads to the VP161 reference genome (63.2% of total reads). The number of unique insertion sites (UIS) and total insertions per insertion site were determined for each gene (see Appendix 1 for full HI TraDIS data). Collectively there were 81,929 unique *Himar1* insertion sites identified in the VP161 genome, giving a unique insertion every 28 bp and an average of 38 unique insertion sites per gene. The number of UIS per gene was compared between the four rich media TraDIS libraries, with high R^2 values showing good reproducibility between libraries (Fig. S1). As such, data from the four rich media TraDIS libraries was combined to identify essential genes for growth in rich media. An insertion index was calculated for each gene by dividing the number of unique insertions by gene length. Insertion indexes for all genes were plotted on a histogram, with the resulting graph showing two normally distributed datasets that represent essential and non-essential genes (Fig. S2). Normal distribution curves were drawn for each dataset and the intersection between the two curves was taken as the cut-off for essentiality. From this, all genes with an insertion index < 0.0134 were identified as essential (Fig. S2 and Fig. S3) A total of 509 genes were identified as putatively essential for *P. multocida* growth in HI broth (Fig. 2; Appendix 2). Of these 509 genes, 473 encoded proteins, 35 encoded tRNAs and one encoded a 5s rRNA.

The predicted essential proteins were characterised by cluster of orthologous groups (COGs) to allow for general function predictions. A large proportion of the proteins predicted to be involved in translational, ribosomal structure and biogenesis were identified as essential (Fig. 3). Specific metabolic pathways for each of the *P. multocida* genes predicted as essential were identified by using KEGG Orthology And Links Annotation (BlastKOALA) tool (36). Several of the essential genes encoded proteins involved in central metabolism pathways, including glycolysis, pyruvate oxidation and gluconeogenesis, biosynthesis pathways of phosphatidylethanolamine, lipid A, peptidoglycan, purine and pyrimidines, lysine, several vitamins and cofactors, and housekeeping processes including ribosome biogenesis, aminoacyl-tRNA synthetases, DNA replication, cell division and the Sec protein export system. The 473 predicted essential protein-encoding genes were compared against the database of essential genes (DEG), that contains a list of genes identified as essential in other bacterial species. Of the 473 proteins, 442 (93%) were homologs of proteins known to be essential in other bacteria, suggesting that the list of essential genes in *P. multocida* is robust. The 31 *P. multocida* proteins with no DEG matches were involved in several different cellular processes, including DNA repair and stress response, gene and protein regulation, nutrient synthesis, and a contact-dependent growth inhibition system. Of the 31 proteins that did not have a homolog in the DEG, 17 were hypothetical with no conserved domains.

Isolation of capsule mutant strains via discontinuous Percoll density gradients. As capsule is an important *P. multocida* virulence factor, we used TraDIS to identify genes

required for HA capsule production. Discontinuous Percoll gradient centrifugation was optimised for separation of acapsular *P. multocida* cells from capsulated cells using a 1:1 mix of capsulated *P. multocida* VP161-Tn*7* and a known acapsular *P. multocida* strain X73 *hexA* mutant (strain PBA930) (6). Gradients with layers of 10%, 40% and 80% Percoll separated the cells into two clear fractions; one fraction was present at the interface between the 10% and 40% layers, and the other fraction was present at the interface between the 40% and 80% layers (Fig. S4A). Both fractions were recovered and plated onto HI supplemented with kanamycin to select for VP161-Tn*7* cells, and onto HI agar supplemented with tetracycline to select for the acapsular mutant cells. The population of cells recovered from the bottom fraction was 99.98% acapsular, whereas the population recovered from the top fraction was 98.7% capsulated, showing good separation of acapsular and capsulated *P. multocida* cells.

To isolate acapsular mutants within the VP161-Tn*7* *Himar1* mutant library, three separate replicates of the VP161-Tn*7* *Himar1* mutant library were fractionated through a 10%, 40% and 80% stepwise Percoll gradient. The VP161-Tn*7* mutant library separated into two fractions at each interface (Fig. S4B) and the bottom fraction containing putative acapsular mutants was recovered and plated onto HI agar plates. Based on colony morphology (small non-mucoid), approximately 43% of the recovered cells displayed an acapsular phenotype. To enrich the acapsular population, cells recovered from the bottom fraction after a single Percoll gradient were cultured for 4 h in HI broth and fractionated through a second Percoll gradient. Cells were recovered from both the top and bottom fractions following centrifugation and plated onto HI agar plates. All colonies from cells recovered from the top fraction appeared capsulated, and approximately 99% of the colonies from cells recovered from the bottom fraction appeared to be acapsular. Genomic DNA was purified from selected individual mutants that displayed an acapsular phenotype and direct Sanger sequencing was performed to identify *Himar1* insertion sites in these strains (see Table S3 in supplementary material). Sequencing analysis revealed that many of the selected mutants had *Himar1* insertions in known capsule biosynthesis genes, namely *hyaC*, *hyaD, hyaE* or *phyA*. Together, these data suggested that Percoll density gradients were successful in separating acapsular mutants from the large VP161-Tn*7* *Himar1* mutant library.

Identification of genes important for capsule production. The genomic DNA extracted from cells recovered from both the top and bottom fractions was used to produce TraDIS libraries. These Percoll TraDIS libraries were sequenced on an Illumina MiSeq and analysed using Bio-TraDIS and modified scripts. For the top fraction TraDIS libraries 3,482,324 reads were generated, 3,271,984 reads contaned a *Himar1* inverted repeat (94.0% of total reads) and 2,573,581 reads mapped to the VP161 reference genome (73.9% of total reads). For the bottom fraction TraDIS libraries 1,998,534 reads were generated, 1,869,079

reads contained a *Himar1* inverted repeat (93.58% of total reads) and 1,437,595 reads mapped to the VP161 reference genome (71.9% of total reads). As above, the number of UIS per gene were compared between library replicates, showing a high correlation between three top and bottom TraDIS library replicates (Fig. S5A). Additionally, principal component analysis was performed using the number of UIS per gene and total *Himar1* insertions per gene for all capsule library replicates (Fig. S5B). The three bottom fraction library replicates clustered together, with the top fractions more distant but not clustering with the bottom fraction (Fig. S5B). The number of transposon insertions was normalised for total read count, with the number of transposon reads for each gene compared between the samples obtained from the top and bottom fractions. As acapsular VP161-Tn*7* *Himar1* mutants were enriched in the bottom fraction, any gene that had > 4.0-fold (> 2.0 $\log_2$) increase in read counts in the bottom fraction, a *q*-value of < 0.001 and an insertion index > 0.8 was identified as important for capsule production. From this analysis, 69 genes were identified as putatively important for capsule production (Table 1, Fig. 4, Fig. S6 and Appendix S3). Importantly, all genes within the capsule biosynthesis locus (*phyAB*, *hyaBCDE*, *hexABCD*) and transcriptional regulators of capsule biosynthesis genes (*fis*, *hfq*) fulfilled these selection criteria.

To validate the TraDIS findings, some of the novel capsule-associated genes (*ppx*, *ptsH* and *spoT*) were selected for directed mutagenesis in wild-type *P. multocida* VP161 using the TargeTron method. As *Himar1* insertions in *spoT* mapped only to the 3' end of the gene in mutants recovered from the bottom fraction (bp 1,416 to 2,124; Fig. S7), the directed TargeTron insertion for this gene was also targeted to the 3' end. Successful insertion of each target gene was confirmed in each TargeTron mutant using PCR amplification of the region and direct Sanger sequencing of putative mutant genomic DNA with TargeTron-specific primer BAP6544. In addition, to these three directed mutants, *Himar1* mutants of *galU* and *pgm*, identified following initial genomic sequencing of putative acapsular *Himar1* mutants were also included for phenotypic validation. Complementation plasmids were constructed for all target genes by cloning an intact wild-type copy of each gene into the *P. multocida* expression vector pAL99S (conferring spectinomycin resistance and suitable for the TargeTron mutants) or pAL99T (conferring tetracycline resistance and suitable for the *Himar1* transposon mutants) and transformed into the corresponding mutant strain. In addition, the appropriate empty vector was provided to each mutant to generate control strains.

The amount of HA capsule produced by each strain was determined by HA capsule absorbance assays and scanning electron microscopy (SEM). All five mutants and corresponding strain harbouring empty vector produced between 3.2 ± 0.8 and 10.9 ± 1.0 µg of HA per mL of culture, significantly less than the wild-type strain VP161 and the *Himar1* library parent strain VP161-Tn*7* that produced 20.5 ± 2.5 and 19.1 ± 4.4 µg of HA per mL of

culture, respectively (Fig. 5). Complementation with an intact wild-type copy of the disrupted gene *in trans* significantly increased capsule production in the mutant strains (Fig. 5). Similarly, SEM showed the wild-type VP161 harbouring empty vector had an average cell width including capsule of 709.4 ± 25.3 nm compared to the *hyaD* and *spoT* mutants harbouring empty vector that had an average cell width of 518 ± 24.3 and 528.4 ± 25.9 nm respectively (Fig. 6B). The *hyaD* and *spoT* mutants complemented with an intact wild-type copy of the disrupted gene *in trans* had an average cell width of 718.9 ± 24.1 and 732.3 ± 41.9 nm, respectively, similar to the cell width of the wild-type parent strain (Fig. 6B). In addition, VP161 had a ruffled outer surface indicative of the presence of capsule that was absent in the *hyaD* and *spoT* mutants but present in the complemented mutants (Fig. 6A). Together these data show that the density gradients coupled with TraDIS successfully identified *P. multocida* acapsular mutants and that further examination of the genes *ppx*, *ptsH*, *spoT*, *galU* and *pgm* showed all five are important for *P. multocida* HA capsule production.

Disruption of *spoT* results in reduced capsule biosynthesis gene expression independent of *fis*. TraDIS analysis of the VP161-Tn*7 Himar1* mutant library following Percoll gradient centrifugation identified *relA* and *spoT* as important for HA capsule production, and that most of the *Himar1* insertions in mutants recovered in the bottom fraction were located within the 3' end of both genes (Fig. S7). The genes *relA* and *spoT* encode a (p)ppGpp synthase and a bifunctional (p)ppGpp synthase/hydrolase, respectively, that together control the levels of the stringent response (SR) alarmone molecules, guanosine 5'-diphosphate-3'-diphosphate and guanosine 5'-triphosphate-3'-diphosphate (collectively referred to as (p)ppGpp) (37). In other bacterial species, RelA and SpoT respond to different stresses by increasing the concentration of (p)ppGpp, activating the SR that results in widespread changes in gene expression and protein activity (38, 39). In *Escherichia coli*, removal or disruption of regulatory domains in the C-terminal region of RelA results in increased (p)ppGpp production, while disruption of the C-terminal region of SpoT is predicted to have the same effect (40, 41). During the SR in *E. coli*, the gene encoding the nucleoid-associated regulatory protein Fis is downregulated (42, 43). To determine if disruption of the C-terminal regulatory region of SpoT alters *fis* and capsule gene expression in *P. multocida*, qRT-PCR was performed to measure *fis* and *hyaD* expression relative to the expression of the control gene, *gyrB*, in wild-type VP161, the *spoT* TargeTron mutant and *spoT* complemented mutant. The *spoT* mutant had significantly reduced *hyaD* expression compared to VP161, with expression partially restored by complementation with an intact copy of *spoT* provided *in trans*. However, there was no significant difference in *fis* expression between these strains (Fig. 7), showing that disrupting the C-terminal regulatory region of SpoT results in reduced capsule gene

expression but does not significantly alter *fis* expression. This *fis*-independent response by SpoT represents a novel mechanism of capsule gene regulation in *P. multocida*.

DISCUSSION

We used TraDIS to identify genes required for the highly virulent *P. multocida* strain VP161 to grow in rich media, and to produce the key virulence factor HA capsule. For TraDIS analysis, a dense transposon mutant library consisting of over 6 x 10^6 mutants was generated by introducing the *Himar1* transposon into *P. multocida* strain VP161-Tn*7*. To identify genes essential for growth in rich media, the VP161-Tn*7* *Himar1* mutant library was grown in HI broth until mid-exponential phase growth phase. Viable mutants were recovered, genomic DNA isolated and used for TraDIS library production. Sequencing of the rich media TraDIS libraries identified 81,929 unique *Himar1* transposon insertion sites, giving a unique insertion every 28 bp, showing good coverage across the entire *P. multocida* genome. A total of 509 genes (473 encoding proteins and 36 encoding tRNA or rRNAs) were identified as putatively essential for *P. multocida* growth in HI broth. The majority of the genes essential for growth in rich media were involved in housekeeping processes, including central metabolism and translation (Fig. 3). Of the 473 protein-encoding genes identified in this analysis, 442 had homologs in the DEG, indicating good conservation of *P. multocida* essential genes with other bacterial species. The proteins that had no DEG matches were predicted to be involved in several different cellular processes, including a number of putative DNA repair and stress response proteins. There were 17 essential genes identified that had no known homologs or putative function. Investigation of these *P. multocida*-specific essential genes would give important insights into *P. multocida* biology.

To identify genes essential for capsule production, acapsular VP161-Tn*7* *Himar1* mutants were separated from capsulated cells by Percoll density gradient centrifugation into acapsular and capsulated cell fractions. TraDIS libraries were generated from cells recovered from the top and bottom fractions, representing capsulated and acapsular VP161-Tn*7* *Himar1* mutants, respectively. Comparison of the normalised *Himar1* insertion sites identified 69 genes as important for *P. multocida* HA capsule production. All *P. multocida* genes previously identified to have a role in HA capsule production or capsule gene regulation were identified, including *hyaBCDE*, *hexABCD*, *phyAB*, *fis* and *hfq* (17, 18, 25, 26). Identification of known *P. multocida* capsule genes and regulators gave strong support that other genes identified by this analysis are also important for capsule production. To confirm that some of the other genes identified by TraDIS were truly important for capsule production, genes not previously associated with HA capsule production in *P. multocida* were disrupted using TargeTron mutagenesis, or mutants with *Himar1* insertions in genes identified by TraDIS were recovered

following Percoll gradients. All mutants and mutants harbouring empty vector produced significantly reduced capsule compared to the corresponding wild-type parent strain, with complementation significantly increasing capsule production in all mutants (Fig. 5). Some complemented strains (*hexA*, *hyaD* and *ptsH*) produced less capsule than the wild-type strain, with incomplete complementation likely due to the *tpiA* promotor in the complementation vectors not giving the same level of expression as the native promotors for these genes during early-exponential growth phase. SEM showed the *hyaD* and *spoT* TargeTron mutants harbouring empty vector had significantly reduced capsule, as measured by cell width, compared to the wild-type VP161 parent strain (Fig. 6). Cell width was restored to wild-type levels when these mutants were provided with an intact copy of the target gene *in trans*. Collectively, these data show that the genes identified by TraDIS are important for HA capsule production in *P. multocida*.

Two capsule-associated genes, *pgm* and *galU*, encode phosphoglucomutase and UTP-glucose-1-phosphate uridylytransferase, respectively. Pgm converts glucose-6-phosphate to glucose-1-phosphate and GalU converts glucose-1-phosphate to UDP-glucose-1-phospate. The UDP-glucose-1-phosphate is then converted to UDP-glucuronic acid (UDP-GlcA) by HyaC (23), a monomer used for the assembly of HA capsule. Both *pgm* and *galU* have been shown to be essential for capsule production in *Streptococcus pneumoniae* serogroups 1 and 3, with both serotypes incorporating UDP-GlcA into their capsule polymers (44, 45). In *P. multocida*, we predict that disruption *pgm* or *galU* would stop UDP-GlcA synthesis and stall HA production. Thus, both *pgm* and *galU* are essential for HA production in *P. multocida*.

The genes *spoT* and *relA* were also identified as essential for HA capsule production in *P. multocida* strain VP161. SpoT and RelA are involved in controlling the SR, which is a stress response known in other bacteria to be induced by nutrient starvation (46). Activation of the SR depends on the concentration of the alarmone molecules (p)ppGpp. These alarmone molecules mediate the SR by binding to protein targets and allosterically regulating their activity. Protein targets of (p)ppGpp are involved in DNA replication, transcription, translation and amino acid biosynthesis (47-53). The (p)(p)ppGpp alarmones can also bind to RNA polymerase (RNAP), either alone or together with DskA, changing promoter affinity of RNAP (54, 55). The precise outcomes of the SR vary between bacteria species, but SR activation generally results in genome wide changes in gene expression, reduced cell growth, activation of amino acid biosynthesis and stress response systems, as well as altered regulation of virulence factors (reviewed extensively in 38, 39, 46). *E. coli* strains unable to synthesise (p)ppGpp, or which have a modified RNAP that abolishes (p)ppGpp binding, do not show regulatory changes seen in cells undergoing the SR (56, 57).

In bacteria, the concentration of (p)ppGpp is controlled by RelA/SpoT homologs (RSH), which either have (p)ppGpp synthase, or bifunctional synthase/hydrolase activity, and small alarmone synthase (SAS) homologs, which have (p)ppGpp synthase activity (58). The presence of RSH and SAS proteins varies between bacterial species, and different RSH and SAS proteins respond to different stimuli (58, 59). RelA and SpoT are the only predicted RSH/SAS homologs in *P. multocida* and share 60% and 67% amino acid identity to $RelA_{Ec}$ and $SpoT_{Ec}$ respectively. $RelA_{Ec}$ catalyses the synthesis of (p)ppGpp from ATP and GDP/GTP following activation by binding to deacylated tRNAs in the aminoacyl site in ribosomes, conditions indicating low amino acid supply (60-62). $SpoT_{Ec}$ hydrolyses (p)ppGpp under normal growth conditions but can transition from hydrolysis to (p)ppGpp synthesis in response to iron, phosphate or fatty acid starvation (63-66). Given that both *relA* and *spoT* were identified as important for capsule in *P. multocida*, it is likely that (p)ppGpp concentration impacts capsule production in this bacterium.

RSH proteins have six conserved domains (41, 58). However, nearly all *Himar1* insertions in *spoT* and *relA* acapsular mutants were located in the 3' end of each gene (Fig. S7). The N-terminal region of RSH proteins contains a (p)ppGpp hydrolase and (p)ppGpp synthase domain (40, 67, 68). Hydrolysis of (p)ppGpp is essential in a $relA^+$ strain, so the lack of *Himar1* insertions in the 5' region of *spoT* is not surprising (37). The data also indicated that disrupting the catalytic activity of $RelA_{Pm}$ is unlikely to result in loss of capsule. The regulatory C-terminal region of RSH proteins contains the domains TGS (for ThrRS GTPase, SpoT), CC/ZFD (for coiled coil/Zinc finger domain) and ACT/RRM (for Aspartokinase, choismate mutase, TyrA/RNA recognition motif) (41, 58). The regulation of $RelA_{Ec}$ (p)ppGpp synthase activity by the C-terminal region is well characterized. The C-terminal region of $RelA_{Ec}$ interacts with uncharged tRNAs and the ribosome, with uncharged tRNAs interacting at the TGS domain (62, 69-71). Removal of the ZFD and RRM domains from the C-terminal end stops RelA binding to rRNAs (71). Removing the entire C-terminal region or just the RRM domain of $RelA_{Ec}$ results in increased (p)ppGpp synthesis compared to cells containing wild-type $RelA_{Ec}$ (40, 41). Self-regulation of SpoT by domains within the C-terminal region in bacterial species that have two RSH proteins is not well characterised. The TGS domain in $SpoT_{Ec}$ can bind both acyl carrier proteins, resulting in increased (p)ppGpp synthesis when fatty acid biosynthesis is low, and Rsd, resulting in increased (p)ppGpp hydrolysis when levels of carbon sources are low (65, 72). No regulatory roles have been shown for the CC or ACT domains in SpoT proteins in bacteria with two RSH proteins. The *Himar1* insertion sites identified *spoT* following growth of the VP161-Tn*7* *Himar1* library in rich media were predominantly in the CC and ACT domains, showing these domains were not required for growth in HI broth. In addition, we could generate a viable TargeTron mutant in the ACT domain. Together, these

data indicate the CC and ACTdomains are not required for $SpoT_{Pm}$ (p)ppGpp hydrolysis. In *P. multocida*, it is likely that disrupting the CC or ACT domains in SpoT, like in RelA, results in increased (p)ppGpp synthesis.

Fis is a known regulator of HA capsule biosynthesis genes in *P. multocida* (26) and was confirmed in this study to be important for capsule synthesis. Fis production is known to be regulated by the SR in *E. coli* via negative regulation of the promotor during SR activation (42, 43, 73-75). As such, disrupting regulatory domains in the C-terminal half of RSH proteins, which likely increases (p)ppGpp production, should result in reduced *fis* expression that in turn reduces capsule biosynthesis gene expression. However, the qRT-PCR analysis of the *P. multocida spoT* mutant with a disrupted ACT domain showed that while there was significantly reduced *hyaD* expression, there was no significant change to *fis* expression (Fig. 7). Thus, disruption of the C-terminal region of *spoT* results in loss of capsule gene expression in a Fis-independent manner.

Several other genes identified as important for capsule production in *P. multocida* were also associated with the SR. The genes *tufA_1* and *tufA_2* both encode elongation factor Tu (EF-Tu), which forms a complex with charged tRNAs and GTP that delivers the charged tRNA to the ribosome (76). The EF-Tu-tRNA-GTP complex competes with RelA-uncharged tRNA for binding to the ribosome A site. Removal of EF-Tu may result in an increased RelA-uncharged tRNA complex interacting with the ribosome, resulting in increased (p)ppGpp synthesis by RelA. The *ptsH* gene encodes HPr, a component of the phosphoenolpyruvate: sugar phosphotransferase (PTS) system. HPr acts in a phosphorylation cascade with EI and sugar specific EIIA proteins, which results in carbohydrate monomer phosphorylation (77). No other PTS system genes were identified as associated with capsule production in the TraDIS analysis, suggesting that import and phosphorylation of carbohydrate monomers by the PTS system is not required for HA production in *P. multocida*. HPr has been shown to have a role in regulation of protein activity by binding to other proteins (72). In *E. coli*, HPr binds to Rsd to stop it interacting with $SpoT_{Ec}$ (72). HPr in *P. multocida* may have a similar role, binding directly to $SpoT_{Pm}$ or possibly another protein that regulates $SpoT_{Pm}$ activity. The *ppx* gene encodes an exopolyphosphatase that cleaves terminal phosphates from long chain polyphosphate (polyP) (78). The alarmones (p)ppGpp bind to and inhibit PPX activity, resulting in increased polyP concentration within the cell (79). PolyP helps relieve amino acid starvation by interacting with and increasing Lon protease activity (80). PPX is structurally similar to the pppGpp phosphatase GppA (81, 82). In *E. coli*, ppGpp has a stronger effect on target proteins than pppGpp (83). In *P. multocida* there appears to be no *gppA* homolog, so PPX may be required for converting pppGpp to ppGpp and therefore correct control of protein activity during the SR.

Several carbohydrate metabolism genes were identified as important for capsule biosynthesis, including *glmS* that encodes a fructose-6-phosphate aminotransferase (84), *pgl* that encodes a phosphogluconolaconase (85), and *rpiA* that encodes a ribose-5-phosphate isomerase. GlmS and Pgl are part of the UDP-N-acetyl-D-glucosamine (UDP-GlcNAc) biosynthesis pathway, with UDP-GlcNAc required as a monomer for peptidoglycan and HA biosynthesis. Genes in the UDP-GlcNAc biosynthesis pathway are essential in several bacterial species and were also identified as essential for *P. multocida* growth in rich media, explaining why no other UDP-N-acetyl-D-glucosamine biosynthesis genes were identified as important for capsule biosynthesis in the TraDIS analysis. RpiA is not involved in pathways required for UDP-GlcA or UDP-GlcNAc synthesis so the role of RpiA in HA capsule biosynthesis is as yet unknown.

Also identified as important for capsule biosynthesis was the gene *efp*, which encodes Elongation Factor P (EF-P) that restores ribosomes stalled during translation of polyproline stretches (86, 87). The genes *empA* and *empB*, which encode an EF-P lysine lysyltransferase that adds R-β-lysine to Lys34 within EF-P and a lysine 2,3-aminomutase that can convert L-lysine to (R)-β-lysine, respectively, were also identified as important for capsule biosynthesis (88, 89). Both *empA* and *empB* are required for correct EF-P function in *E. coli* (88). Mutation of *efp*, *empA* or *empB* in *E. coli* leads to reduced production of several capsule associated genes, some of which are homologs of genes identified as important for *P. multocida* capsule production by TraDIS analysis, including Rnb and several ATP synthase subunit proteins (90). Therefore, EF-P is likely important for capsule production in *P. multocida* as it facilitates the production of proteins required for its biosynthesis.

In conclusion, we have identified 509 genes essential for *P. multocida* growth in rich media and 69 genes important for HA capsule production. The genes essential for growth in rich media were predominantly involved in housekeeping processes as observed with other bacterial species grown in rich media. All known capsule biosynthesis genes and regulators were identified as important for HA capsule production, as well a number of genes not previously known to be associated with capsule production in *P. multocida*. Several genes involved in activation of the SR were identified to be associated with capsule production. Overall, this study has given novel insights into *P. multocida* biology and has comprehensively identified the genes required for HA capsule production in *P. multocida*.

Materials and methods

Bacterial strains, plasmids and culturing conditions. Bacterial strains and plasmids used in this study are listed in Table S1 in the supplementary material. *P. multocida* strains were cultured in HI broth (Oxoid). *E. coli* strains were cultured in Lysogeny Broth (LB). *P. multocida* and *E. coli* cultures were grown at 37ºC, with liquid broth cultures shaken at 200 RPM. Solid media was produced by adding 1.5% (w/v) agar to the media. When required, media was supplemented with appropriate antibiotics at the following concentrations: kanamycin (50 µg/mL), spectinomycin (50 µg/mL) and tetracycline (2.5 µg/mL)

DNA manipulations, Sanger sequencing and oligonucleotides. Genomic DNA was extracted using the HiYield genomic DNA MiniKit (RBC bioscience) and plasmid DNA was extracted using the NucleoSpin plasmid kit (Machery-Nagel), as per the manufacturer's instructions. DNA samples were quantified using a NanoDrop 1000 Spectrophotometer or Qubit fluorometer (Thermo Fisher Scientific). PCRs were performed using *Taq* DNA polymerase (Roche), Phusion high-fidelity DNA polymerase (NEB) or KAPA HiFi HotStart DNA polymerase (Roche) as per manufacturer's instructions. Restriction digestion and ligation reactions were performed using enzymes from Roche or New England Biolabs, as per manufacturer's instructions. Oligonucleotides used in this study were synthesised by Sigma-Aldrich and are listed in Table S2 in the supplementary material. Sanger sequencing reactions were performed using plasmid or genomic DNA as template with BigDye Terminatior version 3.1 (Applied biosystems) as previously described (91). Sequencing data was analysed using Geneious Prime (Biomatters).

Nanopore genome sequencing. *P. multocida* strain VP161 genomic DNA was isolated using the HiYield Genomic DNA MiniKit (RBC bioscience) following manufacturer's instructions ensuring all mixing steps were done with gentle inversion to minimize DNA fragmentation. The genomic DNA was prepared for Nanopore MinION sequencing using a Nanopore Rapid Sequencing Kit (SQK-RAD004) as per manufacturer's instructions. Reads from Nanopore sequencing, and reads previously generated using Illumina sequencing, were assembled *de novo* using SPAdes genome assembler (https://github.com/ablab/spades) and annotated using Prokka (https://github.com/tseemann/prokka).

Transformation and conjugation. Transformation of *P. multocida* was performed as previously described (6), with the electroporation conditions modified to 2 kV, 200 Ω and 25 µF. Conjugation of *P. multocida* was performed by filter mating with donor *E. coli* as follows; donor and recipient cultures were each grown to mid-exponential stage (OD_{600} ~0.5), washed in 1 x PBS, then 1 mL of each donor and recipient were mixed and filter syringed onto a 0.23 nM nylon filter. The filter was then placed (cell-side up) onto a HI plate without antibiotics and

incubated at 37ºC for 1 h. Following incubation, cells were recovered by vortex mixing the nylon filter in 1 mL of HI, then resuspended cells were plated onto HI agar plates containing appropriate antibiotics for selection of donor, recipient or transconjugants.

Site-directed mutagenesis. A kanamycin resistant *P. multocida* VP161 derivative strain was produced using Tn*7* mutagenesis. Tn*7* mutagenesis was performed by filter mating conjugation between *P. multocida* strain AL2232 (harbouring pAL99T, conferring tetracycline resistance for counter selection) and *E. coli* strains AL2017 and AL2621 harbouring the Tn*7* vector pTNS3 and the Tn*7* helper vector pAL1083, respectively (see Table S1 in supplementary material). *P. multocida* Tn*7*-transposants were selected on HI agar supplemented with tetracycline and kanamycin, before being cured of the pAL99T plasmid by growth in non-selective media. Directed inactivation of selected genes was performed using TargeTron mutagenesis as previously described (26, 91), with the *P. multocida* TargeTron vector pAL953 retargeted to specific genes using the intron site finder Clostron (92).

Construction of complementation vectors. Selected *P. multocida* mutants were complemented by providing a wild-type copy of the disrupted gene on one of the *P. multocida* expression plasmids pAL99S or pAL99T (see Table S1 in supplementary material). To produce each of the complementation vectors, the target gene was PCR-amplified using Phusion polymerase with *P. multocida* VP161 genomic DNA as template and oligonucleotides that flanked the target gene (see Table S2 in supplementary material). Flanking oligonucleotides contained 5' restriction enzyme sites. PCR products were digested with the appropriate restriction enzymes and ligated into similarly digested pAL99S or pAL99T. Sanger sequencing was used to confirm each complementation vector contained a wild-type copy of the target gene. Complementation vectors were used to transform the mutant strains by electroporation. As a control, the empty pAL99S or pAL99T vector was used to separately transform each mutant to produce mutant harbouring empty vector control strains.

Generation of the *P. multocida Himar1* transposon mutant library. *P. multocida* transposon mutant libraries were produced by delivery of the *Himar1* transposon-containing suicide vector pAL614 into *P. multocida* by conjugation. Conjugation was performed between donor *E. coli* strain AL2972 and *P. multocida* strain VP161-Tn*7* (see Table S1 in supplementary material). Following filter mating, the conjugation mixture was plated onto HI agar plates supplemented with kanamycin (resistance conferred by Tn*7*) and/or spectinomycin (resistance conferred by *Himar1*), and viable counts were performed to determine CFU/mL of donor, recipient and transconjugants. Following overnight growth, *Himar1* transconjugants were scraped from HI agar plates supplemented with both kanamycin and spectinomycin, pooled in 15% glycerol broth and then stored at -80ºC in 1 mL aliquots containing ~ 6.45×10^6 VP161-Tn*7* *Himar1* mutants.

Growth of the *P. multocida Himar1* mutant library in rich media. For identification of genes essential for growth in rich media, two aliquots of the *P. multocida* VP161-Tn*7 Himar1* mutant library were grown separately in 50 mL of HI broths supplemented with spectinomycin. Following overnight growth at 37ºC with shaking at 200 RPM, a 1 mL aliquot of each overnight culture was subcultured into 50 mL of fresh HI broth supplemented with spectinomycin and grown to mid-exponential stage growth (OD_{600} 0.5-0.6) with the same culturing conditions. To recover surviving mutant strains, 2 mL aliquots were taken from each mid-exponential stage culture and centrifuged at 13,000 x *g* for 5 min. Genomic DNA was then isolated from the pelleted cells for TraDIS library production.

Discontinuous Percoll gradient centrifugation. Separation of capsular and acapsular mutant strains in the VP161-Tn*7 Himar1* mutant library was performed using Percoll density gradient centrifugation as previously described (33, 93). Percoll gradients consisted of a 10% (w/v in 1x PBS) top layer, a 40% middle layer and 80% bottom layer in Polyallomer centrifuge tubes (17 mL, 16 mm x 102 mm, Beckman). Individual strains or the *P. multocida* VP161-Tn*7 Himar1* mutant library were grown to mid-exponential growth stage in 20 mL of HI broth, centrifuged and resuspended in 1 mL of 1 x PBS (20-fold concentration) before being gently pipetted onto the top of the gradient. Cells in the top and bottom fractions were recovered by syringe aspiration. For enrichment of acapsular mutants in the bottom fraction, cells recovered from the bottom fraction were grown to mid-exponential growth stage and separated through a second discontinuous Percoll gradient centrifugation.

TraDIS library production, sequencing and analysis. TraDIS libraries for Illumina sequencing were produced as previously described (94) with minor modifications. Full methodology is provided in supplementary material. Briefly 6 µg of DNA extracted from recovered cells was sheared to ~300 bp using Covaris Adaptive Focused Acoustics (AFA). DNA fragments were then end repaired, A-tailed and Splinkerette adapters were ligated to all DNA fragments using NEBNext® End repair module, NEBNext® dA-tailing module and NEBNext® Quick Ligation Module (NEB) as per manufacturer's instructions, with 100 µM Splinkerette adapter used for adapter ligation. TraDIS libraries were produced by PCR using KAPA HiFi HotStart ReadyMix (Roche) with either 100 ng or 250 ng of adapter-ligated DNA fragments as template, and *Himar1*-specific and adapter-specific oligonucleotides. Between each of the library production steps, DNA was purified using AxyPrep™ MAG PCR Clean-up Kit (Axygen) using manufacturer's instructions, with different sample to bead ratios specific for each purification performed. DNA concentration was determined using Qubit and fragment size analysis was performed using Fragment Analyzer™ after each step of the library production. The concentration of clusterable DNA in the final TraDIS libraries was determined using KAPA Library Quantification Kit (Roche) as per the manufacturer's instructions. TraDIS

libraries were submitted to the Micromon Genomics facility at Monash University and sequenced on an Illumina MiSeq v2 using 150 bp-single-end sequencing with the custom *Himar1* sequencing oligonucleotide BAP8042 (see Table S2 in supplementary material).

Poor quality reads and Illumina adapter sequences were removed from the sequencing data using Trimmomatic v0.38. The sequencing reads were then analysed using Bio-TraDIS v1.4.1 and modified scripts as previously described (33, 94). Scripts used for analysis in this study are available at https://github.com/sanger-pathogens/Bio-Tradis and https://github.com/francesca-short/tradis_scripts. All reads beginning with a correct *Himar1* transposon tag (5'-CAACCTGT-3') were aligned to the closed *P. multocida* VP161 genome sequence, with the first base of each alignment deemed to be a transposon insertion site. Unique insertion sites and total read counts were determined for each gene. An insertion index was calculated for each gene by dividing the number of unique insertions by the gene length. All insertion indexes were plotted as a histogram, generating two normally distributed datasets representing essential and non-essential genes. Normal curves were plotted for each dataset and the intersection between the two curves taken as the cut-off for essentiality, with all genes that have a lower insertion index than the cut-off designated as essential. To compare between two TraDIS libraries, read counts were normalized and the fold-change in read counts for each gene determined. Genes were called as important for capsule production if they had a fold-change > 4.0 (log_2 > 2.0) of *Himar1* insertion sites identified in the TraDIS libraries generated from the bottom fraction compared to the top fraction, with a *q*-value < 0.001 and an insertion index ratio > 0.8. The COGs group numbers and metabolic pathways of *P. multocida* rich media essential genes were identified using the annotation tools eggNOG mapper and BlastKOALA with KEGG mapper, respectively (95, 96). Known essential homologs of essential genes identified for *P. multocida* growth in rich media were were identified using command line BLASTp with the DEG database (97). To determine reproducibility of the TraDIS library replicates, principal component analysis was performed using the prcomp function using R, and plots generated using the ggbiplot package.

Measurement of hyaluronic acid capsule production. Capsular polysaccharide was extracted from *P. multocida* cells grown in broth and measured as previously described (6) with minor modifications. Cells were harvested from 1 mL of mid-exponential culture and cells were pelleted by centrifugation at 13,000 x *g* for 10 min. Mann-Whitney U tests were performed to determine if differences in the amount of hyaluronic acid produced by different strains were significant.

Scanning electron microscopy. SEM was performed on *P. multocida* strains grown to mid-exponential growth phase (OD_{600} 0.5-0.6). Cells were filtered onto a 0.4 µm aperture polycarbonate membrane and prepared for imaging using Lysine acetate-Ruthenium Red-

Osmium fixation, critical point drying, and sputter coating as previously described with minor modifications (98). For the fixation solution, 100 µL of 16% formaldehyde was used in place of 80 µL of 25% formaldehyde solution. Cells were imaged using a field emission scanning electron microscope Nova NanoSEM 450 with an accelerating voltage of 5 kV in secondary electron mode with a working distance of ~4.5 mm. Images were taken under immersion mode through the lens detector. Cell widths were measured from electron micrographs generated by SEM using the FIJI package for ImageJ. Significant differences in cell widths were determined using an unpaired t-test.

RNA extraction and qRT-PCR. RNA was isolated from *P. multocida* cultures grown to mid-exponential phase growth (OD_{600} ~0.4) in biological quadruplicate using TRIzol™ reagent (ThermoFisher Scientific) as per the manufacturer's instructions. RNA extractions were DNAse treated using the RNeasy kit (Qiagen) as per the manufacturer's instructions, then RNA was purified using phenol:choloroform extraction with 5 PRIME Phase Lock Gel tubes (Quanta Biosciences) following the manufacturer's instructions followed by sodium acetate-ethanol precipitation. Copy DNA was synthesised for quantitative reverse-transcriptase PCR (qRT-PCR) using the AffinityScript cDNA synthesis kit (Agilent Technologies) as per manufacturer's instructions. Each sample was used for reactions with reverse transcriptase (+RT) or without reverse transcriptase (-RT). The qRT-PCRs were performed using Brilliant II SYBR Green qPCR Master Mix (Agilent Technologies) as per manufacturer's instructions and primers specific to *P. multocida* strain VP161 *gyrB*, *hyaD* or *fis* (see Table S2 in supplementary material), with reactions performed in an Aria Mx Real-time PCR system (Agilent). Each sample was measured in technical duplicate. The -RT reactions were analysed to ensure no products were amplified within 35 cycles and the melt curves from +RT reactions were analysed to ensure only a single product was amplified in each reaction.

References

1. Wilkie IW, Harper M, Boyce JD, Adler B. 2012. *Pasteurella multocida*: diseases and pathogenesis. Curr Top Microbiol Immunol 361:1-22.
2. Fereidouni S, Freimanis GL, Orynbayev M, Ribeca P, Flannery J, King DP, Zuther S, Beer M, Höper D, Kydyrmanov A, Karamendin K, Kock R. 2019. Mass die-off of saiga antelopes, Kazakhstan, 2015. Emerg Infect Dis 25:1169-1176.
3. Woo YK, Kim JH. 2006. Fowl cholera outbreak in domestic poultry and epidemiological properties of *Pasteurella multocida* isolate. J Microbiol 44:344-53.
4. Giordano A, Dincman T, Clyburn BE, Steed LL, Rockey DC. 2015. Clinical features and outcomes of *Pasteurella multocida* infection. Medicine (Baltimore) 94:e1285.
5. Boyce JD, Adler B. 2000. The capsule is a virulence determinant in the pathogenesis of *Pasteurella multocida* M1404 (B:2). Infect Immun 68:3463-8.
6. Chung JY, Wilkie I, Boyce JD, Townsend KM, Frost AJ, Ghoddusi M, Adler B. 2001. Role of capsule in the pathogenesis of fowl cholera caused by *Pasteurella multocida* serogroup A. Infect Immun 69:2487-92.
7. Boyce JD, Harper M, St Michael F, John M, Aubry A, Parnas H, Logan SM, Wilkie IW, Ford M, Cox AD, Adler B. 2009. Identification of novel glycosyltransferases required for assembly of the *Pasteurella multocida* A:1 lipopolysaccharide and their involvement in virulence. Infect Immun 77:1532-42.
8. Harper M, Cox AD, St Michael F, Wilkie IW, Boyce JD, Adler B. 2004. A heptosyltransferase mutant of *Pasteurella multocida* produces a truncated lipopolysaccharide structure and is attenuated in virulence. Infect Immun 72:3436-43.
9. Tatum FM, Yersin AG, Briggs RE. 2005. Construction and virulence of a *Pasteurella multocida fha*B2 mutant in turkeys. Microb Pathog 39:9-17.
10. Boyce JD, Wilkie I, Harper M, Paustian ML, Kapur V, Adler B. 2002. Genomic scale analysis of *Pasteurella multocida* gene expression during growth within the natural chicken host. Infect Immun 70:6871-9.
11. Paustian ML, May BJ, Cao D, Boley D, Kapur V. 2002. Transcriptional response of *Pasteurella multocida* to defined iron sources. J Bacteriol 184:6714-20.
12. Rimler RB, Rhoades KR. 1987. Serogroup F, a new capsule serogroup of *Pasteurella multocida*. J Clin Microbiol 25:615-8.
13. Pandit KK, Smith JE. 1993. Capsular hyaluronic acid in *Pasteurella multocida* type A and its counterpart in type D. Res Vet Sci 54:20-4.
14. Rimler RB. 1994. Presumptive identification of *Pasteurella multocida* serogroups A, D and F by capsule depolymerisation with mucopolysaccharidases. Vet Rec 134:191-2.
15. DeAngelis PL, Gunay NS, Toida T, Mao WJ, Linhardt RJ. 2002. Identification of the capsular polysaccharides of Type D and F *Pasteurella multocida* as unmodified heparin and chondroitin, respectively. Carbohydr Res 337:1547-52.
16. Michael FS, Cairns CM, Fleming P, Vinogradov EV, Boyce JD, Harper M, Cox AD. 2020. The capsular polysaccharides of *Pasteurella multocida* serotypes B and E: Structural, genetic and serological comparisons. Glycobiology doi:10.1093/glycob/cwaa069.
17. Chung JY, Zhang Y, Adler B. 1998. The capsule biosynthetic locus of *Pasteurella multocida* A:1. FEMS Microbiol Lett 166:289-96.
18. Boyce JD, Chung JY, Adler B. 2000. Genetic organisation of the capsule biosynthetic locus of *Pasteurella multocida* M1404 (B:2). Vet Microbiol 72:121-34.
19. Townsend KM, Boyce JD, Chung JY, Frost AJ, Adler B. 2001. Genetic organization of *Pasteurella multocida* cap loci and development of a multiplex capsular PCR typing system. J Clin Microbiol 39:924-9.
20. Jing W, DeAngelis PL. 2000. Dissection of the two transferase activities of the *Pasteurella multocida* hyaluronan synthase: two active sites exist in one polypeptide. Glycobiology 10:883-9.
21. DeAngelis PL, White CL. 2002. Identification and molecular cloning of a heparosan synthase from *Pasteurella multocida* type D. J Biol Chem 277:7209-13.

22. DeAngelis PL, Padgett-McCue AJ. 2000. Identification and molecular cloning of a chondroitin synthase from *Pasteurella multocida* type F. J Biol Chem 275:24124-9.
23. Chu X, Han J, Guo D, Fu Z, Liu W, Tao Y. 2016. Characterization of UDP-glucose dehydrogenase from *Pasteurella multocida* CVCC 408 and its application in hyaluronic acid biosynthesis. Enzyme Microb Technol 85:64-70.
24. Deangelis PL, White CL. 2004. Identification of a distinct, cryptic heparosan synthase from *Pasteurella multocida* types A, D, and F. J Bacteriol 186:8529-32.
25. Megroz M, Kleifeld O, Wright A, Powell D, Harrison P, Adler B, Harper M, Boyce JD. 2016. The RNA-binding chaperone Hfq is an important global regulator of gene expression in *Pasteurella multocida* and plays a crucial role in production of a number of virulence factors including hyaluronic acid capsule. Infect Immun 84:1361-70.
26. Steen JA, Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD. 2010. Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. PLoS Pathog 6:e1000750.
27. Wilkie IW, Grimes SE, O'Boyle D, Frost AJ. 2000. The virulence and protective efficacy for chickens of *Pasteurella multocida* administered by different routes. Vet Microbiol 72:57-68.
28. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. Genome Res 19:2308-16.
29. van Opijnen T, Camilli A. 2013. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol 11:435-42.
30. Brunner J, Scheres N, El Idrissi NB, Deng DM, Laine ML, van Winkelhoff AJ, Crielaard W. 2010. The capsule of *Porphyromonas gingivalis* reduces the immune response of human gingival fibroblasts. BMC Microbiol 10:5.
31. DeAngelis PL, Weigel PH. 1994. Rapid detection of hyaluronic acid capsules on group A streptococci by buoyant density centrifugation. Diagn Microbiol Infect Dis 20:77-80.
32. Patrick S, Reid JH. 1983. Separation of capsulate and non-capsulate *Bacteroides fragilis* on a discontinuous density gradient. J Med Microbiol 16:239-41.
33. Dorman MJ, Feltwell T, Goulding DA, Parkhill J, Short FL. 2018. The capsule regulatory network of *Klebsiella pneumoniae* defined by density-TraDISort. MBio 9.
34. Murray GL, Morel V, Cerqueira GM, Croda J, Srikram A, Henry R, Ko AI, Dellagostin OA, Bulach DM, Sermswan RW, Adler B, Picardeau M. 2009. Genome-wide transposon mutagenesis in pathogenic *Leptospira* species. Infect Immun 77:810-6.
35. Choi KH, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR, Schweizer HP. 2005. A Tn7-based broad-range bacterial cloning and expression system. Nat Methods 2:443-8.
36. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726-731.
37. Xiao H, Kalman M, Ikehara K, Zemel S, Glaser G, Cashel M. 1991. Residual guanosine 3',5'-bispyrophosphate synthetic activity of *relA* null mutants can be eliminated by *spoT* null mutations. J Biol Chem 266:5980-90.
38. Boutte CC, Crosson S. 2013. Bacterial lifestyle shapes stringent response activation. Trends Microbiol 21:174-80.
39. Dalebroux ZD, Swanson MS. 2012. ppGpp: magic beyond RNA polymerase. Nat Rev Microbiol 10:203-12.
40. Gropp M, Strausz Y, Gross M, Glaser G. 2001. Regulation of *Escherichia coli* RelA requires oligomerization of the C-terminal domain. J Bacteriol 183:570-9.
41. Turnbull KJ, Dzhygyr I, Lindemose S, Hauryliuk V, Roghanian M. 2019. Intramolecular interactions dominate the autoregulation of *Escherichia coli* stringent factor RelA. Front Microbiol 10:1966.

42. Mallik P, Paul BJ, Rutherford ST, Gourse RL, Osuna R. 2006. DksA is required for growth phase-dependent regulation, growth rate-dependent control, and stringent control of fis expression in *Escherichia coli*. J Bacteriol 188:5775-82.
43. Durfee T, Hansen AM, Zhi H, Blattner FR, Jin DJ. 2008. Transcription profiling of the stringent response in *Escherichia coli*. J Bacteriol 190:1084-96.
44. Hardy GG, Caimano MJ, Yother J. 2000. Capsule biosynthesis and basic metabolism in *Streptococcus pneumoniae* are linked through the cellular phosphoglucomutase. J Bacteriol 182:1854-63.
45. Bonofiglio L, García E, Mollerach M. 2005. Biochemical characterization of the pneumococcal glucose 1-phosphate uridylyltransferase (GalU) essential for capsule biosynthesis. Curr Microbiol 51:217-21.
46. Irving SE, Choudhury NR, Corrigan RM. 2021. The stringent response and physiological roles of (pp)pGpp in bacteria. Nat Rev Microbiol 19:256-271.
47. Corrigan RM, Bellows LE, Wood A, Gründling A. 2016. ppGpp negatively impacts ribosome assembly affecting growth and antimicrobial tolerance in Gram-positive bacteria. Proc Natl Acad Sci U S A 113:E1710-9.
48. Wang B, Dai P, Ding D, Del Rosario A, Grant RA, Pentelute BL, Laub MT. 2019. Affinity-based capture and identification of protein effectors of the growth regulator ppGpp. Nat Chem Biol 15:141-150.
49. Zhang Y, Zborníková E, Rejman D, Gerdes K. 2018. Novel (p)ppGpp Binding and Metabolizing Proteins of *Escherichia coli*. mBio 9.
50. Maciag M, Kochanowska M, Lyzeń R, Wegrzyn G, Szalewska-Pałasz A. 2010. ppGpp inhibits the activity of *Escherichia coli* DnaG primase. Plasmid 63:61-7.
51. Wang JD, Sanders GM, Grossman AD. 2007. Nutritional control of elongation of DNA replication by (p)ppGpp. Cell 128:865-75.
52. Diez S, Ryu J, Caban K, Gonzalez RL, Jr., Dworkin J. 2020. The alarmones (p)ppGpp directly regulate translation initiation during entry into quiescence. Proc Natl Acad Sci U S A 117:15565-15572.
53. Milon P, Tischenko E, Tomsic J, Caserta E, Folkers G, La Teana A, Rodnina MV, Pon CL, Boelens R, Gualerzi CO. 2006. The nucleotide-binding site of bacterial translation initiation factor 2 (IF2) as a metabolic sensor. Proc Natl Acad Sci U S A 103:13962-7.
54. Perederina A, Svetlov V, Vassylyeva MN, Tahirov TH, Yokoyama S, Artsimovitch I, Vassylyev DG. 2004. Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. Cell 118:297-309.
55. Ross W, Vrentas CE, Sanchez-Vazquez P, Gaal T, Gourse RL. 2013. The magic spot: a ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. Mol Cell 50:420-9.
56. Traxler MF, Summers SM, Nguyen HT, Zacharia VM, Hightower GA, Smith JT, Conway T. 2008. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. Mol Microbiol 68:1128-48.
57. Sanchez-Vazquez P, Dewey CN, Kitten N, Ross W, Gourse RL. 2019. Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. Proc Natl Acad Sci U S A 116:8310-8319.
58. Atkinson GC, Tenson T, Hauryliuk V. 2011. The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. PLoS One 6:e23479.
59. Battesti A, Bouveret E. 2009. Bacteria possessing two RelA/SpoT-like proteins have evolved a specific stringent response involving the acyl carrier protein-SpoT interaction. J Bacteriol 191:616-24.
60. Payoe R, Fahlman RP. 2011. Dependence of RelA-mediated (p)ppGpp formation on tRNA identity. Biochemistry 50:3075-83.
61. Pedersen FS, Kjeldgaard NO. 1977. Analysis of the *relA* gene product of *Escherichia coli*. Eur J Biochem 76:91-7.

62. Arenz S, Abdelshahid M, Sohmen D, Payoe R, Starosta AL, Berninghausen O, Hauryliuk V, Beckmann R, Wilson DN. 2016. The stringent factor RelA adopts an open conformation on the ribosome to stimulate ppGpp synthesis. Nucleic Acids Res 44:6471-81.
63. Heinemeyer EA, Richter D. 1978. Characterization of the guanosine 5'-triphosphate 3'-diphosphate and guanosine 5'-diphosphate 3'-diphosphate degradation reaction catalyzed by a specific pyrophosphorylase from *Escherichia coli*. Biochemistry 17:5368-72.
64. Vinella D, Albrecht C, Cashel M, D'Ari R. 2005. Iron limitation induces SpoT-dependent accumulation of ppGpp in *Escherichia coli*. Mol Microbiol 56:958-70.
65. Battesti A, Bouveret E. 2006. Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism. Mol Microbiol 62:1048-63.
66. Bougdour A, Gottesman S. 2007. ppGpp regulation of RpoS degradation via anti-adaptor protein IraP. Proc Natl Acad Sci U S A 104:12896-901.
67. Gentry DR, Cashel M. 1996. Mutational analysis of the *Escherichia coli* spoT gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. Mol Microbiol 19:1373-84.
68. Montero M, Rahimpour M, Viale AM, Almagro G, Eydallin G, Sevilla Á, Cánovas M, Bernal C, Lozano AB, Muñoz FJ, Baroja-Fernández E, Bahaji A, Mori H, Codoñer FM, Pozueta-Romero J. 2014. Systematic production of inactivating and non-inactivating suppressor mutations at the *relA* locus that compensate the detrimental effects of complete spot loss and affect glycogen content in *Escherichia coli*. PLoS One 9:e106938.
69. Brown A, Fernández IS, Gordiyenko Y, Ramakrishnan V. 2016. Ribosome-dependent activation of stringent control. Nature 534:277-280.
70. Loveland AB, Bah E, Madireddy R, Zhang Y, Brilot AF, Grigorieff N, Korostelev AA. 2016. Ribosome•RelA structures reveal the mechanism of stringent response activation. Elife 5.
71. Winther KS, Roghanian M, Gerdes K. 2018. Activation of the stringent response by loading of RelA-tRNA complexes at the ribosomal A-site. Mol Cell 70:95-105.e4.
72. Lee JW, Park YH, Seok YJ. 2018. Rsd balances (p)ppGpp level by stimulating the hydrolase activity of SpoT during carbon source downshift in *Escherichia coli*. Proc Natl Acad Sci U S A 115:E6845-e6854.
73. Ninnemann O, Koch C, Kahmann R. 1992. The *E.coli fis* promoter is subject to stringent control and autoregulation. EMBO J 11:1075-83.
74. Walker KA, Atkins CL, Osuna R. 1999. Functional determinants of the *Escherichia coli* fis promoter: roles of -35, -10, and transcription initiation regions in the response to stringent control and growth phase-dependent regulation. J Bacteriol 181:1269-80.
75. Mallik P, Pratt TS, Beach MB, Bradley MD, Undamatla J, Osuna R. 2004. Growth phase-dependent regulation and stringent control of fis are conserved processes in enteric bacteria and involve a single promoter (*fis* P) in *Escherichia coli*. J Bacteriol 186:122-35.
76. Villa E, Sengupta J, Trabuco LG, LeBarron J, Baxter WT, Shaikh TR, Grassucci RA, Nissen P, Ehrenberg M, Schulten K, Frank J. 2009. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. Proc Natl Acad Sci U S A 106:1063-8.
77. Wang G, Louis JM, Sondej M, Seok YJ, Peterkofsky A, Clore GM. 2000. Solution structure of the phosphoryl transfer complex between the signal transducing proteins HPr and IIA(glucose) of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system. EMBO J 19:5635-49.
78. Akiyama M, Crooke E, Kornberg A. 1993. An exopolyphosphatase of *Escherichia coli*. The enzyme and its *ppx* gene in a polyphosphate operon. J Biol Chem 268:633-9.

79. Kuroda A, Murphy H, Cashel M, Kornberg A. 1997. Guanosine tetra- and pentaphosphate promote accumulation of inorganic polyphosphate in *Escherichia coli*. J Biol Chem 272:21240-3.
80. Kuroda A, Nomura K, Ohtomo R, Kato J, Ikeda T, Takiguchi N, Ohtake H, Kornberg A. 2001. Role of inorganic polyphosphate in promoting ribosomal protein degradation by the Lon protease in *E. coli*. Science 293:705-8.
81. Kristensen O, Ross B, Gajhede M. 2008. Structure of the PPX/GPPA phosphatase from *Aquifex aeolicus* in complex with the alarmone ppGpp. J Mol Biol 375:1469-76.
82. Rangarajan ES, Nadeau G, Li Y, Wagner J, Hung MN, Schrag JD, Cygler M, Matte A. 2006. The structure of the exopolyphosphatase (PPX) from *Escherichia coli* O157:H7 suggests a binding mode for long polyphosphate chains. J Mol Biol 359:1249-60.
83. Mechold U, Potrykus K, Murphy H, Murakami KS, Cashel M. 2013. Differential regulation by ppGpp versus pppGpp in *Escherichia coli*. Nucleic Acids Res 41:6175-89.
84. Wu HC, Wu TC. 1971. Isolation and characterization of a glucosamine-requiring mutant of *Escherichia coli* K-12 defective in glucosamine-6-phosphate synthetase. J Bacteriol 105:455-66.
85. Zimenkov D, Gulevich A, Skorokhodova A, Biriukova I, Kozlov Y, Mashko S. 2005. *Escherichia coli* ORF *ybhE* is pgl gene encoding 6-phosphogluconolactonase (EC 3.1.1.31) that has no homology with known 6PGLs from other organisms. FEMS Microbiol Lett 244:275-80.
86. Doerfel LK, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina MV. 2013. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science 339:85-8.
87. Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. 2013. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science 339:82-5.
88. Park JH, Johansson HE, Aoki H, Huang BX, Kim HY, Ganoza MC, Park MH. 2012. Post-translational modification by β-lysylation is required for activity of *Escherichia coli* elongation factor P (EF-P). J Biol Chem 287:2579-90.
89. Roy H, Zou SB, Bullwinkle TJ, Wolfe BS, Gilreath MS, Forsyth CJ, Navarre WW, Ibba M. 2011. The tRNA synthetase paralog PoxA modifies elongation factor-P with (R)-β-lysine. Nat Chem Biol 7:667-9.
90. Peil L, Starosta AL, Lassak J, Atkinson GC, Virumäe K, Spitzer M, Tenson T, Jung K, Remme J, Wilson DN. 2013. Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. Proc Natl Acad Sci U S A 110:15265-70.
91. Harper M, St Michael F, John M, Vinogradov E, Steen JA, van Dorsten L, Steen JA, Turni C, Blackall PJ, Adler B, Cox AD, Boyce JD. 2013. *Pasteurella multocida* Heddleston serovar 3 and 4 strains share a common lipopolysaccharide biosynthesis locus but display both inter- and intrastrain lipopolysaccharide heterogeneity. J Bacteriol 195:4854-64.
92. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, Minton NP. 2010. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. J Microbiol Methods 80:49-55.
93. Feltwell T, Dorman MJ, Goulding DA, Parkhill J, Short FL. 2019. Separating bacteria by capsule amount using a discontinuous density gradient. JoVE (Journal of Visualized Experiments):e58679.
94. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA, Parkhill J. 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. Bioinformatics 32:1109-11.
95. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0:

a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309-d314.
96. Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27-30.
97. Luo H, Lin Y, Gao F, Zhang CT, Zhang R. 2014. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic Acids Res 42:D574-80.
98. Hammerschmidt S, Rohde M. 2019. Electron microscopy to study the fine structure of the pneumococcal cell. Methods Mol Biol 1968:13-33.

Figures

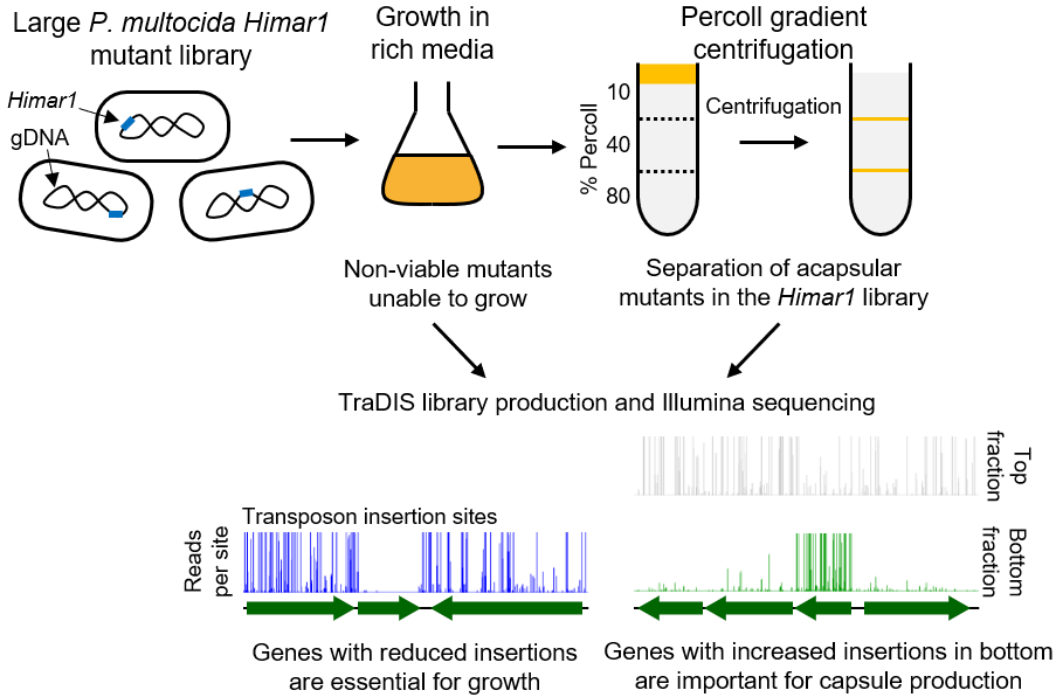


FIG 1. Overview of TraDIS experiments performed in this study. A large, random *P. multocida* mutant library was generated using the *Himar1* transposon. The VP161 *Himar1* mutant library was grown in heart infusion broth to mid-exponential stage with TraDIS analysis identifying rich media essential genes. The mutant library was separated through a discontinuous 10%, 40%, 80% Percoll gradient to isolate acapsular VP161 *Himar1* mutants. TraDIS libraries were sequenced via Illumina sequencing, with reads aligned to the VP161 genome to identify *Himar1* insertion sites. For growth in rich media, genes with reduced unique insertion sites were designated as essential genes. For capsule gene analysis, genes with significantly increased reads in the bottom fraction compared to the top fraction were designated important for capsule production. gDNA: genomic DNA.

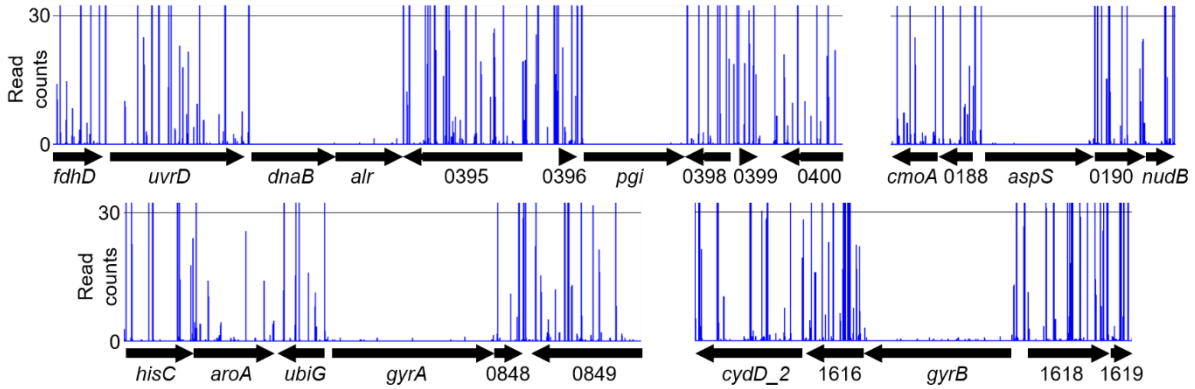


FIG 2. Transposon insertion sites identified in selected genomic regions in *P. multocida* VP161-Tn*7* *Himar1* mutants recovered following growth in heart infusion (HI) broth. Genes essential for growth in HI were identified as those genes showing a highly reduced number of unique insertion sites. Genomic regions encompassing the identified essential genes *dnaB*, *alr*, *pgi*, *aspS*, *gyrA* and *gyrB* are shown. Each position on the horizontal axis represents a unique *Himar1* insertion and the vertical axis shows total read counts for each insertion site. Numbers below horizontal arrows represent PmVP161 locus tag numbers.

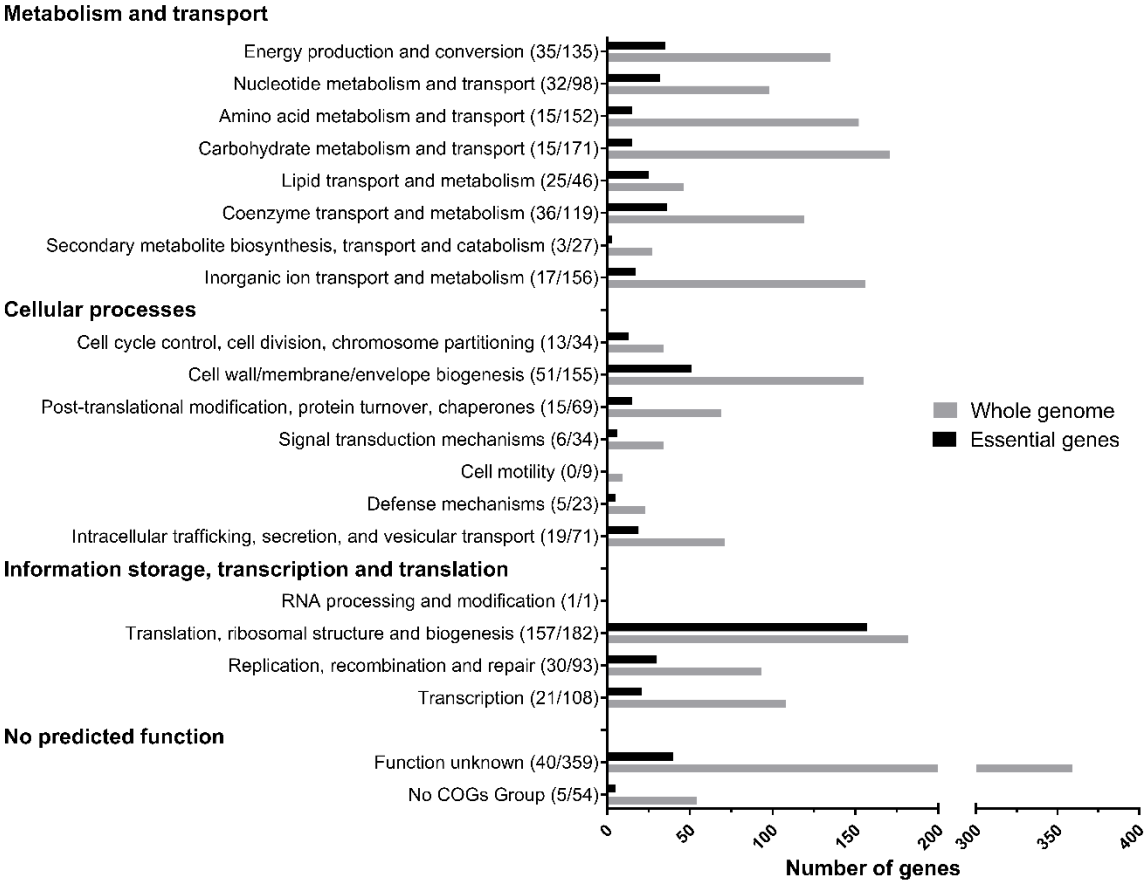


FIG 3. Cluster of orthologous groups (COGs) categories for genes essential for *P. multocida* VP161-Tn*7* growth in rich media, compared to whole genome COGs counts. Essential gene counts are in black and whole genome counts are in grey.

TABLE 1. *P. multocida* genes important for hyaluronic acid capsule production. Genes were predicted as important for capsule production if there was a four-fold increase in *Himar1* insertions (log_2 fold-change > 2.0) in the population recovered from the bottom fraction compared to the population recovered from the top fraction, with a *q*-value < 0.001 and an insertion index ratio > 0.8

Gene name	VP161 locus tag	Function	logFC	*q* value	Insertion index ratio
PmVP161_0059	PmVP161_0059	hypothetical protein	7.6	2.78E-106	1.1
yibF	PmVP161_0060	putative GST-like protein YibF	2.3	1.46E-07	0.9
dus	PmVP161_0075	putative tRNA-dihydrouridine synthase	14.1	6.77E-194	11.4
fis	PmVP161_0076	DNA-binding protein Fis	13.9	5.74E-140	14.0
pgm	PmVP161_0089	Phosphoglucomutase	12.9	1.05E-211	4.8
cpdA	PmVP161_0148	3',5'-cyclic adenosine monophosphate phosphodiesterase CpdA	8.1	1.08E-22	2.1
znuA	PmVP161_0255	High-affinity zinc uptake system protein ZnuA	7.2	1.02E-43	3.1
gmk	PmVP161_0260	Guanylate kinase	7.1	0.000618	3.0
rpoZ	PmVP161_0261	DNA-directed RNA polymerase subunit omega	6.8	3.22E-60	1.5
spoT	PmVP161_0262	Bifunctional (p)ppGpp synthase/hydrolase SpoT	6.0	1.19E-35	1.7
seqA	PmVP161_0336	Negative modulator of initiation of replication	5.6	2.22E-49	0.9
glnD	PmVP161_0442	Bifunctional uridylyltransferase/uridylyl-removing enzyme	7.6	8.78E-115	0.9
prmC	PmVP161_0549	Release factor glutamine methyltransferase	12.5	2.13E-52	26.0
truA	PmVP161_0625	tRNA pseudouridine synthase A	4.1	3.03E-34	0.8
ubiX	PmVP161_0693	Flavin prenyltransferase UbiX	3.9	3.79E-25	0.9
pta	PmVP161_0698	Phosphate acetyltransferase	6.3	4.42E-06	1.6
dnaJ	PmVP161_0735	Chaperone protein DnaJ	9.1	1.50E-52	3.5
rimP	PmVP161_0757	Ribosome maturation factor RimP	9.9	2.84E-83	2.7
phyB	PmVP161_0773	Capsule phospholipid substitution protein	12.7	5.07E-159	6.1
phyA	PmVP161_0774	Capsule phospholipid substitution protein	12.5	1.16E-152	6.2
hyaE	PmVP161_0775	Hyaluronic acid capsule biosynthesis protein	13.7	1.42E-141	6.0
hyaD	PmVP161_0776	Hyaluronic acid synthase	14.0	1.03E-131	8.4
hyaC	PmVP161_0777	UDP-glucose 6-dehydrogenase	13.5	8.45E-125	12.2
hyaB	PmVP161_0778	Hyaluronic acid capsule biosynthesis protein	12.7	5.64E-184	5.8
hexD	PmVP161_0779	Hyaluronic acid capsule export	8.1	1.97E-22	3.7
hexC	PmVP161_0780	Hyaluronic acid capsule export	7.8	2.08E-28	3.2
hexB	PmVP161_0781	Hyaluronic acid capsule transport protein HexB	6.5	1.34E-10	2.2
hexA	PmVP161_0782	Hyaluronic acid capsule transport ATP-binding protein HexA	5.3	7.72E-08	2.0
PmVP161_0825	PmVP161_0825	hypothetical protein	6.3	4.00E-76	0.9
holC	PmVP161_0828	DNA polymerase III subunit chi	2.9	7.33E-09	1.5
gyrA	PmVP161_0847	DNA gyrase subunit A	9.6	1.19E-23	3.0
PmVP161_0878	PmVP161_0878	hypothetical protein	11.9	1.36E-107	18.4
ptsH	PmVP161_0905	Phosphocarrier protein HPr	9.8	6.26E-102	2.9
hfq	PmVP161_0914	RNA-binding protein Hfq	6.8	1.61E-63	2.5
znuB	PmVP161_0986	High-affinity zinc uptake system membrane protein ZnuB	6.0	5.09E-09	1.1
secG	PmVP161_1023	Protein-export membrane protein SecG	10.8	7.65E-36	7.0
epmA	PmVP161_1029	Elongation factor P--(R)-beta-lysine ligase	5.7	2.78E-61	0.9
queD	PmVP161_1046	6-carboxy-5,6,7,8-tetrahydropterin synthase	7.1	1.63E-73	1.0
rnb	PmVP161_1050	Exoribonuclease 2	5.5	2.24E-44	1.2
efp	PmVP161_1135	Elongation factor P	7.9	1.95E-16	7.0
epmB	PmVP161_1136	L-lysine 2,3-aminomutase	7.6	2.12E-32	1.7
alsT_2	PmVP161_1144	Amino-acid carrier protein AlsT	4.6	4.65E-31	1.1
rseP	PmVP161_1260	Regulator of sigma-E protease RseP	5.7	0.00018	1.6
cpxA	PmVP161_1346	Sensor histidine kinase CpxA	2.5	1.94E-14	0.9
relA	PmVP161_1370	GTP pyrophosphokinase	6.3	6.54E-74	1.1
purB	PmVP161_1383	Adenylosuccinate lyase	4.0	4.34E-48	0.9
ppx	PmVP161_1426	Exopolyphosphatase	8.6	9.75E-81	2.2
trmA	PmVP161_1430	tRNA/tmRNA (uracil-C(5))-methyltransferase	7.6	8.83E-92	1.2
PmVP161_1431	PmVP161_1431	hypothetical protein	9.2	1.14E-60	2.5
tufA_1	PmVP161_1480	Elongation factor Tu 1	6.0	1.33E-61	1.5
mioC	PmVP161_1625	Protein MioC	3.0	0.000615	0.9
PmVP161_1628	PmVP161_1628	hypothetical protein	7.1	7.11E-94	2.2
atpB	PmVP161_1629	ATP synthase subunit a	4.8	3.22E-56	4.5
atpH_1	PmVP161_1630	ATP synthase subunit c	7.8	3.43E-63	2.1
atpF	PmVP161_1631	ATP synthase subunit b	8.1	1.32E-34	3.5
atpH_2	PmVP161_1632	ATP synthase subunit delta	4.3	1.10E-50	2.2
atpA	PmVP161_1633	ATP synthase subunit alpha	8.0	1.42E-80	2.5
atpG	PmVP161_1634	ATP synthase gamma chain	7.0	5.52E-68	2.9
atpD	PmVP161_1635	ATP synthase subunit beta	7.7	1.86E-80	2.2
pgl	PmVP161_1693	6-phosphogluconolactonase	5.6	1.24E-30	1.0
rpe	PmVP161_1767	Ribulose-phosphate 3-epimerase	3.8	4.88E-45	1.2
thiE	PmVP161_1798	Thiamine-phosphate synthase	4.0	0.000551	0.9
thiM	PmVP161_1800	Hydroxyethylthiazole kinase	6.6	6.31E-05	2.2
ssuB	PmVP161_1804	Aliphatic sulfonates import ATP-binding protein SsuB	6.6	5.96E-26	2.5
galU	PmVP161_1828	UTP--glucose-1-phosphate uridylyltransferase	13.3	2.19E-175	5.4
PmVP161_1879	PmVP161_1879	hypothetical protein	7.1	7.15E-05	1.5
tufA_2	PmVP161_1902	Elongation factor Tu 1	6.0	9.15E-63	1.4
glmS	PmVP161_1917	Glutamine--fructose-6-phosphate aminotransferase (isomerizing)	5.9	6.37E-23	1.6
rpiA	PmVP161_1987	Ribose-5-phosphate isomerase A	8.6	2.36E-46	2.7

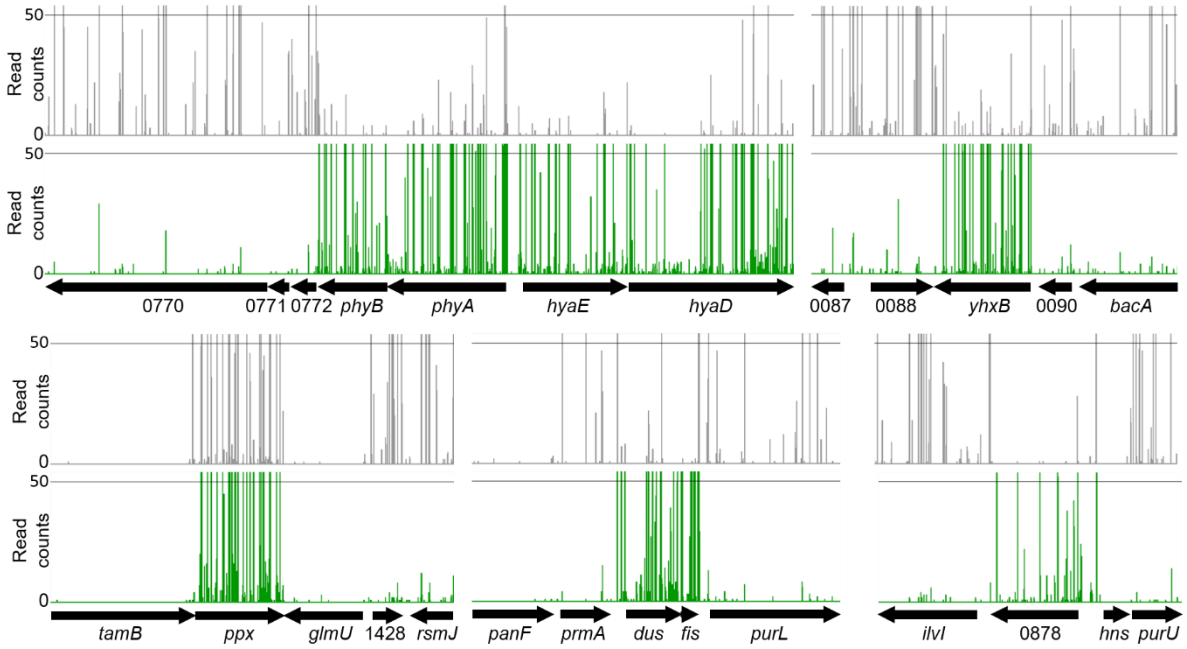


FIG 4. Transposon insertion sites in selected genomic regions in *P. multocida* VP161-Tn*7* *Himar1* mutants recovered from either the top fraction (shown in grey) or bottom fraction (shown in green) of Percoll gradient centrifugation. Genes with a four-fold increase (log_2 fold-change > 2.0) in insertions in the mutant population recovered from the bottom fractions compared to the mutants in the top fraction, a *q*-value < 0.001 and an insertion index ratio > 0.8 were designated as important for HA capsule production. *Himar1* insertion sites in genomic regions surrounding genes identified as important for capsule production (*phyAB*, *yhxB*, *ppx*, *dus*, *fis* and PmVP161_0878) are shown. Each position on the horizontal axis represents a unique *Himar1* insertion and the vertical axis shows total read counts for each insertion site. Numbers below the horizontal arrows represent PmVP161 locus tag numbers.

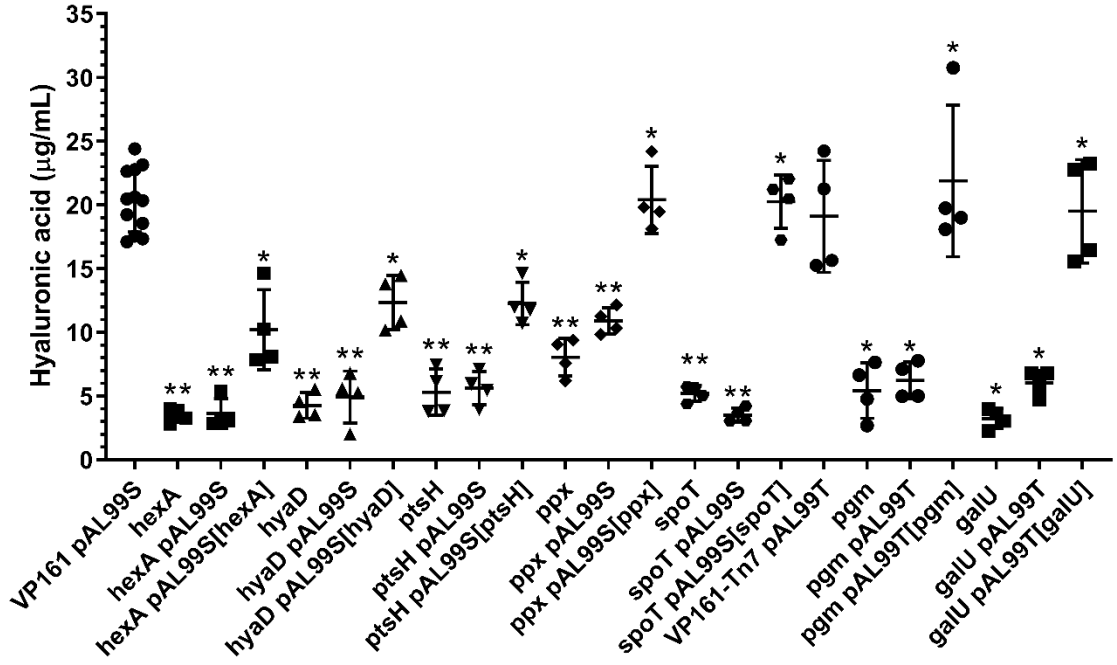


FIG 5. Amount of hyaluronic acid (HA) capsule produced by a range of *P. multocida* VP161 mutants and complemented mutants. HA was measured from cells grown to mid-exponential growth phase (OD_{600} 0.5-0.6). Capsule was measured for control strains VP161 and VP161-Tn*7* harbouring pAL99S or pAL99T empty vector, and each TargeTron or *Himar1* mutant, mutant harbouring pAL99S or pAL99T empty vector, and mutant harbouring pAL99S or pAL99T complementation vectors containing an intact copy of the appropriate gene. HA was measured from each strain in biological quadruplicate. The production of HA in mutants and mutants harbouring empty vector was compared to the amount of HA produced by the wild-type parent strain, and the amount of HA capsule produced by complemented mutants was compared to HA production in mutants harbouring empty vector. Error bars represent the mean ± standard deviation. Strains were compared using a Mann-Whitney U-test, with * representing $p < 0.05$ and ** representing $p < 0.01$.

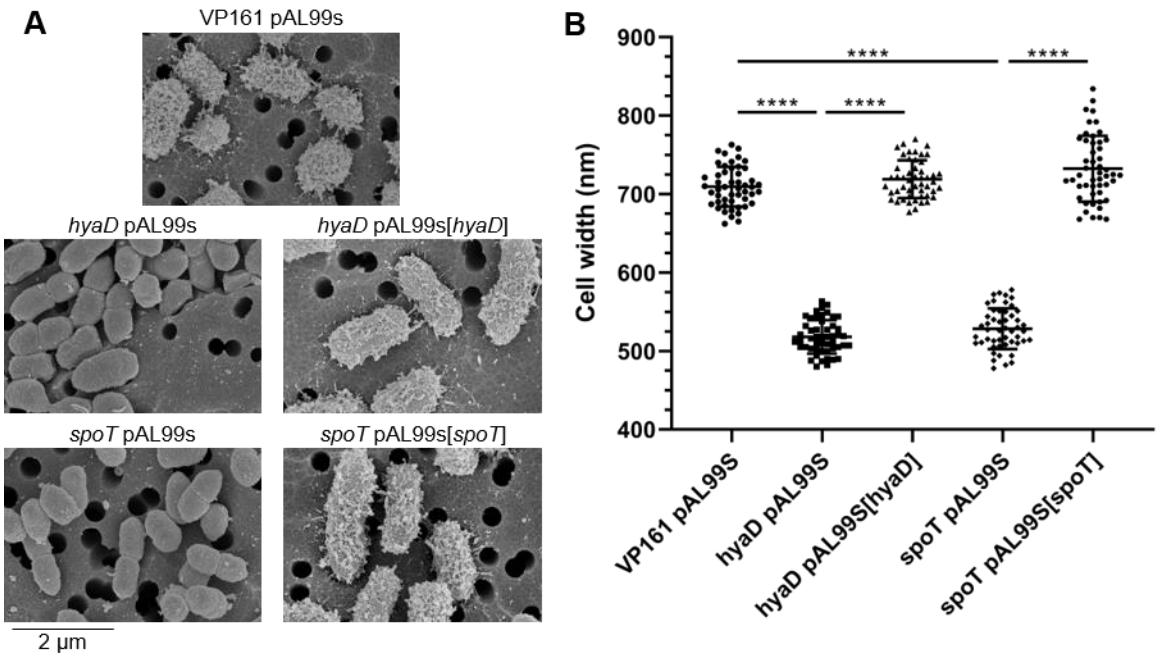


FIG 6. Scanning electron microscopy (SEM) of *P. multocida* strain VP161, and *spoT* and *hyaD* TargeTron mutant strains. SEM was performed using cultures at mid-exponential growth phase (OD_{600} ~0.4), with cells prepared for imaging by lysine-ruthenium red-osmium fixation and critical point drying as previously described (98). Microscopy was performed using wild-type parent strain VP161 harbouring pAL99S empty vector and TargeTron mutant strains harbouring pAL99S empty vector or pAL99S complementation vectors containing an intact copy of the appropriate gene. (A) Electron micrographs of the strains imaged by SEM. (B) Cell widths for 50 individual cells of each strain were measured using the FIJI imaging package in ImageJ, with error bars representing the mean ± standard deviation. **** represents $p < 0.0001$ when strains were compared using an unpaired t-test.

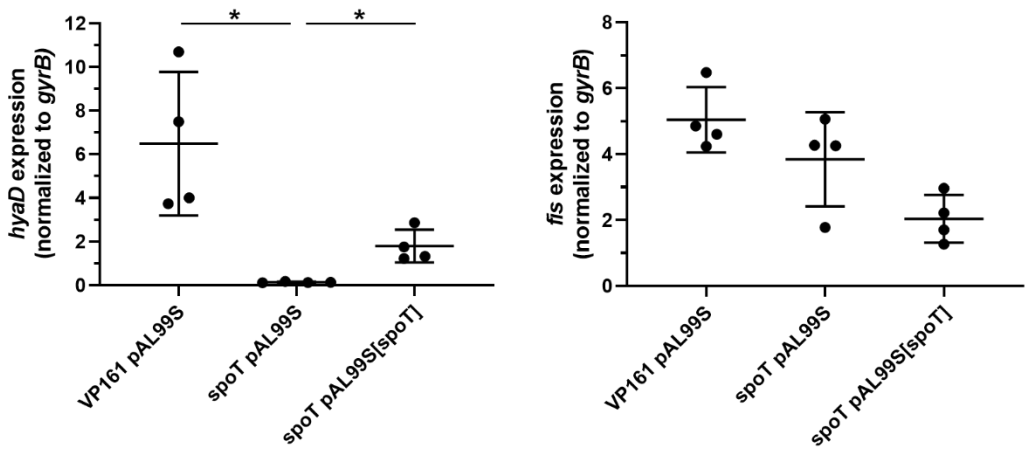


FIG 7. The expression levels of *hyaD* and *fis* in *P. multocida* VP161, the VP161 *spoT* TargeTron mutant and complemented *spoT* mutant. RNA for qRT-PCR was extracted from mid-exponential growth phase cultures (OD_{600} ~0.4). The relative expression of *hyaD* and *fis* was determined by comparing expression of both genes in VP161 harbouring pAL99S empty vector with the VP161 *spoT* mutant harbouring either pAL99S empty vector or pAL99S complementation vector containing an intact copy of *spoT*. Expression of both *hyaD* and *fis* was determined using four biological replicates. Expression of both genes was normalised to the level of housekeeping gene *gyrB*. Error bars represent the mean ± standard deviation, and expression between strains was compared using a Mann-Whitney U-test, with * representing $p < 0.05$.

Manuscript supplementary material

TABLE S1. Strains and plasmids used in this study.

Strain or plasmid	Description	Source or reference
Strains		
P. multocida		
VP161	Avian isolate strain, serotype A:1	(1)
VP161-Tn*7*	VP161 with Tn*7* downstream of *glmS;* Kan^R	This study
PBA930	X73 *hexA*::*tet*(M) mutant	(2)
AL2188	VP161 harbouring pAL99S; $Spec^R$	Akhlaghi, personal communication
AL2232	VP161 harbouring pAL99T; Tet^R	John, personal communication
AL3544	VP161 *ppx* TargeTron® mutant, intron inserted between nucleotides 474-475; Kan^R	This study
AL3574	VP161 *hyaD* TargeTron® mutant	This study
AL3684	AL3145 harbouring pAL99T	This study
AL3685	VP161 *hexA* TargeTron® mutant, intron inserted between nucleotides 123-124; Kan^R	This study
AL3793	VP161 *ptsH* TargeTron® mutant, intron inserted between nucleotides 58-59; Kan^R	This study
AL3818	VP161-Tn*7* *pgm Himar1* mutant, transposon inserted between nucleotides 121-122; $Spec^R$	This study
AL3819	VP161-Tn*7* *galU Himar1* mutant, transposon inserted between nucleotides 476-477; $Spec^R$	This study
AL3874	AL3818 harbouring pAL99T; $Spec^R$ Tet^R	This study
AL3876	AL3818 harbouring pAL1676; $Spec^R$ Tet^R	This study
AL3878	AL3819 harbouring pAL99T; $Spec^R$ Tet^R	This study
AL3880	AL3818 harbouring pAL1678; $Spec^R$ Tet^R	This study
AL3953	AL3793 harbouring pAL99S; Kan^R $Spec^R$	This study
AL3954	AL3793 harbouring pAL1687; Kan^R $Spec^R$	This study
AL3955	AL3544 harbouring pAL99S; Kan^R $Spec^R$	This study
AL3956	AL3544 harbouring pAL1690; Kan^R $Spec^R$	This study
AL3957	AL3685 harbouring pAL99S; Kan^R $Spec^R$	This study
AL3958	AL3685 harbouring pAL1691; Kan^R $Spec^R$	This study
AL3961	AL3574 harbouring pAL99S; $Spec^R$	This study
AL3962	AL3574 harbouring pAL1693; $Spec^R$	This study
AL3991	VP161 *spoT* TargeTron® mutant, intron inserted between nucleotides 1779-1780; Kan^R	This study
AL3993	AL3991 harbouring pAL99S; Kan^R $Spec^R$	This study
AL3994	AL3991 harbouring pAL1700; Kan^R $Spec^R$	This study
E. coli		
DH5α	F^- deoR endA1 gyrA96 hsdR17 (r_K^- m_K^-) recA1 relA1 supE44 thi-1 Δ(lacZYAargFV169) ϕ80lacZΔM15	Bethesda Research Laboratories
S17-1 λ*pir*	(F^-) RP4-2-Tc::Mu *aphA*::Tn*7***recA*λ*pir* lysogen Sm^r Tp^r	(3)
AL1296	DH5α harbouring pAL99S; $Spec^R$	(4)
AL1995	DH5α harbouring pAL953; Kan^R $Spec^R$	(4)
AL2017	S17-1 λ*pir* harbouring pTNS3; donor strain for Tn*7* mutagenesis	Alwis, personal communication
AL2224	DH5α harbouring pAL99T; Tet^R	(4)
AL2621	S17-1 λ*pir* harbouring pAL1083; donor strain for Tn*7* mutagenesis	Alwis, personal communication
AL2972	S17λpir harbouring pAL614, Amp^R $Spec^R$	This study
Plasmids		
pTNS3	Tn*7* helper plasmid encoding the Tn*7* site-specific transposition pathway *tnsABCD*; Amp^R	(5)
pAL1083	Tn*7* delivery vector; mini-Tn*7* containing Kan^R oriColE1 Amp^R	Alwis, personal communication
pAL99S	*P. multocida-E. coli* expression plasmid; $Spec^R$ P*tpiA*	(4)
pAL99T	*P. multocida-E. coli* expression plasmid; Tet^R P*tpiA*	(4)
pAL614	RP4 mobilisable plasmid containing *Himar1*::Spec; RP4 Mob^+ oriR6K Amp^R $Spec^R$ tnpC9	(6)
pAL953	*P. multocida* plasmid containing TargeTron group II intron; $Spec^R$ Kan^R	Harper, 2013 #371}
pAL1337	pAL953 with modified group II intron retargeted to *hyaD*; $Spec^R$	Gulliver, personal communication
pAL1504	pAL953 with modified group II intron retargeted to *hexA*; $Spec^R$ Kan^R	This study
pAL1511	pAL953 with modified group II intron retargeted to *ppx*; $Spec^R$ Kan^R	This study
pAL1600	pAL953 with modified group II intron retargeted to *ptsH*; $Spec^R$ Kan^R	This study
pAL1658	pAL953 with modified group II intron retargeted to *spoT*; $Spec^R$ Kan^R	This study
pAL1676	*yhxB* from VP161 cloned into pAL99T, Tet^R	This study
pAL1678	*galU yhxB* from VP161 cloned into pAL99T, Tet^R	This study

Strain or plasmid	Description	Source or reference
pAL1687	*ptsH* from VP161 cloned into pAL99S, Spec^R	This study
pAL1690	*ppx* from VP161 cloned into pAL99S, Spec^R	This study
pAL1691	*hexA* from VP161 cloned into pAL99S, Spec^R	This study
pAL1693	*hyaD* from VP161 cloned into pAL99S, Spec^R	This study
pAL1700	*spot* from VP161 cloned into pAL99S, Spec^R	This study

TABLE S2. Oligonucleotides used in this study

Oligonucleotide	Sequence (5'-3')	Desription
BAP612	GTAAAACGACGGCCAGT	M13 universal primer, used for sequencing the cloning region in pAL99 and derivatives
BAP2106	GCCCTTTCCGATAAATTGCAA	Forward *gyrB* oligonucleotide used for qRT-PCR
BAP2107	ATCGCGGCTAATGGTGCTT	Reverse *gyrB* oligonucleotide used for qRT-PCR
BAP2679	TTGTGTGGAATTGTGAGCGGA	M13 universal primer, used for sequencing the cloning region in pAL99 and derivatives
BAP3453	CGTCATGGTCTTTGTAGTCTATGG	Forward primer that anneals upstream of the *Himar1* right inverted repeat, used to identify *Himar1* insertion sites in transposon mutants
BAP6544	CGAAATTAGAAACTTGCGTTCAGTAAAC	EBS universal for TargeTron mutagenesis
BAP8034	P-G*ATCGGAAGAGCGGTTCAGCAGGTTTTTTTTTTCAAAAAAA*A	Splinkerette adapter top strand; P- represents a 5' phosphate and * represents a phosphorothioate bond (7)
BAP8035	G*AGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T	Splinkerette adapter bottom strand; P- represents a 5' phosphate and * represents a phosphorothioate bond (7)
BAP8037	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGAGATCGGTCTCGGCATTCC	Splinkerette adapter-specific oligonucleotide for TraDIS library amplification, contains an index sequence to allow multiplexing of libraries and Illumina P7 sequence at the 5' end
BAP8038	CAAGCAGAAGACGGCATACGAGATCGTACTAGGAGATCGGTCTCGGCATTCC	Splinkerette adapter-specific oligonucleotide for TraDIS library amplification, contains an index sequence to allow multiplexing of libraries and Illumina P7 sequence at the 5' end
BAP8039	CAAGCAGAAGACGGCATACGAGATAGGCAGAAGAGATCGGTCTCGGCATTCC	Splinkerette adapter-specific oligonucleotide for TraDIS library amplification, contains an index sequence to allow multiplexing of libraries and Illumina P7 sequence at the 5' end
BAP8040	CAAGCAGAAGACGGCATACGAGATTCCTGAGCGAGATCGGTCTCGGCATTCC	Splinkerette adapter-specific oligonucleotide for TraDIS library amplification, contains an index sequence to allow multiplexing of libraries and Illumina P7 sequence at the 5' end
BAP8042	GGTTCTAGAGACCGGGGACTTATCAGC	Custom *Himar1* transposon-specific Illumina sequencing oligonucleotide
BAP8043	TTCAGCAGGAATGCCGAGACCGATCTC	Custom adapter index-specific Illumina sequencing oligonucleotide
BAP8265	GTAACATCAGAGGGTACCTCG	Sequencing oligonucleotide to check Tn*7* insertion, binds upstream from the right junction and sequences into genomic DNA
BAP8350	AATGATACGGCGACCACCGAGATCTACACCTCTAGAAAGTATAGGAACTTCGAACCG	*Himar1*-specific oligonucleotide for TraDIS library amplification, contains Illumina P5 at the 5' end
BAP8479	AATGATACGGCGACCACCGAGATCTACACCTGGGGTACGCGTCTAGGCGG	*Himar1*-specific oligonucleotide for TraDIS library amplification, contains Illumina P5 at the 5' end
BAP8707	AAAAAAGCTTATAATTATCCTTATGATTCCCCTGCGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *hexA*
BAP8708	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCCCTGCCCTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *hexA*
BAP8709	TGAACGCAAGTTTCTAATTTCGGTTAATCATCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *hexA*
BAP8712	AAAAAAGCTTATAATTATCCTTACGTCACGCCGCAGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ppx*
BAP8713	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGCCGCATTTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ppx*
BAP8714	TGAACGCAAGTTTCTAATTTCGGTTTGACGTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ppx*
BAP8865	AGCGGGGTCGACTTATTTGGTTTGTAATGCATTATTA	Reverse flanking *hexA* oligonucleotide, includes SalI site
BAP8868	GTTAATGAGCTCAGGAGGGAAAATGAATACATTATCACAAGCA	Forward flanking *hyaD* oligonucleotide, includes SacI site

Oligonucleotide	Sequence (5'-3')	Desription
BAP8869	AATGCAGTCGACTTATAGAGTTATACTATTAATAATGA	Reverse flanking *hyaD* oligonucleotide, includes SalI site
BAP8945	TGGGCGGGATCCTTAGTTAAACGTGATATCTAATCC	Reverse flanking *ppx* oligonucleotide, includes BamHI site
BAP8960	AAAAAAGCTTATAATTATCCTTACTCGTCCTGCGGGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ptsH*
BAP8961	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTGCGGCATAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ptsH*
BAP8962	TGAACGCAAGTTTCTAATTTCGATTACGGCTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ptsH*
BAP8986	TGATGGTAGTCAGGAAGATCTA	*hyaD*-specific oligonucleotide used to sequence complementation plasmid
BAP8987	AGGGTACGCATCATGTCTAAA	*hyaD*-specific oligonucleotide used to sequence complementation plasmid
BAP8988	TATTAATAAATTAAGTCAGTTAAATC	*hyaD*-specific oligonucleotide used to sequence complementation plasmid
BAP9070	GCTTAAGTCGACTTAGCAATCCTGCTTACCAA	Reverse flanking *pgm* oligonucleotide, includes SalI site
BAP9072	GCTAACGTCGACTTATTCTAAGGTTGGAATAAGT	Reverse flanking *ptsH* oligonucleotide, includes SalI site
BAP9151	AAAAAAGCTTATAATTATCCTTAAGCTCCGGCAAGGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *spoT*
BAP9152	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGGCAAGGGTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *spoT*
BAP9153	TGAACGCAAGTTTCTAATTTCGATTGAGCTTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *spoT*
BAP9157	TGGCTATCTAGAAGAAGGAAAAAAAATGAAAGCAATT	Forward flanking *galU* oligonucleotide, includes XmaI site
BAP9158	TAAACAGTCGACTTAAAAACTTTTGACGAGCTG	Reverse flanking *galU* oligonucleotide, includes SalI site
BAP9160	TCTTTGTCTAGAAGGAGGAAAAACATGTACTCAAAAGAC	Forward flanking *ptsH* oligonucleotide, includes XmaI site
BAP9162	TTAATATCTAGAAGGAGAATTTATGTCAATCTTT	Forward flanking *pgm* oligonucleotide, includes XmaI site
BAP9172	ATTTGACCCGGGAGGAAGAATATGAAGAGTGAAA	Forward flanking *ppx* oligonucleotide, includes XmaI site
BAP9185	TGTCAATGTTGTTGGCACC	*ppx*-specific oligonucleotide used to sequence complementation plasmid
BAP9186	CTTTTTCTTCTGGATTCGGG	*pgm*-specific oligonucleotide used to sequence complementation plasmid
BAP9187	TGCAGACCCGGGAGGAGGGATAGAGCCTCGATG	Forward flanking *hexA* oligonucleotide, includes XmaI site
BAP9188	ACATTGCCCGGGAGGAGGTGTCCCTTGTATCTTT	Forward flanking *spoT* oligonucleotide, includes XmaI site
BAP9189	GGAATTGTCGACTCATTGATTGATATTACGTTCA	Reverse flanking *spoT* oligonucleotide, includes SalI site
BAP9233	TCACCACCACGTAAATTTTT	*spoT*-specific oligonucleotide used to sequence complementation plasmid
BAP9299	GGGGTGGAGAAGATGTGGAA	Forward *hyaD* oligonucleotide used for qRT-PCR
BAP9300	CCTGGTGGCTCTTGATGGTA	Reverse *hyaD* oligonucleotide used for qRT-PCR
BAP9301	TCGCAGTTAGATGGCCAAGA	Forward *fis* oligonucleotide used for qRT-PCR
BAP9302	GCTGCACGAGTTTGATTTCC	Reverse *fis* oligonucleotide used for qRT-PCR

Detailed TraDIS library production method

The TraDIS library production method used in this study was adapted from L. Barquist et al. (7) for use with a large *P. multocida* strain VP161-Tn*7 Himar1* mutant library. Following growth or Percoll gradient separation of the transposon mutant library, VP161-Tn*7 Himar1* mutants were recovered as a pool and genomic DNA was extracted from all recovered mutants using HiYield genomic DNA MiniKit (RBC bioscience). A total of 6 µg of genomic DNA in 50 µL of dH_2O was sheared to ~300 bp fragments using Covaris Adaptive Focused Acoustics (AFA) with the following settings: Duty Cycle – 20%, Intensity – 5, Cycles – 200. Samples were sheared in Covaris microTUBEs with AFA Fibre and Snap-Caps. The sheared DNA fragments were purified using AxyPrep Mag PCR Purification Kit (Axygen) magnetic beads as per manufacturer's instructions with the exception that 80% EtOH was used instead of 70% EtOH. Sheared DNA samples were purified using a 1.5 x volume of magnetic beads with eluted in 51 µL dH_2O. The Sheared DNA fragments were then end-repaired using NEBNext End Repair module (NEB) as per the manufacturer's instructions using 50 µL of the purified fragmented DNA. The end-repaired DNA fragments were then purified using a 1.5x volume of magnetic beads with elution in 32 µL of dH_2O. The full aliquot of purified, end-repaired fragments was then 3' A-tailed using the NEBNext dA-tailing module (NEB) as per the manufacturer's instructions then purified using a 1.8 x volume of magnetic beads with elution in 20 µL of dH_2O. Splinkerette adapters were then ligated to the ends of purified 3' A-tailed DNA fragments. The top and bottom strand splinkerette oligonucleotides (BAP8034 and BAP8035 in Table S2) were annealed by mixing at a final concentration of 50 µM for each oligo, incubating at 99°C for 15 minutes then allowed to cool at room temperature. The annealed splinkerette adapter was then ligated onto the A-tailed DNA fragments using the NEBNext Quick Ligation Module (NEB) as per the manufacturer's instructions using 1 µL of the annealed adapter. The adapter-ligated fragments were then purified using a 1 x volume of magnetic beads with elution in 50 µL of dH_2O, followed by a 0.8 x volume of magnetic beads with elution in 20 µL of dH_2O

The purified adapter-ligated DNA fragments were then used as the template for PCR amplification to produce TraDIS libraries using KAPA HiFi HotStart ReadyMix PCR Kit (Roche). Each PCR was setup as follows: 25 µL of 2 x KapaHiFi HotStart ReadyMix, 1 µM each of the P5 *Himar1*-specific and P7 Adapter-specific primers (Table S2), either 100 ng or 250 ng of purified adapter-ligated DNA and a final volume of 50 µL. For the growth in rich media study, two TraDIS libraries were generated for each library replicate, and were amplified using 100 ng of DNA as template (HI TraDIS libraries 1 and 3) or 250 ng of DNA as template (HI TraDIS libraries 2 and 4). For the capsule study, all TraDIS libraries were generated using 250 ng of template. The thermocycler amplification conditions were as follows: 95°C for 2 min, 19 x [98°C for 20 sec, 65°C for 30 sec, 72°C for 45 sec], 72°C for 2 min, 12°C hold. All amplified

TraDIS libraries were then purified using a 0.8 x volume of magnetic beads with elution in 32 µL of dH_2O.

Following the shearing, adapter ligation and library PCR amplification steps, 1 µL of the purified DNA was diluted 1:10 in dH_2O and used to determine DNA concentration and size range using Qubit fluorometric quantification (Thermo Fisher) and Fragment Analyzer (Agilent) analyses. DNA size ranging was used to confirm Covaris shearing produced DNA fragments of ~300 bp and that there was minimal adapter dimer produced during the adapter ligation step. Next, the amount of Illumina flow cell clusterable DNA was determined using the KAPA Library Quantification Kit KK4824 (Roche) as per the manufacturer's instructions (universal qPCR Master Mix setup with no ROX). The TraDIS libraries were diluted 1:10,000 and 1:100,000 for this qPCR analysis. Following the above quality controls, all libraries were sequenced on an Illumina MiSeq v2 using 150 bp single end sequencing using the custom *Himar1*-specific oligonucleotides BAP8042 and BAP8043 that were used to sequence *Himar1* transposon-chromosome junctions and index sequences in the splinkerette adapter, respectively.

The TraDIS illumina sequencing data was trimmed to remove adapter sequences and poor-quality data using Trimmomatic v0.38 (8), using all adapter-specific primers and Illumina P5 and P7 sequences. Trimmomatic was run with the following settings: leading:3, trailing:3, slidingwindow:4:15, minlength:1, crop:40. The TraDIS sequencing data was then analysed using Bio-TraDIS v1.4.1 and modified scripts available at https://github.com/sanger-pathogens/Bio-Tradis and https://github.com/francesca-short/tradis_scripts (7, 9). *Himar1* insertion sites were mapped to the VP161 reference genome using the bacteria_tradis script with the following settings: --smalt –smalt_r 0 -m 0. The correlation of the number of unique *Himar1* insertion sites per gene between TraDIS library replicates was high (Fig. S1 and Fig. S5), allowing data from library replicates to be combined using combine_tradis_plots. Genes essential for growth in rich media were identified using the scripts tradis_gene_insert_sites and tradis_essentiality.R, with tradis_gene_insert_sites run with the setting -trim3 0.1 to ignore insertion sites in the final 10% of the genes. Genes important for capsule production in *P. multocida* were determined using the scripts tradis_gene_insert_sites with no additional settings, and TraDIS_comparison_positive_selection.R with the settings -q 0.001 -i 0.8.

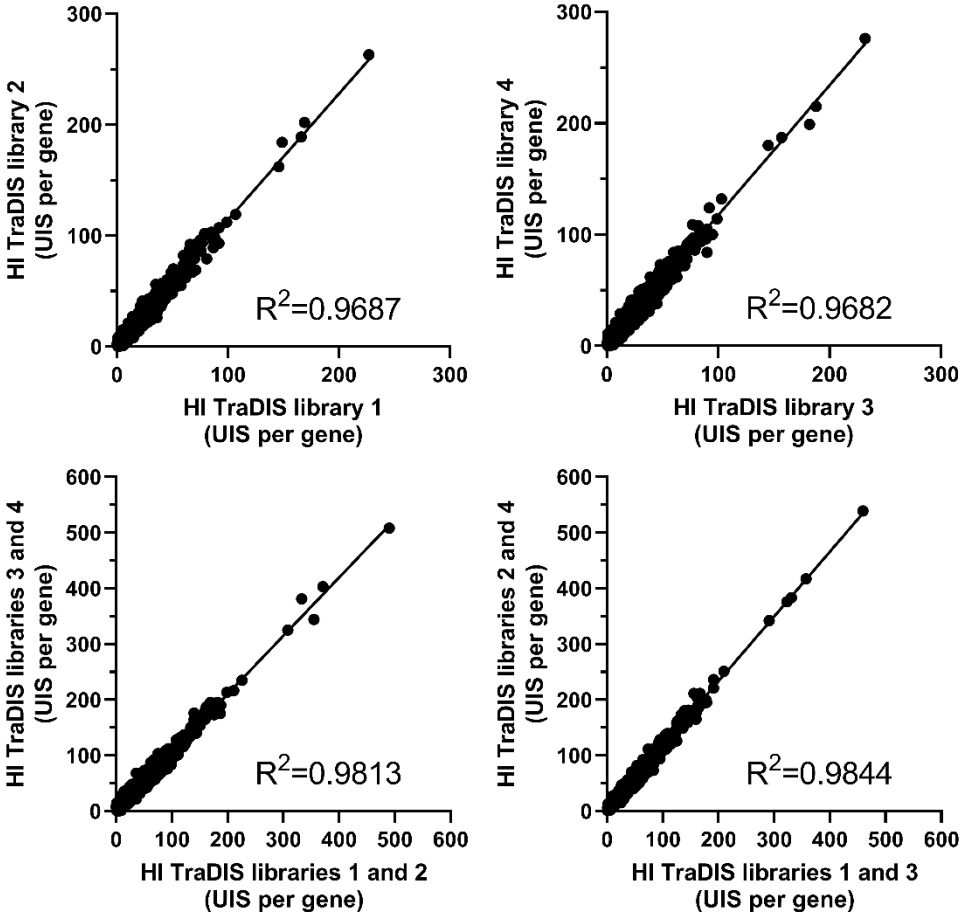


FIG S1. The number of unique *Himar1* insertion sites (UIS) identified in each gene following growth of the VP161-Tn*7* *Himar1* library in rich media. Comparisons were performed using TraDIS data obtained from the same VP161-Tn*7* *Himar1* mid-exponential phase growth culture (library 1 against library 2, library 3 against library 4), data obtained between the two mid-exponential phase growth cultures (libraries 1 and 2 against libraries 3 and 4) and to compare data obtained using two different amounts of starting template for library amplification (libraries 1 and 3 against libraries 2 and 4). The R^2 values were determined by simple linear regression analysis for each of the compared datasets.

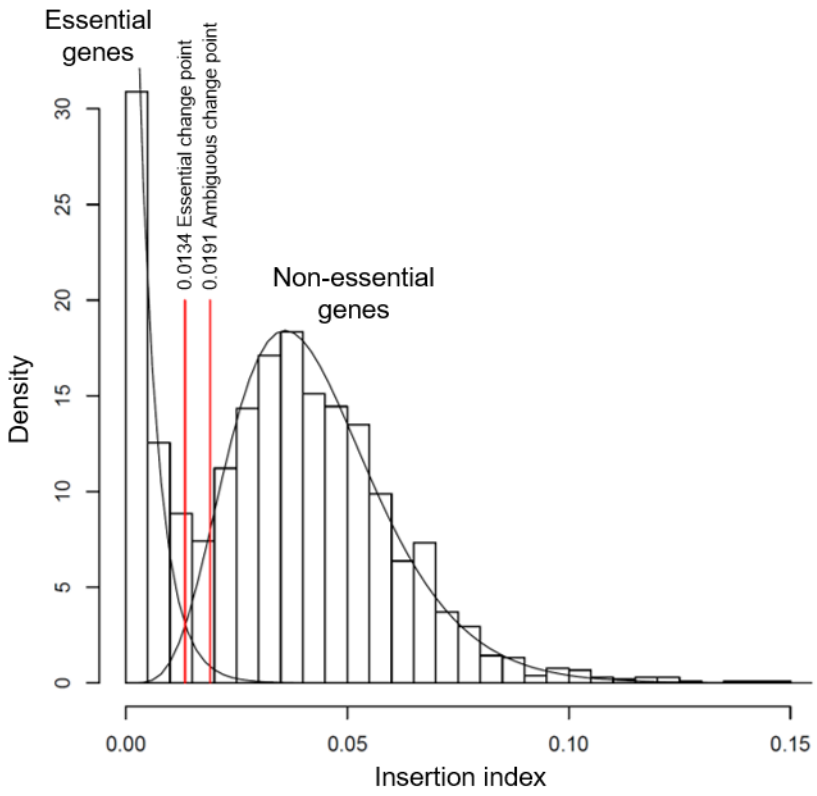


FIG S2. Histogram of insertion indexes for all genes in TraDIS libraries generated from VP161-Tn*7* *Himar1* mutants grown in heart infusion broth. Insertion indexes are calculated for all genes by dividing the number of unique *Himar1* insertion sites by gene length. Normal curves were drawn for the bimodal dataset, with the intersection between the two curves used as the cut-off to call a gene essential for *P. multocida* growth in rich media.

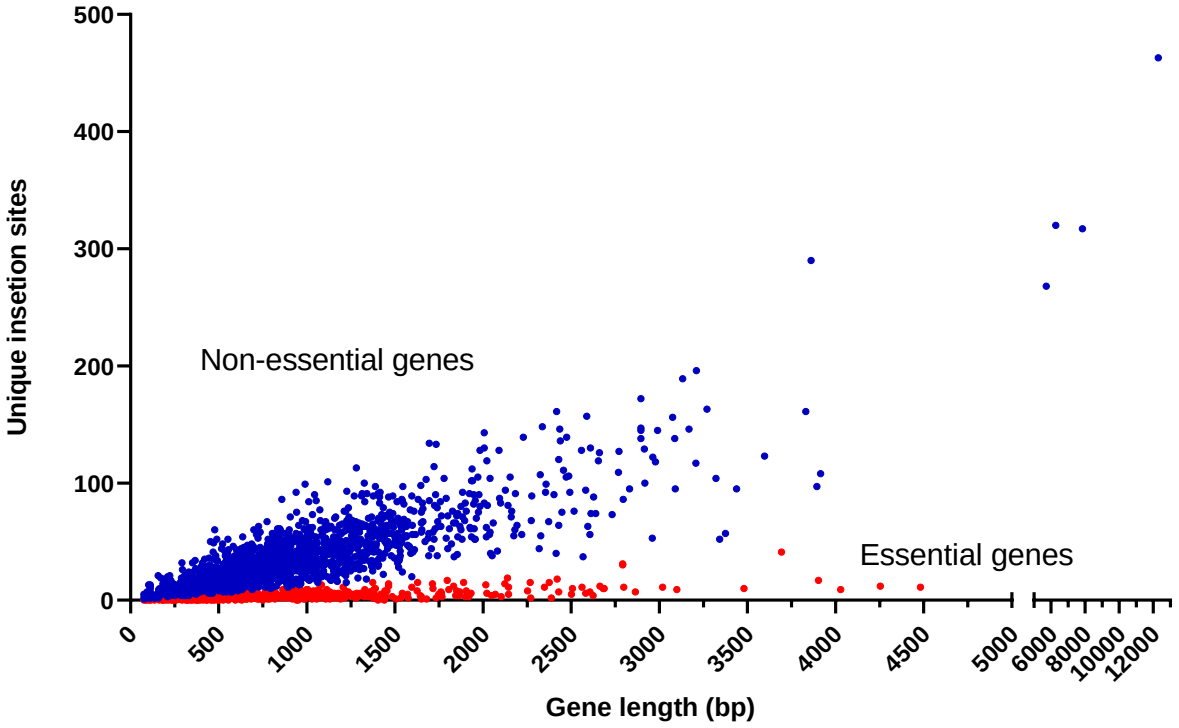


FIG S3. Insertion indexes for all genes in TraDIS libraries generated from VP161-Tn*7* *Himar1* mutants grown in heart infusion broth. All genes were plotted as the number of unique insertion sites against gene length. Genes not essential for *P. multocida* growth in rich media are represented by blue dots, whereas genes essential for *P. multocida* growth in rich media are represented by red dots.

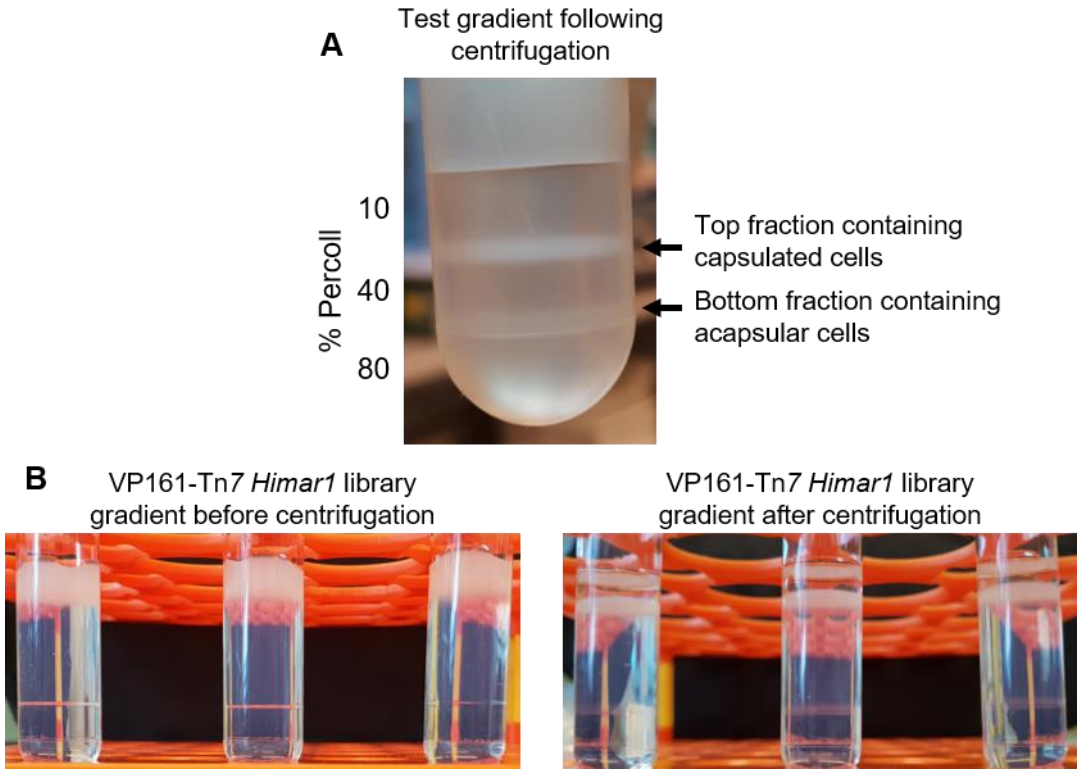


FIG S4. Discontinuous Percoll density gradient separation of acapsular *P. multocida* cells from a mixed population. (A) Test gradient separating a 1:1 mix of capsulated *P. multocida* strain VP161-Tn*7* and the acapsular *hexA* mutant *P. multocida* strain PBA930 separated through a discontinuous 10%, 40%, 80% Percoll density gradient. Following centrifugation, two cell fractions were observed at the 10% to 40% interface containing mostly capsular cells, and at the 40% to 80% interface containing mostly acapsular cells. (B) Separation of acapsular *P. multocida* VP161-Tn*7* *Himar1* mutants from the large VP161-Tn*7* *Himar1* mutant library using a 10%, 40%, 80% Percoll gradient, separating cells into capsulated and acapsular cell fractions.

TABLE S3. Genes disrupted by *Himar1* insertions in putatively acapsular *P. multocida* VP161-Tn*7* *Himar1* mutant strains recovered following discontinuous Percoll gradient centrifugation of the *P. multocida* VP161-Tn*7* *Himar1* mutant library.

Gene	VP161 locus tag	Unique insertion sites	Function
pgm	PmVP161_0089	7	Phosphoglucomutase
phyA	PmVP161_0774	6	Capsule phospholipid substitution protein
ubiG	PmVP161_0846	1	Ubiquinone biosynthesis O-methyltransferase
hyaD	PmVP161_0776	3	Hyaluronan synthase
ppx	PmVP161_1426	6	Exopolyphosphotase
galU	PmVP161_1828	3	UTP--glucose-1-phosphate uridylyltransferase
hyaE	PmVP161_0775	5	Hyaluronic acid capsule biosynthesis protein
cpxA	PmVP161_1346	1	Sensor histidine kinase CpxA
phyB	PmVP161_0773	1	Capsule phospholipid substitution protein
galM	PmVP161_0133	1	Aldose 1-epimerase
hyaC	PmVP161_0777	1	UDP-glucose 6-dehydrogenase

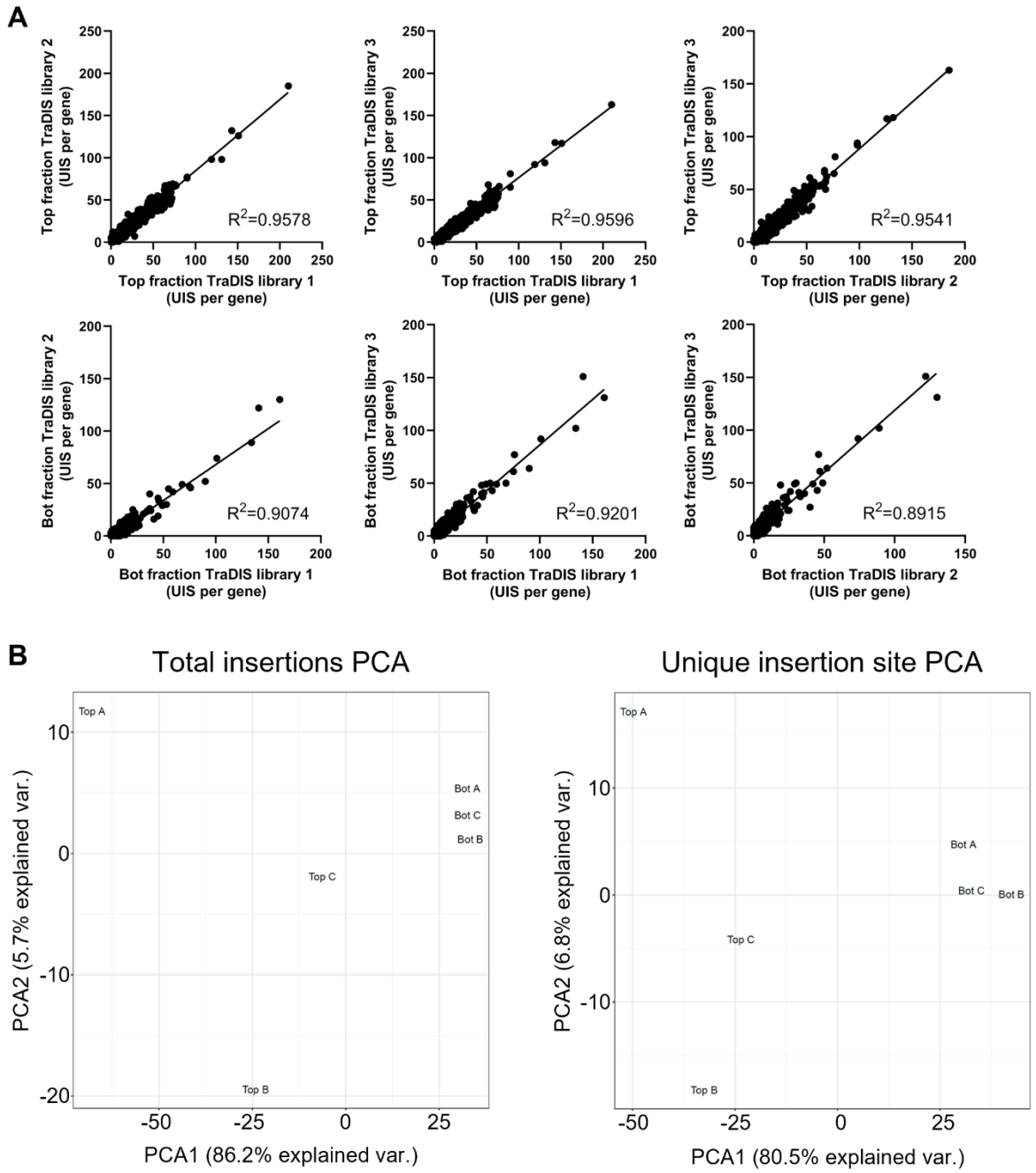


FIG S5. Reproducibility of the capsule TraDIS library replicates. (A) Comparison of the number of unique *Himar1* insertion sites (UIS) identified in each gene between TraDIS library replicates generated from cells recovered from either the top fraction or bottom fraction of the second Percoll gradient. The R^2 values were determined by simple linear regression analysis for each of the compared datasets. (B) Principal component analysis plots for the top fraction and bottom fraction TraDIS libraries generated by comparing either UIS or total *Himar1* insertions per gene.

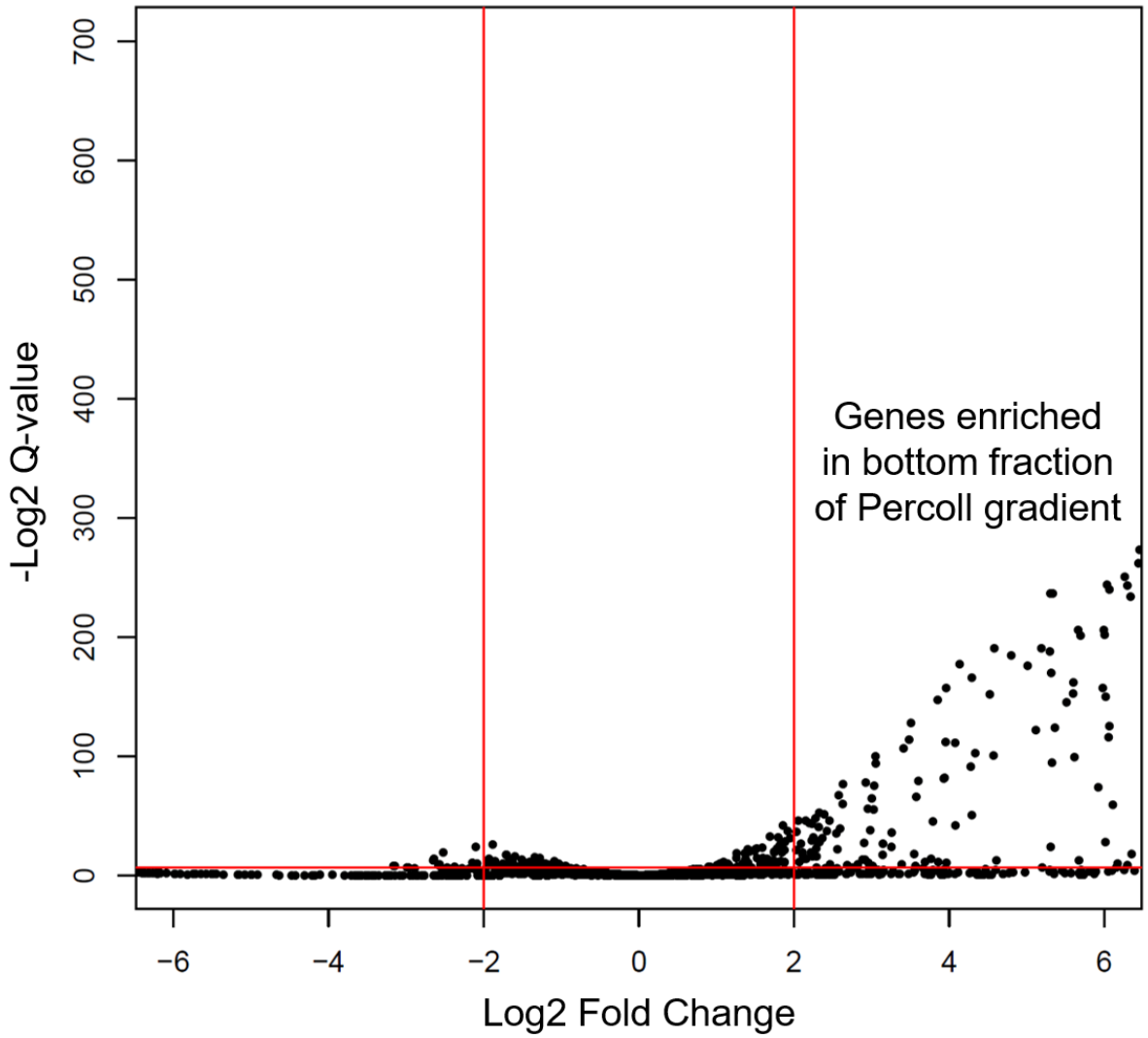


FIG S6. Volcano plot showing the relative fold change in total *Himar1* insertion sites per gene between the top and bottom Percoll gradient fractions. The x-axis shows the $\log_2$ fold change of total *Himar1* insertion sites per gene for the TraDIS libraries generated from the bottom fraction of the Percoll gradients compared to the TraDIS libraries generated from the top fraction. The $-\log_2$ *q*-value is shown on the y-axis. Genes with a $\log_2$ fold change > 2 and a *q*-value < 0.001 were identified as important for *P. multocida* hyaluronic acid capsule production and are on the right of the volcano plot.

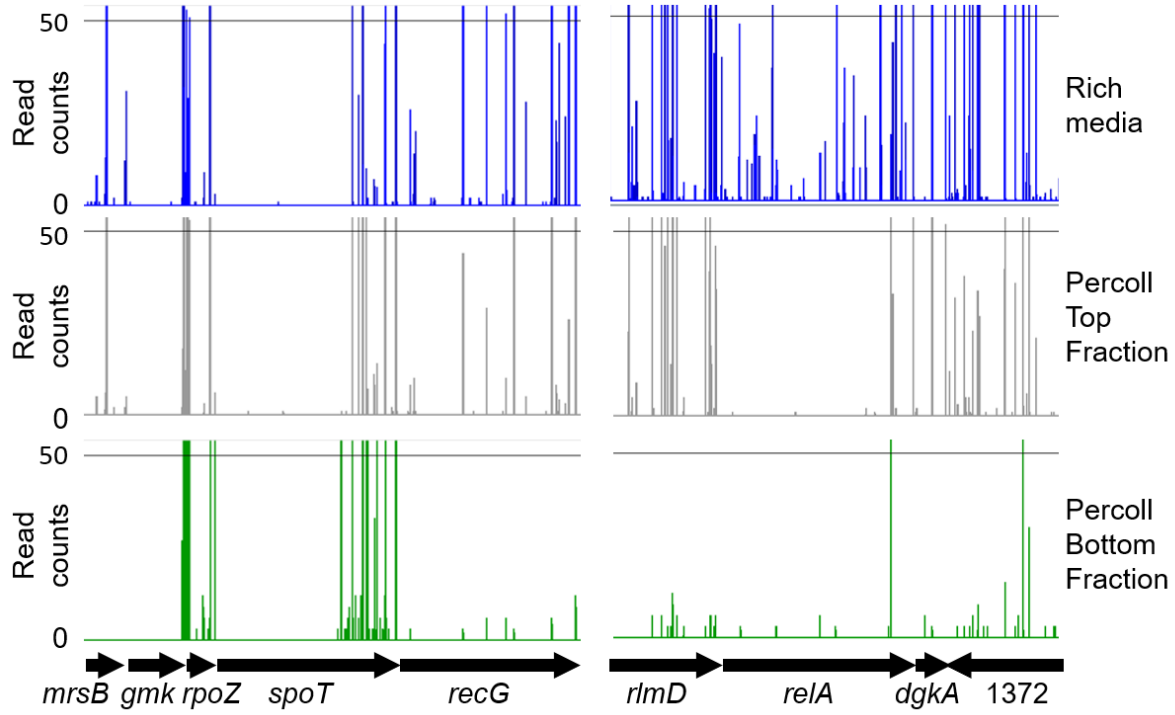


FIG S7. Insertion site plots showing *Himar1* insertions in *spoT* and *relA* with the surrounding genomic region. Insertion sites are shown for data obtained from combined rich media TraDIS libraries (blue), combined Percoll gradient top fraction TraDIS libraries (grey) and combined Percoll gradient bottom fractions (green). The numbers below the arrows represent *P. multocida* strain VP161 locus tag numbers.

References

1. Wilkie IW, Grimes SE, O'Boyle D, Frost AJ. 2000. The virulence and protective efficacy for chickens of *Pasteurella multocida* administered by different routes. Vet Microbiol 72:57-68.
2. Chung JY, Wilkie I, Boyce JD, Townsend KM, Frost AJ, Ghoddusi M, Adler B. 2001. Role of capsule in the pathogenesis of fowl cholera caused by *Pasteurella multocida* serogroup A. Infect Immun 69:2487-92.
3. Simon R, Priefer U, Pühler A. 1983. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. Nat Biotechnol 1:784-791.
4. Harper M, St Michael F, John M, Vinogradov E, Steen JA, van Dorsten L, Steen JA, Turni C, Blackall PJ, Adler B, Cox AD, Boyce JD. 2013. *Pasteurella multocida* Heddleston serovar 3 and 4 strains share a common lipopolysaccharide biosynthesis locus but display both inter- and intrastrain lipopolysaccharide heterogeneity. J Bacteriol 195:4854-64.
5. Choi KH, Mima T, Casart Y, Rholl D, Kumar A, Beacham IR, Schweizer HP. 2008. Genetic tools for select-agent-compliant manipulation of *Burkholderia pseudomallei*. Appl Environ Microbiol 74:1064-75.
6. Murray GL, Morel V, Cerqueira GM, Croda J, Srikram A, Henry R, Ko AI, Dellagostin OA, Bulach DM, Sermswan RW, Adler B, Picardeau M. 2009. Genome-wide transposon mutagenesis in pathogenic *Leptospira* species. Infect Immun 77:810-6.
7. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA, Parkhill J. 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. Bioinformatics 32:1109-11.
8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.
9. Dorman MJ, Feltwell T, Goulding DA, Parkhill J, Short FL. 2018. The capsule regulatory network of *Klebsiella pneumoniae* defined by density-TraDISort. MBio 9.

Additional work performed using TraDIS to investigate *P. multocida* strain VP161 pathogenesis, but not presented in the manuscript above, is presented below.

2.1 Introduction

Pasteurella multocida is the causative agent of a wide range of animal diseases, including fowl cholera in birds, which affects nearly all bird species (13). Fowl cholera ranges from a chronic infection with little effect on the host, to peracute and acute cases that can be fatal within 24 h of infection (25). *P. multocida* colonises the upper respiratory tract before disseminating to the lower respiratory tract. *P. multocida* can enter the bloodstream via host-immune mediated damage to the respiratory mucosa, or via direct trafficking by macrophages (13, 22). In some cases, *P. multocida* may directly enter the bloodstream through skin wounds (13). *P. multocida* strains that are resistant to clearance from the blood by the host immune system cause peracute and acute fowl cholera and colonise the liver and spleen, with bacterial loads increasing during infection (21). The resulting bacteraemia stimulates a robust host immune response that leads to haemorrhaging and necrosis in several tissues, and ultimately host death (20, 22). *P. multocida* type A and D fowl cholera isolates have been shown to resist complement-mediated killing. Of 35 *P. multocida* fowl cholera isolates examined (types A:L1, A:L3 and D:L3), 27 were shown to be resistant to complement-mediated killing by chicken serum (209). Additionally, *P. multocida* type A:L1 strains X73 and VP161 (the focus of this study) have also been shown to be resistant to complement-mediated killing, with no difference observed between growth in untreated or heat-inactivated chicken serum (76, 107).

In *P. multocida* strains belonging to capsule genotype A, the hyaluronic acid (HA) capsule is important for resistance to complement-mediated killing. An acapsular *P. multocida* strain X73 *hexA* (A:L1) mutant unable to export capsule onto the surface of the cell was sensitive to complement-mediated killing, whereas the parent strain displayed normal exponential growth in active serum (76). In addition, enzymatic removal of hyaluronic acid capsule in type A strains that were resistant to complement-mediated killing was able to induce complement sensitivity in nine of the ten strains tested (209). In contrast, the presence or absence of full length lipopolysaccharide (LPS) structure had no effect on the ability of the *P. multocida* strain VP161 (A:L1) to resist complement-mediated killing in chicken serum (107). This indicates that the LPS structure does not have a role in complement resistance for *P. multocida* capsule type A strains. However, *P. multocida* serum resistance appears to be mediated by different mechanisms in strains with different capsule types, as an acapsular *P. multocida* M1404 (B:L2) *cexA* mutant had indistinguishable growth rate compared to the wild-type parent strain in active bovine serum (75). Therefore, while capsule plays an important role in complement

resistance in type A strains, it is likely that there are other complement evasion mechanisms used by *P. multocida* that are yet to be identified.

Bacteria use a number of different mechanisms to resist complement-mediated killing, including surface structures that block complement deposition, or outer membrane proteins that block or degrade complement proteins (210, 211). Whole genome approaches such as transcriptomics and transposon insertion sequencing (TIS) have been used to assess the mechanisms utilised by different bacteria to resistant complement-mediated killing. Transcriptomic analysis of *Leptospira interrogans* serovar Copenhageni grown in active guinea pig serum identified increased expression of LPS biosynthesis genes compared to growth in rich media (212). A similar transcriptomic study investigating *Escherichia coli* strain XM grown in human serum identified that LPS biosynthesis genes, and LPS, capsule and peptidoglycan biosynthesis genes were expressed at increased levels in serum (213). TraDIS analysis of *E. coli* strain EC958 following growth in human serum identified several LPS biosynthesis genes and genes encoding membrane proteins as important for serum resistance (203). Similarly, TraDIS analysis *Acinetobacter baumannii* strain AB5075 grown in human serum identified genes encoding capsule production and transport systems, as well as genes encoding an outer membrane lipid asymmetry system that removes phospholipids from the outer leaflet of the outer membrane (214). Intra-species global analysis can be used to identify how mechanisms of serum resistance differ between strains. TraDIS analysis of four *Klebsiella pneumoniae* strains (B5055, NTUH-K2044, ATCC 43816 and RH201207) following growth in human serum identified that a different cohort of LPS biosynthesis/modification and capsule biosynthesis genes were important for resistance in the different strains, with only three genes identified as essential across all strains (215). Complement binding assays showed strains B5055 and NTUH-K2044 prevented the binding of complement proteins to surface structures, while strains ATCC 43816 and RH201207 bound complement C5b-9 but were still resistant to killing by complement, showing these strains utilised different mechanisms (215). Collectively, these studies show how whole genome methods can identify mechanisms of complement resistance in different bacterial species and strains.

In this study, we utilised TraDIS to comprehensively characterise genes required for HA capsule production and resistance to complement-mediated killing in the *P. multocida* fowl cholera isolate VP161. Genes involved in capsule production were identified by TraDIS examination of an acapsular mutant population isolated from a *P. multocida* VP161-Tn*7* mutant library using Percoll density gradients as is reported in the manuscript above. This section of chapter two reports additional work performed for the Percoll TraDIS experiments that is not included in the above manuscript, and also reports on work performed to identify genes important for *P. multocida* survival in chicken serum. These analyses identified that

capsule and aerobic ATP synthesis are important for *P. multocida* strain VP161 to survive in chicken serum. Moreover, a comparison of the two studies identified several genes that were important for both capsule production and serum resistance, including *cpdA* and PmVP161_0878. Further analysis of these genes would provide greater understanding of pathogenic mechanisms utilised by *P. multocida* and provide novel targets for therapeutics.

2.2 Results

2.2.1 Percoll density gradient isolation of acapsular *P. multocida Himar1* mutants

Discontinuous Percoll density gradients were optimised to allow isolation of acapsular *P. multocida* strain VP161-Tn*7* *Himar1* mutants from a large mutant library, as described in the manuscript above. Individual colonies were recovered from the bottom fraction of the optimised Percoll gradient that appeared to be acapsular based on colony morphology (non-mucoid, small in size, no translucent halo). The *Himar1* insertion site was identified in 35 putatively acapsular mutants using Sanger sequencing with a transposon-specific primer and purified genomic DNA as template (*Himar1* mutant strains in Table 2.4). The amount of HA capsule produced by a selection of these strains was determined by HA capsule absorbance assays and compared to the amount of HA produced by the VP161-Tn*7* parent strain that produces wild-type levels of capsule. A VP161 *hyaD* mutant that is unable to synthesise capsule was included as a negative control. All mutant strains tested produced between 1.4 and 4.7 µg of HA per mL of culture, significantly less than the VP161-Tn*7* parent strain that produced 15.7 µg of HA/mL, confirming that the *Himar1* mutants with an acapsular colony morphology produced less HA capsule than the wild-type parent strain (Figure 2.1).

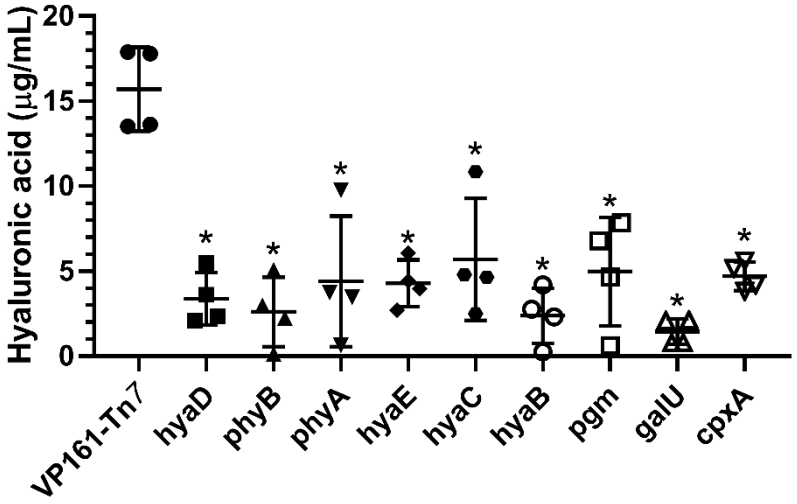


Figure 2.1. Amount of hyaluronic acid (HA) capsule produced by different *P. multocida* VP161-Tn*7* *Himar1* mutant strains. HA was measured from cultures at mid-exponential growth phase (OD_{600} 0.5-0.6). Capsule was measured for the wild-type parent strain VP161-Tn*7* and the known acapsular *hyaD* mutant strain (AL2234) as comparisons. HA was measured from each strain in biological quadruplicate, with error bars representing the mean ± standard deviation. Mutant strains were compared to VP161-Tn*7* using a Mann-Whitney U-test, with * representing $p < 0.05$.

2.2.2 TargeTron mutagenesis of genes important for capsule

TraDIS analysis of VP161-Tn*7* *Himar1* mutants recovered from the bottom fraction of Percoll density gradients identified 69 genes important for HA capsule production. To confirm that genes identified by TraDIS were truly required for capsule production, directed TargeTron mutagenesis was attempted for 14 genes not previously associated with *P. multocida* capsule production in *P. multocida* strain VP161; namely, *prmC*, *rimP*, PmVP161_0878, *ppx*, *phyA*, *cpdA*, *pgm*, *ptsH*, *atpB*, *rnb*, *epmB*, *cpxA*, *spoT* and *relA* (See modified TargeTron vectors in Table 2.4). However, TargeTron mutants were only successfully generated for *ppx*, *ptsH* and *spoT* (despite two independent attempts for other genes). To increase the number of capsule-associated genes examined, two *Himar1* transposon mutants with confirmed insertions in *pgm* and *galU* (genes identified as important for HA capsule by TraDIS) were used to confirm TraDIS accurately identified genes required for capsule production. The growth kinetics of the VP161 *ppx* and *ptsH* TargeTron mutants and VP161-Tn*7* *pgm* and *galU* *Himar1* mutants was assessed. All mutant strains displayed growth rates indistinguishable from wild-type VP161 (Figure 2.2), and there was no difference colony numbers (CFU/mL) observed for any of the cultures at an OD_{600} ~0.5 (data not shown).

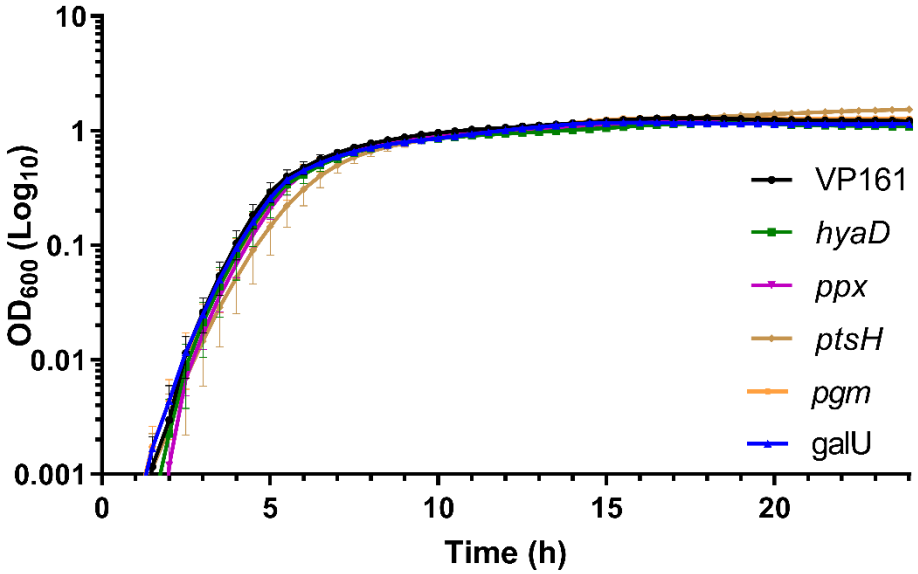


Figure 2.2: Growth kinetics of *P. multocida* VP161 and VP161 mutants in heart infusion broth. Growth of the *hyaD* (AL3574), *ppx* (AL3544) and *ptsH* (AL3793) TargeTron mutants and the *pgm* (AL3818) and *galU* (AL3819) *Himar1* mutants was assessed at 37ºC for 24 hours, shaken at 200 rpm. Data points represent the mean of three biological replicates for each strain with the error bars representing ± standard deviation.

2.2.3 Electron microscopy of putative acapsular *P. multocida* mutants

HA capsule production was investigated in the *ppx*, *ptsH* and *spoT* TargeTron and in the *pgm* and *galU Himar1* mutants using HA capsule absorbance assays. All five mutants produced significantly less HA capsule than the wild-type parent strain. Complementation with a functional copy of the disrupted gene significantly increased capsule production, confirming that inactivation of these genes caused loss of capsule (see Fig. 5 from the manuscript above). Electron microscopy was performed to confirm that these strains had less capsule on the cell surface. Transmission electron microscopy (TEM) was first used to image the cells as TEM allows for direct imaging and measurement of the capsule layer. TEM was first attempted with *P. multocida* VP161 and *hyaD*, *hexA* and *ptsH* TargeTron mutants, with cells prepared by Ruthenium Red-Osmium fixation. Both VP161, and to a lesser extent the *ptsH* TargeTron mutant, had dense aggregations at the cell surface (Figure 2.3). It was predicted these aggregates represented HA capsule that had collapsed during preparation of the cells as they were absent from the surface of *hyaD* and *hexA* mutants. TEM was repeated with lysine added to the ruthenium red fixation solution to stabilise the capsule layer. However, the electron micrographs showed the same dense aggregations present on the cell surface for VP161 and the *ptsH* TargeTron mutant, and these aggregations were again absent from the surface of the *hexA* and *hyaD* TargeTron mutants (data not shown). Given that the TEM did not allow for imaging of HA capsule, scanning electron microscopy (SEM) was performed on the same set of *P. multocida* strains using Lysine acetate-Ruthenium Red-Osmium fixation of cells and hexamethyldisilazane chemical drying. SEM micrographs showed evidence of capsule on the surface of VP161 cells as they had a ruffled texture compared to acapsular *hyaD* and *hexA* mutant cells (Figure 2.4 A). The *hexA* and *hyaD* cells had average cell widths of of 0.528 ± 0.026 µm and 0.524 ± 0.022 µm, respectively, and were all significantly narrower than VP161 parent cells that had an average cell width of 0.731 ± 0.029 µm (Figure 2.4 B). The *ptsH* cells showed an intermediate phenotype with a cell width of 0.579 ± 0.031 µm (Figure 2.4 B). However, SEM using the above method revealed some cells had collapsed and could not be included in cell-width analysis. In subsequent SEM experiments, critical point drying was used in place of hexamethyldisilazane chemical drying, which greatly reduced the number of collapsed cells.

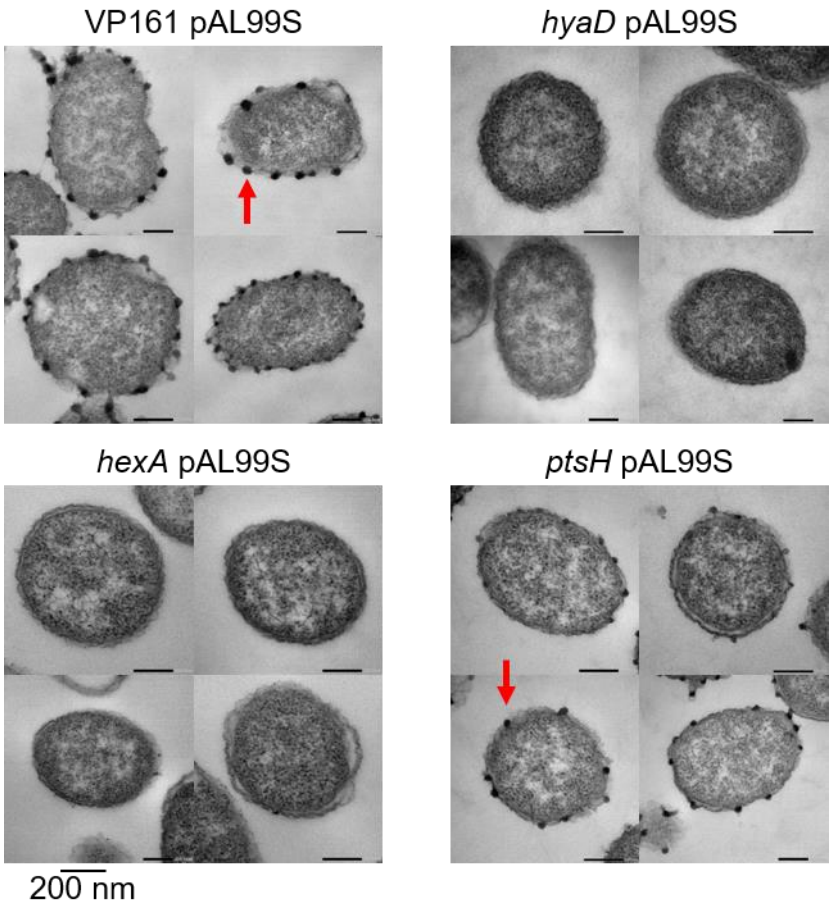


Figure 2.3: Transmission electron microscopy (TEM) of *P. multocida* strain VP161, and the *hyaD* (AL3574), *hexA* (AL3685) and *ptsH* (AL3793) TargeTron mutant strains. TEM was performed using cultures grown to mid-exponential growth phase (OD_{600} ~0.5), with cells prepared for imaging by lysine-ruthenium red-glutaraldehyde fixation as previously described in Jacques *et al* 1989. Microscopy was performed with all strains harbouring the pAL99S *P. multocida* expression (for later experiments involving complemented strains. Red arrows show aggregate material that formed at the cell surface.

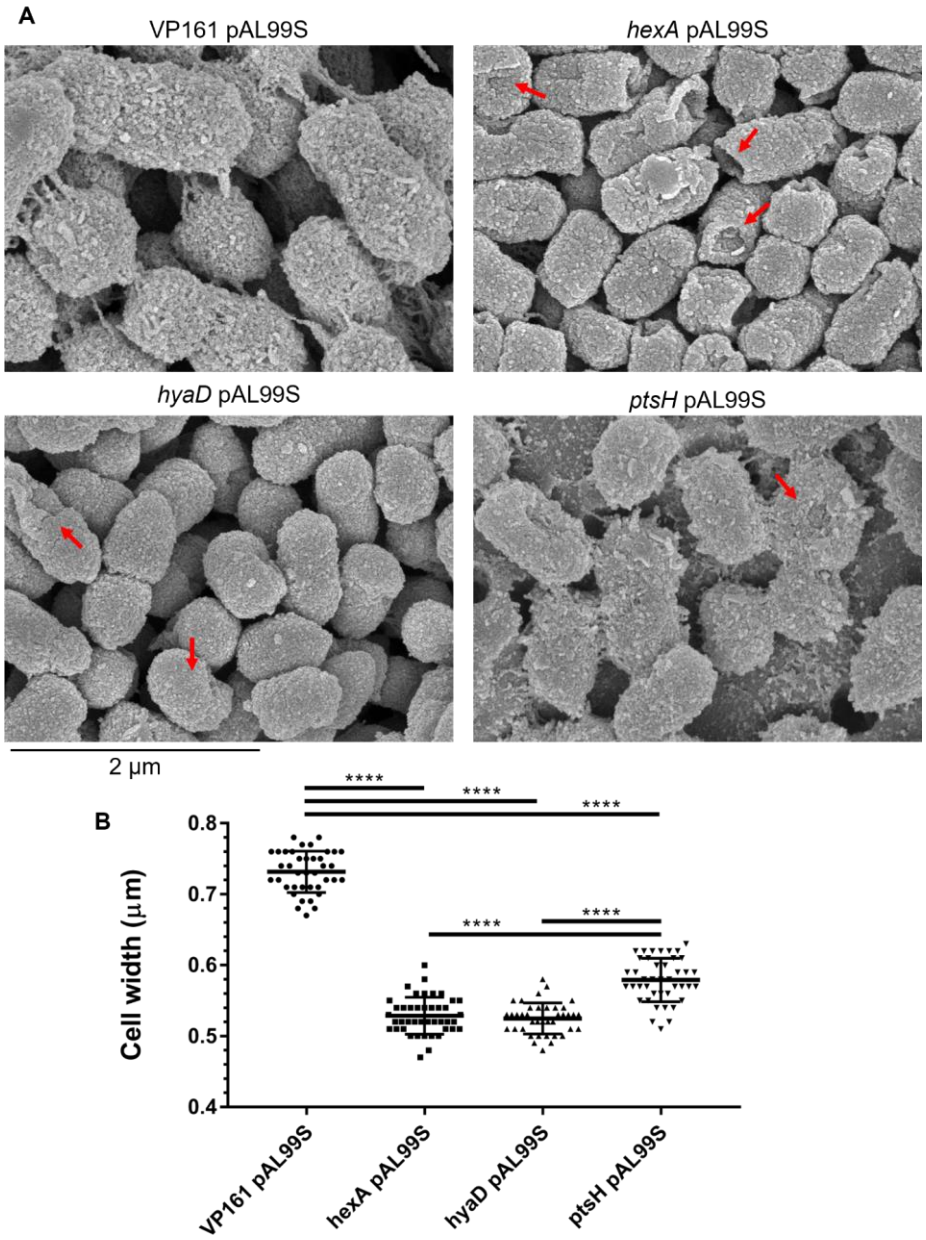


Figure 2.4: Scanning electron microscopy (SEM) and cell-width measurement of *P. multocida* strain VP161, and the *hyaD* (AL3574), *hexA* (AL3685) and *ptsH* (AL3793) TargeTron mutant strains. All strains harboured the *P. multocida* expression vector pAL99S. SEM was performed using cultures at mid-exponential growth phase (OD_{600} ~0.5), with cells prepared for imaging by lysine-ruthenium red-osmium fixation and hexamethyldisilazane chemical drying as previously described Hammerschmidt *et al* 2019. (A) Electron micrographs of the strains imaged by SEM. Red arrows show cells that have collapsed during preparation for SEM. (B) The width of a representative number of cells for each strain was measured (n=39 to 43) using the FIJI package in ImageJ. Datapoints represent individual cells with error bars representing the mean ± standard deviation. **** represents $p < 0.0001$ when strains were compared using an unpaired t-test.

2.2.4 Selection of *P. multocida Himar1* mutants resistant to complement-mediated killing in chicken serum

The TraDIS method was used to identify the full cohort of genes important for *P. multocida* strain VP161 survival in chicken serum. As complement components are heat-liable, control serum sensitivity assays were performed to ensure the chicken serum to be used in the experiments (stored at -80ºC) was active. Growth in untreated or heat-inactivated chicken serum was assessed for the *P. multocida* VP161-Tn*7* strain used to produce the *Himar1* mutant library, *E. coli* DH5α (as a negative control) and *P. multocida* with TargeTron insertions in capsule locus genes (VP161 *hexA* and *hyaD* TargeTron mutants in Table 2.4). As expected, all strains were able to grow in heat-inactivated serum as complement was inactivated (Figure 2.5). In untreated serum, *P. multocida* VP161-Tn*7* grew rapidly, whereas *E. coli* DH5α and the *P. multocida hexA* mutant failed to grow, demonstrating that complement was active and able to kill susceptible strains (Figure 2.5). Unexpectedly, despite producing no HA capsule and having no visible signs of an external capsule, the *P. multocida hyaD* mutant grew in active serum, indicating that *P. multocida* serogroup A strains may be able to resist complement-mediated killing in the absence of HA capsule by an unknown mechanism.

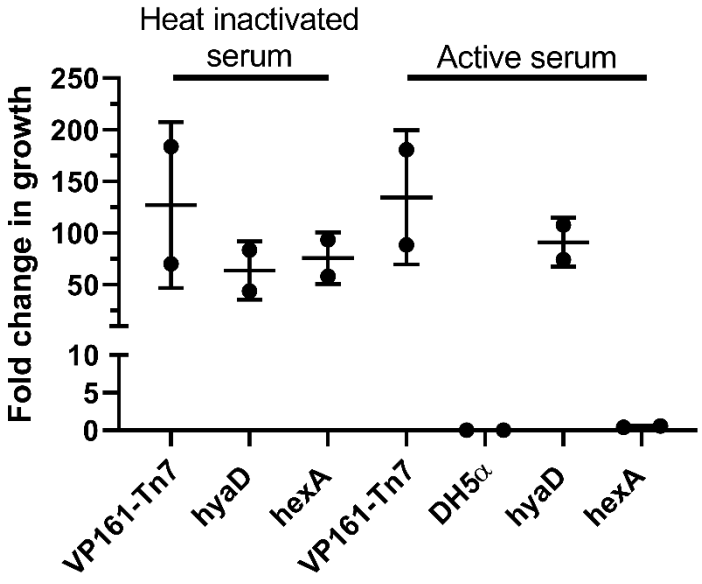


Figure 2.5: Sensitivity of *P. multocida* and *E. coli* strains to complement-mediated killing in chicken serum. *P. multocida* strain VP161-Tn*7*, VP161 *hyaD* and *hexA* TargeTron mutants and *E. coli* strain DH5α were grown to mid-exponential growth phase (OD_{600} ~0.5) before being incubated in untreated serum or heat-inactivated serum for 3 h. Fold change was calculated as (CFU of each strain recovered following 3 h incubation in active or inactive serum) / (CFU of the input culture). Datapoints show the fold change for each biological replicate (n=2) with error bars representing mean ± standard deviation.

As the above preliminary data confirmed complement activity was active in the chicken serum, and that the VP161-Tn*7* parent strain used to construct the *Himar1* library was resistant to complement-mediated killing in chicken serum, TraDIS experiments were performed to identify genes required for VP161 serum resistance. Analysis of the VP161-Tn*7* *Himar1* mutant library was performed in biological triplicate. To ensure the VP161-Tn*7* *Himar1* mutant library was fully represented, the input was increased to ~1 x 10^6 CFU, compared to previous serum sensitivity experiments that used an input of ~1 x 10^5 CFU. To ensure maximum exposure to serum components, a higher volume of serum was also used. Each aliquot of the *Himar1* library was grown in 3 mL total volume consisting of 90% untreated or heat-inactivated chicken serum, with 10% HI media containing the mutant library. As a control for complement activity, *E. coli* DH5α was grown separately under the same conditions. In the untreated chicken serum control experiments, growth of the *E. coli* was below the level of detection, indicating complement was active (Figure 2.6). The *Himar1* mutant library replicates A and B grew in active serum with fold-change growth increases of 2.4 and 14, respectively, but the growth of library replicate C was below the limit of detection (Figure 2.6). The average fold-change in growth of the *Himar1* mutant library replicates was 7-fold lower than the growth of the VP161-Tn*7* parent strain in preliminary studies using a smaller volume assay. We predict this is likely due to increased volume of serum used relative to the cell input. Following growth in active serum, surviving VP161-Tn*7* *HImar1* mutants were recovered from each replicate by centrifugation. Genomic DNA was isolated and used to produce TraDIS libraries that were sequenced on an Illumina MiSeq as described in the manuscript above.

Growth of the VP161-Tn*7* *Himar1* mutant library in heat-inactivated serum was also performed. However, the *E. coli* DH5α control grew poorly (fold-change increase of 1.1) in heat-inactivated serum, indicating that complement may not have been completely inactivated. The VP161-Tn*7* *Himar1* mutant library replicate B and C displayed relatively modest growth, with fold change increases of between 2.3 and 6.4, with library A below the limit of detection (Figure 2.6). The *Himar1* libraries displayed an average of a 4-fold increase in growth, far below the VP161-Tn*7* parent strain that had an average of a 127-fold increase in growth when grown in heat-inactivated serum (Figure 2.5, Figure 2.6). Together, these data suggest the complement activity had not completely been inactivated during heat treatment of the serum. As such, TraDIS libraries were not generated using the *Himar1* mutant libraries grown in heat-inactivated serum.

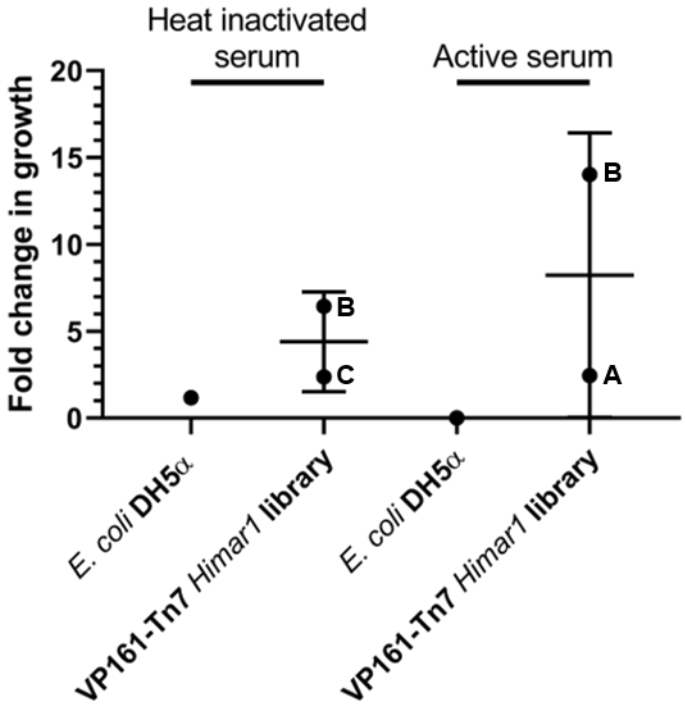


Figure 2.6: Sensitivity of the *P. multocida* VP161-Tn*7* *Himar1* mutant library and *E. coli* DH5α to complement-mediated killing in chicken serum. The *Himar1* mutant libraries and DH5α were grown separately to mid-exponential growth phase (OD_{600} ~0.5) before being incubated in serum or heat-inactivated serum for 3 h. Fold change was calculated as (CFU of each strain recovered following 3 h incubation in active or inactive serum) / (CFU of the input culture). Fold change for *E. coli* DH5α was measured for a single biological replicate. Each point of the *Himar1* mutant library represents a separate library growth. Error bars representing mean ± standard deviation.

2.2.4 TraDIS analysis of *P. multocida Himar1* mutants resistant to complement-mediated killing

The sequencing data from the untreated serum TraDIS libraries was analysed using Bio-TraDIS and modified scripts, as described in the manuscript above. A total of 4,051,495 reads were generated for the serum TraDIS libraries, with 3,373,137 reads containing a *Himar1* inverted repeat (79.6% of total reads) and 2,666,962 reads mapping to VP161 (65.8% of total reads). Comparison of unique *Himar1* insertion sites and total reads per gene showed good reproducibility between TraDIS library replicates A, B and C (Additionally, principal component analysis was performed A). Additionally, principal component analysis (PCA) was performed using both the number of UIS and total *Himar1* insertions per gene (Figure 2.7B). The three serum TraDIS library replicates clustered together away from the three rich media TraDIS library replicates (Figure 2.7B). These comparisons allowed data from library C, that grew relatively poorly in untreated serum compared to replicates A and B, to be included in downstream analysis, and for data from the three replicates to be combined for identification of essential genes. Collectively, a total of 61,429 unique *Himar1* insertion sites were identified, giving a unique *Himar1* insertion every 37 bp (Appendix 2.1). There were 523 genes identified as essential for *P. multocida* growth in chicken serum, with 477 genes encoding proteins and 46 encoding tRNAs (Figure 2.8 and Appendix 2.2). Comparative analysis of the 523 serum-essential genes identified that 52 genes were essential for *P. multocida* growth in chicken serum but not essential for growth in rich media (Table 2.1). Genes that conferred a significant fitness defect when inactivated (but were not essential) in chicken serum were identified by comparing the number of *Himar1* read counts for all genes between the chicken serum and rich media TraDIS libraries. Any gene that had a four-fold decrease in reads in the serum TraDIS libraries ($\log_2 < -2$) with a *q*-value < 0.001 was identified as important for growth in chicken serum (Figure 2.9). This analysis identified 75 genes important for growth in chicken serum, 44 of which were not essential for growth in rich media or chicken serum (Table 2.2). TargeTron mutagenesis was attempted to generate mutants in a select number of genes essential for growth in chicken serum but not for growth in rich media. The TargeTron vector pAL953 was independently retargeted to five genes (aroQ, yfeX, ahpC, znuC, and fadR); however, despite two independent attempts, no mutants were generated.

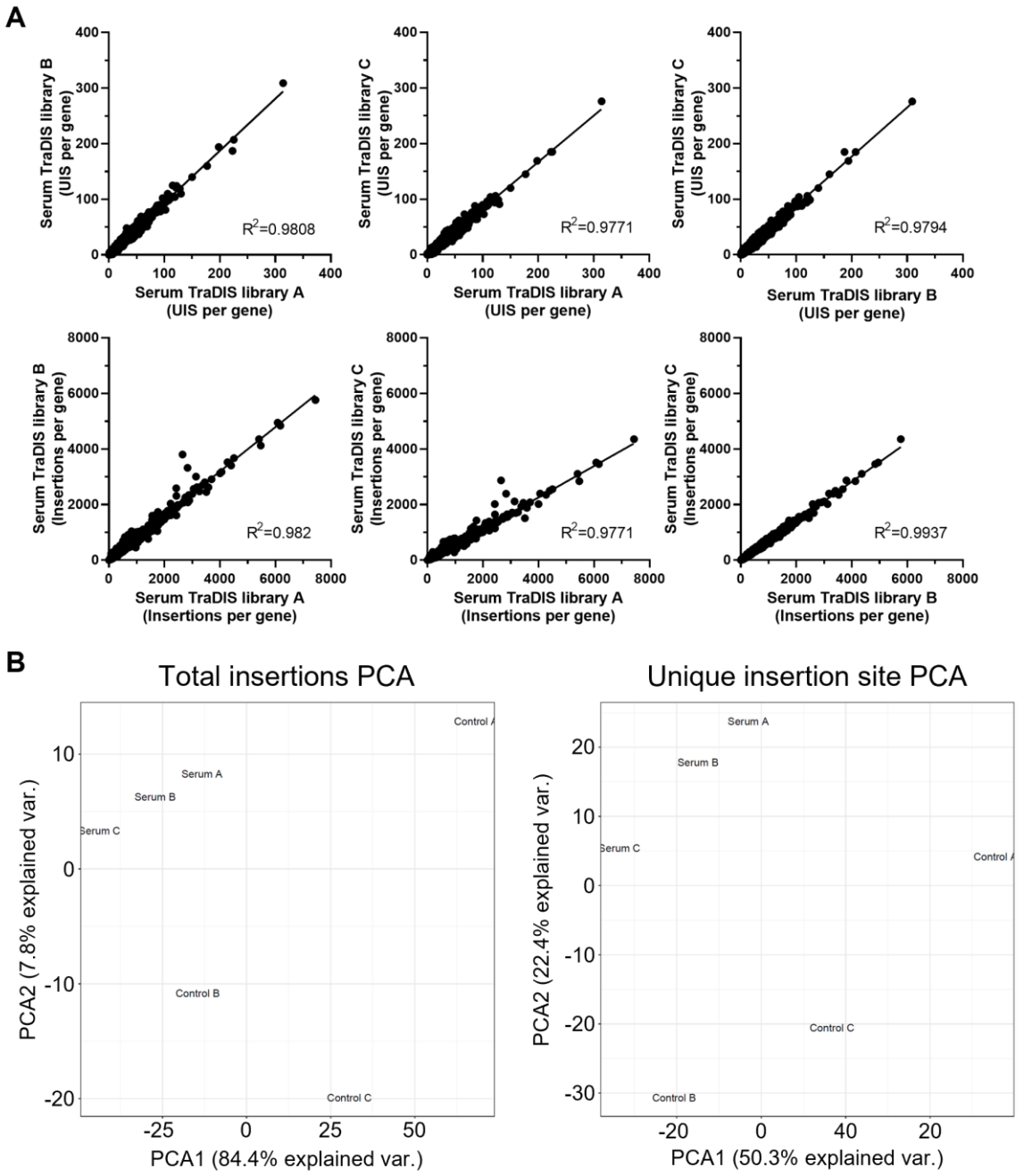


Figure 2.7: Reproducibility of the rich media input and serum TraDIS library replicates. (A) Comparison of the number of unique *Himar1* insertion sites (UIS) and total *Himar1* insertions in each gene between TraDIS library replicates generated from cells recovered following growth in chicken serum. The R^2 values were determined by simple linear regression analysis for each of the datasets. UIS: unique insertion sites. (B) Principal component analysis plots for the rich media and serum TraDIS libraries generated by comparing either UIS or total *Himar1* insertions per gene.

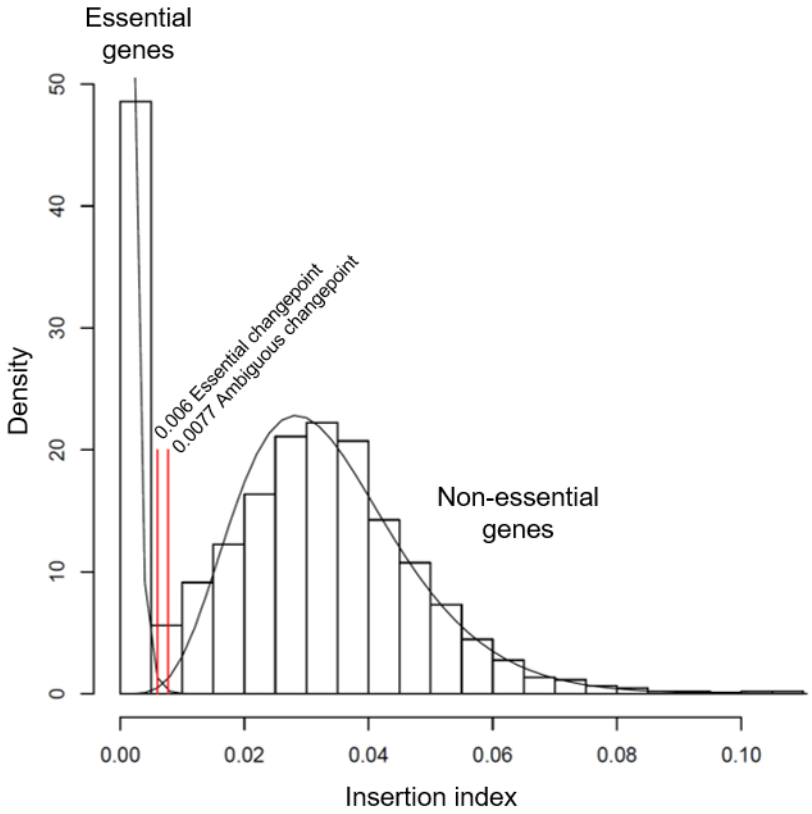


Figure 2.8: Histogram of insertion indexes for all genes following *P. multocida* strain VP161-Tn*7* *Himar1* mutant library growth in active chicken serum. Insertion indexes are calculated for all genes by dividing the number of unique *Himar1* insertion sites by gene length. Normal curves were drawn for the bimodal dataset, with the intersection between the two used as the cut-off to call a gene essential for *P. multocida* growth in chicken serum, shown by the red line closest to the y-axis.

Table 2.1: Genes identified as essential for *P. multocida* strain VP161 growth in untreated chicken serum but not essential for growth in heart infusion broth.

Gene name	PmVP161 locus tag number	Function	Insertion index
PmVP161_0041	0041	tRNA-Val	0
PmVP161_0042	0042	tRNA-Val	0
aroQ	0069	3-dehydroquinate dehydratase	0.004963
yhdT	0072	putative membrane protein YhdT	0
panF	0073	Sodium/pantothenate symporter	0.000775
tolQ	0203	Protein TolQ	0.003221
PmVP161_0208	0208	tRNA-Lys	0
PmVP161_0210	0210	tRNA-Lys	0
PmVP161_0211	0211	tRNA-Lys	0
recC	0217	RecBCD enzyme subunit RecC	0.004937
iscR	0297	HTH-type transcriptional regulator IscR	0.004773
hpf	0380	Putative metal ABC transporter substrate-binding protein Hpf	0.003778
yfeX	0437	putative deferrochelatase/peroxidase YfeX	0.001247
PmVP161_0447	0447	tRNA-Arg	0
PmVP161_0448	0448	tRNA-His	0
PmVP161_0449	0449	tRNA-Pro	0
xseB	0524	Exodeoxyribonuclease 7 small subunit	0.004505
prmC	0549	Release factor glutamine methyltransferase	0
PmVP161_0661	0661	hypothetical protein	0
rimP	0757	Ribosome maturation factor RimP	0.004866
hexD	0779	Hyaluronic acid capsule export	0.00094
hexC	0780	Hyaluronic acid capsule export	0
hexB	0781	Hyaluronic acid capsule transport protein HexB	0
hexA	0782	Hyaluronic acid capsule transport ATP-binding protein HexA	0
ahpC	0800	Alkyl hydroperoxide reductase C	0.005525
PmVP161_0878	0878	hypothetical protein	0
znuC	0985	Zinc import ATP-binding protein ZnuC	0.00561
ykgO	1009	50S ribosomal protein L36 2	0
PmVP161_1022	1022	tRNA-Leu	0
PmVP161_1127	1127	tRNA-Asp	0
PmVP161_1128	1128	tRNA-Asp	0
efp	1135	Elongation factor P	0
fadR	1189	Fatty acid metabolism regulator protein	0.003058
sodA	1237	Superoxide dismutase (Mn)	0.001721
PmVP161_1350	1350	hypothetical protein	0
purB	1383	Adenylosuccinate lyase	0.000812
hlyU	1622	Transcriptional activator HlyU	0
PmVP161_1628	1628	hypothetical protein	0.005435
atpB	1629	ATP synthase subunit a	0.001397
atpH_2	1632	ATP synthase subunit delta	0.002079
atpA	1633	ATP synthase subunit alpha	0.002161
ftsB	1754	Cell division protein FtsB	0
PmVP161_1760	1760	hypothetical protein	0
PmVP161_1762	1762	hypothetical protein	0.003165
rpe	1767	Ribulose-phosphate 3-epimerase	0.004975
PmVP161_1864	1864	tRNA-Phe	0
ubiI	1930	2-octaprenylphenol hydroxylase	0.003799
ubiH	1931	2-octaprenyl-6-methoxyphenol hydroxylase	0.002762
ubiB	1968	putative protein kinase UbiB	0.004208
rpiA	1987	Ribose-5-phosphate isomerase A	0.001684
aroB	2040	3-dehydroquinate synthase	0.004077
PmVP161_2042	2042	hypothetical protein	0

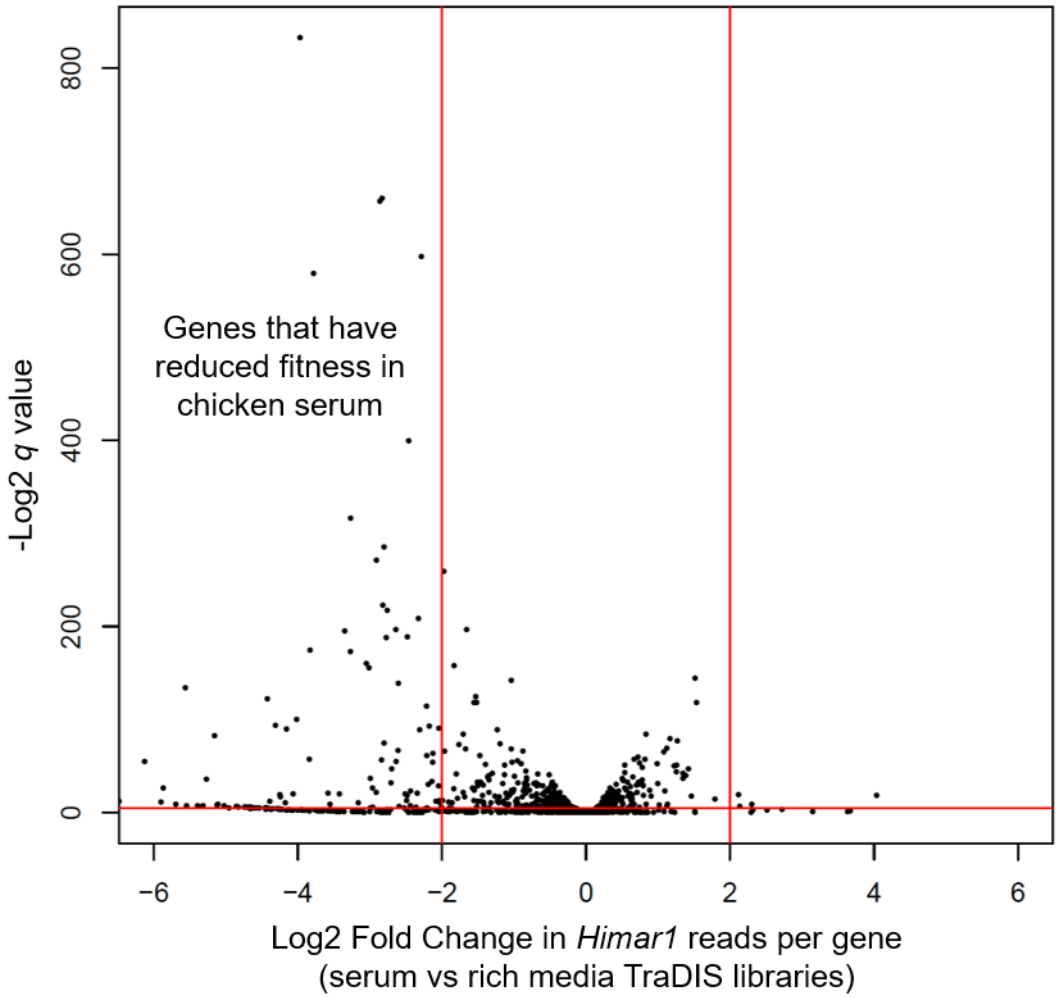


Figure 2.9: Volcano plot showing the change in total *Himar1* insertion sites between growth in active chicken serum and growth in heart infusion broth. The x-axis shows the $\log_2$ fold change of total *Himar1* insertion sites per gene for the three TraDIS libraries generated from cells recovered following growth in chicken serum, compared to TraDIS libraries generated from cells recovered following growth in heart infusion broth. The $-\log_2$ *q*-value is shown on the y-axis. Genes with a $\log_2$ fold change < -2 and a *q*-value < 0.001 were identified as having a fitness defect in untreated chicken serum and are shown on the left of the plot.

Table 2.2: Genes identified as important for fitness of *P. multocida* strain VP161 following growth in untreated chicken serum, but not essential for growth under that condition or growth in heart infusion broth.

Gene name	PmVP161 locus tag number	Function	logFC	*q* value
pgm	0089	Phosphoglucomutase	-3.3	1.89E-59
cpdA	0148	3'-5'-cyclic adenosine monophosphate phosphodiesterase CpdA	-2.6	7.90E-60
ubiF	0165	2-octaprenyl-3-methyl-6-methoxy-1-4-benzoquinol hydroxylase	-3.0	1.92E-47
PmVP161_0183	0183	hypothetical protein	-2.8	2.95E-57
ruvC	0194	Crossover junction endodeoxyribonuclease RuvC	-2.3	7.13E-07
ruvB	0196	Holliday junction ATP-dependent DNA helicase RuvB	-2.1	5.60E-17
tolB	0206	Protein TolB	-2.2	7.67E-10
pal	0207	Outer membrane protein P6	-3.8	5.18E-18
purA	0242	Adenylosuccinate synthetase	-2.9	1.94E-82
mntB_2	0378	Manganese transport system membrane protein MntB	-2.2	4.03E-19
ppnN	0465	Pyrimidine/purine nucleotide 5'-monophosphate nucleosidase	-2.1	8.74E-20
purR	0540	HTH-type transcriptional repressor PurR	-3.8	2.60E-53
aroG	0556	Phospho-2-dehydro-3-deoxyheptonate aldolase- Phe-sensitive	-2.2	1.23E-28
xthA	0683	Exodeoxyribonuclease III	-4.3	8.12E-29
ubiX	0693	Flavin prenyltransferase UbiX	-4.0	7.86E-31
phyB	0773	Capsule phospholipid substitution proteins	-3.8	3.33E-175
phyA	0774	Capsule phospholipid substitution proteins	-2.8	1.99E-199
hyaE	0775	Hyaluronic acid capsule biosynthesis protein	-3.3	5.51E-96
hyaD	0776	Hyaluronan synthase	-2.9	1.71E-198
hyaC	0777	UDP-glucose 6-dehydrogenase	-3.2	9.48E-53
hyaB	0778	Hyaluronic acid capsule biosynthesis	-3.8	2.15E-251
grxD	0783	Glutaredoxin 4	-2.5	1.23E-06
aroA	0845	3-phosphoshikimate 1-carboxyvinyltransferase	-2.7	9.25E-15
ubiG	0846	Ubiquinone biosynthesis O-methyltransferase	-5.6	5.07E-41
hldE	0891	Bifunctional protein HldE	-3.0	4.87E-49
pasT	1066	Persistence and stress-resistance toxin PasT	-2.4	9.40E-08
gmhA_2	1109	Phosphoheptose isomerase	-2.3	1.59E-27
xerD	1142	Tyrosine recombinase XerD	-2.0	2.53E-09
mrdA	1310	Peptidoglycan D-D-transpeptidase MrdA	-2.2	5.30E-35
plsX	1321	Phosphate acyltransferase	-2.3	2.08E-63
recA	1416	Protein RecA	-2.1	6.96E-11
ubiA	1472	4-hydroxybenzoate octaprenyltransferase	-4.1	8.50E-28
yqiC	1519	putative protein YqiC	-3.6	5.84E-07
gpsA	1572	Glycerol-3-phosphate dehydrogenase (NAD(P)+)	-2.8	7.09E-68
mioC	1625	Protein MioC	-4.1	1.24E-06
zwf	1692	Glucose-6-phosphate 1-dehydrogenase	-2.0	5.11E-28
asnA	1721	Aspartate--ammonia ligase	-2.3	1.41E-180
aroE	1809	Shikimate dehydrogenase (NADP(+))	-2.8	1.13E-17
galU	1828	UTP--glucose-1-phosphate uridylyltransferase	-2.8	1.23E-86
metQ	1919	putative D-methionine-binding lipoprotein MetQ	-2.5	6.21E-121
metP	1920	Methionine import system permease protein MetP	-2.7	3.78E-66
metN	1921	Methionine import ATP-binding protein MetN	-2.5	1.70E-57
xerC	1955	Tyrosine recombinase XerC	-2.7	1.97E-10
ubiE_3	1971	Ubiquinone/menaquinone biosynthesis C-methyltransferase UbiE	-4.4	1.56E-37

The general functions of the 84 protein-encoding genes that were either essential for growth in serum or important for serum fitness were determined using the KEGG Orthology And Links Annotation (BlastKOALA) tool (216). Of the genes essential or important for growth in serum, 41 were predicted to encode metabolic enzymes, 16 were predicted to encode proteins involved in transport, six were predicted to encode DNA repair and recombination proteins, three were predicted to encode transcription factors and two were predicted to encode translation proteins. All genes known to be involved in capsule biosynthesis and export were identified as essential or important for growth in serum, including *hexABCD* that were identified as essential, and *hyaBCDE*, *phyAB*, *pgm* and *galU* that were identified as important for wild-type levels of fitness in chicken serum. Genes involved in the Tol-Pal export system (*pal*, *tolB*, *tolQ*), ubiquinone biosynthesis (*ubiABE_3FGHIX*), adenosine monophosphate synthesis (AMP) (*purABR*) and genes encoding ATP synthase subunits (*atpABH_2* and PmVP161_1682 that encodes a homolog of AtpI) were also identified as important for growth in serum. The set of genes identified as essential or important for growth in serum were compared to the genes previously identified as important for hyaluronic acid capsule production (manuscript above). There were 26 genes identified as important for both phenotypes (Table 2.3) and these included genes encoding capsule biosynthesis proteins (*hexABCD*, *hyaBCDE*, *phyAB*, *pgm* and *galU*), translation proteins (*efp*, *rimP*, *prmC*), ATP synthase subunits (*atpABH_2*), central metabolism proteins Rpe and RpiA and the cyclic-AMP phosphodiesterase CpdA. CpdA encodes a cAMP phosphodiesterase, with disruption of *cpdA* likely leading to an increase in cAMP concentration that would alter expression of genes regulated by the cAMP-CRP complex. The promoter sequence of capsule biosynthesis genes and capsule gene regulators were searched for the cAMP-CRP consensus binding sequence 5'-TGTGANNNNNNTCACA-3' via BLASTn (217). The sequence 5'-TCACA-3' was identified 2 bp upstream of the previously identified promoter region for *hyaE*, but the associated palindromic sequence 5'-TGTGA-3' was not present, indicating that only a weak putative cAMP-CRP binding site was potentially in this region (92). Of interest, the gene PmVP161_0878 was identified as important for both capsule production and growth in serum. The encoded protein contains a NhaC domain characteristic of Na^+/H^+ antiporters, and bioinformatic Quick2D analysis identified transmembrane domains formed by α-helices. Together, these data indicate that PmVP161_0878 encodes an inner membrane antiporter.

Table 2.3: Genes important for hyaluronic acid capsule production and essential or important for fitness in untreated chicken serum (but not essential for growth in rich media)

Gene name	PmVP161 locus tag number	Function	Essential in serum	Fitness cost in serum
pgm	0089	Phosphoglucomutase	FALSE	TRUE
cpdA	0148	3'-5'-cyclic adenosine monophosphate phosphodiesterase CpdA	FALSE	TRUE
prmC	0549	Release factor glutamine methyltransferase	TRUE	TRUE
ubiX	0693	Flavin prenyltransferase UbiX	FALSE	TRUE
rimP	0757	Ribosome maturation factor RimP	TRUE	TRUE
phyB	0773	Capsule phospholipid substitution proteins	FALSE	TRUE
phyA	0774	Capsule phospholipid substitution proteins	FALSE	TRUE
hyaE	0775	Hyaluronic acid capsule biosynthesis protein	FALSE	TRUE
hyaD	0776	Hyaluronan synthase	FALSE	TRUE
hyaC	0777	UDP-glucose 6-dehydrogenase	FALSE	TRUE
hyaB	0778	Hyaluronic acid capsule biosynthesis	FALSE	TRUE
hexD	0779	Hyaluronic acid capsule export	TRUE	TRUE
hexC	0780	Hyaluronic acid capsule export	TRUE	TRUE
hexB	0781	Hyaluronic acid capsule transport protein HexB	TRUE	TRUE
hexA	0782	Hyaluronic acid capsule transport ATP-binding protein HexA	TRUE	TRUE
PmVP161_0878	0878	hypothetical protein	TRUE	TRUE
efp	1135	Elongation factor P	TRUE	FALSE
purB	1383	Adenylosuccinate lyase	TRUE	FALSE
mioC	1625	Protein MioC	FALSE	TRUE
PmVP161_1628	1628	hypothetical protein	TRUE	FALSE
atpB	1629	ATP synthase subunit a	TRUE	FALSE
atpH_2	1632	ATP synthase subunit delta	TRUE	FALSE
atpA	1633	ATP synthase subunit alpha	TRUE	TRUE
rpe	1767	Ribulose-phosphate 3-epimerase	TRUE	FALSE
galU	1828	UTP--glucose-1-phosphate uridylyltransferase	FALSE	TRUE
rpiA	1987	Ribose-5-phosphate isomerase A	TRUE	TRUE

2.3 Discussion

TraDIS was successfully used to identify all genes required for HA capsule production in *P. multocida* strain VP161. Acapsular VP161-Tn*7* *Himar1* mutants were isolated by discontinuous Percoll gradients. The majority of mutants collected from the bottom fraction of the gradient generated small, non-mucoid colonies when plated onto solid media, suggesting these mutants were acapsular. In initial testing, individual putatively acapsular *Himar1* mutants were recovered and *Himar1* insertion sites identified by Sanger sequencing. Many of these mutants had transposon insertions in genes within the capsule biosynthesis locus *hyaBCE* and *phyAB*), while others had insertions in genes not previously known to be associated with capsule production in *P. multocida*, including in *pgm*, *galU* and *cpxA*. To confirm these recovered mutants were truly acapsular, HA capsule assays were performed to assess HA capsule production. All mutant strains tested had reduced capsule production compared to the wild-type parent strain, confirming cells fractionated into the bottom layer of the Percoll gradients were acapsular (Figure 2.1).

TraDIS analysis of the fractionated *Himar1* mutant library identified 69 genes important for capsule production. To confirm these genes, TargeTron mutagenesis was attempted for 14 genes not previously associated with capsule production. However, TargeTron mutants were only successfully generated for *ppx*, *ptsH* and *spoT*. Given this, two *Himar1* mutants that had insertions in genes identified as important for HA production by TraDIS, and that were confirmed by Sanger sequencing to have single insertions in *pgm* and *galU*, were included in further phenotypic analyses. Growth curves were performed for the *ppx* and *ptsH* TargeTron mutants and *pgm* and *galU Himar1* mutants, showing the same growth rate and CFU/mL at mid-exponential growth phase as VP161 (Figure 2.3). Each mutant was provided with a wild-type copy of the appropriate gene on *P. multocida* expression vector pAL99S or pAL99T. HA capsule assays showed that all five mutants had reduced capsule compared to the parent strain, and importantly complementation restored HA capsule production.

To determine the amount of capsule on the cell surface, mutants were imaged using electron microscopy. Initially, TEM was performed as previously described (76, 218), and repeated with the addition of lysine, but clear evidence of capsule could not be obtained. However, there were dense aggregates observed on the cell surface of wild-type VP161 cells, and to a lesser extent on the *ptsH* mutant cells that likely represented capsule that had collapsed during preparation of the cells (Figure 2.3). These aggregates were absent in the *hyaD* and *hexA* mutant strains, suggesting a lack of capsule on the surface of these cells (Figure 2.4). To improve capsule imaging, SEM was used to image the same strains. The *hexA* and *hyaD* mutants lacked the surface structures observable in wild-type VP161 and the *ptsH* mutant (Figure 2.4 A), and the *hexA* and *hyaD* cells all had significantly reduced cell width compared

to VP161. Together this indicated the *hexA* and *hyaD* mutants lacked measurable capsule, as expected given HexA is required for capsule export and HyaD is required for HA synthesis (76, 83). The *ptsH* mutant also had significantly reduced width compared to VP161, but also had significantly wider cells compared to the *hexA* and *hyaD* mutants. In addition, the HA capsule absorbance assays suggested that the *ptsH* mutant produced more HA (although not statistically significant) than the *hexA* and *hyaD* mutants (see Fig. 5 in the manuscript above). Together, these data suggest that while *ptsH* is involved in capsule production, loss of this gene does not completely abolish capsule production.

TraDIS was also used to investigate genes important for *P. multocida* survival in chicken serum. Previous serum sensitivity assays performed with *P. multocida* used an input of ~1 x 10^5 CFU (76). We increased the input to ~1 x 10^6 CFU to ensure the *Himar1* mutant library was fully represented, and the volume of the assay was also increased to accommodate this increased inoculum. Control assays in untreated and inactive serum were first performed to show that the VP161-Tn*7* mutant library parent strain was resistant to complement-mediated killing by chicken serum, and that known capsule mutants were sensitive. Both *E. coli* DH5α and the *P. multocida* VP161 *hexA* mutant were sensitive to complement-mediated killing; *E. coli* DH5α was not viable following growth in active serum and the *P. multocida hexA* mutant had a 2-fold reduction in viability. The VP161-Tn*7* grew equally well in untreated or heat-inactivated chicken serum, confirming the mutant library parent strain was resistant to chicken complement-mediated killing (Figure 2.5). Interestingly, the VP161 *hyaD* mutant also grew equally well in untreated or heat-inactivated chicken serum. Serum sensitivity assays previously performed on this mutant also showed it was resistant to complement-mediated killing (fold-change increase of growth of ~35 when using an input of 5 x 10^5 CFU) (Mégroz pers. comm.). Together, these assays suggest *hyaD* mutants are not killed in serum despite a lack of capsule. It is known that *P. multocida* type A strains have a functional heparin synthase encoded on a gene outside the capsule biosynthesis locus (90). Therefore, it is possible that VP161 *hyaD* mutants produce a low level of heparin capsule that is not detected by imaging but could confer resistance to complement-mediated killing. Any alternative capsule would likely be exported through the same HexABCD capsule ABC-transport system. Disruption of *hexA* would lead to an inability to export both HA and heparin capsule.

Serum sensitivity assays were performed with the VP161-Tn*7* *Himar1* mutant library, with the volume of the assay increased to 3 mL to ensure all sensitive cells were exposed to complement components in the serum. Both the *E. coli* DH5α and the mutant libraries grown in heat-inactivated serum showed only very modest increases in growth, indicating that the chicken serum used for this assay had not been successfully inactivated. As such TraDIS libraries were generated for mutant libraries grown in heat-inactivated chicken serum (Figure

2.5). All *Himar1* libraries grown in untreated serum had lower fold-change increase in growth compared to that previously seen for the VP161-Tn*7* parent strain (Figure 2.5 and Figure 2.6), likely due to the increase in serum volume used for the assay. However, the *E. coli* DH5α control was completely killed by untreated chicken serum, showing serum had active complement activity (Figure 2.5). Surviving *Himar1* mutants were recovered from growth in untreated chicken serum and used to generate TraDIS libraries, with the libraries sequenced via Illumina sequencing and analysed using Bio-TraDIS and modified scripts.

A total of 61,429 *Himar1* insertion sites were identified in the serum TraDIS libraries, with an insertion every 37 bp. Comparison of unique *Himar1* insertion sites and total insertions per gene by both PCA and simple linear regression analysis showed good reproducibility between the serum TraDIS libraries, allowing all libraries to be combined for downstream analysis (Figure 2.7). Analysis of the data identified 523 genes that were putatively essential and 75 genes that were important for wild-type levels of fitness in untreated chicken serum (Appendix 2.1 and Appendix 2.2). When genes essential for growth in rich media were removed, there were 52 genes putatively essential and 44 genes important for fitness for *P. multocida* growth in serum but not essential for *P. multocida* growth in rich media (Table 1 and Table 2). Genes identified by TraDIS are only putatively essential and need to be confirmed; attempts were made to generate TargeTron mutants for five of the genes identified as essential for growth in active serum but these attempts were unsuccessful. General functions of proteins encoded by genes important for survival of *P. multocida* in serum were identified by comparison with proteins in the KEGG database. Most proteins identified as essential or important for growth in chicken serum were predicted to be involved in metabolic pathways or transport, including capsule biosynthesis, the Tol-Pal export system and several pathways involved in ATP synthesis. All genes encoding proteins involved in capsule biosynthesis, capsule export and UDP-glucuronic acid production were identified by TraDIS as essential or important for fitness, confirming previous research that capsule is crucially important for serum resistance displayed by *P. multocida* type A strains (75, 76). Identification of *hyaD* as important for wild-type levels of fitness did not agree with the results from growth of the VP161 *hyaD* mutant individually in 90% serum; the results likely differ due to the increased volume of 90% serum used to grow the VP161-Tn*7* *Himar1* mutant libraries resulting in increased killing by complement compared to the initial assays performed with individual strains. No other genes involved in the biosynthesis of other surface structures such as LPS were identified as important for serum resistance. This supports previous data showing a VP161 mutant with severely truncated LPS was still fully resistant to complement-mediated killing (107). Although capsule was identified as important for *P. multocida* VP161 growth in serum, several genes shown to be involved in capsule production in VP161 were not identified as important for growth in serum, including

fis, *hfq*, *ptsH*, *ppx* and *spoT*. Fis, Hfq and SpoT are all involved in the positive regulation of capsule biosynthesis genes. Separate inactivation of *fis*, *hfq* and *spoT* resulted in reduced capsule biosynthesis gene expression and HA capsule production, with inactivation of *ptsH* and *ppx* predicted to also reduce capsule gene expression (92, 96 and Fig. 7 in the manuscript abovet). It is possible that mutations in these genes still allow for a small amount of HA capsule to be produced given the capsule biosynthesis locus is still intact, or possibly an alternative heparin capsule, that may be sufficient to confer resistance to complement-mediated killing

The precise roles of other systems identified as important for serum resistance in *P. multocida* is less clear. Three genes encoding proteins in the Tol-Pal export system were important for *P. multocida* serum resistance. Several genes encoding Tol-Pal system components are important for serum resistance in other bacterial species, including *K. pneumoniae* NTUH-K2044 and ATCC 43816, and *E. coli* EC958 (203, 215). The Tol-Pal system has several roles, including stabilisation of the outer membrane in Gram negative bacteria (219, 220). Destabilization of the *P. multocida* outer membrane may increase serum sensitivity by increasing deposition of complement components onto the cell surface. Several genes involved in aerobic ATP biosynthesis were also identified as important for *P. multocida* chicken complement resistance, including genes encoding the ATP synthase subunits α, δ, a and c, and genes encoding ubiquinone biosynthesis proteins. In *E. coli*, deletion of the entire ATP synthase operon, or mutagenesis of individual amino acids in ATP synthase subunits disrupted aerobic ATP synthesis (221, 222), with mutagenesis of ATP synthase subunit genes in *P. multocida* also likely to stop aerobic respiration. Ubiquinone is an electron carrier in the electron transport chain and is required for aerobic respiration (223). Studies in *E. coli* have demonstrated that synthesis of ubiquinone occurs in a metabolon, with deletion of *ubiA* and *ubiG* reducing aerobic respiration and deletion of *ubiJ* significantly reducing ubiquinone production (223-225). Disrupting genes in either of these pathways would likely stop aerobic respiration, pushing the cell into anaerobic respiration or fermentation for ATP production. This suggests that ATP synthesis by aerobic respiration is important for *P. multocida* resistance to chicken complement-mediated killing.

Comparison of the Percoll and chicken serum TraDIS data identified 26 genes important for both conditions, including all capsule biosynthesis genes, *cpdA* and PmVP161_0878. The *cpdA* gene encodes a predicted cylic-AMP (cAMP) phosphodiesterase that catabolises the regulatory molecule cAMP (226). cAMP binds to CRP, forming a regulatory complex that causes global expression changes in response to changes in available nutrient sources (217, 227). Mutagenesis of *cpdA* in *E. coli* leads to increased cellular cAMP concentration, as mutants are unable to hydrolyse cAMP (226, 228). The cAMP-CRP complex has been shown to regulate several virulence factors, including biofilm in *A. baumannii* and capsule in

K. pneumoniae (228-230). In *K. pneumoniae*, the cAMP-CRP complex binds to capsule biosynthesis gene promotors downregulating their expression (229). Mutagenesis of *cpdA* resulted in decreased capsule gene expression due to the increase in cAMP concentration (229). The identification of *cpdA* as important for *P. multocida* HA capsule production suggests cAMP-CRP downregulates capsule biosynthesis genes in *P. multocida*. The downregulation of capsule biosynthesis genes would likely result in reduced capsule on the surface of *cpdA* mutants and could possibly increase susceptibility to complement-mediated killing. A short putative cAMP-CRP binding site was identified near the *hyaE* promotor, although this sequence only represented the 5' end of the consensus cAMP-CRP binding sequence.

The hypothetical protein PmVP161_0878 contains a NhaC domain that is found in sodium proton antiporters and antibiotic transporters (231), with *in silico* analysis identifying putative transmembrane domains, indicative of a transporter. The gene encoding PmVP161_0878 had a 675-fold decrease in *Himar1* insertions in cells recovered form untreated serum compared to cells recovered from rich media, and a 3,900-fold increase in *Himar1* insertions in cells recovered from the bottom fraction of the Percoll gradients compared to the top fraction, collectively showing PmVP161_0878 is critical for capsule production and serum resistance. Na^+/H^+ antiporters have been shown to be important for virulence in other bacterial species. A *Yersinia pestis* strain with mutations in Na^+/H^+ antiporter genes *nhaA* and *nhaB* had attenuated virulence in mice, as well as reduced survival in sheep blood and serum, with complementation of one of the antiporters restoring survival in blood (232). It is possible that *P. multocida* PmVP161_0878 may be required for salt tolerance in chicken serum. However, *P. multocida* has homologs of Na^+/H^+ antiporters NhaA and NhaB that would likely provide a level of functional redundancy. Further work is required to elucidate the precise role of PmVP161_0878 in capsule production and serum resistance.

2.4 Conclusion

We utilised TraDIS to investigate genes important for *P. multocida* strain VP161 hyaluronic acid capsule production, and for resistance to complement-mediated killing in chicken serum. Percoll density gradient centrifugation allowed separation of acapsular VP161-Tn*7* *Himar1* mutants from the mutant library, with cells in the bottom fraction confirmed to have reduced capsule by HA absorbance assays. A select number of novel capsule-associated genes identified by TraDIS were confirmed to have reduced capsule by HA capsule absorbance assays and SEM. TraDIS analysis of *P. multocida* mutants that survived growth in active chicken serum identified 52 essential genes and 44 genes important for fitness in chicken serum. Pathways identified as important for *P. multocida* growth in serum included capsule biosynthesis genes, but not regulators of capsule biosynthesis, the Tol-Pal export system and anaerobic respiration. Comparison of the Percoll and chicken serum TraDIS data identified 26 genes important for both capsule biosynthesis and survival in chicken serum, including capsule biosynthesis locus genes, *cpdA* encoding the cAMP phosphodiesterase and the PmVP161_0878 encoding a putative Na^+/H^+ antiporter.

2.5 Materials and methods

The materials and methods given in the manuscript above are applicable to work in this section. However, experiments performed in this section but not covered by the materials and methods in the above manuscript are described below.

2.5.1 Bacterial strains, plasmids and oligonucleotides

Several strains and oligonucleotides used for this work are listed in Table S1 and Table S2 in the manuscript above. Additional strains and oligonucleotides used are listed in Table 2.4 and Table 2.5, respectively.

Table 2.4: Strains and plasmids used in this study.

Strain or plasmid	Description	Source or reference
Strains		
P. multocida		
AL2234	VP161 *hyaD* TargeTron mutant; Kan^R	(96)
AL3821	VP161-Tn*7* *Himar1* mutant, *Himar1* inserted in *phyB* inserted between nucleotides 731-732	This study
AL3822	VP161-Tn*7* *Himar1* mutant, *Himar1* inserted in *phyA* between nucleotides 732-333	This study
AL3823	VP161-Tn*7* *Himar1* mutant, *Himar1* inserted in *hyaE* between nucleotides 795-796	This study
AL3824	VP161-Tn*7* *Himar1* mutant, *Himar1* inserted in *hyaC* between nucleotides 450-451	This study
AL3825	VP161-Tn*7* *Himar1* mutant, *Himar1* inserted in *hyaB* between nucleotides 590-591	This study
AL3818	VP161-Tn*7* *Himar1* mutant, *Himar1* inserted in *pgm* between nucleotides 119-120	This study
AL3819	VP161-Tn*7* *Himar1* mutant, *Himar1* inserted in *galU* between nucleotides 475-476	This study
AL3820	VP161-Tn*7* *Himar1* mutant, *Himar1* inserted in *cpxA* between nucleotides 7766-767	This study
E. coli		
AL3699	DH5α harbouring pAL1586	This study
AL3700	DH5α harbouring pAL1587	This study
AL3701	DH5α harbouring pAL1588	This study
AL3702	DH5α harbouring pAL1589	This study
AL3703	DH5α harbouring pAL1590	This study
AL3705	DH5α harbouring pAL1592	This study
AL3706	DH5α harbouring pAL1593	This study
AL3707	DH5α harbouring pAL1594	This study
AL3747	DH5α harbouring pAL1596	This study
AL3748	DH5α harbouring pAL1597	This study
AL3749	DH5α harbouring pAL1598	This study
AL3751	DH5α harbouring pAL1601	This study
AL3752	DH5α harbouring pAL1602	This study
AL3753	DH5α harbouring pAL1603	This study
AL3754	DH5α harbouring pAL1604	This study
Plasmids		
pAL1586	pAL953 with modified group II intron retargeted to *aroQ*; $Spec^R$ Kan^R	This study
pAL1587	pAL953 with modified group II intron retargeted to *yfeX*; $Spec^R$ Kan^R	This study
pAL1588	pAL953 with modified group II intron retargeted to *prmC*; $Spec^R$ Kan^R	This study
pAL1589	pAL953 with modified group II intron retargeted to *rimP*; $Spec^R$ Kan^R	This study
pAL1590	pAL953 with modified group II intron retargeted to *ahpC*; $Spec^R$ Kan^R	This study
pAL1592	pAL953 with modified group II intron retargeted to PmVP161_0878; $Spec^R$ Kan^R	This study
pAL1593	pAL953 with modified group II intron retargeted to *znuC*; $Spec^R$ Kan^R	This study
pAL1594	pAL953 with modified group II intron retargeted to *fadR*; $Spec^R$ Kan^R	This study
pAL1596	pAL953 with modified group II intron retargeted to *phyA*; $Spec^R$ Kan^R	This study
pAL1597	pAL953 with modified group II intron retargeted to *cpdA*; $Spec^R$ Kan^R	This study
pAL1598	pAL953 with modified group II intron retargeted to *pgm*; $Spec^R$ Kan^R	This study
pAL1601	pAL953 with modified group II intron retargeted to *atpB*; $Spec^R$ Kan^R	This study
pAL1602	pAL953 with modified group II intron retargeted to *rnb*; $Spec^R$ Kan^R	This study
pAL1603	pAL953 with modified group II intron retargeted to *epmB*; $Spec^R$ Kan^R	This study
pAL1604	pAL953 with modified group II intron retargeted to *cpxA*; $Spec^R$ Kan^R	This study

Table 2.5. Oligonucleotides used in this study

Oligonucleotide	Sequence (5'-3')	Description
BAP8875	AAAAAAGCTTATAATTATCCTTAATTCCCTTTGTAGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *aroQ*
BAP8876	CAGATTGTACAAATGTGGTGATAACAGATAAGTCTTTGTAGATAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *aroQ*
BAP8877	TGAACGCAAGTTTCTAATTTCGGTTGGAATCCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *aroQ*
BAP8878	AAAAAAGCTTATAATTATCCTTAGTTGTCATTGTGGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *yfeX*
BAP8879	CAGATTGTACAAATGTGGTGATAACAGATAAGTCATTGTGCTTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *yfeX*
BAP8880	TGAACGCAAGTTTCTAATTTCGATTACAACTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *yfeX*
BAP8881	AAAAAAGCTTATAATTATCCTTAAATGCCATTGCCGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *prmC*
BAP8882	CAGATTGTACAAATGTGGTGATAACAGATAAGTCATTGCCCCTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *prmC*
BAP8883	TGAACGCAAGTTTCTAATTTCGGTTGCATTCCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *prmC*
BAP8884	AAAAAAGCTTATAATTATCCTTAATGTTCACCCTGGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *rimP*
BAP8885	CAGATTGTACAAATGTGGTGATAACAGATAAGTCACCCTGATTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *rimP*
BAP8886	TGAACGCAAGTTTCTAATTTCGGTTAACATCCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *rimP*
BAP8887	AAAAAAGCTTATAATTATCCTTAGGCGTCGTACGTGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ahpC*
BAP8888	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTACGTCATAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ahpC*
BAP8889	TGAACGCAAGTTTCTAATTTCGATTACGCCTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *ahpC*
BAP8893	AAAAAAGCTTATAATTATCCTTACATGTCCTCTATGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to PmVP161_0878
BAP8894	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTCTATTTTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 PmVP161_0878
BAP8895	TGAACGCAAGTTTCTAATTTCGGTTACATGTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 PmVP161_0878
BAP8896	AAAAAAGCTTATAATTATCCTTAACGCACGGCGTCGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *znuC*
BAP8897	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGGCGTCGATAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *znuC*
BAP8898	TGAACGCAAGTTTCTAATTTCGATTTGCGTTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *znuC*
BAP8899	AAAAAAGCTTATAATTATCCTTAGAGTTCGCAGAGGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *fadR*
BAP8900	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGCAGAGAATAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *fadR*
BAP8901	TGAACGCAAGTTTCTAATTTCGATTAACTCTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *fadR*
BAP8948	AAAAAAGCTTATAATTATCCTTATATGGCGATATGGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *phyA*

BAP8949	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGATATGGCTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *phyA*
BAP8950	TGAACGCAAGTTTCTAATTTCGGTTCCATATCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *phyA*
BAP8951	AAAAAAGCTTATAATTATCCTTATGTGTCATTCAAGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *cpdA*
BAP8952	CAGATTGTACAAATGTGGTGATAACAGATAAGTCATTCAAGCTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *cpdA*
BAP8953	TGAACGCAAGTTTCTAATTTCGGTTACACATCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *cpdA*
BAP8954	AAAAAAGCTTATAATTATCCTTAGTCGTCGATGCCGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *pgm*
BAP8955	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGATGCCTATAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *pgm*
BAP8956	TGAACGCAAGTTTCTAATTTCGATTACGACTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *pgm*
BAP8963	AAAAAAGCTTATAATTATCCTTATATACCCTTCATGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *atpB*
BAP8964	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTTCATCCTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *atpB*
BAP8965	TGAACGCAAGTTTCTAATTTCGGTTAAAATCCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *atpB*
BAP8966	AAAAAAGCTTATAATTATCCTTACTTTACCTTGAGGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *rnb*
BAP8967	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTTGAGCCTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *rnb*
BAP8968	TGAACGCAAGTTTCTAATTTCGGTTTAAAGTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *rnb*
BAP8969	AAAAAAGCTTATAATTATCCTTAGCTGTCGTGCCAGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *epmB*
BAP8970	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTGCCAAGTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *epmB*
BAP8971	TGAACGCAAGTTTCTAATTTCGATTACAGCTCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *epmB*
BAP8972	AAAAAAGCTTATAATTATCCTTAATAGCCCTTTCTGTGCGCCCAGATAGGGTG	IBS TargeTron oligonucleotide for retargeting group II intron in pAL953 to *cpxA*
BAP8973	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTTTCTCGTAACTTACCTTTCTTTGT	EBS1d TargeTron oligonucleotide for retargeting group II intron in pAL953 to *cpxA*
BAP8974	TGAACGCAAGTTTCTAATTTCGGTTGCTATCCGATAGAGGAAAGTGTCT	EBS2 TargeTron oligonucleotide for retargeting group II intron in pAL953 to *cpxA*
BAP8990	CAACTTGAGCTCAGGAAGGCATTATGTCGCAAC	Forward flanking *aroQ* oligonucleotide
BAP8991	TTTAATGTCGACCTATTTTTTCAGGAAACTGACCG	Reverse flanking *aroQ* oligonucleotide
BAP8992	TTGATTTCTAGAAGGAGGAACAAAATGACAGCGCAAAGTG	Forward flanking *yfeX* oligonucleotide
BAP8993	TGAATAGTCGACTTAGAGAGTTTTTAATTTCTCTAA	Reverse flanking *yfeX* oligonucleotide
BAP8994	TAAACAGAGCTCAGGAGGTAAATAATGACGTATCAAGAATG	Forward flanking *prmC* oligonucleotide
BAP8995	AGTTCAGTCGACTCACCTCCAACAACCTAAGG	Reverse flanking *prmC* oligonucleotide
BAP8996	ATCCTATCTAGAAGAGAGGTAATTTTGGCAACCTTA	Forward flanking *rimP* oligonucleotide
BAP8997	TGTTTGGTCGACTTAGAATTTAGGAATAATATTGGC	Reverse flanking *rimP* oligonucleotide
BAP8998	TTAATTGAGCTCAGGAGATTAATACTATGGTATTAG	Forward flanking *ahpC* oligonucleotide
BAP8999	TAACTGGTCGACTTACAGTTGATCAGCATGCT	Reverse flanking *ahpC* oligonucleotide
BAP9002	TCCTTCGAGCTCAGGAGGAAAACTTATGGAACTTATTGATT	Forward flanking PmVP161_0878 oligonucleotide
BAP9003	TTTTTAGTCGACTTGTGAACTATCTCACAATTTA	Reverse flanking PmVP161_0878 oligonucleotide
BAP9004	TTAAGCGAGCTCAGGAGGCAAATAATGCACATTCATTCTAT	Forward flanking *znuC* oligonucleotide

BAP9005	ACATGAGTCGACCTATTCTTTATGAGTACATTGT	Reverse flanking *znuC* oligonucleotide
BAP9064	ATTTAAGAGCTCAGGAGGAACGTATGAATAATGATCAA	Forward flanking *fadR* oligonucleotide
BAP9007	GTTTAAGTCGACCTATTCATTAAAATTAGCTGGT	Reverse flanking *fadR* oligonucleotide
BAP9065	AACTAGGAGCTCAGGAGGTATGATATGTCAACTTTATATGA	Forward flanking *phyA* oligonucleotide
BAP9066	CATATTGTCGACCTACCTCAAAACAGAATAAAC	Reverse flanking *phyA* oligonucleotide
BAP9067	TATTTTGAGCTCAGGAGTTCGCTTTGTTCAGC	Forward flanking *cpdA* oligonucleotide
BAP9068	GGCTGTGTCGACTTAATAGCCATCTTCATGCA	Reverse flanking *cpdA* oligonucleotide
BAP9162	TTAATATCTAGAAGGAGAATTTATGTCAATCTTT	Forward flanking *pgm* oligonucleotide
BAP9070	GCTTAAGTCGACTTAGCAATCCTGCTTACCAA	Reverse flanking *pgm* oligonucleotide
BAP9073	AAATATGAGCTCAGGAGGAGTTATGGCAGCAGAGCTTA	Forward flanking *atpB* oligonucleotide
BAP9074	GTTTATGTCGACTTAATGTTCTGCTTTGTTATAAC	Reverse flanking *atpB* oligonucleotide
BAP9161	TTTCATTCTAGAAGGAGGGCCTCAATGTTCCAAAATAAT	Forward flanking *rnb* oligonucleotide
BAP9076	CCCTTTGTCGACTTAACAAAGAGTGCCGACAAT	Reverse flanking *rnb* oligonucleotide
BAP9077	CCTCAAGAGCTCAGGAGGTTGTTTTCAAAAGTGCATATTTT	Forward flanking *epmB* oligonucleotide
BAP9078	TTGTTGGTCGACCTATGCTGTATATAGGGTTT	Reverse flanking *epmB* oligonucleotide
BAP9163	ATAGCCTCTAGAAG GAGGGCAACTATGCCATTTG	Forward flanking *cpxA* oligonucleotide
BAP9080	AACATGGTCGACTTAACTGGTAATCCAAAGCG	Reverse flanking *cpxA* oligonucleotide

2.5.2 Growth curves

The growth kinetics of different *P. multocida* strains were assessed using growth in heart infusion broth. Strains were grown overnight with appropriate antibiotics, diluted 1:50 in fresh HI broth without antibiotic selection and grown to an OD_{600} of 0.2, before being diluted 1:100 in fresh HI broth. The cultures were then grown for 24 h, with OD_{600} readings taken every 30 min. The number of viable cells (CFU/mL) was determined for each biological replicate at mid-exponential growth phase (OD_{600} of 0.5).

2.5.3 Serum sensitivity assays

Serum sensitivity assays were performed as previously described, with minor modifications (76). The input culture was increased from ~10^5 to ~10^6 CFU to ensure the full representation of the *Himar1* mutant library, and the volume of the assay was increased to 3 mL to ensure maximum contact of cells to the complement components in the serum. Preliminary assays examining single strains were performed in a total volume of 50 µL of 90% untreated or heat-inactivated chicken serum, with 10% HI broth.

2.5.4 TraDIS sequencing data analysis

Genes essential for *P. multocida* growth in untreated chicken serum were identified using the TraDIS analysis pipeline as described in the manuscript above. For identification of fitness defects in serum, the tradis_gene_insert_sites script was run with no trimming of reads from the 3' end of the gene. Genes were called as having a fitness defect in if they showed a fold-change in *Himar1* reads in the serum TraDIS data compared to the HI TraDIS data < -4 (log_2 < -2) and had a *q*-value < 0.001.

2.5.5 Transmission electron microscopy

TEM was performed using *P. multocida* strains grown to mid-exponential phase (OD_{600} ~0.5). For initial TEM analysis, cells were submitted to the Ramaciotti Centre for Cryo-Electron Microscopy for preparation and imaging. Cells were initially fixed in 2.5% glutaraldehyde, 2% paraformaldehyde, 0.075% ruthenium red in 0.1 M sodium cacodylate buffer for 2 h at RT. For the second TEM analysis, cells were fixed in the same solution with the addition of 50 mM of lysine to help stabilize the capsule. The cells were then washed three times in 0.1 M sodium cacodylate buffer for 15 mins. Cells were postfixed in 1% osmium tetroxide buffer for 1 h at RT, then washed three times in milli-Q water for 15 mins. The cells were then embedded in 4% agarose and dissected into 1 mm^3 squares. The agarose-embedded cells were dehydrated by incubating in a graded ethanol series (30, 50, 70, 90, 100%). The embedded cells were then incubated in a series as follows: 30 min in 100% propylene oxide, 6 h in a 1:1 mixture of Epon resin and propylene oxide, overnight in a 2:1 mixture of Epon resin and propylene oxide, 6 h in 100% Epon resin, in fresh 100% Epon resin overnight. The embedded cell pellets were placed into Beem capsules with Epon resin, with the resin polymerised at

60°C for 48 h. Sections (80-90 nm) of the resin-embedded cell pellets were cut using a Leica UCS ultramicrotome, then sections were added to formvar-coated 100 mesh copper grids and then stained with uranyl acetate for 10 min. The cell sections were then imaged on a JEOL 1400 Plus Transmission Electron Microscope.

2.5.6 Scanning electron microscopy

Scanning electron microscopy was performed as described in the manuscript above with minor modifications. Samples were dried using hexamethyldisilazane chemical drying in place of critical point drying.

2.6 Appendices

Appendix 2.1 – Serum TraDIS data

Appendix 2.2 – Genes identified as essential for growth in untreated chicken serum

Appendix 2.3 – Genes identified as having a fitness defect when grown in untreated chicken serum

Chapter 3 – Genomic analysis of *Pasteurella* isolates from humans, cats and dogs

3.1 Introduction

Bacteria from the genus *Pasteurella* are Gram-negative coccobacilli that are commensals in the upper respiratory tract of many mammal and bird species (233). The *Pasteurella* genus currently comprises 13 species; namely, *aerogenes*, *bettyae*, *caballi*, *caecimuris*, *canis*, *dagmatis*, *langaaensis*, *mairii*, *multocida*, *oralis*, *skynesis*, *stomatis* and *testudinis* (https://lpsn.dsmz.de/genus/pasteurella) (170). However, phylogenetic analysis of 16s rRNA sequence suggests that only *Pasteurella* spp. *multocida*, *canis*, *dagmatis* and *stomatis* are members of the *Pasteurella sensu stricto* (177, 178). Several *Pasteurella* spp. are pathogens, causing a wide range of human and animal diseases. Species *aerogenes*, *bettyae*, *caballi*, *canis*, *dagmatis* and *stomatis* have been associated with several human diseases including skin infections following animal bites and scratches, urethritis, peritonitis and septicaemia (59, 171-175). The most well characterised of the *Pasteurella* species is *P. multocida*. *P. multocida* is responsible for several distinct animal disease syndromes, including fowl cholera in birds, haemorrhagic septicaemia in ungulates, atrophic rhinitis in pigs and rabbits, and bovine respiratory disease in livestock animals (13). Collectively these animal diseases have a large economic impact on several agricultural industries.

P. multocida strains can be differentiated into three subspecies (subsp. *multocida*, *gallicida* and *septica*) that have differential sorbitol and dulcitol fermentation, and by their capsule (A, B, D, E and F) and LPS (L1-L8) genotypes, with strains from different types producing structurally different capsule or LPS outer core polysaccharides (9, 11, 234). *P. multocida* strains from specific capsule and/or LPS types are often associated with specific diseases. Fowl cholera is generally caused by capsule types A and F, and LPS types L1 and L3 (6, 18-20), HS is exclusively caused by capsule type B and E strains (16, 34, 35), atrophic rhinitis is almost exclusively caused by toxigenic group A and D strains (42, 43), and bovine respiratory disease is primarily caused by type A strains (43). Despite this strong association between capsule and LPS type and disease, the molecular basis of *P. multocida* host predilection and ability to cause different animal diseases remains poorly understood. In addition to animal diseases, *P. multocida* also causes several human diseases, commonly causing skin wound infections following animal bites and scratches, as well as pneumonia, meningitis, and septicaemia in immunocompromised individuals (61, 64, 65, 235). Apart from some limited experiments using PCR (58), very little is known about the types of *P. multocida* strains that cause human infections.

Several studies have used bioinformatic analysis of *P. multocida* genomes in an attempt to identify determinants of *P. multocida* pathogenesis and host specificity. Analysis of individual genome sequences has allowed identification of several virulence factors and antibiotic resistance genes. The first *P. multocida* genome sequenced and analysed was strain Pm70 (capsule genotype F:LPS genotype L3), an avian isolate (118). Homology to other proteins allowed for the prediction of two filamentous haemagglutinins, and 53 iron scavenging and uptake proteins (118). Sequence analysis of the *P. multocida* strain 36950 genome identified an 82-kb integrative conjugative element (ICE), denoted ICE*Pmu1*, that contained 11 antibiotic resistance genes (162). Comparative genomics has also been used to predict virulence factors, and host- or disease-specific genes. Comparison of *P. multocida* strain HN07 (F:L3) and Pm70 allowed for identification of ICE*pmcn07* that contained a type IV secretion system (165). Comparison of the avian-virulent strains X73 (A:L1) and P1059 (A:L3) with the putatively avian-avirulent strain Pm70 identified 336 genes with possible roles in avian virulence, including two putative filamentous haemagglutinins, an L-fucose metabolism system, ribose specific ABC-transporters and a trimethylamine-N-oxide reductase system (122). However, it is now known that the original Pm70 strain is indeed virulent in birds, but some isolates of this strain were avirulent due to a single point mutation in an LPS biosynthesis gene (Harper pers. comm.). Comparison of genomes from *P. multocida* HS isolates with avian and porcine isolates identified 92 genes unique to HS strains, indicating possible HS-specific virulence factors (164). This analysis also identified four large genomic regions conserved in HS isolates that contained two intact bacteriophages and two putative bacteriophages, as well as a putative ICE carrying seven antibiotic resistance genes conserved in HS isolates (164). These studies identified genes that may be associated with host and disease specificity; however, these studies used only limited numbers of different *P. multocida* strains. A very recent study performing comparative genomics with 656 *P. multocida* strains identified several virulence factors associated with strains from particular capsule types, including PtfA that was associated with capsule type B and F strains, HgbA that was associated with capsule type F strains, and OmpH1, OmpH3, PlpE and PfhB1 that were associated with capsule type A strains, indicating possible roles for these proteins in host predilection or disease specificity (167). In the same study, single nucleotide variants (SNVs) were identified in 653 *P. multocida* strains when compared to *P. multocida* strain ATCC 43137, with a phylogeny based on these SNVs showing separation of strains into two clades, with human and avian isolates clustering away from other animal isolates into an outer clade (167). Pan-genome analysis identified several genes associated with the human and avian outer clade, representing possible virulence genes associated with disease in humans and birds (167). In addition, comparison with known antibiotic resistance genes identified some mobile genetic elements containing antibiotic resistance genes in the chromosome of some *P. multocida* strains (167).

Since the first *P. multocida* genome sequence was published in 2001, the number of sequenced *P. multocida* strains has lagged behind other species, with only 21 publicly available whole-genome sequences released by 2015. However, since 2015, an additional 254 whole-genome sequences have been made available. However, of the more than 270 publicly available *P. multocida* whole-genome sequences, only eight have come from human isolates. In addition, other *Pasteurella* spp. have no more than three publicly available genome sequences. To better understand genetic determinants of *P. multocida* host and disease specificity, in this study 35 human, 16 feline and 8 canine *Pasteurella* spp. isolates were sequenced and analysed by a range of different bioinformatic methods to identify putative virulence factors and trait-associated genes. These combined investigations indicated that these human, cat and dog isolates had a similar virulence factor profile to other *P. multocida* species; however, loss of the capsule biosynthesis loci and divergence of the Tad-locus genes was observed in several isolates. In addition, comparative genomics was performed using 260 publicly available *P. multocida* reference whole genome sequences, identifying genes over-represented in type F and type B:L2 strains, indicating putative host- and disease-specific virulence genes. Furthermore, a *P. multocida* phylogeny was generated that showed two clear clades; one containing mostly human, cat and dog isolates (clade A) and one containing mostly animal isolates (clade B). Pan-genome analysis identified a putative fucose uptake and utilisation system over-represented in strains from clade A, and anaerobic respiration and carbohydrate uptake genes over-represented in strains from clade B.

3.2 Results

3.2.1 Sequencing and species identification *Pasteurella* spp. isolated from humans, cats and dogs

Pasteurella isolates from human infections were donated by Monash Medical Centre (Tony Korman and Despina Kotsanas pers. comm.) and *Pasteurella* isolates from dog infections and cat normal flora were donated by the University of Queensland (Conny Turni and Justine Gibson, pers. comm.) (Table 3.8). Human isolates were recovered from blood culture, peritoneal dialysis or skin infections following dog or cat bites. Human isolates were initially identified as *P. multocida* (22), *P. canis* (8) or *P. dagmatis* (5) by MALDI-TOF mass spectrometry (Table 3.8). Cat isolates were recovered from the upper respiratory tract of wild cats near chicken farms in Queensland and typed as *P. multocida* by a Gram-negative Microbact kit (Table 3.8). Dog isolates were recovered from skin wound swabs, URT washes, ear or eye and typed as *P. multocida* by biochemical tests (Table 3.8). Genomic DNA was extracted from overnight cultures grown in HI broth. Prior to sequencing, *P. multocida*-specific KMT PCR was used to confirm identification of isolates as *P. multocida* as indicated by MALDI-TOF, Microbact or biochemical tests (excluding strain PI38). Eight of the isolates typed as *P. multocida* did not produced a KMT-specific product, six of which were typed by biochemical tests (Figure 3.1).

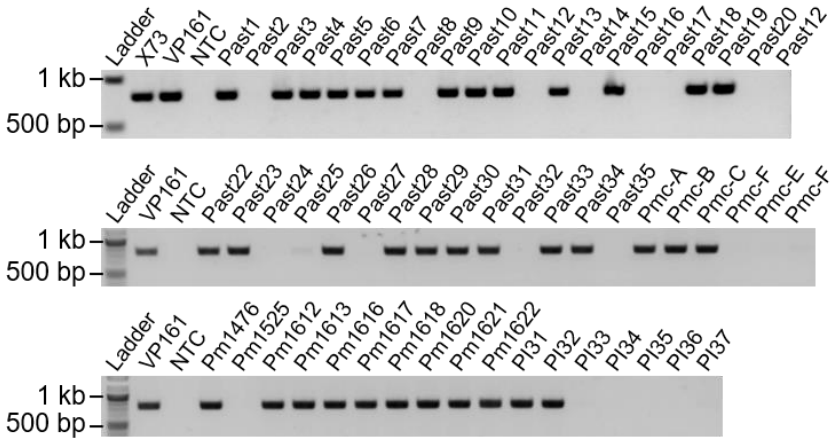


Figure 3.1: Identification of *P. multocida* strains using KMT-specific PCR. Genomic DNA from *Pasteurella* isolates that were isolated from humans, cats or dogs was used as a template with primers BAP511 and BAP512 that target the *P. multocida*-specific KMT genomic fragment. *P. multocida* strains X73 and VP161 were used as positive controls. PCR products were run alongside a no-template control reaction (NTC) and were compared to a 1 kb or 100 bp DNA ladder (NEB) with relevant ladder markers given. Reactions were run on a 1% w/v agarose gel at 110 V for 45 min.

Genomic DNA isolated from each strain was submitted to Micromon Genomics (Monash University) and sequenced on an Illumina MiSeq. Sequencing reads were first assessed using FASTQ trimmer, where poor quality reads and Illumina adapter sequences were removed, before being assembled into draft genomes *de novo* using Unicycler (236). Draft genomes were then annotated using Prokka. Prokka identifies open reading frames (ORFs) then assigns gene function to ORFs based on similarity to genes in highly curated databases (UniProtKB, NCBI AMRFinder and ISfinder) (237). Quast was used to determine quality statistics for the annotated draft genomes (Table 3.1). All isolate draft genomes had between 29x and 96x coverage and contained between 15 and 71 contigs. The smallest contig size required for generating half the genome (N50) was between 64,204 and 490,318, and the number of contigs required for generating half the genome (L50) was between 2 and 13 contigs (Table 3.1). Collectively, these data show draft genomes were good quality, with most isolate genomes containing few contig breaks over 50% of the genome.

Average nucleotide identity (ANI) was used to identify the species of each isolate. ANI compares the number of commonly present kmers between two genomes, with any ANI score over 95 between query and reference indicating that the two genomes come from the same species (238, 239). ANI species identification has a similar accuracy compared with DNA-DNA hybridisation species identification (239). Given all isolates were initially identified as *P. multocida*, *P. canis* or *P. dagmatis*, *Pasteurella* spp. type strains were used as the reference in fastANI analysis (238). No whole-genome sequences for *P. caballi* and *P. stomatis* were publicly available, therefore no ANI comparisons could be made to these species. Nearly all isolates had ANI values greater than 95 against *P. multocida*, *P. canis* or *P. dagmatis* reference genomes, allowing isolates to be called as one of these three species (Table 3.2). All strains that were KMT-PCR positive had a high ANI value when compared to the *P. multocida* reference genome and these strains had an average %GC content of 40.43 ± 0.08.

Table 3.1: Isolate genome assembly quality control data.

Strain[1]	Genome Size	Sequencing depth	Contigs[2]	N50	L50	GC content (%)	Number of ORFs
Past1^{Pm}	2 290 030	81x	30	150 782	6	40.40	2159
Past2^{Pd}	2 433 423	73x	67	71 469	12	37.47	2297
Past3^{Pm}	2 251 995	63x	30	135 802	6	40.43	2134
Past4^{Pm}	2 222 762	72x	43	91 867	9	40.52	2079
Past5^{Pm}	2 209 709	80x	22	133 647	5	40.42	2088
Past6^{Pm}	2 179 954	78x	30	147 327	6	40.54	2052
Past7^{Pm}	2 312 025	75x	22	151 557	4	40.38	2217
Past8^{Pc}	2 147 754	89x	22	132 576	5	36.56	2028
Past9^{Pm}	2 373 692	82x	36	161 100	5	40.49	2302
Past10^{Pm}	2 442 576	69x	48	90 591	9	40.25	2374
Past11^{Pm}	2 287 048	71x	41	101 177	8	40.44	2174
Past12^{Pc}	2 150 109	84x	33	93 381	9	36.66	2015
Past13^{Pm}	2 284 711	73x	27	155 275	5	40.37	2173
Past14^{Pc}	2 185 832	75x	57	81 062	9	36.78	2073
Past15^{Pm}	2 294 665	69x	28	120 190	6	40.33	2201
Past16^{Pc}	2 023 144	82x	28	94 244	8	36.54	1879
Past17^{Pc}	2 162 587	77x	41	84 804	9	36.77	2011
Past18^{Pm}	2 272 683	70x	28	161 774	6	40.36	2157
Past19^{Pm}	2 162 151	67x	32	113 211	7	40.40	2047
Past20^{Pc}	2 174 935	59x	71	47 100	13	36.67	2050
Past21^{Pd}	2 316 545	78x	36	124 330	7	37.51	2154
Past22^{Pm}	2 264 618	29x	63	101 116	7	40.48	2149
Past23^{Pm}	2 308 982	27x	62	70 506	12	40.43	2186
Past24^{Pc}	2 145 938	36x	66	53 907	14	36.74	2004
Past25^{Ps}	2 474 472	28x	87	77 361	10	38.92	2277
Past26^{Pm}	2 180 995	43x	54	82 630	10	40.61	2053
Past27^{Pc}	2 112 546	31x	60	62 319	13	36.65	1956
Past28^{Pm}	2 248 839	96x	34	139 426	7	40.51	2131
Past29^{Pm}	2 222 654	31x	47	103 954	7	40.42	2083
Past30^{Pm}	2 227 555	34x	24	173 135	5	40.42	2079
Past31^{Pm}	2 254 775	38x	49	92 486	9	40.58	2141
Past32^{Pd}	2 277 138	44x	51	103 305	8	37.47	2097
Past33^{Pm}	2 268 144	40x	31	129 213	7	40.40	2152
Past34^{Pm}	2 360 301	43x	53	85 958	8	40.58	2264
Past35^{Pc}	2 143 777	51x	35	104 034	6	36.6	2009
Pm1476^{Pm}	2 306 050	29x	59	161 858	6	40.5	2196
Pm1525^{Fc}	2 264 289	38x	16	278 170	2	42.7	2187
Pm1612^{Pm}	2 189 281	45x	31	108 374	6	40.38	2040
Pm1613^{Pm}	2 193 562	31x	23	192 146	4	40.35	2047
Pm1616^{Pm}	2 216 246	43x	41	121 108	7	40.34	2084
Pm1617^{Pm}	2 240 225	23x	64	69 675	11	40.44	2119
Pm1618^{Pm}	2 212 068	56x	30	154 291	6	40.49	2097
Pm1620^{Pm}	2 196 796	48x	21	167 963	4	40.34	2047
Pm1621^{Pm}	2 226 418	44x	50	185 480	5	40.32	2091
Pm1622^{Pm}	2 227 463	29x	28	196 480	4	40.44	2111
Pmc-A^{Pm}	2 279 063	36x	25	318 478	3	40.43	2183
Pmc-B^{Pm}	2 268 061	54x	29	185 538	5	40.47	2167
Pmc-C^{Pm}	2 277 767	28x	25	311 494	3	40.43	2182
Pmc-D^{Ps}	2 472 039	38x	28	221 898	4	38.9	2267
Pmc-E^{Ps}	2 468 801	42x	37	162 923	6	38.91	2275
Pmc-F^{Ps}	2 469 333	49x	39	122 316	8	38.93	2274
PI31^{Pm}	2 228 956	53x	25	219 012	3	40.42	2118
PI32^{Pm}	2 240 863	49x	25	207 556	4	40.43	2110
PI33^{Pc}	2 133 587	76x	59	64 204	13	36.64	2010
PI34^{Pc}	2 168 297	39x	27	158 469	4	36.63	2032
PI35^{Pc}	2 145 706	57x	22	273 147	3	36.64	2016
PI36^{Pc}	2 150 366	37x	60	65 641	11	36.67	2002
PI37^{Fc}	2 073 365	43x	16	313 761	3	42.53	1982
PI38^{Fc}	2 092 205	41x	15	490 318	2	42.47	1978

[1]Isolates identified as *P. multocida* ^{Pm}, *P. canis* ^{Pc}, *P. dagmatis* ^{Pd}, *P. stomatis* ^{Ps} or *Frederiksenia canicola* ^{Fc}

[2]Only contigs >500 nucleotides in length were included in the count

Table 3.2: Best average nucleotide identity (ANI) match for isolate genomes compared against *Pasteurella* spp. type strain reference genomes. ANI values were generated using fastANI.

Query genome	Reference genome	ANI value
Past1	*Pasteurella multocida* NCTC10322	96.79
Past2	*Pasteurella dagmatis* NCTC11617	98.82
Past3	*Pasteurella multocida* NCTC10322	98.60
Past4	*Pasteurella multocida* NCTC10322	96.95
Past5	*Pasteurella multocida* NCTC10322	96.89
Past6	*Pasteurella multocida* NCTC10322	96.81
Past7	*Pasteurella multocida* NCTC10322	96.84
Past8	*Pasteurella canis* NCTC11621	98.83
Past9	*Pasteurella multocida* NCTC10322	96.93
Past10	*Pasteurella multocida* NCTC10322	96.82
Past11	*Pasteurella multocida* NCTC10322	96.94
Past12	*Pasteurella canis* NCTC11621	98.93
Past13	*Pasteurella multocida* NCTC10322	96.85
Past14	*Pasteurella canis* NCTC11621	98.85
Past15	*Pasteurella multocida* NCTC10322	98.78
Past16	*Pasteurella canis* NCTC11621	98.90
Past17	*Pasteurella canis* NCTC11621	98.95
Past18	*Pasteurella multocida* NCTC10322	96.93
Past19	*Pasteurella multocida* NCTC10322	96.99
Past20	*Pasteurella canis* NCTC11621	98.82
Past21	*Pasteurella dagmatis* NCTC11617	99.97
Past22	*Pasteurella multocida* NCTC10322	97.05
Past23	*Pasteurella multocida* NCTC10322	96.74
Past24	*Pasteurella canis* NCTC11621	98.87
Past25	*Pasteurella dagmatis* NCTC11617	90.15
Past26	*Pasteurella multocida* NCTC10322	96.98
Past27	*Pasteurella canis* NCTC11621	98.99
Past28	*Pasteurella multocida* NCTC10322	96.89
Past29	*Pasteurella multocida* NCTC10322	98.38
Past30	*Pasteurella multocida* NCTC10322	96.87
Past31	*Pasteurella multocida* NCTC10322	96.89
Past32	*Pasteurella dagmatis* NCTC11617	98.94
Past33	*Pasteurella multocida* NCTC10322	98.76
Past34	*Pasteurella multocida* NCTC10322	96.88
Past35	*Pasteurella canis* NCTC11621	98.91
Pm1476	*Pasteurella multocida* NCTC10322	96.85
Pm1525	*Pasteurella dagmatis* NCTC11617	77.95
Pm1612	*Pasteurella multocida* NCTC10322	98.88
Pm1613	*Pasteurella multocida* NCTC10322	98.84
Pm1616	*Pasteurella multocida* NCTC10322	98.81
Pm1617	*Pasteurella multocida* NCTC10322	98.71
Pm1618	*Pasteurella multocida* NCTC10322	96.94
Pm1620	*Pasteurella multocida* NCTC10322	98.87
Pm1621	*Pasteurella multocida* NCTC10322	98.82
Pm1622	*Pasteurella multocida* NCTC10322	96.87
Pmc-A	*Pasteurella multocida* NCTC10322	96.82
Pmc-B	*Pasteurella multocida* NCTC10322	96.78
Pmc-C	*Pasteurella multocida* NCTC10322	96.72
Pmc-D	*Pasteurella dagmatis* NCTC11617	90.14
Pmc-E	*Pasteurella dagmatis* NCTC11617	90.27
Pmc-F	*Pasteurella dagmatis* NCTC11617	90.21
PI31	*Pasteurella multocida* NCTC10322	96.87
PI32	*Pasteurella multocida* NCTC10322	98.75
PI33	*Pasteurella canis* NCTC11621	98.87
PI34	*Pasteurella canis* NCTC11621	98.92
PI35	*Pasteurella canis* NCTC11621	98.83
PI36	*Pasteurella canis* NCTC11621	98.84
PI37	*Pasteurella dagmatis* NCTC11617	77.74
PI38	*Pasteurella dagmatis* NCTC11617	78.01

The best match for isolates Past25, Pm1525, Pmc-D, Pmc-E, Pmc-F, PI37 and PI38 had ANI scores lower than 95 (Table 3.2). Isolates with no matches were compared by fastANI to all complete Pasteurellaceae genomes in the NCBI database. Isolates Pm1525, PI37 and PI38 matched to *Frederiksenia canicola* strain HPA 21 with ANI values greater than 98, allowing identification of these isolates as *F. canicola* strains (Table 3.3, Appendix 3.2). The top match for isolates Pmc-D, Pmc-E, Pmc-F and Past25 was *P. dagmatis*, but with an ANI value between 90.1 and 90.3, indicating these isolates are likely a closely related species to *P. dagmatis* (Table 3.3). To determine the species of these isolates, the complete 16s rRNA gene sequence was compared against the NCBI reference RNA sequence database and complete 16s rRNA gene sequences from *Pasteurella* type strains that included *P. stomatis* and *P. caballi* by BLASTn. Due to contig breaks in the draft assembly for these isolates, only Pmc-E had a full length 16s rRNA gene. ANI values between Pmc-D, Pmc-E, Pmc-F and Past25 were over 98, indicating these isolates were the same species, allowing identification of Pmc-E by 16s rRNA identity to be indicative of all four unidentified isolates. The highest identity match was to *P. stomatis* 16s rRNA with 98.61% identity over 88% of the gene, identifying these isolates as *P. stomatis* strains

Table 3.3: Best average nucleotide identity (ANI) match for isolate genomes compared against complete Pasteurellaceae reference genomes from NCBI assembly database. ANI values were generated using fastANI.

Query genome	Reference genome	ANI value
Pmc-D	*Pasteurella dagmatis* NCTC8282	90.2048
Pmc-E	*Pasteurella dagmatis* NCTC8282	90.3853
Pmc-F	*Pasteurella dagmatis* NCTC11617	90.2196
Past25	*Pasteurella dagmatis* NCTC8282	90.1746
Pm1525	*Frederiksenia canicola* HPA 21	98.0332
PI37	*Frederiksenia canicola* HPA 21	98.3419
PI38	*Frederiksenia canicola* HPA 21	98.3742

The genome sequences for *P. dagmatis*, *P. canis* and *F. canicola* isolates greatly expand the number of whole genome sequences available for these species, and the sequencing of *P. stomatis* isolates represent the first genome sequences available for this species. The *P. canis* genomes were all between 2,023,144 and 2,185,832 bp in length, with a %GC content between 36.54 and 36.77 (36.66 ± 0.07) and contained between 1,879 to 2,073 ORFs (Table 5.1). *P. dagmatis* genomes were between 2,277,138 and 2,433,423 bp in length, with a %GC content between 37.47 and 37.51 (37.48 ± 0.02) and had between 2,097 and 2,297 ORFs (Table 3.1). *P. stomatis* genomes were between 2,468,801 and 2,474,472 bp in length, with a %GC content between 38.9 and 38.93 (38.92 ± 0.1) and 2,267 and 2,277 ORFs (Table 3.1). *F. canicola* genomes were between 2,092,205 and 2,264,289 bp in length, with a %GC content between 42.47 and 42.7 (42.57 ± 0.12) and between 1,978 and 2,187 ORFs. Assemblies from the same species had small ranges for total length, percentage GC content and number of ORFs, indicating good quality genome assembly and annotation.

3.2.2 Identification of virulence and resistance genes

Genome assemblies were searched for known virulence and antibiotic resistance genes using ABRicate (https://github.com/tseemann/abricate). ABRicate searches query genomes against highly curated virulence factor and antibiotic resistance factor gene databases. Databases used for ABRicate searches included the virulence factor database (VFDB), comprehensive antibiotic resistance database (CARD), ResFinder and PlasmidFinder (240-243). Matches were designated as positive hits if the region in the query genome had >50% nucleotide identity over >50% of the reference sequence. ABRicate searches against the CARD and ResFinder databases produced only a few matches and no matches were seen against the PlasmidFinder database. All isolates had matches for *tet34* and *hmrM* that confer oxytetracycline and fluroquinolone resistance respectively, and several *P. multocida* isolates also had a match to *vatE*, which confers streptogramin resistance. All CARD and ResFinder matches were between 68% and 72% nucleotide identity with several gaps, indicating that these genes were not transferred by horizontal gene transfer recently and given the low identity may not confer antibiotic resistance.

All strains contained several genes that matched reference genes in the VFDB (Full results in Appendix 3.3). The majority of matches for all isolates were against reference genes from *Haemophilus influenzae* that were involved in LPS biosynthesis. All isolates also had matches to conserved stress response genes, such as *htpB* from *Legionella pneumophila*, *clpP* from *Listeria monocytogenes*, and *katA* from *Neisseria meningitidis* that encode a chaperone, Clp protease and catalase, respectively; these genes are highly conserved across most bacterial species. All isolates had a match to *luxS* from *Vibrio cholerae*, suggesting they have a quorum sensing system. Homologs of genes in the *lsr* operon (*lsrABCDFGKR*) from *E. coli* K-12 strain

MG1655 were identified in all *Pasteurella* spp. isolates by BLASTn, indicating these strains possibly have a functional quorum sensing system.

All *P. multocida* isolates, excluding strains Past6, Past33, Past1616, Past1617 and Past1621, had matches to a capsule specific ABC-transporter system from *Haemophilus influenzae* (Appendix 3.3). Most *P. multocida* isolates, all *P. dagmatis* and *F. canicola* isolates had matches to *ilpA* from *Vibrio vulnificus* that encodes a lipoprotein. The IlpA protein showed 58% amino acid identity to PlpB from *P. multocida* strain Pm70, suggesting matches to *ilpA* are likely *plpB* homologs. *P. stomatis* isolates and the strain Past2 had matches to *hcp-2* and *vipB* from *Vibrio cholerae*, which encode components of the type VI secretion system (T6SS) tube suggesting a possible T6SS; however, these isolates had no matches for other T6SS genes in the VFDB. These matches to the VFDB had between 67% and 72% nucleotide identity over 59% to 91% of the gene and had up to 15 gaps in the query match, suggesting possible frameshift mutations in the query. Alignments between reference and query sequences are not given as an output of ABRicate, so the effect of gaps on the protein sequence could not be investigated. There were several other hits for the isolates against the VFDB, however most of these matches had between 40 and 120 gaps, indicating the query was unlikely to have the same function as the reference. Although these databases give useful information about certain well-studied virulence and antibiotic resistance genes, there is little data from species that are closely related to *Pasteurella* spp., which limits the identification of virulence and resistance genes in the sequenced isolates.

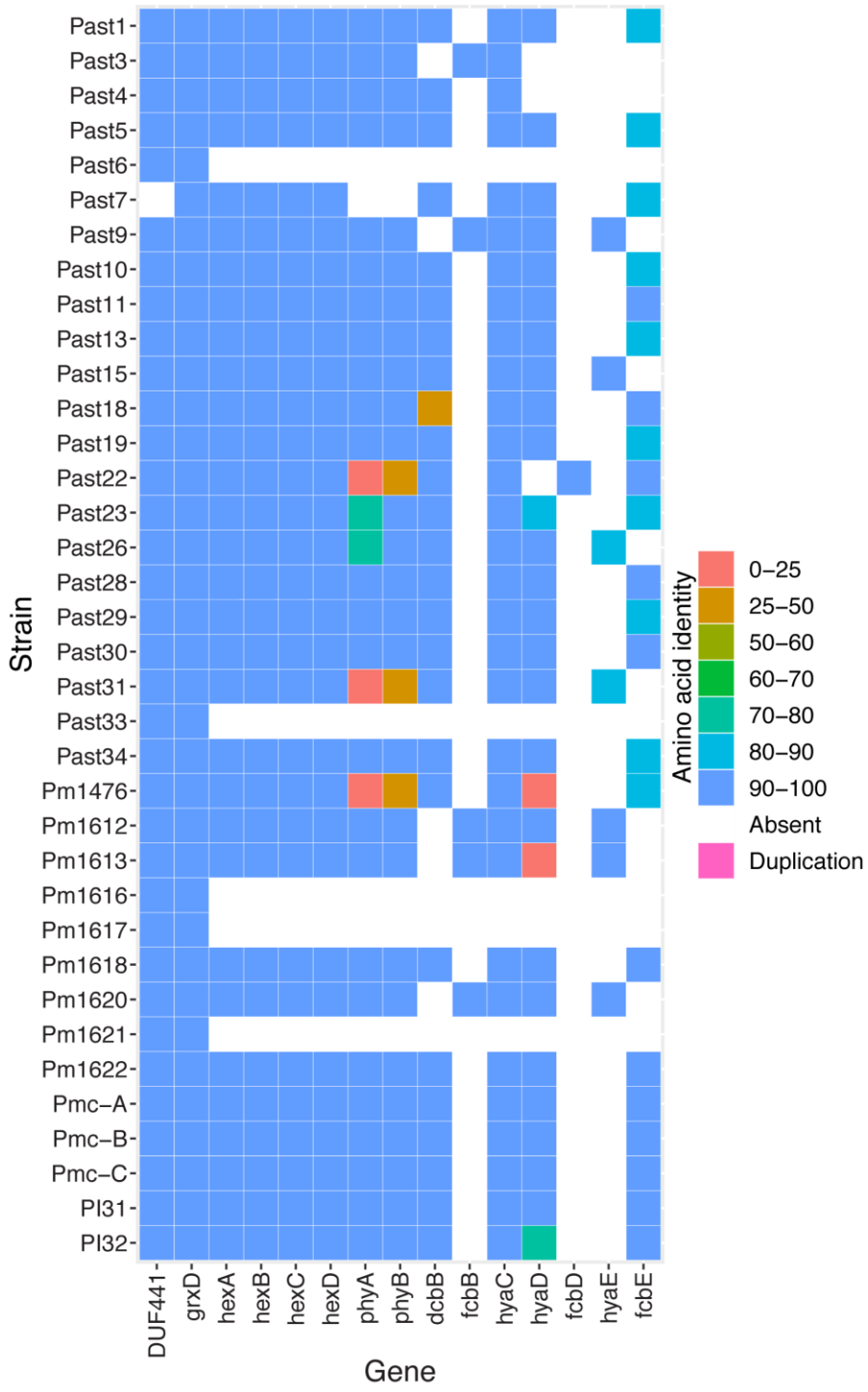
3.2.3 Identification of *P. multocida*-specific virulence and resistance genes

For a more targeted identification of *P. multocida*-specific virulence and antibiotic resistance genes in the newly sequenced isolates, a database of known *P. multocida* virulence and antibiotic resistance genes was compiled (designated *Pasteurella* virulence and resistance database or PastyVRDB). Genes included in the PastyVRDB fell into five categories: capsule biosynthesis; LPS biosynthesis; outer membrane proteins, adhesins and virulence factors; iron scavenging and uptake proteins; and mobile genetic elements and antibiotic resistance genes (including integrative conjugative element (ICE) and plasmid genes). The full PastyVRDB list is given in appendix 3.6. Genes were included in the PastyVRDB if there was direct evidence of a role in virulence or antibiotic resistance in *P. multocida*, or the protein had high amino acid identity to a known virulence factor or antibiotic resistance protein from another species. All isolate genomes were compared with the PastyVRDB by Assembly2Gene. Assembly2Gene allows identification of genes in query genomes that are split by contig breaks, making it ideal for searching draft genomes. In addition, Assembly2Gene gives detailed data about the relative position of missense mutations, insertions, deletions, and truncations compared to the reference, as well as possible gene

duplications. Data generated by Assembly2Gene was summarised into heat maps (Figures 5.2 to 5.5). Full datasets are given in appendices 3.2 to 3.6. All matches had nucleotide and amino acid identity normalised for reference gene length and any matches with <50% nucleotide identity were excluded from the heat maps. All strains had several hits for all categories in the PastyVRDB except for plasmid, ICE and antibiotic resistance genes. Only Past33 had high amino acid identity hits to mobilisation genes *mobA* and *mobC* from pB1000, indicating this strain may harbour a plasmid. No heat map was generated for plasmid, ICE and antibiotic resistance gene searches. In addition, non-*P. multocida* strains were also assessed by Assembly2Gene analysis; however, these genomes contained very few matches to genes in the PastyVRDB.

Capsule biosynthesis genes are co-localised in the *P. multocida* genome, with the locus flanked by a conserved DUF441-containing gene and *grxD*. The capsule biosynthesis locus can be broken into three regions that encode proteins known or predicted to be involved in polysaccharide lipid attachment (region 1), polysaccharide synthesis (region 2) and capsule export (region 3) (Figure 1.2) (77, 78). Genes required for capsule export and lipid attachment have high nucleotide identity between genotypes A, D and F, and between genotypes B and E (9). Therefore, only capsule export and lipidation genes from genotypes A (*phyAB* and *hexABCD*) and B (*lipAB* and *cexABCD*) were included as references. All capsule polysaccharide biosynthesis genes from all capsule genotypes were included in the PastyVRDB. Several of the polysaccharide biosynthesis genes from genotypes A, D and F have >97% nucleotide identity (*hyaB*/*dcbB*/*fcbB*, *hyaC*/*dcbC*/*fcbC* and *hyaE*/*dcbE*/*fcbE*), which resulted in isolates showing positive matches to capsule biosynthesis genes from multiple capsule types. However, all capsule genotypes have either unique genes, or genes with significant nucleotide differences that allow for capsule genotype to be accurately determined from the Assembly2Gene output. Most (31/36) *P. multocida* isolates were predicted to encode a fully functional capsule biosynthesis system (Figure 3.2). These strains encoded proteins with high amino acid identity to HexABCD and PhyAB, as well as HyaC, either DcbB or FcbB, either HyaD or FcbD, and either HyaE or FcbE. All of the 31 capsule-positive isolates were genotype A, except Past22, which was genotype F. Several strains of *P. multocida* had low amino acid identity matches, or no matches for PhyAB, HyaD, HyaE or FcbE (strains Past3, Past4, Past7, Past22, Past23, Past26, Past31, Pm1476 and Pm1613); however, these strains had contig breaks in the capsule biosynthesis locus with nucleotide sequences missing between the contigs that corresponded to the predicted positions of the absent genes. Several variant and truncated capsule biosynthesis genes were identified in the *P. multocida* isolates containing a predicted functional capsule biosynthesis system. Twenty isolates had a query sequence matching to *hyaD* that contained a 14 bp and 7 bp gap at

nucleotides 154 and 159 respectively, as well as 30 conserved nucleotide polymorphisms that introduced missense mutations. Ten isolates contained a match to *hyaE* or *fcbE* that contained a 135 bp gap at nucleotide 1,148 and 54 conserved missense mutations. Past18 had two gaps in *dcbB* at nucleotide 676 and 686 that result in a nonsense mutation, truncating the encoded protein and likely resulting in a non-functional capsule biosynthesis system given *hyaB* is required for capsule biosynthesis (92).



Figures 3.2: Heatmap showing presence of *P. multocida* capsule biosynthesis proteins in the *P. multocida* isolates sequenced in this study identified by Assembly2Gene (See Appendix 3.6 for full PastyVRDB). Coloured squares represent the amino acid identity between proteins encoded by query and reference sequences, or for when there are multiple matches in a query genome for a reference gene, a gene duplication. Isolate genomes and/or reference genes with no matches were excluded from the heat map.

Five of the *P. multocida* isolate strains (Past6, Past33, Pm1616, Pm1617 and Pm1621) had matches to *grxD* and the DUF441-containing gene, but no matches to any of the capsule biosynthesis genes (Figure 3.2). These capsule negative strains had *grxD* and the DUF441-gene adjacent to each other on an intact contig, and no genes with identity to the capsule biosynthesis locus were identified elsewhere in the genomes. There are two main types of capsule biosynthesis systems in Gram-negative bacteria, the ABC-transport-dependant and Wzy-polymerase-dependant systems (244). Wzy-polymerase dependant systems have five conserved genes (*wza*, *wzb*, *wzc*, *wzx* and *wzy*). A BLASTn search of these five conserved genes obtained from *E. coli* strain K12 did not identify matches in the five capsule negative strains, indicating that these strains do not contain a Wzy-polymerase-dependant capsule biosynthesis system. Together this suggests these strains are unlikely to produce a capsule layer.

The genes involved in the biosynthesis of *P. multocida* LPS have been characterised previously (12). These include genes encoding proteins involved in LPS inner core biosynthesis that are found at various genomic locations, and genes encoding LPS outer core polysaccharide transferases that are found at a conserved locus between *priA* and *fpg* (Figure 1.3) (12). Two different LPS inner core structures have been identified (glycoform A and B), with most characterised strains able to simultaneously produce both types (105). Eight different outer core biosynthesis loci have been identified and characterised (denoted L1-L8) (11). Six inner core and all outer core biosynthesis genes were included in the PastyVRDB. All *P. multocida* isolates had matches for the inner core biosynthesis genes, indicating that all strains would be able to produce both inner core glycoform A and B. Almost all strains had contig breaks in the LPS outer core biosynthesis locus; however, most of the *P. multocida* genomes had matches to outer core biosynthesis genes allowing them to be accurately typed as one of the eight known LPS genotypes. There were 19 strains were designated type L1 with a match to *hptE*_X73, 1 strain was designated type L2 with a match to *hptE*_M1404, 14 strains designated type L3 with a match to *hptE*_P1059 and one strain designated L4 with a match to *natE* (Figure 3.3). Only seven strains had a complete LPS Past5, Past13, Pm1613, Pm1620, Pmc-A, Pmc-C, PI32

Interestingly, Past29 was typed as L3 due to a match for *hptE*_P1059, but also had matches for *gatH* and *gctD* from genotype L6, suggesting possible horizontal transfer of LPS genes in this strain. Most strains had contig breaks in the LPS outer core biosynthesis locus, causing several missing genes or low amino acid identity matches, with only seven strains having a complete LPS outer core locus (Figure 3.3). Strains Pm1613, Pm1620 and PI32 had matches to GatG (Pm1613 and Pm1620) and/or NatC (Pm1613, Pm1620 and PI32) that had low amino acid identity. Matches to *gatC* and/or *natC* in these strains contained deletions that introduced

premature nonsense or frameshift mutations. As such, the encoded protein would likely be non-functional, and would result in altered LPS outer core structure.

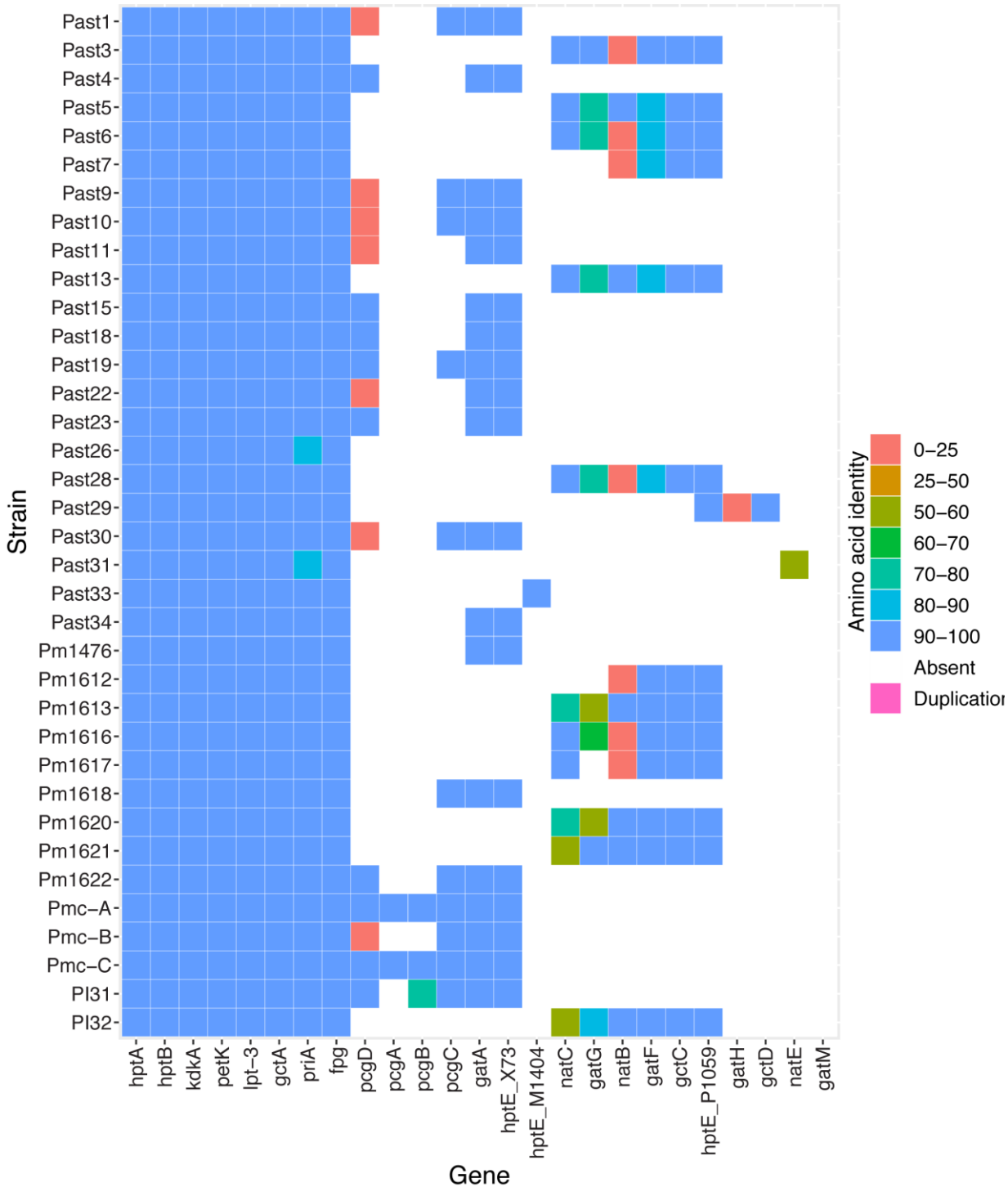


Figure 3.3: Heatmap showing presence of *P. multocida* LPS biosynthesis proteins in the *P. multocida* isolates sequenced in this study identified by Assembly2Gene (See Appendix 3.6 for full PastyVRDB). Coloured squares represent the amino acid identity between proteins encoded by query and reference sequences, or for when there are multiple matches in a query genome for a reference gene, a gene duplication. Isolate genomes and/or reference genes with no matches were excluded from the heat map.

Virulence factor and outer membrane protein genes in the PastyVRDB included encoding filamentous haemagglutinin and secretion partners (*fhaB1-fhaC1, fhaB2-fhaC2*), type four fimbriae (*ptfA_1* and *ptfA_2*), Fimbrial low molecular weight (Flp) pili assembly genes from the Tad-locus (*tadGFEDCBAZ-rcpBAC-tadV-flp12*), outer membrane proteins (*ompA*, *ompH_1*, *ompH_2*, *oma87*, *comE1*, *plpB*, *plpE*), sialic acid scavenging and uptake (*nanB*, *nanH*, *nanU*, *nanP*), *Pasteurella multocida* toxin (*toxA*),superoxide dismutase (*sodA*, *sodC*), and regulatory protein Fis. Most of the *P. multocida* strains encoded orthologs of multiple adhesin proteins (Figure 3.4). All *P. multocida* isolates had matches to one filamentous haemagglutinin gene and corresponding secretion partner (Figure 3.4). Past15 had a C3451T nucleotide mutation in *fhaB1* resulting in a nonsense mutation. All matches to FhaC1 and FhaC2 were between 61% and 68% amino acid identity. Alignments indicated query *fhaC1* genes were 279 bp longer than the Pm70 *fhaC1* gene, and query *fhaC2* genes were 315 bp longer than the Pm70 *fhaC2* gene. The Pm70 *fhaC1* and *fhaC2* genes have in-frame ORFs upstream of each gene, suggesting the genes are longer in the isolate strains than in Pm70, or that the Pm70 genes are mis-annotated. All strains had a match for putative type IV fimbrial genes *ptfA_1* and *ptfA_2* (Figure 3.4). A variant PtfA_1 was identified in eight of the isolates that was identical to reference PtfA_1 for the first 67 amino acids but contained 30 missense mutations in the remaining third of the protein.

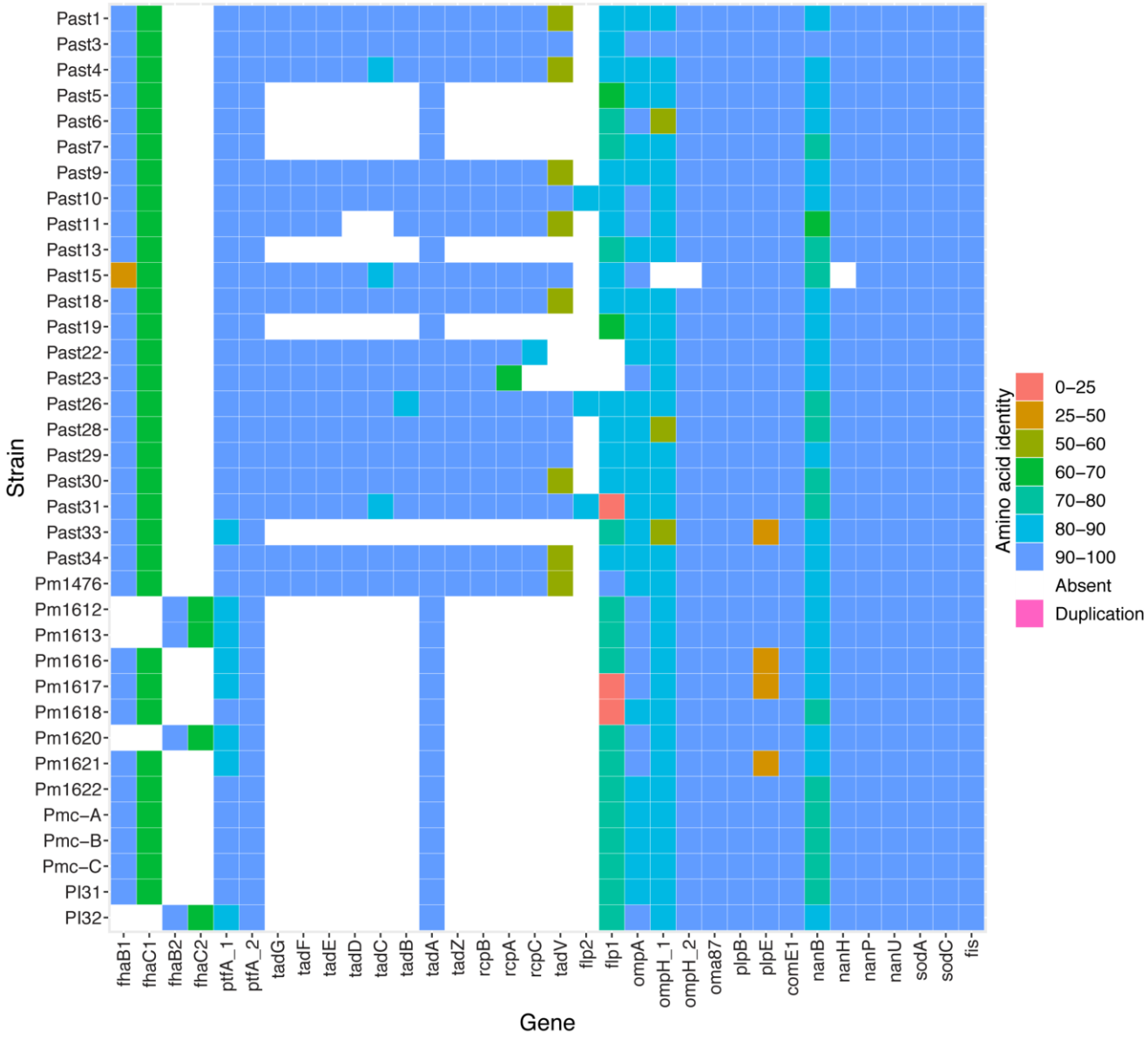


Figure 3.4: Heatmap showing presence of *P. multocida* adhesins, outer membrane proteins and virulence factors in the *P. multocida* isolates sequenced in this study identified by Assembly2Gene (See Appendix 3.6 for full PastyVRDB). Coloured squares represent the amino acid identity between proteins encoded by query and reference sequences, or for when there are multiple matches in a query genome for a reference gene, a gene duplication. Isolate genomes and/or reference genes with

A putative Tad-locus was identified in 15 isolates (Figure 3.4). Several strains with a putative Flp-pili biosynthesis system had contig breaks within the Tad-locus corresponding to missing genes. All strains, excluding Pm1476 and Past31, contained a Flp1 variant with between 81% and 87% amino acid identity to reference Flp1 from strain HB03. The query sequence from these strains matching *flp1* contained between 9 and 11 nucleotide polymorphisms in the 3' end of the gene that caused missense mutations. Eight isolates encoded a predicted TadV with 53% amino acid identity to reference TadV from strain HB03. The predicted *tadV* gene in these isolates contained an additional 102 bp at the start of the gene compared to reference *tadV* from strain HB03. Eighteen of the isolates encoded putative homologs of Flp1 and TadA, but no homologs of other components encoded by the Tad-locus. Genomes of these strains were searched using the reference *tad*-genes using BLASTn to determine if other *tad*-genes were present at below the 80% nucleotide identity cut-off used for the Assembly2Gene analysis. Indeed, *tadBCDEF* and *rcpCA* were identified in these strains with between 65% and 75% nucleotide identity. Although matches were found for most of the *tad*-genes, no matches were identified for *tadV* and *rcpB* in these strains. In addition, the Tad-locus identified in these isolates was intact, so lack of *tadV* and *rcpB* was not due to contig breaks. The lack of *tadV* and *rcpB* suggests these strains do not have a functional Flp-pili biosynthesis system.

Almost all the *P. multocida* strains encoded orthologs of OmpH_2, Oma87, PlpB, PlpE and ComE1 with high amino acid identity (>90%) (Figure 3.4). Four isolates had a predicted PlpE with only 26% amino acid identity to reference PlpE from Pm70. The query sequence in these strains displayed 95% nucleotide identity to reference *plpE* but have several conserved indels over the entire sequence resulting in a frameshift mutation. There was considerable variability observed for matches to outer membrane proteins OmpA and OmpH_1 in the *P. multocida* isolates. All strains encoded a predicted OmpA that had > 87% amino acid identity with the reference OmpA; however, four conserved OmpA variants were identified. In each variant, the query sequence matching the reference *ompA* gene contained conserved indels and nucleotide polymorphisms causing missense mutations in sections of the gene encoding external loops (133). Investigation of *ompA* sequence in 190 *P. multocida* isolates identified 25 unique *ompA* sequences, showing *ompA* variability is common in *P. multocida* (245). Similarly, several OmpH_1 variants were identified, with the query sequences matching reference *ompH_1* containing indels and nucleotide polymorphisms causing missense mutations in regions encoding external protein loops (246).

All *P. multocida* strains, except for Past15, encoded predicted orthologs of the sialic acid scavenging and uptake proteins NanB, NanH, NanU, and NanP; Past15 encoded a predicted ortholog of all but NanH (Figure 3.4). The sequences matching NanH, NanP and NanU all displayed >90% amino acid identity to their corresponding reference protein, while orthologs

of NanB were all between 70% and 83% identical to the NanB reference. Three variant NanB orthologs were identified. The query sequence matching to reference *nanB* for each variant contained either 13, 15 or 9 conserved indels and several nucleotide polymorphisms resulting in missense mutations. All *P. multocida* isolates had high amino acid identity homologs of the super oxide dismutases SodA and SodC, and for the global transcriptional regulatory protein Fis. None of the isolates encoded a *Pasteurella multocida* toxin (PMT) homolog, indicating that these strains do not carry the lysogenic phage that carries *toxA* (109).

Iron acquisition genes included in the PastyVRDB included genes encoding outer membrane iron receptors (*hgbA*, *hgbB*, *hpbA*, *pfhR*, *hasR*, *hemR*, Pm0741, Pm1081, Pm1282, Pm1428, *tpbA*), the TonB energy transduction system (*tonB-exbBD*) and iron-specific ABC-transport systems (*afuA_1A_2A_3BC*, *fbpABC*, *fecBCDE*, *yfeABCD*). All *P. multocida* isolates encoded proteins with >90% amino acid identity to TonB energy transduction system proteins TonB, ExbB, ExbD, all iron-specific ABC-transport system proteins and iron receptors HpbA, PfhR, Pm0741 and Pm1081 (Figure 3.5). Several isolates did not have matches to between one and four of the reference iron receptors (Figure 3.5). Most of the iron receptor orthologs identified in the *P. multocida* isolates had high amino acid identity to the corresponding reference, excluding HgbA. Several variant HgbA proteins were identified in the isolates. The query sequence matching *hgbA* showed several indels and nucleotide polymorphisms starting from nucleotide 810 until nucleotide 2,019, with the 5' and 3' ends of predicted *hgbA* having high nucleotide identity with reference *hgbA*. Four isolates had a match to *hgbA* that had high nucleotide identity over the first 432 bp of the gene compared to the *hgbA* reference but contained a large number of nucleotide polymorphisms causing missense mutations over the remainder of the gene. In addition, Strains Past6 and Past33 had mutations in matches for iron-receptor genes that would result in non-functional proteins. Past6 had a C1054T single nucleotide polymorphism (SNP) in the query sequence matching to reference *Pm1081* causing a nonsense mutation, as well as a one bp deletion in query sequence matching to reference *Pm1428* at bp 52 resulting in a nonsense mutation shortly downstream. Past15 had a 1037 bp gap in the query sequence matching to reference *Pm1428* starting at nucleotide 357.

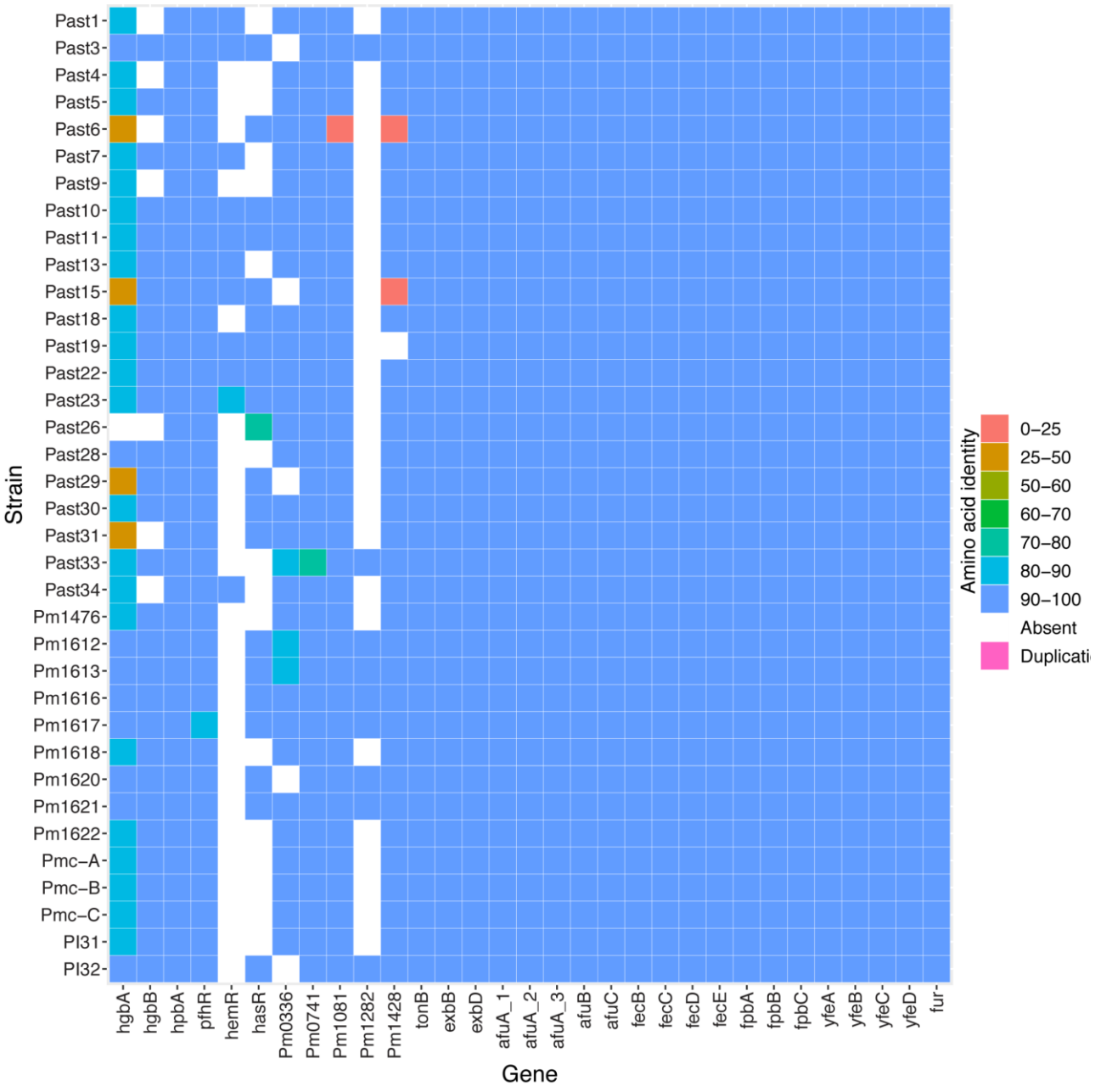


Figure 3.5: Heatmap showing presence of *P. multocida* iron-acquisition proteins in the *P. multocida* isolates sequenced in this study identified by Assembly2Gene (See Appendix 3.6 for full PastyVRDB). Coloured squares represent the amino acid identity between proteins encoded by query and reference sequences, or for when there are multiple matches in a query genome for a reference gene, a gene duplication. Isolate genomes and/or reference genes with no matches were excluded from the heat map.

3.2.4 *P. multocida* trait-associated genes

P. multocida strains can cause a wide range of diseases in different hosts. Apart from capsule and LPS type, little is known about the mechanisms that determine *P. multocida* host predilection. In an attempt to elucidate mechanisms of host specificity, genes/proteins associated with different *P. multocida* traits were identified. Traits included capsule and LPS genotypes, as well as the host each *P. multocida* strain was isolated from. Trait-associated genes were identified using all 36 *P. multocida* isolates sequenced in this study, and 260 draft and complete reference *P. multocida* reference genomes downloaded from the PATRIC database (downloaded Dec 2020). Capsule and LPS genotypes for all genomes was first identified *in silico* using Assembly2Gene. Unique capsule and LPS outer core biosynthesis genes that are used as targets in multiplex PCR identification of capsule and LPS type were used as references for genotyping (9, 11). Of the 288 *P. multocida* strains that were successfully capsule genotyped, 208 were type A, 30 were type B, 27 were type D and 23 were type F, with no strains typed as capsule genotype E. For the 277 LPS genotyped strains, 41 were L1, 32 were L2, 139 were L3 and 62 were L6. There was only one strain typed as L4, L5 and L7 respectively, with no strains typed as L8. There was a strong correlation between capsule types B, D and F with LPS genotypes L2, L6 and L3 respectively, with only five genotype D and F strains having a different LPS genotype.

Genes associated with different traits were identified using Roary and Scoary. Roary performs core- and pan-genome identification by grouping all genes with greater than 95% amino acid identity into gene groups. Roary requires an annotated genome as an input, so all reference genomes were re-annotated using Prokka to ensure all genome annotations were consistent. The core *P. multocida* genome consisted of 1,622 genes, with the pan-genome containing 7,537 genes. Scoary uses gene presence/absence tables from Roary analysis and identifies associations between traits and genes in the pan-genome. A gene was called as associated with a trait if there was a greater than 90% presence of the gene in strains with that trait, and lower than 10% presence for strains without that trait. To ensure trait-associated genes did not have homologs below 95% amino acid identity in other gene groups, trait-associated genes were compared to the pan-genome using BLASTn. Any trait-associated gene with a match to another gene group at > 50% nucleotide identity over 50% of the full gene length were not included as trait-specific genes. All capsule and LPS biosynthesis genes were excluded from the analysis. Capsule genotypes A and D; LPS genotypes L1, L3, L4, L6 and L7; and all hosts had no associated genes under these parameters. Functions of trait-associated genes were predicted by comparison with the UniProt database using BLASTp, or to the NCBI Conserved Domain Database. There were 10 genes associated with capsule genotype F strains (Table 3.4). Genes associated with capsule genotype F included two gene

groups containing tyrosine recombinases including group_961 (p = 2.44 x 10^{-16}) and xerC_1 (p = 6.74 x 10^{-15}), gene groups involved in arsenate or heavy metal resistance, groups *arsC* (p = 5.13 x 10^{-19}), *hmrR* (p = 3.59 x 10^{-16}) and *zitB* (p = 3.59 x 10^{-16}), and group_3956 (p = 8.23 x 10^{-22}) that encodes a protein containing a DNA binding domain. Capsule genotype B and LPS genotype L2 had 21 and 26 trait-associated gene groups, respectively. All of the genotype L2-associated genes were also associated with genotype B genes (Table 3.5). Most of the type B:L2-associated genes were predicted to encode hypothetical proteins, or bacteriophage components. Group_2899 (p =8.75 x 10^{-37}) encoded a protein with 30% amino acid identity to the putative lipoprotein Pm1514 from *P. multocida* strain Pm70. The protein encoded by the group_2899 reference gene contained a lipoprotein signal peptide identified by SignalP-5.0, further indicating this gene group encodes a lipoprotein. Group *wbnI* (p = 8.75 x 10^{-37}) encoded a protein containing a glycosyltransferase A family domain.

Table 3.4: Genes associated with *P. multocida* capsule genotype type F strains

Roary gene group	Gene name	Predicted protein function	Percentage presence in type F strains	Percentage presence non-type F strains	Bonferroni adjusted P-value
group_3924		Hypothetical protein, contains BrnA antitoxin domain	95.65	2.26	1.07E-22
group_2272		Hypothetical protein	95.65	3.01	8.23E-22
group_3956		Hypothetical protein, contains LysR DNA binding transcriptional regulator domain	95.65	3.77	6.00E-21
arsC	*arsC*	Arsenate reductase	95.65	5.66	5.13E-19
group_379		Hypothetical protein	95.65	9.43	2.44E-16
group_961		Tyrosine recombinase	95.65	9.43	2.44E-16
group_3806		Hypothetical protein, contains a DUF5376 domain	95.65	9.81	3.59E-16
hmrR		Hypothetical protein, contains a MerR domain found in transcriptional regulators of genes involved in heavy metal resistance or oxidative stress	95.65	9.81	3.59E-16
zitB		Hypothetical protein, contains a CzcD domain found in Zn, Co and Cu efflux proteins	95.65	9.81	3.59E-16
xerC_1		Tyrosine recombinase	91.30	9.43	6.74E-15

Table 3.5: Genes associated with *P. multocida* capsule genotype B and LPS genotype L2 strains

Roary gene group	Gene name	Predicted protein function	Percentage presence in B:L2 strains	Percentage presence in non-B:L2 strains	Bonferroni adjusted P-value
group_1186		Hypothetical protein	96.67	6.20	1.73E-25
group_1489		Hypothetical protein, contains a DUF2075 domain	96.67	1.55	2.75E-32
group_1490		Hypothetical protein	93.34	7.36	1.60E-22
group_1495		Putative bacteriophage tape measure protein	96.67	1.93	1.73E-31
group_1810		Hypothetical protein, contains a DUF2681 domain	93.34	5.03	5.02E-25
group_1813	*gp47*	Mu-like bacteriophage baseplate protein	93.34	5.03	5.02E-25
group_1814		Hypothetical protein	96.67	2.32	9.50E-31
group_1815		Hypothetical protein	93.34	5.81	4.02E-24
group_2285		Hypothetical protein	96.67	5.03	8.63E-27
group_251		Phage head maturation protease	96.67	8.13	1.32E-23
group_2874		Hypothetical protein	93.34	5.03	5.02E-25
group_2881		Hypothetical protein, contains a DUF2313 domain	93.34	5.03	5.02E-25
group_2893		Hypothetical protein, contains a phage tail U domain	96.67	2.71	4.48E-30
group_2899		Putative lipoprotein	96.67	0	8.75E-37
group_3976		Hypothetical protein, contains a DUF3486 domain	90	0	8.04E-33
group_3980		Hypothetical protein	96.67	0	8.75E-37
group_3981		Hypothetical protein	96.67	5.03	8.63E-27
group_3982		Hypothetical protein	96.67	5.03	8.63E-27
group_3983		Hypothetical protein	96.67	5.03	8.63E-27
group_3989		Hypothetical protein	96.67	3.48	7.68E-29
group_601		Mu-like bacteriophage DNA circularisation	96.67	5.03	8.63E-27
group_757		Hypothetical protein	96.67	7.36	2.52E-24
group_938		Bacteriophage portal protein	96.67	8.13	1.32E-23
group_964		Putative Mu-like bacteriophage tail fibre protein	96.67	2.32	9.50E-31
intA_4	*intA*	Prophage integrase	96.67	5.03	8.63E-27
wbnI		Glycosyltransferase family A	96.67	0	8.75E-37
group_592		Bacteriophage capsid protein	96.67	6.20	1.14E-23

3.2.5 Phylogenies of *P. multocida*, the *Pasteurella* genus and the Pasteurellaceae family

Roary analysis of all isolate and reference *P. multocida* genomes described above produced a core-genome alignment of 290,493 bp in regions common to all *P. multocida* strains. The *P. multocida* core-genome alignment was used to produce a phylogenetic tree using IQ-TREE, generating a maximum-likelihood phylogeny (Figure 3.6 and Appendix 3.11). The country of isolation, host strain, capsule genotype and LPS genotype for all strains were annotated on the tree (full annotation file in Appendix 3.12). Given the number of strains included, bootstrapping values could not be annotated on the tree but are given in appendix 3.11. The *P. multocida* core-genome phylogeny showed two main clades, designated clade A and clade B (Figure 3.6). A previous 16S rRNA *Pasteurellaceae* family phylogeny showed *P. multocida* subsp. *septica* strains were distinct from subsp. *multocida* and *gallicida* strains (234). *P. multocida* subsp. *septica* strains were identified by 16S rRNA gene sequence comparison; however, this could not be performed as several of the whole-genome sequences used were incomplete and lacked an intact 16s rRNA gene sequence. ANI analysis can be used to identify subspecies (247, 248), with values over 98 between a reference and query suggesting both strains are the same subspecies. To determine if this was true for *P. multocida* subspecies, fastANI was run to compare *P. multocida* subsp. type strains NCTC 10204 (subsp. *gallicida*), NCTC 10322 (subsp. *multocida*), and NCTC 11995 (subsp. *septica*) (234). The fastANI analysis gave ANI values below 98 when comparing NCTC 11995 (subsp. *septica*) to both NCTC 10204 (subsp. *gallicida*) and NCTC 10322 (subsp. *multocida*), but ANI values over 98 when comparing NCTC 10204 (subsp. *gallicida*) and NCTC 10322 (subsp. *multocida*) (data not shown). This allowed fastANI to identify subsp. septica strains, but did not allow differentiation between subsp. *multocida* and *gallicida* strains. All *P. multocida* genomes used in this study were compared using fastANI to the subsp. type strains NCTC 10204, NCTC 10322, and NCTC 11995, identifying all strains in clade A as subsp. *septica* with no strains in clade B identified as subsp. *septica* (data not shown). Strains in clade A were isolated primarily from humans, cats and dogs, with 26 out of 32 *P. multocida* human isolates clustering into clade A. Clade B contained primarily animal isolates. Five subclades were observed in clade B that contained closely related strains with the same capsule and LPS groups (subclades A:L3, B:L2, A:L6, D:L6, A:L1). Core genome SNPs were identified in all 296 *P. multocida* strains compared to strain VP161 using Parsnp. Strains that clustered into each subclade had similar core-genome SNP counts when compared to VP161. Core-genome SNPs were identified between the most distant strains from each subclade: strains 2597PM and P1062 from subclade A:L3 had 506 core-genome SNPs; strains FDAARGOS_217 and Islm from subclade B:L2 had 175 SNPs; strains SH03 and 4407 from subclade A:L6 had 175 SNPs; strains SH02 and HND04 from subclade D:L6 had 228 SNPs; and strains PMTB2.1

and HN141014 had 117 SNPs. Collectively, these SNP counts showed strains were highly related in each of the subclades. Most of these subclades contained isolates recovered from multiple animal hosts, further suggesting that closely related strains of *P. multocida* can cause similar diseases in different animal hosts.

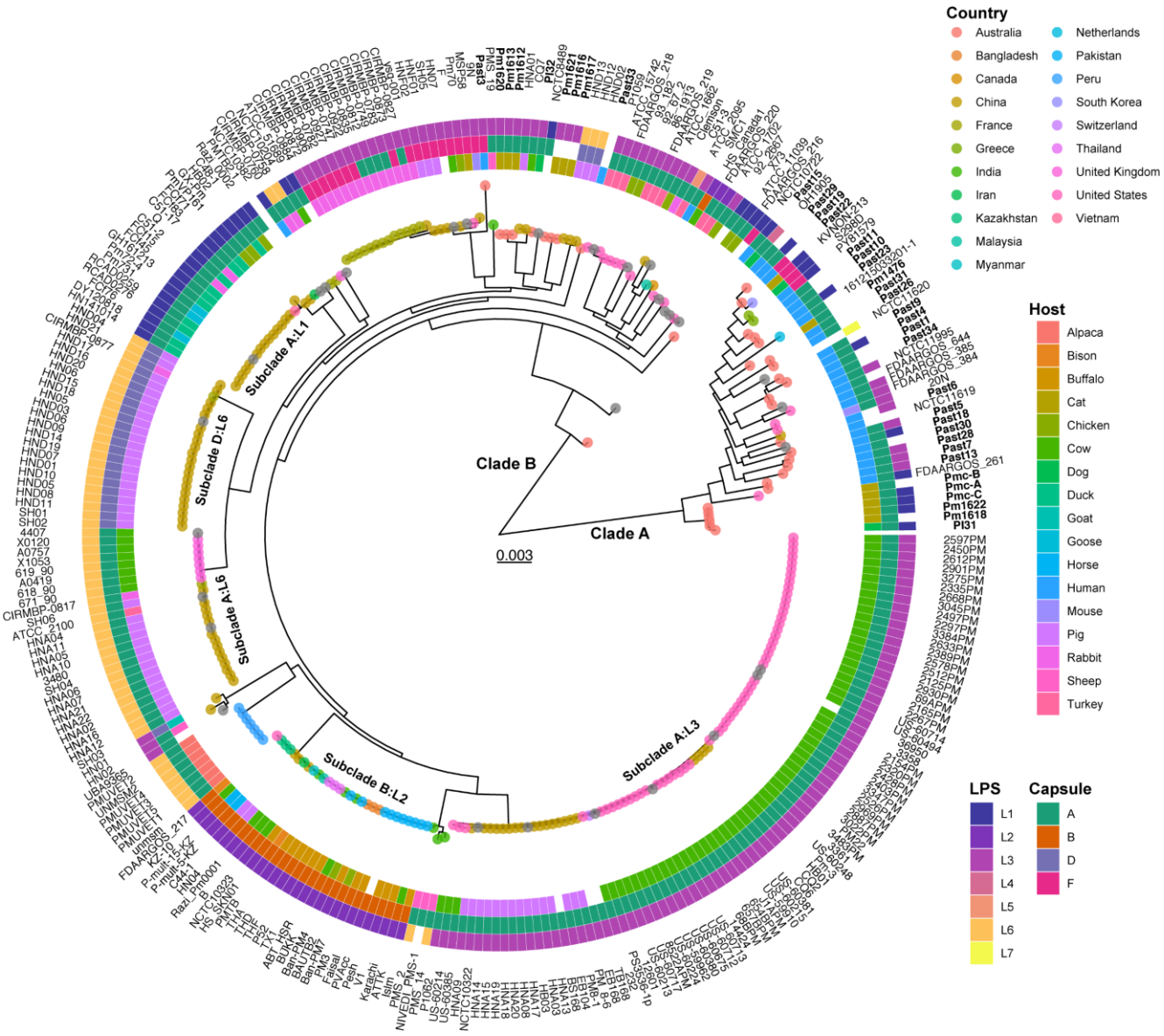


Figure 3.6: Maximum likelihood phylogenetic tree of the core-genome alignment from 296 *P. multocida* reference and isolate genomes. Roary was used to generate a core-genome alignment of 260 *P. multocida* complete and draft reference genomes from the PATRIC database and all 36 *P. multocida* isolates generated in this study, with the strain names of isolates sequenced in this study in bold. The core-genome alignment was used to generate a phylogenetic tree using IQ-TREE with a general time reversible nucleotide substitution model with rate heterogeneity modelled with empirical base frequencies and a FreeRate model distribution (GTR+F+R8). Branch supports were estimated using 1,000 bootstrap replicates (Appendix 3.11). References and isolates were capsule and LPS typed *in silico* using Assembly2Gene to identify capsule and LPS type-specific genes. The inner ring shows the host the strain was isolated from, the middle ring shows the capsule genotype, and the outer ring shows the LPS genotype. Colour on tips shows country of origin. Scale bar represents number of nucleotide substitutions per site.

Given most of the human isolates clustered into clade A, Scoary analysis was used to identify genes associated with either clade A or clade B. There were 14 genes associated with clade A (Table 3.6). Most of the clade A-associated genes were involved in carbohydrate uptake or utilisation. The genes *mglA_1* (p = 2.76 x 10^{-27}), *rbsB_3* (p = 2.76 x 10^{-27}) and group_2709 (p = 2.76 x 10^{-27}) all encoded proteins that contained domains found in monosaccharide ABC-transport systems. Several L-fucose utilisation genes were also identified as associated with clade A, including *fucUARKI* and *aldA* (p-values between 2.76 x 10^{-27} and 8.90 x 10^{-41}). The ability to utilise L-fucose and possibly other carbohydrates as a carbon source may give clade A strains a fitness advantage over clade B strains in humans. There were 12 genes associated with clade B (Table 3.7). Four of the clade B-associated genes were involved in anaerobic respiration and carbon source uptake. The *torCDA* operon was associated with clade B (all had a p-value of 2.17 x 10^{-39}), which is predicted to encode a trimethylamine-N-oxide anaerobic respiration system. The *dcuD_1* gene (p = 1.87 x 10^{-29}) encoded a protein with 74% amino acid identity to an anaerobic C4-carboxylate transporter from *E. coli* K12. The *macA* (p = 2.09 x 10^{-26}) gene, which encoded a predicted component of a MacAB-TolC macrolide efflux system was also associated significantly with clade B.

Table 3.6: Genes associated with *P. multocida* clade A strains

Roary gene group	Gene name	Predicted protein function	Percentage presence in clade A strains	Percentage presence in clade B strains	Bonferroni adjusted P-value
group_3551		Hypothetical protein, contains dinD and Bro-N domains found in DNA binding proteins	97.22	0.38	8.90E-41
fucU	*fucU*	L-fucose mutarotase	97.22	0.38	8.90E-41
hscC	*hscC*	Sigma70 regulatory protein	97.22	0.38	8.90E-41
group_6074		Hypothetical protein	97.22	0.38	2.76E-27
group_3579		Hypothetical	97.22	0.38	2.76E-27
aldA	*aldA*	Aldehyde dehydrogenase A	97.22	8.46	2.76E-27
mglA_1		Hypothetical protein, contains a MglA domain found in monsaccharide ABC-transport system proteins	97.22	8.46	2.76E-27
rbsB_3		Hypothetical protein, contains an ABC sugar binding-like domain found in monosaccharide ABC-transport system proteins	97.22	8.46	2.76E-27
group_2709		Hypothetical protein, contains an AraH permease domain found in monosaccharide ABC-transport system proteins	97.22	8.46	2.76E-27
fucA	*fucA*	L-fuculose-phosphate aldolase	97.22	8.46	2.76E-27
ydjF	*fucR*	L-fucose operon regulator	97.22	8.46	5.57E-37
fucK	*fucK*	L-fuculokinase	97.22	8.46	8.90E-41
fucI	*fucI*	L-fucose isomerase	97.22	8.84	6.84E-27
group_2118	*tnpA1*	IS200 transposase	91.66	0.38	8.90E-41

Table 3.7: Genes associated with *P. multocida* clade B strains

Roary gene group	Gene name	Predicted protein function	Percentage presence in clade B	Percentage presence in clade A	Bonferroni adjusted P-value
torA	*torA*	Trimethylamine-N-oxide reductase	98.84	0	2.17E-39
torD	*torD*	Trimethylamine-N-oxide reductase-specific chaperone	98.84	0	2.17E-39
torC	*torC*	C-type cytochrome	98.84	0	2.17E-39
group_1879		Hypothetical protein	98.46	0	2.17E-38
group_3340		Hypothetical protein	98.46	0	2.17E-38
group_44		Hypothetical protein	99.61	5.55	2.82E-37
cah	*cah*	Putative alpha carbonic anhydrase	99.61	8.33	2.40E-35
group_838		Hypothetical protein	97.30	2.77	1.63E-33
dcuD_1		Anaerobic C4-dicarboxylate transporter	96.15	5.55	1.87E-29
btuB_4		Hypothetical protein, contains a CirA domain found in outer membrane receptor proteins	93.07	0	2.30E-29
yeaD_2	*yeaD*	D-hexose-6-phosphate mutarotase	90	0	4.99E-26
macA	*macA*	Macrolide efflux system protein	90.83	0	2.09E-26

To assess the phylogenetic relationships of all strains sequenced in this project, a *Pasteurella* genus sketch-distance-based tree was generated using Mashtree. Mashtree compiles the most common 10,000 kmers present in a genome into a sketch and compares sketches between all genomes to determine the sketch-distance between two genomes. All isolates sequenced in this study and 338 complete and draft *Pasteurella* spp. reference genomes downloaded from the PATRIC database (downloaded Dec 2020) were used. References sequences included [*P*.] *aerogenes*, [*P*.] *bettyae*, [*P*.] *caecimuris*, *P. canis*, *P. dagmatis*, [*P*.] *langaaenisis*, [*P*.] *mairii*, *P. oralis*, [*P*.] *skynesis* and [*P*.] *testudinis*. The *F. canicola* isolates sequenced in this study were included as out-groups. The *Pasteurella* genus tree showed *P. canis*, *P. dagmatis*, *P. multocida, P. oralis* and *P. stomatis* clustered together away from [*P*.] *aerogenes*, [*P*.] *bettyae*, [*P*.] *caecimuris*, [*P*.] *langaaenisis*, [*P*.] *mairii*, [*P*.] *skynesis* and [*P*.] *testudinis* and *F. canicola* strains, supporting the proposal that *Pasteurella* spp. *canis*, *dagmatis*, *multocida*, *oralis* and *stomatis* are likely the only members of *Pasteurella sensu stricto* (Figure 3.7). Reference strain OH4807 was listed as *P. multocida* in the PATRIC database, and although clustered within the *Pasteurella sensu stricto* clade, was not closely related to any of the other species in this clade. In addition, ANI comparison of strain OH4807 with other *Pasteurella sensu stricto* strains used for this phylogeny did not have a match over > 95 (data not shown). These data suggest that strain OH4807 may be [*P.*] *caballi* that may be a true *Pasteurella* species, or a novel *Pasteurella* species. To further assess whether strains typed as [*Pasteurella*] spp. *aerogenes*, *bettyae*, *caecimuris*, *langaaensis*, *mairii*, *skynesis, testudinis,* and *F. canicola* represented true *Pasteurella* spp. a Pasteurellaceae sketch-distance-phylogeny was generated using Mashtree. The Pasteurellaceae tree was generated using all isolates sequenced in this study, all complete Pasteurellaceae reference genomes from NCBI (downloaded Dec 2020) and all *Pasteurella* genus type strains. The Pasteurellaceae tree showed good separation of genera (Figure 3.8). As with the *Pasteurella* genus distance-based tree, *Pasteurella* spp. *multocida*, *canis*, *dagmatis*, *stomatis* and *oralis* and strain OH4807 clustered together away from other bacterial species, strongly suggesting these are the only true *Pasteurella* species.

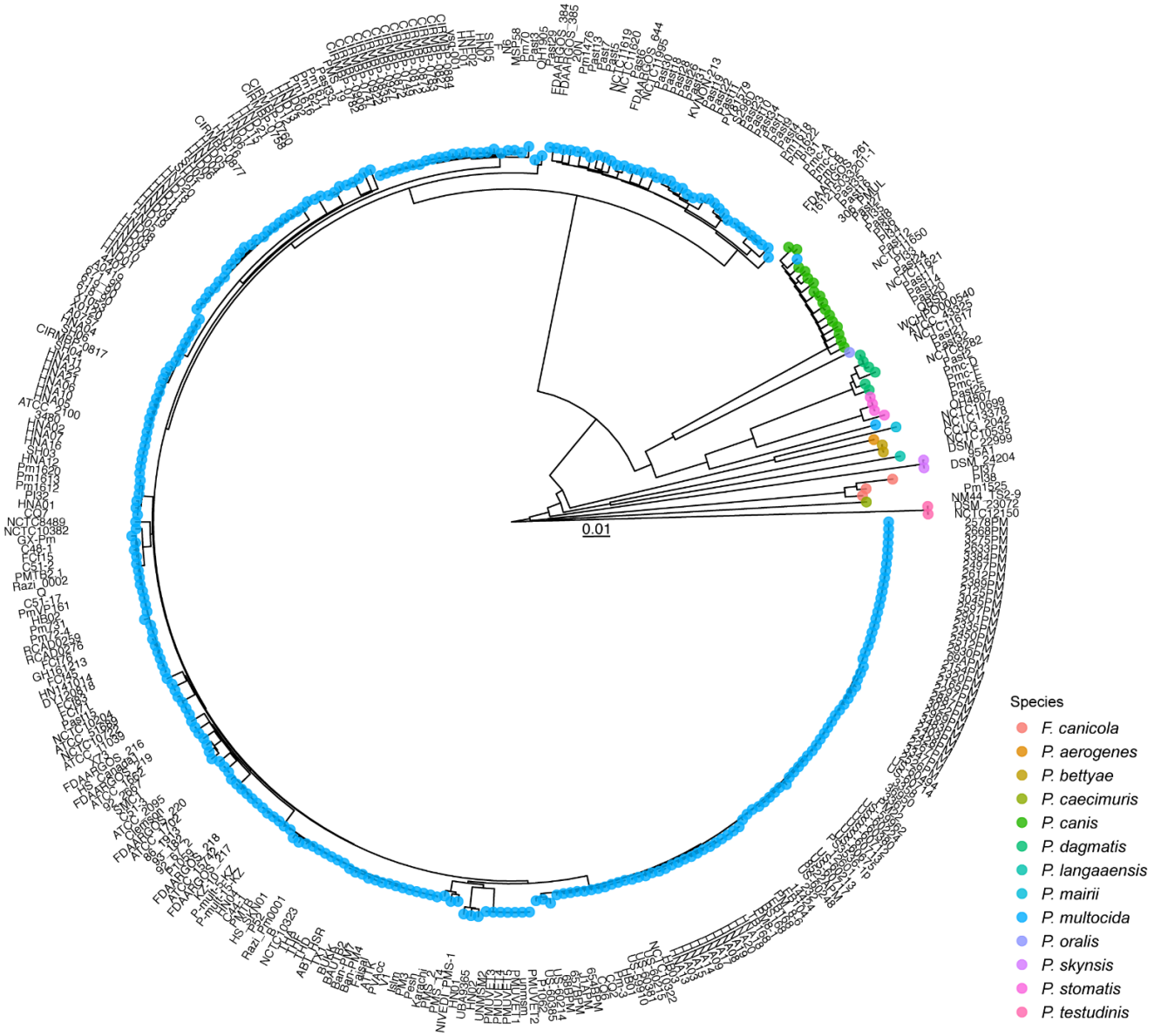


Figure 3.7: Neighbour-joining tree of sketch-distances between 338 *Pasteurella* spp. reference and isolate genomes generated using Mashtree. Sketches were generated by Mash for all genomes based on 10,000 21 bp kmers, and distances between sketches identified using the min-hash algorithm. The neighbour-joining tree was generated using the QuickTree neighbour joining algorithm. 279 complete and draft *Pasteurella* spp. reference genomes downloaded from the PATRIC database, and all 56 *Pasteurella* spp. isolate genomes generated in this study were used. The three *F. canicola* isolates were included as out groups. Branch supports were estimated using 1,000 bootstrap replicates (Appendix 3.13). Scale bar represents number of nucleotide substitutions per site.

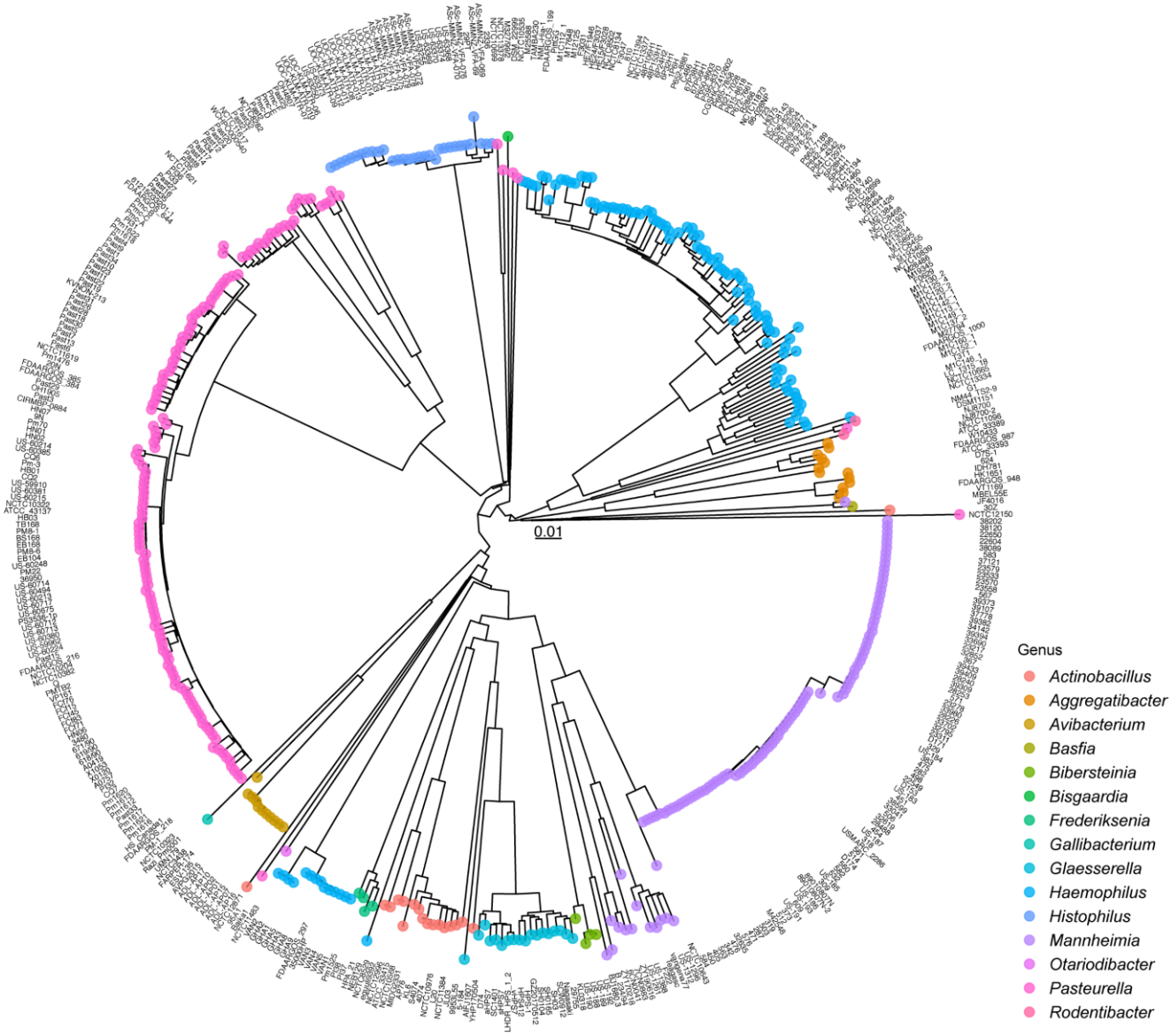


Figure 3.8: Neighbour joining tree of sketch-distances between 448 Pasteurellaceae reference and isolate genomes generated using Mashtree. Sketches were generated by Mash for all genomes based on 10,000 kmers, and distances between sketches identified using the min-hash algorithm. Neighbour joining tree was generated using the QuickTree neighbour joining algorithm. 389 complete Pasteurellaceae reference genomes downloaded from NCBI, and all isolate genomes generated in this study were used. Branch supports were estimated using 1,000 bootstrap replicates (Appendix 3.14). Scale bar represents number of nucleotide substitutions per site.

3.3 Discussion

This study aimed to investigate *P. multocida* strains capable of causing human disease, and to identify possible mechanisms behind *P. multocida* animal host predilection. Bioinformatic analysis was performed using 59 human and cat *Pasteurella* spp. isolates, and 275 publicly available *P. multocida* reference genomes. There are currently eight publicly available genome sequences for *P. multocida* isolates from human infections. To allow for more robust bioinformatic analysis, 22 human *P. multocida* isolates were obtained from Monash Medical Centre. In addition, eight *P. canis* and five *P. dagmatis* human isolates from Monash Medical Centre, 16 cat and eight dog *P. multocida* isolates were obtained from the University of Queensland. All isolates were whole genome sequenced via Illumina sequencing. *De novo* genome assembly produced high quality genomes, indicated by high coverage and high quality contig data (Table 3.1). Initially, isolates were identified to species level using MALDI-TOF, Microbact or biochemical tests. These identifications were checked using *P. multocida*-specific KMT PCR (249), with ten of the isolates originally typed as *P. multocida* not producing a KMT PCR product (Figure 3.1). Species identification was also performed *in silico* using a range of methods. FastANI was first used to call bacterial species. Comparisons to *Pasteurella* spp. type strains, or all complete Pasteurellaceae reference genomes in the NCBI database identified all strains as either *P. multocida*, *P. canis*, *P. dagmatis* or *F. canicola* (Table 3.2, Table 3.3). Strains Pmc-D, Pmc-E, Pmc-F and Past25 had a highest ANI score with *P. dagmatis* type strain NCTC8282, but lower than the required score of 95 to be confident that these strains were indeed *P. dagmatis*. Whole-gene 16s rRNA identity was used to identify these species. The 16s rRNA from Pmc-E displayed 98.61% identity to 16s rRNA from *P. stomatis*. Other studies investigating species identification using whole-gene 16s rRNA identity have suggested a 98.65% to 99% identity for species calling (250, 251). Although the match was just under these thresholds, the lack of ANI matches to other Pasteurellaceae references suggests these strains are indeed *P. stomatis*.

ABRicate was used to identify homologs of well characterised virulence factor and resistance genes in the newly sequenced isolates. All isolates had matches for *vatE*, *tet34*, and *hmrM* when searched against the ResFinder and CARD databases. The genes *vatE*, *tet34*, and *hmrM* encode proteins that confer resistance to streptogramin, oxytetracycline and several antibiotics, respectively (252-254). However, the matches for these genes displayed low nucleotide identity, indicating these genes may no longer function in antibiotic resistance in the *Pasteurella* spp. isolates. Searches against the VFDB identified several virulence factors that have yet to be identified in *P. multocida*. Several of the isolates encoded a protein with identity to IlpA from *V. vulnificus*, which is required for bacterial adhesion to human intestinal epithelial cells (255). IlpA displays 58% amino acid identity to lipoprotein PlpB from

P. multocida Pm70, indicating that these matches to IlpA were likely PlpB homologs. PlpB has been reported to function as a methionine uptake protein (132), but given the identity to IlpA, it is possible that PlpB could also function as an adhesin in *P. multocida*. All strains had a 71% to 74% nucleotide identity hit to *luxS*. The LuxS protein is a type-2 autoinducer synthase (256). Homologs of the *lsr* quorum sensing operon from *E. coli* K12 were identified in all *Pasteurella* isolates, indicating these strains likely have a functional Lsr-quorum sensing system. All isolates encoded proteins with significant identity to HtpB from *Legionella pneumophila* and ClpP from *Lysteria monocytogenes.* HtpB (GroEL) and ClpP are a protein chaperone and protease respectively, required for correct protein folding and degradation of a large number of bacterial proteins (257, 258). Both HtpB and ClpP are included in the VRDB as these proteins have been shown to be required for *L. monocytogenes* and *L. pneumophila* cellular invasion (259, 260). As there is currently limited evidence of *P. multocida* invasion and survival within host cells, HptB and ClpP are unlikely to have important roles in virulence in *P. multocida*.

To identify known *P. multocida* virulence factor and antibiotic resistance genes in *P. multocida* isolates, a curated database was compiled, denoted the *Pasteurella* virulence and resistance database. The PastyVRDB was used as a reference for BLAST analysis using Assembly2Gene. Assembly2Gene analysis showed virulence factors were well conserved within the *P. multocida* isolates sequenced in this study. Apart from Past33, no strains had matches for antibiotic resistance genes or mobile genetic element genes included in the PastyVRDB, suggesting none of the *P. multocida* isolates sequenced in this study contain any known *P. multocida* plasmids or ICEs. Past33 had matches for plasmid mobilisation genes *mobA* and *mobB* suggesting Past33 may harbour vector pB1000; however, no matches for any antibiotic resistance genes were identified for Past33.

Most of the *P. multocida* isolates encoded homologs of capsule biosynthesis proteins, suggesting they would produce a capsule layer (Figure 3.2). All strains with an intact capsule biosynthesis locus were genotype A, excluding Past22, which was genotype F. This data agrees with previous publications where only capsule type A, D and F strains have been isolated from human disease (58, 261). Several strains encoded a variant HyaD, with seven amino acids deleted between amino acids 50 and 60 and 30 missense mutations compared to HyaD from *P. multocida* strain X73. HyaD is the glycosyltransferase that synthesises hyaluronic acid (HA) capsule in *P. multocida* type A strains (262). HyaD is a class II synthase, capable of adding both N-acetylglucosamine (GlcNAc) and N-glucuronic acid (GlcA) to a HA polymer (262, 263). Functional characterisation of HyaD showed amino acids 1-117 and 704-972 are not required for glycosyltransferase function (83, 87). As such, the amino acid deletions near the start of the protein may not abrogate protein function. HyaD has two

conserved glycosyltransferase domains, A1 between residues 152-325 responsible for GlcNAc transferase activity, and A2 between residues 432-604 responsible for GlcA transferase activity (87). Both domains have a conserved DGS and DxD motifs required for transferase activity, and a WGGED domain at residues 366-370 is also required for GlcNAc transferase activity (83, 87). None of the missense mutations in HyaD from the *P. multocida* isolates were within the conserved motifs, suggesting that these HyaD proteins are likely to be functional. Although HyaE and FcbE have some similarity to glycosyltransferases, there is data on the function of homologs or domains and motifs in these proteins. As such the effect of the 135 bp gap and 54 missense mutations in these proteins could not be predicted. Five strains (Past6, Past33, Pm1616, Pm1617 and Pm1621) did not have any matches to *P. multocida* capsule biosynthesis genes. In addition, no homologs were identified for conserved Wzy-polymerase capsule biosynthesis system genes in these strains. Given both ABC-transport and Wzy-polymerase capsule biosynthesis systems have well conserved genes (74, 244), absence of these components in these strains strongly suggests they do not produce extracellular capsule.

Genes encoding proteins required for transfer of different carbohydrates and decorations were included in the PastyVRDB (12). *P. multocida* produces two structurally different LPS inner core glycoforms (105). All *P. multocida* isolates encoded homologs of proteins required for synthesis of both LPS inner core glycoforms and are therefore predicted to produce both inner core glycoforms (Figure 3.3). Only seven strains had a predicted intact outer core LPS biosynthesis locus. Despite this, isolates were genotyped using the presence of LPS outer core biosynthesis genes unique for a particular LPS genotype (101). This analysis identified 19 type L1 isolates, one type L2, 15 type L3 and one type L4 (Figure 3.3). Past29 was genotyped as L3 due to a high nucleotide identity hit for *hptE* from *P. multocida* strain P1059 but had hits for *gatH* and *gctD* from genotype L6, suggesting possible horizontal transfer of LPS outer core biosynthesis genes. Three strains (Pm1613, Pm1620 and PI32) typed as L3 with a predicted complete outer core locus encoded proteins that had missense or nonsense mutations in GatG and/or NatC that likely result in a non-functional protein. Investigation of *P. multocida* genotype L3 strains showed disrupted outer core transferases were common in L3 strains and resulted in a truncated LPS outer core structure (101). The full length L3 outer core structure consists of inner core-glucose I-heptose IV-glcose IV-galactose I-galactose II-GalNAc I-GalNAc II (101). GatG and NatC are responsible for the addition of galactose II and GalNAc II, respectively. As such, these isolates would likely have LPS outer core structures truncated at galactose I (Pm1613 and Pm1620) or GalNAc I (PI32).

All *P. multocida* isolates encoded homologs of filamentous haemagglutinin and type IV fimbriae, and several strains were predicted to produce functional Flp-pili (Figure 3.4). Four

filamentous haemagglutinin and secretion partner pairs have been identified in *P. multocida* strains (FhaB1C1 to FhaB4C4), with different strains having different combinations of filamentous haemagglutinins (122) Only FhaB1 and FhaB2 were included in the PastyVRDB as sequences for FhaB3 and FhaB4 could not be obtained. All *P. multocida* isolates encoded a single filamentous haemagglutinin and secretion partner. All strains encoded homologs of putative type IV fimbriae subunit proteins PtfA_1 and PtfA_2. Eight strains had a variant PtfA_1 that contained 30 missense mutations in the final third of the protein compared to reference PtfA_1 from Pm70. *P. multocida* PtfA is a type IVa pilin with high identity to type IVa pilins from other bacterial species (124). Type IVa pilins have a well conserved N-terminal section that contains a signal peptide and a hydrophobic α-helix domain (264). The C-terminal half of PtfA contains variable αβ-loop and D-region domains, with conserved cystine residues flanking the D-region that form an essential disulphide bridge (264). The variant PtfA_1 identified have the well conserved signal peptide, α-helix domain and D-region with flanking cystine residues, suggesting the variant PtfA_1 would still function as a fimbrial subunit protein

Fifteen of the *P. multocida* isolates were predicted to encode functional Flp-pili as they contained all genes normally present in the *P. multocida* Tad-locus (Figure 3.4). All isolates contained homologs of TadV, which is a prepilin peptidase required for maturation of the Flp1 pilin subunit (265). TadV contains two aspartic-acid containing domains required for Flp1 cleavage, a D(I/L)XXRXL motif at residues 23-29 and a (G/A)(G/A)GDXKL motif at residues 73-80; disruption of the aspartic-acid residues results in a non-functional TadV (265). These domains were present in all isolates, suggesting that the TadV homologs were functional. However, the TadV from strain HB03 used as a reference was 34 amino acids shorter than TadV from the newly sequenced isolates. TadV homologs from several other bacterial species were of a similar length to TadV seen in the *P. multocida* isolates. The HB03 TadV was missing the D(I/L)XXRXL motif, suggesting that the HB03 TadV may be inactive or incorrectly annotated. All but one of the Flp-pili-positive strains encoded a Flp1 with 9 to 11 missense mutations in the C-terminal half of the protein. Flp1 has a N-terminal signal peptide and a hydrophobic region with a conserved T(A/M)IEY(G/A)LI domain in the N-terminal half of the protein (128). All of the Flp1 proteins showed high amino acid identity up to the end of the hydrophobic region, suggesting they are likely to form functional Flp-pili subunits. The Assembly2Gene analysis indicated that 18 *P. multocida* isolates encoded high identity matches to Flp1 and TadA, but were missing genes encoding other important components in Flp-pili biosynthesis (Figure 3.4). These isolates contained homologs of several other Tad-locus genes, but most were below 80% nucleotide identity compared to the reference HB03 reference sequences and were therefore not identified by Assembly2Gene. BLASTn analysis identified matches for all genes in the Tad-locus with between 65% and 75% nucleotide

identity, excluding *tadV* and *rcpB*. Both *tadV* and *rcpB* are required for Flp-pili assembly in *A. actinomycetemcomitans*, suggesting isolates lacking *tadV* and *rcpB* would not have a functional Tad-locus (129). Production of a Flp-pili is important for *P. multocida* to cause disease. Signature tagged mutagenesis (STM) of strain VP161 in chickens and strain TF5 in mice identified *flp1* and *tadD* as required for disease, respectively. (97, 106). Given the importance of Flp-pili for disease, strains with an apparent non-functional Tad-locus should be investigated further to confirm these strains do not produce Flp-pili.

All isolates encoded homologs of OmpA and OmpH_1; however, there was major variability observed in amino acid sequences of these homologs (Figure 3.4). OmpA is an outer membrane protein that has been shown to play roles in membrane stability and transport, as well as acting as a receptor for bacteriophage attachment (266). Comparison of OmpA from *P. multocida* and other Gram-negative bacteria showed there was high amino acid conservation in sections predicted to form β-strands, but low amino acid conservation in sections predicted to form the four external loops (133). The variability seen in different OmpA variants identified in this study is consistent with previous studies that have used DNA sequence analysis or molecular mass heterogeneity of outer membrane proteins (245, 267, 268). Given the missense mutations in variant OmpA proteins were observed in predicted external loop regions and high amino acid identity to reference OmpA in six of the eight regions predicted to encode β-strands it is likely that these variant OmpA proteins are functional. OmpH is a porin (269), and most *P. multocida* strains produce two related OmpH proteins. OmpH_1 from *P. multocida* isolates showed high amino acid identity in sections of the protein predicted to form 16 antiparallel β-strands, but sequence variation in sections predicted to form external loops. As with OmpA, the variant OmpH_1 proteins would likely be functional. The variation of OmpA and OmpH_1 in predicted external loops suggests that these proteins may be under host immune selection pressure. Four strains had PlpE genes with low amino acid identity (Figure 3.4). These strains had several deletions in *plpE* causing frameshift mutations. PlpE is a lipoprotein that has been shown to stimulate protective host immunity, even against strains from different LPS serotypes (138). Loss of *plpE* from different *P. multocida* strains would lower the effectiveness of recombinant vaccines targeting *plpE*.

All *P. multocida* isolates encoded homologs of sialic acid scavenging and uptake proteins, superoxide dismutase and the virulence regulator Fis (Figure 3.4). All strains, excluding Past3, encoded a NanB homolog with between 70% and 83% amino acid identity to NanB from Pm70. NanB is an outer membrane sialidase, capable of cleaving 2-3’ and 2-6’ sialyl lactose; the cleaved sialic acids can then be imported into the bacteria and used as a carbon source or as a surface decoration for immune avoidance (150). Alignments of *P. multocida* NanB with other sialidases showed several conserved domains, including a FRIP domain near the N-

terminus, four aspartate box domains and conserved residues at the C-termius predicted to be involved in membrane anchoring (150, 270). The FRIP domain and first aspartate box were well conserved between the different NanB proteins; however, the remaining aspartate boxes and putative membrane anchoring domain were not identified. Given this, the variant NanB proteins are likely non-functional, with the sialic acid scavenging system relying on a functional NanH sialidase.

All *P. multocida* strains encoded homologs of the five iron-specific ABC-transport systems, the TonB energy transduction system proteins and the iron-responsive regulatory protein Fur (Figure 3.5). The TonB energy transduction system provides energy for import of iron by outer membrane receptors (145) and is required for *P. multocida* virulence in mice (97), so it is unsurprising that all strains encode TonB-system components. Given all isolates encoded homologs of all iron ABC-transport systems, there is likely strong selective pressure for strains to retain functional copies of all iron-transport systems. All *P. multocida* isolates encoded homologs with high amino acid identity to reference iron receptors from Pm70; however, all isolates were missing between one and four of the iron receptors included in the PastyVRDB. Several of these iron receptors have been shown to have similar iron binding profiles (139-141, 143), with redundancy in protein function likely allowing for loss of one or more iron-receptors without a significant fitness cost. No strains encoded a homolog of the bovine transferrin receptor TbpA. In addition, a number of strains encoded homologs of various receptors that are likely to be non-functional due to frameshift or nonsense mutations.

Core and pan genome analysis of *P. multocida* strains (36 isolates sequenced in this study and 260 reference *P. multocida* genomes) using Roary (271) identified a core-genome of 1,622 genes and a pan-genome of 7,537 genes across this group of strains. This is a smaller core-genome compared to previous studies where core genomes of 1,806 and 1,780 were reported (166, 272); however, the analysis from this study included more genomes, which likely leads to the small decrease in core-genome size. The pan-genome identified in this study is smaller than similar analysis performed using 656 *P. multocida* genomes that identified a pan-genome of 37,854 (167), and is likely due to the analysis in this study using less strains. Capsule and LPS genotyping were performed using Assembly2Gene. Compared to previous studies that have used *in silico* PCRs with capsule and LPS multiplex primers, Assembly2Gene identified the same capsule and LPS genotypes (166). Most strains identified were capsule type A, and LPS types L1, L2, L3 and L6. Capsule genotypes B, D and F were mostly identified with LPS types L2, L6 and L3, respectively, whereas capsule genotype A was found with all LPS genotypes identified. Identification of multiple capsule types for most LPS types, but lack of variability for capsule types B, D and F suggests LPS diversity was present before capsule diversity. Identification of one D:L3 and four F:L1 type strains suggests

homologous transfer of LPS outer core biosynthesis genes between different *P. multocida* strains.

Trait-associated genes were assessed using Scoary, which assesses gene groups identified by Roary for association with different bacterial traits (273). Scoary analysis identified no host-assocated genes, suggesting that diverse *P. multocida* strains can cause disease within the same host, and that there are no genes that confer a specific advantage for a single host. Several genes were associated with capsule and LPS genotypes, but all capsule- and LPS-specific biosynthesis genes were excluded from further analysis. As such, no other genes were identified as being associated with capsule genotypes A or D, or LPS genotypes L1, L3, L4, L6 and L7. Several genes found in capsule genotype F strains were involved in resistance, including *asrC* ($p < 2.44 \times 10^{-19}$), the *hmmR* gene group ($p < 3.59 \times 10^{-16}$) and the *zitB* gene group ($p < 6.74 \times 10^{-15}$) (Table 3.4). ArsC breaks down arsenate to arsenite, which is then exported from the cell (274). The *hmrR* gene group encodes proteins with a MerR family domain that is found in transcriptional regulators that respond to increases in heavy metal concentration or oxidative stress (275). The *zitB* gene group encodes proteins with a CzcD domain. Proteinscontaining a CzcD domain often mediate efflux of heavy metals via cation diffusion (276). Bacterial CzcD proteins have specificity for Zn, Cu, Co and Ni (277). These data suggest the presence of heavy metal resistance genes may confer an advantage specifically to *P. multocida* genotype F strains that infect pigs and bids. Additional work would be required to determine the role of these heavy metal resistance proteins in virulence.

Most genes identified as associated with B:L2 strains were predicted to encode proteins associated with bacteriophages, indicating the presence of genotype B:L2-specific bacteriophage. Previous *P. multocida* comparative genomic analysis has identified 96 genes unique to B:L2 strains that were associated with multiple bacteriophages (164). Other B:L2 associated genes included gene group 2899 ($p < 8.75 \times 10^{-37}$) and WbnI ($p < 8.75 \times 10^{-37}$). Gene group 2899 encoded a protein that had 30% amino acid identity to the lipoprotein Pm1514 from *P. multocida* strain Pm70, which is exported via a Slam lipoprotein exporter (278). WbnI is a glycosyltransferase from *E. coli* strain O86 that functions to add galactose onto the *E. coli* LPS O-antigen (279). As type B strains cause HS, these proteins are strong candidates for HS-specific virulence factors that could contribute to the severity of the disease. Lipoproteins have a wide range of virulence-associated functions in Gram-negative bacteria, including acting as iron receptors, adhesins, and being involved in immune evasion (280-282). Given the role of WbnI in *E. coli*, the WbnI homolog in *P. multocida* may also have a role in galactose transfer onto LPS, altering antigen presentation, although further work would be needed to confirm this.

The core-genome alignment generated by Roary was used to produce a *P. multocida* maximum likelihood core-genome phylogeny (Figure 3.6). The *P. multocida* core-genome phylogeny showed two clear clades that we have designated clade A and clade B (Figure 3.6). Comparison of ANI to reference subsp. *multocida*, *gallicida* and *septica* strains showed all subsp. *septica* strains clustered into clade A, whereas all strains in clade B were subsp. *multocida* or *gallicida*. Clade A contained 26 out of the 34 *P. multocida* strains isolated from humans, with other isolates in clade A recovered from cats, dogs and mice, animals likely to be vectors of *P. multocida* transmission to humans. Clustering human and animal vector isolates together into clade A suggests these isolates are genetically distinct, and that strains from clade A are more capable of causing disease in humans compared to strain B. Clade B contained several subclades with strains that belonged to the same capsule and LPS type. Strains within these subclades had low core-genome SNP counts when compared to strains from the same subclade, showing these strains were closely related. Identification of closely related strains from different countries indicates possible transmission of *P. multocida* between countries. Evidence of inter-country transmission was particularly prevalent for the B:L2 clade, where closely related strains had been isolated from 10 different countries (Figure 3.6). This indicates the possibility of type B:L2 strains of *P. multocida* being transported to non-endemic countries and causing widespread HS outbreaks. Whole-genome single nucleotide variant (SNV) phylogenies have shown a similar tree structure to clade B, with several strains from the same capsule and LPS type clustering together (166, 167). The *P. multocida* SNV phylogeny generated with 653 strains also showed two distinct phylogenetic clades, with the smaller clade containing 13 out of the 16 *P. multocida* strains isolated from humans included in the analysis (167), similar to clade A identified in the core-genome phylogeny from this study. The smaller clade in the previous work contained several avian isolates (167); however, no avian *P. multocida* isolates included in this study clustered into clade A.

Scoary analysis was performed to identify genes associated with strains from clade A or clade B. Several clade A-associated genes were predicted to be involved in carbohydrate uptake and utilisation (Table 3.6). MglA was predicted to function as an ATP-binding component of ribose and/or galactose transport. Similarly, RsbB and the protein encoded by group_2709 contained encoded proteins with an ABC-sugar-binding-like domain and an AraH domain, respectively; these domains are found in predicted ribose ABC-transport system proteins. This suggests these proteins function as a carbohydrate uptake system, and may allow uptake of L-fucose. Several genes involved in L-fucose utilisation were found to be associated with clade A, including *fucUARKI* and *aldA* (283, 284). The L-fucose metabolism proteins produced by these clade A strains would be predicted to break down L-fucose to to L-lactaldehyde. *aldA*

encodes an aldehyde dehydrogenase, and has been shown to break down L-lactaldehyde to L-lactate in *E. coli* (285, 286). L-lactate is broken down into pyruvate, which then enters central metabolism pathways. L-fucose is commonly added to the terminal end of human cell surface glycoproteins and is used to commonly decorate mucins (287, 288). The ability to utilise L-fucose has been shown to increase bacterial fitness *in vivo*. *Campylobacter jejuni* strain NCTC 11168 with a disrupted L-fucose permease had a 15- to 100-fold reduction in colonisation of pig intestine compared to the wild-type parent strain (289). Previous comparative genomic analysis of pathogenic and non-pathogenic avian *P. multocida* isolates showed L-fucose and ribose ABC transporters were associated with pathogenic *P. multocida* avian isolates (122). Roary and Scoary analysis in this study showed 22 avian isolate strains from clade B contained L-fucose metabolism and ribose ABC-transport uptake systems, indicating these systems may also contribute to pathogenicity within birds. Investigation of the ability of these strains to utilise L-fucose as a carbon source is warranted.

Similar Roary and Scoary analysis has been performed on human and avian *P. multocida* strains that clustered into an outer clade, away from most animal isolates in a recently published whole-genome SNV phylogeny (167). The previous analysis identified several genes as over-represented (*esiB*, *cdiA*, *oatA*, *etk* and *tetD*) or under-represented (*wzc*) in this outer clade (167). The Roary analysis performed here disagrees with these results, showing instead that most of these genes were present in almost all *P. multocida* strains (*esiB*, *etk* and *wzc*), absent in almost all *P. multocida* strains (*oatA*) or did not have a good match to any of the gene groups called by Roary, suggesting the gene is not present in any of the *P. multocida* genomes (*cdiA*). The analysis from this study was performed using Roary with default settings and as such, any homologs with < 95% amino acid identity would have been included in different gene groups. Indeed, the nucleotide reference sequence for *etk* and *wzc* displayed > 99% nucleotide identity. The only gene found to be associated with the human clade from both analyses was *tetD*, however this was present in only 26 of the 36 strains that clustered into clade B in this study.

Several clade B-associated genes were predicted to be important for anaerobic growth. The *torCAD* operon encodes a trimethylamine-N-oxide anaerobic respiration system. TorC is a c-type cytochrome anchored to the inner membrane that transfers electrons from menaquinones to TorA (290). TorA is a reductase that reduces trimethylamine-N-oxide to trimethylamine, with TorD a TorA-specific chaperone (291, 292). Trimethylamine-N-oxide reductase system genes have been identified in previous *P. multocida* comparative genomic analyses and proposed to be found only in strains capable of causing disease in birds (122, 165, 293). In this study, Roary analysis showed that the *torCAD* system is present in all but three strains (FDDARGOS_384, FDDARGOS_385 and FDDARGOS_644) from clade B, indicating this

system is likely to be important for disease in all animals, not only for disease in birds and pigs. Another clade B-specific gene, *dcuC* is predicted to encode an anaerobic C4-dicarboxylate transporter (294). Anerobic energy production and carbon source transporters may confer greater fitness to clade B strains during systemic disease, where bacteria would encounter anaerobic environments in host tissue. The *macA* gene was also significantly associated with clade B strains. The MacAB-TolC system exports macrolide antibiotics from the cell (295). Studies of macrolide resistance in *P. multocida* did not identify *macA* or *macB* as resistance determinants, indicating the MacAB-TolC system may have a different function in *P. multocida* (163). All genes required for a functional MacAB-TolC system were present in most clade B strains; however, *macA* was largely absent in clade A strains. Thus, a functional system may give clade B strains a fitness advantage during infections in animal hosts, however further work is required to identify the exact role during disease.

In addition to the *P. multocida* phylogeny, sketch-based trees were generated for strains from within the current *Pasteurella* genus and Pasteurellaceae family. Previous Pasteurellaceae family phylogenies produced by comparison 16s rRNA or housekeeping genes have shown that several *Pasteurella* spp. cluster away from a consistent core *Pasteurella* clade containing the species *multocida*, *canis*, *dagmatis* and *stomatis* (5, 177). Both the *Pasteurella* genus and Pasteurellaceae family tree generated in this study showed a strict *Pasteurella* clade containing spp. *multocida*, *canis*, *dagmatis*, *stomatis* and *oralis* indicating that these species are all part of *Pasteurella sensu stricto* (Figure 3.7 and Figure 5.8). However, the current [*Pasteurella*] species *aerogenes*, *bettyae*, *caecimuris*, *langaaensis*, *mairii*, *skynesis* and *testudinis* clustered away from the *Pasteurella* clade on the *Pasteurella* spp. tree and away from all other genera on the Pasteurellaceae family tree, suggesting these species may represent novel Pasteurellaceae genera (Figure 3.7 and Figure 5.8).

3.4 Conclusion

We sequenced 56 *Pasteurella* spp. isolates from human and dog infections, and from cat normal flora. This generated 22 human *P. multocida* isolates, provided the first whole-genome sequences for *P. stomatis* and greatly increased the number of *P. canis* and *P. dagmatis* whole-genome sequences. In-depth virulence factor and antibiotic resistance identification was performed using ABRicate and Assembly2Gene searches. Most *P. multocida* strains encoded homologs of known virulence factors that were include in a new database designated the PastyVRDB. Five of the human *P. multocida* isolates were predicted not to produce surface capsule, and 18 isolates had a predicted inactive Tad-locus. All *P. multocida* isolates encoded a single filamentous haemagglutinin and secretion partner pair, and were missing between one and four outer membrane iron-receptors. In addition, a putative LuxS-Lsr quorum

sensing system was identified in all *Pasteurella* spp. isolates. Pan-genome analysis using Roary and Scoary identified several trait-specific genes, including capsule type F strains encoding several putative heavy metal resistance proteins and type B:L2 strains encoding a specific lipoprotein and glycosyltransferase. We also generated an extended *P. multocida* core-genome phylogeny that clearly showed two clades; 26 of the 32 *P. multocida* strains recovered from humans included in this study clustered into clade A, suggesting clade A strains are better suited to causing human disease. Genes associated with clade A encoded L-fucose utilisation proteins and a putative ribose-specific ABC-transporter. Clade B-associated genes encoded a trimethylamine-N-oxide anaerobic respiration system and an anaerobic C4-dicarboxylate transporter. The identified trait-specific genes represent excellent candidates for factors that uniquely confer fitness for human or animal disease. Finally, *Pasteurella* genus and Pasteurellaceae family trees were generated, suggesting that the *Pasteurella* genus contains only the species *multocida*, *canis*, *dagmatis*, *stomatis* and oralis, and that the species *aerogenes*, *bettyae*, *caecimuris*, *langaaensis*, *mairii*, *skynesis* and *testudinis* should be reclassified into new genera.

3.6 Materials and methods

3.6.1 Bacterial strains and growth conditions

Human and Cat isolates used in this study are listed in Table 3.8 *Pasteurella* isolates were cultured on heart infusion (HI) agar plates (Oxoid) or horse blood agar plates (Oxoid), or in HI broth for liquid cultures. Culturing conditions for solid media were 37ºC for 24 h to 48 h. Broth cultures were grown at 37ºC for up to 24 hours, shaking at 200 rpm.

Table 3.8: Bacterial strains used in this study

Strain	Description	Reference
VP161	Avian isolate strain, *P. multocida* KMT fragment positive strain	(26)
X-73	Avian isolate strain, *P. multocida* KMT fragment positive strain	(296)
Past1	Human blood culture isolate, MALDI-TOF typed as *P. multocida*	This study
Past2	Human blood culture isolate, MALDI-TOF typed as *P. dagmatis*	This study
Past3	Human blood culture isolate, MALDI-TOF typed as *P. multocida*	This study
Past4	Human blood culture isolate, MALDI-TOF typed as *P. multocida*	This study
Past5	Human blood culture isolate, MALDI-TOF typed as *P. multocida*	This study
Past6	Human peritoneal dialysis isolate, MALDI-TOF typed as *P. multocida*	This study
Past7	Human peritoneal dialysis isolate, MALDI-TOF typed as *P. multocida*	This study
Past8	Human peritoneal dialysis isolate, MALDI-TOF typed as *P. canis*	This study
Past9	Human peritoneal dialysis isolate, MALDI-TOF typed as *P. multocida*	This study
Past10	Human cat bite isolate, MALDI-TOF typed as *P. multocida*	This study
Past11	Human skin tissue isolate, MALDI-TOF typed as *P. dagmatis*	This study
Past12	Human skin swab isolate, MALDI-TOF typed as *P. multocida*	This study
Past13	Human blood culture isolate, MALDI-TOF typed as *P. multocida*	This study
Past14	Human skin swab isolate, MALDI-TOF typed as *P. canis*	This study
Past15	Human respiratory tract isolate, MALDI-TOF typed as *P. multocida*	This study
Past16	Human skin swab isolate, MALDI-TOF typed as *P. canis*	This study
Past17	Human skin swab isolate, MALDI-TOF typed as *P. canis*	This study
Past18	Human cat bite isolate, MALDI-TOF typed as *P. multocida*	This study
Past19	Human cat bite abcess isolate, MALDI-TOF typed as *P. multocida*	This study
Past20	Human skin swab isolate, MALDI-TOF typed as *P. canis*	This study
Past21	Human skin tissue isolate, MALDI-TOF typed as *P. dagmatis*	This study
Past22	Human cat bite isolate, MALDI-TOF typed as *P. multocida*	This study
Past23	Human peritoneal dialysis isolate, MALDI-TOF typed as *P. multocida*	This study
Past24	Human dog bite isolate, MALDI-TOF typed as *P. canis*	This study
Past25	Human skin tissue isolate, MALDI-TOF typed as *P. dagmatis*	This study
Past26	Human skin tissue isolate, MALDI-TOF typed as *P. multocida*	This study
Past27	Human dog bite isolate, MALDI-TOF typed as *P. canis*	This study
Past28	Human cat bite isolate, MALDI-TOF typed as *P. multocida*	This study
Past29	Human blood culture isolate, MALDI-TOF typed as *P. multocida*	This study
Past30	Human cat bite isolate, MALDI-TOF typed as *P. multocida*	This study
Past31	Human skin tissue isolate, MALDI-TOF typed as *P. multocida*	This study
Past32	Human skin tissue isolate, MALDI-TOF typed as *P. dagmatis*	This study
Past33	Human skin tissue isolate, MALDI-TOF typed as *P. multocida*	This study
Past34	Human dog bite isolate, MALDI-TOF typed as *P. multocida*	This study
Past35	Human skin tissue isolate, MALDI-TOF typed as *P. canis*	This study
Pm1476	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1525	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1612	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1613	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1616	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1617	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1618	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1620	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1621	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pm1622	Cat upper respiratory tract isolate, Microbact typed as *P. multocida*	This study
Pmc-A	Cat upper respiratory tract isolate	This study
Pmc-B	Cat upper respiratory tract isolate	This study
Pmc-C	Cat upper respiratory tract isolate	This study
Pmc-D	Cat upper respiratory tract isolate	This study
Pmc-E	Cat upper respiratory tract isolate	This study
Pmc-F	Cat upper respiratory tract isolate	This study
PI31	Dog wound isolate, typed as *P. multocida* by biochemical tests	This study
PI32	Dog tissue isolate, typed as *P. multocida* by biochemical tests	This study
PI33	Dog animal bite wound isolate, typed as *P. multocida* by biochemical tests	This study
PI34	Dog salivary gland isolate, typed as *P. multocida* by biochemical tests	This study
PI35	Dog tracheal wash isolate, typed as *P. multocida* by biochemical tests	This study

Strain	Description	Reference
PI36	Dog ear isolate, typed as *P. multocida* by biochemical tests	This study
PI37	Dog wound isolate, typed as *P. multocida* by biochemical tests	This study
PI38	Dog retrobulbar isolate, typed as *P. multocida* by biochemical tests	This study

3.6.2 DNA extraction, PCR and whole genome sequencing

Genomic DNA was extracted using the HiYield Genomic DNA Mini Kit (Real Biotech Corporation) as per manufacturer's instructions. DNA quantity was assessed using agarose gel electrophoresis and Qubit Fluorometry (Life Technologies). PCR amplification of the KMT fragment was performed using *Taq* polymerase as described in chapter two, primers BAP511 (5'-TGGCGGATCCGACCAACAAACCTATTGG-3') and BAP512 (5'-CAGTGAATTCACATCAAATCCCGGACCG-3'), and genomic DNA from isolates as the template. PCR reactions were analysed by electrophoresis through a 1% (w/v) agarose gel, compared to 1 kb and 100 bp DNA ladders (New England Biolabs). Electrophoresis was performed at 100 V for 50 min and DNA was visualised using a Amersham Imager RGB 680. Genomic DNA from all isolates was submitted to Micromon Genomics (Monash University, Australia) and sequenced using 150 bp-paired-end sequencing protocol on an Illumina MiSeq v2.

3.6.3 Genome Assembly and species identification

Quality of Illumina sequences was assessed using FastQC, before low quality sequencing reads and Illumina adapter sequences were trimmed using FASTQ Trimmer from the FASTX-Toolkit v0.0.13 (https://github.com/Debian/fastx-toolkit). Isolate genomes were *de novo* assembled using Unicycler v0.4.8 (236) and annotated using Prokka v1.14.5 (237). Genome assembly quality control data was generated using Quast v5.1.0 with the *de novo* assembled genomes and trimmed reads as input (297). Isolate bacterial species identification was performed using fastANI v1.32, with a kmer size of 21 (238). Reference genomes used for fastANI analysis were *Paseturella* spp. type strains downloaded from the PATRIC database (https://lpsn.dsmz.de/genus/pasteurella), and 380 complete Pasteurellaceae genomes from the NCBI database (downloaded Dec 2020). For isolates where species identification was not possible using fastANI, the whole 16s rRNA gene from that strain was used as a query against the NCBI reference RNA sequence database using BLASTn. A match to a reference 16s rRNA with identity >98.5% was used to call the isolate as the same bacterial species as the reference.

3.6.4 Identification of virulence, resistance, and trait-specific genes

Putative virulence genes in the isolate genomes were identified using two search strategies. ABRicate v1.0.1 (https://github.com/tseemann/abricate) was used to search the comprehensive antimicrobial resistance database, the virulence factor database, and

ResFinder and PlasmidFinder databases (240-243). Positive matches in ABRicate required 50% nucleotide identity over 50% of the query sequence. A targeted *P. multocida* virulence factor and antibiotic resistance gene search was performed using Assembly2Gene (https://github.com/LPerlaza/Assembly2Gene). Assembly2Gene identifies genes in input genomes and compares these genes to a reference gene database by BLASTn. A database of known *P. multocida* virulence and resistance genes (PastyVRDB) was compiled and used as reference sequences with Assembly2Gene, with *P. multocida* isolate genomes used as the input genome sequences. An 80% nucleotide identity between query and reference was used to identify a match, and there was no minimum coverage required for a match. Heat maps for Assembly2Gene outputs were generated using the ggplot2 package in R-Studio. Trait-specific genes were identified using Roary v3.11.2 and Scoary v1.6.16 (271, 273). Roary analysis of the *P. multocida* core- and pan-genome was performed using the 36 *P. multocida* isolates sequenced from this study, and 260 reference *P. multocida* genomes from the PATRIC database (downloaded Dec 2020). Roary was run using default 95% amino acid identity cut-off for inclusion into the same group, and with paralogs allowed in the core-genome alignment. Scoary was run with an adjusted P-value cutoff of $1E^{-4}$ and no pairwise adjustment. For identification of protein function, protein sequences were searched against the UniProt database using BLASTp to identify homologs, and conserved domains identified using NCBI conserved domain search.

3.6.5 Generation of phylogenetic trees and core-genome single nucleotide polymorphism analysis

IQ-TREE v1.6.12 was used to generate a maximum-likelihood phylogenetic tree from the *P. multocida* core-genome alignment produced by Roary. IQ-TREE was run with 1,000 bootstrap replicates. Sketch distance-based trees of the *Pasteurella* genus and Pasteurellaceae family were generated using Mashtree v0.57. Mashtree was run with 100 bootstrap replicates. Trees were annotated using packages ggplot2 and ggtree in R-Studio. Strain names were used as node labels, with strain names over 12 characters shortened (ABT_RAWAL_2015_HSR replaced with ABT_HSR and USDA-ARS-USMARC with US). Core-genome single nucleotide polymorphisms were identified using Parsnp v1.5.6 (298), comparing all *P. multocida* whole genome sequences used in this study to *P. multocida* strain VP161. The Parsnp core-genome SNP output table was used to compare the number of conserved or unique core-genome SNPs between selected strains.

3.5 Appendices

Appendix 3.1 – ANI values for isolates sequenced in this study against *Pasteurella* type strains

Appendix 3.2 – ANI values for isolates sequenced in this study against complete Pasteurellaceae genomes

Appendix 3.3 – ABRicate hits for isolate genomes searched against the Virulence Factor Database

Appendix 3.4 – ABRicate hits for isolate genomes searched against the Comprehensive Antimicrobial Resistance Database

Appendix 3.5 – ABRicate hits for isolate genomes searched against the ResFinder database

Appendix 3.6 – Genes included in the Pasteurella Virulence and Resistance Database

Appendix 3.7 – Assembly2Gene hits for isolates when searched against capsule biosynthesis genes

Appendix 3.8 – Assembly2Gene hits for isolates when searched against lipopolysaccharide biosynthesis genes

Appendix 3.9 – Assembly2Gene hits for isolates when searched against adhesins, outer membrane proteins and virulence factor genes

Appendix 3.10 – Assembly2Gene hits for isolates when searched against iron scavenging and uptake genes

Appendix 3.11 – Nexus file of the *P. multocida* core-genome phylogeny

Appendix 3.12 - Annotation file for the *P. multocida* core-genome phylogeny

Appendix 3.13 – Nexus file of the *Pasteurella* genus sketch-distance based tree

Appendix 3.14 – Nexus file of the Pasteurellaceae family sketch-distance based tree

Chapter 4 – General discussion and future directions

Pasteurella multocida is responsible for several distinct animal diseases, including fowl cholera in birds, haemorrhagic septicaemia in ungulates, atrophic rhinitis in pigs and shipping fever in livestock animals (13). *P. multocida* can also cause a range of zoonotic diseases in humans (61). Despite the range of animal and human diseases, the precise molecular pathogenic mechanisms of many *P. multocida* diseases are poorly understood, with almost nothing known about the pathogenic mechanisms or virulence factors used by *P. multocida* human isolates. A number of *P. multocida* virulence factors have been identified as discussed in chapter one, however most of these are yet to be completely characterised, with many only identified bioinformatically. One key *P. multocida* virulence factor is capsule. *P. multocida* type A strains produce a hyaluronic acid (HA) capsule that is required for causing disease in the natural host and has been shown to protect the bacterium against chicken complement-mediated killing (76). Despite the importance of *P. multocida* capsule, a complete understanding of the regulatory network and biosynthesis pathway of this polysaccharide has yet to be fully elucidated. In this study, we aimed to fully characterise the *P. multocida* hyaluronic acid capsule biosynthesis pathway, and to characterise the ability of *P. multocida* strain VP161 to resist chicken complement-mediated killing. Genes involved in HA production, growth in active chicken serum and growth in rich media were investigated using transposon-directed insertion site sequencing (TraDIS). In a separate study, we used whole-genome sequencing and comparative bioinformatics to characterise *P. multocida* human disease isolates, and investigated the genetic that underly the ability of different *P. multocida* strains to cause different animal diseases.

Chapter two details work performed to establish a TraDIS method for use in *P. multocida* strain VP161 and subsequent TraDIS assays performed to investigate genes essential for growth in rich media or in active serum, and genes important for HA capsule production. TraDIS is a whole-genome forward genetics approach that couples random transposon mutagenesis with whole genome sequencing to identify genes required for a particular phenotype. A very large, random transposon mutant library must first be generated in the organism of interest, with the mutant library ideally containing multiple independent insertions in all genes in the genome. The mutant library is then grown in a condition of interest, either negatively selecting against transposon mutants unable to grow in the condition of interest or positively selecting transposon mutants with a phenotype of interest. All surviving or isolated mutants are recovered and used to prepare a TraDIS library that allows sequencing of transposon insertion sites. Analysis of insertion sites in the recovered mutants identifies genes essential for growth in the condition of interest, or genes important for the phenotype of interest. TraDIS is ideal to identify all genes important for *P. multocida* hyaluronic acid capsule production, and all genes

essential for *P. multocida* survival in chicken serum. Initially, a *Himar1* mutant library was generated in the highly virulent *P. multocida* strain VP161 (A:L1) via conjugation, with each conjugation producing well in excess of 10^6 *Himar1* mutants. A large pool of ~6.45 x 10^6 VP161-Tn*7* *Himar1* mutants was used for all subsequent TraDIS experiments.

Initially, the VP161-Tn*7* *Himar1* mutant library was used to identify genes essential for *P. multocida* growth in rich media. The mutant library was grown in HI broth, with surviving mutants recovered and used to produce TraDIS libraries, using a previously described method with minor modifications to improve library production efficiency (196). The rich media TraDIS libraries were sequenced via Illumina sequencing and data analysed using Bio-TraDIS scripts (196). Combined, the library contained 81,929 unique insertion sites (UIS), giving a UIS every 28 bp and an average of 38 UIS per gene. The VP161-Tn*7* mutant library has greater resolution than most mutant libraries generated using *mariner* transposons in other bacterial species (189). A total of 509 genes were identified as essential for *P. multocida* growth in rich media, including 473 protein encoding genes, 35 tRNAs and one 5s rRNA. The number of genes shown to be essential for *P. multocida* growth in rich media was similar to the number of essential genes identified by transposon insertion sequencing (TIS) methods in other bacterial species, ranging from 358 essential genes in *Escherichia coli* K-12 strain BW25113 to 721 essential genes in *Pseudomonas aeruginosa* strain BL23 (199, 299). The 473 *P. multocida* proteins identified as essential for growth in rich media were compared against the database of essential genes (DEG) (300), with 442 (93%) identified as homologs of proteins essential in another bacterial species. Most of the entries in the DEG database from other bacterial species were also identified as essential for growth in rich media, giving us confidence that the genes identified by our TraDIS analysis are truly essential for *P. multocida* growth in rich media. The general function and pathways each essential gene was involved in was identified using COGs group assignment and comparison to the KEGG database. Most of the essential *P. multocida* genes encoded proteins involved in housekeeping processes, with a large proportion involved in translation and cell wall biosynthesis. Metabolic pathways and processes identified by KEGG analysis included the central metabolic pathways of glycolysis, pyruvate oxidation and gluconeogenesis, as well as biosynthesis pathways for cell wall components phosphatidylethanolamine, lipid A, peptidoglycan, nucleotides, lysine, and several vitamins and cofactors, and housekeeping processes such as ribosome production, tRNA synthetases, DNA replication, cell division and Sec protein export. Several of these pathways have also been identified in other bacterial species as essential for growth in rich media (187, 202, 301, 302).

Although *Himar1* mutagenesis generated a large, high resolution VP161 mutant library, the overall number of UIS was still too low to confidently investigate very small DNA elements

such as non-coding RNAs and regulatory elements. The TraDIS analysis method used in this study identifies essential genes by comparing insertion indexes (number of UIS normalised for gene length) between all genes. A histogram of all insertion indexes should show a bimodal distribution; non-essential genes with low insertion indexes cluster to the left and essential genes with high insertion indexes cluster to the right (see Figure S2 and 2.8 from chapter two). Two normal curves are drawn from the dataset, with the intersect between the two used as the cut off for essentiality. Ideally, there should be a clear gap between essential and non-essential genes in the insertion index histogram. The histogram generated from our TraDIS analysis of the VP161-Tn*7* *Himar1* mutant library grown in rich media displayed a narrow gap between essential and non-essential genes, with several genes near the cut off for essentiality. This increases the likelihood that genes near the cut-off may be incorrectly called as essential or non-essential. Shorter genes have higher variability in their insertion index and are therefore more likely to be incorrectly called as essential or non-essential from small differences in the number of UIS. This was observed for copies of the same tRNA being differentially called as essential or non-essential for growth in rich media due to a difference of just one UIS. Although disruption of a tRNA gene may cause a fitness defect, multiple copies of tRNA genes are usually present, so intact copies of the same tRNA gene should provide functional redundancy. Therefore, a single tRNA gene with a copy elsewhere on the genome is unlikely to be truly essential.

Increasing the resolution of the mutant library would increase the number of UIS per gene, increasing the insertion index of non-essential genes, and would likely decrease the number of false positives. The *Himar1* target sequence is a TA dinucleotide (191), giving 148,690 possible *Himar1* insertion sites in the VP161 genome. Despite generating more than 10^6 VP161-Tn*7* *Himar1* mutants per conjugation, only 81,929 (55.1%) of the possible insertion sites contained a *Himar1* insertion. It is possible that many of the predicted *Himar1* sites may not be accessible to the transposon due to chromosomal secondary structure. Bacterial DNA is condensed into a nucleoid by DNA binding proteins, including the histone-like nucleoid structuring protein H-NS that binds to AT-rich regions (303-305). Indeed, *Himar1* mutant libraries generated in *Vibrio cholerae* strain C6706 had AT-rich genes under-represented compared to *Himar1* libraries generated in a strain C6706 Δ*hns* mutant, suggesting H-NS driven DNA condensation blocks *Himar1* transposition (306). Other TIS studies have successfully identified several essential non-coding RNAs, regulatory sequences and protein domains using Tn*5* mutant libraries with a UIS every 5 to 13 bp (199, 202, 301). Overall, the Tn*5* transposon has been shown to produce higher resolution mutant libraries compared to *Himar1* (189). As such, future experiments using Tn*5* in place of *Himar1* may allow for generation of higher resolution mutant libraries in *P. multocida* that would allow investigation

of non-coding RNA and regulatory elements. In addition, culturing conditions used may have also biased the results. For example, mutants with insertions in genes not essential for growth in rich media but that result in a severe fitness defect may be outcompeted during culturing and would therefore be present at lower abundance in the recovered mutant pool. The VP161 *Himar1* mutant library was grown overnight in heart infusion broth, subcultured, and grown to mid-exponential growth phase. Shortening the initial culturing time from overnight to a few hours, so that all cultures only reach early-exponential phase growth may reduce competition and lower the impact of fitness defects. Therefore, in future, increasing the resolution of the mutant library and altering growth conditions should result in a more accurate assessment of essential and non-essential genes and allow detailed analysis of essential regulatory regions and non-coding sequences.

Genes identified by TraDIS are only putatively essential and their role must be confirmed in directed mutants and complemented mutant strains. In *P. multocida*, the most effective mutagenesis method involved TargeTron technology that utilises a group II intron that can be retargeted to specific sequences (307, 308). However, fully inactivating mutants cannot be done in truly essential genes, as the mutant strain would be non-viable. To overcome this, conditional mutants can be made. This involves first providing the parent strain with an intact copy of the target gene on a plasmid. However, if TargeTron mutagenesis is to be used to inactivate the chromosomal copy of the gene, the intron targeted sequence in the gene provided *in trans* must be first removed by introducing silent mutations. This setup would provide a functional copy of the essential gene during mutagenesis, ensuring cell viability, and would allow for TargeTron mutagenesis to specifically target the chromosomal copy of the gene. An inability to cure the mutant of the complementation vector would confirm the target gene is essential for growth in rich media. Alternatively, utilisation of a λ red homologous recombineering may allow high frequency homologous recombination in *P. multocida* (309). Homologous recombination at sequences up and downstream of the target gene would allow for specific targeting of the chromosomal copy of the gene, but not the plasmid copy of the gene. However, initial attempts at λ recombineering in *P. multocida* have so far been unsuccessful (Boyce laboratory unpublished data).

With a TraDIS method optimised for use in *P. multocida* strain VP161, we investigated genes important for capsule production and for growth in active chicken serum. For investigation of capsule production by TraDIS, the large VP161-Tn*7* *Himar1* mutant library was separated into capsular and acapsular populations by discontinuous Percoll density gradient centrifugation. TraDIS libraries were generated from the two fractions, with sequencing and data analysis identifying 69 genes important for capsule production. All previously known capsule biosynthesis and regulatory proteins were identified by this TraDIS analysis. Five of the novel

capsule biosynthesis genes identified (*ptsH*, *ppx*, *spoT*, *pgm* and *galU*) were confirmed to be required for capsule production by HA capsule absorbance assays, with the involvement of *spoT* also confirmed by scanning electron microscopy. TargeTron mutagenesis was attempted for an additional 11 novel capsule-associated genes; however, despite two attempts no mutants were generated for these genes. Ideally, several genes close to the $\log_2$-fold change and *q*-value cut off should be confirmed by directed mutagenesis for further validation of the genes identified by TraDIS.

Two genes identified as important for capsule biosynthesis, *relA* and *spoT*, encode proteins that control the level of the alarmones guanosine tetraphosphate and guanosine pentaphosphate (collectively termed (p)ppGpp) (310). Guanosine alarmones control the activation of the stringent response (SR), a stress response that responds to lack of nutrients (311). During the SR, (p)ppGpp allosterically regulates protein activity to halt cell growth, and to alter RNAP activity resulting in global changes in gene expression (311-313). *Himar1* insertion sites in *relA* and *spoT* acapsular mutants were almost all within the 3' region of the genes, which are predicted to encode regulatory domains of the proteins. Disruption of these regulatory domains in RelA in other bacteria result in increased (p)ppGpp synthesis (310, 314, 315). In *E. coli*, the gene encoding the global regulatory protein Fis is downregulated during the SR (93, 316). Fis positively regulates capsule biosynthesis in *P. multocida* (92). Together, this suggested that disrupting regulatory domains of RelA and SpoT results in increased (p)ppGpp production in *P. multocida* that would reduce *fis* expression, which would then lead to reduced capsule gene expression. To determine if this was true, expression of *fis* and *hyaD* was assessed by qRT-PCR in the *spoT* acapsular mutant that had a TargeTron insertion in the 3' regulatory region. The *spoT* mutant displayed significantly reduced *hyaD* expression but no change to *fis* expression, suggesting increased (p)ppGpp causes downregulation of capsule biosynthesis genes independently of Fis. Several other genes identified by the Percoll TraDIS analysis (*tufA_1*, *tufA_2*, *ptsH* and *ppx*) encode proteins that have been associated with regulating the activity of RelA and SpoT, with mutants in these genes also likely to result in an increase in (p)ppGpp concentration.

Future work on the SR in *P. multocida* would involve confirming (p)ppGpp downregulation of capsule genes and to elucidate the mechanism of regulation. This work would involve generating mutants in *tufA_1*, *tufA_2*, and constructing a *relA* mutant with an insertion in the 3' regulatory region. All strains with mutations in SR-related genes would have capsule production and capsule gene expression assessed as described above. In addition, the concentration of (p)ppGpp in these mutant strains should be assessed by a high-performance liquid chromatography-based method (317), and compared to levels of (p)ppGpp in wild-type VP161 at early exponential growth phase. If these mutants had reduced capsule production,

reduced capsule gene expression, and increased (p)ppGpp levels, this would confirm that activation of the SR downregulates capsule in *P. multocida* VP161. Additionally, a *relA* mutant with an insertion in the 5' end of the gene could also be produced and assessed using the same methods. Such a *relA* mutant with an insertion disrupting (p)ppGpp synthase activity should have similar or reduced (p)ppGpp concentration compared to wild-type VP161, and therefore should have no change in capsule gene expression and capsule production. To identify the mechanism of (p)ppGpp regulation of capsule production, protein targets of (p)ppGpp in *P. multocida* could be identified via a cross-linking mass spectrometry approach (318). Conjugation of (p)ppGpp with a diazirine cross-linker and a biotin tag, followed by incubation with *P. multocida* VP161 cell lysates would allow any proteins interacting with conjugated (p)ppGpp to be cross-linked to the diazirine tag via UV-treatment. The crosslinked (p)ppGpp target proteins could be recovered via biotin co-immunoprecipitation. Mass spectrometry would then be used to identify the precipitated (p)ppGpp targets (318). The genes encoding the (p)ppGpp targets could then be inactivated in a mutant strain that has high (p)ppGpp concentration compared to wild-type VP161 during early exponential phase growth, or individual mutants of (p)ppGpp targets could be grown in conditions that activate the SR. Inactivating a gene encoding a protein required for (p)ppGpp downregulation of capsule biosynthesis genes should restore capsule production to wild-type levels, despite high (p)ppGpp concentration. The most likely mechanism of reduced capsule gene expression is reduced affinity of RNAP to the capsule biosynthesis locus promotor due to (p)ppGpp binding to RNAP. This could be assessed by altering the (p)ppGpp binding site on RNAP to inhibit (p)ppGpp binding, and then comparing capsule gene expression in the mutant and wild-type strains under levels of normal or increased levels of (p)ppGpp (313).

Previous TIS studies successfully used Percoll density gradients to isolate mutants with both more and less capsule production compared to the amounts produced by the wild-type parent strains (319). To identify genes that are predicted to be required for the downregulation of capsule production, density gradients using different percentage Percoll layers above 40% could be assessed to determine if hyper-capsulated mutants can be separated from mutants with wild-type levels of capsule. TraDIS analysis of hyper-capsulated strains would allow for the identification of genes that negatively impact capsule production. In addition, genes associated with capsule production could be confirmed by an alternative selection method for acapsular mutants. Other TIS studies investigating capsule have isolated acapsular mutants by treatment of the mutant library with lytic bacteriophages that target the capsule layer (320, 321). Co-incubation of capsule-targeting phage and a mutant library would allow the phage to enter and lyse mutant cells that have a capsule layer, but any mutant cells with insertions in genes required for capsule production would not produce a capsule layer and therefore would

not be killed by the phage. A lytic bacteriophage, PHB02, has been shown to be specific for *P. multocida* type A strains and is unable to target acapsular *P. multocida* type A mutant strains (322). Performing TraDIS with bacteriophage PHB02 selection of the VP161-Tn*7* *Himar1* mutant library would provide confirmation of genes identified via the Percoll gradient TraDIS analysis.

To identify genes required for growth in chicken serum with active complement activity, the VP161-Tn*7* *Himar1* mutant library was incubated in 90% active chicken serum for 3 h, before surviving mutants were recovered. TraDIS libraries were then generated, and analysis identified 523 genes essential for growth in active serum, and 75 genes important for fitness. Comparative analysis with genes essential for growth in heart infusion broth revealed that 52 genes were uniquely essential for growth in active chicken serum and 44 were important for fitness in active chicken serum but were not essential for growth in either serum or rich media. As expected, all capsule biosynthesis genes were identified as essential or important for fitness in active serum. However, both *fis* and *hfq* were not identified as important for growth in serum, indicating that loss of Fis and Hfq, which are both positive regulators of capsule production, had no effect on viability in active serum. This indicates that Fis and Hfq-dependent upregulation of capsule biosynthesis genes is not required for growth in serum, and suggests that the amount of capsule produced without Fis and Hfq is sufficient for growth in active serum. Given the low levels of capsule observed in *hfq* and *fis* mutants grown in rich media (92, 96), this may indicate that there is an alternative capsule regulator active during growth in serum. An important next step would be to compare the amounts of capsule produced by *P. multocida* VP161, a capsule biosynthesis mutant and *fis* and *hfq* mutants grown in active serum and heat-inactivated serum using HA absorbance assays and scanning electron microscopy, as levels of capsule production by specific mutants during growth in serum has not previously been assessed. Several genes identified as important for serum were involved in aerobic respiration including ubiquinone biosynthesis and ATP synthase subunit genes, suggesting aerobic growth is required for resistance to chicken serum in VP161. To identify if these genes are involved specifically in complement resistance, directed mutants representing these genes should be tested for growth in active and inactive serum. If aerobic respiration is confirmed to be specifically linked to complement resistance, then VP161 sensitivity to complement under both aerobic and anaerobic conditions should be assessed, with anaerobic growth predicted to result in increased sensitivity. Of the genes essential or important for fitness in active serum but not essential for growth in HI, 26 genes were also identified as important for capsule production. Genes important for both survival in serum and production of HA capsule included all capsule biosynthesis genes, *cpdA* that encodes a cAMP phosphodiesterase, PmPV161_0878 that encodes a putative sodium proton antiporter and

efp, *rimP* and *prmC*, which all encode proteins involved in translation. Given these genes were important for serum resistance and capsule production, these genes should be confirmed as important for both conditions by directed mutagenesis. As these genes are likely important for virulence, future work should involve assessing the virulence of these strains in chickens.

Chapter three details bioinformatic work performed to characterise several human and animal *Pasteurella* spp. isolates, and to elucidate putative pathogenic mechanisms that contribute to different *P. multocida* diseases. Draft genome sequences were generated for 35 *Pasteurella* spp. isolates causing disease in humans, 16 isolates from the upper respiratory tract of cats, and 8 isolates causing disease in dogs. Species of these isolates was identified using average nucleotide identity or identity of whole 16s rRNA to reference sequences, with *P. multocida* strains also confirmed by identification of the *P. multocida* specific KMT fragment by PCR (249). These analyses identified isolates as either *P. multocida*, *P. canis*, *P. dagmatis*, *P. stomatis* or *Frederiksenia canicola*. Virulence factors and antibiotic resistance genes were identified by identifying homologs in curated databases using ABRicate, or to the PastyVRDB that contained known *P. multocida* virulence and resistance genes using Assembly2gene. ABRicate analysis identified a putative quorum sensing system in all *Pasteurella* isolates sequenced in this study. Future work could confirm if this system is active, and what genes are regulated by this system when population density is high. The Assembly2Gene analysis showed that the *P. multocida* isolates contained orthologs of most virulence factors in the PastyVRDB but did not show evidence of mobile genetic elements or antibiotic resistance genes. *P. dagmatis*, *P. canis* and *P. stomatis* isolates had few matches in the Assembly2Gene analysis. All *P. multocida* isolates only contained a single filamentous haemagglutinin gene and cognate secretion partner gene, and between one and four of the known iron-receptor genes were absent in the *P. multocida* isolates. While genes encoding *flp1*, *ompA*, *ompH_1* and *nanB* were identified, they displayed variable levels of shared sequence identify with reference genes. Importantly, this analysis showed several *P. multocida* isolates lacked important genes required for capsule and Flp-pili biosynthesis. No other capsule biosynthesis systems were identified elsewhere on the genomes of these strains, suggesting that these strains cannot produce capsule. Given *P. multocida* capsule has been identified as critical for virulence in animal infections, future experiments should be focussed on confirming the lack of any known capsule biosynthesis genes in these isolates using PCR analysis. Isolates should also be assessed for capsule production using HA capsule assays, scanning electron microscopy and by crude polysaccharide extraction followed by Alcian blue staining (323). Confirmation that these strains lack capsule would indicate that capsule is not required for *P. multocida* to cause diseases in humans, although detailed information on the isolation site, disease presentation and extent of disease is required as it is possible many of these strains

were isolated from self-limiting infections. Lastly, the lack of the Flp pili biosynthesis genes should be confirmed by PCR and lack of the Flp1 subunit protein should be confirmed by mass spectrometry of outer membrane proteins.

The presence of putative virulence factors of *P. multocida* was assessed by pan-genome analysis. All homologs were identified between the 36 *P. multocida* isolates sequenced in this study and 260 *P. multocida* reference genomes via Roary. Genes over-represented for specific *P. multocida* strains were identified via Scoary. Given that previous publications have suggested that specific capsule and LPS types are predominantly associated with different animal diseases and host predilection (13), we used Roary and Scoary analysis to identify genes over-represented in particular capsule and LPS types. This analysis identified several putative virulence factors and pathogenic mechanisms. A putative glycosyltransferase with homology to an LPS O-antigen modifying enzyme from *E. coli* and a putative lipoprotein were over-represented in type B:L2 strains. A putative heavy metal efflux protein and putative regulator of heavy metal resistance genes was over-represented in capsule type F strains. Further studies would involve disrupting these genes and assessing virulence in mice. If mutants have reduced viability, the function of the proteins should be investigated. *P. multocida* type B:L2 strains are responsible for haemorrhagic septicaemia, so if the putative glycosyltransferase and lipoprotein were shown to attenuate virulence in mice, virulence of these mutants could be assessed in buffalo, which are highly susceptible to haemorrhagic septicaemia (36). Such experiments could not easily be performed in Australia, as *P. multocida* type B strains are not endemic to Australia.

A maximum likelihood core genome phylogeny of the 36 *P. multocida* isolates sequenced in this study and 260 *P. multocida* reference genomes was generated using Roary and IQ-TREE. The *P. multocida* phylogeny showed two distinct clades, with 37 strains in clade A and 269 strains in clade B. Strains in clade A were recovered from humans, cats, dogs and mice, with 26 of the 37 strains in clade A recovered from humans. Clade B only contained 6 human isolates, suggesting *P. multocida* strains that cause severe disease in humans are genetically distinct from strains that cause disease in most other animals. Genes over-represented in clade A and B were identified by Roary and Scoary analysis. A putative fucose uptake and utilisation pathway was identified as over-represented in clade A. L-fucose found in polysaccharides and glycoproteins on the surface of mammalian cells, and is commonly used to decorate mucins (287, 288). The ability of strains in clade A to utilise L-fucose as a carbon source may provide a competitive advantage for these strains, allowing them to cause severe disease within humans. Further experiments would involve generating directed mutants in uptake and utilisation genes followed by growth on modified *P. multocida* minimal media with L-fucose as the sole carbon source (324). Inability of these mutants to grow on L-fucose

minimal media would confirm these genes encode an L-fucose uptake and utilisation system. These mutants would then be assessed by *in vivo* competition assays in mice to confirm these systems are important for virulence. The putative L-fucose uptake and utilisation system has previously been predicted as a virulence factor in fowl cholera (122). While significantly under-represented in clade B strains, several *P. multocida* bird isolates contained the entire uptake and utilisation system, suggesting L-fucose utilisation may also contribute to fowl cholera severity. This could be confirmed by assessing virulence of L-fucose utilisation pathway mutants in chickens. Several of the genes over-represented in clade B were involved with anaerobic growth, suggesting these strains are more adapted to growth in anaerobic conditions. Confirmation of the importance of these genes could be confirmed by performing *in vitro* competition assays between wild-type and mutant strains in rich media both under aerobic and anaerobic growth conditions. Mutants in these anaerobic systems should have a fitness defect in anaerobic conditions but have no difference during aerobic growth.

In conclusion, the aim of this work was to investigate *P. multocida* virulence factors and pathogenesis mechanisms both through TraDIS and comparative genomics. TraDIS analysis fully characterised genes involved in *P. multocida* HA capsule production and for the first time identified the SR as a regulator of capsule production. Furthermore, this analysis fully characterised the genes required for a *P. multocida* capsule type A strain to grow in complement-active chicken serum, and showed that capsule is an important requirement. Bioinformatic analysis of novel isolates from human infections identified several putative virulence factors and pathogenic mechanisms, including a glycosyltransferase and lipoprotein possibly required for haemorrhagic septicaemia, and an L-fucose uptake and utilisation system that may mediate severe disease within both humans and birds. This knowledge will help form a strong theoretical base for the development of novel treatment and prophylactic strategies for *P. multocida* diseases, including providing live-attenuated vaccine candidates and identifying excellent vaccine targets, and will help to refine our understanding of transmission dynamics and severity of *P. multocida* outbreaks caused by different strains.

References

1. Lariviere S, Leblanc L, Mittal KR, Martineau GP. 1993. Comparison of isolation methods for the recovery of *Bordetella bronchiseptica* and *Pasteurella multocida* from the nasal cavities of piglets. J Clin Microbiol 31:364-7.
2. Dewhirst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L, Buckley CM, Davis IJ, Bennett ML, Marshall-Jones ZV. 2012. The canine oral microbiome. PLoS One 7:e36067.
3. Lowe BA, Marsh TL, Isaacs-Cosgrove N, Kirkwood RN, Kiupel M, Mulks MH. 2012. Defining the "core microbiome" of the microbial communities in the tonsils of healthy pigs. BMC Microbiol 12:20.
4. Muhairwa AP, Christensen JP, Bisgaard M. 2000. Investigations on the carrier rate of *Pasteurella multocida* in healthy commercial poultry flocks and flocks affected by fowl cholera. Avian Pathol 29:133-42.
5. Wilson BA, Ho M. 2013. *Pasteurella multocida*: from zoonosis to cellular microbiology. Clin Microbiol Rev 26:631-55.
6. Hirsh DC, Jessup DA, Snipes KP, Carpenter TE, Hird DW, McCapes RH. 1990. Characteristics of *Pasteurella multocida* isolated from waterfowl and associated avian species in California. J Wildl Dis 26:204-9.
7. Dousse F, Thomann A, Brodard I, Korczak BM, Schlatter Y, Kuhnert P, Miserez R, Frey J. 2008. Routine phenotypic identification of bacterial species of the family *Pasteurellaceae* isolated from animals. J Vet Diagn Invest 20:716-24.
8. Rimler RB, Rhoades KR. 1987. Serogroup F, a new capsule serogroup of *Pasteurella multocida*. J Clin Microbiol 25:615-8.
9. Townsend KM, Boyce JD, Chung JY, Frost AJ, Adler B. 2001. Genetic organization of *Pasteurella multocida* cap Loci and development of a multiplex capsular PCR typing system. J Clin Microbiol 39:924-9.
10. Heddleston KL, Gallagher JE, Rebers PA. 1972. Fowl cholera: gel diffusion precipitin test for serotyping *Pasteruella multocida* from avian species. Avian Dis 16:925-36.
11. Harper M, John M, Turni C, Edmunds M, St Michael F, Adler B, Blackall PJ, Cox AD, Boyce JD. 2015. Development of a rapid multiplex PCR assay to genotype *Pasteurella multocida* strains by use of the lipopolysaccharide outer core biosynthesis locus. J Clin Microbiol 53:477-85.
12. Harper M, Boyce JD. 2017. The myriad properties of *Pasteurella multocida* lipopolysaccharide. Toxins (Basel) 9.
13. Wilkie IW, Harper M, Boyce JD, Adler B. 2012. *Pasteurella multocida*: diseases and pathogenesis. Curr Top Microbiol Immunol 361:1-22.
14. Benkirane A, De Alwis M. 2002. Haemorrhagic septicaemia, its significance, prevention and control in Asia. VETERINARNI MEDICINA-PRAHA- 47:234-240.
15. De Alwis MC. 1999. Haemorrhagic septicaemia.
16. Fereidouni S, Freimanis GL, Orynbayev M, Ribeca P, Flannery J, King DP, Zuther S, Beer M, Höper D, Kydyrmanov A, Karamendin K, Kock R. 2019. Mass die-off of saiga antelopes, Kazakhstan, 2015. Emerg Infect Dis 25:1169-1176.
17. Woo YK, Kim JH. 2006. Fowl cholera outbreak in domestic poultry and epidemiological properties of *Pasteurella multocida* isolate. J Microbiol 44:344-53.
18. Shivachandra SB, Kumar AA, Gautam R, Saxena MK, Chaudhuri P, Srivastava SK. 2005. Detection of multiple strains of *Pasteurella multocida* in fowl cholera outbreaks by polymerase chain reaction-based typing. Avian Pathol 34:456-62.
19. Zhang P, Fegan N, Fraser I, Duffy P, Bowles RE, Gordon A, Ketterer PJ, Shinwari W, Blackall PJ. 2004. Molecular epidemiology of two fowl cholera outbreaks on a free-range chicken layer farm. J Vet Diagn Invest 16:458-60.
20. Strugnell BW, Dagleish MP, Bayne CW, Brown M, Ainsworth HL, Nicholas RA, Wood A, Hodgson JC. 2011. Investigations into an outbreak of corvid respiratory disease associated with *Pasteurella multocida*. Avian Pathol 40:329-36.

21. Matsumoto M, Strain JG, Engel HN. 1991. The fate of *Pasteurella multocida* after intratracheal inoculation into turkeys. Poult Sci 70:2259-66.
22. Bojesen AM, Petersen KD, Nielsen OL, Christensen JP, Bisgaard M. 2004. *Pasteurella multocida* infection in heterophil-depleted chickens. Avian Dis 48:463-70.
23. Hassanin HH, Toth TE, Eldimerdash MM, Siegel PB. 1995. Stimulation of avian respiratory phagocytes by *Pasteurella multocida*: effects of the route of exposure, bacterial dosage and strain, and the age of chickens. Vet Microbiol 46:401-13.
24. Mbuthia PG, Njagi LW, Nyaga PN, Bebora LC, Minga U, Kamundia J, Olsen JE. 2008. *Pasteurella multocida* in scavenging family chickens and ducks: carrier status, age susceptibility and transmission between species. Avian Pathol 37:51-7.
25. Petersen KD, Christensen JP, Permin A, Bisgaard M. 2001. Virulence of *Pasteurella multocida* subsp. multocida isolated from outbreaks of fowl cholera in wild birds for domestic poultry and game birds. Avian Pathol 30:27-31.
26. Wilkie IW, Grimes SE, O'Boyle D, Frost AJ. 2000. The virulence and protective efficacy for chickens of *Pasteurella multocida* administered by different routes. Vet Microbiol 72:57-68.
27. Mbuthia PG, Njagi LW, Nyaga PN, Bebora LC, Minga U, Christensen JP, Olsen JE. 2011. Time-course investigation of infection with a low virulent *Pasteurella multocida* strain in normal and immune-suppressed 12-week-old free-range chickens. Avian Pathol 40:629-37.
28. Samuel MD, Shadduck DJ, Goldberg DR, Johnson WP. 2005. Avian cholera in waterfowl: the role of lesser snow and ross's geese as disease carriers in the Playa Lakes Region. J Wildl Dis 41:48-57.
29. Poermadjaja B, Frost A. 2000. Phagocytic uptake and killing of virulent and avirulent strains of *Pasteurella multocida* of capsular serotype A by chicken macrophages. Vet Microbiol 72:163-71.
30. Botzler RG. 1991. Epizootiology of avian cholera in wildfowl. J Wildl Dis 27:367-95.
31. Blackall PJ, Pahoff JL, Marks D, Fegan N, Morrow CJ. 1995. Characterisation of *Pasteurella multocida* isolated from fowl cholera outbreaks on turkey farms. Aust Vet J 72:135-8.
32. Crawford R, Allwright D, He C. 1992. High mortality of Cape Cormorants (Phalacrocorax capensis) off western South Africa in 1991 caused by *Pasteurella multocida*. Colonial Waterbirds:236-238.
33. Orynbayev M, Sultankulova K, Sansyzbay A, Rystayeva R, Shorayeva K, Namet A, Fereidouni S, Ilgekbayeva G, Barakbayev K, Kopeyev S, Kock R. 2019. Biological characterization of *Pasteurella multocida* present in the Saiga population. BMC Microbiol 19:37.
34. Davies RL. 2004. Genetic diversity among *Pasteurella multocida* strains of avian, bovine, ovine and porcine origin from England and Wales by comparative sequence analysis of the 16S rRNA gene. Microbiology 150:4199-210.
35. Biswas A, Shivachandra SB, Saxena MK, Kumar AA, Singh VP, Srivastava SK. 2004. Molecular variability among strains of *Pasteurella multocida* isolated from an outbreak of haemorrhagic septicaemia in India. Vet Res Commun 28:287-98.
36. Annas S, Zamri-Saad M, Jesse FF, Zunita Z. 2014. New sites of localisation of *Pasteurella multocida* B:2 in buffalo surviving experimental haemorrhagic septicaemia. BMC Vet Res 10:88.
37. Horadagoda NU, Hodgson JC, Moon GM, Wijewardana TG, Eckersall PD. 2001. Role of endotoxin in the pathogenesis of haemorrhagic septicaemia in the buffalo. Microb Pathog 30:171-8.
38. Eriksen L, Aalbaek B, Leifsson PS, Basse A, Christiansen T, Eriksen E, Rimler RB. 1999. Hemorrhagic septicemia in fallow deer (Dama dama) caused by *Pasteurella multocida multocida*. J Zoo Wildl Med 30:285-92.
39. De Alwis MC. 1999. Haemorrhagic septicaemia. Australian Centre for International Agricultural Research.
40. Baalsrud KJ. 1987. Atrophic rhinitis in goats in Norway. Vet Rec 121:350-3.

41. DiGiacomo RF, Xu YM, Allen V, Hinton MH, Pearson GR. 1991. Naturally acquired *Pasteurella multocida* infection in rabbits: clinicopathological aspects. Can J Vet Res 55:234-8.
42. Bessone FA, Pérez MLS, Zielinski G, Dibarbora M, Conde MB, Cappuccio J, Alustiza F. 2019. Characterization and comparison of strains of *Pasteurella multocida* associated with cases of progressive atrophic rhinitis and porcine pneumonia in Argentina. Vet World 12:434-439.
43. Davies RL, MacCorquodale R, Baillie S, Caffrey B. 2003. Characterization and comparison of *Pasteurella multocida* strains associated with porcine pneumonia and atrophic rhinitis. J Med Microbiol 52:59-67.
44. Chanter N, Magyar T, Rutter JM. 1989. Interactions between *Bordetella bronchiseptica* and toxigenic *Pasteurella multocida* in atrophic rhinitis of pigs. Res Vet Sci 47:48-53.
45. Aminova LR, Luo S, Bannai Y, Ho M, Wilson BA. 2008. The C3 domain of *Pasteurella multocida* toxin is the minimal domain responsible for activation of Gq-dependent calcium and mitogenic signaling. Protein Sci 17:945-9.
46. Wilson BA, Ho M. 2012. *Pasteurella multocida* toxin interaction with host cells: entry and cellular effects. Curr Top Microbiol Immunol 361:93-111.
47. Felix R, Fleisch H, Frandsen PL. 1992. Effect of *Pasteurella multocida* toxin on bone resorption in vitro. Infect Immun 60:4984-8.
48. Mullan PB, Lax AJ. 1998. *Pasteurella multocida* toxin stimulates bone resorption by osteoclasts via interaction with osteoblasts. Calcif Tissue Int 63:340-5.
49. Martineau-Doize B, Dumas G, Larochelle R, Frantz JC, Martineau GP. 1991. Atrophic rhinitis caused by *Pasteurella multocida* type D: morphometric analysis. Can J Vet Res 55:224-8.
50. Dillehay DL, Paul KS, DiGiacomo RF, Chengappa MM. 1991. Pathogenicity of *Pasteurella multocida* A:3 in Flemish giant and New Zealand white rabbits. Lab Anim 25:337-41.
51. Nikunen S, Hartel H, Orro T, Neuvonen E, Tanskanen R, Kivela SL, Sankari S, Aho P, Pyorala S, Saloniemi H, Soveri T. 2007. Association of bovine respiratory disease with clinical status and acute phase proteins in calves. Comp Immunol Microbiol Infect Dis 30:143-51.
52. Welsh RD, Dye LB, Payton ME, Confer AW. 2004. Isolation and antimicrobial susceptibilities of bacterial pathogens from bovine pneumonia: 1994--2002. J Vet Diagn Invest 16:426-31.
53. Bethe A, Wieler LH, Selbitz HJ, Ewers C. 2009. Genetic diversity of porcine *Pasteurella multocida* strains from the respiratory tract of healthy and diseased swine. Vet Microbiol 139:97-105.
54. Autio T, Pohjanvirta T, Holopainen R, Rikula U, Pentikainen J, Huovilainen A, Rusanen H, Soveri T, Sihvonen L, Pelkonen S. 2007. Etiology of respiratory disease in non-vaccinated, non-medicated calves in rearing herds. Vet Microbiol 119:256-65.
55. Gagea MI, Bateman KG, van Dreumel T, McEwen BJ, Carman S, Archambault M, Shanahan RA, Caswell JL. 2006. Diseases and pathogens associated with mortality in Ontario beef feedlots. J Vet Diagn Invest 18:18-28.
56. Hirose K, Kobayashi H, Ito N, Kawasaki Y, Zako M, Kotani K, Ogawa H, Sato H. 2003. Isolation of *Mycoplasmas* from nasal swabs of calves affected with respiratory diseases and antimicrobial susceptibility of their isolates. J Vet Med B Infect Dis Vet Public Health 50:347-51.
57. Sudaryatma PE, Nakamura K, Mekata H, Sekiguchi S, Kubo M, Kobayashi I, Subangkit M, Goto Y, Okabayashi T. 2018. Bovine respiratory syncytial virus infection enhances *Pasteurella multocida* adherence on respiratory epithelial cells. Vet Microbiol 220:33-38.
58. Ujvári B, Weiczner R, Deim Z, Terhes G, Urbán E, Tóth AR, Magyar T. 2019. Characterization of *Pasteurella multocida* strains isolated from human infections. Comp Immunol Microbiol Infect Dis 63:37-43.

59. Talan DA, Citron DM, Abrahamian FM, Moran GJ, Goldstein EJ. 1999. Bacteriologic analysis of infected dog and cat bites. Emergency Medicine Animal Bite Infection Study Group. N Engl J Med 340:85-92.
60. Abrahamian FM, Goldstein EJ. 2011. Microbiology of animal bite wound infections. Clin Microbiol Rev 24:231-46.
61. Giordano A, Dincman T, Clyburn BE, Steed LL, Rockey DC. 2015. Clinical features and outcomes of *Pasteurella multocida* infection. Medicine (Baltimore) 94:e1285.
62. Chang K, Siu LK, Chen YH, Lu PL, Chen TC, Hsieh HC, Lin CL. 2007. Fatal *Pasteurella multocida* septicemia and necrotizing fasciitis related with wound licked by a domestic dog. Scand J Infect Dis 39:167-70.
63. Klein NC, Cunha BA. 1997. *Pasteurella multocida* pneumonia. Semin Respir Infect 12:54-6.
64. Kimura R, Hayashi Y, Takeuchi T, Shimizu M, Iwata M, Tanahashi J, Ito M. 2004. *Pasteurella multocida* septicemia caused by close contact with a domestic cat: case report and literature review. J Infect Chemother 10:250-2.
65. Kofteridis DP, Christofaki M, Mantadakis E, Maraki S, Drygiannakis I, Papadakis JA, Samonis G. 2009. Bacteremic community-acquired pneumonia due to *Pasteurella multocida*. Int J Infect Dis 13:e81-3.
66. Lion C, Conroy MC, Carpentier AM, Lozniewski A. 2006. Antimicrobial susceptibilities of *Pasteurella* strains isolated from humans. Int J Antimicrob Agents 27:290-3.
67. Pandit KK, Smith JE. 1993. Capsular hyaluronic acid in *Pasteurella multocida* type A and its counterpart in type D. Res Vet Sci 54:20-4.
68. Rimler RB. 1994. Presumptive identification of *Pasteurella multocida* serogroups A, D and F by capsule depolymerisation with mucopolysaccharidases. Vet Rec 134:191-2.
69. DeAngelis PL, Gunay NS, Toida T, Mao WJ, Linhardt RJ. 2002. Identification of the capsular polysaccharides of Type D and F *Pasteurella multocida* as unmodified heparin and chondroitin, respectively. Carbohydr Res 337:1547-52.
70. Michael FS, Cairns CM, Fleming P, Vinogradov EV, Boyce JD, Harper M, Cox AD. 2020. The capsular polysaccharides of *Pasteurella multocida* serotypes B and E: Structural, genetic and serological comparisons. Glycobiology doi:10.1093/glycob/cwaa069.
71. Frantz C, Stewart KM, Weaver VM. 2010. The extracellular matrix at a glance. J Cell Sci 123:4195-200.
72. DeAngelis PL. 2002. Evolution of glycosaminoglycans and their glycosyltransferases: Implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. Anat Rec 268:317-26.
73. Cuthbertson L, Mainprize IL, Naismith JH, Whitfield C. 2009. Pivotal roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in gram-negative bacteria. Microbiol Mol Biol Rev 73:155-77.
74. Willis LM, Whitfield C. 2013. Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. Carbohydr Res 378:35-44.
75. Boyce JD, Adler B. 2000. The capsule is a virulence determinant in the pathogenesis of *Pasteurella multocida* M1404 (B:2). Infect Immun 68:3463-8.
76. Chung JY, Wilkie I, Boyce JD, Townsend KM, Frost AJ, Ghoddusi M, Adler B. 2001. Role of capsule in the pathogenesis of fowl cholera caused by *Pasteurella multocida* serogroup A. Infect Immun 69:2487-92.
77. Boyce JD, Chung JY, Adler B. 2000. Genetic organisation of the capsule biosynthetic locus of *Pasteurella multocida* M1404 (B:2). Vet Microbiol 72:121-34.
78. Chung JY, Zhang Y, Adler B. 1998. The capsule biosynthetic locus of *Pasteurella multocida* A:1. FEMS Microbiol Lett 166:289-96.
79. Larue K, Ford RC, Willis LM, Whitfield C. 2011. Functional and structural characterization of polysaccharide co-polymerase proteins required for polymer

export in ATP-binding cassette transporter-dependent capsule biosynthesis pathways. J Biol Chem 286:16658-68.
80. Steenbergen SM, Vimr ER. 2008. Biosynthesis of the *Escherichia coli* K1 group 2 polysialic acid capsule occurs within a protected cytoplasmic compartment. Mol Microbiol 68:1252-67.
81. Willis LM, Whitfield C. 2013. KpsC and KpsS are retaining 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) transferases involved in synthesis of bacterial capsules. Proc Natl Acad Sci U S A 110:20753-8.
82. McNulty C, Thompson J, Barrett B, Lord L, Andersen C, Roberts IS. 2006. The cell surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. Mol Microbiol 59:907-22.
83. Jing W, DeAngelis PL. 2000. Dissection of the two transferase activities of the *Pasteurella multocida* hyaluronan synthase: two active sites exist in one polypeptide. Glycobiology 10:883-9.
84. DeAngelis PL, White CL. 2002. Identification and molecular cloning of a heparosan synthase from *Pasteurella multocida* type D. J Biol Chem 277:7209-13.
85. DeAngelis PL, Padgett-McCue AJ. 2000. Identification and molecular cloning of a chondroitin synthase from *Pasteurella multocida* type F. J Biol Chem 275:24124-9.
86. Kane TA, White CL, DeAngelis PL. 2006. Functional characterization of PmHS1, a *Pasteurella multocida* heparosan synthase. J Biol Chem 281:33192-7.
87. Jing W, DeAngelis PL. 2003. Analysis of the two active sites of the hyaluronan synthase and the chondroitin synthase of *Pasteurella multocida*. Glycobiology 13:661-71.
88. DeAngelis PL. 1999. Molecular directionality of polysaccharide polymerization by the *Pasteurella multocida* hyaluronan synthase. J Biol Chem 274:26557-62.
89. Chu X, Han J, Guo D, Fu Z, Liu W, Tao Y. 2016. Characterization of UDP-glucose dehydrogenase from *Pasteurella multocida* CVCC 408 and its application in hyaluronic acid biosynthesis. Enzyme Microb Technol 85:64-70.
90. Deangelis PL, White CL. 2004. Identification of a distinct, cryptic heparosan synthase from *Pasteurella multocida* types A, D, and F. J Bacteriol 186:8529-32.
91. Watt JM, Swiatlo E, Wade MM, Champlin FR. 2003. Regulation of capsule biosynthesis in serotype A strains of *Pasteurella multocida*. FEMS Microbiol Lett 225:9-14.
92. Steen JA, Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD. 2010. Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. PLoS Pathog 6:e1000750.
93. Ninnemann O, Koch C, Kahmann R. 1992. The *E.coli fis* promoter is subject to stringent control and autoregulation. EMBO J 11:1075-83.
94. Bradley MD, Beach MB, de Koning APJ, Pratt TS, Osuna R. 2007. Effects of Fis on *Escherichia coli* gene expression during different growth stages. Microbiology (Reading) 153:2922-2940.
95. Vogel J, Luisi BF. 2011. Hfq and its constellation of RNA. Nat Rev Microbiol 9:578-89.
96. Megroz M, Kleifeld O, Wright A, Powell D, Harrison P, Adler B, Harper M, Boyce JD. 2016. The RNA-binding chaperone Hfq is an important global regulator of gene expression in *Pasteurella multocida* and plays a crucial role in production of a number of virulence factors including hyaluronic acid capsule. Infect Immun 84:1361-70.
97. Fuller TE, Kennedy MJ, Lowery DE. 2000. Identification of *Pasteurella multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis. Microb Pathog 29:25-38.
98. Raetz CR, Whitfield C. 2002. Lipopolysaccharide endotoxins. Annu Rev Biochem 71:635-700.

99. Harper M, St Michael F, John M, Vinogradov E, Adler B, Boyce JD, Cox AD. 2011. *Pasteurella multocida* Heddleston serovars 1 and 14 express different lipopolysaccharide structures but share the same lipopolysaccharide biosynthesis outer core locus. Vet Microbiol 150:289-96.
100. St Michael F, Harper M, Parnas H, John M, Stupak J, Vinogradov E, Adler B, Boyce JD, Cox AD. 2009. Structural and genetic basis for the serological differentiation of *Pasteurella multocida* Heddleston serotypes 2 and 5. J Bacteriol 191:6950-9.
101. Harper M, St Michael F, John M, Vinogradov E, Steen JA, van Dorsten L, Steen JA, Turni C, Blackall PJ, Adler B, Cox AD, Boyce JD. 2013. *Pasteurella multocida* Heddleston serovar 3 and 4 strains share a common lipopolysaccharide biosynthesis locus but display both inter- and intrastrain lipopolysaccharide heterogeneity. J Bacteriol 195:4854-64.
102. Harper M, St Michael F, Steen JA, John M, Wright A, van Dorsten L, Vinogradov E, Adler B, Cox AD, Boyce JD. 2015. Characterization of the lipopolysaccharide produced by *Pasteurella multocida* serovars 6, 7 and 16: identification of lipopolysaccharide genotypes L4 and L8. Glycobiology 25:294-302.
103. Harper M, St Michael F, Vinogradov E, John M, Boyce JD, Adler B, Cox AD. 2012. Characterization of the lipopolysaccharide from *Pasteurella multocida* Heddleston serovar 9: identification of a proposed bi-functional dTDP-3-acetamido-3,6-dideoxy-α-D-glucose biosynthesis enzyme. Glycobiology 22:332-44.
104. Harper M, St Michael F, John M, Steen J, van Dorsten L, Parnas H, Vinogradov E, Adler B, Cox AD, Boyce JD. 2014. Structural analysis of lipopolysaccharide produced by Heddleston serovars 10, 11, 12 and 15 and the identification of a new *Pasteurella multocida* lipopolysaccharide outer core biosynthesis locus, L6. Glycobiology 24:649-59.
105. Harper M, Boyce JD, Cox AD, St Michael F, Wilkie IW, Blackall PJ, Adler B. 2007. *Pasteurella multocida* expresses two lipopolysaccharide glycoforms simultaneously, but only a single form is required for virulence: identification of two acceptor-specific heptosyl I transferases. Infect Immun 75:3885-93.
106. Harper M, Boyce JD, Wilkie IW, Adler B. 2003. Signature-tagged mutagenesis of *Pasteurella multocida* identifies mutants displaying differential virulence characteristics in mice and chickens. Infect Immun 71:5440-6.
107. Harper M, Cox AD, St Michael F, Wilkie IW, Boyce JD, Adler B. 2004. A heptosyltransferase mutant of *Pasteurella multocida* produces a truncated lipopolysaccharide structure and is attenuated in virulence. Infect Immun 72:3436-43.
108. Lax AJ, Chanter N. 1990. Cloning of the toxin gene from *Pasteurella multocida* and its role in atrophic rhinitis. J Gen Microbiol 136:81-7.
109. Pullinger GD, Bevir T, Lax AJ. 2004. The *Pasteurella multocida* toxin is encoded within a lysogenic bacteriophage. Mol Microbiol 51:255-69.
110. Peng Z, Wang X, Zhou R, Chen H, Wilson BA, Wu B. 2019. *Pasteurella multocida*: Genotypes and Genomics. Microbiol Mol Biol Rev 83.
111. iDali C, Foged NT, Frandsen PL, Nielsen MH, Elling F. 1991. Ultrastructural localization of the *Pasteurella multocida* toxin in a toxin-producing strain. J Gen Microbiol 137:1067-71.
112. Clemons NC, Bannai Y, Haywood EE, Xu Y, Buschbach JD, Ho M, Wilson BA. 2018. Cytosolic Delivery of Multidomain Cargos by the N Terminus of *Pasteurella multocida* Toxin. Infect Immun 86.
113. Kitadokoro K, Kamitani S, Miyazawa M, Hanajima-Ozawa M, Fukui A, Miyake M, Horiguchi Y. 2007. Crystal structures reveal a thiol protease-like catalytic triad in the C-terminal region of *Pasteurella multocida* toxin. Proc Natl Acad Sci U S A 104:5139-44.
114. Kamitani S, Kitadokoro K, Miyazawa M, Toshima H, Fukui A, Abe H, Miyake M, Horiguchi Y. 2010. Characterization of the membrane-targeting C1 domain in *Pasteurella multocida* toxin. J Biol Chem 285:25467-75.

115. Oldham WM, Hamm HE. 2008. Heterotrimeric G protein activation by G-protein-coupled receptors. Nat Rev Mol Cell Biol 9:60-71.
116. Orth JH, Preuss I, Fester I, Schlosser A, Wilson BA, Aktories K. 2009. *Pasteurella multocida* toxin activation of heterotrimeric G proteins by deamidation. Proc Natl Acad Sci U S A 106:7179-84.
117. Tatum FM, Yersin AG, Briggs RE. 2005. Construction and virulence of a *Pasteurella multocida fha*B2 mutant in turkeys. Microb Pathog 39:9-17.
118. May BJ, Zhang Q, Li LL, Paustian ML, Whittam TS, Kapur V. 2001. Complete genomic sequence of *Pasteurella multocida*, Pm70. Proc Natl Acad Sci U S A 98:3460-5.
119. Serra DO, Conover MS, Arnal L, Sloan GP, Rodriguez ME, Yantorno OM, 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:e28811.
120. Scheller EV, Melvin JA, Sheets AJ, Cotter PA. 2015. Cooperative roles for fimbria and filamentous hemagglutinin in *Bordetella* adherence and immune modulation. MBio 6:e00500-15.
121. Vakevainen M, Greenberg S, Hansen EJ. 2003. Inhibition of phagocytosis by *Haemophilus ducreyi* requires expression of the LspA1 and LspA2 proteins. Infect Immun 71:5994-6003.
122. Johnson TJ, Abrahante JE, Hunter SS, Hauglund M, Tatum FM, Maheswaran SK, Briggs RE. 2013. Comparative genome analysis of an avirulent and two virulent strains of avian *Pasteurella multocida* reveals candidate genes involved in fitness and pathogenicity. BMC Microbiol 13:106.
123. Ruffolo CG, Tennent JM, Michalski WP, Adler B. 1997. Identification, purification, and characterization of the type 4 fimbriae of *Pasteurella multocida*. Infect Immun 65:339-43.
124. Doughty SW, Ruffolo CG, Adler B. 2000. The type 4 fimbrial subunit gene of *Pasteurella multocida*. Vet Microbiol 72:79-90.
125. Boyce JD, Wilkie I, Harper M, Paustian ML, Kapur V, Adler B. 2002. Genomic scale analysis of *Pasteurella multocida* gene expression during growth within the natural chicken host. Infect Immun 70:6871-9.
126. Tomich M, Planet PJ, Figurski DH. 2007. The *tad* locus: postcards from the widespread colonization island. Nat Rev Microbiol 5:363-75.
127. Kachlany SC, Planet PJ, Bhattacharjee MK, Kollia E, DeSalle R, Fine DH, Figurski DH. 2000. Nonspecific adherence by *Actinobacillus actinomycetemcomitans* requires genes widespread in bacteria and archaea. J Bacteriol 182:6169-76.
128. Kachlany SC, Planet PJ, Desalle R, Fine DH, Figurski DH, Kaplan JB. 2001. flp-1, the first representative of a new pilin gene subfamily, is required for non-specific adherence of *Actinobacillus actinomycetemcomitans*. Mol Microbiol 40:542-54.
129. Perez BA, Planet PJ, Kachlany SC, Tomich M, Fine DH, Figurski DH. 2006. Genetic analysis of the requirement for *flp-2*, *tadV*, and *rcpB* in *Actinobacillus actinomycetemcomitans* biofilm formation. J Bacteriol 188:6361-75.
130. Planet PJ, Kachlany SC, Fine DH, DeSalle R, Figurski DH. 2003. The Widespread Colonization Island of *Actinobacillus actinomycetemcomitans*. Nat Genet 34:193-8.
131. Spinola SM, Fortney KR, Katz BP, Latimer JL, Mock JR, Vakevainen M, Hansen EJ. 2003. *Haemophilus ducreyi* requires an intact *flp* gene cluster for virulence in humans. Infect Immun 71:7178-82.
132. Hatfaludi T, Al-Hasani K, Boyce JD, Adler B. 2010. Outer membrane proteins of *Pasteurella multocida*. Vet Microbiol 144:1-17.
133. Dabo SM, Confer AW, Quijano-Blas RA. 2003. Molecular and immunological characterization of Pasteurella multocida serotype A:3 OmpA: evidence of its role in *P. multocida* interaction with extracellular matrix molecules. Microb Pathog 35:147-57.

134. Mullen LM, Nair SP, Ward JM, Rycroft AN, Williams RJ, Robertson G, Mordan NJ, Henderson B. 2008. Novel adhesin from *Pasteurella multocida* that binds to the integrin-binding fibronectin FnIII9-10 repeats. Infect Immun 76:1093-104.
135. Mullen LM, Bossé JT, Nair SP, Ward JM, Rycroft AN, Robertson G, Langford PR, Henderson B. 2008. Pasteurellaceae ComE1 proteins combine the properties of fibronectin adhesins and DNA binding competence proteins. PLoS One 3:e3991.
136. Al-Hasani K, Boyce J, McCarl VP, Bottomley S, Wilkie I, Adler B. 2007. Identification of novel immunogens in *Pasteurella multocida*. Microb Cell Fact 6:3.
137. Hatfaludi T, Al-Hasani K, Gong L, Boyce JD, Ford M, Wilkie IW, Quinsey N, Dunstone MA, Hoke DE, Adler B. 2012. Screening of 71 *P. multocida* proteins for protective efficacy in a fowl cholera infection model and characterization of the protective antigen PlpE. PLoS One 7:e39973.
138. Wu JR, Shien JH, Shieh HK, Chen CF, Chang PC. 2007. Protective immunity conferred by recombinant *Pasteurella multocida* lipoprotein E (PlpE). Vaccine 25:4140-8.
139. Bosch M, Garrido ME, Pérez de Rozas AM, Badiola I, Barbé J, Llagostera M. 2004. *Pasteurella multocida* contains multiple immunogenic haemin- and haemoglobin-binding proteins. Vet Microbiol 99:103-12.
140. Cox AJ, Hunt ML, Boyce JD, Adler B. 2003. Functional characterization of HgbB, a new hemoglobin binding protein of *Pasteurella multocida*. Microb Pathog 34:287-96.
141. Garrido ME, Bosch M, Medina R, Bigas A, Llagostera M, Pérez de Rozas AM, Badiola I, Barbé J. 2003. *fur*-independent regulation of the *Pasteurella multocida hbpA* gene encoding a haemin-binding protein. Microbiology (Reading) 149:2273-2281.
142. Prado ME, Dabo SM, Confer AW. 2005. Immunogenicity of iron-regulated outer membrane proteins of *Pasteurella multocida* A:3 in cattle: molecular characterization of the immunodominant heme acquisition system receptor (HasR) protein. Vet Microbiol 105:269-80.
143. Bosch M, Garrido ME, Llagostera M, Pérez De Rozas AM, Badiola I, Barbé J. 2002. Characterization of the *Pasteurella multocida hgbA* gene encoding a hemoglobin-binding protein. Infect Immun 70:5955-64.
144. Ogunnariwo JA, Schryvers AB. 2001. Characterization of a novel transferrin receptor in bovine strains of *Pasteurella multocida*. J Bacteriol 183:890-6.
145. Krewulak KD, Vogel HJ. 2008. Structural biology of bacterial iron uptake. Biochim Biophys Acta 1778:1781-804.
146. Troxell B, Hassan HM. 2013. Transcriptional regulation by Ferric Uptake Regulator (Fur) in pathogenic bacteria. Front Cell Infect Microbiol 3:59.
147. Liu Q, Hu Y, Li P, Kong Q. 2019. Identification of Fur in *Pasteurella multocida* and the Potential of Its Mutant as an Attenuated Live Vaccine. Front Vet Sci 6:5.
148. Paustian ML, May BJ, Kapur V. 2001. *Pasteurella multocida* gene expression in response to iron limitation. Infect Immun 69:4109-15.
149. Paustian ML, May BJ, Cao D, Boley D, Kapur V. 2002. Transcriptional response of *Pasteurella multocida* to defined iron sources. J Bacteriol 184:6714-20.
150. Mizan S, Henk A, Stallings A, Maier M, Lee MD. 2000. Cloning and characterization of sialidases with 2-6' and 2-3' sialyl lactose specificity from Pasteurella multocida. J Bacteriol 182:6874-83.
151. Steenbergen SM, Lichtensteiger CA, Caughlan R, Garfinkle J, Fuller TE, Vimr ER. 2005. Sialic Acid metabolism and systemic pasteurellosis. Infect Immun 73:1284-94.
152. Tatum FM, Tabatabai LB, Briggs RE. 2009. Sialic acid uptake is necessary for virulence of Pasteurella multocida in turkeys. Microb Pathog 46:337-44.
153. Bouchet V, Hood DW, Li J, Brisson JR, Randle GA, Martin A, Li Z, Goldstein R, Schweda EK, Pelton SI, Richards JC, Moxon ER. 2003. Host-derived sialic acid is incorporated into Haemophilus influenzae lipopolysaccharide and is a major virulence factor in experimental otitis media. Proc Natl Acad Sci U S A 100:8898-903.

154. Katsuda K, Hoshinoo K, Ueno Y, Kohmoto M, Mikami O. 2013. Virulence genes and antimicrobial susceptibility in *Pasteurella multocida* isolates from calves. Vet Microbiol 167:737-41.
155. Shivachandra SB, Kumar AA, Biswas A, Ramakrishnan MA, Singh VP, Srivastava SK. 2004. Antibiotic sensitivity patterns among Indian strains of avian *Pasteurella multocida*. Trop Anim Health Prod 36:743-50.
156. Ares-Arroyo M, Bernabe-Balas C, Santos-Lopez A, Baquero MR, Prasad KN, Cid D, Martin-Espada C, San Millan A, Gonzalez-Zorn B. 2018. PCR-Based analysis of ColE1 plasmids in clinical isolates and metagenomic samples reveals their importance as gene capture platforms. Front Microbiol 9:469.
157. Kehrenberg C, Catry B, Haesebrouck F, de Kruif A, Schwarz S. 2005. Novel spectinomycin/streptomycin resistance gene, *aadA14*, from *Pasteurella multocida*. Antimicrob Agents Chemother 49:3046-9.
158. Kehrenberg C, Wallmann J, Schwarz S. 2008. Molecular analysis of florfenicol-resistant *Pasteurella multocida* isolates in Germany. J Antimicrob Chemother 62:951-5.
159. Kehrenberg C, Schwarz S. 2005. Plasmid-borne florfenicol resistance in *Pasteurella multocida*. J Antimicrob Chemother 55:773-5.
160. Wu JR, Shieh HK, Shien JH, Gong SR, Chang PC. 2003. Molecular characterization of plasmids with antimicrobial resistant genes in avian isolates of *Pasteurella multocida*. Avian Dis 47:1384-92.
161. San Millan A, Escudero JA, Gutierrez B, Hidalgo L, Garcia N, Llagostera M, Dominguez L, Gonzalez-Zorn B. 2009. Multiresistance in *Pasteurella multocida* is mediated by coexistence of small plasmids. Antimicrob Agents Chemother 53:3399-404.
162. Michael GB, Kadlec K, Sweeney MT, Brzuszkiewicz E, Liesegang H, Daniel R, Murray RW, Watts JL, Schwarz S. 2012. ICEPmu1, an integrative conjugative element (ICE) of *Pasteurella multocida*: analysis of the regions that comprise 12 antimicrobial resistance genes. J Antimicrob Chemother 67:84-90.
163. Kadlec K, Brenner Michael G, Sweeney MT, Brzuszkiewicz E, Liesegang H, Daniel R, Watts JL, Schwarz S. 2011. Molecular basis of macrolide, triamilide, and lincosamide resistance in *Pasteurella multocida* from bovine respiratory disease. Antimicrob Agents Chemother 55:2475-7.
164. Moustafa AM, Seemann T, Gladman S, Adler B, Harper M, Boyce JD, Bennett MD. 2015. Comparative Genomic Analysis of Asian Haemorrhagic Septicaemia-Associated Strains of *Pasteurella multocida* Identifies More than 90 Haemorrhagic Septicaemia-Specific Genes. PLoS One 10:e0130296.
165. Peng Z, Liang W, Wang Y, Liu W, Zhang H, Yu T, Zhang A, Chen H, Wu B. 2017. Experimental pathogenicity and complete genome characterization of a pig origin *Pasteurella multocida* serogroup F isolate HN07. Vet Microbiol 198:23-33.
166. Peng Z, Liang W, Wang F, Xu Z, Xie Z, Lian Z, Hua L, Zhou R, Chen H, Wu B. 2018. Genetic and Phylogenetic Characteristics of *Pasteurella multocida* Isolates From Different Host Species. Front Microbiol 9:1408.
167. Smith E, Miller E, Aguayo JM, Figueroa CF, Nezworski J, Studniski M, Wileman B, Johnson T. 2021. Genomic diversity and molecular epidemiology of *Pasteurella multocida*. PLoS One 16:e0249138.
168. Bisgaard M, Petersen A, Christensen H. 2013. Multilocus sequence analysis of *Pasteurella multocida* demonstrates a type species under development. Microbiology (Reading) 159:580-590.
169. Hurtado R, Maturrano L, Azevedo V, Aburjaile F. 2020. Pathogenomics insights for understanding *Pasteurella multocida* adaptation. Int J Med Microbiol 310:151417.
170. Parte AC. 2014. LPSN--list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 42:D613-6.
171. Akahane T, Nagata M, Matsumoto T, Murayama T, Isaka A, Kameda T, Fujita M, Oana K, Kawakami Y. 2011. A case of wound dual infection with *Pasteurella*

dagmatis and *Pasteurella canis* resulting from a dog bite -- limitations of Vitek-2 system in exact identification of Pasteurella species. Eur J Med Res 16:531-6.
172. Ejlertsen T, Gahrn-Hansen B, Søgaard P, Heltberg O, Frederiksen W. 1996. *Pasteurella aerogenes* isolated from ulcers or wounds in humans with occupational exposure to pigs: a report of 7 Danish cases. Scand J Infect Dis 28:567-70.
173. Escande F, Vallee E, Aubart F. 1997. *Pasteurella caballi* infection following a horse bite. Zentralbl Bakteriol 285:440-4.
174. Freeman AF, Zheng XT, Lane JC, Shulman ST. 2004. *Pasteurella aerogenes* hamster bite peritonitis. Pediatr Infect Dis J 23:368-70.
175. Holst E, Rollof J, Larsson L, Nielsen JP. 1992. Characterization and distribution of *Pasteurella* species recovered from infected humans. J Clin Microbiol 30:2984-7.
176. Church S, Harrigan KE, Irving AE, Peel MM. 1998. Endocarditis caused by Pasteurella caballi in a horse. Aust Vet J 76:528-30.
177. Christensen H, Kuhnert P, Olsen JE, Bisgaard M. 2004. Comparative phylogenies of the housekeeping genes *atpD*, *infB* and *rpoB* and the 16S rRNA gene within the *Pasteurellaceae*. Int J Syst Evol Microbiol 54:1601-9.
178. Korczak B, Christensen H, Emler S, Frey J, Kuhnert P. 2004. Phylogeny of the family *Pasteurellaceae* based on *rpoB* sequences. Int J Syst Evol Microbiol 54:1393-9.
179. Paustian ML, May BJ, Kapur V. 2002. Transcriptional response of *Pasteurella multocida* to nutrient limitation. J Bacteriol 184:3734-9.
180. Gulliver EL, Wright A, Lucas DD, Mégroz M, Kleifeld O, Schittenhelm RB, Powell DR, Seemann T, Bulitta JB, Harper M, Boyce JD. 2018. Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen *Pasteurella multocida* and identification of the GcvB seed binding region. RNA 24:704-720.
181. He F, Zhao Z, Wu X, Duan L, Li N, Fang R, Li P, Peng Y. 2021. Transcriptomic Analysis of High- and Low-Virulence Bovine *Pasteurella multocida* in vitro and in vivo. Front Vet Sci 8:616774.
182. Cummins J, Gahan CG. 2012. Signature tagged mutagenesis in the functional genetic analysis of gastrointestinal pathogens. Gut Microbes 3:93-103.
183. Chao MC, Abel S, Davis BM, Waldor MK. 2016. The design and analysis of transposon insertion sequencing experiments. Nat Rev Microbiol 14:119-28.
184. van Opijnen T, Camilli A. 2013. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol 11:435-42.
185. Gawronski JD, Wong SM, Giannoukos G, Ward DV, Akerley BJ. 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc Natl Acad Sci U S A 106:16422-7.
186. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI. 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe 6:279-89.
187. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. Genome Res 19:2308-16.
188. van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6:767-72.
189. Barquist L, Boinett CJ, Cain AK. 2013. Approaches to querying bacterial genomes with transposon-insertion sequencing. RNA Biol 10:1161-9.
190. Shevchenko Y, Bouffard GG, Butterfield YS, Blakesley RW, Hartley JL, Young AC, Marra MA, Jones SJ, Touchman JW, Green ED. 2002. Systematic sequencing of cDNA clones using the transposon Tn*5*. Nucleic Acids Res 30:2469-77.
191. Plasterk RH, Izsvak Z, Ivics Z. 1999. Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet 15:326-32.
192. Lampe DJ, Churchill ME, Robertson HM. 1996. A purified mariner transposase is sufficient to mediate transposition in vitro. EMBO J 15:5470-9.

193. Kwon YM, Ricke SC, Mandal RK. 2016. Transposon sequencing: methods and expanding applications. Appl Microbiol Biotechnol 100:31-43.
194. Morgan RD, Dwinell EA, Bhatia TK, Lang EM, Luyten YA. 2009. The *Mme*I family: type II restriction-modification enzymes that employ single-strand modification for host protection. Nucleic Acids Res 37:5208-21.
195. Klein BA, Tenorio EL, Lazinski DW, Camilli A, Duncan MJ, Hu LT. 2012. Identification of essential genes of the periodontal pathogen *Porphyromonas gingivalis*. BMC Genomics 13:578.
196. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA, Parkhill J. 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. Bioinformatics 32:1109-11.
197. Gallagher LA, Shendure J, Manoil C. 2011. Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. MBio 2:e00315-10.
198. Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J, van Opijnen T. 2020. A decade of advances in transposon-insertion sequencing. Nat Rev Genet 21:526-540.
199. Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, Lund PA, Cole JA, Henderson IR. 2018. The essential genome of *Escherichia coli* K-12. mBio 9.
200. Remmele CW, Xian Y, Albrecht M, Faulstich M, Fraunholz M, Heinrichs E, Dittrich MT, Muller T, Reinhardt R, Rudel T. 2014. Transcriptional landscape and essential genes of *Neisseria gonorrhoeae*. Nucleic Acids Res 42:10579-95.
201. Lee SA, Gallagher LA, Thongdee M, Staudinger BJ, Lippman S, Singh PK, Manoil C. 2015. General and condition-specific essential functions of *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 112:5189-94.
202. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ, McAdams HH, Shapiro L. 2011. The essential genome of a bacterium. Mol Syst Biol 7:528.
203. Phan MD, Peters KM, Sarkar S, Lukowski SW, Allsopp LP, Gomes Moriel D, Achard ME, Totsika M, Marshall VM, Upton M, Beatson SA, Schembri MA. 2013. The serum resistome of a globally disseminated multidrug resistant uropathogenic *Escherichia coli* clone. PLoS Genet 9:e1003834.
204. Dong TG, Ho BT, Yoder-Himes DR, Mekalanos JJ. 2013. Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. Proc Natl Acad Sci U S A 110:2623-8.
205. Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF, Fagan RP. 2015. High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. MBio 6:e02383.
206. Kakkanat A, Phan MD, Lo AW, Beatson SA, Schembri MA. 2017. Novel genes associated with enhanced motility of *Escherichia coli* ST131. PLoS One 12:e0176290.
207. Cowley LA, Low AS, Pickard D, Boinett CJ, Dallman TJ, Day M, Perry N, Gally DL, Parkhill J, Jenkins C, Cain AK. 2018. Transposon insertion sequencing elucidates novel gene involvement in susceptibility and resistance to phages T4 and T7 in *Escherichia coli* O157. mBio 9.
208. Hassan KA, Cain AK, Huang T, Liu Q, Elbourne LD, Boinett CJ, Brzoska AJ, Li L, Ostrowski M, Nhu NT, Nhu Tdo H, Baker S, Parkhill J, Paulsen IT. 2016. Fluorescence-based flow sorting in parallel with transposon insertion site sequencing identifies multidrug efflux systems in *Acinetobacter baumannii*. mBio 7.
209. Diallo IS, Frost AJ. 2000. Survival of avian strains of *Pasteurella multocida* in chicken serum. Vet Microbiol 72:153-61.
210. Doorduijn DJ, Rooijakkers SH, van Schaik W, Bardoel BW. 2016. Complement resistance mechanisms of *Klebsiella pneumoniae*. Immunobiology 221:1102-9.
211. Blom AM, Hallström T, Riesbeck K. 2009. Complement evasion strategies of pathogens-acquisition of inhibitors and beyond. Mol Immunol 46:2808-17.

212. Patarakul K, Lo M, Adler B. 2010. Global transcriptomic response of *Leptospira interrogans* serovar Copenhageni upon exposure to serum. BMC Microbiol 10:31.
213. Ma J, An C, Jiang F, Yao H, Logue C, Nolan LK, Li G. 2018. Extraintestinal pathogenic *Escherichia coli* increase extracytoplasmic polysaccharide biosynthesis for serum resistance in response to bloodstream signals. Mol Microbiol 110:689-706.
214. Sanchez-Larrayoz AF, Elhosseiny NM, Chevrette MG, Fu Y, Giunta P, Spallanzani RG, Ravi K, Pier GB, Lory S, Maira-Litrán T. 2017. Complexity of Complement Resistance Factors Expressed by *Acinetobacter baumannii* Needed for Survival in Human Serum. J Immunol 199:2803-2814.
215. Short FL, Di Sario G, Reichmann NT, Kleanthous C, Parkhill J, Taylor PW. 2020. Genomic profiling reveals distinct routes to complement resistance in *Klebsiella pneumoniae*. Infect Immun 88.
216. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726-731.
217. Shimada T, Fujita N, Yamamoto K, Ishihama A. 2011. Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. PLoS One 6:e20081.
218. Jacques M, Graham L. 1989. Improved preservation of bacterial capsule for electron microscopy. J Electron Microsc Tech 11:167-9.
219. Cascales E, Bernadac A, Gavioli M, Lazzaroni JC, Lloubes R. 2002. Pal lipoprotein of *Escherichia coli* plays a major role in outer membrane integrity. J Bacteriol 184:754-9.
220. Yeh YC, Comolli LR, Downing KH, Shapiro L, McAdams HH. 2010. The *Caulobacter* Tol-Pal complex is essential for outer membrane integrity and the positioning of a polar localization factor. J Bacteriol 192:4847-58.
221. Jensen PR, Michelsen O. 1992. Carbon and energy metabolism of atp mutants of *Escherichia coli*. J Bacteriol 174:7635-41.
222. Vik SB, Long JC, Wada T, Zhang D. 2000. A model for the structure of subunit a of the *Escherichia coli* ATP synthase and its role in proton translocation. Biochim Biophys Acta 1458:457-66.
223. Wallace BJ, Young IG. 1977. Role of quinones in electron transport to oxygen and nitrate in *Escherichia coli*. Studies with a *ubiA- menA-* double quinone mutant. Biochim Biophys Acta 461:84-100.
224. Wu G, Williams HD, Zamanian M, Gibson F, Poole RK. 1992. Isolation and characterization of *Escherichia coli* mutants affected in aerobic respiration: the cloning and nucleotide sequence of *ubiG*. Identification of an S-adenosylmethionine-binding motif in protein, RNA, and small-molecule methyltransferases. J Gen Microbiol 138:2101-12.
225. Hajj Chehade M, Pelosi L, Fyfe CD, Loiseau L, Rascalou B, Brugière S, Kazemzadeh K, Vo CD, Ciccone L, Aussel L, Couté Y, Fontecave M, Barras F, Lombard M, Pierrel F. 2019. A soluble metabolon synthesizes the isoprenoid lipid ubiquinone. Cell Chem Biol 26:482-492.e7.
226. Imamura R, Yamanaka K, Ogura T, Hiraga S, Fujita N, Ishihama A, Niki H. 1996. Identification of the *cpdA* gene encoding cyclic 3',5'-adenosine monophosphate phosphodiesterase in *Escherichia coli*. J Biol Chem 271:25423-9.
227. Franchini AG, Ihssen J, Egli T. 2015. Effect of global regulators RpoS and Cyclic-AMP/CRP on the catabolome and transcriptome of *Escherichia coli* K12 during carbon- and energy-limited growth. PLoS One 10:e0133793.
228. Fuchs EL, Brutinel ED, Klem ER, Fehr AR, Yahr TL, Wolfgang MC. 2010. In vitro and in vivo characterization of the *Pseudomonas aeruginosa* cyclic AMP (cAMP) phosphodiesterase CpdA, required for cAMP homeostasis and virulence factor regulation. J Bacteriol 192:2779-90.

229. Lin CT, Chen YC, Jinn TR, Wu CC, Hong YM, Wu WH. 2013. Role of the cAMP-dependent carbon catabolite repression in capsular polysaccharide biosynthesis in *Klebsiella pneumoniae*. PLoS One 8:e54430.
230. Hufnagel DA, Evans ML, Greene SE, Pinkner JS, Hultgren SJ, Chapman MR. 2016. The catabolite repressor protein-cyclic AMP complex regulates *csgD* and biofilm formation in uropathogenic *Escherichia coli*. J Bacteriol 198:3329-3334.
231. Ito M, Guffanti AA, Zemsky J, Ivey DM, Krulwich TA. 1997. Role of the *nhaC*-encoded Na+/H+ antiporter of alkaliphilic *Bacillus firmus* OF4. J Bacteriol 179:3851-7.
232. Minato Y, Ghosh A, Faulkner WJ, Lind EJ, Schesser Bartra S, Plano GV, Jarrett CO, Hinnebusch BJ, Winogrodzki J, Dibrov P, Häse CC. 2013. Na+/H+ antiport is essential for *Yersinia pestis* virulence. Infect Immun 81:3163-72.
233. Boyce J, Harper M, Wilkie I, Adler B. 2010. *Pasteurella*. Pathogenesis of Bacterial Infections in Animals, Fourth Edition:325-346.
234. Kuhnert P, Boerlin P, Emler S, Krawinkler M, Frey J. 2000. Phylogenetic analysis of *Pasteurella multocida* subspecies and molecular identification of feline *P. multocida* subsp. *septica* by 16S rRNA gene sequencing. Int J Med Microbiol 290:599-604.
235. O'Neill E, Moloney A, Hickey M. 2005. *Pasteurella multocida* meningitis: case report and review of the literature. J Infect 50:344-345.
236. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.
237. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.
238. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.
239. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126-31.
240. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895-903.
241. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res 44:D694-7.
242. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FS, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566-d573.
243. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640-4.
244. Whitfield C. 2006. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. Annu Rev Biochem 75:39-68.
245. Ujvári B, Makrai L, Magyar T. 2019. Virulence gene profiling and *ompA* sequence analysis of *Pasteurella multocida* and their correlation with host species. Vet Microbiol 233:190-195.
246. Luo Y, Glisson JR, Jackwood MW, Hancock RE, Bains M, Cheng IH, Wang C. 1997. Cloning and characterization of the major outer membrane protein gene (*ompH*) of *Pasteurella multocida* X-73. J Bacteriol 179:7856-64.
247. Pearce ME, Langridge GC, Lauer AC, Grant K, Maiden MCJ, Chattaway MA. 2021. An evaluation of the species and subspecies of the genus *Salmonella* with whole genome sequence data: Proposal of type strains and epithets for novel *S. enterica* subspecies VII, VIII, IX, X and XI. Genomics 113:3152-3162.

248. Minias A, Żukowska L, Lach J, Jagielski T, Strapagiel D, Kim SY, Koh WJ, Adam H, Bittner R, Truden S, Žolnir-Dovč M, Dziadek J. 2020. Subspecies-specific sequence detection for differentiation of *Mycobacterium abscessus* complex. Sci Rep 10:16415.
249. Townsend KM, Frost AJ, Lee CW, Papadimitriou JM, Dawkins HJ. 1998. Development of PCR assays for species- and type-specific identification of *Pasteurella multocida* isolates. J Clin Microbiol 36:1096-100.
250. Edgar RC. 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics 34:2371-2375.
251. Kim M, Oh HS, Park SC, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346-351.
252. Jensen LB, Hammerum AM, Aarestrup FM. 2000. Linkage of *vat(E)* and *erm(B)* in streptogamin-resistant *Enterococcus faecium* isolates from Europe. Antimicrob Agents Chemother 44:2231-2.
253. Nonaka L, Suzuki S. 2002. New Mg2+-dependent oxytetracycline resistance determinant *tet_34* in *Vibrio isolates* from marine fish intestinal contents. Antimicrob Agents Chemother 46:1550-2.
254. Xu XJ, Su XZ, Morita Y, Kuroda T, Mizushima T, Tsuchiya T. 2003. Molecular cloning and characterization of the HmrM multidrug efflux pump from *Haemophilus influenzae* Rd. Microbiol Immunol 47:937-43.
255. Lee KJ, Lee NY, Han YS, Kim J, Lee KH, Park SJ. 2010. Functional characterization of the *IlpA* protein of Vibrio vulnificus as an adhesin and its role in bacterial pathogenesis. Infect Immun 78:2408-17.
256. Pei D, Zhu J. 2004. Mechanism of action of S-ribosylhomocysteinase (LuxS). Curr Opin Chem Biol 8:492-7.
257. Ortega J, Lee HS, Maurizi MR, Steven AC. 2004. ClpA and ClpX ATPases bind simultaneously to opposite ends of ClpP peptidase to form active hybrid complexes. J Struct Biol 146:217-26.
258. Zahn R, Buckle AM, Perrett S, Johnson CM, Corrales FJ, Golbik R, Fersht AR. 1996. Chaperone activity and structure of monomeric polypeptide binding domains of GroEL. Proc Natl Acad Sci U S A 93:15024-9.
259. Gaillot O, Pellegrini E, Bregenholt S, Nair S, Berche P. 2000. The ClpP serine protease is essential for the intracellular parasitism and virulence of *Listeria monocytogenes*. Mol Microbiol 35:1286-94.
260. Chong A, Lima CA, Allan DS, Nasrallah GK, Garduño RA. 2009. The purified and recombinant Legionella pneumophila chaperonin alters mitochondrial trafficking and microfilament organization. Infect Immun 77:4724-39.
261. Donnio PY, Lerestif-Gautier AL, Avril JL. 2004. Characterization of *Pasteurella* spp. strains isolated from human infections. J Comp Pathol 130:137-42.
262. DeAngelis PL. 1996. Enzymological characterization of the *Pasteurella multocida* hyaluronic acid synthase. Biochemistry 35:9768-71.
263. Sze JH, Brownlie JC, Love CA. 2016. Biotechnological production of hyaluronic acid: a mini review. 3 Biotech 6:67.
264. Craig L, Pique ME, Tainer JA. 2004. Type IV pilus structure and bacterial pathogenicity. Nat Rev Microbiol 2:363-78.
265. Tomich M, Fine DH, Figurski DH. 2006. The TadV protein of *Actinobacillus actinomycetemcomitans* is a novel aspartic acid prepilin peptidase required for maturation of the Flp1 pilin and TadE and TadF pseudopilins. J Bacteriol 188:6899-914.
266. Confer AW, Ayalew S. 2013. The OmpA family of proteins: roles in bacterial pathogenesis and immunity. Vet Microbiol 163:207-22.
267. Davies RL, MacCorquodale R, Caffrey B. 2003. Diversity of avian *Pasteurella multocida* strains based on capsular PCR typing and variation of the OmpA and OmpH outer membrane proteins. Vet Microbiol 91:169-82.

268. Katoch S, Sharma M, Patil RD, Kumar S, Verma S. 2014. In vitro and in vivo pathogenicity studies of *Pasteurella multocida* strains harbouring different *ompA*. Vet Res Commun 38:183-91.
269. Chevalier G, Duclohier H, Thomas D, Shechter E, Wróblewski H. 1993. Purification and characterization of protein H, the major porin of *Pasteurella multocida*. J Bacteriol 175:266-76.
270. Kim S, Oh DB, Kang HA, Kwon O. 2011. Features and applications of bacterial sialidases. Appl Microbiol Biotechnol 91:1-15.
271. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691-3.
272. Boyce JD, Seemann T, Adler B, Harper M. 2012. Pathogenomics of *Pasteurella multocida*. Curr Top Microbiol Immunol 361:23-38.
273. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 17:238.
274. Gladysheva TB, Oden KL, Rosen BP. 1994. Properties of the arsenate reductase of plasmid R773. Biochemistry 33:7288-7293.
275. Brown NL, Stoyanov JV, Kidd SP, Hobman JL. 2003. The MerR family of transcriptional regulators. FEMS Microbiol Rev 27:145-63.
276. Montanini B, Blaudez D, Jeandroz S, Sanders D, Chalot M. 2007. Phylogenetic and functional analysis of the Cation Diffusion Facilitator (CDF) family: improved signature and prediction of substrate specificity. BMC Genomics 8:107.
277. Anton A, Weltrowski A, Haney CJ, Franke S, Grass G, Rensing C, Nies DH. 2004. Characteristics of zinc transport by two bacterial cation diffusion facilitators from *Ralstonia metallidurans* CH34 and *Escherichia coli*. J Bacteriol 186:7499-507.
278. Hooda Y, Lai CCL, Moraes TF. 2017. Identification of a large family of Slam-dependent surface lipoproteins in Gram-negative bacteria. Front Cell Infect Microbiol 7:207.
279. Yi W, Zhu L, Guo H, Li M, Li J, Wang PG. 2006. Formation of a new O-polysaccharide in *Escherichia coli* O86 via disruption of a glycosyltransferase gene involved in O-unit assembly. Carbohydr Res 341:2254-60.
280. Fleury C, Su YC, Hallström T, Sandblad L, Zipfel PF, Riesbeck K. 2014. Identification of a *Haemophilus influenzae* factor H-Binding lipoprotein involved in serum resistance. J Immunol 192:5913-23.
281. Dashper SG, Hendtlass A, Slakeski N, Jackson C, Cross KJ, Brownfield L, Hamilton R, Barr I, Reynolds EC. 2000. Characterization of a novel outer membrane hemin-binding protein of *Porphyromonas gingivalis*. J Bacteriol 182:6456-62.
282. Jin S, Joe A, Lynett J, Hani EK, Sherman P, Chan VL. 2001. JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. Mol Microbiol 39:1225-36.
283. Autieri SM, Lins JJ, Leatham MP, Laux DC, Conway T, Cohen PS. 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 75:5465-75.
284. Chen YM, Zhu Y, Lin EC. 1987. The organization of the fuc regulon specifying L-fucose dissimilation in *Escherichia coli* K12 as determined by gene cloning. Mol Gen Genet 210:331-7.
285. Baldomà L, Aguilar J. 1987. Involvement of lactaldehyde dehydrogenase in several metabolic pathways of *Escherichia coli* K12. J Biol Chem 262:13991-6.
286. Chen YM, Zhu Y, Lin EC. 1987. NAD-linked aldehyde dehydrogenase for aerobic utilization of L-fucose and L-rhamnose by *Escherichia coli*. J Bacteriol 169:3289-94.
287. Ma B, Simala-Grant JL, Taylor DE. 2006. Fucosylation in prokaryotes and eukaryotes. Glycobiology 16:158r-184r.
288. Tailford LE, Crost EH, Kavanaugh D, Juge N. 2015. Mucin glycan foraging in the human gut microbiome. Front Genet 6:81.

289. Stahl M, Friis LM, Nothaft H, Liu X, Li J, Szymanski CM, Stintzi A. 2011. L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. Proc Natl Acad Sci U S A 108:7194-9.
290. Gon S, Giudici-Orticoni MT, Méjean V, Iobbi-Nivol C. 2001. Electron transfer and binding of the c-type cytochrome TorC to the trimethylamine N-oxide reductase in *Escherichia coli*. J Biol Chem 276:11545-51.
291. Pommier J, Méjean V, Giordano G, Iobbi-Nivol C. 1998. TorD, a cytoplasmic chaperone that interacts with the unfolded trimethylamine N-oxide reductase enzyme (TorA) in *Escherichia coli*. J Biol Chem 273:16615-20.
292. Méjean V, Iobbi-Nivol C, Lepelletier M, Giordano G, Chippaux M, Pascal MC. 1994. TMAO anaerobic respiration in *Escherichia coli*: involvement of the tor operon. Mol Microbiol 11:1169-79.
293. Yu C, Sizhu S, Luo Q, Xu X, Fu L, Zhang A. 2016. Genome sequencing of a virulent avian *Pasteurella multocida* strain GX-Pm reveals the candidate genes involved in the pathogenesis. Res Vet Sci 105:23-7.
294. Zientz E, Janausch IG, Six S, Unden G. 1999. Functioning of DcuC as the C4-dicarboxylate carrier during glucose fermentation by *Escherichia coli*. J Bacteriol 181:3716-20.
295. Graille M, Baltaze JP, Leulliot N, Liger D, Quevillon-Cheruel S, van Tilbeurgh H. 2006. Structure-based functional annotation: yeast ymr099c codes for a D-hexose-6-phosphate mutarotase. J Biol Chem 281:30175-85.
296. Heddleston KL. 1962. Studies on pasteurellosis. V. Two immunogenic types of *Pasteurella multocida* associated with fowl cholera. Avian Dis 6:315-321.
297. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-5.
298. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15:524.
299. Poulsen BE, Yang R, Clatworthy AE, White T, Osmulski SJ, Li L, Penaranda C, Lander ES, Shoresh N, Hung DT. 2019. Defining the core essential genome of *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 116:10072-10080.
300. Luo H, Lin Y, Gao F, Zhang CT, Zhang R. 2014. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic Acids Res 42:D574-80.
301. Barquist L, Langridge GC, Turner DJ, Phan MD, Turner AK, Bateman A, Parkhill J, Wain J, Gardner PP. 2013. A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. Nucleic Acids Res 41:4549-64.
302. Narayanan AM, Ramsey MM, Stacy A, Whiteley M. 2017. Defining Genetic Fitness Determinants and Creating Genomic Resources for an Oral Pathogen. Appl Environ Microbiol 83.
303. Dillon SC, Dorman CJ. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nat Rev Microbiol 8:185-95.
304. Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, Gualerzi CO, Mavathur R, Muskhelishvili G, Pon CL, Rimsky S, Stella S, Babu MM, Travers A. 2007. High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. Nucleic Acids Res 35:6330-7.
305. Dame RT, Wyman C, Goosen N. 2000. H-NS mediated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Res 28:3504-10.
306. Kimura S, Hubbard TP, Davis BM, Waldor MK. 2016. The Nucleoid Binding Protein H-NS Biases Genome-Wide Transposon Insertion Landscapes. mBio 7.
307. Frazier CL, San Filippo J, Lambowitz AM, Mills DA. 2003. Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. Appl Environ Microbiol 69:1121-8.

308. Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM. 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat Biotechnol 19:1162-7.
309. Thomason LC, Sawitzke JA, Li X, Costantino N, Court DL. 2014. Recombineering: genetic engineering in bacteria using homologous recombination. Curr Protoc Mol Biol 106:1.16.1-39.
310. Atkinson GC, Tenson T, Hauryliuk V. 2011. The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. PLoS One 6:e23479.
311. Irving SE, Choudhury NR, Corrigan RM. 2021. The stringent response and physiological roles of (pp)pGpp in bacteria. Nat Rev Microbiol 19:256-271.
312. Dalebroux ZD, Swanson MS. 2012. ppGpp: magic beyond RNA polymerase. Nat Rev Microbiol 10:203-12.
313. Ross W, Vrentas CE, Sanchez-Vazquez P, Gaal T, Gourse RL. 2013. The magic spot: a ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. Mol Cell 50:420-9.
314. Gropp M, Strausz Y, Gross M, Glaser G. 2001. Regulation of *Escherichia coli* RelA requires oligomerization of the C-terminal domain. J Bacteriol 183:570-9.
315. Turnbull KJ, Dzhygyr I, Lindemose S, Hauryliuk V, Roghanian M. 2019. Intramolecular interactions dominate the autoregulation of *Escherichia coli* stringent factor RelA. Front Microbiol 10:1966.
316. Walker KA, Atkins CL, Osuna R. 1999. Functional determinants of the *Escherichia coli fis* promoter: roles of -35, -10, and transcription initiation regions in the response to stringent control and growth phase-dependent regulation. J Bacteriol 181:1269-80.
317. Varik V, Oliveira SRA, Hauryliuk V, Tenson T. 2017. HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. Sci Rep 7:11022.
318. Wang B, Dai P, Ding D, Del Rosario A, Grant RA, Pentelute BL, Laub MT. 2019. Affinity-based capture and identification of protein effectors of the growth regulator ppGpp. Nat Chem Biol 15:141-150.
319. Dorman MJ, Feltwell T, Goulding DA, Parkhill J, Short FL. 2018. The capsule regulatory network of *Klebsiella pneumoniae* defined by density-TraDISort. MBio 9.
320. Pickard D, Kingsley RA, Hale C, Turner K, Sivaraman K, Wetter M, Langridge G, Dougan G. 2013. A genomewide mutagenesis screen identifies multiple genes contributing to Vi capsular expression in *Salmonella enterica* serovar Typhi. J Bacteriol 195:1320-6.
321. Goh KGK, Phan MD, Forde BM, Chong TM, Yin WF, Chan KG, Ulett GC, Sweet MJ, Beatson SA, Schembri MA. 2017. Genome-wide discovery of genes required for capsule production by uropathogenic *Escherichia coli*. mBio 8.
322. Chen Y, Sun E, Yang L, Song J, Wu B. 2018. Therapeutic application of bacteriophage PHB02 and its putative depolymerase against *Pasteurella multocida* capsular type A in mice. Front Microbiol 9:1678.
323. Pan YJ, Lin TL, Chen YY, Lai PH, Tsai YT, Hsu CR, Hsieh PF, Lin YT, Wang JT. 2019. Identification of three podoviruses infecting *Klebsiella* encoding capsule depolymerases that digest specific capsular types. Microb Biotechnol 12:472-486.
324. Jablonski PE, Jaworski M, Hovde CJ. 1996. A minimal medium for growth of *Pasteurella multocida*. FEMS Microbiol Lett 140:165-9.