Whole genome sequence analysis of Avian Pathogenic *E. coli* to assess their potential as reservoirs of antimicrobial resistance and as potential zoonotic pathogens.

Max Laurence Cummins

Bachelor of Biotechnology (Hons)

University of Technology Sydney

Faculty of Science

School of Life Sciences

The i3 Institute

September 2020

Certificate of Authorship

I, Max Laurence Cummins declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science, School of Life Sciences at the University of Technology Sydney. This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Signature:

Date: **29/09/2020**

Acknowledgements

Firstly, I would like to thank my supervisor Steve, for whom I have a tremendous amount of respect and gratitude. Your contagious enthusiasm has been an endless source of motivation over the years and I greatly appreciate your support and mentorship. I will fondly remember the conference trips I shared with you and other members of our lab, particularly our trip to Italy. Thanks to Tiz for helping make that trip so special for all of us, as well as for her comradery and general role as life coach during her and my time studying together. Mohammad, you played a critical role in guiding me through the development of much of this thesis – thanks a lot!

Thank you to our collaborators in Australia and abroad, particularly Marc Marenda and Glenn Browning for their help with the collection which I studied for most of my thesis. Thanks also to Ausgem, UTS and the i3 Institute for funding and hosting my research, and the Australian Government Research Training Program.

Also, a big thank you to Piklu, who helped greatly in showing me the ropes in the lab, and also Ethan, who helped me learn some of the computational aspects of my research. Thanks to Veronica, I've found your relentless work ethic to be something to aspire to. Thanks to Dmitriy, I have been grateful to have your companionship and have enjoyed our nerdy chats about analytical pipelines and videogames!

I would also like to thank my colleague and dear friend Cameron. I have learnt a lot from you, as a person and an academic, and I look forward to us both being old and presumably bald (or at least balding) bioinformatician types when our time comes.

Thanks also to Carla for helping me through some difficult times during my studies.

Lastly, I would like to thank my family. Thank you to my parents for letting me rehearse my presentations to you, listening to my rants at the dinner table, for your financial and emotional support and for always reminding me how proud you are of me. Thanks to my grandparents for helping me through tough times and for their kindness and tremendous generosity over the years. Thanks to my brother Elliot for always taking an interest in my work and for always having been there for me when I need him. I love you all, thank you for helping me get here.

Statement

This thesis is by publication. Chapters 4, 5 and 6 constitute the results chapters and each has been published in a peer reviewed journal. Please note that numbers of figure, table and sections within these chapters have been changed to maintain consistency and flow with the thesis document as a whole. Figures from each results chapter have been included as additional files and readers may prefer to view these rather than those present in this document.

List of publications

Paper 1 - Chapter 4

Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene

Max L. Cummins[1], Cameron J. Reid[1], Piklu Roy Chowdhury[1], Rhys N. Bushell[2], Nicolas Esbert[2], Kelly A. Tivendale[2], Amir H. Noormohammadi[2], Shaiful Islam[2], Marc S. Marenda[2], Glenn F. Browning[2], Philip F. Markham[2] and Steven P. Djordjevic[1*]

[1] The ithree Institute, University of Technology Sydney, Ultimo, NSW, Australia

[2] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, and Werribee, Victoria 3030, Australia.

[*] Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Paper 2 - Chapter 5

Salmonella Genomic Island 1B Variant Found in a Sequence Type 117 Avian Pathogenic *Escherichia coli* Isolate

Max Laurence Cummins[a], Piklu Roy Chowdhury[a], Marc Serge Marenda[c], Glenn Francis Browning[b], Steven Philip Djordjevic[a*]

[a] The ithree institute, University of Technology Sydney, Ultimo, New South Wales, Australia

[b] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia

[c] Asia-Pacific Centre for Animal Health, Department of Veterinary Biosciences, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Werribee, Victoria, Australia

[*] Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Paper 3 - Chapter 6

Salmonella Genomic Island 1 is Broadly Disseminated within Gammaproteobacteriaceae

Max L. Cummins[1,2], Mohammad Hamidian[1] and Steven P. Djordjevic[1,2*]

[1] The ithree Institute, University of Technology Sydney, Ultimo, NSW, Australia

[2] Australian Centre for Genomic Epidemiological Microbiology, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia

[*] Correspondence: Steven P. Djordjevic, steven.djordjevic@uts.edu.au

Table of Contents

Abbreviations

3`-CS	3`-conserved segment
5`-CS	5`-conserved segment
AMR	Antimicrobial resistance
AMS	Antimicrobial susceptible
APEC	Avian pathogenic *E. coli*
cgMLST	Core-genome multi-locus sequence type
CIA	Critically important antimicrobial
ESBL	Extended-spectrum beta-lactamase
ExPEC	Extraintestinal pathogenic *E. coli*
FUTI	Foodborne-urinary tract infection
HC	Hierarchical cluster
HierCC	Hierarchical Clustering of cgMLST
IME	Integrative mobilisable element
IPEC	Intestinal pathogenic *E. coli*
IS	Insertion sequence
KGI	*Klebsiella* genomic island
LB	Lysogeny broth
LGT	Lateral gene transfer
MDR	Multidrug resistant
MGE	Mobile genetic element
MLST	Multi-locus sequence type
NMEC	Neonatal meningitis-associated *E. coli*
ONT	Oxford Nanopore Technologies
ORF	Open reading frame

PCR   Polymerase chain reaction

PGI1/2   *Proteus* genomic island

SGI1-RE   *Salmonella* genomic island 1-related element

SGI1/2   *Salmonella* genomic island 1/2

SNP   Single nucleotide polymorphism

SRA   Sequence read archive

ST   Sequence type

STEC   Shiga-toxigenic *E. coli*

Tn   Transposon

UPEC   Uropathogenic *E. coli*

UTI   Urinary-tract infection

VAG   Virulence-associated gene

VGI   *Vibrio* genomic island

WGS   Whole genome sequencing

XDR   Extensively drug resistant

Abstract

Antimicrobial resistance (AMR) is a complex, global health challenge requiring a One Health framework, acknowledging the interconnectedness of human, animal and environmental microbiospheres, in order to understand its reservoirs, transfer mechanisms and evolution. *Escherichia coli* is a model 'One Health organism' as it is a gastrointestinal commensal of diverse vertebrate species, a common contaminant of foodstuffs and natural environments, increasingly multidrug resistant (MDR) and a major human pathogen.

Within humans, extraintestinal pathogenic *E. coli* (ExPEC) cause severe disease, including urinary tract infections (UTIs) and blood sepsis. Similarly, avian pathogenic *E. coli* (APEC), a subset of ExPEC, cause significant economic losses in the poultry sector. Despite known phylogenetic and genotypic similarities between the two, it is still not possible to define them with genomic data alone and their transit between poultry and humans is poorly characterised. Genomic data on Australian APEC in particular is lacking, hampering efforts to assess and mitigate their threat to human and animal health.

Accordingly, we characterised the genomes of 95 Australian MDR APEC sourced from four Australian states. ST117 and clonal-complex 350 were abundant, however pandemic human ExPEC lineages also featured, including ST95, ST73 and ST131. Almost all strains carried large ColV-like virulence plasmids, however no strains carried genes conferring resistance to critically important antimicrobials (CIAs), such as carbapenems, extended-spectrum β–lactams, fluoroquinolones and colistin, except one exhibiting a fluoroquinolone-resistance associated mutation. This attests to the value of strict antimicrobial use regulations in food animals in Australia.

One strain carried a novel variant of *Salmonella* genomic island 1 (SGI1) named SGI1-B-Ec1. SGI1 is an integrative mobilisable element that confers resistance to five different classes of antimicrobials and may confer pathogenicity traits. It was originally characterised in outbreak strains of *Salmonella enterica* serovar Typhimurium, however SGI1-related elements (SGI1-REs) have been reported in diverse genera, some carrying CIA-resistance genes. The first identification of an SGI1-RE in *E. coli* is alarming as it indicates other more clinically significant SGI1-REs may exist within Australian food-

animals. Subsequently, high-throughput analysis of 455,632 bacterial genomes revealed previously undescribed SGI1-REs in *E. coli, Klebsiella pneumoniae, Vibrio cholerae* and *Cronobacter sazakii.*

This data provides valuable insight into the current AMR status of Australian MDR APEC, as well as the spread and evolution of SGI1 and SGI1-REs within Gammaproteobacteriaceae. Large-scale, One Health-oriented genomic epidemiological studies are urgently required into APEC and bacterial populations more broadly to identify and mitigate their threat to human and animal health.