University of Glasgow

Investigation of colibactin regulation by exogenous amino acids using molecular analyses and colorectal cancer models

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

Sofia Sandalli BSc (Hons)

Submitted September 2024

School of Infection and Immunity

College of Medical, Veterinary and Life Sciences

University of Glasgow

Declaration

I hereby declare that this thesis is the result of my own work and has been composed for the degree of PhD at the University of Glasgow. This work has not been submitted for any other degree at this or any other institution. All work presented was performed by myself unless otherwise stated. All sources of information and contributions to the work have been specifically acknowledged in the text.

Sofia Sandalli

September 2024

Acknowledgements

Firstly, I wish to thank my PhD supervisors, Prof. Andrew Roe and Dr. Gillian Douce, for their continuous guidance and support over the past four years. I am grateful for the dedication and enthusiasm with which you have helped me develop my research and for all that I have learnt from you.

A truly heartfelt thank you goes to the past and present members of the Roe and Douce research groups: in particular, Dr. Jennifer Hallam, Dr. Nicky O'Boyle, Dr. Kabo Wale, Dr. Rebecca McHugh, Dr. Ester Serrano, Dr. David Mark, Dr. Emily Addington, Patricia Rimbi, Emily Horsburgh and Samantha Tucker. Thank you for your teachings, encouragement, kindness and friendship – I can't imagine a better group of colleagues. A special thank you to Dr. Emily Addington for her thoughtfulness and support: it's been a pleasure to work alongside you on our overlapping areas of research and I am excited knowing you'll be tackling the questions we have not yet answered. Many thanks also go to the wider Bacteriology community at the University of Glasgow for making the department such a wonderful environment.

I wish to acknowledge my assessors, Prof. Donald Wall and Dr. Catherine Berry, for the advice they offered throughout my PhD; and my collaborators for the helpful discussions we shared and for their contributions presented in this thesis. In particular, thank you to Prof. Eric Oswald for hosting me in his group and graciously providing bacterial strains, and to Dr. Min Tang and Dr. Nadège Bousset for supervising me while I was in Toulouse. I also wish to thank Dr. Angelika Rupp, Dr. Xiang Li, Rachel Ridgway, Dr. Johan Vande Voorde, Colin Nixon, and Dr. Christopher Bigley for their invaluable help in Glasgow.

Thank you to the Wellcome Trust for funding my studentship, and to Dr. Megan MacLeod who has been a cherished mentor. I would also like to acknowledge the school teachers, university lecturers, and internship supervisors who over the years have conveyed their love for Science and fostered by enthusiasm for research.

Thank you to Dr. Hua Wang, for inspiring me to act with positivity and creativity, both inside and outside the lab; and to all those who have been involved in Bio-Lit Talks, an important part of my time as a PhD student.

Finally, I am grateful to have had the constant encouragement of my parents, Elvira and Vittorio, and my sister Maria Elena. I am thankful to those who make my days lighter and brighter with their unwavering warmth: Ferruccio, Emma J., Magda, Emma D., Sevgi, Elia, Niccolò and Sabrina – thank you.

Abstract

Colorectal cancer (CRC) is the third most common malignant disease and the second deadliest form of cancer worldwide. Risks of CRC development can arise from genetic predispositions, intestinal disorders, environmental factors like lifestyle and diet, and from oncogenic microorganisms, which are increasingly being investigated for their role in cancer aetiology. In 2006, a potent genotoxic metabolite produced by strains of *Escherichia coli* was identified, and subsequently named colibactin. The biosynthetic machinery for production of this genotoxin is encoded on the 54 kb polyketide-synthase (*pks*) pathogenicity island, which is comprised of 19 genes (*clbA-S*). The *pks* island has been identified in *Klebsiella pneumoniae* strains, in a *Citrobacter koseri* isolate, and in *Escherichia coli* strains of the B1 phylogroup, however the vast majority of colibactin-producing bacteria are *E. coli* strains belonging to the B2 phylogroup. Colibactin behaves as both a genotoxin and a cyclomodulin and is capable of inducing DNA damage, such as interstrand cross links (ICLs), double strand breaks (DSBs), single base pair substitutions (SBSs), and insertion-deletion (indel) mutations. These genomic aberrations lead to cell cycle arrest in the G2/M phase, chromosomal instability, senescence, and over time have the potential of triggering CRC. Although several *pks*+ *E. coli* strains are commensals of the human gut, they are heavily overrepresented in patients suffering from CRC, with studies indicating that, on average, 50-60% of CRC patients harbour colibactin producing *E. coli*, as opposed to 20% of healthy individuals. Interestingly, the *pks* island is also carried by the widely administered probiotic strain, *E. coli* Nissle 1917 (EcN), which is sold under the name Mutaflor® and used to manage several inflammatory intestinal disorders. Indeed, the probiotic activity of EcN is tightly linked to its genotoxic one, with studies showing that deletion of *clb* genes results in a lack of antagonistic activity.

As research has demonstrated a strong interplay between CRC, diet, and the gut microbiome, this project aimed to investigate the impact of dietary components on colibactin expression and subsequent downstream effects on CRC. Previously, D-Serine has been shown to downregulate transcription of colibactin, but it was unclear whether this effect is unique to D-Serine or could be driven by different isomers of multiple AAs. The impact of inclusion of L- and D- isomers of all

proteinogenic AAs on colibactin expression was screened in four different types of growth media using a transcriptional GFP reporter assay. RT-qPCR was then used to confirm L-Histidine, L-Isoleucine, L-Serine, D-Serine, and D-Tyrosine as the most significantly inhibitory AAs. The functional impact of colibactin transcriptional repression was further validated through *in cellulo* assays and DNA crosslinking activity. Results showed that L-Serine and D-Serine reduced the megalocytotic, senescent phenotype induced by EcN, and that D-Serine and D-Tyrosine decreased DNA crosslinking caused by EcN. Moreover, the combination of D-Serine and D-Tyrosine greatly repressed the transcription of colibactin biosynthesis genes, indicating additive activity between the two AAs. The activity of D-Serine on genotoxin activity was further validated in several colibactin-producing clinical isolates and two different EcN stocks. *In vivo* work conducted in C57BL/6 mice indicated that EcN elicited probiotic activity, driving down colonic inflammation, while EcN-gavaged mice who were fed D-Serine did not seem to display the probiotic effects of EcN. This was likely because, by acting on colibactin, D-Serine counteracted EcN's probiotic activity. D-Serine was also found to have ameliorative effects on the intestinal mucosa of AOM/DSS mice by reducing inflammation. Stool sampling performed to analyse the impact of DSS, D-Serine, and EcN administration highlighted that DSS treatment was responsible for the biggest changes to microbiota composition. Finally, data from this PhD project also showed that the nitrogen regulatory protein *glnG* may play a role in colibactin transcription.

Overall, this thesis advances what is currently understood about colibactin's endogenous and exogenous regulation and contributes to elucidating how D-Serine may reduce colibactin production. Furthermore, it lays the groundwork for future research that could focus on further probing D-Serine's effect on *pks*+ strains *in vivo*, on understanding the mechanism by which D-Serine reduces colibactin expression, and on investigating how D-Tyrosine downregulates colibactin transcription.

Table of contents

List of figures

List of tables

Abbreviations

AA	Amino acid
ACC	1-aminocyclopropane-1-carboxylic acid
AIEC	Adherent invasive *Escherichia coli*
AM	Aminomalonyl
Amp^R	Ampicillin-resistant
ANOVA	Analysis of variance
AOM	Azoxymethane
BBEC	Bloodborne meningitic *Escherichia coli*
BCAA	Branch chained amino acid
BCFA	Branch chained fatty acids
BFT	*Bacteroides fragilis* toxin
BLAST	Basic local alignment search tool
bp	Base pair
CagA	Cytoxin-associated antigen A
Cdh	Cadherin
cDNA	Complementary DNA
CDT	Cytolethal distending toxin
CEACAM6	Carcinoembryonic antigen-related cell adhesion molecule 6
CFU	Colony forming units
Chl^R	Chloramphenicol-resistant
Cif	Cycle-inhibiting factor
Cldn	Claudin
CNF	Cytotoxic necrotising factor
CNV	Copy number variation
CoPEC	Colibactin-producing *Escherichia coli*
CRC	Colorectal cancer
CRUK	Cancer Research UK
Csr	Carbon storage regulator
DAAO	D-amino acid oxidase
DAEC	Diffuse-adherent *Escherichia coli*
dd	Double-distilled
DME	Drug/Metabolite Exporter
DMEM	Dulbecco's Modified Eagle Medium
DMT	Drug/Metabolite Transporter
DNT	Dermonecrotic toxin
DSB	Double strand break
DSS	Dextran sulfate sodium
EAEC	Enteroaggregative *Escherichia coli*
EcN	*Escherichia coli* Nissle 1917
EDTA	Ethylenediaminetetraacetic acid

EHEC	Enterohaemorrhagic *Escherichia coli*
EIEC	Enteroinvasive *Escherichia coli*
EPEC	Enteropathogenic *Escherichia coli*
ETBF	Enterotoxigenic *Bacteroides fragilis*
EXPEC	Extraintestinal pathogenic *Escherichia coli*
FBS	Foetal bovine serum
FIP	*Fusobacterium nucleatum* inhibitory protein
FMT	Faecal microbiome transplants
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
GI	Gastrointestinal
GTP	Guanosine triphopshate
H&E	Hematoxylin and Eosin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGF	Hepatocyte growth factor
HPI	High pathogenicity island
Htp	High temperature protein
Hyg^R	Hygromycin-resistant
IBD	Inflammatory bowel disease
ICL	Interstrand crosslink
Ifng	Interferon gamma
IL	Interleukin
IP	Intraperitoneal
ISCLA	Interstrand crosslinking assay
Kan^R	Kanamycin-resistant
kb	Kilobase
LB	Lysogeny Broth
LEE	Locus of enterocyte effacement
MALT	Mucosa-associated lymphoid tissue
MAP	Mitogen activated protein
MATE	Multidrug and toxic compound extrusion
Mcc	Microcin
MEM	Minimum Essential Medium Eagle
MNP	Multiple nucleotide polymorphism
MOI	Multiplicity of infection
MS	Mass spectrometry
nf	Nuclease-free
NMDA (colibactin biosynthesis)	N-myristoyl-D-Asparagine
NMDA (neuron receptor)	N-methyl D-Aspartate
NMEC	Neonatal meningitic *Escherichia coli*
NRPS	Nonribosomal peptide synthase
NTBF	Nontoxigenic *Bacteroides fragilis*

NTC	No template control
Ntr	Nitrogen regulation
OD	Optical density
OTU	Operational taxonomic unit
PCA	Principal component analysis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PKS	Polyketide synthase
PMT	*Pasteurella multocida* toxin
PPK	Polyphosphate kinase
PPTase	Phosphopantetheinyl transferase
PRPP	phosphoribosyl-pyrophosphate
RFU	Relative fluorescence unit
rpm	Revolutions per minute
RT	Reverse transcriptase
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
SA	Senescence associated
SAM	S-adenosylmethionine
SCFA	Short chain fatty acid
SDS	Sodium dodecyl sulfate
sil	Streptococcal invasion locus
SMR	Small multidrug resistance
SNP	Single nucleotide polymorphisms
Src	Sarcoma
Str^R	Streptomycin-resistant
T3SS	Type 3 secretion system
TAE	Tris-acetate EDTA
TE	Tris-EDTA
Tet^R	Tetracycline-resistant
Tjp	Tight junction protein
TNF	Tumour necrosis factor
UDG	Uracil-DNA glycosylases
UPEC	Uropathogenic *Escherichia coli*
VacA	Vacuolating toxin A
Vir	Virulent
WGS	Whole genome sequencing
WHO	World Health Organisation
x g	Centrifugal force
Δ	Deletion

1 Introduction

1.1 Oncogenic microorganisms and infectious causes of cancer

The aetiopathogenesis of cancer is known to be multifactorial and complex. While genetic predispositions and physical and chemical carcinogens are frequently attributed as leading causes of cancer, it is only recently that an appreciation for biological carcinogens has emerged. Well-recognised examples of oncogenic microorganisms include oncoviruses such as the Epstein-Barr virus, linked to Burkitt's lymphoma (Pannone *et al.*, 2014) and Hodgkin's lymphoma (Gandhi, Tellam and Khanna, 2004); human papillomaviruses, infections with which can result in cancerous lesions in cervical, genital, and oropharyngeal tracts (Jensen *et al.*, 2024); and hepatitis B and C viruses, both associated to hepatocellular carcinomas (Arbuthnot and Kew, 2000), (Khatun, Ray and Ray, 2021). The most well-known example of an oncogenic bacterium is *Helicobacter pylori*, which is linked to gastric cancer development by inducing chronic gastritis and peptic ulcers. Multiple cohort and case-control studies collated for the World Health Organisation (WHO) over thirty years ago confirm that *H. pylori* can be carcinogenic in humans (World Health Organisation International Agency for Research on Cancer, 1994). Parasites have also been classified as carcinogens, with *Schistosoma haematobium* connected to bladder cancer (Mostafa, Sheweita and O'Connor, 1999), and *Opisthorchis viverrini* linked to cancer of the bile ducts (Sripa *et al.*, 2012). Compared to bacteria, viruses, and parasites, the link between fungi and human cancers is less clear although there is evidence to suggest that some fungal species could be involved in the progression of cancers. For example, *Candida albicans* produces carcinogenic levels of acetaldehyde from nitrosamine (Gainza-Cirauqui *et al.*, 2013) and the mycotoxins produced by *Fusarium* fungi are capable of causing chromosomal aberrations (Knasmüller *et al.*, 1997).

Broadly, the molecular mechanisms through which microbes promote the onset of cancer include: (1) genomic integration and modification, whereby microbial DNA becomes inserted in the host genome and which is, for example, the mode of action of several oncoviruses (Shen *et al.*, 2017); (2) genotoxicity, which consists of the microbial production of a toxin capable of causing DNA damage in the host, disrupting cell-cycle progression, and leading to tumour development; (3) inflammation, whereby persistence of specific microbial virulence factors could

promote chronic inflammation and lead to unregulated cellular proliferation; (4) immunity, whereby the microbiota could negatively interfere with the host's anticancer immunosurveillance; and (5) metabolism, whereby in an environment like the digestive tract, some microorganisms could negatively affect the correct metabolism of nutrients (Scott *et al.*, 2018).

1.2 Colorectal cancer

Colorectal cancer (CRC) is the third most frequent malignant disease worldwide – following breast cancer and lung cancer – with 1.85 million new cases and 0.88 million deaths reported every year (Mattiuzzi, Sanchis-Gomar and Lippi, 2019). While incidences of CRC have been decreasing in high-income countries – largely because of effective screening - according to current predictions, the global burden of CRC is expected to increase by 20-25%, reaching 2.2 million new cases and 1.1 million deaths per year by 2030 (Arnold *et al.*, 2017). Furthermore, the prevalence of CRC in adults younger that 50 years is greatly increasing, with current trends estimating that in 2030 the incidence rates for colon and rectal cancers will increase respectively by 90% and 124.2% (Bailey *et al.*, 2015). This is particularly concerning given that screening programmes are not routinely offered in age groups that were previously considered at low risk for cancer development. Genetics and certain diseases are also risk factors for CRC development. For example, people whose first degree relatives have a history of CRC have a twofold higher risk of developing the same cancer compared to the general population (Butterworth, Higgins and Pharoah, 2006), and patients suffering from obesity (Thrift *et al.*, 2015) or inflammatory bowel diseases (IBD) - in particular ulcerative colitis (Zhou *et al.*, 2019) – are more predisposed to developing CRC. Exogenous factors can also positively influence CRC development. While the most obvious examples are alcohol and tobacco consumption – considered risk factors for a plethora of diseases, including cancer (Ezzati *et al.*, 2002) – nutrition has also emerged as playing a significant role.

1.2.1 The role of nutrition in colorectal cancer

In investigating the role of nutrition in CRC development, several positive and negative correlations have been identified. Generally, the diet most strongly linked to CRC development is the 'Western-style diet', consisting of red and processed meats, added sugars, and refined carbohydrates. On the other hand, a so-called 'prudent' diet – abundant in vegetables, fruits, fish and whole-grains – is negatively correlated with a risk of CRC (Thanikachalam and Khan, 2019). Numerous epidemiological studies have shown the positive correlation between

red and processed meats and CRC, with an average high (>160 g/day) intake of red meat predisposing to a 1.35-fold greater risk, compared to a low (<20 g/day) intake (Norat *et al.*, 2005). Linear dose-response models have shown that a total increase of red and processed meats by 100 g/day is associated to risks of colorectal, colon, and rectal cancers, increasing by respectively 14%, 25%, and 31% (Chan *et al.*, 2011). In contrast, studies looking at fish consumption have either shown non-conclusive results, or shown beneficial associations, with meta-analyses demonstrating that a daily increment of 50 g of fish is associated with a 4% decrease in the risk of developing CRC (Caini *et al.*, 2022).

Fibre is probably the most well-studied, protective dietary component, with numerous studies showing its beneficial effects on gastrointestinal (GI) health (Gill *et al.*, 2021). Its functionality in the GI tract is determined by multiple physiochemical characteristics, such as fermentability and solubility that result in facilitating the absorption of other nutrients, in improving glycaemic and lipemic responses, facilitating gut transit, and enabling the production of short chain fatty acids (SCFAs) which are crucial in GI health as the colonocytes' primary source of energy (Fu *et al.*, 2022). Furthermore, in a gut that is deprived of fibre, the intestinal microbiota resort to using host-secreted mucus glycoproteins as a source of nutrients, causing an erosion of the intestinal mucosa and resulting in increased intestinal permeability, a condition linked to a variety of disorders and to increased pathogen susceptibility (Desai *et al.*, 2016). While some early studies failed to demonstrate a significant protective effect of fibre against CRC, several other studies showed that fibre decreased risks of distal adenomas (Peters *et al.*, 2003), reduced the risk of CRC by 40% in populations with low fibre intakes (Bingham *et al.*, 2003), and that even just a 10 g/day intake of dietary fibre was able to cause a 20% reduction in CRC risk (Chan *et al.*, 2011). Currently, there is robust evidence demonstrating that fibre has a preventative effect against CRC onset (Celiberto *et al.*, 2023). Beyond CRC, epidemiological cohort studies have shown that fibre intake is correlated to a reduced onset of various disorders, including coronary heart disease, stroke, and type 2 diabetes mellitus. Taken together, these well-studied associations have resulted in most nations recommending a daily intake of 25-35 g of fibre in adults (Gill *et al.*, 2021). Finally, there are also studies that have shown dairy products, vitamin D, and calcium inhibit CRC, with particularly strong evidence in support of calcium, whose intake

appears to reduce the risk of both colon and rectal cancer by up to 45% (Huncharek, Muscat and Kupelnick, 2008).

1.2.2 Intestinal bacteria associated to colorectal cancer

Given that the human GI tract is estimated to harbour 10-100 trillion microbes (Turnbaugh *et al.*, 2007) recognised to be extensively connected to an array of physiological conditions of the host, the gut becomes a fascinating environment in which to study carcinogenic microbial-host interactions.

The most well-recognised oncogenic bacterium in the gut is *Fusobacterium nucleatum*, a Gram-negative bacterium that commonly resides in the oral cavity and reaches the GI tract through the bloodstream. It has been found to have higher colonisation rates in colon tumours than in healthy tissues. Although the mechanism through which *F. nucleatum* promotes tumour development is still unknown, the bacterium has been shown to selectively recruit tumour-infiltrating immune cells that promote tumour progression (Kostic *et al.*, 2013). Its tumorigenic activity also appears to be linked to FadA, a virulence factor that exhibits amyloid properties, facilitating the bacterium's survival and allowing it to form biofilms, bind to host cells, and colonise CRC tumours where it enhances tumour progression (Guo *et al.*, 2020). Amyloid-like FadA has been detected in bacteria associated with colorectal carcinomas but not in those colonising healthy tissues, suggesting that this virulence factor plays a key role in *F. nucleatum's* pro-tumorigenic activity (Meng *et al.*, 2021).

Probably the second most well-known group of oncogenic, intestinal bacteria are *Escherichia coli* strains containing a pathogenicity locus known as the polyketide synthase (*pks*) island. These *E. coli* strains are commonly referred to as *pks*+ *E. coli* or colibactin-producing *E. coli* (CoPEC) due to the fact that the *pks* island encodes the biosynthetic machinery for the synthesis of a genotoxin called colibactin. Colibactin was first identified in 2006 in the *E. coli* meningitis strain IHE3034, and subsequently in the intestinal commensal strain Nissle 1917 (also referred to as EcN), in the uropathogenic strain CFT073, and in several clinical isolates (Nougayrède *et al.*, 2006). Numerous studies in the literature have shown

that colibactin acts as a genotoxin by damaging DNA. Compared to controls, *pks+ E. coli* are significantly over-represented in the intestine of patients with CRC and other types of intestinal disorders. For example, compared to the 20% of healthy individuals harbouring *pks+ E. coli,* 68% of patients suffering from familial adenomatous polyposis were found to be colonised by CoPEC (Dejea *et al.*, 2018). Another study showed that while only 19.3% of patients with diverticulosis appear to carry *pks+* strains, colibactin-producing *E. coli* were found in 55.3% of patients with CRC (Buc *et al.*, 2013). As the work described in this thesis focuses on the modulation of colibactin by amino acids, colibactin and *pks+ E. coli* will be discussed in greater detail in later sections of this introduction.

Another bacterium that is considered to be implicated in the onset of CRC is *Bacteroides fragilis,* which can be classified into nontoxigenic *B. fragilis* (NTBF) or enterotoxigenic *B. fragilis* (ETBFT) depending on the capacity to express and secrete the *B. fragilis* toxin (BFT) (Li *et al.*, 2022). ETBF have been found in approximately 90% of CRC patients compared to 50% in healthy individuals (Boleij *et al.*, 2015). The main mechanisms through which the toxin appears to elicit CRC development are by inducing chronic intestinal inflammation leading to increased levels of IL-17, which is implicated in cancer cell survival, and by inducing degradation of E-cadherin, which leads to an alteration of signalling pathways and a promotion of carcinogenesis (Cheng, Kantilal and Davamani, 2020).

Other bacteria that appear to be associated to the development of CRC in the literature are *Salmonella enteridis, Peptostreptococcus anaerobius, Enterococcus faecalis, Streptococcus bovis, and Clostridium septicum* (Li *et al.*, 2022).

1.2.3 The interplay between diet, the gut microbiota and host physiology

The gut microbiota has long been known to critically influence host homeostasis and dietary components are increasingly investigated for their ability to modulate the microbial composition of the gut in an effort to improve human health.

Dietary components act as a link between bacterial metabolism and physiological states of the host. Nutrients ingested by the host facilitate proliferation of

microbes able to utilise these components for growth and, in doing so, the gut microbiota digests and decomposes complex subunits, releasing by-products that are absorbable by the host and that act both locally and systemically to influence host metabolism (Ma, Liu and Wang, 2022). In addition to breaking down material ingested by the host, the gut microbiota can also synthesise *de novo* nutrients such as amino acids, as demonstrated by studies conducted on ruminal bacterial species (Atasoglu *et al.*, 1998), and other bioactive compounds that are necessary for whole-organismal health. Furthermore, microbiota-derived metabolites can in turn encourage growth of other microbial species, resulting in a diet-microbiota-host interplay that is complex, simultaneous, bilateral, and multi-layered.

Another layer of intricacy is added by the fact that dietary components can be equally and concurrently used independently by the host and its microbes. For example, hosts and pathogens share similar nutritional substrates, required respectively for supporting host immune responses, and for maintaining pathogen physiology. The competitive interaction occurring at the host-pathogen interface can be seen in nutritional immunity, a process by which hosts restrict the availability of minerals, in particular iron and zinc, in an effort to limit a pathogen's proliferation. For example, host calprotectin chelates manganese to starve *Acinetobacter baumanii* which, in contrast, has evolved to coordinate transcription of an Mn transporter and of a urea carboxylase to combat the host's Mn sequestration (Juttukonda, Chazin and Skaar, 2016).

Given the increasing evidence of the involvement of the gut microbiota in an array of diseases such as colorectal carcinogenesis, several forms of 'microbiome-targeted therapy' are under development, including the use of probiotics, faecal microbiome transplants (FMTs), and dietary interventions. The latter could prove particularly effective as a cheap, reversible, straightforward, and personalised form of treatment.

1.3 Classification of *Escherichia coli*

The vast majority of *E. coli* strains are commensal, colonising the infant gut as early as within 48 hours from birth, and playing a critical role in the maintenance of intestinal physiology (Mackie, Sghir and Gaskins, 1999). However, some strains have acquired virulence factors that enable them to induce an array of diseases in both humans and animals. Pathogenic *E. coli* are commonly classified into pathotypes based on the type of disease they cause or the sites they colonise: enteropathogenic *E. coli* (EPEC), associated with diarrhoea; enterohaemorrhagic *E. coli* (EHEC), associated with haemorrhagic colitis or haemolytic-uremic syndrome, and including strains that produce the Shiga toxin; enteroaggregative *E. coli* (EAEC), associated with persistent diarrhoea; diffuse-adherent *E. coli*, (DAEC) associated with acute diarrhoea; enteroinvasive *E. coli* (EIEC) and *Shigella*, associated with invasive infections and dysentery; and adherent-invasive *E.coli* (AIEC), associated with chronic intestinal inflammatory conditions. Strains that colonise tissues outside of the digestive tract are classified as extraintestinal *E. coli* (ExPEC) and categorised as: uropathogenic *E. coli* (UPEC) which infect the urinary tract; neonatal meningitic *E. coli* (NMEC) causing meningitis in new-borns, and bloodborne strains of *E. coli* (BBEC), causing sepsis (Yu, Banting and Neumann, 2021).

Another common way to classify *E. coli* is by their serotype, which is based on the major surface antigens that the bacterium expresses. More specifically, the letter O denominates antigens on the cell wall lipopolysaccharide, H designates the bacterial flagella, and K refers to capsular polysaccharides (Orskov *et al.*, 1977). At present among *E. coli* there are 187 recognised O antigens (Liu *et al.*, 2020), 53 H antigens (Wang *et al.*, 2003), and 72 K antigens (Reddy Kunduru, Nair and Rathinavelan, 2015).

In addition to their serotype, *E. coli* can be classified according to their phylogeny, which is reconstructed through whole genome sequencing. Historically, they have been classified into four phylogroups: A, B1, B2, and D, with most virulent extra-intestinal strains categorised in groups B2 and D, and most commensal strains belonging to group A (Herzer *et al.*, 1990). These four phylogroups are also distinguishable via PCR using genes *chuA*, *yjaA*, and an anonymous DNA fragment (Clermont, Bonacorsi and Bingen, 2000). More recently, three more phylogroups

have been proposed: E, containing a small group of unassigned strains of which O157:H7 is the most well-known member; F, consisting of a sister group to group B2; and C, a group of strains similar to B1. A lineage phenotypically identical to *E. coli* but genetically distinct, was also identified, and classified as a separate clade (Clermont *et al.*, 2013), (Tenaillon *et al.*, 2010), (Luo *et al.*, 2011). Although these seven groups remain largely accepted as the framework for the phylogenetic classification of *E. coli*, a recent, comprehensive Mash-based analysis has revealed 14 distinct phylogroups (Abram *et al.*, 2021). Indeed, due to the mobility of certain genetic elements and the continuous evolution and adaptation of virulence factors, it is likely that the current phylogroup divisions may be revisited.

1.4 Bacterial genotoxins and cyclomodulins

Bacterial genotoxins are defined as polyketide-peptide metabolites or multi-protein virulent factors capable of altering DNA. The term 'cyclomodulin' was instead first proposed in 2005 to describe bacterial toxins and effectors that interfere with the eukaryotic cell cycle (Nougayrède *et al.*, 2005). With most bacterial genotoxins also being cyclomodulins, this section describes some well-characterised cellular-proliferating and cellular-inhibiting cyclomodulins (Figure 1-1) in order to contextualise colibactin.

1.4.1 Cytotoxic necrotising factor and the dermonecrotic toxin

Cytotoxic necrotising factors (CNFs), dermonecrotic toxin (DNT) and the *Pasteurella multocida* toxin (PMT) are known to stimulate proliferation of eukaryotic cells (Nougayrède *et al.*, 2005). CNF was first identified in an *E. coli* strain isolated from a case of infant enteritis (Caprioli *et al.*, 1983). Since then CNF has been identified in the genomes of multiple EPEC strains and in *Yersinia pseudotuberculosis*. CNFs share a highly similar intracellular trafficking mechanism and cytopathic mode of action: they are secreted directly by the bacterial cells or bound to outer membrane vesicles and, once internalised into the mammalian cells by endocytosis, are transported by endosomes, before releasing their catalytic fragment which deamidates Rho GTPases, triggering downstream cell responses. More specifically, the continuous activation of Rho-GTPases by CNFs induces actin cytoskeleton rearrangements, impairing cytokinesis without disrupting replication, and thus resulting in uncontrolled DNA synthesis and multinucleated cells (Chaoprasid and Dersch, 2021). To date, three different types of CNFs have been described: CNF1, the first identified CNF encoded for by the chromosomal gene *cnf1*; CNF2, which is 86% identical to CNF1 and encoded by a transmissible virulent (Vir) plasmid, causing similar cytopathic effects to CNF1 (De Rycke *et al.*, 1990); and CNF3, which shares about a 70% sequence identity with CNF1 and CNF2, but induces a 5-10 fold stronger activation of RhoA (Stoll *et al.*, 2009). The CNF variant of *Y. pseudotuberculosis* is known as CNF_Y.

DNT produced by *Bordetella* species shares the same mechanism of action as CNF, involving the activation of Rho family proteins by deamidation of glutamine to glutamic acid at specific sites. They are consequently both described as Rho GTPase-activating toxins. Both CNFs and DNT share homology, are lethal in experimental animal models, and induce a characteristic dermonecrosis upon injection that confers their name (Chaoprasid and Dersch, 2021).

1.4.2 The *Pasteurella multocida* toxin and other cellular-proliferating cyclomodulins

Among the bacterial toxins known to induce cellular proliferation, *Pasteurella multocida* toxin (PMT) and the cytoxin-associated antigen A (CagA) produced by *H. pylori* have also been identified. PMT is a highly potent mitogen that has been shown to activate both G proteins and MAP kinases, indirectly inducing Rho-GTPase activation, leading to actin stress fibre formation, and sustained DNA synthesis (Morgan *et al.*, 2022). PMT also shows lethal, dermonecrotic activities in animal models and leads to similar outcomes as DNT, although it shares a lower degree of homology with CNF and DNT (Horiguchi, 2001). CagA instead is injected into gastric epithelial cells by a type 4 secretion system (T4SS) encoded for on the *cag*-pathogenicity island. Once CagA gets phosphorylated by Src kinases, it activates SHP-2, an oncoprotein that, in turn, activates a MAPK cascade, leading to uncontrolled cellular proliferation (Hatakeyama, 2004).

1.4.3 Cytolethal distending toxin

Cytolethal distending toxins (CDTs) are a class of tripartite toxins produced by several Gram-negative, pathogenic bacteria including *Shigella dysenteriae*, *Actinobacillus actinomycetemcomitans,* some *E. coli* strains, and some *Salmonella*, *Campylobacter* and *Haemophilus* species (Nougayrède *et al.*, 2005). CDT inhibits cell proliferation, with affected cells dying by apoptosis, and was the first bacterial toxin shown to block the eukaryotic cell cycle in the G2/M phase (Pérès *et al.*, 1997). It is composed of three subunits, CdtA, CdtB, and CdtC, where, in most cases, the B subunit is enzymatically active, and the other two

mediate binding to host cells, facilitating internalisation of CdtB. The internalisation pathway is dependent on intact lipid rafts in cellular membranes and toxin uptake occurs via dynamin-dependent endocytosis. Endosomes then transport CdtB to the Golgi apparatus, the endoplasmic reticulum, and the nucleus through a retrograde transport pathway (Guerra *et al.*, 2011). Although some mechanisms regulating CdtB-induced DNA damage are still unclear, it is known that the damage-provoking pathway involves the phosphorylation of CDC25C, preventing it from activating the CyclinB-Cdk1 complex, and leading to a cell-cycle arrest in G2 (Escalas *et al.*, 2000).

1.4.4 Other cellular-inhibiting cyclomodulins

The cycle-inhibiting factor (Cif) produced by enteropathogenic *E. coli* (EPEC) and enterohaemorrhagic *E. coli* (EHEC) is another toxin known to inhibit the eukaryotic cell cycle. It is not encoded by the locus of enterocyte effacement (LEE) , but by a lamboid prophage present in the strains (Oswald *et al.*, 2005). Cif induces and accumulation of p21 and p27, inhibiting CyclinB-Cdk1 and Cyclin A/E-Cdk2 complexes, preventing cell cycle progression, and eventually leading to an accumulation of cells arrested in G1 and G2 phases (Dé Ric Taieb, Nougayrè De and Oswald, 2011). Other well-characterised, inhibitory cyclomodulins include: the vacuolating toxin A (VacA) produced by *H. pylori* which increases p53 expression, interferes with T cell activation, and induces mitochondrial apoptosis, leading to cell cycle arrest in the G1 phase; the *F. nucleatum* inhibitory protein (FIP) which – by altering the expression of the proliferating cell nuclear antigen (PCNA) – induces cell cycle arrest in mid-G1 phase; and mycolactone, produced by *Mycobacterium ulcerans*, which is a polyketide-derived macrolide that induces cell cycle arrest in G0-G1 phases (Oswald *et al.*, 2005), (Nougayrède *et al.*, 2005), (Morgan *et al.*, 2022). Another well-characterised, inhibitory cyclomodulin is colibactin, around which this PhD project was centred.

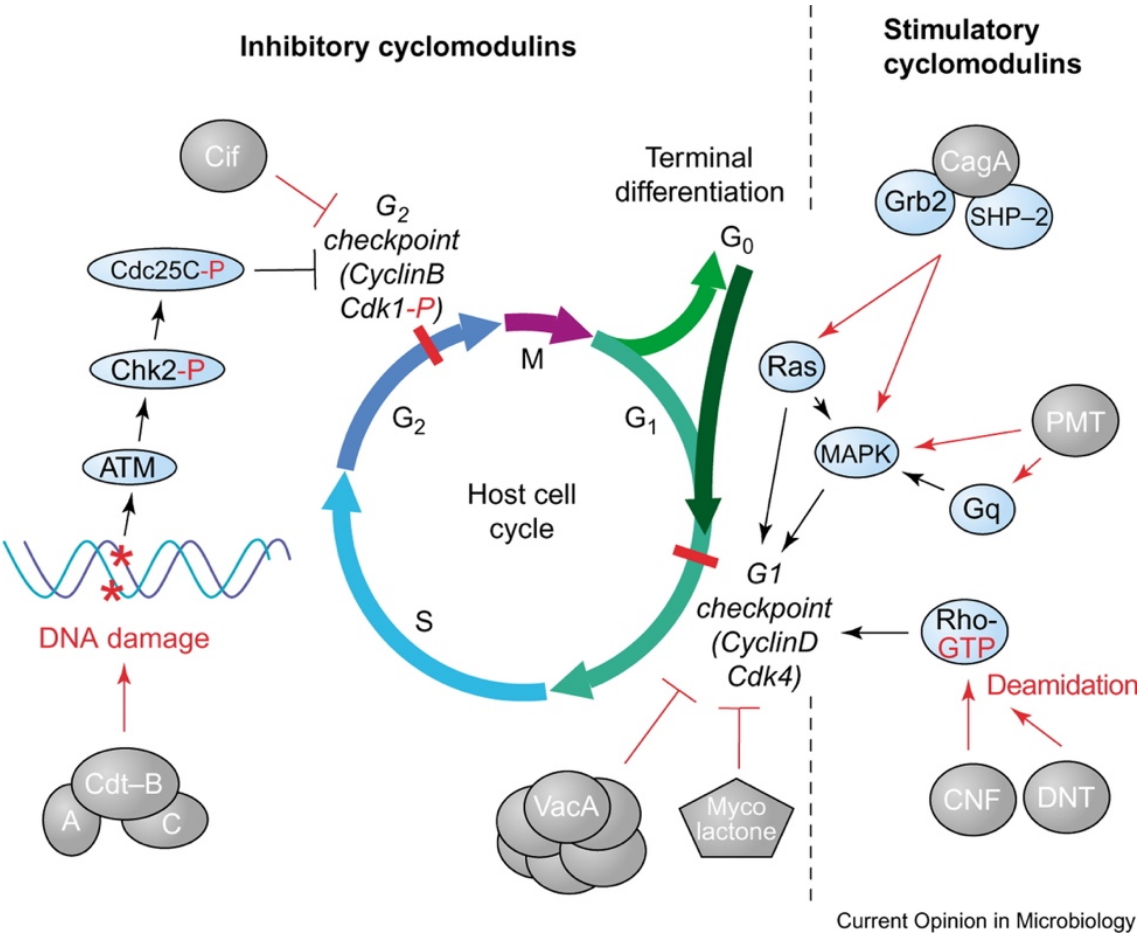


Figure 1-1 Bacterial cyclomodulins influencing the host cell cycle. Several bacterial toxins are capable of exerting cell cycle modulating effects, rendering the bacterial species that produce them potentially carcinogenic to their eukaryotic hosts. Cyclomodulins can be broadly classified into those that stimulate cellular proliferation and those that inhibit it. The schematic represents inhibitory (Cif, CDT, VacA, and mycolactone) and stimulatory (CagA, CNF, DNT, and PMT) cyclomodulins and the pathways through which they act on phases of the eukaryotic cell cycle (Oswald *et al.*, 2005).

1.5 Colibactin

Colibactin is a powerful genotoxin and cyclomodulin present in different members of the bacterial family *Enterobacteriaceae*. It was first discovered in 2006 and since then its mode of action and implications in the development of CRC have been extensively studied. The sections below provide further information on colibactin biosynthesis, distribution, mechanisms of action, and regulation.

1.5.1 Structure and genotoxic activity

In 2019 two groups separately determined the structure of colibactin (Jiang *et al.*, 2019), (Xue *et al.*, 2019). It had already been hypothesised that colibactin's genotoxicity was likely to arise from DNA alkylation (Wilson *et al.*, 2019) and that DNA strands are cross-linked covalently when exposed to colibactin both *in vitro* and *in cellulo* (Bossuet-Greif *et al.*, 2018). Elucidating the structure of colibactin provided greater insight into the underlying mechanism of the genotoxin. Colibactin is made up of two, nearly symmetrical subunits, each containing an electrophilic cyclopropane 'warhead' that binds to DNA residues. The cyclopropane rings appear to have high affinity for adenine-rich sequences; in particular, one of the two cyclopropane rings is currently well-known to react at the N3 of adenine, whereas the specific alkylation site of the second warhead is still unknown (Wilson *et al.*, 2019). The alkylation of residues on opposite DNA strands requires opening of the cyclopropane rings, which are adducted with adenine residues, triggering the formation of interstrand crosslinks and (ICLs), and activating a cascade of host responses (Dougherty and Jobin, 2021). Since its discovery, colibactin has also been reported to induce double-strand breaks (DSBs), which are likely derived from the repair pathway of the ICLs, activating the DNA damage checkpoint pathway, leading to cell cycle arrest and eventually causing cell death (Nougayrède *et al.*, 2006). There is also a report that one form of colibactin can directly exert DSBs through oxidate cleavage via copper mediation (Li *et al.*, 2019). The way in which colibactin molecules translocate to the host nucleus is still unknown. A diagram of colibactin's biosynthetic cluster molecular structure can be found below (Figure 1-2).

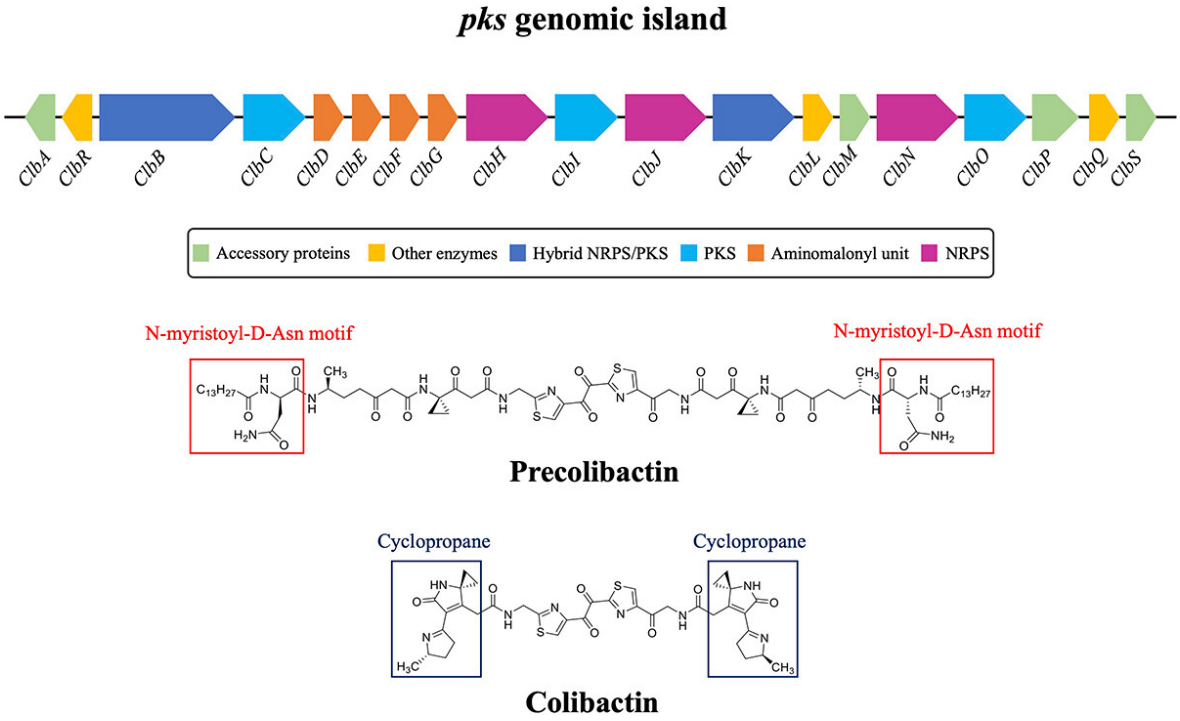


Figure 1-2 Structure of the *pks* genomic island and of colibactin. The illustration represents the 54 kb *pks* island, constituted by 19 genes (*clbA*-*clbS*) responsible for colibactin biosynthesis, with the function of each gene is indicated by colour. The figure also represents the proposed structure of precolibactin, depicting the prodrug motif that is cleaved off (red square) to from mature colibactin, and highlighting the cyclopropane rings (black squares) that alkylate DNA (Falzone *et al.*, 2024).

Several studies have investigated host cells responses to colibactin exposure. Nougayrède and colleagues were the first to show that the DSBs induced by colibactin lead to cell-cycle arrest in the G2/M phase, inducing megalocytosis and subsequent cell death (Nougayrède *et al.*, 2006). Others have shown that short-term exposure to colibactin induces the formation of phosphorylated histones (γH2AX) and of anaphase bridges, leading to chromosomal abnormalities in eukaryotic cells (Cuevas-Ramos *et al.*, 2010). Research has also shown that following infection with *pks*+ *E. coli*, cells exhibit cellular senescence, characterised by an increase in senescence associated (SA)-β-galactosidase activity and by distinctive SA heterochromatin foci (SAHF). Affected cells were also able to induce γH2AX foci formation in neighbouring, uninfected cells (Secher *et al.*, 2013). This mechanism was later shown to depend on microRNA-20a-5p, which targets SENP1, a key protein in the regulation of p53 SUMOylation. The accumulation of SUMO-conjugated p53 was linked to the secretion of hepatocyte growth factors (HGFs), which trigger cell proliferation in neighbouring cells *in vitro* and promote colonic tumours *in vivo* (Cougnoux *et al.*, 2014).

Errors derived from host DNA repair following colibactin-induced damage result in a distinct, mutational signature. One study performed on human organoids found that chronic exposure to *pks*+ *E. coli* results in a mutational signature identifiable by numerous single base substitutions (SBSs) and short insertion-deletions (indels), particularly in AT-rich regions of DNA (Pleguezuelos-Manzano *et al.*, 2020). While another study performing short-term exposure of Caco-2 cells to *pks*+ strains found enrichment of SBS in AT-abundant regions of the DNA minor groove (Dziubańska-Kusibab *et al.*, 2020). The presence of a specific DNA-damage signature further supports the designation of *pks*+ *E. coli* as oncogenic microorganisms. Experiments carried out in colonic human organoids have also shown the transformative activity of *pks*+ *E. coli*: infected organoids displayed major chromosomal aberrations, with over 650 million base pairs affected by copy number variations (CNVs), although these lacked the colibactin mutational signature (Iftekhar *et al.*, 2021). From a clinical perspective, this indicates that chromosomal aberrations precede genomic imprinting and are sufficient to cause malignant transformations, possibly suggesting that the incidence of colibactin-driven transformations in CRC patients may be higher than the presence of mutational signatures would suggest.

1.5.2 Colibactin biosynthesis

The genes responsible for colibactin biosynthesis are encoded within a 54-kilobase genomic cluster known as the polyketide synthase (*pks)* island. It consists of 19 genes (*clbA* to *clbS*) forming a hybrid non-ribosomal peptide synthetase-polyketide synthase (NRPS-PKS) 'assembly line' (Nougayrède *et al.*, 2006). More specifically, the biosynthetic gene cluster is composed of three NRPSs (ClbH, ClbJ, ClbN), three PKSs (ClbC, ClbI, ClbO), and two NRPS/PKS hybrids (ClbB, ClbK), as well as accessory enzymes (ClbA, ClbD, ClbE, ClbF, ClbL, ClbP, ClbQ, ClbS) involved in the synthesis, modification, self-resistance, and transport of colibactin (Faïs *et al.*, 2018). Systematic mutagenesis conducted upon discovery of the *pks* island showed that all genes with the exception of *clbM* are required for functional colibactin production (Nougayrède *et al.*, 2006). One explanation for this observation is that other transporters encoded in other regions of the genome compensate for the deletion of the efflux pump. Similarly, *clbS* is not specifically required for synthesis of the genotoxin, as it encodes a protein preventing auto-toxicity from the mature genotoxin (Tang *et al.*, 2022).

NRPS-PKS enzymes are among the largest found in bacteria. They have been shown to work in 'assembly lines' with the forming metabolite progressively transferred between enzymes. Examples of other macromolecules produced by NRPS-PKS enzyme complexes include other anti-infectious agents like penicillin, erythromycin, avermectin, and vancomycin (Faïs *et al.*, 2018).

Since its discovery, colibactin has proven to be a challenging molecule to isolate. The fact that genotoxicity is not observed when human cells are treated with cell lysates or supernatants and that cell-to-cell contact is required for genotoxic activity (Nougayrède *et al.*, 2006), indicates that colibactin is highly unstable. Hence, colibactin biosynthesis has been studied over the past two decades by indirect approaches, with the activity of the biosynthetic enzymes being predicted and confirmed, instead of performing molecule isolation and structure elucidation (Balskus, 2015). Colibactin synthesis begins with ClbA, a phosphopantetheinyl transferase (PPTase), that activates the NRPS and PKS proteins through the addition of phosphopantetheinyl arms which will then carry the metabolite throughout its elongation. Homology studies with pre-xenocoumacins – secondary metabolites produced by the Gram-negative bacterium *Xenorhabdus nematophila*

- revealed that the prodrug motif, which is removed in the final stages of biosynthesis to produce the active toxin, consists of an N-acyl chain appended to a D-Asparagine residue (Brotherton and Balskus, 2013), (Bian *et al.*, 2013). Biochemical characterisation studies showed that, following ClbA activity, ClbN utilises L-Asparagine and the fatty acid substrate myristoyl-CoA for acetylation of the N-terminus, synthesising an N-myristoyl-D-asparagine (NMDA), a prodrug scaffold that is cleaved off during conversion of precolibactin (the immature form of the genotoxin) to colibactin (the active form of the genotoxin). The intermediate generated by ClbN is further elongated by the NRPS module of ClbB, containing a C domain linking an L-AA to a growing peptide chain. ClbB was shown to exercise elongation through the addition of L-Alanine (preferred) or L-Valine, (Brotherton and Balskus, 2013), (Bian *et al.*, 2013).

Whereas the mechanism of other biosynthetic enzymes did not become clear until later, ClbP had been shown to exhibit homology to XcnG and ZmaM, membrane-anchored, periplasmic peptidases respectively involved in prodrug removal during the final production stages of the antibiotics xenocoumacin and zwittermicin. Experiments involving synthetic substrates with the colibactin prodrug motif revealed that ClbP only hydrolysed substrates containing the full length sequence (Brotherton and Balskus, 2013). Further structural and functional studies on ClbP confirmed its role in cleaving precolibactin from the drug scaffold (Dubois *et al.*, 2011), (Cougnoux *et al.*, 2012). Elucidation of the function of ClbP leading to the isolation of precolibactin has been possible through deletion of *clbP* in *pks+* strains of interest and subsequent tandem mass-spectrometry (MS) analyses. Since the first isolation of precolibactin (Vizcaino *et al.*, 2014), over 40 precolibactins have been identified to date differing in the number of fatty acyl substituents and in molecular weight (Tang *et al.*, 2022).

Subsequent studies characterising the PKS components of the colibactin biosynthetic cluster showed that synthesis of precolibactins continues with ClbC, ClbH, ClbI, ClbJ, and ClbK, which use malonyl-CoA and other amino acids as substrates for building upon the precolibactin metabolite scaffold (Zha *et al.*, 2016). While it was proposed that ClbH may use the L-Methionine-derived cyclopropane-containing amino acid 1-aminocyclopropane-1-carboxylic acid (ACC) as a building block (Bian *et al.*, 2015), other studies suggest that *S*-

adenosylmethionine (SAM) – a non-proteinogenic AA – is instead directly activated as an NRPS building block and loaded onto ClbH (Zha *et al.*, 2017), leading to subsequent generation of cyclopropane rings, which are essential for colibactin's mode of action (Vizcaino and Crawford, 2015). Although SAM can be directly converted in ACC, an enzymatic reaction that has for example been witnessed in plants (Thibodeaux, Chang and Liu, 2011), there is yet no evidence to suggest that this reaction can take place in *E. coli*. ClbI has also been found to be essential for cyclopropane formation, potentially promoting cyclisation, or tethering an intermediate during construction (Zha *et al.*, 2016).

The synthesis of colibactin was also found to involve an aminomalonyl (AM) unit (Brachmann *et al.*, 2015), synthesised under the regulation of ClbD, ClbE, ClbF and ClbG. More specifically, it appears that ClbH loads a molecule of serine onto ClbE, which gets converted into AM by dehydrogenases ClbD and ClbF, and then transferred by ClbG to ClbK, which is responsible for incorporating AM into the structure of colibactin (Zha *et al.*, 2016). Interestingly, mutants lacking the AM biosynthetic machinery are not genotoxic, implying these enzymes are critical in the production of the genotoxic metabolite (Brachmann *et al.*, 2015). ClbK appears to also be responsible for the formation of a thiazole ring, with current analyses indicating that the Ox domain in ClbK oxidises the thiazoline introduced by ClbJ into thiazol (Vizcaino *et al.*, 2014).

The catalytic machinery of ClbO, ClbQ, and ClbL still remains unclear. ClbO is the final enzymatic component of the colibactin assembly chain. It is thought to participate in the extension of precolibactin through the incorporation of a yet uncharacterised building block (Brachmann *et al.*, 2015). Work suggests that ClbQ is responsible for offloading the precolibactin metabolites from the NRPS-PKS assembly line by acting on the AM unit (Sandhya Guntaka *et al.*, 2017). ClbL instead appears to be involved in the final step for the formation of intact precolibactin by linking to intermediates from the NRPS-PKS scaffold together. More specifically, it appears to link a ClbI-derived thioester and a ClbO-bound α-aminoketone, and in so doing forms a pseudodimeric precolibactin (Jiang *et al.*, 2019), (Tripathi *et al.*, 2023). Precolibactin is brought into the periplasm by ClbM, a multidrug and toxic compound extrusion (MATE) transporter (Mousa *et al.*, 2016), where ClbP, a membrane-bound peptidase, removes the NMDA prodrug

motif, inducing cyclisation of linear precolibactins and catalysing the formation of mature, active colibactin (Brotherton and Balskus, 2013).

ClbS has been identified as the self-resistance enzyme which, by hydrolysing the cyclopropane rings, prevents the bacterial cell from experiencing self-induced genotoxicity (Tripathi *et al.*, 2017). Indeed, the entire prodrug production strategy is conserved in the biosynthetic pathway of multiple toxins and can be considered an evolutionary strategy to prevent bacterial auto-intoxication. Finally, ClbR has been established as the LuxR-type transcriptional activator of the colibactin cluster, binding both to a site upstream of *clbB* and to a region upstream of *clbR* itself. This suggests that, in addition to being a transcriptional regulator, ClbR is also an auto-activator (Wallenstein *et al.*, 2020).

Several internal factors expressed in the host bacterial cell are known to influence colibactin synthesis (Addington, Sandalli and Roe, 2024). These include the *E. coli* high-temperature protein G (HtpG/Hsp90EC), which acts as a molecular chaperone and is crucial for the synthesis of the genotoxin, with mutations/deletions of HtpG resulting in the absence of colibactin production (Garcie *et al.*, 2016). Current findings suggest that the direct interaction between HtpG and DnaK prevents misfolding of colibactin biosynthetic proteins and may protect them from degradation induced by proteases (Corteggiani *et al.*, 2022). It has also been shown that inactivation of polyphosphate kinase (PPK) in *pks+ E. coli* results in downregulation of *clbB* transcription and decreased colibactin synthesis, suggesting that PPK is required for colibactin production. Beyond *in vitro* mutagenesis, this was also demonstrated through bacterial treatment with mesalamine, an inhibitor of PPK enzymatic activity, which decreased genotoxicity of colibactin. The phenotype was reversed when PPK was supplemented. (Tang-Fichaux *et al.*, 2020).

The BarA-UrvY two-component system, which modulates adaptive responses in *E. coli*, has also been shown to be involved in colibactin expression through mediation of the carbon storage regulator (Csr) system. In the absence of an environmental stimulus, the BarA-UrvY system is inactive, and the Csr carbon storage regulator, CsrA, binds to the mRNA of *clb* genes, preventing colibactin expression; in the presence of an appropriate external cue detected by the BarA sensor kinase, UvrY is activated and in turn triggers other members of the Csr

system, which block CsrA from binding to the *pks* locus, thus enabling colibactin biosynthesis (Rehm *et al.*, 2022), (Wong *et al.*, 2022).

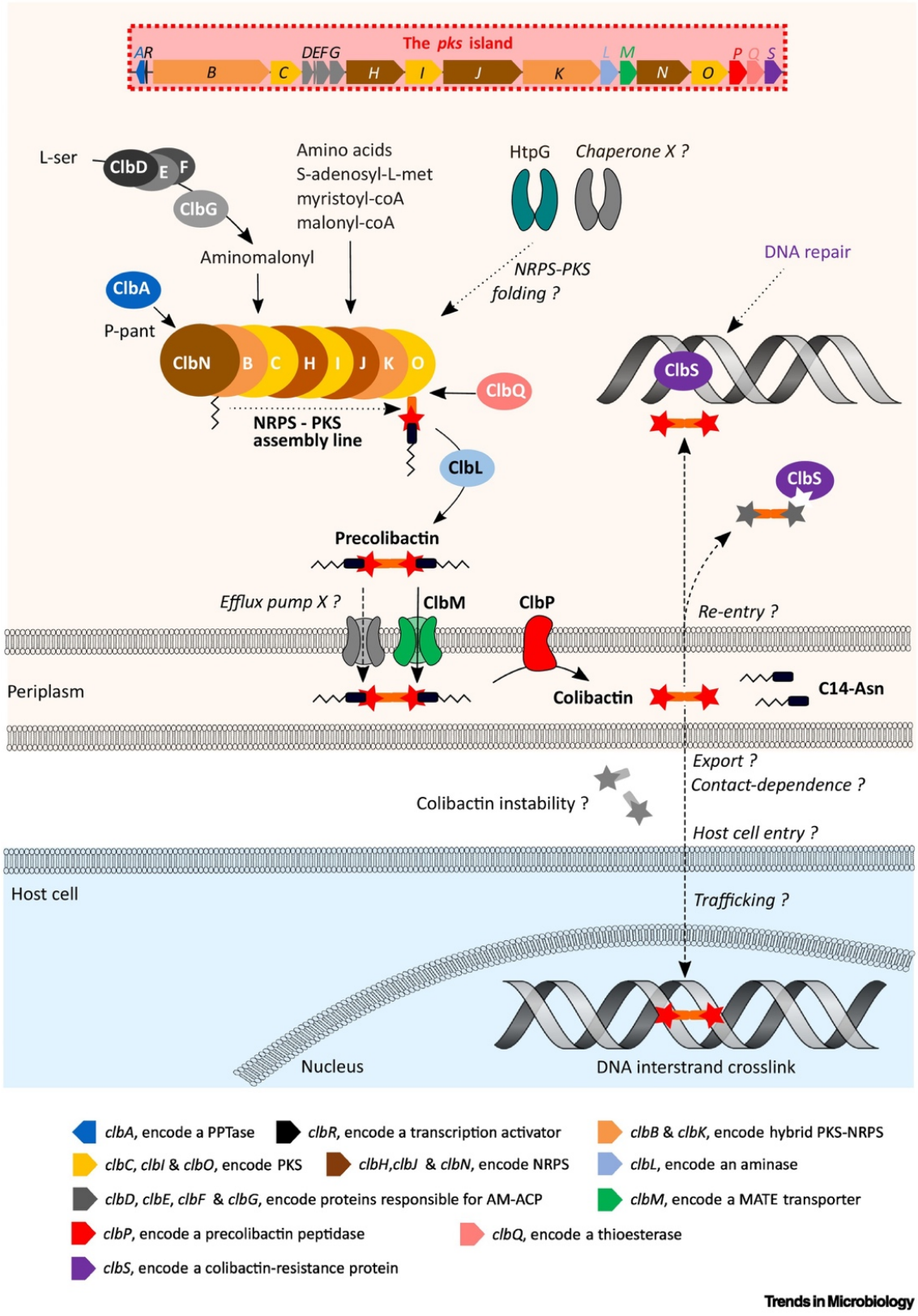


Figure 1-3 Model of colibactin biosynthesis and transport. Schematic representing the 54 kb *pks* island *clbA-clbS*, the function of the of the 19 NRPS, PKS, NRPS-PKS enzymes responsible for colibactin biosynthesis; and the involvement of regulators, such as the HtpG chaperone protein, in regulating colibactin synthesis (Chagneau *et al.*, 2022)

1.5.3 Phylogenetic distribution of the *pks* island

Several studies looking at the phylogenetic distribution of the *pks* island within *E. coli* and among other *Enterobacteriaceae* have revealed clinical correlations between *pks* carriage in pathogenic *E. coli* and other *E. coli* virulence factors. A study in 2008 found that the colibactin biosynthetic cluster was present in 58% of *E. coli* bloodstream isolates, exclusively belonging to the B2 phylogroup, and was in particular associated to a subset exhibiting elevated virulence scores (Johnson *et al.*, 2008). Carriage of *clb* genes was also significantly associated with the presence of virulence genes (Johnson *et al.*, 2008), such as *pic*, involved in intestinal mucus layer penetration and colonisation (Henderson *et al.*, 1999). Another study screened 1565 isolates from different *Enterobacteriaceae* genera and found that the *pks* island was not exclusive to B2 phylogroup *E. coli* but was also found in an *E. coli* isolate from the B1 phylogroup. Moreover, out of the 9.5% *E. coli* strains that tested positive for the *pks* island, 73.1% were clinical ExPEC strains and 26.9% were commensal faecal isolates from healthy volunteers. *Clb* genes were also detected in 3.5% of the tested *Klebsiella pneumoniae* strains, in 27.3% of the analyse *Enterobacter aerogenes* strains, and in a single *Citrobacter koseri* isolate (Putze *et al.*, 2009). A study looking at 785 *E. coli* human and bovine isolates identified the *pks* island in 109 strains, with the majority belonging to the B2 phylogroup, although some *pks*+ isolates also belonged to phylogroups A, B1, and D. This highlights the *pks* island's capacity to disseminate through horizontal gene transfer (Auvray *et al.*, 2021).

1.5.4 Clinical relevance of colibactin

There is extensive evidence to suggest that colibactin contributes to cancer development, particularly CRC. Numerous CRC animal models have shown that *pks*+ *E. coli* enhance tumorigenesis (Cougnoux *et al.*, 2014), (Veziant *et al.*, 2016), (Bonnet *et al.*, 2014), (Lopès *et al.*, 2020), and that deletion of the *pks* island decreases tumour multiplicity and invasion (Arthur *et al.*, 2012). Cohort studies performed on CRC patients have shown that colonisation of *pks*+ *E. coli* is over three times higher in CRC patients, compared to healthy controls (Buc *et al.*, 2013), (Arthur *et al.*, 2012). Interestingly, it was also discovered that the

association of a Western-style diet with CRC is stronger in tumours that are heavily colonised by colibactin-producing *E. coli*, suggesting a link between diet, the intestinal microbiota, and CRC aetiology (Arima *et al.*, 2022). Moreover, the fact that CoPEC leave a mutational fingerprint identified in CRC biopsies, supports that *pks+ E.coli* may have a causative role in CRC onset (Pleguezuelos-Manzano *et al.*, 2020), (Dziubańska-Kusibab *et al.*, 2020). Furthermore, *pks+ E. coli* are able to modify the tumour microenvironment to their advantage: they are able to induce proliferation in neighbouring cells (Cougnoux *et al.*, 2014), decrease host expression of CD3+ and CD45+ T-cells (Lopès *et al.*, 2020), and confer resistance to chemotherapeutic drugs (Sogari *et al.*, 2024).

1.5.5 Other metabolites produced by *pks+ E. coli*

Beyond colibactin, the *pks* island is responsible for the synthesis of an array of other metabolites. For example, it has been discovered that the PPTase ClbA, in addition to being essential for the synthesis of colibactin, also contributes to the synthesis of enterobactin, salmochelin, and yersiniabactin (Martin *et al.*, 2013), which are all siderophores, necessary for scavenging iron and copper (Garénaux, Caza and Dozois, 2011). The fact that the colibactin pathogenicity island (*pks*) and the yersiniabactin high pathogenicity island (HPI) are highly conserved, usually coexisting within the genome, suggests that these genomic clusters are intertwined on a biosynthetic level and are evolutionary co-selected for high bacterial virulence (Wami *et al.*, 2021). Another example is that of the probiotic, *pks+ E. coli* strain, Nissle 1917 (EcN), which produces two microcins (Mcc), MccH47 and MccM, that exhibit strong bactericidal activity against phylogenetically related bacteria (Chagneau *et al.*, 2022). Interestingly, it was demonstrated that ClbP is involved both in the synthesis of colibactin and in that of Mccs, reflecting once again the co-evolution of virulence and fitness determinants for strain adaptation (Massip *et al.*, 2019), (Massip *et al.*, 2020). In EcN the *pks* island has also been found to synthesise lipopeptides with analgesic activity, such as C12AsnGABAOH, which crosses the epithelial barrier and acts on sensory neurons (Pérez-Berezo *et al.*, 2017).

1.6 *E. coli* Nissle 1917

The vast majority of experiments in this PhD project have been performed using the *E. coli* B2 phylogroup strain Nissle 1917, commonly referred to as EcN (O6:K5:H1). This strain was isolated for the first time in 1917 by the physician Alfred Nissle from a solider who, unlike his companions, displayed resistance to *Shigella*-induced diarrhoea. After curing patients with laboratory-grown cultures of the strain, the strain became patented for commercialisation. Since then, it has been sold under the trademark of Mutaflor® and is a well-established and successful probiotic. Given its strong antagonistic activity against other bacteria, EcN has been used extensively to treat an array of intestinal disorders (Schultz and Burton, 2017), including ulcerative colitis (Kruis *et al.*, 2004) and pouchitis (Kuzela, Kascak and Vavrecka, 2001). Numerous *in vivo* experiments have shown that EcN reduces DSS-induced colitis (Rodríguez-Nogales *et al.*, 2018), (Schultz *et al.*, 2004) and that it decreases the secretion of proinflammatory cytokines, while stimulating the production of anti-inflammatory ones (Sturm *et al.*, 2005). In addition to the production of Mccs, of siderophores (Massip *et al.*, 2019), and of an analgesic lipopeptide (Pérez-Berezo *et al.*, 2017) – previously described – other factors that confer EcN its antagonistic activity include: the motility due to its H1 flagellum allowing it to access the mucin layer of the intestine (Troge *et al.*, 2012); the formation of biofilms due to type-1 fimbriae, F1C and Curli-fimbriae (Hancock, Dahl and Klemm, 2010); the induction of anti-inflammatory cytokines and defensins (Wehkamp *et al.*, 2004) and, naturally, the production of colibactin, which, beyond its procarcinogenic effects in mammalian systems, has been found to be an effective bacteriocin, inducing lytic development in bacteria that contain prophages (Silpe *et al.*, 2022). However, despite its probiotic use, EcN's carriage of the *pks* island confers a strong genotoxic activity. Accordingly, studies have shown that EcN induces DNA damage *in vivo* (Nougayrède *et al.*, 2021) and that colonic organoids exposed to EcN display characteristics that are typical of CRC cells (Iftekhar *et al.*, 2021). Furthermore, deletion of the colibactin biosynthesis genes in EcN results in a loss of probiotic activity (Olier *et al.*, 2012), immutably linking the probiotic and genotoxic effects of EcN and making it an exceptionally fascinating bacterium to investigate.

1.7 Amino acids

Amino acids (AAs) can broadly be distinguished by those which are proteinogenic, and are incorporated biosynthetically into proteins during translation, or non-coded AAs, which are usually formed through post-translational modifications (Ahluwalia, Kumar and Kumar, 2022). Of the over 500 AAs existing in nature, only 20 are proteinogenic, with food-derived proteins typically made up of L-AAs. All AAs, excepting glycine, exist as stereoisomers, where they may occur in a L-form (left-handed form) or D-form (right-handed form) as mirror images of each other. The distinction between L- and D- isomers does not depend on physical or chemical properties, but simply on the spatial orientation of the AA macromolecule (Genchi, 2017).

AAs have a vital function in intestinal homeostasis by playing roles in immune regulation, epithelium renewal, and gut hormone secretion. For example, glutamate and glutamine are the main contributors to energy metabolism in the mammalian intestinal epithelium, with research showing that these two AAs are extensively metabolised by enterocytes and colonocytes (Beaumont and Blachier, 2020). Beyond directly affecting host physiology, the acquisition and synthesis of AAs also influences the crosstalk between the host and its microbiota; driving the expression of virulence factors or contributing to microbial blooms of certain species (Lin *et al.*, 2017). For instance, the addition of asparagine to Dulbecco's Modified Eagle Medium appears to drive expression of the streptococcal invasion locus (*sil)* in Group A *Streptococcus* (Baruch *et al.*, 2014); similarly, glutamine appears to act as a switch for the virulence genes in *Listeria monocytogenes* (Haber *et al.*, 2017). Typically, significant metabolic alterations of AAs can be observed at the site of infection, such as following ETEC infection, where the relative abundances of six AAs are altered in the jejunum (Macia *et al.*, 2017). Studies carried out on milk-fed piglets and soybean-fed piglets showed that those in the latter group, fed a greater protein content, displayed a higher gastrointestinal pH and lower *Lactobacillus*-to-coliform ratio (Partanen and Mroz, 1999). Another consequence of the interplay occurring between the host, its microbiota, and AAs, is that the concentration of certain metabolites drives microbial niche specificity. By looking at the published literature, it is clear that dietary factors, particularly protein consumption, can alter gastrointestinal health

and consequently whole organismal physiology by shaping the gut microbiota's composition.

In trying to understand how diet can influence the gut microbiota to improve gastrointestinal health in the host, most studies have concentrated on the benefits of dietary fibre, on antibiotic-derived dysbiosis, and on the health risks posed by certain food additives. While several studies have been carried out looking at the role of AAs in host-pathogen interactions, more specifically at the effect of AAs on virulence factors, few have looked at the impact of specific AAs on the metabolism of commensal bacteria. The aim of this PhD project was to assess whether L- and D- isoforms of proteinogenic AAs in the diet can alter colibactin expression in EcN, reducing the risks of colibactin-driven CRC.

1.7.1 D-Amino acids

Currently, we still have no explanation as to why L-AAs evolved to become more prevalent in nature than D-AA. Interestingly, D-AA usage is often seen in prokaryotes, with their inclusion fundamental to the production of peptidoglycan, a key component in the cell wall. Other D-AAs are found in antibiotics, opioids, hormones, and neuropeptides (Grishin *et al.*, 2019). Interestingly, a lot of toxic compounds contain D-AAs: for example, the therapeutic capacity of antibiotics like gramicidin S and polymyxin B is modulated by D-Phenylalanine contained in their structure (Solanas *et al.*, 2009); gastropods of the family *Conidae* produce venomous secretions containing D-Leucine and D-Tryptophan that are fatal to fish and mammals (Grishin *et al.*, 2019); and the venomous North-American spider *Agelenopsis aperta* produces agatoxins that can block calcium channels and differ only by the presence of D-Serine, compared to the less toxic analogues (Murkin and Tanner, 2002).

D-AAs can be acquired in the human diet through ingestion of foods containing D-AAs, as by-products produced by the microbiota, and derived from L-AAs through the process of racemisation. Food-treatment processes that can induce racemisation include exposure to high temperatures, changes in pH, or fermentation processes. For example, whereas fruits like apples, grapes and oranges naturally contain D-Alanine, D-Asparagine, D-Arginine and D-Glutamine

(Gandolfi *et al.*, 1994); their presence in dairy products is usually an indicator of thermal or alkaline treatment, or of ageing.

1.7.2 D-Serine

D-Serine is one of the most abundant D-AAs found in mammals and is often studied in the context of the nervous system. In the human body, it is synthesised by the serine racemases and degraded by the D-amino acid oxidase (DAAO). D-Serine has been detected in the frontal brain area of rats at a particularly high concentration, representing up to 25% of the total serine levels (Hashimoto *et al.*, 1992). Subsequent studies have reported localisation within frontal brain areas, including the hippocampus and hypothalamus, and *in vitro* and *in vivo* studies have demonstrated that D-Serine serves as a co-agonist for the N-methyl D-Aspartate (NMDA) receptor of the mammalian brain (Shleper, Kartvelishvily and Wolosker, 2005), (Wolosker *et al.*, 2008). D-Serine continues to be studied for its potential therapeutic role in neurological disorders like Alzheimer's disease, amyotrophic lateral sclerosis, and schizophrenia (Bastings *et al.*, 2019).

In addition to being abundant in the mammalian brain, D-Serine is also abundant in the human bladder, having been detected at concentrations of up to 1 mM in urine (Anfora *et al.*, 2007). No relevant literature on the exact concentration of D-Serine in the human intestine has been found (Nishikawa, 2011), but it is estimated to be lower than in the bladder, with data from the Roe Group showing that D-Serine is found in the GI tract of mice at a concentration of approximately 1 µM (Connolly *et al.*, 2015). In part, this is also suggested by the fact that *E. coli* strains which are limited to the gastrointestinal tract, such as the EHEC strain O157:H7, lack the genes required for D-Serine catabolism. UPEC strains carry a complete *dsdCXA* locus, encoding for a system regulator (DsdC), a D-Serine inner membrane transporter (DsdX) and a D-Serine deaminase (DsdA), allowing UPEC strains to covert D-Serine into pyruvate and ammonia, overcoming the metabolite's high concentrations in the human urinary tract and using D-Serine as a sole carbon source. In contrast, EHEC contains a truncated version of the *dsdCXA* locus, in which a portion of *dsdC* and *dsdX* are substituted by the *csrRAKB* sucrose utilization locus, and fails to use D-Serine as a sole carbon source (Sabri, Nielsen

and Vickers, 2013). Based on this evidence, the Roe Group investigated the role of D-Serine on the virulence factors of different *E. coli* pathotypes. It discovered that D-Serine activates the SOS response and inhibits expression of the Type 3 Secretion System (T3SS) in EHEC O157:H7 (Connolly *et al.*, 2015). This suggests an 'evolutionary incompatibility', whereby inhibition of virulence by D-Serine, prompted EHEC adaptation to non-D-Serine rich environments such as the gut (Connolly *et al.*, 2015). Indeed, while UPEC can harmlessly travel through the digestive system (Chen *et al.*, 2013), EHEC – despite having the capacity to attach to different cell types – only causes infections in the GI tract (Nguyen and Sperandio, 2012). It was in fact demonstrated that of the 1591 strains tested, only 1.6% carried both the locus of enterocyte effacement (LEE), which encodes for the T3SS, and the *dsdCXA* locus (Connolly *et al.*, 2015). Interestingly, the Roe Group also showed that the transcriptional responses to D-Serine are distinct, even among *E. coli* pathotypes that can metabolise the AA: exposure to D-Serine led to the differential expression of 140 genes in UPEC and 55 genes in NMEC, with only 12 genes in common to both strains (Connolly *et al.*, 2021). Previous research from the Roe Group, also showed that D-Serine represses expression of colibactin biosynthesis genes in the UPEC strain CFT073 (Hallam *et al.*, 2023).

1.8 PhD project aims

CRC is among the deadliest malignant diseases worldwide, and in addition to being linked to genetic and dietary factors, it can be driven by oncobacteria in the intestinal flora. In particular, bacteria containing a genomic cluster known as the *pks* island produce a potent genotoxin, colibactin, that plays a role in colorectal carcinogenesis. Given the powerful capacity of dietary amino acids to influence both organismal physiology and bacterial metabolism, and given the Roe Group's long-standing interest in the effects of amino acids on bacterial virulence, the aims of this PhD project were the following:

1. To investigate which amino acids have a modulatory effect on colibactin expression, with the particular goal of identifying those with repressive activity against colibactin.

2. To establish whether specific combinations of amino acids offer synergistic effects, further reducing colibactin expression.

3. To investigate whether colibactin-producing bacteria can enhance colorectal tumours *in vivo*, and whether their carcinogenic capacity can be hindered by the administration of dietary amino acids.

4. To investigate the molecular mechanisms underlying amino acid-driven downregulation of colibactin.

Most of the work in this thesis was carried out using the *E. coli* strains Nissle 1917 (EcN) and CFT073. Stocks of these two strains were already present in the Andrew Roe (AR) Group at the start of the PhD project, but new stocks were acquired from the Eric Oswald (EO) Group at the Institut de Recherche en Santé Digestive in Toulouse, France. Strains were subsequently referred to as EcN_AR, CFT073_AR, EcN_EO, and CFT073_EO. Experiments performed within both the Roe and Oswald laboratories suggested that EO strains expressed more colibactin than AR, hence the choice to carry out most experiments with EO strains. When a specific strain was used for an experiment, it is specified in the text. If not specified, experiments were carried out with both strains.

2 Materials and Methods

2.1 Chemicals and growth media

2.1.1 Chemicals and molecular reagents

Chemicals were purchased from Thermo Fisher Scientific, Merck, and their respectively-owned brands, unless otherwise specified. Most of the kits used were from Qiagen, GoTaq DNA polymerase was purchased from Promega, restriction enzymes were purchased from Promega and New England Biolabs, oligonucleotides were purchased from Invitrogen, cDNA synthesis kits and qPCR master mixes were purchased from New England Biolabs.

2.1.2 Growth media, antibiotics, gel solutions, and buffers

Growth media, antibiotics, gel solutions, and buffers were prepared using nuclease-free water (nf H_2O), or distilled deionised water (dd H_2O), and sterilised by autoclaving or, if heat sensitive, by filtration with 0.2 µm filters. Minimum Essential Medium Eagle (MEM-HEPES) Dulbecco's Modified Eagle Medium (DMEM) were purchased from Sigma-Aldrich.

Table 2-1 Lysogeny Broth (LB) – 1 L

Component	Quantity
Tryptone	10 g
Yeast Extract	5 g
NaCl	10 g

*Stored at room temperature.

Table 2-2 M9 Minimal Media – 100 mL

Component	Quantity
dd H_2O (autoclaved)	77.8 mL
M9 salts (5X)	20 mL
20% glucose	2 mL
$MgSO_4$ (1 M)	200 µm
$CaCl_2$ (1 M)	100 µm

*M9 (5X) salt solution was sterilised by autoclaving, while glucose, $MgSO_4$ and $CaCl_2$ were filter-sterilised. The components were added together to make fresh M9 Minimal Media on the day of experiment. All components were stored at room temperature.

Table 2-3 MacConkey Agar – 100 mL

Component	Quantity
MacConkey agar	5.2 g
dd H_2O	Make up to 100 mL

*The agar was dissolved in water by microwaving and then sterilised by autoclaving. It was stored at room temperature.

Table 2-4 1X Phosphate-buffered saline (PBS) – 1 L

Component	Quantity
NaCl	8 g
KCl	0.2 g
Na_2HPO_4	1.44 g
KH_2PO_4	0.24 g
dd H_2O	Make up to 1 L

*PBS was made up by dissolving all reagents in 800 mL of dd H_2O and then making up the solution to a final volume of 1L. It was then sterilised by autoclaving. Note that commercial PBS (Invitrogen) was used for tissue culture. Both types of PBS were stored at room temperature.

Table 2-5 50X Tris-acetate-ED) Buffer – 1 L

Component	Quantity
Tris -base	242 g
Glacial acetic acid	57.1 mL
EDTA (0.5 M)	18.61 g
dd H_2O	Make up to 1 L

*TAE Buffer was made up by dissolving Tris-base and EDTA in 700 mL of dd H_2O, adding glacial acetic acid, and then making up the buffer to a final volume of 1 L. To create and run agarose gels, TAE Buffer was diluted and used at a concentration of 1X. It was stored at room temperature.

Table 2-6 4% Paraformaldehyde (4% PFA) – 100 mL

Component	Quantity
Paraformaldehyde	4 g
PBS (10X)	10 mL
NaOH (1 M)	1 mL
dd H_2O	Make up to 100 mL

*PFA was made up by dissolving 4 g of PFA in 50 mL of dd H_2O by adding 1 mL of NaOH and stirring continuously at 60°C. 10X PBS was then added, the pH was adjusted to 7.4 with HCl and the final volume was made up to 100 mL with dd H_2O. The 4% PFA was then left to cool before filter-sterilising (0.2 µm) and storing at -20°C.

Table 2-7 Methylene blue staining solution - 500 mL

Component	Quantity
Methylene blue powder	5 g
Tris-base solution (0.01 M, pH = 8.5)	500 mL

*Methylene blue was stirred continuously for 3-4 hrs until dissolved, and the staining solution was stored at room temperature.

Table 2-8 Methylene blue wash buffer – 500 mL

Component	Quantity
Tris-base solution (1 M)	5 mL
dd H_2O	495 mL

*Stored at room temperature.

Table 2-9 Alkaline gel solution – 500 mL

Component	Quantity
NaCl	2.92 g
EDTA (0.5 M)	2 mL
Autoclaved dd H_2O	Make up to 500 mL

*The alkaline gel solution was made up by dissolving NaCl in 300 mL of autoclaved dd H_2O, adding EDTA (Sigma-Aldrich), and then adding the remaining dd H_2O to a final volume of 500 mL. The solution was stored at room temperature.

Table 2-10 Alkaline gel running buffer – 800 mL

Component	Quantity
NaOH	32 mL
EDTA (0.5 M)	1.6 mL
Autoclaved dd H_2O	Make up to 800 mL

*The buffer was stored at room temperature.

Table 2-11 Alkaline gel neutralisation buffer – 1 L

Component	Quantity
NaCl	8.77 g
Trizma hydrochloride solution (1 M)	100 mL
Autoclaved dd H_2O	Make up to 1 L

*The neutralisation buffer was made up by dissolving NaCl in Trizma hydrochloride solution (Sigma-Aldrich), and then adding autoclaved dd H_2O to a final volume of 1 L. The buffer was stored at room temperature.

Table 2-12 Antibiotic stocks for selective bacterial growth in liquid culture and on agar

Antibiotic	Stock (mg/mL)	Solvent	Final concentration (µg/mL)
Chloramphenicol	25	Ethanol	25
Ampicillin	100	dd H_2O	100

*Other antibiotics used, include gentamicin solution (Sigma-Aldrich) for *in cellulo* infection assays (200 µL/mL), pen-strep solution (Sigma-Aldrich) for tissue culture media (10 mL/L), streptomycin in dd H_2O (2 mg/mL) prior to infections in mouse models and hygromycin B (Thermo Fisher Scientific) used for lambda red recombineering (100 µg/mL).

2.1.3 Amino acid stocks

If performing large screenings involving all 40 amino acids (AAs) in 96-well plates and/or if working with small volumes (250 µL), soluble AAs were dissolved in autoclaved dd H_2O at a concentration of 5 mM , filter-sterilised (2 µm filters), and combined at a 1:4 dilution with growth medium to achieve a final AA concentration of 1 mM. Insoluble AAs (Seleno-L-Cysteine, L-Tyrosine, D-Tyrosine) were instead made up at a concentration of 1.25 mM in the growth medium of interest and combined at a 4:1 dilution with sterile, dd H_2O to achieve a final AA concentration of 1 mM. When working with larger volumes and screening few, individual AAs these were made up at 5 mM in the growth medium of interest and, if soluble, filter-sterilised. If working with D-Serine alone (Thermo Fisher Scientific), the AA was dissolved at a concentration of 1 M, filter-sterilised, and used at a 1:1000 dilution to achieve a final concentration of 1 mM.

2.2 Bacterial strains, plasmids, and primers

2.2.1 Bacterial strains

The bacterial strains used in this project are compiled in the table below.

Table 2-13 Bacterial strains used in this project

Strain	Description	Source
EcN_AR	Wild type *E. coli* Nissle 1917 O6:K5:H1	Roe laboratory inventory
EcN_EO	Wild type *E. coli* Nissle 1917 O6:K5:H1	Prof. Eric Oswald
CFT073_AR	Wild type UPEC O6:K2:H1	Roe laboratory inventory
CFT073_EO	Wild type UPEC O6:K2:H1	Prof. Eric Oswald
EcN_AR Δ*lacI*	EcN_AR *lacI* knockout	This study
EcN_EO Δ*lacI*	EcN_EO *lacI* knockout	This study
EcN_AR Δ*clbB*	EcN_AR *clbB* knockout	This study
EcN_EO Δ*clbB*	EcN_EO *clbB* knockout	This study
EcN_AR Δ*glnG*	EcN_AR *glnG* knockout	This study
EcN_EO Δ*glnG*	EcN_EO *glnG* knockout	This study
DH10B	K12 strain hosting pBAC-*pks*	Prof. Eric Oswald
EC0_32	Clinical isolate O1:H7 (catheter-associated, patient diagnosed with cholecystitis).	Prof. Tom Evans
EC0_22	Clinical isolate O6:H1 (catheter-associated, patient diagnosed with UTI).	Prof. Tom Evans
EC0_62	Clinical isolate O2:H7 (catheter-associated, patient diagnosis unrecorded).	Prof. Tom Evans
EC0_79	Clinical isolate O6:H1 (catheter-associated, patient diagnosis unrecorded).	Prof. Tom Evans
EC1_93	Clinical isolate O6:H31 (catheter-associated, patient diagnosed with urosepsis).	Prof. Tom Evans
EC1_91	Clinical isolate O18:H1 (catheter-associated, patient diagnosed with urosepsis).	Prof. Tom Evans
EC0_58	Clinical isolate O18:H7 (catheter-associated, patient diagnosis unrecorded).	Prof. Tom Evans
EC1_57	Clinical isolate O18:H31 (catheter-associated, patient diagnosed with urosepsis).	Prof. Tom Evans

EC1_60	Clinical isolate O22:H1 (catheter-associated, patient diagnosed with urosepsis).	Prof. Tom Evans

2.2.2 Bacterial stocks and maintenance

Overnight cultures were grown by inoculating a single colony from a solid agar plate, supplemented with antibiotics if necessary for plasmid expression, into 5 mL of LB medium, supplemented with the same antibiotic if required. Cultures were grown at 30°C, 37°C or 42°C depending on the bacterial strain, with 200 rpm shaking (New Brunswick Scientific controlled environment shaker), for 12-16 hours. The following day, if long-term storage of the bacterial strain was necessary, 0.5 mL of culture were added to a cryo-vial containing 1 mL of sterile glycerol (40%) and peptone (2%) and the stock was placed at -80 °C. Bacterial strains from -80°C were streaked on LB agar plates (supplemented with appropriate antibiotics if necessary), grown at the relevant temperature for 12-16 hours, and stored at 4°C for up to 2 weeks. Single colonies from these plates were then used for overnight culture inoculation, described above.

2.2.3 Bacterial growth curves

Growth curves were generated by inoculating growth medium with overnight cultures to give a starting OD_{600} of 0.06. These cultures were then incubated at 37°C, 200 rpm shaking. The OD_{600} was measured hourly using a spectrophotometer (BioPhotometer Eppendorf) and plotted against time. A sterile sample of growth medium was used to calibrate the spectrophotometer prior to each measurement. When the OD_{600} reached a value greater than 1.0, the bacterial culture was diluted 1:10 in growth media in order to obtain a more accurate measurement.

2.2.4 Bacterial plasmids

The bacterial plasmids used in this project are compiled in the table below.

Table 2-14 Plasmids used in this project

Plasmid	Description	Source
pclbB:gfp	*clbB* promoter translational fusion of pAJR70 to eGFP (Chl^R)	Roe laboratory inventory
pACYC184	Multicopy number plasmid (Chl^R)	Roe laboratory inventory
pSIM18	Temperature-inducible red recombinase system, 42°C inducible (Tet^R, Hyg^R)	Chan *et al*., 2007
pKD3	Template plasmid for Lambda Red mutagenesis (Chl^R, Amp^R)	Datsenko and Wanner, 2000
pCP20	Encoding for FLP recombinase to remove antibiotic-resistance cassettes (Amp^R)	Cherepanov and Wackernagel, 1995
pUC19	Plasmid used for interstrand crosslinking assays	Roe laboratory inventory

2.3 Molecular techniques

2.3.1 Polymerase chain reaction (PCR)

Primers used in this project (Table 2-15) were designed in Primer3web, with lengths of approximately 20 bp, melting temperatures of 55-60 °C, and a GC content of approximately 50%. Check primers and qRT-PCR primers were designed using regions complementary to template DNA. Lambda Red recombination primers were designed bearing approximately 20 bp complementary to pKD3 and 50 bp at the 5'-end flanking regions, homologous to the 50 bp directly upstream and downstream of the target gene. Oligonucleotides were purchased from Invitrogen, reconstituted with nf H_2O at a final concentration of 100 µM. They were then diluted to 10 µM using nf H_2O of use in PCR reactions and both stocks were stored at -20°C.

Table 2-15 Primers used in this project.

Name	Description	Sequence (5' to 3')	Amplicon size (bp)
gapA_Q_F	RT-qPCR *gapA* forward	TTTCCGTGCTGCTCAGAAAC	107
gapA_Q_R	RT-qPCR *gapA* reverse	GGCCGTGAGTGGAGTCATAT	

clbB_Q_F	RT-qPCR *clbB* forward	TTGTCTCCGGATGTGTGTCA	123
clbB_Q_R	RT-qPCR *clbB* reverse	CACATCGTCAGCATAGCACC	
lacI.red_F	Lambda Red *lacI* forward	AAGACTGGGGTGCCTAATGAGTGAGCTGACTCACATTAATTGTGTTGCGCGTGTAGGCTGGAGCTGCTTC	1115
lacI.red_R	Lambda Red *lacI* reverse	GCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATATATGAATATCCTCCTTAG	
lacI_F	Lambda Red forward check (*lacI*)	GTTGTAATACGACGGCCAGA	1340 if KO successful (pKD3)
lacI_R	Lambda Red reverse check (*lacI*)	ACCTGTAAACGCCGTAGGTG	1409 if KO unsuccessful 325 if KO successful and clean (no Chl^R)
rpsL_F	Forward for *rpsL*	AATTCGGCGTCCTCATATTG	555
rpsL_R	Reverse for *rpsL*	GTGGCATGGAAATACTCCGT	
27F	Forward for 16S sequencing	AGAGTTTGATCCTGGCTCAG	Variable
1392R	Reverse for 16S sequencing	GGTTACCTTGTTACGACTT	
clbB.red_F	Lambda Red *clbB* forward	TAGAATAATCATGTTAAATAATCTATAAATCCGATAATAAGGTGATGGTTGTGTAGGCTGGAGCTGCTTC	1115
clbB.red_R	Lambda Red *clbB* reverse	CGTATTCCATAAACTTCATTACCTTTCTTCAGGGCGAGGAGGCGCCCGGAATATGAATATCCTCCTTAG	
clbB_F	Lambda Red forward check (*clbB*)	TCACGCAGAGATAACGGGTT	1667 if KO successful (pKD3)
clbB_R	Lambda Red reverse check (*clbB*)	CTCGCTAAAGAAGGTGACGC	10274 if KO unsuccessful 701 if KO successful and clean (no Chl^R)
glnG.red_F	Lambda Red *glnG* forward	CATACCGAGTTCTCGGTTTACCTGCCTATCAGGAAATAAAGGTGACGTTTGTGTAGGCTGGAGCTGCTTC	1115
glnG.red_R	Lambda Red *glnG* reverse	AATACCAGCAATTTGCGCTCAATAATCAATCTTTACACACAAGCCGTGAAATATGAATATCCTCCTTAG	

glnG_F	Lambda Red forward check (*glnG*)	CGCGGATTGATGTGGAAGAT	1514 if KO successful (pKD3) 1909 if KO unsuccessful 548 if KO successful and clean (no Chl^R)
glnG_R	Lambda Red reverse check (*glnG*)	CTGGTTGCGTCTTATCAGGC	
Rps29_F	RT-qPCR *Rps29* forward	ACGGTCTGATCCGCAAATAC	138
Rps29_R	RT-qPCR *Rps29* reverse	CATGATCGGTTCCACTTGGT	
Tnf_F	RT-qPCR *Tnf* forward	CCTGTAGCCCACGTCGTAG	442
Tnf_R	RT-qPCR *Tnf* reverse	GGGAGTAGACAAGGTACAACCC	
Ifng_F	RT-qPCR *Ifng* forward	TGAGTATTGCCAAGTTTGAG	159
Ifng_R	RT-qPCR *Ifng* reverse	CTTATTGGGACAATCTCTTCC	
Il1b_F	RT-qPCR *Il1b* forward	TTTTCCTCCTTGCCTCTGAT	104
Il1b_R	RT-qPCR *Il1b* reverse	GAGTGCTGCCTAATGTCCCC	
Il6_F	RT-qPCR *Il6* forward	CTCTGGGAAATCGTGGAAATG	75
Il6_R	RT-qPCR *Il6* reverse	AAGTGCATCATCGTTGTTCATACA	
Cldn2_F	RT-qPCR *Cldn2* forward	GTGGCTGTAGTGGGTGGAGT	178
Clnd2_R	RT-qPCR *Cldn2* reverse	CCAAAGAAAACAGGGCTGAG	
Tjp1_F	RT-qPCR *Tjp1* forward	ACTCCCACTTCCCCAAAAAC	166
Tjp1_R	RT-qPCR *Tjp1* reverse	CCACAGCTGAAGGACTCACA	
Cdh1_F	RT-qPCR *Cdh1* forward	GGATCAGGACCAGGACTACG	169
Cdh1_R	RT-qPCR *Cdh1* reverse	AGGGAAGGAGCTGAAAGACC	
F11r_F	RT-qPCR *F11r* forward	CACCTTCTCATCCAGTGGCATC	442
F11r_R	RT-qPCR *F11r* reverse	CTCCACAGCATCCATGTGTGC	

S100a8_F	RT-qPCR *S100a8* forward	TGTCCTCAGTTTGTGCAGAATATAAA	103
S100a8_R	RT-qPCR *S100a8* reverse	TCACCATCGCAAGGAACTCC	

PCR reactions were set up in nuclease-free tubes on ice, typically at a 20 µL or 25 µL volume. Two different types of polymerases were used, based on the length of the gene to be amplified. GoTaq Hot Start Green Master Mix (Promega) was used for standard PCR reactions (Table 2-16), amplifying targets of approximately 100 bp – 4 kb, while GoTaq Long PCR Master Mix (Promega) was used for PCR reactions, amplifying targets above 4 kb (Table 2-17). Q5 High-Fidelity DNA Polymerase Reaction Buffer (New England Biolabs) was used when amplifying PCR targets to be sequenced(Table 2-17). Genomic or plasmid template DNA was typically used at a concentration of 100 ng/µL. If using a bacterial colony as template, the colony was picked from a plate, mixed into 50 µL of nf H_2O, of which 1 µL was then added to the PCR reaction mix.

Table 2-16 Standard PCR reaction mix using GoTaq Hot Start Green Master Mix

Component	Volume (total 20 µL)
nf H_2O	8 µL
Forward primer (10 µM)	0.5 µL
Reverse primer (10 µM)	0.5 µL
GoTaq Hot Start Green Master Mix (2X)	10 µL
Template DNA (100 ng/µL)	1 µL

Table 2-17 PCR reaction mix using GoTaq Long PCR Master Mix or Q5 High-Fidelity DNA Polymerase Reaction Buffer

Component	Volume (total 25 µL)
nf H_2O	9 µL
Forward primer (10 µM)	1.25 µL
Reverse primer (10 µM)	1.25 µL
GoTaq Long PCR Master Mix (2X) or Q5 DNA Polymerase Reaction Buffer (2X)	12.5 µL
Template DNA (100 ng/µL)	1 µL

PCR conditions were optimised based on the length of the PCR target to be amplified and on the polymerase being used. Amplification was performed using the SimpliAmp Thermal Cycler. The tables below indicate the thermocycling parameters for the different polymerases used (Table 2-18 to 2-20).

Table 2-18 Standard PCR cycle using GoTaq Hot Start Green Master Mix (35 cycles)

Step	Temperature	Time
Initial Denaturation	95°C	3 min
Denaturation	95°C	15 sec
Annealing	55°C	30 sec
Extension	72°C	1 min/kb
Final Extension	72°C	5 min
Hold	4°C	Hold

Table 2-19 PCR cycle using GoTaq Long PCR Master Mix (35 cycles)

Step	Temperature	Time
Initial Denaturation	95°C	2 min
Denaturation	94°C	15 sec
Annealing	59°C	30 sec
Extension	65°C	11 min
Final Extension	72°C	10 min
Hold	4°C	Hold

Table 2-20 PCR cycle using Q5 High-Fidelity DNA Polymerase Reaction Buffer (30 cycles)

Step	Temperature	Time
Initial Denaturation	98°C	30 sec
Denaturation	98°C	10 sec
Annealing	60°C	30 sec
Extension	72°C	20 sec
Final Extension	72°C	5 min
Hold	4°C	Hold

2.3.2 Purification of PCR products

PCR products were purified using the QIAquick PCR purification kit (Qiagen), eluted in 30 µL of nf H_2O and combined with BlueJuice loading buffer (Invitrogen) for visualisation by agarose gel electrophoresis.

2.3.3 Preparation of genomic DNA

Genomic DNA (gDNA) was extracted using the DNeasy Blood & Tissue kit (Qiagen) from 1 mL of overnight bacterial culture. DNA was resuspended in nf H_2O and

stored at -20°C. DNA concentration was measured using the DS-11 FX spectrophotometer (DeNovix).

2.3.4 Agarose gel electrophoresis

Amplified PCR products were typically visualised on 1-1.5% agarose gels made using 1X TAE buffer. Agarose was dissolved in TAE buffer by microwaving and stirring. Once cooled, GelRed Nucleic Acid Stain was added at a ratio of 1:10000 per volume of buffer and the molten agarose was poured into a cast to set. Agarose gels were placed into tanks filled with TAE buffer, and typically run at 100 V for 60 mins, or adjusted depending on gel % and PCR product size. The DNA was visualised using a UV transilluminator (Azure biosystems).

2.3.5 Plasmid DNA purification

An overnight culture of bacteria carrying the desired plasmid was centrifuged at 3750 g for 5 minutes (Beckman Coulter Allegra X-12R Centrifuge). The volume of bacterial culture grown depended on the amount of plasmid DNA required. Plasmids were purified using either the QIAprep Spin Miniprep kit (Qiagen) or the Qiagen Plasmid Maxi kit (Qiagen). Purified plasmid DNA was eluted in 30-100 µL of nf H_2O and stored at -20°C. Whenever possible, plasmid DNA purification was performed from DH5α cells.

2.3.6 Preparation of competent cells

An overnight culture of the bacterial strain of interest was prepared as described above. The following day, 50 µL of the overnight culture were withdrawn and dispensed in 5 mL of LB. The new culture was grown to an OD_{600} ranging between 0.3 and 0.4, and immediately centrifuged at 3750 g (Beckman Coulter Allegra X-12R Centrifuge) for 5 mins at 4°C. The supernatant was then discarded and the bacterial cell pellet was washed three times with 1 mL of nf ice-cold water,

centrifuging at 21130 g (Centrifuge 5424 R Eppendorf) for 1 min at 4°C between washes. Cells were then resuspended in 50 µL of nf ice-cold water and used in transformations.

2.3.7 Transformations by heat-shock

A total of 50 µL of competent cells were mixed with 5 µL of plasmid and incubated on ice for 30 mins. The samples were then heat-shocked at 42°C for 45 seconds and immediately placed back on ice for 2 mins. A volume of 1 mL of pre-warmed LB was added to the tube, and cells were left to recover for 2 hrs at 37°C in a shaking incubator, before plating 100 µL of the reaction onto solid agar plates containing the appropriate selective antibiotic.

2.3.8 Transformations by electroporation

A total of 50 µL of competent cells were mixed with 1 µL of plasmid (100 ng/µL) and transferred to a pre-chilled electroporation cuvette. The samples were then shocked at 2500 V using an electroporator (Eporator Eppendorf) and 1 mL of pre-warmed LB was added to each sample. Cells were left to recover for 2 hrs at 37°C in a shaking incubator before plating 100 µL of the reaction onto solid agar plates which contained the appropriate selective antibiotic, necessary for plasmid expression.

2.3.9 Lambda Red recombineering

A linear deletion fragment was prepared by amplifying the chloramphenicol resistance cassette from pKD3 with oligonucleotides bearing 50 bp homologous to the regions immediately upstream and downstream of the gene intended for deletion. Four 50 µL reactions were prepared, pooled into a single tube, and purified with the QIAquick PCR purification kit (Qiagen). The PCR product was

eluted in nf H_2O, and an appropriate concentration of approximately 100 ng/µL was confirmed using the DS-11 FX spectrophotometer (DeNovix).

Prior to recombination, strains were transformed with pSIM18 by electroporation at 2500 V, as described above. The plasmid pSIM18, contains the three Red genes (*exo*, *bet*, and *gam*), required for recombineering and the λ-CI857 repressor, which is active at 32°C and becomes inactive at 42°C (Lee, Wang and Liu, 2009). A single pSIM18-transformed colony was inoculated in LB supplemented with 100 µg/ml hygromycin at 30°C, 200 rpm and kept dark until an OD_{600} of 0.3-0.4 was reached. The culture was placed in a water bath at 42°C for 15 minutes to activate the temperature-inducible red recombinase system, and then centrifuged at 3750 x g for 5 minutes at 4°C. The pellet was washed three times with nuclease-free H_2O, resuspended in a final volume of 50 µL, mixed with 1 µL of PCR product (≅ 100 ng/µL) and electroporated at 2500 V. Cells were recovered at 37°C, 200 rpm for 2 hrs and plated onto agar containing 25 µg/ml chloramphenicol. Insertional mutants were verified by PCR using check primers and, if successful, further confirmed by Sanger sequencing. The pKD3 resistance cassette was eliminated through pCP20, a helper plasmid expressing the FLP recombinase, which acts on the FLP recognition target (FRT) sites flanking the resistance cassette. Since pCP20 is temperature sensitive, it was cured by growth at 42°C.

2.4 Genome and transcriptome analyses

2.4.1 Transcriptional reporter fusion assays

Gene expression was measured indirectly using a promoter:GFP reporter fusion construct (Figure 9-1). Electrocompetent cells were transformed with *pclbB:gfp* (to measure *pclbB* fluorescence) and pACYC184 (to measure background fluorescence) plasmids (Table 2-14). The transformants were inoculated in LB, grown overnight as previously described, and made into stocks the following day. If performing reporter assays to screen numerous conditions (such as screening 40 AAs simultaneously), growth media supplemented with chloramphenicol (25 µg/mL) was directly dispensed in the 96-well plate. AAs were then added to each well to obtain a final concentration of 1 mM and wells were directly inoculated with the overnight cultures of EcN_AR *pclbB:gfp* or EcN_AR pACYC184. Bacterial cultures were grown in a humidity chamber, placed in an incubator at 37°C, 200 rpm shaking. Readings were taken hourly, as described below.

Instead, if performing reporter assays with a smaller number of samples, overnight cultures were sub-cultured at a starting OD_{600} of 0.06 in 10 mL of the appropriate growth medium supplemented with chloramphenicol (25 µg/mL) and with 1 mM of the AA of interest. To determine overall promoter activity, 250 µL of culture were transferred into a clear 96-well plate for hourly readings. Absorbance was measured with wavelength of 600 nm, and fluorescence was read with excitation of 485 nm and emission of 520 nm using the Varioskan LUX plate reader (Thermo Fisher Scientific). Sterile growth medium was dispensed in one well of the 96-well plate and its absorbance was measured hourly to normalise absorbance values at each time point. Fluorescence values were normalised by subtracting background fluorescence intensity. Further normalisation was performed by dividing fluorescence by absorbance to obtain relative fluorescence units (RFUs). All assays were performed in triplicate from separate growth cultures. Statistical comparisons were performed on RFU values obtained in exponential phase.

2.4.2 RNA extraction and DNAse treatment

Bacterial cultures were grown in MEM-HEPES supplemented with 1 mM of the AAs of interest for 5-7 h at 37 °C, 200 rpm shaking. Cells were then harvested by centrifugation, resuspended in 1 mL of RNA protect (Qiagen), incubated at room temperature for 10 mins, and centrifuged once more before removing the supernatant. Pellets were then resuspended in 100 µL lysozyme solution, made up at 10 mg/mL in Tris-EDTA (TE) buffer (Invitrogen) and 0.5 µL of 10% sodium dodecyl sulfate (SDS) solution (Invitrogen) were added to each sample, before vortexing to resuspend the pellet. After a 5 min incubation, 350 µL of lysis buffer (Invitrogen) supplemented with 1% B-mercapthoethanol were dispensed in each tube. Samples were vortexed, 250 µL of 100% ethanol were added to each tube, and the total volume (700 µL) was transferred to a cartridge with collection tube. Bacterial total RNA extraction was performed using the PureLink RNA Mini kit (Invitrogen) according to manufacturer's instructions. RNA was eluted in a final volume of 100 µL of RNA-free H_2O. DNA contamination was removed using the Turbo DNase kit (Invitrogen), with 4 µL of Turbo DNase enzyme solution and 10 µL of Turbo DNase buffer added to each samples. Samples were incubated for 1 h at 37 °C on a heat-block, and vortexed and centrifuged every 15 mins.

2.4.3 Phenol-chloroform extraction and ethanol precipitation

Prior to phenol-chloroform extractions, 300 µL of RNA-free H_2O were added to each sample to make up a total, final volume of 400 µL. An equal volume of phenol:chloroform:isoamyl alcohol (PCIA, Sigma Aldrich) was added to each sample. Tubes were then centrifuged for 5 min at 17000 x g (Heraeus Pico 17 Centrifuge, Thermo Fisher Scientific), the top layer was transferred to a new tube, to which 400 µL of chloroform:isoamyl alcohol (CIA, Sigma Aldrich) were then added. The tubes were inverted several times and centrifuged as above for 1 min. The top layer was once again removed and dispensed into a new Eppendorf tube containing 800 µL 100% ethanol, 40 µL sodium acetate, 1 µL GlycoBlue Coprecipitant (Thermo Fisher Scientific). Samples were inverted several times and stored at -20°C overnight or -80°C for 1 h. Following freezer incubation, samples were centrifuged at 17000 x g at 4°C for 20 min. The supernatant was carefully

removed, and 1 mL of 70% ethanol was then added to wash the pellet. Samples were centrifuged at 17000 x g at 4°C for 5 min, and the supernatant was then aspirated. Pellets were left to airdry on ice before being resuspended in 50 µL of nf H_2O (if proceeding with RT-qPCR) or of TE buffer (if proceeding with RNA-seq). Concentrations were measured using the DS-11 FX spectrophotometer (DeNovix). For RT-qPCR processing samples were then diluted in nf H_2O to a final concentration of 10 ng/µL. DNA contamination was assessed by PCR (using primers of a gene known to be present in the bacterial strain whose RNA was extracted) and subsequent agarose gel electrophoresis. RNA integrity was assessed by agarose gel electrophoresis.

2.4.4 Transcriptomic analyses by RNA-Seq

EcN was cultured in M9 in the presence or absence of 1 mM D-Serine, 1 mM D-Tyrosine, or 1 mM of both AAs for 7 h at 37°C, 200 rpm shaking. For each growth condition, four extractions were performed from separately grown cultures. RNA was extracted as described above and was checked for DNA contamination and RNA integrity before being delivered to the University of Glasgow Polyomics facility, which performed depletion and cDNA library preparation. RNA sequencing was performed to a depth of 100 bp single-end reads using a NextSeq2000 sequencing system (Illumina). Data analysis was performed using CLC Genomics Workbench version 21.0.5 (Qiagen), with reads mapped to the EcN reference genome (NCBI accession number NZ_CP007799.1). The Differential Expression for RNA-seq tool in the CLC Genomics Workbench 21 was used to calculate differential gene expression. Genes were considered significantly differentially expressed with a fold change of ≥1.5 or ≤1.5 and a false-discovery rate (FDR)-corrected *P* value of ≤0.05. Volcano plots were created using GraphPad Prism Version 9.2.0.

2.4.5 RT-qPCR of RNA extracted from bacterial cells

As previously mentioned, for RT-qPCR processing, RNA samples were diluted in nf H_2O to a final concentration of 10 ng/µL. RNA was then converted into cDNA using

the LunaScript RT SuperMix kit (New England Biolabs) as per manufacturer's instructions, indicated below (Table 2-21).

Table 2-21 cDNA synthesis reaction mix using the LunaScript SuperMix

Component	Volume (10 µL)
LunaScript RT+ or RT- Supermix 5x	2 µL
nf H_2O	7 µL
Extracted RNA (10 ng/µL)	1 µL

The samples were then subjected to thermocycling under the following conditions (Table 2-22).

Table 2-22 Thermocycling conditions for cDNA synthesis from bacterial RNA

Step	Temperature	Time
Annealing	25°C	2 min
cDNA synthesis	55°C	15 min
Heat inactivation	95°C	1 min
Hold	4°C	Hold

For RT-qPCR the Luna Universal qPCR master mix (New England Biolabs) was used according to manufacturer's instructions, indicated below (Table 2-23). Each sample of cDNA was analysed in technical duplicates and in biological triplicates. For each primer set a no template control (NTC) was added to screen whether the reagents had any nucleic acid contamination or primer-dimer formation. For each sample, a no reverse-transcriptase (RT-) reaction was included to check for contaminating DNA. RT-qPCR reactions were carried out using the CFX Connect Real-Time System (Bio Rad) with the thermocycling conditions indicated below (Table 2-24). Data was analysed according to the $2^{-\Delta\Delta cq}$ method. Melt curves were included in the protocol to check the specificity of the amplified product.

Table 2-23 Reaction mix using the Luna Universal qPCR mix

Component	Volume (20 µL)
2X Luna qPCR mix	10 µL
nf H_2O	7 µL
Forward primer (10 µM)	1 µL
Reverse primer (10 µM)	1 µL
cDNA	1 µL

Table 2-24 Bacterial RT-qPCR thermocycling conditions (39 cycles)

Step	Temperature	Time
Initial Denaturation	95°C	3 min
Denaturation	95°C	10 sec
Annealing	60°C	30 sec
Extension	65°C	5 sec
Final Extension	95°C	1 min

2.4.6 Sanger Sequencing

Primers were designed around the sequence of interest using Primer3web. The sequence was amplified by PCR and purified using a the QIAquick PCR purification kit (Qiagen). Tubes containing the purified PCR product and primers were sent to Eurofins Genomics for Sanger sequencing. Returned FASTA files were opened with SnapGene Viewer Version 7.2.1 and alignment was performed using the EMBOSS Needle Pairwise Sequence Alignment tool (EMBL-EBI).

2.4.7 Whole genome sequencing (WGS)

Genomic DNA (gDNA) from EcN_AR and EcN_EO was extracted using the DNeasy Blood & Tissue Kit (Qiagen) and sent to MicrobesNG at the University of Birmingham, UK. The genome was sequenced using the Illumina Short Reads 2 x 250 bp Paired End sequencing platform. Computational analyses were performed on Galaxy Australia (Jalili *et al.*, 2020) using the EcN genome (NCBI accession number NZ_CP007799.1) for reference. The tool 'snippy' was used to identify nucleotide mutations compared to the reference genome, and the tools 'SPAdes', 'Prokka' and 'JBrowse' were used to respectively assemble, annotate, and visualise the sequenced genomes. GenBank files containing the sequenced *pks* region were uploaded on the Comparative Gene Cluster Analysis Toolbox (CAGECAT) and the clinker tool (van den Belt *et al.*, 2023) was used to align the colibactin biosynthetic gene clusters of EcN_EO and EcN_AR.

2.5 *In cellulo* infections, microscopy techniques, and DNA crosslinking assays

2.5.1 Maintaining HeLa and HT29 cells

HeLa cells were routinely cultured in T75 flasks in 10 mL of DMEM containing 10% foetal bovine serum (FBS), (Sigma-Aldrich) and 1% pen-strep solution. They were stored in an incubator set to 37°C, 5% CO_2. When reaching about 80% confluency, cells were washed three times with Dulbecco's phosphate buffered saline (PBS) solution (Sigma-Aldrich) and then trypsonised by adding 2 mL of Trypsin-EDTA 0.25% (Sigma-Aldrich) to the flask and incubating for 2 minutes in the incubator with the above-mentioned conditions. Cells were then resuspended in 8 mL pre-warmed DMEM media and either passed into a fresh flask - typically at 1:10 ratio of cells to fresh media - or seeded into plates for infection assays.

HT29 cells were routinely cultured in McCoy's 5A Medium containing 10% FBS, 1% pen-strep solution, and 1% L-Glutamine. They were trypsonised, passed, and seeded as described above.

2.5.2 Infection assays in HeLa or HT29 cells

Infection assays were performed in either 24-well plates or 6-well plates. Approximately 24 h prior to infection, 5 x 10^4 cells were seeded in each well when using 24-well plates, or 3 x 10^5 cells when using 6-well plates. Seeded cells were incubated overnight at 37°C, 5% CO_2. Infection assays were performed with EcN_AR, EcN_EO or DH10B to observe colibactin-driven megalocytosis and senescence, and to investigate whether AAs could reduce these processes. Bacterial cultures for infection were grown in MEM-HEPES with or without 1 mM of the AAs of interest at 37°C, 200 rpm shaking for 4.5 h. DMEM was aspirated from the seeded cells, which were washed with PBS before being replenished with MEM containing, if required, 1 mM AAs of interest. The OD_{600} of bacterial cultures was measured to perform cell infections: cells were infected with varying volumes of bacterial culture, calculated to achieve a multiplicity of infection (MOI) equal to 400. Plates were centrifuged at 200 x g for 1 minute and then incubated at 37°C, 5% CO_2 for 4 h. After inoculation, cells were washed with PBS, replenished

with DMEM containing 200 µg/mL gentamicin, and incubated for 72 h at 37°C, 5% CO_2.

2.5.3 Methylene blue staining and light microscopy

After incubation, cells were washed with PBS and fixed with 4% PFA for 20 min. Afterwards, cells were washed once with methylene blue wash buffer (Table 2-8), completely covered with methylene blue stain (Table 2-7), and then immediately rinsed several times with methylene blue wash buffer until the liquid came out clear. The well-plate was left to dry overnight or at 37°C for 1-2 h. Cell morphology was observed by light microscopy using the Evos FL Auto 2 microscope (Thermo Fisher Scientific). Images were acquired at 20X magnification. To quantify individual cell areas, a total of 5 images were captured in each well, in the centre, top, bottom, left and right section of the well. A pipeline was then developed on CellProfiler software (McQuin *et al.*, 2018) to measure cell areas. To assess how many cells survived following infection under the different AA regimes, the central 10% of the well area was measured using the Evos FL Auto 2 microscope (Thermo Fisher Scientific). A total of 181 snapshots were captured at 20X magnification and a pipeline developed on ImageJ was employed to calculate the percentage of the well covered by cells. Different replicates were used to calculate cell area and cell survival under the different AA conditions.

2.5.4 Fluorescent staining and microscopy

Fluorescent microscopy was only performed when cells were seeded onto glass coverslips. After incubation, cells were washed with PBS and fixed with 4% PFA for 20 min. Cells were permeabilised with 0.1% Triton X-100 for 5 minutes and, after being washed with PBS, each coverslip was stained with 5 units/mL of Alexa Fluor 555 Phalloidin (Thermo Fisher Scientific). Cells were incubated in the dark for 1 h on a rocker. They were then washed twice with PBS and coverslips were mounted onto sterile microscope slides with 4 µL of Vectashield mounting media with DAPI (Vector Laboratories). Cover slips were sealed with clear nail polish and

slides were left to dry overnight in the dark. Images were acquired at 10X magnification using the Zeiss Axio Imager M1 microscope (Zeiss) and processed using the Zen Blue Version 3.2 software. For long-term storage, slides were kept at 4°C.

2.5.5 Interstrand crosslink assays

A 50 mL culture of DH5α cells carrying the pUC19 plasmid was cultured overnight. The following day a maxiprep of pUC19 was performed using the Plasmid Maxi Kit (Qiagen) and the plasmid was linearised using the restriction enzyme BamHI, with the reaction indicated below (Table 2-25). Both the restriction enzyme and the buffer were purchased from New England Biolabs.

Table 2-25 Reaction mix to linearise pCU19

Component	Volume (total 50 µL)
nf H_2O	22 µL
CutSmart Buffer (10X)	5 µL
BamHI	3 µL
pUC19 DNA (450 ng/µL)	20 µL

EcN was cultured overnight as described above. The following morning, fresh M9 media was made and 9.5 mL of media were dispensed into a 50 mL falcon tube. If required, 1 mM of the AA of interested was added to the tube and the media was inoculated with 500 µL of overnight culture. The samples were then incubated for 1.5 h at 37°C, 200 rpm shaking and following incubation the $OD6_{00}$ was measured. Bacterial cultures were adjusted to contain 1.5 x 10^6 CFU in 100 µL of M9 +/- 1 mM AA and dispensed in 500 µL microcentrifuge tubes. Adjusted cultures were performed in duplicates so that, in parallel, one set of samples could be used for CFU counts to confirm appropriate cell density. Samples were placed at 37°C in a static incubator for 5 h. Afterwards, 450 ng/µL of linearised pUC19 and 0.2 µL of 0.5 M EDTA (Sigma-Aldrich) were added to each sample, and tubes were incubated

for 40 min at 37°C in a static incubator. Tubes were then centrifuged at 17000 x g (Heraeus Pico 17 Centrifuge, Thermo Fisher Scientific) for 2 mins. The supernatant was transferred to a new tube and purified using the PCR purification kit (Qiagen), while the bacterial pellet was discarded. DNA concentration was measured using the DS-11 FX spectrophotometer (DeNovix) and samples were stored at -20°C overnight. In the meantime, a 1% agarose gel was prepared using the alkaline gel solution (Table 2-9) and, once set, soaked overnight in alkaline gel running buffer (Table 2-10). The following morning, 4.5 µL of purified pUC19 DNA were thoroughly combined with 1.5 µL of loading dye and loaded onto the denaturing agarose gel alongside the 1 kB plus DNA ladder (Invitrogen). The gel was run initially at 25V for 45 min and then 50V for 150 min. It was afterwards soaked in neutralisation buffer (Table 2-11) for 30 min on a rocker and, once neutralisation had lapsed, it was visualised using a UV transilluminator (Azure biosystems). The percentage of crosslinked DNA was calculated relative to the total DNA content in each lane. It was calculated by densitometry using ImageJ and statistical analyses were performed in GraphPad Prism Version 9.2.0.

2.5.6 Single cell gene expression assay

EcN_AR *pclbB:gfp* was cultured overnight in 5 mL of LB supplemented with chloramphenicol (25 µg/mL). The following morning, MEM-HEPES supplemented with chloramphenicol (25 µg/mL) was inoculated with overnight culture and grown to an OD_{600} of 0.4 at 37°C, 200 rpm shaking. A total of 1 mL was withdrawn from the culture and centrifuged at 17000 x g (Heraeus Pico 17 Centrifuge, Thermo Fisher Scientific) for 1 min. The supernatant was discarded and bacterial cells were resuspended in 400 µL of PBS. Afterwards, 400 µL of 4% PFA were added to the tube, which was thoroughly vortexed to break up bacterial aggregates. A total of 25 µL of culture were placed on a welled, microscope slide and left to dry on a pre-warmed 37°C heat-block. Once dry, 25 µL of the *E. coli* O-6 agglutinating antiserum (Mast Group) – made up at a ratio of 1:100 in PBS – were added to the well, and the slide was incubated for 1 h in a humidity chamber placed on a rocker. After the hour had passed, the well was washed three times with PBS and 25 µL of the secondary antibody - the Goat anti-Rabbit IgG, Alexa Fluor™ 555 (Invitrogen) antibody made up at a ratio of 1:400 in PBS – were added to each well. The slide

was once again incubated for 1 h in a humidity chamber placed on a rocker. Following incubation, the well was washed three times with PBS and a coverslip was mounted on the slide using clear mounting media. The slide was left to dry overnight and imaged the following day. Images were acquired at 100X magnification using the Zeiss Axio Imager M1 microscope (Zeiss) and processed using the Zen Blue Version 3.2 software. EcN_AR *pclbB:gfp* fluoresced at 488 nm (green) and the bacterial surface fluoresced at 555 nm (red). Green and red channels were used both individually and merged to capture the same cells and check colocalisation of individual bacteria and their *clbB* expression. For long-term storage, slides were kept at 4°C.

2.6 *In vivo* experiments

2.6.1 Home Office animal licence

All procedures were carried out in strict accordance with the project licence number PP9552821, granted by the Home Office on 24 January 2022 and held by Dr. Gillian Douce.

2.6.2 Mouse maintenance

C57BL/6 mice were purchased from Charles River Laboratories aged between 6 and 8 weeks old. They were housed in groups of 3 or 4 at a controlled room temperature (22°C) with a 12 h light/dark cycle. Mice had *ad libitum* access to water and sterile food pellets and cages were changed at least once per week. Health status and weight checks were conducted and recorded daily. Mice were culled by exposure to carbon dioxide gas, followed by cervical dislocation for confirmation of death.

2.6.3 Infection of C57BL/6 mice with EcN in an AOM/DSS background

Azoxymenthane (AOM) was purchased from Sigma-Aldrich and made up according to manufacturer's instructions. On the first day of the experiment, all mice received an AOM intraperitoneal (IP) injection corresponding to 10 mg/kg of their body weight. IP injections of AOM were performed under mild mouse anaesthesia by Dr. Gillian Douce. For the first three days mice were also administered streptomycin (2 mg/mL) in their drinking water to reduce diversity of the gut microbiota. On the fourth day, mice belonging to infection groups received by oral gavage 10^9 CFUs of EcN WT or EcN Δ*clbB* in a 200 µL volume. Bacterial strains for mice infections were grown overnight in 30 mL of LB; the following day they were centrifuged at 3273 x g and resuspended in 3 mL of PBS, with CFU counts performed to confirm appropriate dosage. Mice who did not belong to infection groups received an oral gavage of 200 µL of PBS. Between days 5 to 9 mice were

maintained on a regimen of regular food and drinking water. On day 10, all mice were placed on 1.5% or 2% Dextran sulfate sodium (DSS) treatment. DSS (Mw ca 40,000) was purchased from Thermo Fisher Scientific and made up fresh on first day of administration. Mice were maintained on a 5-7 day cycle of DSS, and, in the last three animal experiments, a fresh solution of DSS was replaced mid-cycle to encourage colitis. For the following two weeks, mice were maintained once again on regular food and water before being placed under a second 1.5% or 2% DSS cycle, lasting 5-7 days. Following the second DSS cycles, mice were administered regular food and drinking water for 4-5 weeks, before being culled and dissected. Throughout the duration of the entire experiment, health records and weight checks were conducted daily. When mice lost more than 10% of their body weight, they were placed on a 'soft food' diet consisting of food pellets that had been softened by soaking in water. Excess water was removed and softened food pellets were placed on a petri dish at the bottom of the cage. When mice lost more than 20% of their body weight they were culled, as described above. Mice who were under AA treatment received fresh, daily replacement of 1 mM AA. When under administration of DSS, the AA was not replaced daily, but upon substitution of DSS.

2.6.4 Faecal shedding counts of infected mice

Faecal pellets were collected on alternate days, and homogenised at a ratio of 0.1 g of stool per 1 ml of PBS by softening and vortexing to allow release of bacteria. The homogenised solution was serially diluted at a ratio of 1:10, and, for each sample, 20 µl of each dilution were spotted on solid agar plates containing 100 µg/ml streptomycin. The plates were allowed to dry and incubated overnight at 30°C. Single colonies were counted within each spot, and multiplied by the relevant dilution factor to calculate CFU/ml. Serial dilutions were performed in technical replicates for each sample and the CFU/ml was determined as the average number of colonies at a given dilution across all samples within a group

2.6.5 Tissue collection and histology

Ralwax plates were prepared for pinning down the colons during dissections. Pellets of paraffin wax (Thermo Fisher Scientific) were melted in a water bath using a graduated glass beaker. Once all the wax had melted, 10% mineral oil (Sigma-Aldrich) was incorporated into the wax. The mixture was quickly poured into 150 mm dishes (Corning) and allowed to set. During dissection the colon was removed, washed thoroughly with PBS, cut longitudinally and pinned laterally onto ralwax plates. Colons were then submerged with 10% buffered formalin for approximately 24 h. Following fixation macroscopic analyses were performed on the colon, by counting the number, size and distribution of the polyps along the colon length. Swiss rolls were prepared and sent either to the Cancer Research UK (CRUK) Scotland Institute, or to the University of Glasgow School of Biodiversity, One Health, and Veterinary Medicine for embedding, sectioning, and staining with hematoxylin and eosin (H&E). Expert advice on colon pathology was provided by Christopher Bigley and Dr. Angelika Frances Rupp. Blind scoring of colonic tissues for inflammation and neoplasia was performed by Dr. Angelika Frances Rupp using the grading scale provided in the appendix (page 261).

2.6.6 Mouse microbiome analyses

Genomic DNA from the collected stool samples was extracted using the DNeasy PowerSoil Pro Kit (Qiagen) and DNA concentration was measured using the DS-11 FX spectrophotometer (DeNovix). Samples were sent to Novogene Europe for 16S amplicon metagenomic sequencing by amplification of the V3-V4 region at a sequencing depth of 30000 raw tags. Novogene Europe provided both raw data and analysed outputs, based on the comparisons requested by the researcher during sample submission.

2.6.7 Total RNA extraction from colonic mouse tissues

During dissection a 1 cm section of the upper colon, directly below the caecum, was collected, submerged in RNAlater (Qiagen), and placed on ice. Tissues were

kept at -20°C until required for RNA extractions. Upon extraction, a piece of tissue was transferred to microcentrifuge tube, containing a sterile 2 mm stainless steel bead (Qiagen). To each tube 750 µL of Trizol (Invitrogen) were added and the tissue was homogenised using the TissueLyser II (Qiagen) at 25 Hz for 2 min. The tubes were then centrifuged at 12,000 x g for 5 min at 4°C. A total of 700 µL of the supernatant were withdrawn and transferred to a new microcentrifuge tube to which 140 µL of chlorophorm (Sigma-Aldrich) were added. Tubes were shaken until cloudy and then left to sit at room temperature for 3 min. Afterwards, they were centrifuged at 12,000 x g for 15 min at 4°C. The aqueous layer (approximately 350 µL) was withdrawn and transferred to a new microcentrifuge tube to which 100% ethanol was added. The amount of ethanol added corresponded to 1.5X the volume of aqueous layer withdrawn (usually 525 µL of ethanol). From this step onwards, the RNeasy Mini Kit (Qiagen) was used to perform the rest of the extraction, as per manufacturer's instructions. The DNase digestion was performed on-column using the RNase-Free DNase Set (Qiagen): a ttotal of 10 µL of DNAse stock solution and 70 µL of Buffer RDD were added to each sample, and tubes were incubated at room temperature for 15 min. The washing steps indicated in the RNeasy Mini Kit (Qiagen) were followed and the RNA was eluted in 30 µL of nf H_2O. The concentration of samples was measured using the DS-11 FX spectrophotometer (DeNovix).

2.6.8 RT-qPCR of RNA extracted from colonic mouse tissues

The extracted RNA was diluted with nf H_2O to a concentration of 250 ng in a 15 µL volume. RNA was converted to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). A cDNA synthesis reaction mix was prepared as per manufacturer's instructions, indicated in the table below (Table 2-26). Afterwards, 10 µL of diluted RNA and 10 µL of the cDNA synthesis reaction mix were combined in a PCR tube to yield a total final volume of 20 µL. To execute conversion to cDNA, the samples were then subjected to thermocycling under the parameters indicated below (Table 2-27).

Table 2-26 cDNA synthesis reaction mix using the High-Capacity cDNA Reverse Transcription Kit

Component	Volume (10 µL)
10X RT Buffer	2 µL
25X dNTP Mix (100 mM)	0.8 µL
10X RT Random Primers	2 µL
MultiScribe RT	1 µL
nf H_2O	4.2 µL

Table 2-27 Thermocycling conditions for cDNA synthesis from mouse RNA (1 cycle)

Step	Temperature	Time
Annealing	25°C	10 min
cDNA synthesis	37°C	120 min
Heat inactivation	85°C	5 min
Hold	4°C	Hold

Once cDNA synthesis was complete, samples were further diluted 1:10 with nf H_2O (20 µL of cDNA and 180 µL of nf H_2O). These samples were stored at -20°C until execution of RT-qPCRs. For RT-qPCR the PowerUp SYBR Green Master Mix for qPCR (Thermo Fisher Scientific) was used according to manufacturer's instructions, indicated below (Table 2-28). Each sample of cDNA was analysed in technical duplicates and in biological quadruplicates. For each primer set a no template control (NTC) was added to screen whether the reagents had any nucleic acid contamination or primer-dimer formation. Given the large number of samples, the RT-qPCR reactions were carried out using 384-well plates and readings were taken with the QuantStudio 7 Flex Real-Time PCR machine. The thermocycling conditions are indicated below (Table 2-29). Data was analysed according to the $2^{-\Delta\Delta cq}$ method. Melt curves were included in the protocol to check the specificity of the amplified product.

Table 2-28 Reaction mix using the PowerUp SYBR Green Master Mix for qPCR

Component	Volume (10 µL)
2X PowerUp SYBR Green Master Mix for qPCR	5 µL
Forward primer (10 µM)	0.5 µL
Reverse primer (10 µM)	0.5 µL
cDNA (diluted)	4 µL

Table 2-29 Mouse RT-qPCR thermocycling conditions (40 cycles)

Step	Temperature	Time
UDG Activation	50°C	2 min
Dual-Lock DNA polymerase	95°C	2 min
Denature	95°C	15 sec
Anneal/Extend	60°C	1 min

2.7 Statistical analyses

Processing of raw data was done in Microsoft Excel Version 16.77.1. Statistical analyses were carried out using GraphPad Prism Version 9.2.0. Statistical analyses for two samples were performed using a Student's unpaired T-test, analyses between three samples were performed using an ordinary one-way ANOVA, and comparisons of multiple samples to one control condition were performed using a Dunnett's multiple comparison test. Statistical significance of non-parametric data, such as that derived from histology scoring, was analysed using a Mann-Whitney test. Significance levels are indicated in figures as ns, *, **, ***, **** indicating respectively not significant, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, and $p \leq 0.0001$.

3 Screening Amino Acids for their Modulatory Effect on Bacterial Growth and Colibactin Transcription

3.1 Introduction

Extensive studies have shown the extraordinary effects that the presence of certain AAs can have on a pathogen's physiological functions and virulence factors. This is unsurprising as a pathogen's preferred nutritional substrates reflect its distinctive physiological needs and subsequent tissue preferences within the host. Studies conducted in Gram-positive bacteria have shown that branched chain amino acids (BCAA) are linked to promoting pathogen virulence. In *Staphylococcus aureus*, the BCAA transporter BrnQ1 is required for transporting leucine and valine into the bacterial cell and both BCAA transporters BrnQ1 and BcaP are required for *S. aureus* virulence *in vivo* (Kaiser *et al.*, 2016). For intracellular pathogens the need to compete against hosts for AAs is even stronger: a BCAA auxotroph like *Legionella pneumophila* requires PhtJ as a valine transporter to carry out its intracellular replication. However, AA starvation can be as impactful as AA availability in activating and enhancing virulence. For example, isoleucine and valine act as co-repressors by increasing the affinity of CodY, a negative regulator of toxin gene expression, to its target DNA in *Clostridioides difficile*, demonstrating a link between AA limitation and toxin production (Shivers and Sonenshein, 2004), (Dineen *et al.*, 2007).

D-AAs are of particular interest because of their role in inhibition of biofilm formation and quorum sensing among certain bacteria. For example, a study showed that D-Phenylalanine, D-Proline, and D-Tyrosine were all effective in disrupting biofilm formation in *S. aureus*, without affecting bacterial growth in shaking cultures. The authors observed that there was a lack of protein localisation on bacterial cell surface in D-AA treated cultures and hypothesised that D-AAs may prevent localisation of adhesion proteins, disrupting bacterial clustering and inhibiting biofilm development (Hochbaum *et al.*, 2011). Another study showed that synthetic peptides designed with D-Glutamic acid were particularly effective at inhibiting PlcR regulon activity, and consequently disrupting quorum sensing in *Bacillus cereus* (Defoirdt *et al.*, 2019). Certain D-AAs are also found in natural and synthetic peptides with strong antimicrobial properties: D-Valine is found in penicillin; D-Alanine, D-Leucine and D-Valine are found in actinomycin, gramicidin and valinomycin; D-Aspartic acid and D-Glutamic acid are found in mycobacillin (Genchi, 2017).

The Roe Group at the University of Glasgow has a longstanding interest in the effects of AA – particularly D-AAs – on *E. coli* virulence factors. Previous research in the group has shown that D-Serine activates the SOS response and inhibits expression of the Type 3 Secretion System (T3SS) in EHEC O157:H7 (Connolly *et al.*, 2015). Interestingly, the group also showed that the modulatory effects of D-Serine on different *E. coli* pathotypes result in distinct transcriptomes: research conducted on EHEC, UPEC and NMEC – which respectively colonise the gut, bladder, and brain – showed that gene expression affected by D-Serine was regulatory for transport, metabolism, and other functional categories, but there was no single differentially expressed gene common to all three strains (Connolly *et al.*, 2021). Previous research from the Roe Group, also showed that in M9 minimal media, several AAs – particularly D-Serine – can modulate colibactin expression (Hallam *et al.*, 2023).

Of the over 500 AAs existing in nature, only 20 are incorporated into proteins during translation and are known as proteinogenic (Ahluwalia, Kumar and Kumar, 2022). It was decided that all proteinogenic AAs should be screened for their capacity to downregulate colibactin expression as these are naturally acquired through diet. Although food proteins are typically made up of L-AAs, both L- and D- isomers were tested because D-AAs can be found in some foods, acquired through processes of racemisation and obtained as by-products of the microbiota (Genchi, 2017).

Consequently, given that the overarching objective of this PhD project was to find AAs capable of inhibiting colibactin-driven CRC, the aims of this chapter were to: (1) assess if repression was limited to D-Serine, or whether other L- and D- forms of proteinogenic AAs modulated colibactin transcription and (2) identify which AAs offer potential as novel therapeutics *in vivo*.

3.2 Results

3.2.1 Optimisation of concentration at which to test AAs

Previous research from the Roe Group reported that D-Serine has a repressive effect on colibactin transcription when used at a 1 mM concentration (Hallam *et al.*, 2023). This concentration was originally chosen as it corresponds to the upper concentration range detected in the human body. More specifically, D-Serine has been measured at a concentration of up to 1 mM in urine (Anfora *et al.*, 2007) and of up to 107 nM/g in the brain (Nagata *et al.*, 1995). No relevant literature on the concentration of D-Serine in the human intestine has been found (Nishikawa, 2011), but – since *pks*+ *E. coli* have been isolated from both the urinary tract and faecal samples – 1 mM was deemed an appropriate concentration to use *in vitro*. However, it had not been reported whether the effect of D-Serine on colibactin transcription was concentration-dependent. The following experiment was carried out to first determine whether D-Serine's repressive activity is concentration-dependent, and subsequently to use this information to select the appropriate concentration at which to test all other AAs. Given that previous work from the Roe Group had tested the effect of D-Serine on the UPEC strain CFT073, initial studies focused on the same strain across a wider range of D-Serine concentrations. This allowed repetition and extension of previous observations from the group. The majority of future experiments were performed with the commensal intestinal strain EcN, selected with the ultimate aim of investigating the effects of colibactin modulation on CRC *in vivo*. The Roe Group had a stock of EcN, referred to as EcN_AR. Later on, a second stock of the EcN strain was acquired from Eric Oswald's (EO) Group at the Institut de Recherche en Santé Digestive in Toulouse, France, referred to as EcN_EO. Experiments performed within both the Roe and Oswald laboratories suggested that EcN_EO expressed more colibactin than EcN_AR. This is explained in greater detail in Chapter 5.

Colibactin transcription was measured indirectly using a *clbB* promoter:GFP reporter fusion construct, where *clbB* is an essential gene for colibactin biosynthesis by the *pks* island. The reporter fusion construct is referred to as *pclbB:gfp*, and all reporter assays mentioned in this chapter were performed on strains transformed with this plasmid. A map of the plasmid can be found in the

appendix (Figure 10-1). To determine the impact of different concentrations of D-Serine, a 5-fold serial dilution of D-Serine was tested from 1 M to 0.512 µM in MEM-HEPES. Conditions with no D-Serine supplementation and with a concentration of 1 mM were added for comparison. This work showed that a significantly inhibitory effect on colibactin transcription is observed when the medium is supplemented with 0.32 mM D-Serine (0.52-fold decrease, $p = 0.0046$), with an even stronger repression at 1 mM (0.44-fold decrease, $p = 0.0013$). At lower concentrations, ranging between 0.512 µM and 64 µM, D-Serine fails to produce significant repression, whereas at higher concentrations, ranging between 1.6 mM and 0.04 M, *clbB* transcription is comparable to that observed at 1 mM (Figure 3-1A). At concentrations ranging between 1 mM and 0.04 M, D-Serine appears to boost CFT073's growth (Figure 3-1B). At highest concentrations (0.2 M and 1 M) D-Serine appears toxic, limiting bacterial growth (Figure 3-1B). These data taken together provided a strong rationale for using D-Serine at a concentration of 1 mM and support the decision to test other AAs at this concentration.

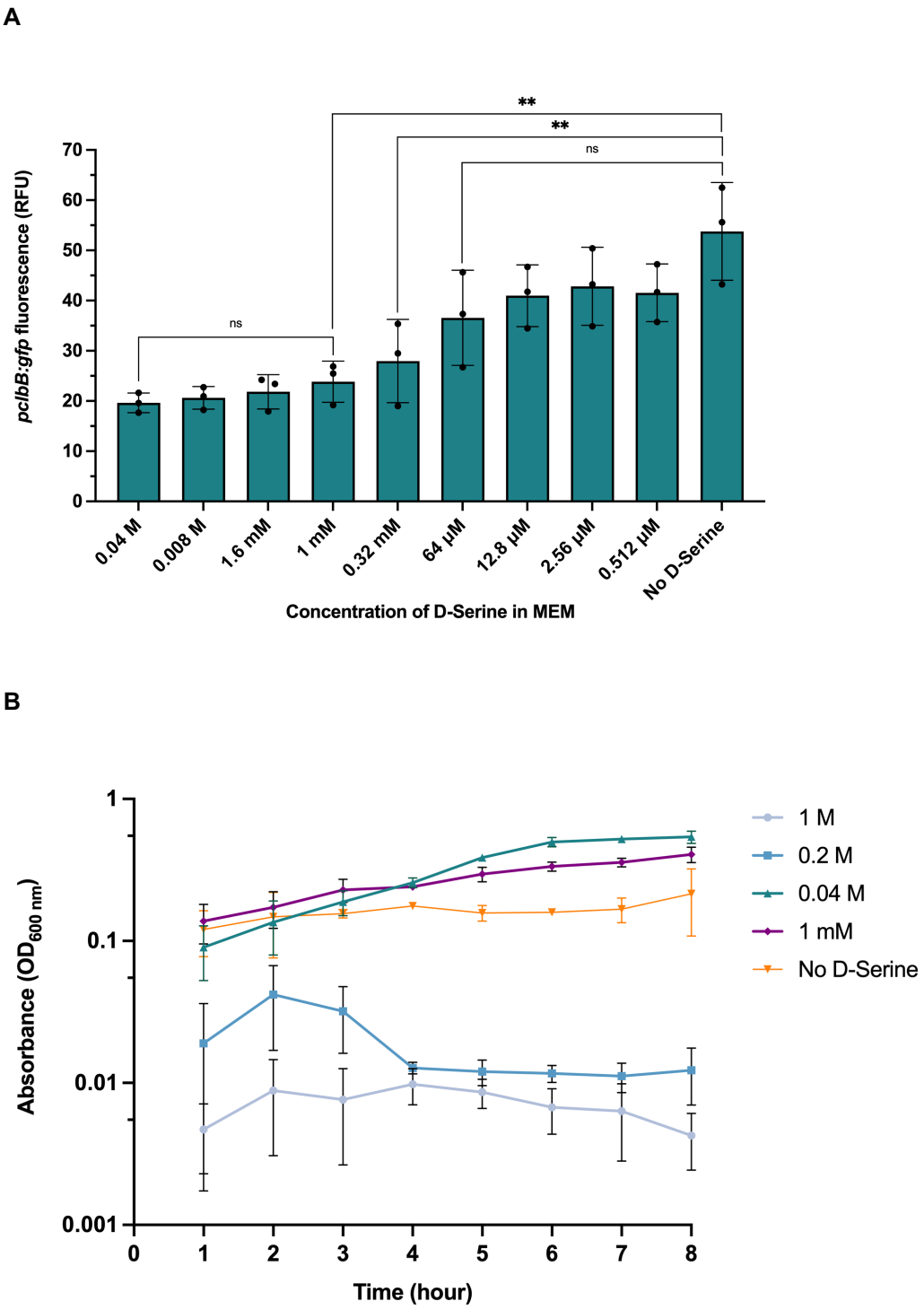


Figure 3-1 Testing varying concentrations of D-Serine on CFT073's growth and *clbB* promoter expression. (**A**) The *pclbB:gfp* reporter activity of CFT073 in MEM-HEPES supplemented with increasing concentrations of D-Serine. Fluorescence was measured hourly over an 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$ in exponential phase, at 7 h from bacteria cultured in a 96-well plate at 37°C, 200 rpm. Columns represent means +/- standard deviation calculated from three replicates. Statistical significance was assessed by a Dunnett's multiple comparison test with ** indicating significance, $p < 0.01$. (**B**). Growth curves of CFT073 in MEM-HEPES alone or supplemented with D-Serine at 1M, 0.2 M, 0.04 M, and 1 mM. CFT073 was cultured in MEM-HEPES supplemented with increasing concentrations of D-Serine in a 96-well plate at 37°C, 200 rpm. Absorbance ($OD_{600\ nm}$) was measured hourly for 8 h and values were plotted on a log-10 scale. Data points represent average absorbance values and error bars indicate standard deviation calculated from three replicates.

3.2.2 Modulation of colibactin transcription by L- and D-AAs in varied media

Proteinogenic AAs alone were selected for screening because they are the only ones incorporated into proteins during translation and are acquired naturally through diet. Of the known 22 proteinogenic AAs, all 20 encoded by the standard genetic code were selected for the capacity to inhibit colibactin production. Selenocysteine (which has the same structure as cysteine, but with replacement of the sulphur atom with a selenium one) was also considered because it can be incorporated into proteins through translational recoding. Pyrrolisine was excluded from study as, although it can be used for the biosynthesis of proteins in some bacteria and archaea, it is absent in humans. Given pre-existing evidence of the antivirulence potential of D-AAs, both L- and D- isomers of the selected proteinogenic AAs were screened. Furthermore, it was decided that AAs would be screened in different types of media. The rationale being that the gut is a particularly metabolically-dynamic environment, hence, if an AA shows inhibitory activity in multiple types of media, it is also more likely to do so in the gut. In order to comprehensively assess inhibitory activity of specific AAs, culture media containing different nutritional compositions, and used for different purposes were selected, including: LB (relatively rich growth medium used for routine-culturing of *E. coli*), M9 (a commonly chemically-defined growth medium), MEM-HEPES (minimum essential medium used in the culture of eukaryotic cells), and DMEM (Dulbecco's modified eagle medium, also used as a basal medium for growth of eukaryotic cells). Standard media for cell culture was selected to reflect the longer term intention to test inhibitory AAs in eukaryotic cell infections. Each AA was tested at a final concentration of 1 mM, providing consistency with research previously performed in the Roe Group and following testing of D-Serine concentrations as described above.

To determine the impact of these different AAs on colibactin expression, EcN_AR was cultured in media +/- 1 mM AA in a 96-well plate, incubated at 37 °C in shaking conditions. To understand the impact of supplementation on growth specifically, absorbance ($OD_{600\ nm}$) of each well was measured hourly and the specific growth rate ($\mu = \Delta \ln OD_{600nm}/\Delta t$) calculated. To measure the impact of AAs on colibactin expression, EcN_AR was transformed with the transcriptional reporter plasmid *pclbB:gfp*, as described above, allowing indirect assessment through measurement

of fluorescence. Fluorescence units were divided by OD_{600} in order to calculate relative fluorescence units (RFUs), normalising transcriptional activity to growth. Both specific growth rates and RFUs were normalised to a control well in which there was no AA supplementation in the growth medium. The control threshold in specific growth rate calculations was standardised to 1, and in RFU calculations it is indicated by 100. Values above and below these thresholds respectively indicated upregulation and downregulation of EcN_AR growth and *clbB* promoter activity.

Supplementation with the majority of AAs in LB largely resulted in comparable or decreased growth rates, relative to non-supplemented controls. Seleno-L-Cysteine appeared to have the greatest inhibitory effect on growth rate (0.13-fold decrease), as did L- and D-Tyrosine (0.59-fold decrease, 0.55-fold decrease respectively). Interestingly, while L- and D- Tyrosine were the only AAs to downregulate promoter activity of *clbB* in this growth medium – by 17.7% and 12.6% respectively – supplementation with Seleno-L-Cysteine appears to enhance transcription with a 411.5% increase in RFU (Figure 3-2). However, this is likely to reflect the failure of the strain to grow biasing RFU calculation, rather than being a true reflection of *clbB* promoter activity. As RFU was calculated as *pclbB:gfp*/$OD_{600\ nm}$, division with a smaller number, leads to the inflated RFU observed.

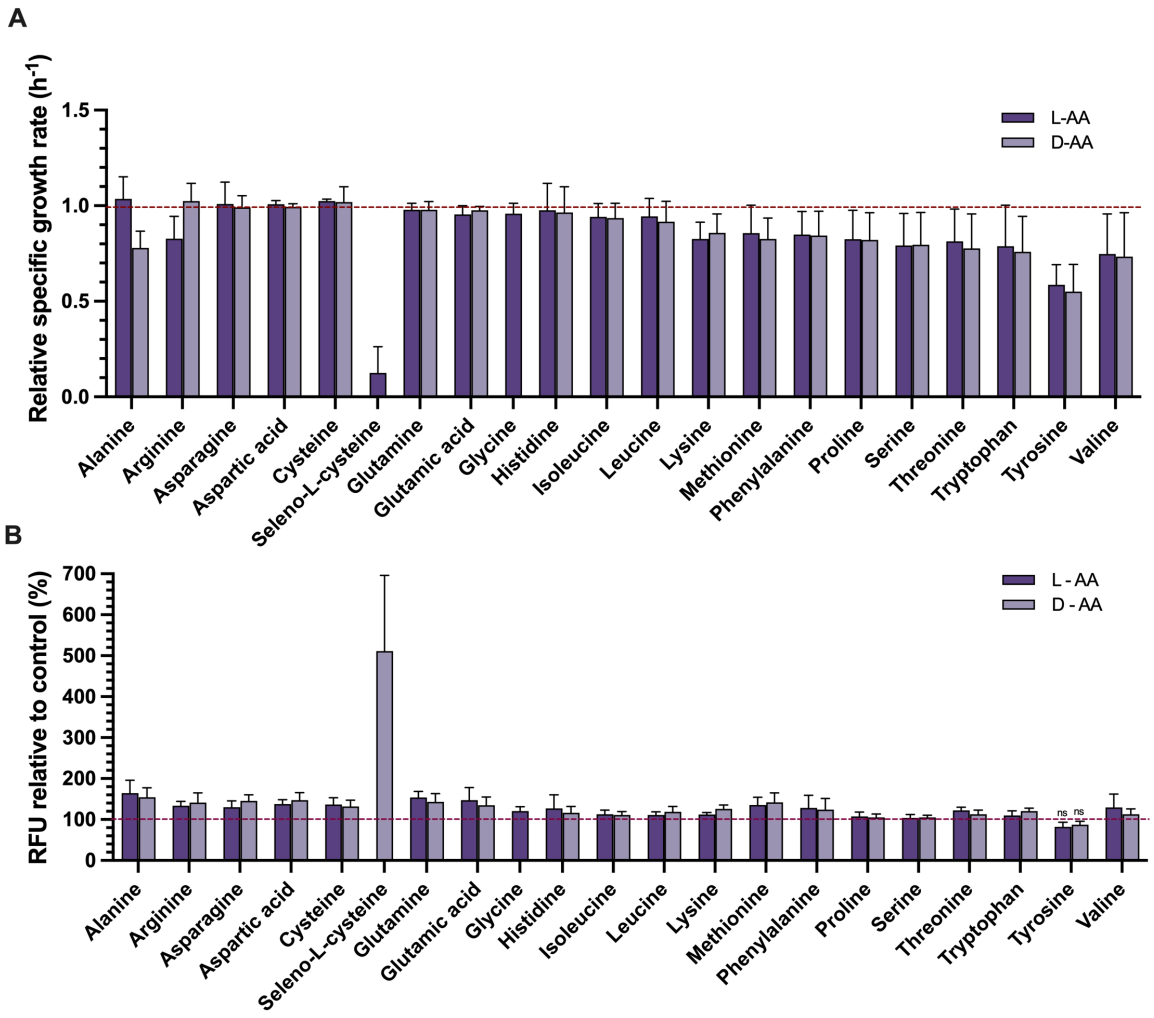


Figure 3-2 Modulatory effect of L- and D- AAs on EcN_AR growth and *clbB* promoter expression in LB. (**A**) The specific growth rates of EcN_AR in LB in the presence of 1 mM amino acids. Bacteria were cultured in LB +/- 1 mM amino acid in a 96-well plate at 37°C, 200 rpm. Absorbance ($OD_{600\ nm}$) was measured hourly and the specific growth rate ($\mu = \Delta \ln OD_{600nm}/\Delta t$) was calculated during exponential growth. The data is represented relative to the control condition (no amino acid supplementation), which is represented by the dashed red line. Columns represent means and error bars indicated standard deviation, calculated from three independent biological replicates. (**B**) EcN_AR *pclbB:gfp* reporter activity at exponential phase when grown in LB supplemented with 1 mM amino acids. Fluorescence was measured hourly over an 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$ in exponential phase, at 7 h in bacteria cultured in a 96-well plate at 37°C, 200 rpm. Columns represent means +/- standard deviation calculated from three independent biological replicates. Statistical significance was assessed for amino acids able to repress transcription by a Dunnett's multiple comparison test with ns, *, **, ***, **** indicating no significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively. The dashed line represents expression in control conditions, without any amino acid supplementation.

In M9, supplementation with the majority of AAs (twenty-nine out of forty of the L- and D-AAs) led to growth rates that were higher or comparable to those of the standard control growth conditions. Growth was again strongly inhibited by addition of Seleno-L-Cysteine (0.02-fold decrease), while upregulating *clbB* expression by 68.1%. This upregulation is again likely to be a calculation-based artifact, as described above. Eleven AAs caused a downregulation of *clbB* promoter activity when added to M9, but none showed statistical significance. Here, there was some overlap between the AAs that decreased both growth and *clbB* expression, with L-Leucine, D-Serine, and D-Tyrosine downregulating *clbB* promoter activity by 8.9%, 10.9% and 20.4%, while simultaneously showing decreasing growth of 0.94-, 0.94-, and 0.80-fold (Figure 3 -3).

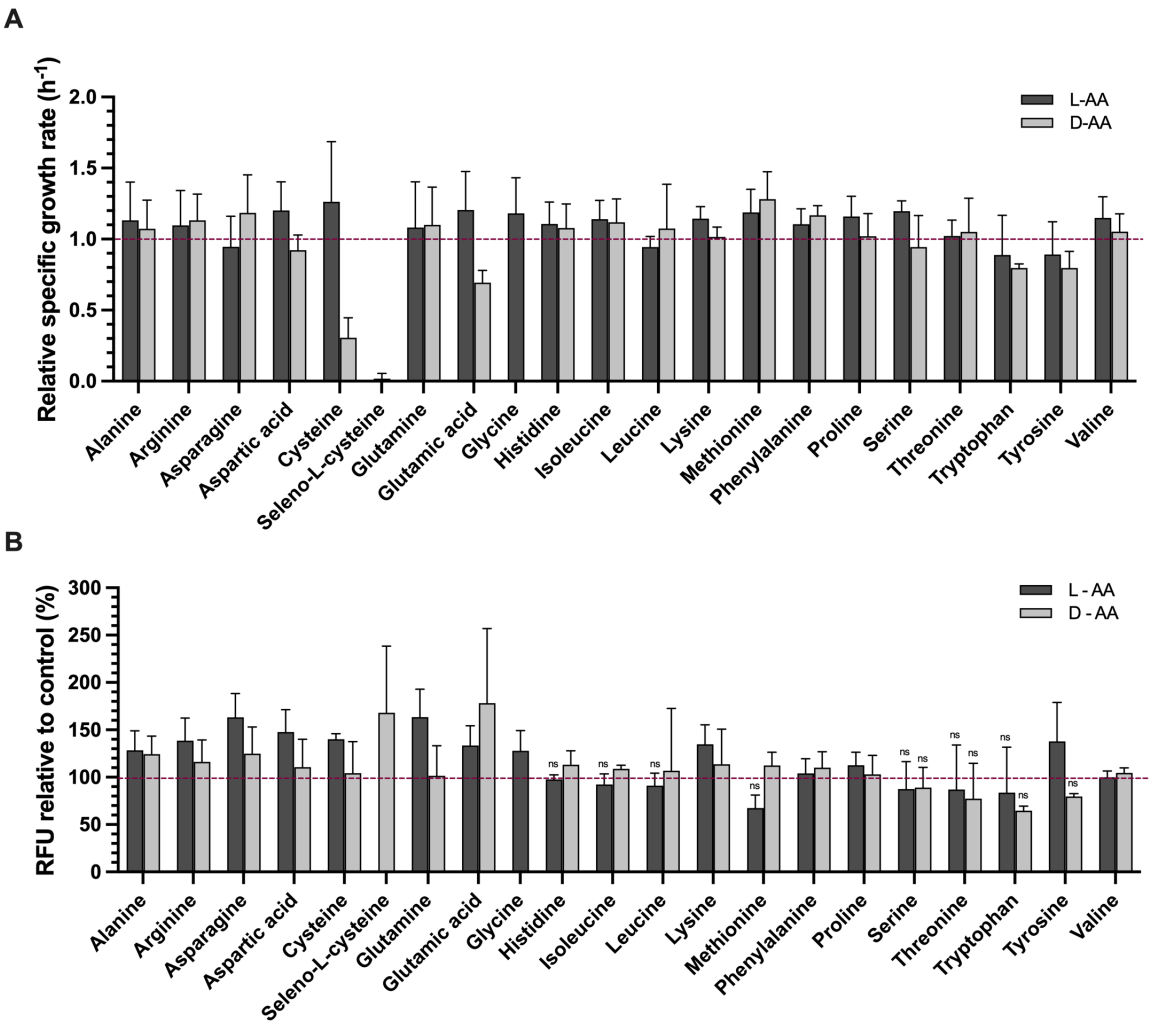


Figure 3-3 Modulatory effect of L- and D- AAs on EcN_AR growth and *clbB* promoter expression in M9. (**A**) The specific growth rates of EcN_AR in M9 in the presence of 1 mM amino acids. Bacteria were cultured in M9 +/- 1 mM amino acid in a 96-well plate at 37°C, 200 rpm. Absorbance ($OD_{600\ nm}$) was measured hourly and the specific growth rate ($\mu = \Delta \ln OD_{600nm}/\Delta t$) was calculated during exponential growth. The data is represented relative to the control condition (no amino acid supplementation), which is represented by the dashed red line. Columns represent means and error bars indicated standard deviation, calculated from three independent biological replicates. (**B**) EcN_AR *pclbB:gfp* reporter activity at exponential phase when grown in M9 supplemented with 1 mM amino acids. Fluorescence was measured hourly over an 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$ in exponential phase, at 7 h in bacteria cultured in a 96-well plate at 37°C, 200 rpm. Columns represent means +/- standard deviation calculated from three independent biological replicates. Statistical significance was assessed for amino acids able to repress transcription by a Dunnett's multiple comparison test with ns *, **, ***, **** indicating no significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively. The dashed line represents expression in control conditions, without any amino acid supplementation.

In MEM-HEPES, unlike M9 and similarly to LB, supplementation with the vast majority of AAs (thirty-three of the forty L- and D-AAs) led to decreased growth rates compared to control conditions. L-Serine and D-Serine enhanced EcN_AR's growth by 1.46-fold and 2.06-fold respectively, while still downregulating *clbB* promoter activity by 5.4% and 19.89%. The inhibitory effect of Seleno-L-Cysteine was less stark compared to that seen in LB and M9, with a 0.5-fold decrease in growth. In this instance, Seleno-L-Cysteine also appeared to significantly repress *clbB* promoter transcription by 60.3% ($p < 0.0001$). Given that growth inhibition triggered by Seleno-L-Cysteine is less severe, the calculated RFU value is considered more accurate. A total of 14 AAs downregulated transcription of the *clbB* promoter, including L-Asparagine, which was the only other AA in which this 38.9% change was statistically significant ($p = 0.002$); (Figure 3-4).

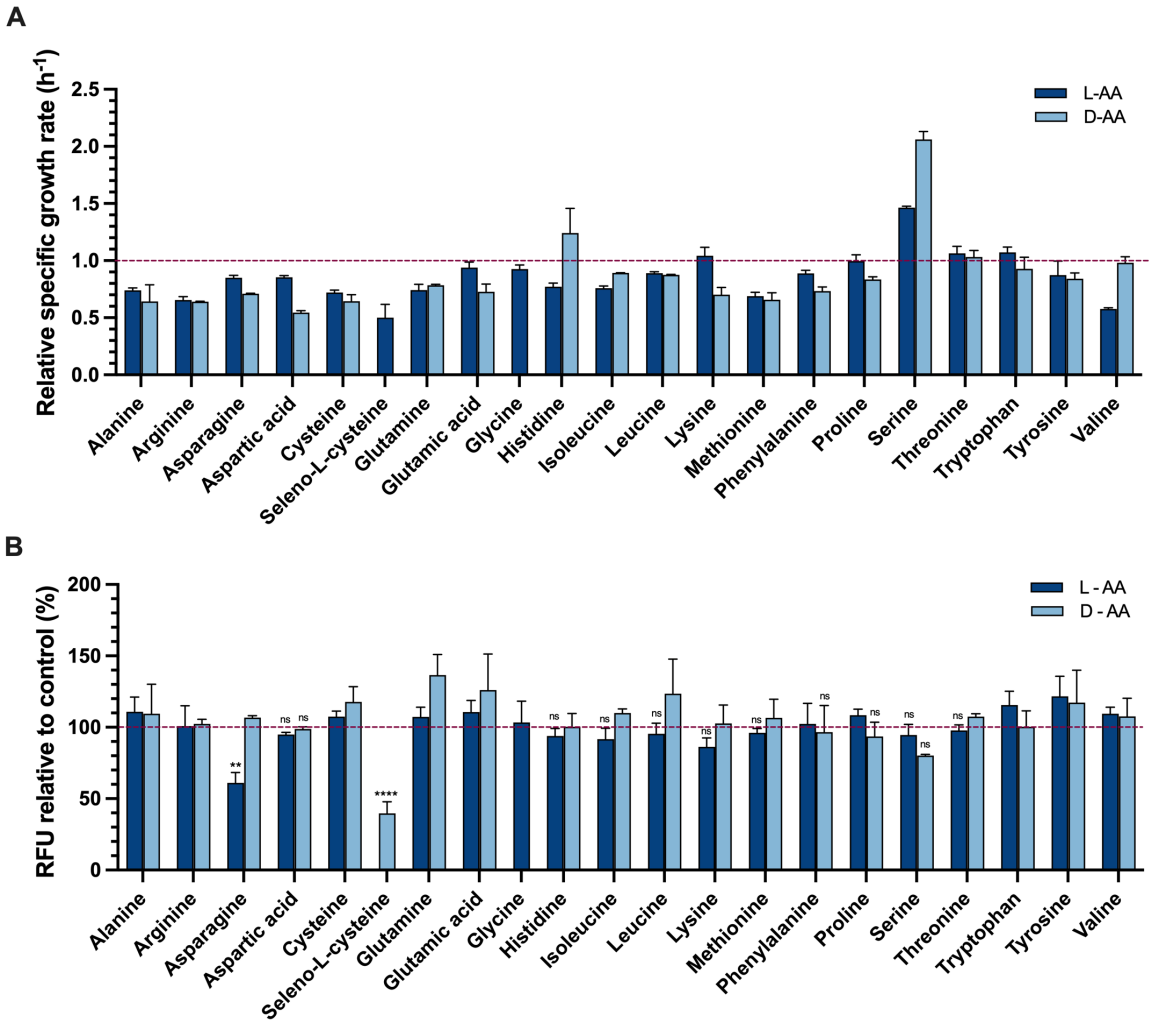


Figure 3-4 Modulatory effect of L- and D- AAs on EcN_AR growth and *clbB* promoter expression in MEM-HEPES. (**A**) The specific growth rates of EcN_AR in MEM-HEPES in the presence of 1 mM amino acids. Bacteria were cultured in MEM-HEPES +/- 1 mM amino acid in a 96-well plate at 37°C, 200 rpm. Absorbance ($OD_{600\ nm}$) was measured hourly and the specific growth rate ($\mu = \Delta \ln OD_{600nm}/\Delta t$) was calculated during exponential growth. The data is represented relative to the control condition (no amino acid supplementation), which is represented by the dashed red line. Columns represent means and error bars indicated standard deviation, calculated from three independent biological replicates. (**B**) EcN_AR *pclbB:gfp* reporter activity at exponential phase when grown in MEM-HEPES supplemented with 1 mM amino acids. Fluorescence was measured hourly over an 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$ in exponential phase, at 7 h in bacteria cultured in a 96-well plate at 37°C, 200 rpm. Columns represent means +/- standard deviation calculated from three independent biological replicates. Statistical significance was assessed for amino acids able to repress transcription by a Dunnett's multiple comparison test with ns, *, **, ***, **** indicating no significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively. The dashed line represents expression in control conditions, without any amino acid supplementation.

Supplementation with most AAs resulted in little change in the growth rate of EcN_AR in DMEM compared to the control wells. Some AAs caused an increase in growth, coupled with a downregulation of *clbB*, including L-Asparagine (1.11-fold), L-Aspartic acid (1.31-fold), D-Histidine (1.37), L-Serine (1.28-fold), and D-Serine (1.28-fold). The downregulation of *clbB* by L-Serine and D-Serine was significant by 27.1% ($p = 0.0026$) and 33.7% ($p < 0.0001$) respectively. Similar to what was observed in MEM-HEPES, Seleno-L-Cysteine decreased EcN_AR's growth by 0.38-fold, and significantly inhibited *clbB* expression significantly by 34.6% when grown in DMEM as a growth medium ($p < 0.0001$); (Figure 3-5).

To summarise the large amount of data generated during these experiments and to help identify which AAs reduced colibactin transcription, a Venn diagram was created, highlighting which AAs downregulated *clbB* transcription in each medium (Figure 3-6). Those AAs that decreased expression in at least two different types of media were selected for further validation of transcriptional repression using RT-qPCR. A total of 13 AAs were chosen for further testing. These included L-Asparagine, L-Aspartic acid, L-Histidine, L-Isoleucine, L-Leucine, L-Lysine, L-Methionine, L-Serine, D-Serine, Seleno-L-Cysteine, L-Threonine, D-Tryptophan, and D-Tyrosine (Figure 3-9).

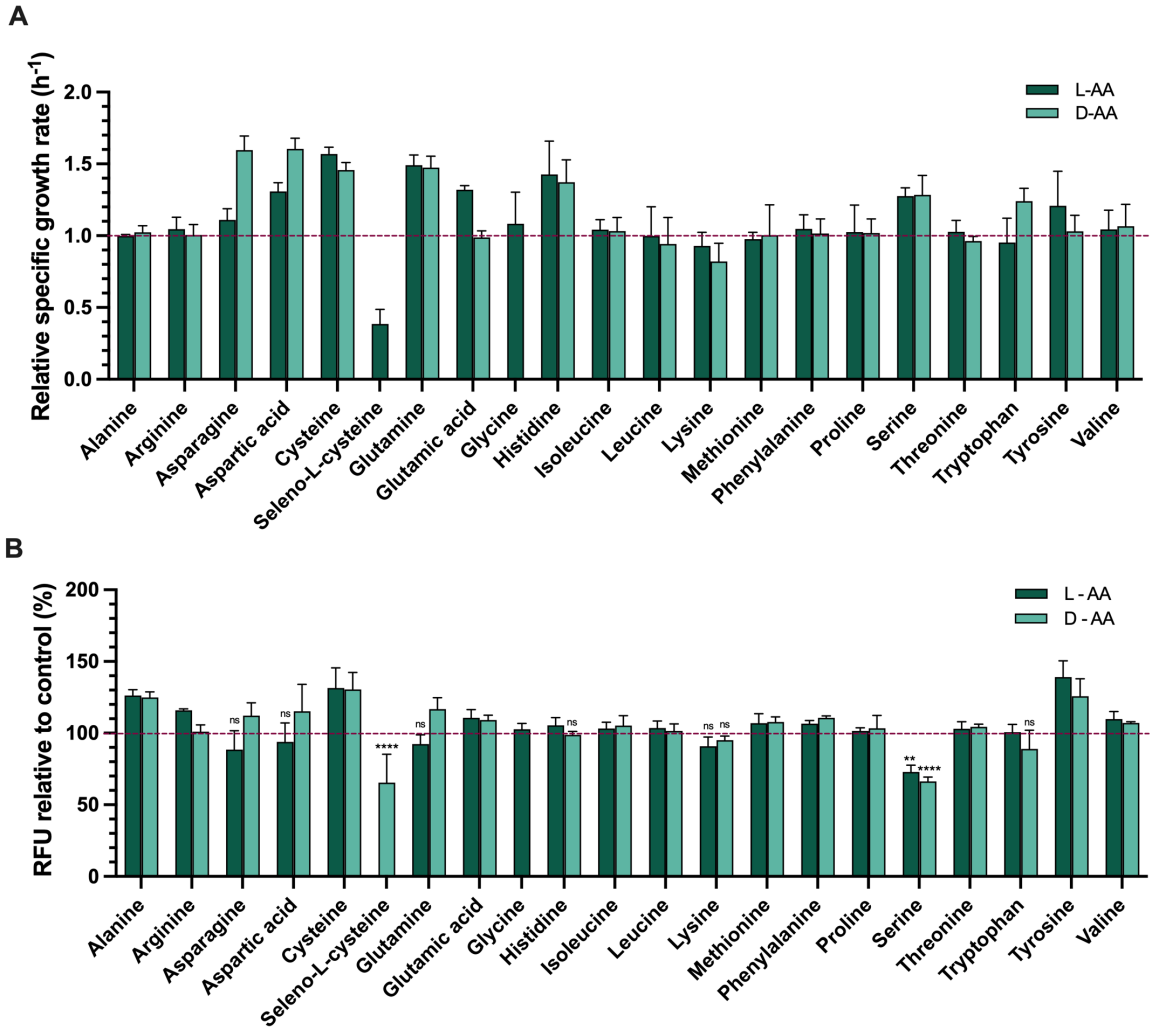


Figure 3-5 Modulatory effect of L- and D- AAs on EcN_AR growth and *clbB* promoter expression in DMEM. (**A**) The specific growth rates of EcN_AR in DMEM in the presence of 1 mM amino acids. Bacteria were cultured in DMEM +/- 1 mM amino acid in a 96-well plate at 37°C, 200 rpm. Absorbance ($OD_{600\ nm}$) was measured hourly and the specific growth rate ($\mu = \Delta \ln OD_{600nm}/\Delta t$) was calculated during exponential growth. The data is represented relative to the control condition (no amino acid supplementation), which is represented by the dashed red line. Columns represent means and error bars indicated standard deviation, calculated from three independent biological replicates. (**B**) EcN_AR *pclbB:gfp* reporter activity at exponential phase when grown in DMEM supplemented with 1 mM amino acids. Fluorescence was measured hourly over an 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$ in exponential phase, at 7 h in bacteria cultured in a 96-well plate at 37°C, 200 rpm. Columns represent means +/- standard deviation calculated from three independent biological replicates. Statistical significance was assessed for amino acids able to repress transcription by a Dunnett's multiple comparison test with ns, *, **, ***, **** indicating no significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively. The dashed line represents expression in control conditions, without any amino acid supplementation.

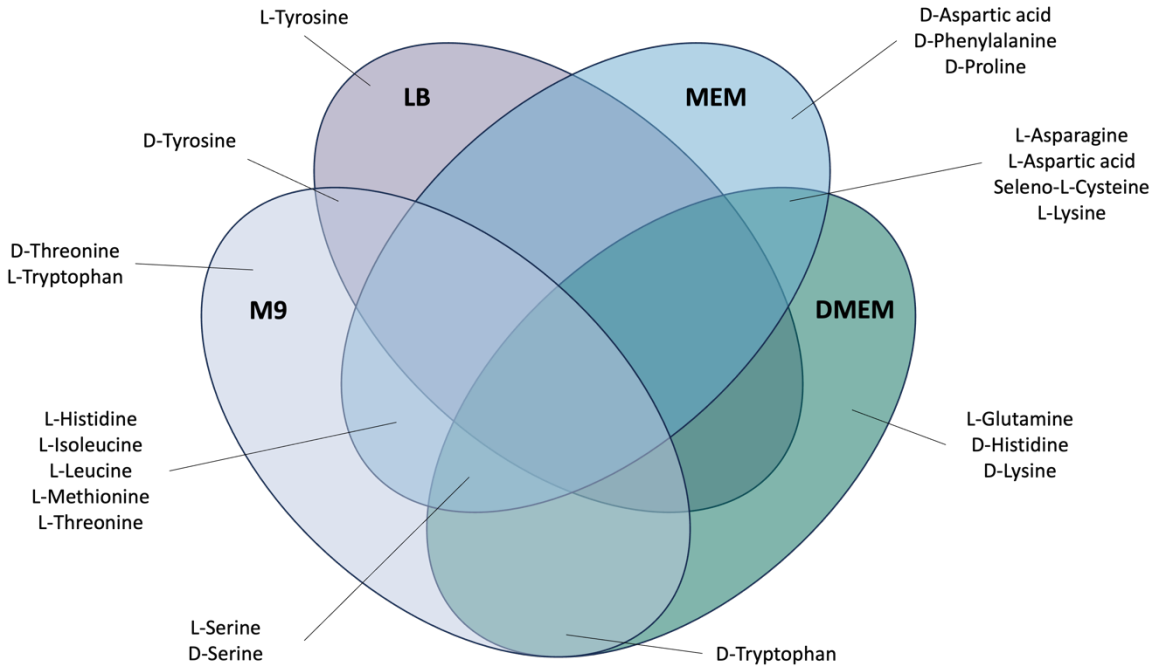


Figure 3-6 Amino acids classified by growth media-based *pclbB:gfp* repression. A *clbB* promoter:GFP reporter-fusion construct was used to assess the ability of proteinogenic L- and D-amino acids to repress colibactin production in different types of media. Amino acids were classified into a Venn diagram based on the growth medium in which they caused *clbB* downregulation. Amino acids that resulted effective in at least two types of media were chosen for further validation.

3.2.3 Addressing single-cell variability in EcN_AR

One observation made while performing reporter assays, was variation in *pclbB:gfp* expression, which although occurred infrequently, generated the large error bars of measured fluorescence shown in Figures 3-2, 3-3, 3-4, and 3-5. Consequently, it was hypothesised that variation in colibactin transcription could be due to heterogeneity in expression at the single cell level. This has been seen for many virulence factors, including protein effectors in *E. coli* O157:H7, such as Tir (Roe *et al.*, 2004) and NleA (Roe *et al.*, 2007). To test this hypothesis, a single-cell gene expression assay was performed, in which EcN_AR *pclbB:gfp* was grown in MEM-HEPES supplemented with 25 µg/mL of chloramphenicol at 37°C, 200 rpm, until an $OD_{600\ nm}$ of approximately 0.4 was reached. After being washed and fixed on a microscope slide, bacterial cells were labelled with the *E. coli* O-6 agglutinating antiserum (Mast Group) as a primary antibody and Goat anti-Rabbit IgG, Alexa Fluor™ 555 (Invitrogen) as a secondary label. In this way, fluorescent microscopy could be used to determine colocalization of individual bacteria and their *clbB* expression. These results revealed that *clbB* expression appeared to be occurring at a population level in EcN_AR across the three biological replicates tested (Figure 3-7).

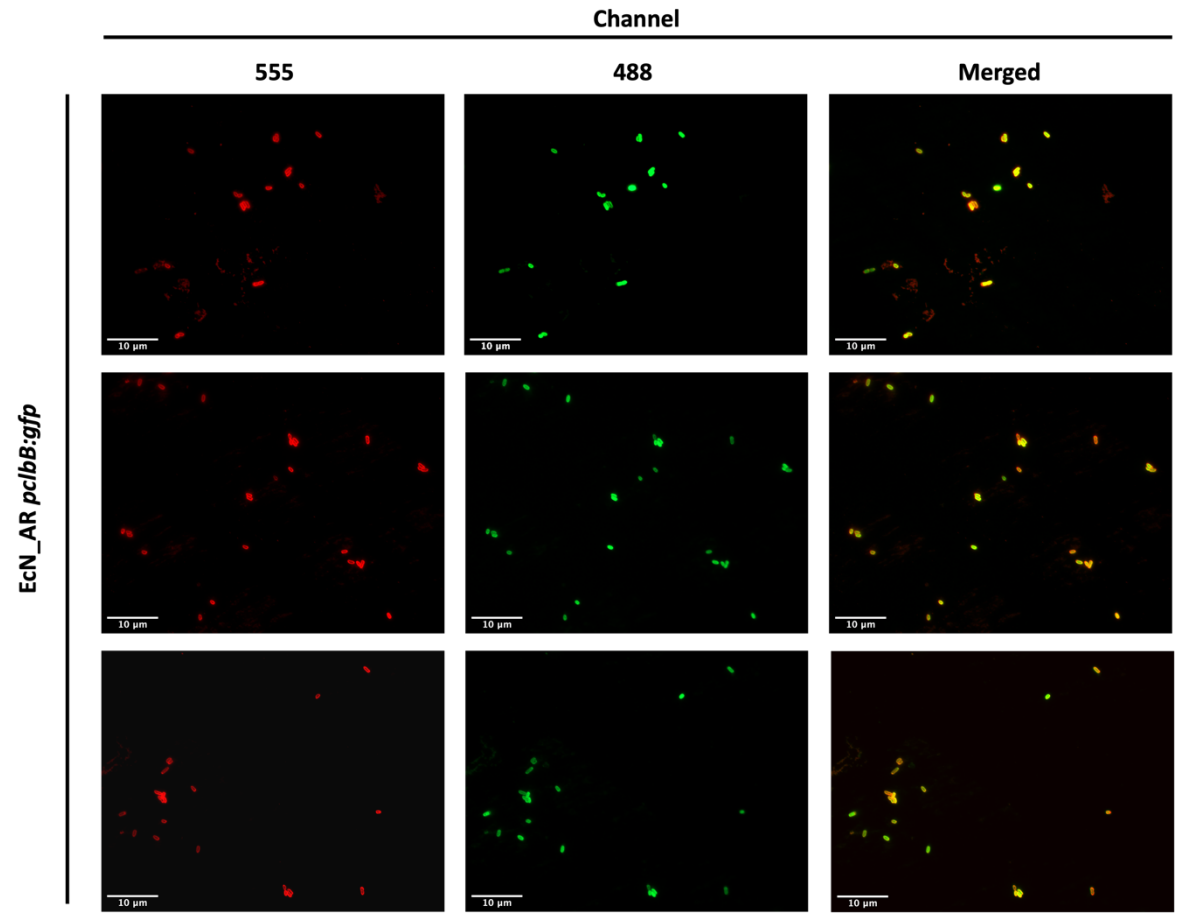


Figure 3-7 Immunofluorescent microscopy of EcN_AR *pclbB:gfp* in MEM-HEPES. EcN_AR *pclbB:gfp* was grown in MEM-HEPES supplemented with 25 µg/mL of chloramphenicol at 37°C, 200 rpm, until reaching an $OD_{600\ nm}$ of approximately 0.4. Bacterial cells were pelleted, washed with PBS, fixed with 4% PFA, and left to dry on a welled microscope slide. Afterwards, they were incubated for 1 h with *E. coli* O-6 agglutinating antiserum (Mast Group) at a 1:100 dilution and, following 3 washes, they were incubated with the Goat anti-Rabbit IgG, Alexa Fluor™ 555 (Invitrogen) antibody at a 1:400 dilution for 1 h. Bacteria were imaged at 100x magnification by fluorescent microscopy, scale bar represents 10 µm, three biological replicates are represented.

3.2.4 Comparing expression of colibactin in MEM-HEPES and DMEM

AAs that displayed a repressive effect on transcription of the *clbB* promoter in at least two types of growth media were considered to offer the greatest potential for repression of colibactin production. To support the observations made using the transcriptional reports, downregulation of expression was further validated by RT-qPCR. As ultimately these AAs would be used both in *in cellulo* assays (Chapter 4) and *in vivo* (Chapter 6), RNA for subsequent RT-qPCR validation was extracted from EcN_AR grown in medium compatible with tissue culture assays: MEM-HEPES or DMEM. Initially work focused on identifying the growth medium in which colibactin expression was enhanced so that any AA-driven downregulation could be more easily observed. Four biological replicates of EcN_AR were cultured in parallel in MEM-HEPES and DMEM. After 5 h samples were harvested for RNA extraction, and *pclbB* expression in subsequently purified RNA was assessed by RT-qPCR relative to the housekeeping gene *gapA*. Results showed that *clbB* expression was significantly higher when bacteria were grown in MEM-HEPES, by 20.3-fold, compared to DMEM ($p = 0.0007$). Moreover, growth in MEM-HEPES was more consistent than in DMEM, in which EcN_AR displayed an earlier exponential phase and death phase (Figure 3-8). Examination of the chemical composition of both growth media revealed that DMEM contained several AAs previously shown to have *clbB* inhibitory activity (highlighted in yellow), and at a higher concentration than those also present in MEM-HEPES, possibly explaining why colibactin expression appears significantly lower in DMEM (Table 3-1).

Consequently, RT-qPCRs, performed to evaluate the effects of different AAs on colibactin expression, were carried out using RNA extracted from EcN_AR grown in MEM-HEPES in the presence or absence of the selected, inhibitory AAs. Additionally, on the basis of this data, subsequent *in cellulo* assays which are described in Chapter 4, focusing on the functional action of colibactin, were also performed using MEM-HEPES rather than DMEM as the growth medium.

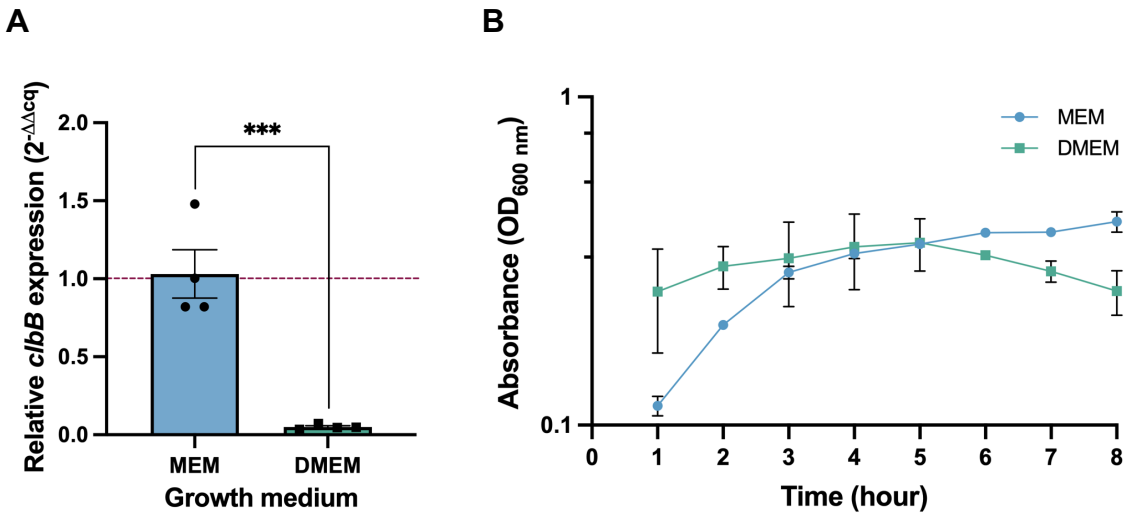


Figure 3-8 Identification of the most suitable growth medium for assessing colibactin expression. (**A**) Relative *clbB* expression in MEM-HEPES and DMEM measured by RT-qPCR. EcN_AR was grown in MEM-HEPES and DMEM for 5 h before extraction and conversion of RNA to cDNA. The expression of *clbB* was measured relative to the housekeeping gene *gapA*, using MEM-HEPES as the calibrator for baseline expression, as indicated by the dashed red line. Statistical significance was determined from four biological replicates, using a Student's t-test, with *** indicating significance, $p < 0.001$. (**B**) Growth curves of EcN_AR in MEM-HEPES and DMEM. EcN_AR was cultured in flasks in MEM-HEPES and DMEM at 37°C, 200 rpm. Absorbance ($OD_{600\ nm}$) was measured hourly for 8 h and values were plotted on a log-10 scale. Data points represent average absorbance values and error bars indicate standard deviation, calculated from three replicates.

Table 3-1 Comparison of amino acid composition in MEM-HEPES-HEPES and DMEM

Amino acid	MEM HEPES - M7278 (g/L)	DMEM - D5546 (g/L)
L-Arginine	0.126	0.084
L-Cysteine	0.031	0.062
Glycine	N/A	0.03
L-Histidine	0.042	0.042
L-Isoleucine	0.052	0.105
L-Leucine	0.052	0.105
L-Lysine	0.073	0.146
L-Methionine	0.015	0.03
L-Phenylalanine	0.032	0.066
L-Serine	N/A	0.042
L-Threonine	0.048	0.095
L-Tryptophan	0.01	0.016
L-Tyrosine	0.052	0.104
L-Valine	0.046	0.094

3.2.5 RT-qPCR validation of the most inhibitory AAs

To confirm the observation that specific AAs could repress colibactin expression, cultures of EcN_AR were grown for 5 h in MEM-HEPES supplemented with 1 mM of the AAs of interest and the RNA was harvested for extraction. The expression of *clbB* was assessed by RT-qPCR relative to the housekeeping gene *gapA* and measurements were normalised to a control condition of no AA supplementation in MEM-HEPES. This confirmed that several AAs inhibited *clbB* expression significantly (reduction of *clbB* expression and significance indicated in parentheses), including: L-Histidine (0.42-fold decrease, $p = 0.0059$), L-Isoleucine (0.39-fold decrease, $p = 0.0039$), L-Leucine (0.45-fold decrease, $p = 0.0108$), L-Lysine (0.54-fold decrease, $p = 0.0395$), L-Methionine (0.45-fold decrease, $p = 0.0094$), L-Serine (0.58-fold decrease, $p = 0.0293$), D-Serine (0.43-fold decrease, $p = 0.0069$), Seleno-L-Cysteine (0.05-fold decrease, $p < 0.0001$), L-Threonine (0.43-fold decrease, $p = 0.0074$), and D-Tyrosine (0.38-fold, $p = 0.003$). In contrast to the transcriptional reporter studies, L-Asparagine, L-Aspartic acid, and D-Tryptophan showed non-significant, minimal fold reductions of 0.99, 0.95 and 0.89 respectively (Figure 3-9). Although Seleno-L-Cysteine is the AA that elicits the most repressive effect on *clbB* expression, it was excluded from further consideration as a therapeutic candidate as it strongly inhibits EcN_AR's growth, suggesting that changes in gene expression could reflect limited growth, rather than specific *clbB* inhibition.

The two L-AAs, L-Isoleucine and L-Histidine, and the two D-AAs, D-Tyrosine and D-Serine, that showed the highest level of transcriptional repression, were chosen for further validation by *in cellulo* and *in vitro* assays, described in the following chapter. L-Serine was also included for further study, not only because it displayed significant repressive activity, but also because it allowed direct comparison with the activity of D-Serine, which had previously been identified to downregulate colibactin expression (Hallam *et al.*, 2023).

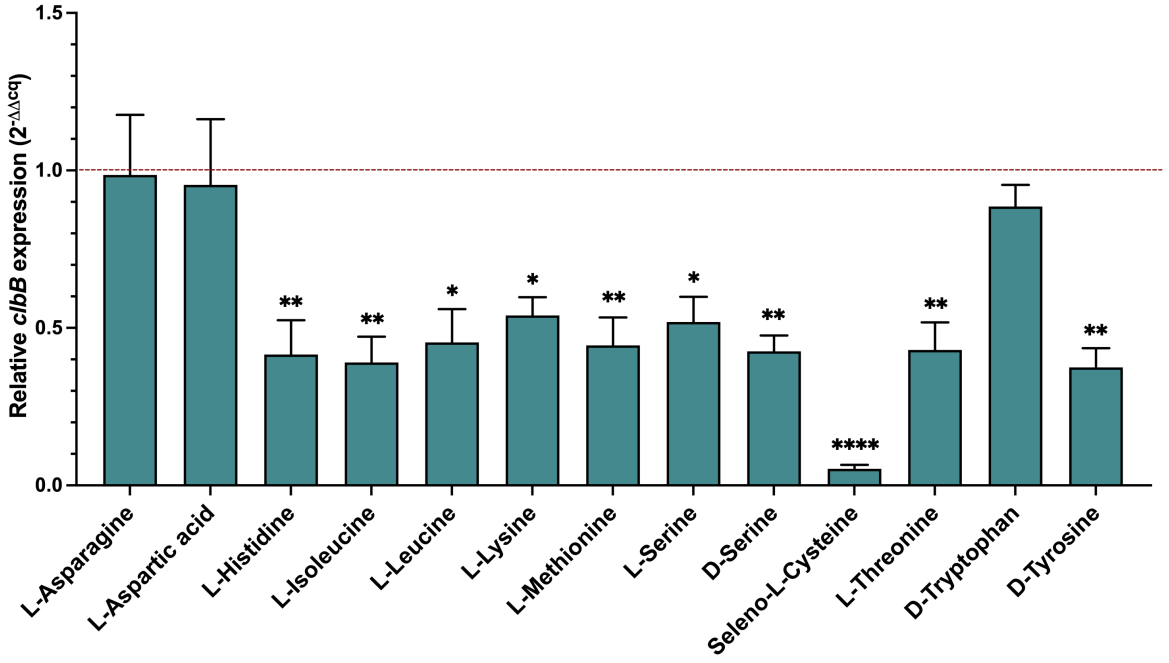


Figure 3-9 The amino acids able to repress transcription screened by RT-qPCR in MEM-HEPES. EcN_AR was grown in flasks at 37°C, 200 rpm for 5 h in MEM-HEPES supplemented with a 1 mM concentration of the amino acids of interest. RNA was extracted from bacterial cultures and converted to cDNA. The expression of *clbB* was measured relative to the housekeeping gene *gapA*, using a control condition of no amino acid supplementation as the calibrator for baseline expression, indicated by the dashed red line. Statistical significance was determined from three biological replicates by a Dunnett's multiple comparison test with *, **, ***, **** indicating significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively.

3.3 Discussion

Using a combination of approaches, data acquired in this chapter revealed that a number of L- and D- AAs appear to repress colibactin transcription. The use of the transcriptional reporter assay was convenient for screening forty AAs simultaneously, in multiple replicates, and in various types of growth media. However, growth of bacterial cultures in a 250 µL volume provides strong limitations, with accuracy, precision, and reproducibility being common issues when working with smaller volumes. Indeed, when researching existing literature, several studies that investigate colibactin regulation were found, but none focused on heterogeneous expression of *clb* genes. It is therefore probable that the variations in colibactin expression observed across different replicates were due to reporter assays being performed in 96-well plates. RT-qPCR is a more sensitive technique, allowing for the measurement of gene expression, instead of the measurement of a reporter gene's promoter activity. A challenge in RT-qPCRs is the selection of an appropriate housekeeping gene – that should remain unvaried across different experimental conditions – required for normalisation. Moreover the transcriptional reporter assays and the RT-qPCR, only provided information on *clbB*. While this is an essential gene in colibactin biosynthesis and its regulation appears to be uniform at a population level, it does not reflect transcription or expression of the whole *clb* cluster, highlighting the need to complement *in vitro* assays with functional ones.

Interestingly, when investigating the modulatory effects of L- and D-AAs using the transcriptional reporter plasmid, it was notable that only L- and D-Serine elicited repression of *clbB* in three types of media, M9, MEM-HEPES, and DMEM. In contrast, downregulation by other AAs was limited to one, or two types of supplemented growth media, suggesting that L- and D-Serine modulate colibactin expression more broadly (Figure 3-6). The lack of repression observed in LB might be due to the fact that it is a rich culture medium, so supplementation with further AAs has little impact. L- and D-Tyrosine were unique in causing a downregulation of *clbB* transcription in LB (Figure 3-6).

Explanations for the specific inhibitory activity of the AAs identified in this study are difficult to provide. Initially, it was investigated whether these AAs had structural features or biochemical properties in common, however, these are limited. For example, isoleucine, and leucine are BCAAs, and are nonpolar, hydrophobic, and aromatic, while histidine, lysine, serine, threonine, and tyrosine are polar and hydrophilic. Histidine and tyrosine are both aromatic, but while histidine has a positive charge, tyrosine is uncharged. Lysine is basic, while tyrosine is classified as acidic (Ahluwalia, Kumar and Kumar, 2022). It was therefore concluded that the inhibitory AAs are likely to act on colibactin through distinct pathways, which may end up converging into a common mechanism of *clbB* repression.

In the mammalian gut, dietary AAs are largely absorbed in the small intestine, specifically in the proximal jejunum by transporters present on the membranes of enterocytes. This occurs after enzymes from the stomach (pepsins) and from the pancreas (pancreatic proteases) have hydrolysed peptide linkages in proteins. AAs are partially absorbed directly by splanchnic tissues, but the majority are released into systemic circulation and absorbed by peripheral tissues. Despite the efficient digestion occurring in the upper digestive tract, some proteins, peptides, AAs, and other metabolites containing nitrogen, reach the large intestine, where they are broken down by the microbiota (van der Wielen, Moughan and Mensink, 2017). A key role of the intestinal microbiota is the degradation of complex carbohydrates – mostly fibres – which results in the production of short-chain fatty acids (SCFAs), such as acetate, propionate, and butyrate. In particular, some *Clostridia* species perform so-called Stickland Fermentation, a paired oxidation-reduction of AAs, using them as a source of nutrients to support anaerobic growth in the absence of glucose (Pavao *et al.*, 2022). The catabolism of AAs results in the generation of various metabolites, including SCFAs, branch-chained fatty acids (BCFAs), neurotransmitters, amines, and phenolic compounds (Beaumont and Blachier, 2020) (Trommelen, Tom and van Loon, 2021).

While this work focused on individual AAs, there are several compounds, with specific AAs as precursors, that could play a role in colibactin synthesis. Spermidine is an example of an abundant polyamine found in the gut, which has been highly associated with cellular proliferation and cancerous tissues (Gerner,

Bruckheimer and Cohen, 2018). Spermidine has also been shown to modulate colibactin production by *E. coli*; specifically, exogenous spermidine has been shown to support colibactin biosynthesis (Chagneau *et al.*, 2019). In this study, L-Arginine, which is a precursor of spermidine, was not one of those AAs identified to downregulate colibactin expression. Polyamines are abundant in cancerous tissues and associated to cell proliferation (Gerner, Bruckheimer and Cohen, 2018). Spermidine's stimulation of colibactin biosynthesis is coherent with clinical findings that *pks+ E. coli* are abundant in colonic tumour biopsies (Buc *et al.*, 2013). Furthermore, since polyamines can be produced by the microbiota – in addition to being acquired through diet – the use of exogenous spermidine by CoPEC, may be an indicator of niche adaptation or symbiosis with other bacterial species. In contrast, a study looking at *pks+ E. coli* strains, showed that the SCFA's acetate, butyrate and propionate were all able to inhibit *E. coli* growth and reduce expression of virulence genes associated with genotoxicity, motility and adhesion (Zhang *et al.*, 2020). In this case, L-Lysine, a precursor of acetate and butyrate, and L-Threonine, a precursor of acetate and propionate, were both AAs that were identified to have significantly repressive effects on colibactin transcription.

Nutrient availability has also been demonstrated to affect colibactin biosynthesis, with nutrient-rich media appearing to repress colibactin expression, and less complex media enhancing expression (Wallenstein *et al.*, 2020). Interestingly, one study showed that in DMEM the expression of *clb* genes is high, with the authors hypothesising that this is due to high glucose levels and the presence of AAs and vitamins in the growth media (Homburg *et al.*, 2007). In contrast, our findings suggest that in MEM – which is less rich than DMEM – expression of colibactin is higher. This could be due to MEM-HEPES composition having fewer AAs able to cause colibactin repression and, when present, these AAs were at a lower concentration than in DMEM. An interesting point to note is that, unlike MEM-HEPES, DMEM contains ferric nitrate. Iron availability is a well-known regulator of the *pks* island, with *clb* genes upregulated in low-iron conditions (Tronnet *et al.*, 2016). Another possible explanation for the higher expression of colibactin in MEM-HEPES could therefore be the absence of ferric nitrate in its composition.

Overall, data obtained in this chapter shows that several proteinogenic AAs have a modulatory effect on colibactin transcription in EcN and that, in particular, a

number of L- and D- AAs appear to repress colibactin transcription. The ability of these AAs to repress colibactin appears to be dependent on the type of growth medium in which EcN was cultured, with the expression of *clbB*, one of the essential genes in colibactin biosynthesis, appearing significantly higher in MEM-HEPES than in DMEM. The AAs that displayed repressive effects on colibactin transcription by reporter fusion assays were further validated by RT-qPCR. The two L-AAs, L-Isoleucine and L-Histidine, and the two D-AAs, D-Tyrosine and D-Serine, that showed the highest level of *clbB* repression, were chosen for further validation by *in cellulo* and *in vitro* assays described in the following chapter.

4 Validation of the most inhibitory amino acids through *in cellulo* and *in vitro* assays with prototypical strains and clinical isolates

I would like to acknowledge Sam McAllister and Patricia Rimbi for their help in operating the PhyloBuild pipeline in Figure 4-12, and Iris Floria who performed the assays in Figure 4-13 under my supervision in the Roe Lab.

4.1 Introduction

Results from the previous chapter showed that L-Histidine, L-Isoleucine, L-Serine, D-Serine, and D-Tyrosine display downregulatory activity on colibactin expression. Consequently, the work described in this chapter aimed to (1) validate the five AAs *in cellulo* and *in vitro* to confirm further their capacity to inhibit colibactin, (2) assess whether combinations of AAs could have a synergistic effect in reducing colibactin expression, and (3) select which AA to test *in vivo* against EcN.

This introduction provides an overview of the impact of the selected five AAs on *E.coli* physiology, with a particular focus on their antivirulence potential.

L-Histidine

The histidine catabolism pathway is highly conserved among bacteria, and is commonly referred to as the Hut system (histidine utilisation). The first three steps consist of the elimination of ammonia from histidine, followed by hydration and ring cleavage. Final products result in ammonia, glutamate, formamide or formate, depending on the bacterial species (Bender, 2012). *E. coli* lacks the *hut* operon and, therefore, cannot grow on L-Histidine due to the lack of degradative enzymes (Zhang *et al.*, 2012). However, it does possess the Histidine transporter complex HisJQMP, which is part of the more extensive ABC transporter system, involved in the import of lysine, arginine and ornithine (Caldara *et al.*, 2007). Its genome also encodes the *his* operon, a cluster of nine genes that synthesise histidine starting from phosphorylation of 5-phosphoribosyl-pyrophosphate (PRPP) (Grisolia, Carlomagno and Bruni, 1982). One study showed that L-Histidine can suppress the proliferation of *recA*-deficient *E. coli* strains in aerobic conditions, and that L-Histidine has an augmenting effect on the DNA-degradation induced by hydrogen peroxide on *E. coli* cells, suggesting that the presence of L-Histidine increases oxidative DNA damage (Nagao *et al.*, 2018). An *in vivo* mouse model of colitis showed that *Faecalibaculum rodentium*, whose bloom is induced by enteral nutrition – a dietary strategy for patients with Crohn's disease – appears to protect mice from colitis through histidine biosynthesis (Zeng *et al.*, 2023).

L-Isoleucine

In *E. coli*, L-Isoleucine is biosynthesised from the deamination of threonine, generating 2-ketobutyrate, which is then condensed with pyruvate to produce 2-aceto-2-hydroxybutanoate. Subsequently, isomerisation, reduction, dehydration, and amination yield L-Isoleucine as the final product. Pyruvate also acts as the precursor for L-Valine and L-Leucine and the biosynthesis of all three AA relies on enzymes encoded by the *ilv* operon, hence why the entwined biosynthetic pathways of the three proteinogenic BCAAs are often represented as a common biosynthetic superpathway (Smith, Smolin and Edwin Umbarger, 1976), (Rhee, Parekh and Hatfield, 1996), (Cotton *et al.*, 2020). In *E. coli* the main transporter for BCAA is BrnQ, though evidence suggests that CycA is also involved in their import (Hook *et al.*, 2022). Although there is currently no literature demonstrating the potential of L-Isoleucine to exert antimicrobial activity on *E. coli*, studies have shown that in human colon cells, L-Isoleucine can upregulate the production of human beta-defensin 1, the major antimicrobial peptide in human epithelia (Sherman, Chapnik and Froy, 2006). This effect does not appear to be exclusive to L-Isoleucine, and is also exerted by its analogues, like L-Isoleucine Hydroxamate, L-Norvaline, and L-Alloisoleucine. The study's authors hypothesise that L-Isoleucine is secreted by bacteria as an intercellular communication molecule, hence acting as a microbial marker and prompting the induction of host defensins (Fehlbaum *et al.*, 2000).

L-Serine

The biosynthesis of L-Serine in *E. coli* is a three-step process that starts with the oxidation of 3-phosphoglycerate and is facilitated sequentially by the gene products of *serA*, *serC*, and *serB*. A secondary pathway for serine biosynthesis is via the Tut cycle, whereby L-Threonine is converted to glycine, which is converted to L-Serine through a reversible reaction (Stauffer, 2004). *E. coli* genes involved in the transport of exogenous L-Serine include the following: *sstT*, *cycA*, *sdaC*, and *tdcC* (Wang *et al.*, 2020). The subsequent degradation of L-Serine consists of a single enzymatic reaction catalysed by L-Serine ammonia-lyases encoded by *sdaA* and *sdaB*, producing ammonia and pyruvate. A metabolic profile conducted on *A.*

baumannii strains showed that low-virulent strains have reduced levels of L-Serine production, compared to their high-virulent counterparts. The same study also showed that L-Serine reduces virulence of *A. baumannii in cellulo* by upregulating SIRT1 and reducing mortality of *Galleria mellonella* to *A. baumannii* (Zhou *et al.*, 2016). A study conducted on *E. coli* instead, showed that L-Serine sensitised clinical isolates that were fluoroquinolone-resistant to the antibiotic through an increase in the production of reactive oxygen species. It also showed that L-Serine could potentiate the action of ofloxacin and moxifloxacin, with fewer bacterial cells persisting when antibiotics were combined with the AA (Duan *et al.*, 2016).

D-Serine

Within *E. coli* cells, D-Serine can be produced through racemisation of L-Serine by the amino acid racemase YgeA (Miyamoto *et al.*, 2017). It can also be directly imported by CycA – a previously-mentioned permease employed in the transport of several AAs across the cytoplasmic membrane (Cosloy, 1973), by DsdX – the D-Serine transporter encoded within the *dsdCXA* locus (Anfora and Welch, 2006) – and by YhaO, a functional D-Serine transporter in both EHEC and UPEC (Connolly *et al.*, 2016). The enzyme D-Serine ammonia lyase, encoded by *dsdA*, catalyses the deamination of D-Serine, yielding pyruvate and ammonia and allowing *E. coli* to use D-Serine as a sole carbon and nitrogen source (Metzler and Snell, 1952), while *dsdC* acts as a LysR-type transcriptional regulator (LTTR) that controls system response upon exposure to D-Serine (Norregaard-Madsen, McFall and Valentin-Hansen, 1995). D-Serine has been found to modulate bacterial virulence through activation of the SOS response and inhibition of T3SS expression in EHEC O157:H7 (Connolly *et al.*, 2015), to reduce colibactin expression in prototypical and clinical isolates of *E. coli* belonging to the B2 phylogroup (Hallam *et al.*, 2023), and to reduce *A. baumannii* attachment to human alveolar cells (Rumbo *et al.*, 2016).

D-Tyrosine

In E.coli D-Tyrosine is synthesised through racemisation of L-Tyrosine, catalysed by the AA racemases MetC and MalY (Miyamoto *et al.*, 2018). Although generally only L-AAs are incorporated into protein synthesis, a few peptide-forming D-AAs have been found in soluble fractions of bacteria, archaea and eukaryotes. These D-AAs act as substrates for aminoacyl-tRNA synthetases which transfer the appropriate AAs to their corresponding tRNAs. This has been seen in D-Tyrosine (Calendar and Berg, 1967), D-Aspartic acid, D-Tryptophan, D-Histidine and D-Lysine (Takayama *et al.*, 2005), but catalytic efficiency with the D-enantiomers is considerably lower than with the L-enantiomers. Among the aminoacyl-tRNA synthases, tyrosyl-tRNA synthetase, also known as tyrosine-tRNA ligase, is the enzyme that catalyses the esterification of both L- and D-Tyrosine to tRNA. The deacylase D-Tyr-tRNA^{Tyr}, encoded by the *dtd* gene in *E. coli* (Soutourina *et al.*, 1999), is capable of hydrolysing the D-Tyrosine-tRNA complex, freeing tRNA, which is why – in the absence of the deacylase – D-Tyrosine is toxic to *E. coli* cells (Soutourina *et al.*, 2004). In terms of antivirulence potential, D-Tyrosine has been shown to inhibit biofilm formation of *S. aureus* (Hochbaum *et al.*, 2011), *P. aeruginosa*, and *B. subtilis* (Yu *et al.*, 2016). Studies performed on *E. coli* showed that D-Tyrosine reduces the secretion of extracellular polymeric substances and downregulates genes involved in adhesion (Li *et al.*, 2023).

4.2 Results

4.2.1 Determination of the most suitable cell line for EcN infection assays

As previously described, colibactin exerts a cytopathic effect on eukaryotic cells, characterised by a progressive enlargement of the cell body and nucleus, known as megalocytosis. The work in the first part of this chapter focuses on whether the megalocytotic phenotype driven by EcN-derived colibactin, can be ameliorated by the availability of AAs that repress colibactin transcription. To study the *in cellulo* effect of the selected repressive AAs identified in the previous chapter, mammalian cells were infected with EcN in the presence or absence of these AAs, with quantification of cell size used as a correlate of cellular senescence. To achieve this objective, an appropriate cell line, in which megalocytosis could be observed, had to be selected. To test this, HT29, a colonic cell line (Martínez-Maqueda, Miralles and Recio, 2015), and HeLa (Landry *et al.*, 2013), a commonly used transformed epithelial cell line, were chosen. The rationale being that HT29 was a good model of the site where CRC is found; and HeLa cells had already been used in previous papers looking at colibactin-induced megalocytosis (Hallam *et al.*, 2023). Infections were performed on HeLa and HT29 cells with the laboratory strain DH10B, an *E. coli* strain hosting an artificial plasmid that bears the complete *pks* island. This strain constitutively expresses colibactin, acting as a positive control for toxin expression. DH10B was cultured for 4.5 h in MEM-HEPES, before infecting cells at an MOI of 400. After 4 h, the cells were washed and the infection was interrupted by the addition of gentamicin, which kills extracellular organisms. Following an incubation of three days, which allows the megalocytotic phenotype to develop, cells were fixed, stained, and imaged by fluorescent microscopy. Although both cell lines exhibited cellular and nuclear enlargement, megalocytosis was undoubtedly more appreciable in HeLa cells (Figure 4-1). Moreover, given that HT29 cells seemed to grow in aggregated clumps rather than even monolayers and excreted mucin, which made cell separation difficult, it was reasoned that quantification would be more accurate with HeLa cells rather than HT29 cells. Consequently, HeLa cells were chosen as a suitable cell line for subsequent analysis of the colibactin-associated megalocytosis phenotype.

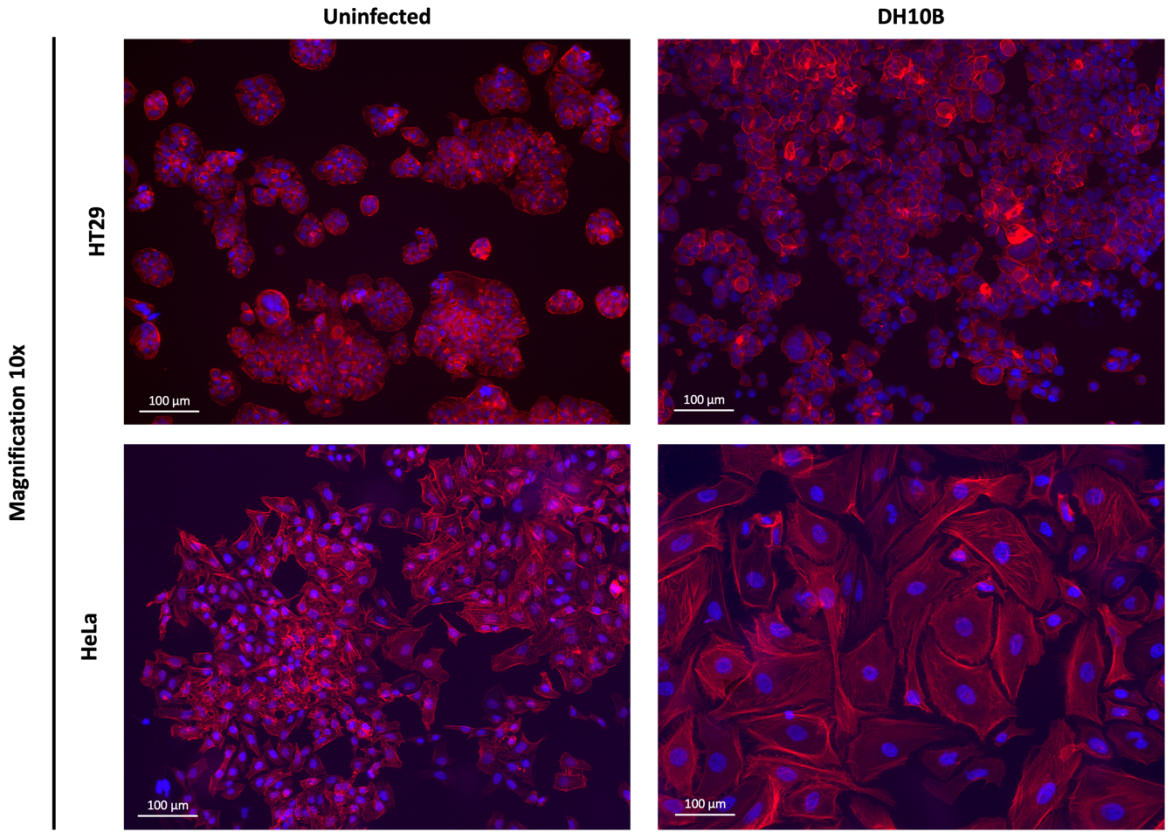


Figure 4-1 Comparison of megalocytosis in HT29 and HeLa cells. HT29 and HeLa cells were infected in MEM-HEPES with DH10B (MOI = 400), which had been previously cultured at 37°C, 200 rpm for 4.5 h in MEM-HEPES supplemented with 25 µg/mL of chloramphenicol. At 4 h after infection, cells were washed and incubated for 72 h in DMEM supplemented with 200 µg/mL of gentamicin. Cells were fixed with 4% PFA; actin cytoskeleton was stained with phalloidin (red) and nuclei were stained with DAPI (blue). Cell morphology was observed by fluorescent microscopy; images were acquired at 10X magnification, scale bar = 100 µm.

4.2.2 The effects of selected amino acids on colibactin-driven megalocytosis in HeLa cells infected with EcN

To investigate whether the selected, repressive AAs could ameliorate colibactin-associated megalocytosis, HeLa cells were infected with EcN in the presence of 1mM L-Histidine, L-Isoleucine, L-Serine, D-Serine, or D-Tyrosine. Infections with both EcN_AR and EcN_EO were performed as four biological replicates, with no differences observed between the two strains during *in cellulo* assays. To quantify cellular senescence, cells were infected for 4 h, treated with gentamicin, and then left to incubate for 72 h. Images were acquired by light microscopy at 20X, and 5 images were systematically acquired at the centre, top, bottom, left, and right of each well, for a total of 20 images per infection condition, across all biological replicates. A pipeline developed on CellProfiler (Stirling *et al.*, 2021) was employed to measure individual cell areas in pixels. HeLa cells infected with EcN under no AA supplementation had an average area of 49440 (100%) pixels. Supplementation during infection with 1mM L-Histidine, L-Isoleucine, L-Serine, D-Serine, or D-Tyrosine instead resulted in average cell areas of respectively 49831 (100.79%), 38224 (77.31%), 32458 (65.65%), 29060 (58.78%), and 39464 (79.82%) pixels. The only AAs that prompted a significant reduction in cell area were L-Serine and D-Serine, which induced a decrease in cell area of respectively 34.35% ($p = 0.0184$) and 41.22% ($p < 0.0001$); (Figure 4-2).

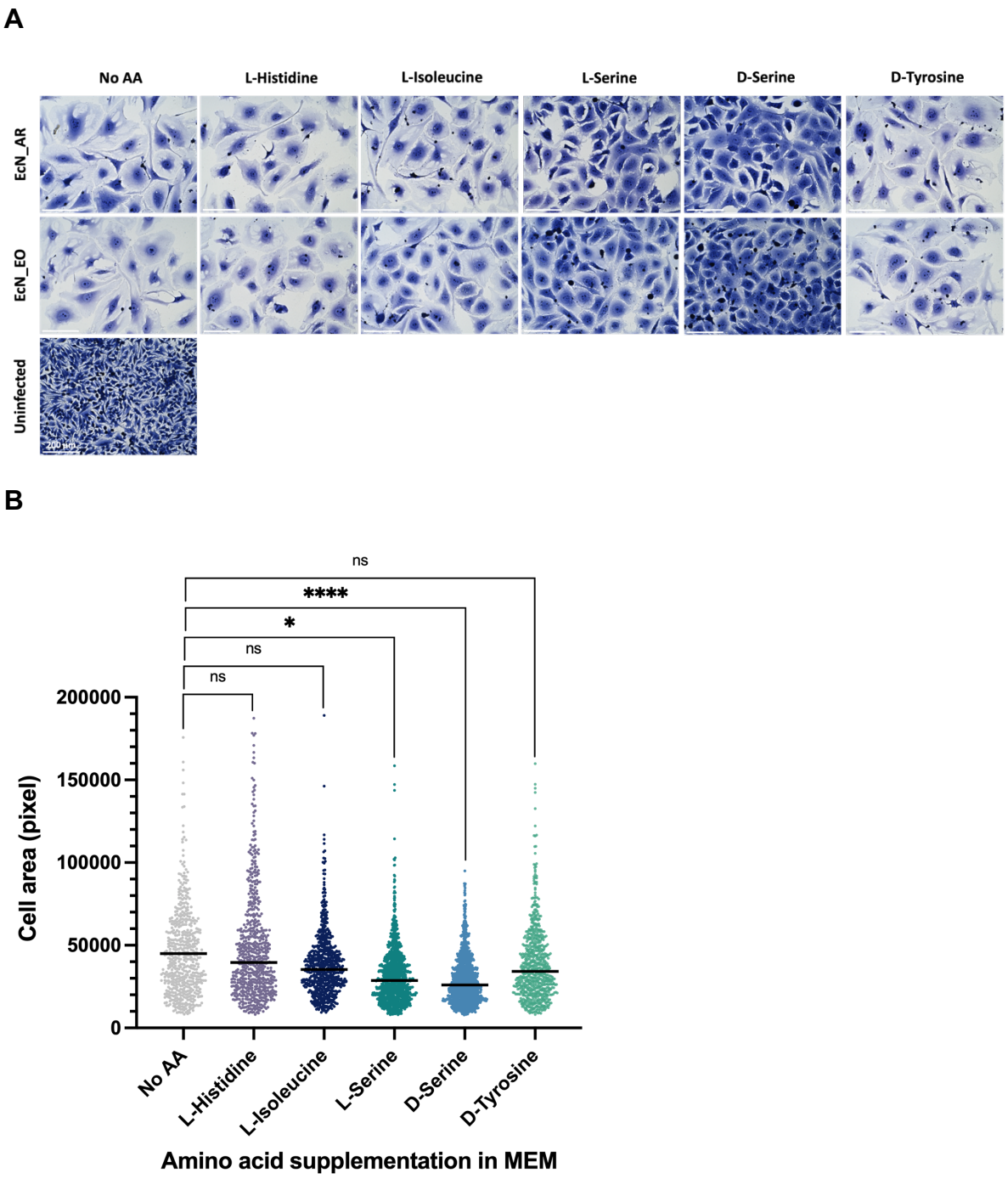


Figure 4-2 Assessment of the selected, inhibitory amino acids against colibactin-driven megalocytosis. HeLa cells were infected with EcN (MOI = 400) in MEM-HEPES alone or MEM-HEPES supplemented with 1 mM AAs in 24-well plates. EcN strains had been previously cultured at 37°C, 200 rpm for 4.5 h in MEM-HEPES alone or MEM-HEPES supplemented with 1 mM AAs. At 4 h after infection, cells were washed and incubated for 72 h in DMEM supplemented with 200 µg/mL of gentamicin. (**A**) Acquisition of cell images by light microscopy. Cells were fixed with 4% PFA and stained with methylene blue. Cell morphology was observed by light microscopy; images were acquired at 20X magnification, scale bar = 200 µm. A total of 5 images were captured in each well, in the centre, top, bottom, left and right. Two infections were performed with EcN_AR and two with EcN_EO for a total of four biological replicates. (**B**) Quantification of cell areas under different infection regimes. A pipeline was developed on CellProfiler software to measure the area of the cells, whose images had been previously captured. As described, measurements were taken from images acquired from four independent experiments. Statistical significance was determined using a Dunnett's multiple comparison test, with * and **** respectively indicating $p < 0.05$ and $p < 0.0001$.

Since colibactin triggers cell-cycle arrest that eventually leads to cell death, another way of quantifying colibactin's activity was by measuring the number of cells that survived following infection. In this instance, larger 6-well plates were employed, and cells were infected and incubated as previously described. Images were acquired by light microscopy at 20X: 10% of the well area was imaged from the centre, capturing a total of 181 snapshots that were stitched together into a single image. Once again, data was generated using four biological replicates, with two EcN_AR and two EcN_EO infections. ImageJ (Rasband, 1997-2018) was employed to calculate the percentage of the well covered by cells. Analyses showed that L-Serine and D-Serine were the only two AAs that induced a significantly greater survival of EcN-infected HeLa cells. Compared to control conditions which were not supplemented with an AA, survival increased by 24.98% ($p = 0.0018$) and 22.81% ($p = 0.0040$) for L-Serine and D-Serine respectively. It should be noted these calculated values look lower than the images suggest. This is because the ImageJ analysis used included the 'fill holes' function, which deletes connected elements when filling in the background. However – as the same pipeline was employed to quantify all images – variations in calculations were consistent across all wells and the data is consequently comparable (Figure 4-3).

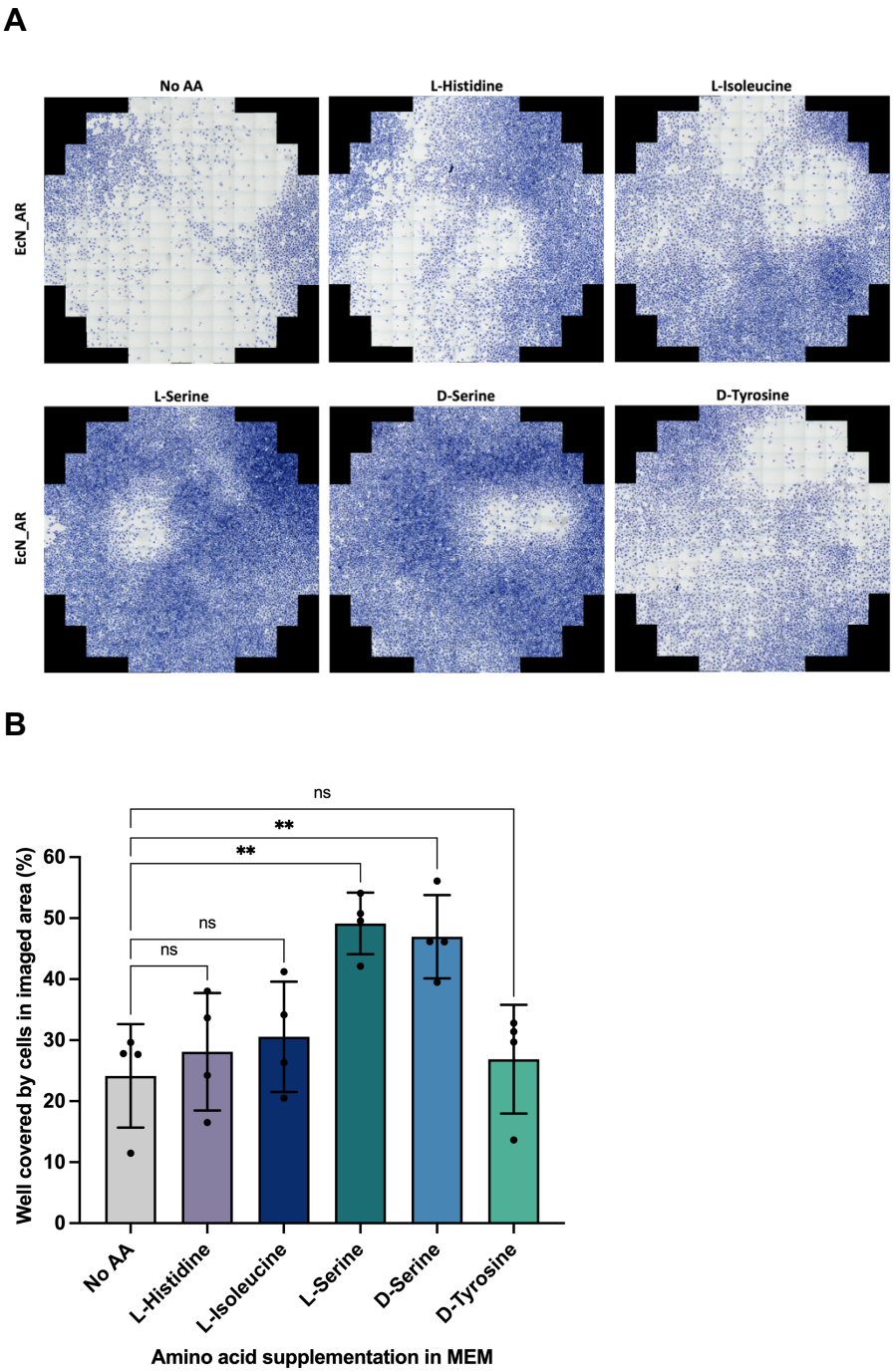
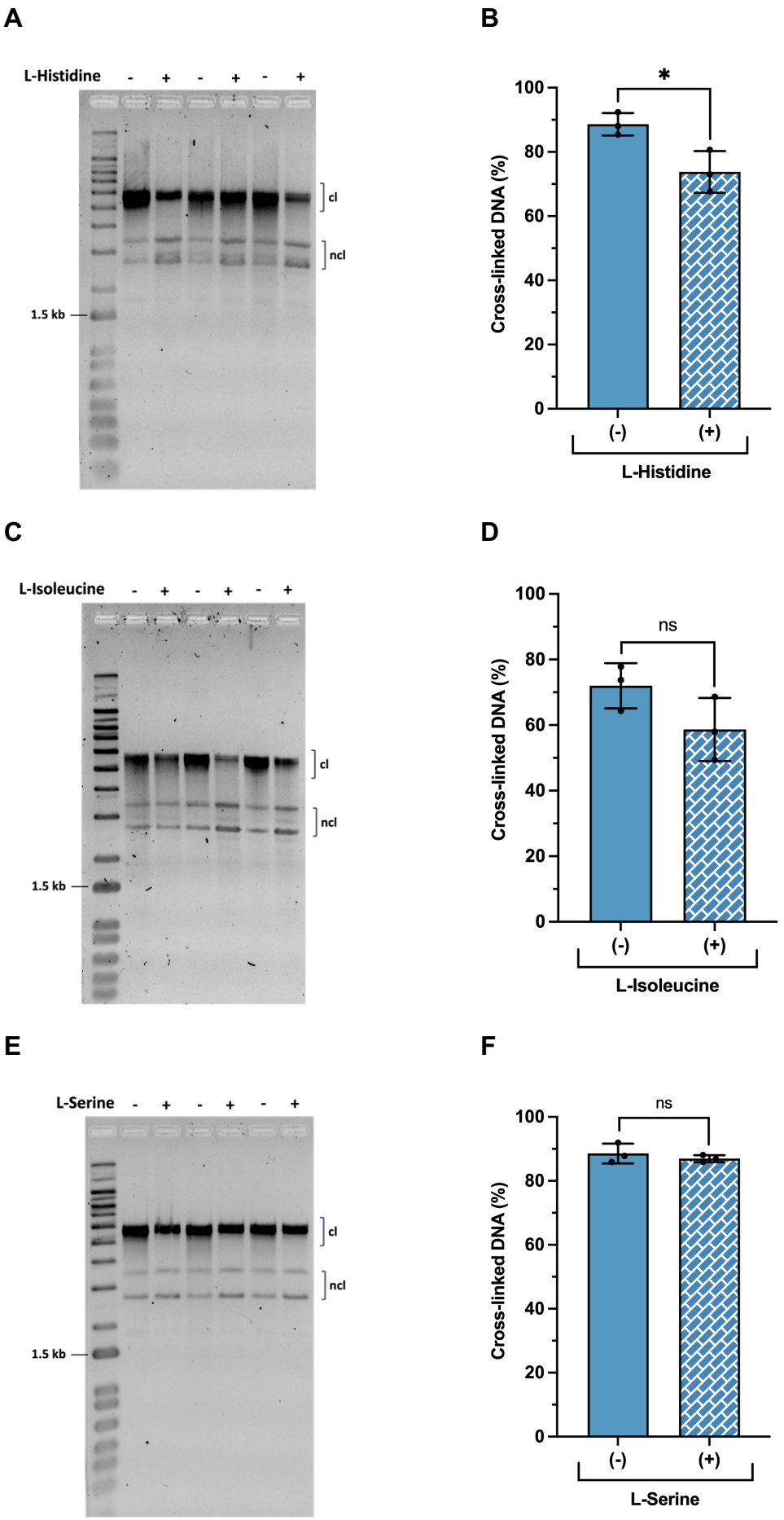


Figure 4-3 Assessment of the selected, inhibitory amino acids against colibactin-driven cellular senescence. HeLa cells were infected with EcN (MOI = 400) in MEM-HEPES alone or MEM-HEPES supplemented with 1 mM AAs in 6-well plates. EcN strains had been previously cultured at 37°C, 200 rpm for 4.5 h in MEM-HEPES alone or MEM-HEPES supplemented with 1 mM AAs. At 4 h after infection, cells were washed and incubated for 72 h in DMEM supplemented with 200 µg/mL of gentamicin. (**A**) Acquisition of the central well area by light microscopy. Cells were fixed with 4% PFA and stained with methylene blue. An automation setting was used to capture 10% of the well area, corresponding to the central part of each well. For each well, 181 images were acquired at 20X magnification and stitched together. Two infections were performed with EcN_AR and two with EcN_EO for a total of four biological replicates. (**B**) Quantification of the percentage of the well covered by cells within the imaged area. A pipeline was developed on ImageJ software to measure what percentage of the captured well area was covered by cells. As described, measurements were taken from stitched images acquired from four independent experiments. Statistical significance was determined using a Dunnett's multiple comparison test, with ** indicating $p < 0.01$.

4.2.3 The effects of selected amino acids on colibactin-driven DNA crosslinking

Since colibactin induces cell-cycle arrest by interstrand crosslinking of DNA, the impact of AAs on reduction of colibactin-driven DNA damage can be assessed by alkaline agarose gel electrophoresis. This is because the hydrogen bonds that maintain the double-stranded nature of DNA are disrupted at the high pH used during electrophoresis, allowing the DNA to migrate through the gel according to its size. Therefore, if interstrand crosslinking occurs, the migration of DNA is expected to be delayed. To determine the impact of AAs on crosslinking activity, an interstrand crosslinking assay (ISCLA) was performed. First, linearised pUC19 was incubated with EcN_EO for 5 h in the presence or absence of 1 mM of the five selected AAs. Then, the plasmid DNA was purified and observed by alkaline gel electrophoresis. Quantification of DNA crosslinking was examined via densitometry, using ImageJ (Rasband, 1997-2018). Results showed that three AAs decreased DNA crosslinking significantly: L-Histidine by 16.74% ($p = 0.0254$), D-Serine by 51.40% ($p = 0.0252$); and D-Tyrosine by 70.88% ($p = 0.004$). In contrast, L-Isoleucine caused a nonsignificant reduction of 18.51%, while surprisingly L-Serine had a negligible effect, inducing a trivial decrease of crosslinking by only 1.82% (Figure 4-4).


A
L-Histidine - + - + - +
cl
ncl
1.5 kb
B
*
Cross-linked DNA (%)
100
80
60
40
20
0
(-)
(+)
L-Histidine
C
L-Isoleucine - + - + - +
cl
ncl
1.5 kb
D
ns
Cross-linked DNA (%)
100
80
60
40
20
0
(-)
(+)
L-Isoleucine
E
L-Serine - + - + - +
cl
ncl
1.5 kb
F
ns
Cross-linked DNA (%)
100
80
60
40
20
0
(-)
(+)
L-Serine

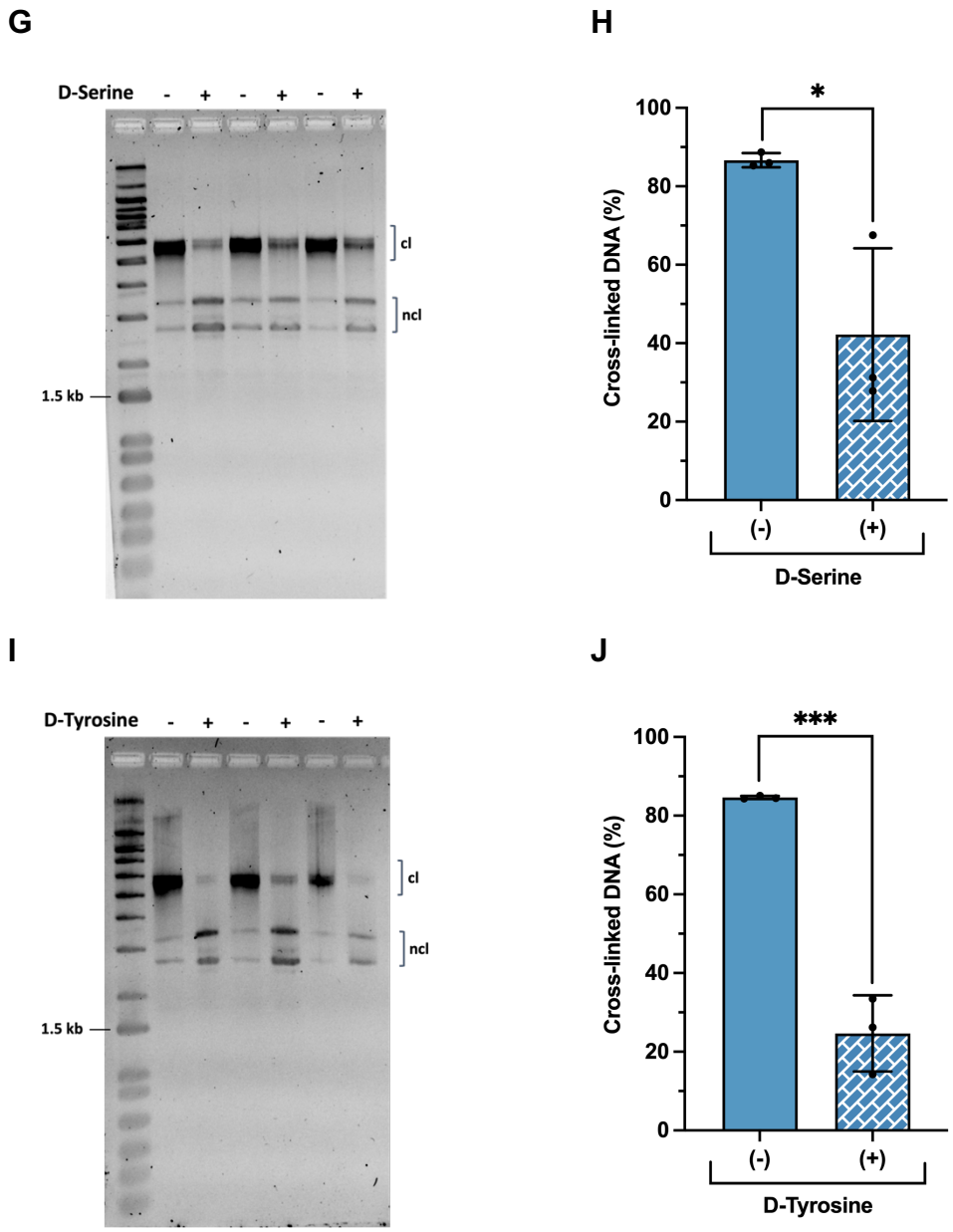


Figure 4-4 Assessment of the selected inhibitory amino acids against colibactin-driven DNA crosslinking. (**A**, **C**, **E**, **G**, **I**) Visualisation of cross-linked DNA after migration under alkaline gel electrophoresis. EcN_EO was cultured for 5 h in M9 alone (-) or M9 supplemented with 1 mM AAs (+) at 37°C, 200 rpm. Linearised pUC19 was then exposed to 1.5 x 10^6 CFUs of EcN_EO for 40 mins, in M9 alone or with the AA of interest. Plasmid DNA was then purified and crosslinking (cl) and non-crosslinking (ncl) activities were visualised by gel electrophoresis in alkaline conditions. The gel was run at 25 V for 45 mins, and then at 50 V for 150 mins; it was soaked in neutralisation buffer for 20 mins prior to image acquisition; the DNA size marker is the 1 kB DNA ladder (Invitrogen). (**B**, **D**, **F**, **H**, **J**) Percentage of cross-linked DNA relative to total DNA signal under different AA regimes. Signal intensities for both crosslinked (cl) and non-crosslinked (ncl) DNA were quantified by ImageJ, and the percentage of crosslinked DNA was calculated relative to the total DNA in each lane. Columns represent means +/- standard deviation calculated from three replicates. Statistical significance was assessed by a Student's t-test with * and *** respectively indicating $p < 0.05$ and $p < 0.001$.

Since D-Tyrosine's ability to reduce DNA damage was so striking, it was wondered whether in combination D-Serine and D-Tyrosine would act to further reduce DNA crosslinking. To evaluate this, linearised pUC19 was again incubated with EcN_EO for 5 h in M9 supplemented with 1 mM of both D-Serine and D-Tyrosine. Quantification by densitometry revealed that when combined, D-Serine and D-Tyrosine elicited a significant reduction, with crosslinking activity being 93.05% ($p < 0.0001$) lower than under no AA conditions (Figure 4-5).

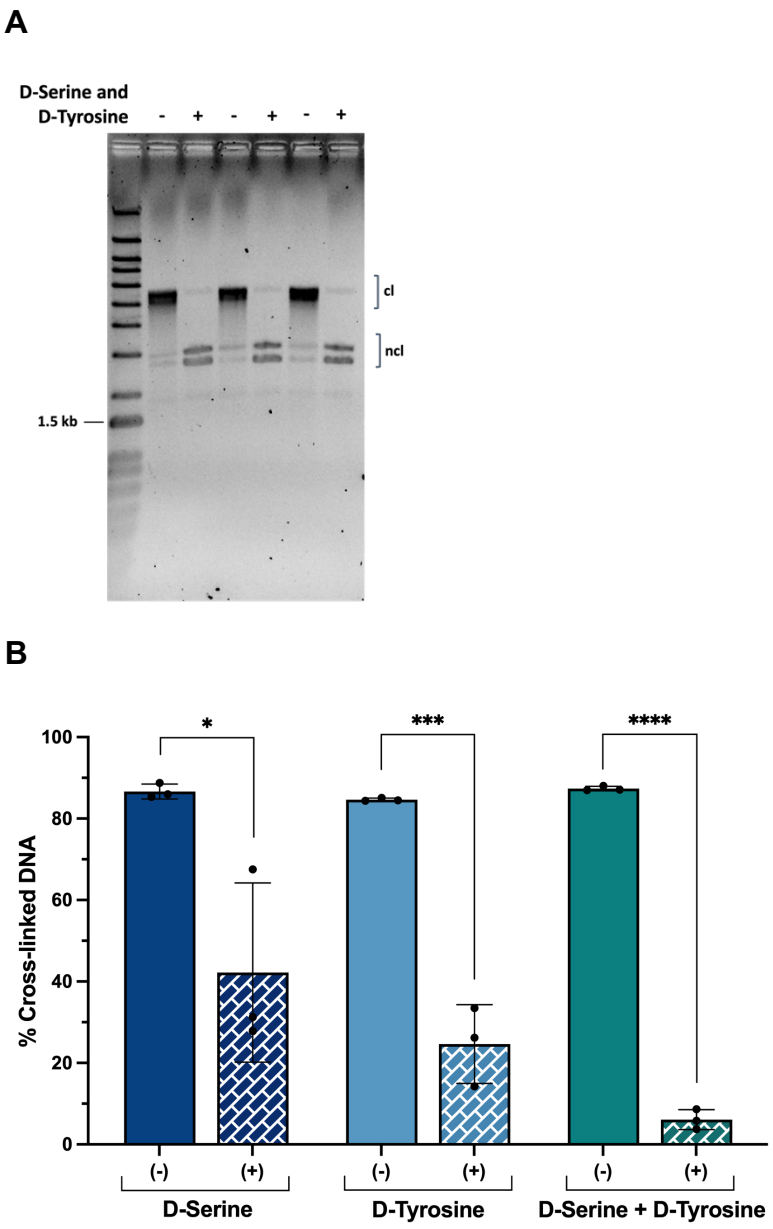


Figure 4-5 Assessment of the combinatorial effect of D-Serine and D-Tyrosine against colibactin-driven DNA crosslinking. (**A**) Visualisation of cross-linked DNA after migration under alkaline gel electrophoresis. EcN_EO was cultured for 5 h in M9 alone (-) or M9 supplemented with 1 mM D-Serine and 1 mM D-Tyrosine (+) at 37°C, 200 rpm. Linearised pUC19 was then exposed to 1.5 x 10^6 CFUs of EcN_EO for 40 mins in M9 alone or with the AA combination of interest. Plasmid DNA was then purified and crosslinking (cl) and non-crosslinking (ncl) activities were visualised by gel electrophoresis in alkaline conditions. The gel was run at 25 V for 45 mins, and then at 50 V for 150 mins; it was soaked in neutralisation buffer for 20 mins prior to image acquisition; the DNA size marker is the 1 kB DNA ladder (Invitrogen). (**B**) Percentage of cross-linked DNA relative to total DNA signal under different AA regimes. Signal intensities for both crosslinked (cl) and non-crosslinked (ncl) DNA were quantified by ImageJ, and the percentage of crosslinked DNA was calculated relative to the total DNA in each lane. Columns represent means +/- standard deviation calculated from three replicates. Statistical significance was assessed by a Student's t-test with * and *** respectively indicating $p < 0.05$, $p < 0.001$, and $p < 0.0001$. To note: the data represented for D-Serine and D-Tyrosine alone is the same that has been represented in Figure 4-4.

4.2.4 The combinatorial effect of D-Serine and D-Tyrosine on *clbB* transcription

Given the noticeable, additive effect that D-Serine and D-Tyrosine had on the reduction of colibactin-driven DNA crosslinking, it was important to determine whether this effect was also observed transcriptionally in colibactin-producing strains EcN_EO and CFT073_EO. To achieve this, the strains which had been previously transformed with the *clbB* transcription reporter (described in section 3.2.1), were cultured in MEM-HEPES, or in M9, in the presence or absence of 1 mM D-Serine, 1 mM D-Tyrosine, or 1 mM of both AAs. Both growth media were used in this instance as D-Tyrosine – unlike D-Serine which had consistently shown repressive activity on colibactin transcription – had downregulated *clbB* by reporter assay only in LB and M9.

Results in M9 showed that in the presence of D-Serine, D-Tyrosine and D-Serine plus D-Tyrosine, there was a reduced expression of *gfp* from the *clbB* promoter, with significant fold-decreases of 0.83 ($p = 0.0395$), 0.65 ($p = 0.0007$), and 0.68 ($p = 0.0012$) compared to no AA supplementation controls. In CFT073_EO the repressive effects of the AAs appeared to be stronger than in EcN_EO, with *clbB* promoter activity reduced by 0.62- ($p = 0.0158$), 0.60- ($p = 0.0115$) and 0.50- ($p = 0.0032$) fold by supplementation of D-Serine, D-Tyrosine, and D-Serine plus D-Tyrosine respectively. Unlike DNA cross-linking, no additive effects of combining the AAs were observed in either EcN_EO or CFT073_EO (Figure 4-6).

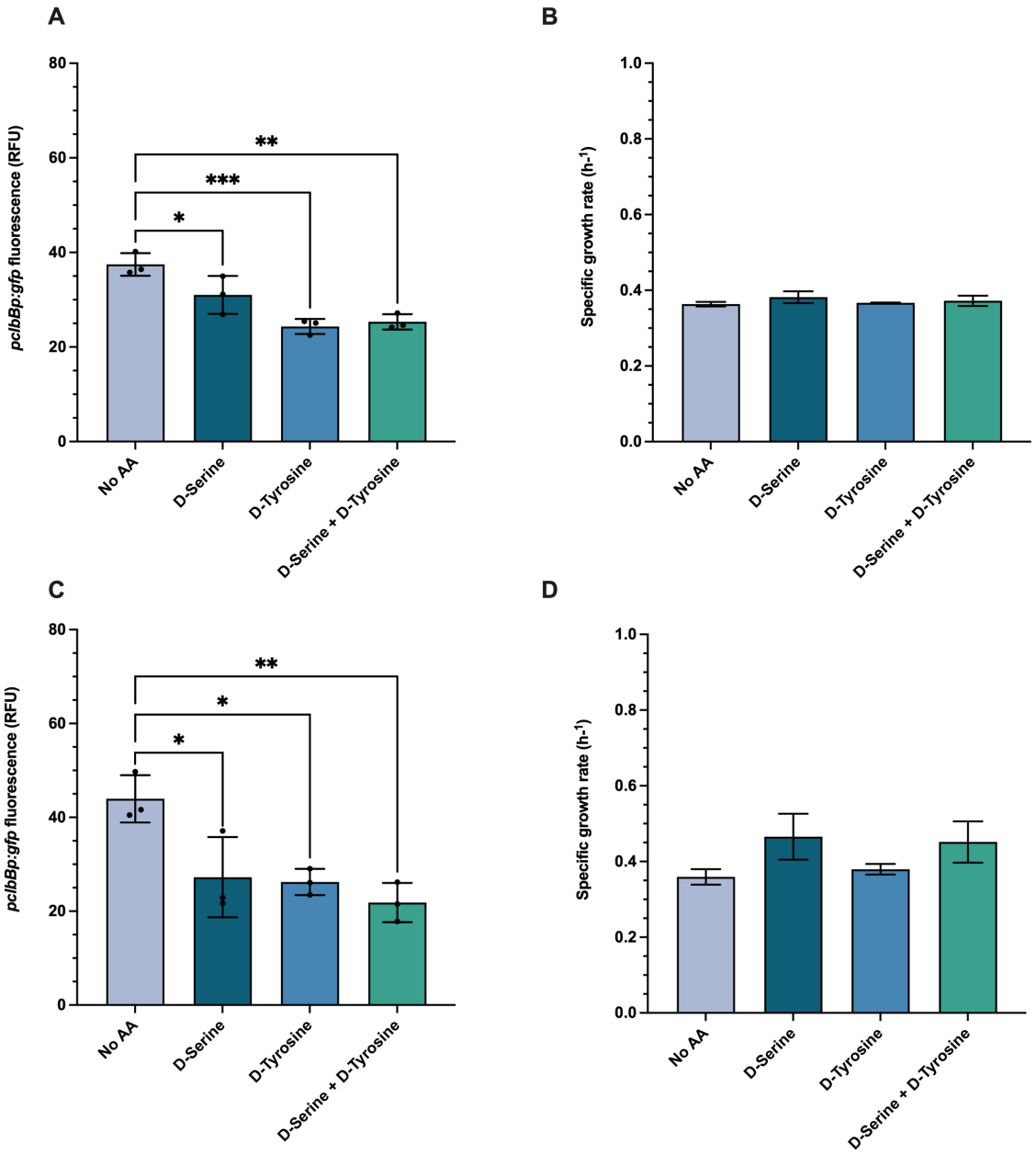


Figure 4-6 Assessment of the combinatorial effect of D-Serine and D-Tyrosine on *clbB* promoter expression in M9. (**A** and **C**) The *clbBp:gfp* reporter activity of EcN_EO (A) and CFT073_EO (C) in M9 supplemented with individual or combined D-Serine and D-Tyrosine. EcN_EO and CFT073_EO were cultured in M9 alone, or in M9 supplemented with 1 mM of D-Serine, 1 mM of D-Tyrosine, or 1 mM of each of the two AAs combined. Bacteria were cultured in 50 mL falcon tubes at 37°C, 200 rpm. Fluorescence was measured hourly over a 3 to 8-hour period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *clbBp:gfp*/$OD_{600\ nm}$ in exponential phase, at 7 hours. Statistical significance was assessed from three replicates by a Dunnett's multiple comparison test with *, **, ***, **** indicating significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively. (**B** and **D**) Growth curves of EcN_EO (B) and CFT073_EO (D) in M9 supplemented with individual or combined D-Serine and D-Tyrosine. The strains were cultured as described above. Absorbance ($OD_{600\ nm}$) was measured hourly between 3 and 8 hours and values were plotted on a log-10 scale. Data points represent average absorbance values and error bars indicate standard deviation calculated from three replicates.

Conversely, results in MEM-HEPES showed that D-Tyrosine failed to repress *clbB* transcription in both EcN_EO and CFT073_EO, whilst actually provoking slight fold-increases of 1.06 and 1.01 respectively. In contrast, D-Serine alone induced a 0.69-fold reduction in EcN_EO ($p < 0.0001$) and a 0.66-fold reduction in CFT073_EO ($p = 0.0009$). In combination, the two AAs elicited an 0.79-fold decrease ($p = 0.0007$) and an 0.63-fold decrease ($p = 0.0006$) of *clbB* promoter activity in EcN_EO and CFT073_EO respectively. Given that D-Tyrosine showed no effect in MEM-HEPES, one can conclude that the activity seen in combinatorial conditions is largely driven by D-Serine (Figure 4-7).

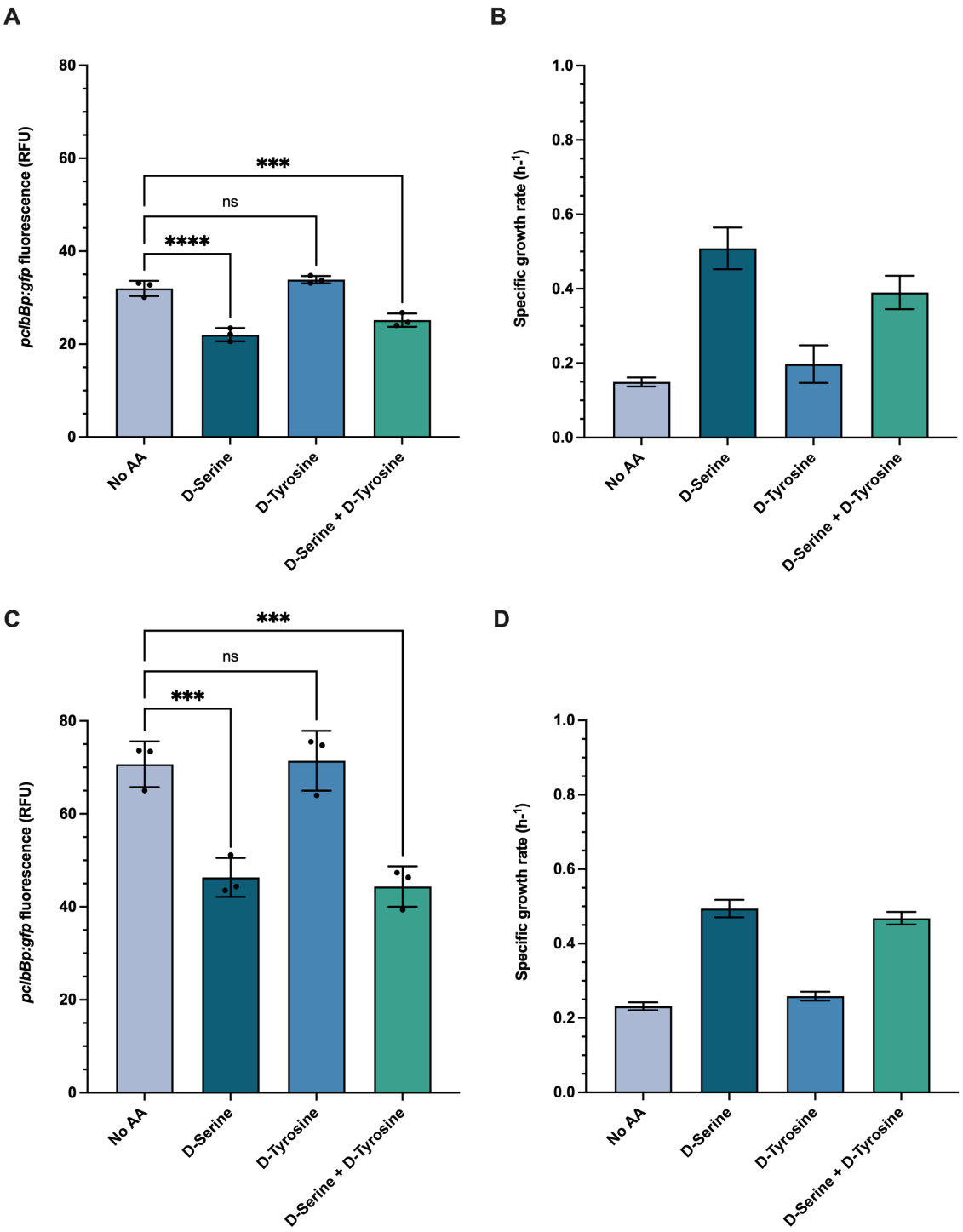


Figure 4-7 Assessment of the combinatorial effect of D-Serine and D-Tyrosine on *clbB* promoter expression in MEM-HEPES. (**A** and **C**) The *clbBp:gfp* reporter activity of EcN_EO (A) and CFT073_EO (C) in MEM-HEPES supplemented with individual or combined D-Serine and D-Tyrosine. EcN_EO and CFT073_EO were cultured in MEM-HEPES alone, or in MEM-HEPES supplemented with 1 mM of D-Serine, 1 mM of D-Tyrosine, or 1 mM of each of the two AAs combined. Bacteria were cultured in 50 mL falcon tubes at 37°C, 200 rpm. Fluorescence was measured hourly over a 3 to 8-hour period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *clbBp:gfp*/$OD_{600\ nm}$ in exponential phase, at 7 hours. Statistical significance was assessed from three replicates by a Dunnett's multiple comparison test with *, **, ***, **** indicating significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively. (**B** and **D**) Growth curves of EcN_EO (B) and CFT073_EO (D) in MEM-HEPES supplemented with individual or combined D-Serine and D-Tyrosine. The strains were cultured as described above. Absorbance ($OD_{600\ nm}$) was measured hourly between 3 and 8 hours and values were plotted on a log-10 scale. Data points represent average absorbance values and error bars indicate standard deviation calculated from three replicates.

4.2.5 Transcriptional responses of EcN_EO to D-Serine, D-Tyrosine, and their combined supplementation in M9

To further investigate the effects elicited by D-Serine and D-Tyrosine, alone or in combination, the individual and combinatorial effect of the two AAs on gene expression was tested by RT-qPCR. Cultures of EcN_EO grown in M9 and supplemented with 1 mM of D-Serine, 1 mM D-Tyrosine, or 1 mM of both AAs were harvested for RNA extraction. Expression of *clbB* was assessed by RT-qPCR relative to the housekeeping gene *gapA* and measurements were normalised to cultures grown in the absence of AA supplementation. All media supplementations displayed significant inhibitory activity on *clbB* expression and are listed as follows, with reduction of *clbB* expression and significance indicated in parentheses: D-Serine (0.88-fold decrease, $p = 0.0443$), D-Tyrosine (0.67-fold decrease, $p = 0.0318$), and D-Serine plus D-Tyrosine (0.49-fold decrease, $p = 0.0067$); (Figure 4-8).

These results show that at the transcriptional level, the combinatorial effect of D-Serine plus D-Tyrosine is stronger than the effect of either of the two AAs alone, corroborating the findings reported for the DNA crosslinking assays. These data, alongside the reporter assays results, suggest that D-Tyrosine's repression is only observed when cells are grown in M9. Whilst no additive effect was measured using the transcriptional reporter to GFP, a significant difference was observed using RT-qPCR and at a functional level via DNA crosslinking assays. These distinctions furthermore suggest that D-Serine and D-Tyrosine are likely to reduce the activity of colibactin through different pathways.

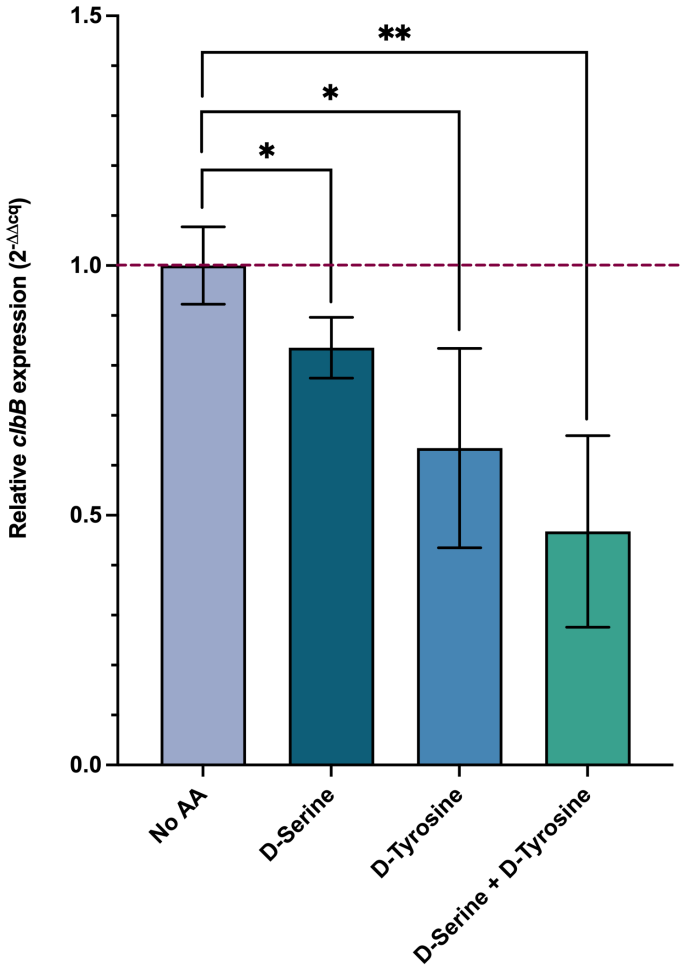


Figure 4-8 The combinatorial effect of D-Serine and D-Tyrosine on *clbB* expression screened by RT-qPCR. EcN_EO was grown in 50 mL falcon tubes at 37°C, 200 rpm for 7 h in M9 supplemented with 1 mM D-Serine, 1 mM of D-Tyrosine, or 1 mM of each of the two AAs combined. RNA was extracted from bacterial cultures and converted to cDNA. The expression of *clbB* was measured relative to the housekeeping gene *gapA*, using a control condition of no amino acid supplementation as the calibrator for baseline expression, as indicated by the dashed red line. Statistical significance was determined from four replicates through a Student's t-tests test with *, **, ***, **** indicating significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively.

To elucidate which pathways might be affected by D-Serine, D-Tyrosine, and their combination, the same RNA samples – harvested and tested in parallel for RT-qPCR – were also sent for RNA sequencing, to more broadly understand the impact of these AAs on gene expression. It should be noted that the fold-changes described in the following paragraphs relate to Log_2 fold-changes. Furthermore, only the genes that met a false-discovery rate (FDR)-corrected *P* value of ≤0.05 are included in this analysis. The only exception to this were the *clb* genes which were always analysed and plotted. All significantly differentially expressed genes are compiled in tables 10-1, 10-2 and 10-3 found in the appendix.

Transcriptomics analyses from EcN_EO cultured in M9 with 1 mM D-Serine revealed minor shifts in gene expression compared to control growth conditions in M9 alone, with only 7 genes significantly downregulated, and only 26 genes significantly upregulated. The function of the significantly up- and down- regulated genes clearly indicated the activation of a chemotaxis response. The *btsT* gene, responsible for encoding the pyruvate/proton symporter BtsT, was the most downregulated gene (-3.22 fold change). When nutrients are limited, such as in a minimal growth medium like M9, BtsT is required to take up pyruvate from the medium for use as a carbon source by the bacteria. Hence it is unsurprising that in the presence of D-Serine – which is metabolised into ammonia and pyruvate by the bacterial cell – *btsT* expression is no longer required necessary and expression of the gene was downregulated. In contrast, the following genes associated with chemotaxis were upregulated, with their respective fold changes indicated in parentheses: *tar* (0.73), *tsr* (0.74), *cheR* (0.63), *cheZ* (0.67), *cheY* (0.71), *cheB* (0.71), and *cheW* (0.71). Briefly, chemotaxis depends on the methylation and demethylation of membrane-bound chemotaxis receptors, such as Tar and Tsr by the enzymes CheR and CheB. Methylation causes the receptors to dimerise and associate with CheW, a linker protein, and CheA, a histidine kinase that autophosphorylates and transfers the phosphoryl group to CheY, the overall response regulator. CheY then binds to flagellar motors, before dephosphorylation by CheZ (Hansen, Endres and Wingreen, 2008).

In accordance with the process of chemotaxis, flagellar genes were also upregulated in the presence of D-Serine, with their respective fold changes indicated in parentheses: *flgN* (0.65), *fliT* (0.70), *fliZ_1* (0.76), *fliZ_3* (0.75), *fliA*

(0.76), *flgL* (0.82), *fliD* (0.83), *fliS* (0.86), *ycgR* (0.87), and *flgK* (0.91). Interestingly, none of the genes involved in colibactin biosynthesis were significantly downregulated, but – upon closer examination of their fold-change – it was observed that all *clb* genes, albeit non-significantly, are downregulated in the presence of 1 mM D-Serine, with the exception of *clbA* (the PPTase responsible for activating enzymes involved in colibactin biosynthesis), and *clbR* (the transcriptional regulator of the colibactin biosynthetic gene cluster), which were upregulated (Figure 4-9).

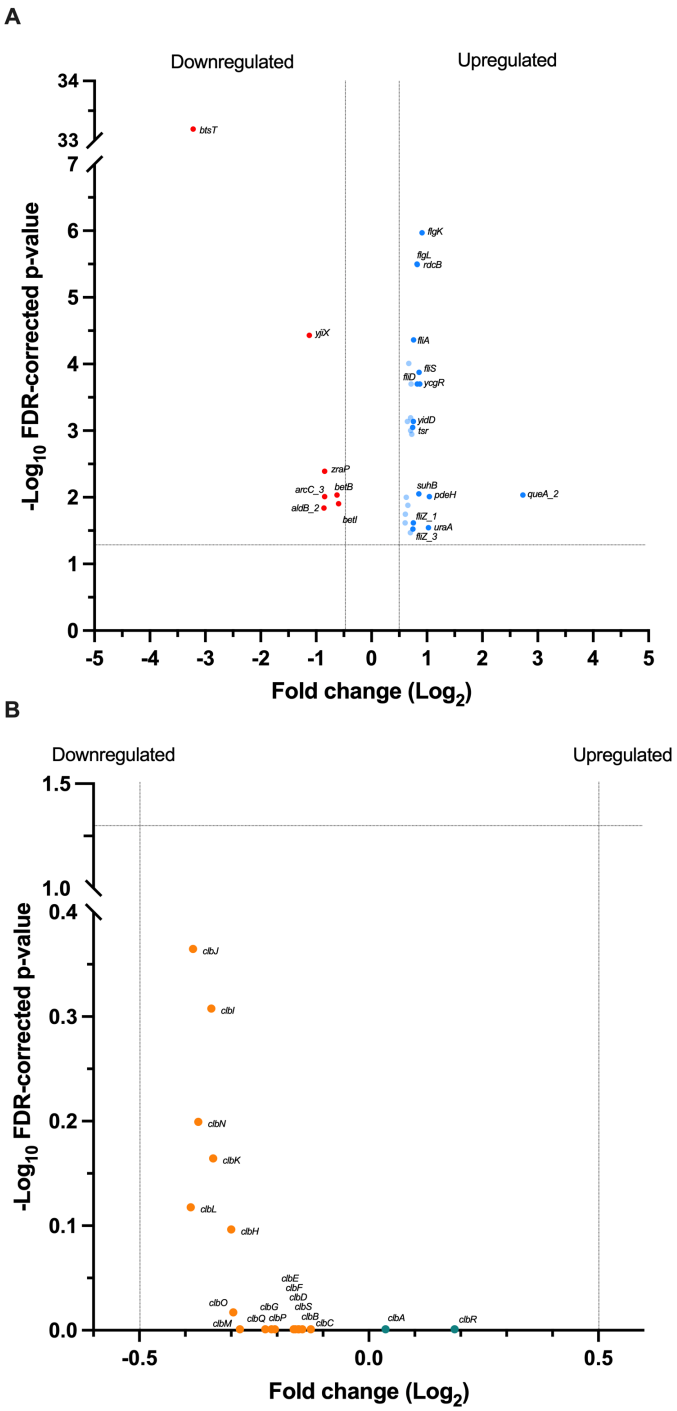


Figure 4-9 Transcriptomic profile of EcN_EO cultured in M9 supplemented with 1 mM D-Serine. EcN_EO was cultured in M9 with 1 mM D-Serine at 37°C, 200 rpm. Samples were harvested at 7 h, RNA was extracted and checked for DNA contamination and integrity before being sent for ribosomal depletion and cDNA library preparation. RNA sequencing was performed to a depth of 100 bp single-end reads using a NextSeq2000 sequencing system (Illumina). Data analysis was performed using CLC Genomics Workbench version 21.0.5 (Qiagen), with reads mapped to the EcN reference genome (NCBI accession number NZ_CP007799.1). The Differential Expression for RNA-seq tool in the CLC Genomics Workbench was used to calculate differential gene expression. Genes were considered significantly differentially expressed with a fold change of ≥1.5 or ≤1.5 and a false-discovery rate (FDR)-corrected *P* value of ≤0.05. (**A**) Volcano plot representing differentially expressed genes. All significantly differentially upregulated (blue) and downregulated (red) genes are plotted, but only the 15 most significant ones are labelled. (**B**) Volcano plot representing differential expression of colibactin biosynthesis genes. Both non-significantly downregulated genes (orange) and non-significantly upregulated genes (teal) are plotted; dashed lines represent thresholds of significance.

Transcriptomics analyses from EcN_EO cultured in M9 with 1 mM D-Tyrosine, revealed significant downregulation of 38 genes and significant upregulation of 55 genes, compared to control cells grown in M9 alone. Out of the 19 *clb* genes, 14 were significantly downregulated; *clbB*, *clbC*, and *clbS* were non-significantly downregulated, and *clbA* and *clbR* were once again upregulated. This was interesting, given that these are the only two genes orientated in the opposite DNA strand to the rest on the colibactin biosynthetic gene cluster.

In analysing this transcriptomics data set, no particular metabolic pathways were highlighted although the change in expression of some genes is worth noting. The two most downregulated genes were *csgB* (-2.48 fold change) and *csgA* (-2.12), encoding for the two subunits that constitute curli filaments, which are, in turn, amyloid aggregates involved in host cell adhesion and biofilm formation. Given that the literature has reported that D-Tyrosine is capable of disrupting biofilm formation (Hochbaum *et al.*, 2011), downregulation of *csg* genes is not surprising. It was also interesting to note upregulation of the following gene family, encoding outer membrane proteins and porins, with their respective fold-changes in parentheses: *ompX* (1.00), *ompC_1* (1.55*)*, *ompC_2* (1.23), *ompA* (1.41), *ompT* (1.53), *ompW* (2.11), *ompF* (2.23). Outer membrane proteins and porins act as channels through which molecules can diffuse. In this instance, given that Tyrosine is 181 Da in weight and that, on average, substances smaller than 600 Da can diffuse through porins (Novikova and Solovyeva, 2009), this gene expression profile could reflect an adaptation to increase the uptake mechanism of D-Tyrosine by bacterial cells. Beyond simple transport though, outer membrane proteins, like OmpX, promote adhesion to eukaryotic cells. This is interesting to note as it might explain the failure to observe a significant reduction in cellular senescence even though D-Tyrosine reduces colibactin expression (though, as described further on, this might also have been due to infections being performed in MEM-HEPES).

Given that tyrosine is classified as weakly acidic among proteinogenic AAs, it was surprising to find that genes encoding for acid-activated periplasmic chaperones, *hdeA* and *hdeB*, were significantly downregulated by growth in its presence (-1.48 fold and -1.46 fold respectively). Genes encoding for other acid-resistance proteins, like *hdeD* (-1.07) and *gadE* (-0.87) were also downregulated. Other potential pathways of interest that emerged from the transcriptomics data relate

to sulphur metabolism, with significant changes in the expression of *ssuB*_1 (-0.87), *thiS* (-0.76), and *trhO* (0.61); and to nitrogen metabolism, with significant changes in the expression of *arcC_3* (-1.10), *glnG* (-0.63), and *narZ_1* (-0.61); (Figure 4-10).

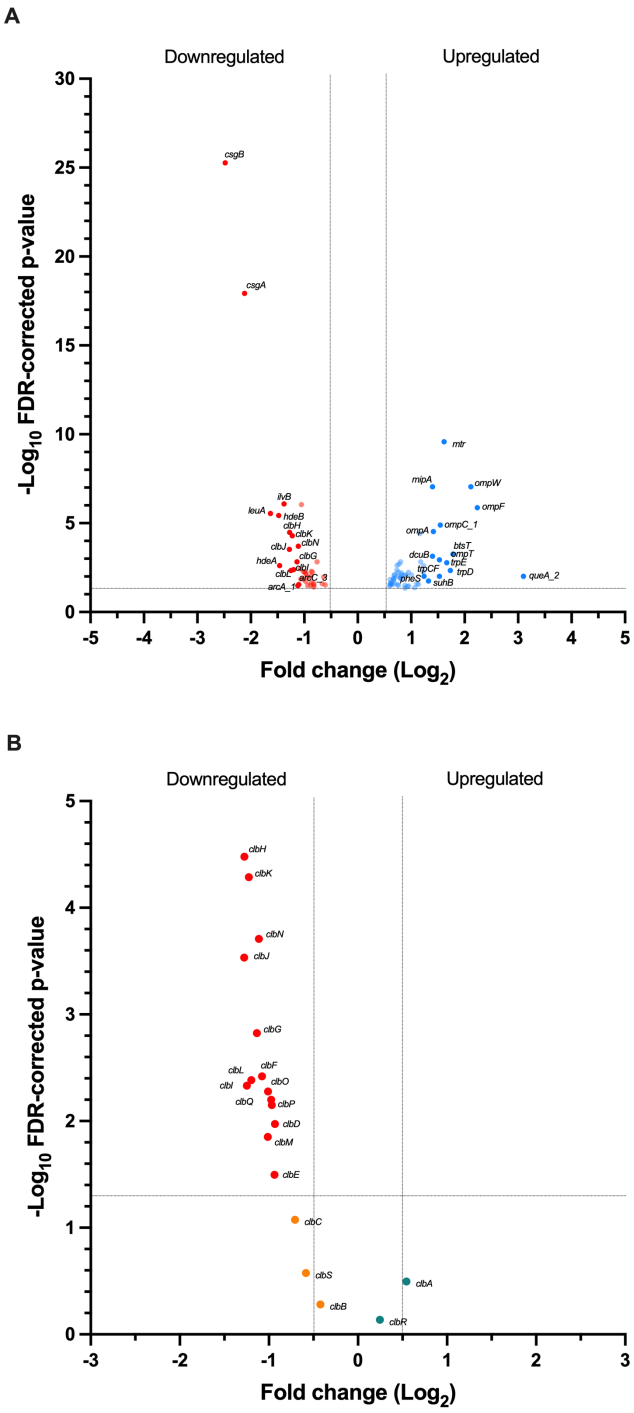


Figure 4-10 Transcriptomic profile of EcN_EO cultured in M9 supplemented with 1 mM D-Tyrosine. EcN_EO was cultured in M9 with 1 mM D-Tyrosine at 37°C, 200 rpm. Samples were harvested at 7 h, RNA was extracted and checked for DNA contamination and integrity before being sent for ribosomal depletion and cDNA library preparation. RNA sequencing was performed to a depth of 100 bp single-end reads using a NextSeq2000 sequencing system (Illumina). Data analysis was performed using CLC Genomics Workbench version 21.0.5 (Qiagen), with reads mapped to the EcN reference genome (NCBI accession number NZ_CP007799.1). The Differential Expression for RNA-seq tool in the CLC Genomics Workbench was used to calculate differential gene expression. Genes were considered significantly differentially expressed with a fold change of ≥1.5 or ≤1.5 and a false-discovery rate (FDR)-corrected *P* value of ≤0.05. (**A**) Volcano plot representing differentially expressed genes. All significantly differentially upregulated (blue) and downregulated (red) genes are plotted, but only the 15 most significant ones are labelled. (**B**) Volcano plot representing differential expression of colibactin biosynthesis genes. Significantly downregulated (red), non-significantly downregulated (orange), and non-significantly upregulated genes (teal) are plotted; dashed lines represent thresholds of significance

As expected, the transcriptomics data from EcN_EO cultured in M9 with 1 mM of both D-Serine and D-Tyrosine, revealed the largest change in gene expression. In total 101 genes were downregulated and 194 were upregulated, with several of these overlapping with the gene variations described in the previous data sets. With the double AA supplementation, all *clb* genes were significantly downregulated, excepting *clbB* – which fell beneath the significance threshold – and *clbA* and *clbR* which were, as always, non-significantly upregulated. The most upregulated genes included those encoding outer membrane proteins and porins (more specifically *ompF*, *ompC_1*, *ompT*, and *ompA*), those encoding for bacterial flagella (in particular *flgB* and *flgC*), and those encoding for phenylalanine tRNA ligase subunits (*pheS* and *pheT*). Conversely, the most downregulated genes comprised curli encoding subunits (*csgA* and *csgB*), several colibactin biosynthesis genes, and – exclusively in this combinatorial culturing condition – *yhaV*; encoding for the type II toxin-antitoxin system ribonuclease toxin. The presence of both AAs broadly displayed the transcriptomic effect of both AAs individually, and mostly altered expression of genes involved in movement and adhesion (Figure 4-11).

In examining the genes whose expression was universally altered across all three culturing conditions, the carbamate kinase encoding gene *arcC* was consistently significantly downregulated. The following genes, with product indicated in parentheses, instead showed constant upregulation: *eptA* (phosphoethanolamine transferase EptA), *yidD* (Membrane protein insertion efficiency factor YidD), *suhB* (inositol-1-monophosphatase), *uraA* (uracil permease), and *queA* (tRNA preQ1(34) S-adenosylmethionine ribosyltransferase-isomerase QueA), which was the consistently most upregulated gene.

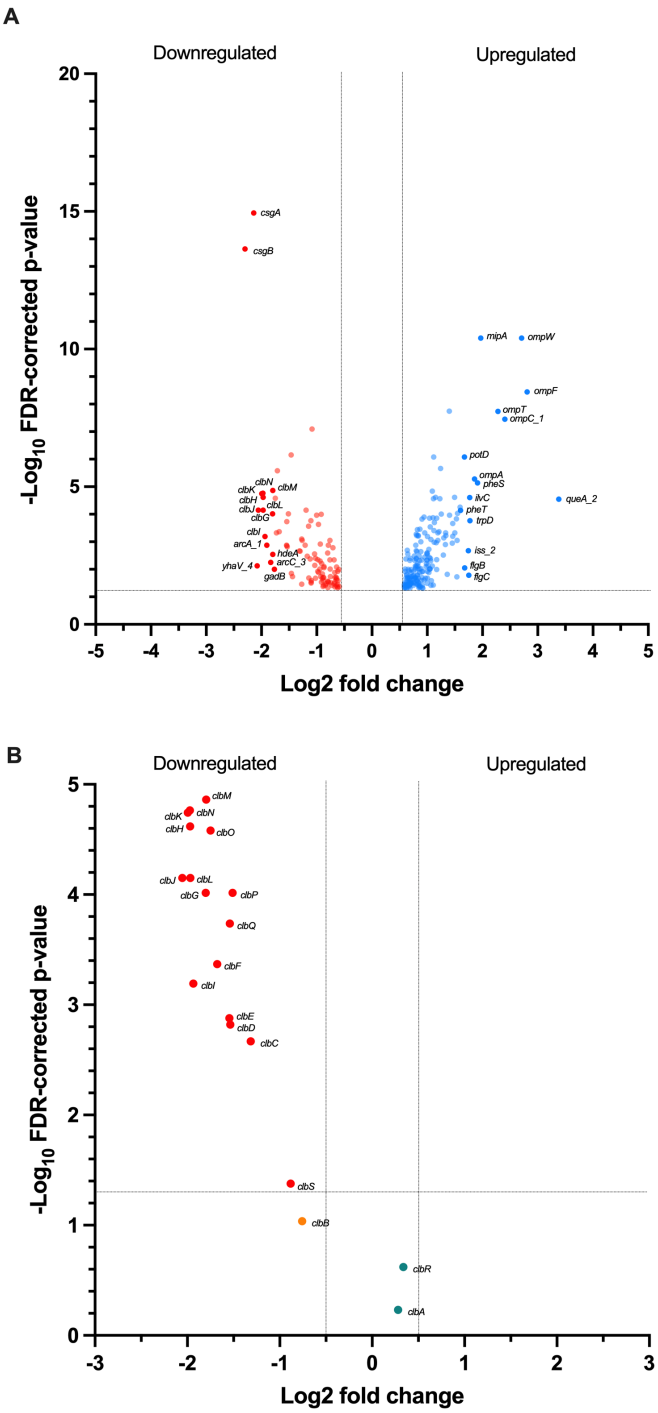
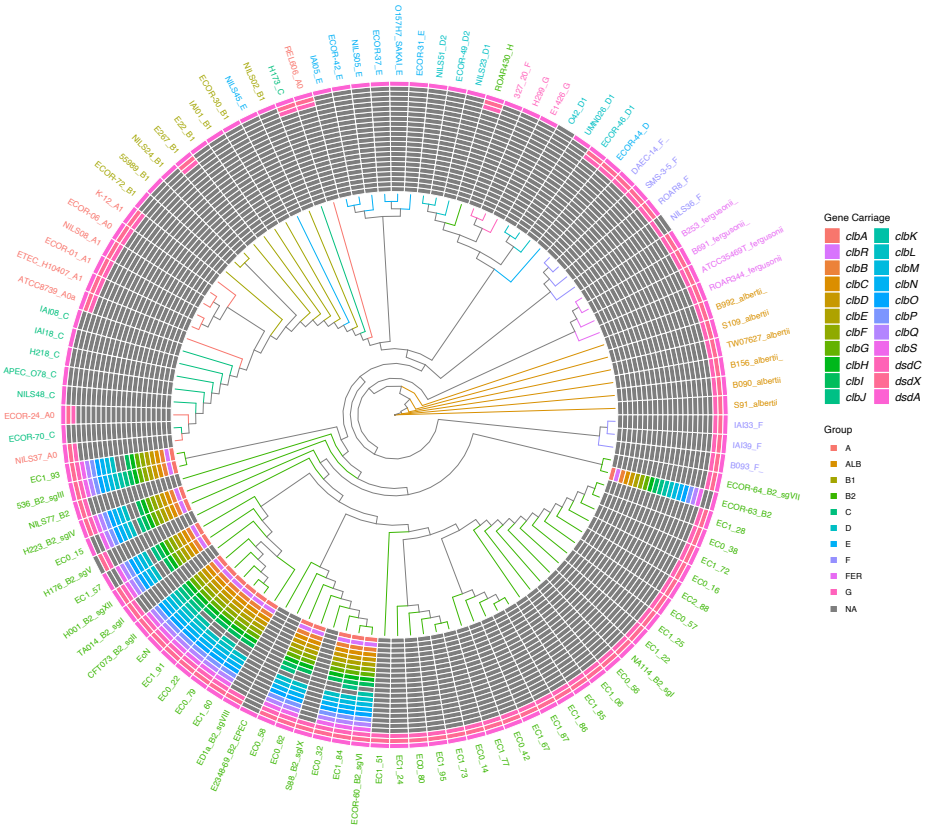


Figure 4-11 Transcriptomic profile of EcN_EO cultured in M9 supplemented with 1 mM D-Serine and D-Tyrosine. EcN_EO was cultured in M9 with 1 mM D-Serine and 1 mM D-Tyrosine at 37°C, 200 rpm. Samples were harvested at 7 h, RNA was extracted and checked for DNA contamination and integrity before being sent for ribosomal depletion and cDNA library preparation. RNA sequencing was performed to a depth of 100 bp single-end reads using a NextSeq2000 sequencing system (Illumina). Data analysis was performed using CLC Genomics Workbench version 21.0.5 (Qiagen), with reads mapped to the EcN reference genome (NCBI accession number NZ_CP007799.1). The Differential Expression for RNA-seq tool in the CLC Genomics Workbench was used to calculate differential gene expression. Genes were considered significantly differentially expressed with a fold change of ≥1.5 or ≤1.5 and a false-discovery rate (FDR)-corrected *P* value of ≤0.05. (**A**) Volcano plot representing differentially expressed genes. All significantly differentially upregulated (blue) and downregulated (red) genes are plotted, but only the 15 most significant ones are labelled. (**B**) Volcano plot representing differential expression of colibactin biosynthesis genes. Significantly downregulated (red), non-significantly downregulated (orange), and non-significantly upregulated genes (teal) are plotted; dashed lines represent thresholds of significance.

4.2.6 Validation of the downregulatory effect of D-Serine on colibactin expression in prototypical strains and clinical isolates

When considering both *in vitro* and *in cellulo* assays, the AA that consistently stood out as having repressive activity against colibactin was D-Serine As further confirmation, of its broader inhibitory activity, D-Serine was tested by reporter assay against several *pks*+ clinical isolates. A total of 33 clinical isolates and their genomes were obtained from a collaborator and tested for carriage of the colibactin biosynthesis *clb* gene cluster and of the D-Serine tolerance locus, *dsdCXA* (Figure 4-12). Screening was done through the PhyloBuild pipeline that is available on GitHub (https://github.com/samandmac/PhyloBuild). Of the 11 strains that had an almost complete *clb* cluster, 9 were selected for further testing. Most of these strains seemed to either lack or possess incomplete copies of *clbJ* and *clbK* with the exception of EC0_79 and EC1_91 that have *clbK*, but not *clbJ*; and EC1_84, EC1_57, EC0_58, and EC0-15 that have *clbJ*, but not *clbK*. All strains – even the ones lacking the *clb* operon – carried a complete *dsdCXA* locus, except for EC1_57 and EC0_15, that both lacked *dsdC* and *dsdX*. The strains selected for validating D-Serine were all transformed with the reporter plasmid (*pclbB:gfp*) as described previously. Interestingly, all these strains displayed a significant reduction in *clbB* promoter activity when grown in the presence of 1 mM D-Serine. The fold-change reductions of gene expression observed for each of the strains, with significance indicated in parentheses, were as follows: EC0-32 (0.47, p = 0.000736), EC0_22 (0.61, p = 0.026377), EC0_62 (0.62, p = 0.001726), EC0_79 (0.56, p = 0.000906), EC1_93 (0.53, p = 0.02282), EC1_91 (0.62, p = 0.014831), EC0_58 (0.52, p = 0.000623), EC1_57 (0.65, p = 0.045144), and EC1_60 (0.78, p = 0.048165). Interestingly, all strains displayed an increase in growth rate when cultured in the presence of 1 mM D-Serine, even EC1_57 that lacked a complete *dsdCXA* locus (Figure 4-13).

A



B

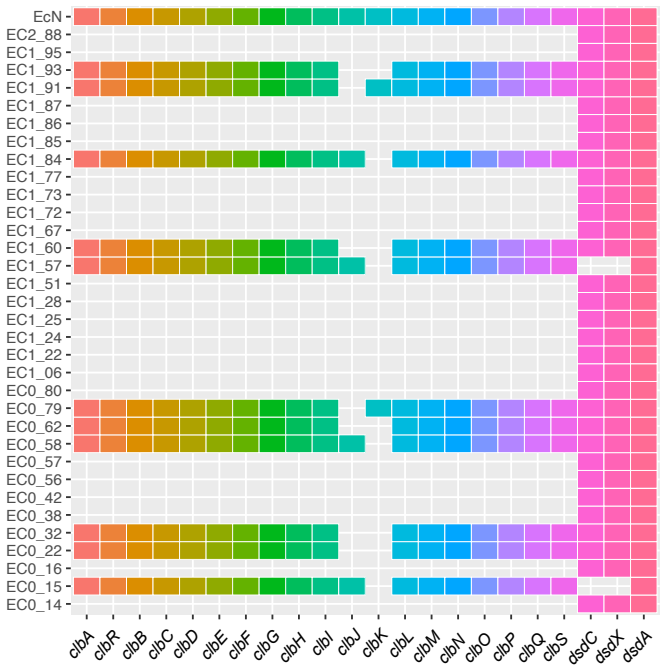


Figure 4-12 **Phylogenetic tree and heatmap displaying carriage of the *clb* operon and of the *dsd* locus in *pks*+ clinical isolates.** The presence of the *clb* operon and of the *dsd* tolerance locus was assessed in 33 B2 *E. coli* clinical strains following the steps illustrated in the PhyloBuild pipeline (https://github.com/samandmac/PhyloBuild).

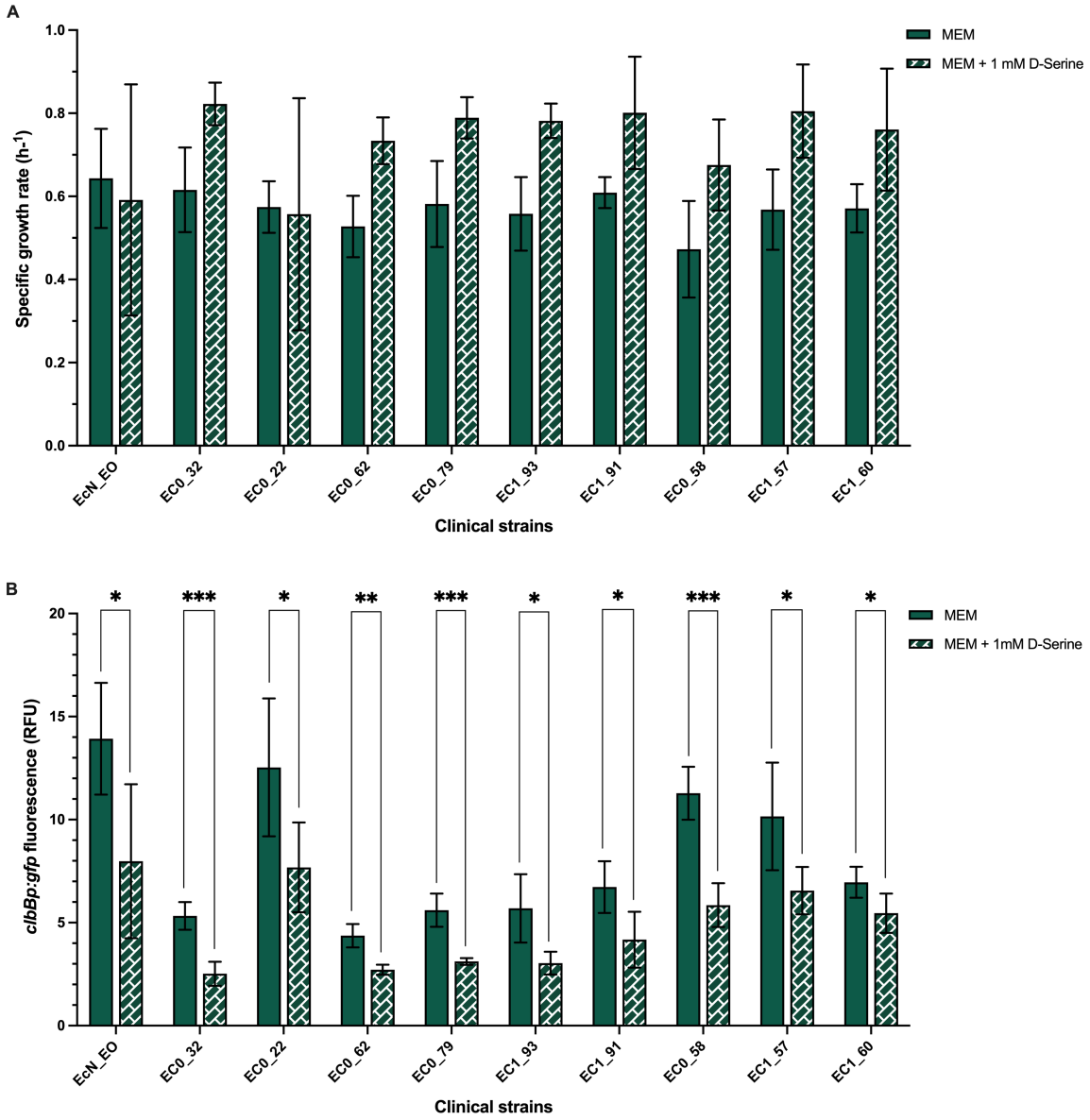


Figure 4-13 Assessment of the effect of D-Serine on *clbB* promoter expression in *pks+* clinical strains. (**A**) The specific growth rates of clinical isolates in MEM-HEPES alone and supplemented with 1 mM D-Serine. Clinical strains were transformed with reporter gene fusion plasmid, *clbBp:gfp* and with an empty vector for control. Bacteria were cultured in MEM-HEPES with 25 µg/mL chloramphenicol, and with or without supplementation of 1 mM D-Serine at 37°C, 200 rpm in 96-well plates. Absorbance ($OD_{600\ nm}$) was measured hourly and the specific growth rate ($\mu = \Delta \ln OD_{600nm}/\Delta t$) was calculated in exponential phase. Columns represent means and error bars indicated standard deviation, calculated from at least three independent biological replicates. (**B**) The *clbB:gfp* reporter activity of clinical isolates in MEM-HEPES alone and supplemented with 1 mM D-Serine. Clinical strains were cultured as described above. Fluorescence was measured hourly over an 8-hour period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *clbBp:gfp*/$OD_{600\ nm}$ at 4 h. Statistical significance was assessed from at least three replicates by a Dunnett's multiple comparison test with *, **, ***, **** indicating significance, $p < 0.05$, < 0.01, < 0.001, and < 0.0001 respectively.

4.3 Discussion

The aim of this chapter was to validate L-Histidine, L-Isoleucine, L-Serine, D-Serine, and D-Tyrosine – AAs that in Chapter 3 had displayed downregulatory activity on colibactin – through *in cellulo* experiments and DNA crosslinking assays. Out of these, only D-Serine consistently reduced colibactin activity in all assays performed. L-Histidine significantly reduced the DNA crosslinking elicited by EcN_EO, but had no effect on colibactin-driven cellular senescence. L-Isoleucine failed to reduce colibactin activity in both assays. L-Serine, was incapable of reducing the percentage of crosslinked DNA *in vitro,* but significantly reduced cellular senescence. D-Tyrosine, prompted the most robust reduction of DNA crosslinking, but had no effect in M9 and *in cellulo*. Taken together, these results demonstrate that D-Serine appears unique in its capacity to modulate colibactin expression.

It's curious to note that while L-Serine had a significantly ameliorative effect on HeLa cells' megalocytosis, it did not reduce DNA crosslinking. A possible explanation for that is that L-Serine has a unique and prominent role in cell proliferation. It has been shown to be required for optimal cell growth, managing by itself to replace complete mixtures of nutritionally nonessential AAs in cell culture media (Eagle, 1958). Hence, the reduction in megalocytosis seen with L-Serine could be a result of its effect on HeLa cells, rather than on EcN's colibactin expression.

Conversely, D-Tyrosine was unique in eliciting colibactin downregulation only in M9 medium. As previously mentioned, D-Tyrosine is known to be involved in disrupting biofilm formation (Hochbaum *et al.*, 2011) (Yu *et al.*, 2016), and EcN is known to be able to form biofilms, considered one of its fundamental characteristics for eliciting probiotic effects (Chen *et al.*, 2017). While D-Tyrosine prompted the most robust repression of colibactin transcription in M9, both by ISCLA (Figure 4-4) and reporter assay (Figure 4-6), it had no effect in MEM-HEPES (Figure 4-7). This likely explains why no effects on HeLa megalocytosis were observed *in cellulo* in the presence of D-Tyrosine. Differences in colibactin expression upon cultivation in varied types of media have not only been reported in this project, but also in the literature with the *clbR* promoter activity, for example, known to be higher in M9, compared to richer media (Wallenstein *et al.*,

2020). To summarise, the data in this chapter demonstrated that D-Serine and D-Tyrosine both elicit repression of colibactin activity, but, while D-Tyrosine was able to do so more strongly in M9 media, D-Serine was able to do so more comprehensively.

Upon comparing L-Serine and D-Serine, it's interesting to speculate on the mechanism underlying only D-Serine's suppression of colibactin activity, given that L-Serine and D-Serine are both metabolised into ammonia and pyruvate by *E. coli*. The differences in activity could either lie in the isomeric forms of the breakdown products from each AA, or in the different genes that are involved in L- and D-Serine degradation and that consequently activate different biochemical pathways. Stark differences in the outcomes deriving from different chiral forms of the same molecule is a well-established phenomenon, with the most common example probably being that of chiral drugs (Brooks, Guida and Daniel, 2011).

One hypothesis explaining the effect of D-Serine on colibactin expression is that this AA is activating a stringent response mechanism, whereby an abundance of D-Serine induces a stress response in the bacterial cell. However, while D-Serine is toxic at high concentrations – as displayed by the harmful effect that concentrations above 0.2 M had on EcN's growth – when used at the described concentration of 1 mM it would appear to boost growth in MEM-HEPES (Figure 3-1). Interestingly, while CFT073 displayed enhanced growth in both MEM-HEPES and M9 when culture media were supplemented with 1 mM D-Serine, EcN exhibited a growth increase in response to D-Serine only in MEM-HEPES, and not in M9, where instead it grew similarly to control conditions (Figure 4-6 and 4-7). Furthermore, while D-Serine is toxic to *E. coli* strains that lack the D-Serine metabolising *dsdCXA* locus – such as EHEC (Moritz and Welch, 2006), which activates its SOS response in the presence of D-Serine (Connolly *et al.*, 2015) – this is not the case for EcN, which possesses a complete *dsdCXA* gene cluster. Interestingly, even EC1_57, which lacks both *dsdC* and *dsdX*, displays an increased growth rate in the presence of 1 mM D-Serine (Figure 4-13). This could potentially be explained by the presence of a different, non-specific transporter like CycA that, coupled with DsdA (D-Serine deaminase), might still facilitate the breakdown of D-Serine, or by the fact that GlyA can also catalyse the conversion of D-Serine to ammonium and pyruvate. Lastly, transcriptomics data revealed that EcN had an obvious

chemotaxis response to D-Serine, in line with the literature demonstrating *E. coli*'s ability to use D-Serine as a sole carbon source (Bloom and McFall, 1975), and none of the significantly differentially expressed genes corresponded to ones involved in the stringent response.

It is also interesting that D-Serine represses colibactin transcription in the tested *pks+ E. coli* belonging to the to the B2 phylogroup. This includes strains designated as UPEC, which live in a D-Serine rich environment (Anfora *et al.*, 2007). Interestingly, *pks+ E. coli* have been implicated as a risk factor for both colorectal and bladder carcinogenesis (Chagneau *et al.*, 2021). Previously D-Serine has been reported to repress transcription of the T3SS in EHEC and it was considered that its lack of a complete *dsdCXA* locus could indicate an 'evolutionary incompatibility' whereby inhibition of virulence by D-Serine prompted bacterial adaptation to non-D-Serine rich environments such as the gut (Connolly *et al.*, 2015). The presence of both the colibactin biosynthetic gene cluster and of D-Serine metabolising genes is difficult to explain from an evolutionary perspective, given that D-Serine is capable of repressing colibactin transcription (Hallam *et al.*, 2023). However, the advantage of D-Serine metabolism in UPEC is not so clearcut. For example the CFT073 *dsdA* mutant was found to display greater flagellation and motility when compared to wildtype, and it was recovered 300-times more frequently than its wildtype counterpart in infected mice bladders, suggesting an increase in virulence stemming from the inability to break down D-Serine (Roesch *et al.*, 2003). This was also observed following the mutation of *dsdC* in CFT073 (Anfora *et al.*, 2007). It is also worth noting that, although the 'urine is sterile' dogma has long been dismantled, the urinary microbiota is estimated to contain 10^4 to 10^5 CFU/ml compared to 10^{12} CFU/g in faeces, suggesting perhaps that the competitive advantage offered by being able to metabolise D-Serine is less relevant in the urinary tract, compared to the gut, especially as D-Serine is more abundant in urine (Perez-Carrasco *et al.*, 2021).

It was interesting to notice that in the RNA-seq data, *clbR* and *clbA* are consistently upregulated – albeit not significantly – when all other *clb* genes are downregulated (Fig 4-9 to 4-11). Genes encoding the PPTase ClbA and the transcriptional regulator ClbR are organised in one orientation. In contrast, the remaining 17 genes, encoding for biosynthesis, transport and resistance, are

transcriptionally orientated in the opposite direction, with absent or short (~50 bp) intergenic regions between them. ClbA and ClbR make up the initiating apparatus for colibactin biosynthesis. ClbA activates NRPS and PKS enzymes by transferring a 4′-phosphopantetheine residue. ClbR is the established LuxR-type transcriptional activator of the colibactin cluster, functioning by binding both to a site upstream of *clbB* and to a region upstream of *clbR* itself, suggesting that, in addition to being a transcriptional regulator, ClbR is also an auto-activator (Wallenstein *et al.*, 2020). It was also noted that ClbR is highly homologous to the transcriptional regulator of *B. subtilis*, GerE (Wallenstein *et al.*, 2020), which can act as both an activator and a repressor via a complex, hierarchical cascade initiated by two different sigma factors (Ichikawa, Halberg and Kroos, 1999). Both regulators lack the N-terminal regulatory receiver (REC) domain and are accordingly classified as effector domain regulators and not response regulators. If ClbR can also function as both an activator and a repressor of transcription, then an upregulation of ClbR can be coupled with both an upregulation or a downregulation of the remaining *clb* genes. Potential explanations as to why *clbR* and *clbA* are consistently upregulated in the presence of D-Serine and D-Tyrosine while other *clb* genes are downregulated, are that either ClbR is acting as a negative control in this instance, or that the downstream pathway triggered by the AAs acts on only the genes orientated in one direction within the biosynthetic cluster. Furthermore, the fact that under D-Tyrosine and under combined AA supplementation *clbB* is the only gene to not be significantly downregulated (among the downregulated *clb* genes) might suggest that its regulation differs from other members of the *pks* cluster and that its transcriptional activity is not the best indicator of overall colibactin biosynthesis.

Among the clinical isolates, most strains seemed to lack or possess incomplete *clbJ* and *clbK* genes. Both genes are large, encoding for polyketide megasynthases, containing repetitive sequences that short-read sequencing technologies – such as Illumina, used in this instance – may struggle to read and assemble effectively. Other studies have observed similar findings with *clbJ* and *clbK* genes and reported that long-read technologies were able to correctly sequence the full genes (Wami *et al.*, 2021). Consequently, it is plausible to conclude that the clinical isolates likely possess full copies of *clbJ* and *clbK* and that their absence is an artificial result from erroneous assemblies of short-read sequencing. Moreover, the *pks*

island is known to be well conserved among strains, with studies reporting that *E. coli* cohorts maintain an over 99% identity in each of the 19 *clb* gene products (Auvray *et al.*, 2021).

When considering both *in vitro* and *in cellulo* assays, the AA that consistently stood out as having repressive activity against colibactin was D-Serine. Even though D-Tyrosine's effect appeared stronger in both the DNA crosslinking assays and in the transcriptomics results, D-Tyrosine had no repressive effect in MEM-HEPES and has exceptionally low solubility, making it problematic for *in vivo* administration. It was therefore decided that the AA with which to test colibactin modulation in future mouse models would be D-Serine.

5 Investigation of the phenotypic and genotypic differences between two isolates of *E. coli* Nissle 1917 (EcN_AR and EcN_EO)

I would like to acknowledge Dr. Emily Addington for having performed the *lacI* deletions and the plasmid purification described in section 5.2.3.

5.1 Introduction

As previously described, this project focused on the contribution of colibactin production by EcN to the development of CRC. While early studies described in Chapter 3 made use of a strain previously used by this group, subsequently referred to as EcN_AR, a second stock of EcN strain was also acquired from Eric Oswald's (EO) Group at the Institut de Recherche en Santé Digestive in Toulouse, France, designated EcN_EO. Experiments performed as part of this PhD project both within the Roe and Oswald laboratories suggested that EcN_EO expressed more colibactin than EcN_AR. Hence the reason why from thesis Chapter 4 onwards the majority of experiments were carried out with EcN_EO.

The work described in this chapter aimed to understand the potential genotypical and phenotypical differences between these two strains. Specifically, using whole genomic sequences (WGS), it was investigated whether differences in activity reflected: (1) genetic differences in the colibactin biosynthetic gene cluster, (2) whether the sequence from each strain exhibited individual genetic mutations compared to the EcN reference genome, and (3) whether the strains displayed differences in colibactin production at a transcriptional and functional level.

5.2 Results

5.2.1 Phenotypic differences between EcN_AR and EcN_EO

When visiting the Oswald laboratory to learn how to perform interstrand crosslinking assays (ISCLAs), it was observed that the EcN strain in the Oswald Group (EcN_EO) appeared more genotoxic than the one in the Roe Group (EcN_AR). As DNA crosslinking is hypothesised to impact how colibactin induces cell transformation, an ISCLA was performed to determine whether these two strains were indeed capable of inducing differing levels of DNA cross-linking. This data revealed a stark difference in the level of crosslinking induced by these two strains with EcN_EO responsible for 84% more crosslinking ($p = 0.0008$) than that observed with EcN_AR (Figure 5-1).

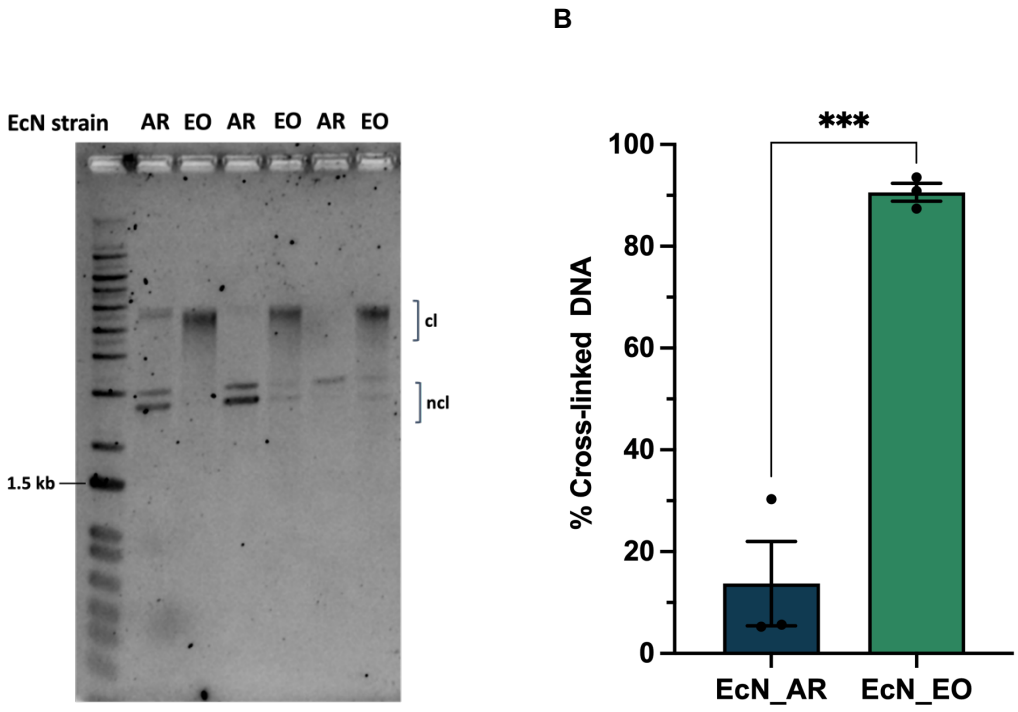


Figure 5-1 Comparison of DNA crosslinking caused by EcN_AR and EcN_EO. (**A**) Visualising cross-linked DNA after migration under alkaline gel electrophoresis. EcN_AR (AR) and EcN_EO (EO) were cultured for 5 h in M9 at 37°C, 200 rpm. Linearised pUC19 was then exposed to 1.5 x 10^6 CFUs of EcN_AR or EcN_EO for 40 mins in M9. Plasmid DNA was purified and crosslinking (cl) and non-crosslinking (ncl) activities were visualised by agarose gel electrophoresis in alkaline conditions. The gel was run at 25 V for 45 mins, and then at 50 V for 150 mins; it was soaked in neutralisation buffer for 20 mins prior to image acquisition; the DNA size marker is the 1 kB DNA ladder (Invitrogen). (**B**) Percentage of cross-linked DNA relative to total DNA. Signal intensities for both crosslinked (cl) and non-crosslinked (ncl) DNA were quantified by densitometry in ImageJ, and the percentage of crosslinked DNA was calculated relative to the total DNA in each lane. Data points represent individual values, columns represent means, and error bars represent standard deviation calculated from three biological replicates. Statistical significance was assessed by a Student's t-test with *** indicating $p < 0.001$.

It was questioned whether EcN_AR and EcN_EO might differ in the phase of growth in which they express colibactin, and whether the greater crosslinking activity of EcN_EO would be reflected at the transcriptional level. Consequently, the two strains were transformed with the *clbB* promoter:GFP reporter fusion plasmid (*pclbB:gfp),* and monitored over 8 h for both growth rate and fluorescence from the *clbB* promoter. Results revealed that EcN_AR *pclbB:gfp* and EcN_EO *pclbB:gfp* were highly comparable in the transcription of *clbB* across all time points in both M9 (Figure 5-2A) and MEM-HEPES (Figure 5-3A), although the overall transcriptional activity tends to be much higher in M9, with RFU values reaching up to 80, rather than in MEM-HEPES, where RFU values reach up to 40. Growth of both strains was also almost identical in both types of media (Figure 5-2B and Figure 5-3B). Both strains also show lower RFUs at 3 h than at 1 h or 2 h, probably reflecting carry-over from accumulation of GFP expression within the overnight culture.

Overall, data suggests that while the strains display no difference at a transcriptional level – at least not at the level of transcriptional activation of the biosynthetic gene *clbB* (Figure 5-2 and 5-3) – there are clear functional differences, with EcN_EO being capable of eliciting far more DNA damage than EcN_AR (Figure 5-1).

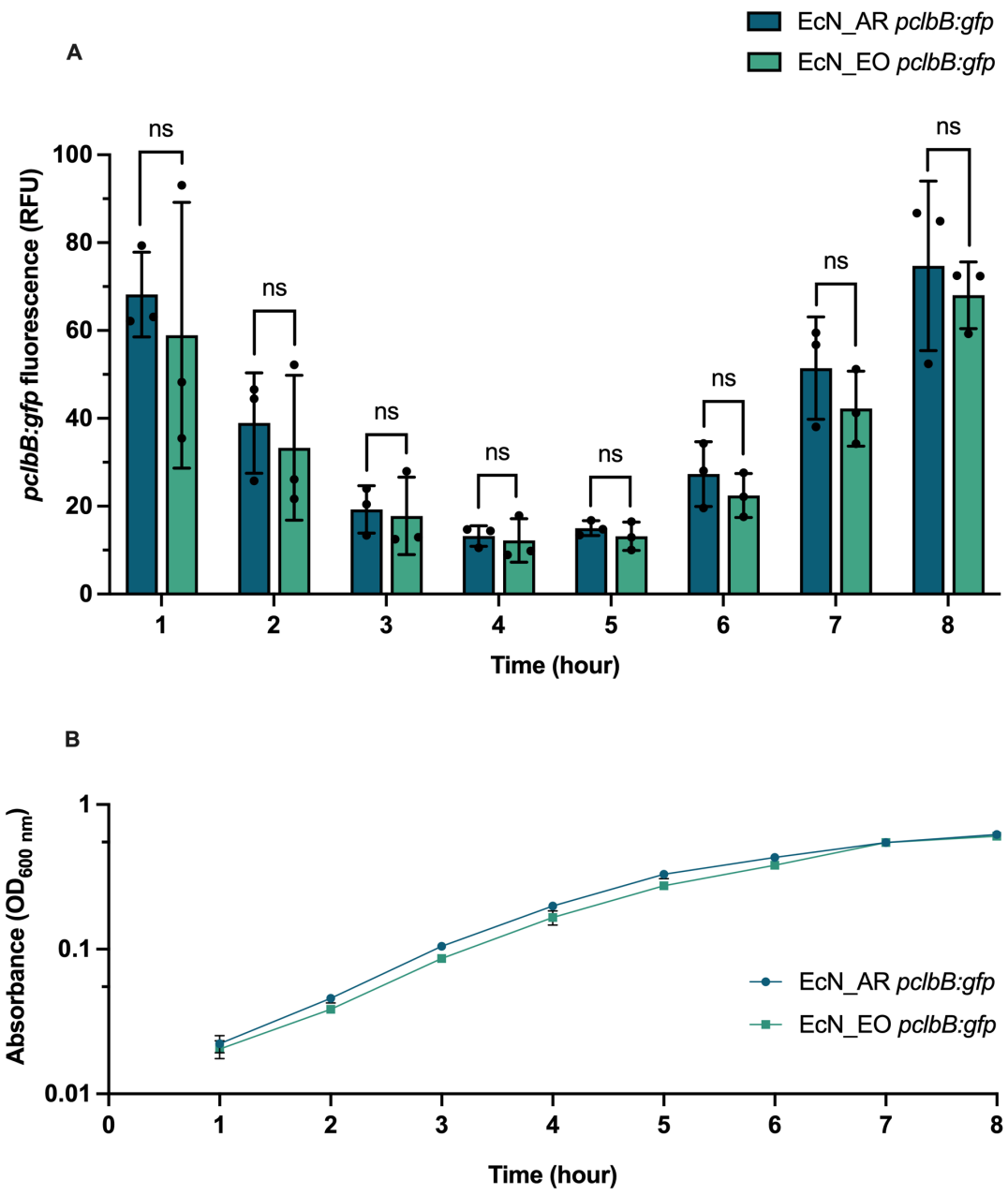


Figure 5-2 Comparison of *clbB* transcription in EcN_AR and EcN_EO and of their growth in M9. (**A**) The *pclbB:gfp* reporter activity of EcN_AR and EcN_EO strains in M9. EcN_AR and EcN_EO were cultured in 50 mL falcon tubes in M9 at 37°C, 200 rpm. Fluorescence was measured hourly over an 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$. Columns represent means, data points represent individual values, and error bars indicated standard deviation, calculated from three biological replicates. Statistical significance was assessed from multiple Student's t-tests with ns indicating lack of significance. (**B**) Growth curves of EcN_AR and EcN_EO in M9. Strains were cultured as described above. Absorbance ($OD_{600\ nm}$) was measured hourly for 8 h and values were plotted on a log-10 scale. Data points represent average absorbance values and error bars indicate standard deviation, calculated from three biological replicates.

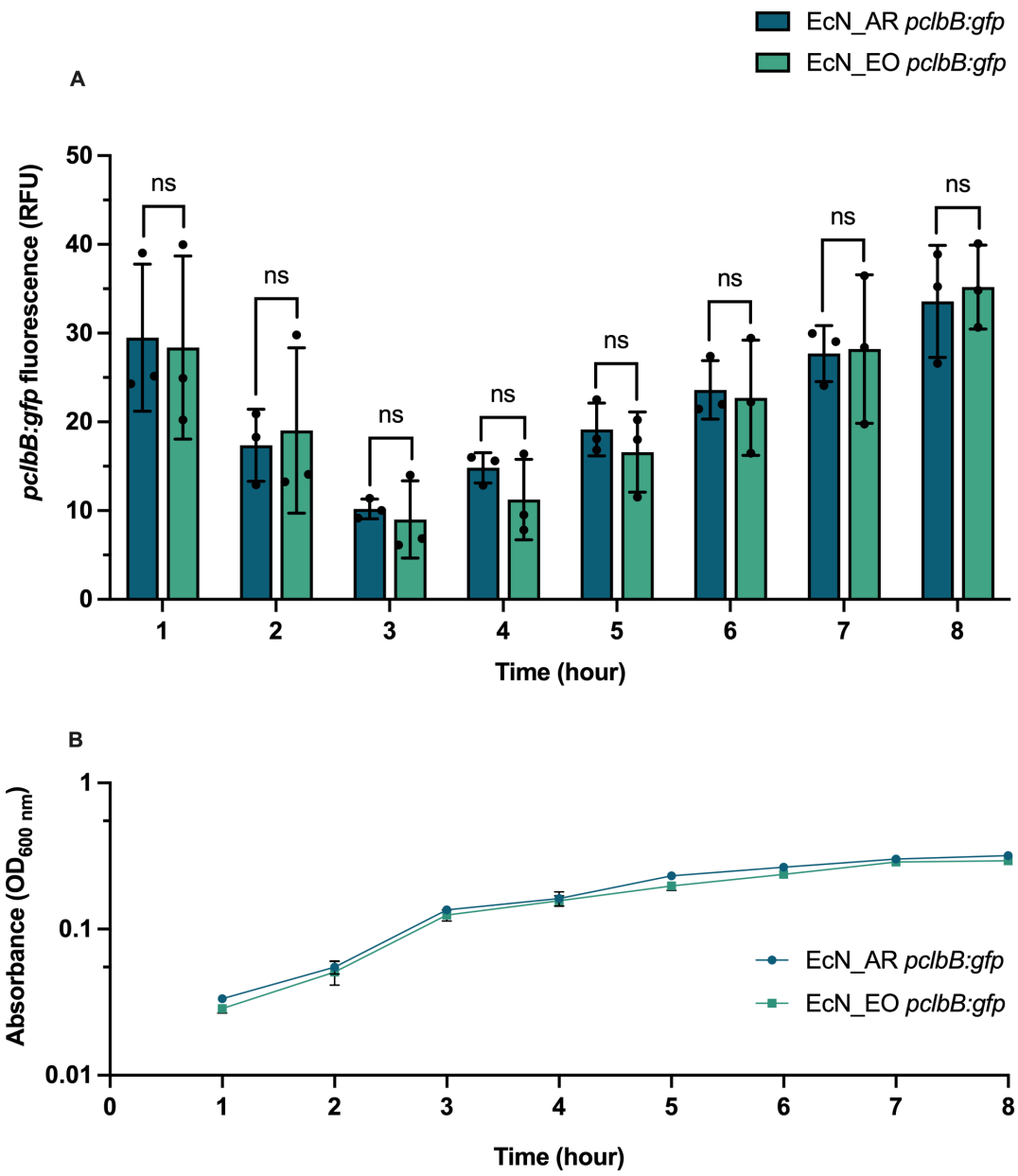


Figure 5-3 Comparison of *clbB* transcription in EcN_AR and EcN_EO and of their growth in MEM-HEPES. (**A**) The *pclbB:gfp* reporter activity of EcN_AR and EcN_EO strains in MEM-HEPES. EcN_AR and EcN_EO were cultured in 50 mL falcon tubes in MEM-HEPES at 37°C, 200 rpm. Fluorescence was measured hourly over an 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$. Columns represent means, data points represent individual values, and error bars indicated standard deviation, calculated from three biological replicates. Statistical significance was assessed from multiple Student's t-tests with ns indicating lack of significance. (**B**) Growth curves of EcN_AR and EcN_EO in MEM-HEPES. Strains were cultured as described above. Absorbance ($OD_{600\ nm}$) was measured hourly for 8 h and values were plotted on a log-10 scale. Data points represent average absorbance values and error bars indicate standard deviation, calculated from three biological replicates.

Next, it was questioned whether colibactin transcription in EcN_AR and EcN_EO would be equally sensitive to the colibactin-repressing effect of D-Serine. To achieve this, a transcriptional reporter fusion assay was used to monitor *clbB* transcription of the two strains in the absence and presence of 1 mM D-Serine. Analyses revealed that at 7 h EcN_AR and EcN_EO had respectively, average RFUs of 43.03 and 40.48 in the absence of D-Serine and of 28.07 and 26.45 in the presence of D-Serine, suggesting highly similar levels of transcriptional control under both conditions (Figure 5-4A). The growth of these strains was also very similar, with both strains displaying an increase in growth when cultured in MEM-HEPES supplemented with 1 mM D-Serine (Figure 5-4B).

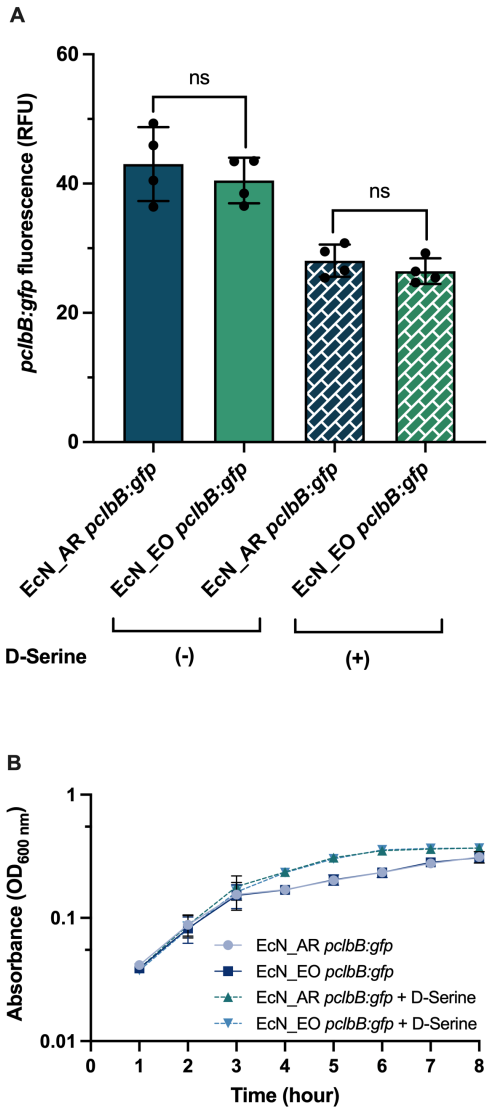


Figure 5-4 Comparison of *clbB* transcription in EcN_AR and EcN_EO and of their growth in MEM-HEPES supplemented with D-Serine. (**A**) The *pclbB:gfp* reporter activity of EcN_AR and EcN_EO strains in MEM-HEPES supplemented with 1 mM D-Serine. EcN_AR and EcN_EO were cultured in 50 mL falcon tubes in MEM-HEPES alone or supplemented with 1 mM D-Serine at 37°C, 200 rpm. Fluorescence was measured hourly over an 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$ at 7 h. Columns represent means, data points represent individual values, and error bars indicated standard deviation, calculated from three biological replicates. Statistical significance was assessed from two Student's t-tests with ns indicating lack of significance. (**B**) Growth curves of EcN_AR and EcN_EO in MEM-HEPES alone and supplemented with 1 mM D-Serine. Strains were cultured as described above. Absorbance ($OD_{600\ nm}$) was measured hourly for 8 h and values were plotted on a log-10 scale. Data points represent average absorbance values, dashed lines represent growth in the presence of D-Serine, and error bars indicate standard deviation, calculated from three biological replicates.

Given the apparent difference in DNA cross-linking observed between the two strains, it was investigated whether this difference would translate into the megalocytotic phenotype observed during *in cellulo* experiments . To analyse this, HeLa cells were exposed to both strains in the presence and absence of the previously identified five AAs (Chapter 4). The resultant *in cellulo* data is represented in Figure 5-5, and has been re-calculated to display the average cell area measurements of each strain under all culturing conditions. While the data suggests that both EcN_AR and EcN_EO can trigger comparable cell enlargement, infection with EcN_EO seemed to elicit a more consistent phenotype. As illustrated in the previous chapter, the average cellular expansion triggered by EcN is significantly reduced with supplementation of 1 mM L-Serine and 1 mM D-Serine by 34.35% ($p = 0.0220$) and 41.22% ($p = 0.0057$) respectively (Figure 5-5).

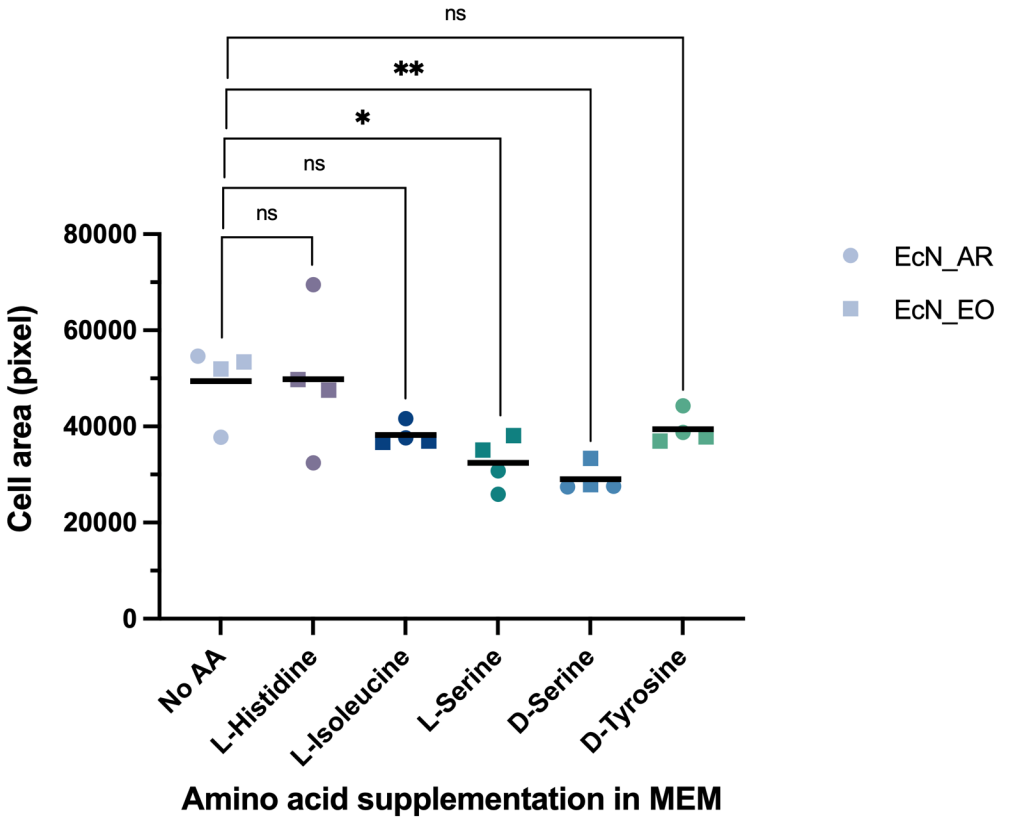


Figure 5-5 Comparison of EcN_AR and EcN_EO – driven megalocytosis under different amino acid regimes. HeLa cells were infected with EcN_AR or EcN_EO (MOI = 400) in MEM-HEPES alone or MEM-HEPES supplemented with 1 mM AAs in 24-well plates. EcN strains had been previously cultured at 37°C, 200 rpm for 4.5 h in MEM-HEPES alone or MEM-HEPES supplemented with 1 mM AAs. At 4 h after infection, cells were washed and incubated for 72 h in DMEM supplemented with 200 µg/mL of gentamicin. Two infections were performed with EcN_AR and two with EcN_EO for a total of four independent biological replicates. Cells were fixed with 4% PFA and stained with methylene blue. Cell morphology was observed by light microscopy at 20X magnification and a total of 5 images were captured in each well, in the centre, top, bottom, left and right. A pipeline was developed on CellProfiler software to measure cell area. Average measurements of infections with EcN_AR are represented with circles, while average measurements of infections with EcN_EO are represented with squares. Statistical significance was determined using a Dunnett's multiple comparison test, with * and ** respectively indicating $p < 0.05$ and $p < 0.01$.

5.2.2 Whole-genome comparison between EcN_AR and EcN_EO

Despite behaving almost identically at a transcriptional level, EcN_AR and EcN_EO displayed a significant difference at a functional level in the amount of DNA damage that they elicited (Figure 5-1). Consequently, it was important to determine whether any genotypic differences existed between the two strains. To determine whether there were any genotypic differences between EcN_AR and EcN_EO, whole genomic DNA (gDNA) from both strains was extracted and sequenced at MicrobesNG at the University of Birmingham, UK.

Following initial computational analyses on Galaxy Australia (Jalili *et al.*, 2020) – the full workflow description of which can be found in Figure 9-3 - GenBank files containing the regions of interest were uploaded on the Comparative Gene Cluster Analysis Toolbox (CAGECAT) and the clinker tool (van den Belt *et al.*, 2023) was used to align the colibactin biosynthetic gene clusters of EcN_EO and EcN_AR. The comparison revealed 100% identity between the two strains across all 19 colibactin-producing genes (Figure 5-6A).

Alignment of the individual strains to the EcN reference genome (NCBI accession number NZ_CP007799.1) revealed a small number of nucleotide mutations. In EcN_AR, most variation was observed in the gene encoding for LacI, with 15 single-nucleotide polymorphisms (SNPs) leading to synonymous variants, 1 multiple nucleotide polymorphism (MNP), and 1 complex mutation (complex), the latter both leading to missense variants. EcN_AR also displayed a SNP in the region encoding for CysT and within the coding region of a hypothetical, unidentified protein, respectively causing a synonymous variant and a missense variant (Figure 5-6 B). In contrast, EcN_EO showed only 3 SNPs: one in the region encoding for RpsL, one in the region encoding for a Drug/Metabolite Transporter (DMT), and another within the coding region of a hypothetical, unidentified protein, all three leading to missense variants (Figure 5-6 C).

The large number of nucleotide mutations in the LacI-encoding sequence of EcN_AR was deemed significative, and consequently *lacI* mutants were made in both EcN_AR and EcN_EO backgrounds to investigate the role that the DNA-binding transcriptional repressor LacI might play in colibactin expression. Results obtained with these mutants are described in a later section of the chapter. The SNP found

in the RpsL-encoding region of EcN_EO was unsurprising given that EcN_EO had been previously reported to be resistant to streptomycin due to a single base change from A to G, which results in a codon substitution of lysine (AAA) to arginine (AGA), known as a K43R mutation (Barnard *et al.*, 2010). Further details on the nucleotide mutations that emerged from the comparison of EcN_AR and EcN_EO genomes can be found in Tables 10-4 and 10-5 of the appendix.

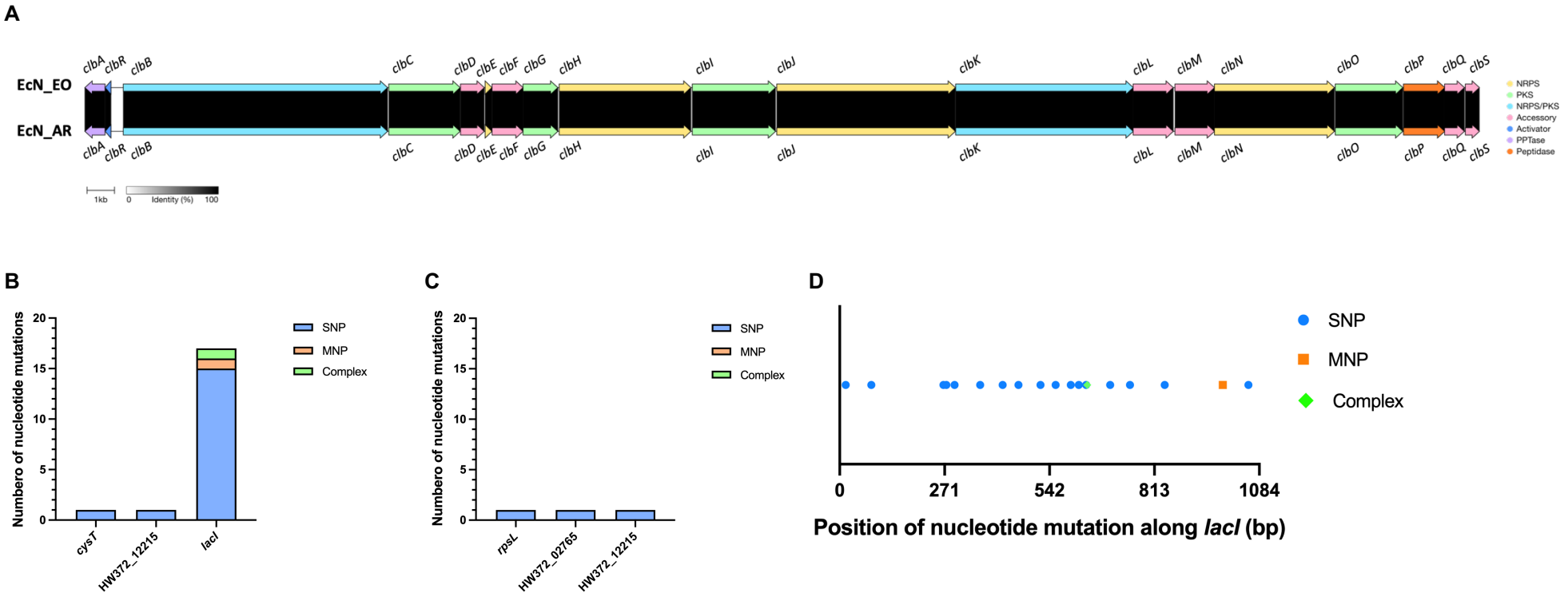


Figure 5-6 Summary of results from WGS of EcN_EO and EcN_AR. Strains were sent for WGS and Galaxy Australia was used to identify nucleotide mutations within the genomes and to assemble and annotate the sequenced genomes. (**A**) Alignment of the colibactin biosynthetic apparatus. The CAGECAT clinker tool was used to align the EcN_EO and EcN_AR colibactin biosynthetic gene clusters. Genes encoding NRPSs are displayed in yellow, PKSs in green, NRPS/PKS hybrids in light blue, accessory enzymes in pink, and the activator, PPTase, and peptidase respectively in dark blue, purple and orange. The black lines connecting the aligned genes represent 100% identity, the scale bar is equal to 1 kb. (**B and C**) Nucleotide mutations in EcN strains compared to the EcN reference genome. Output tables generated by the snippy tool in Galaxy Australia were used to plot the number and types of variants detected in genes. Blue bars represent SNPs, orange bars represent MNPs, and green bars represent complex nucleotide mutations, panel (**B**) corresponds to EcN_AR and (**C**) to EcN_EO; HW372_02765 corresponds to a DMT transporter, and HW372_12215 corresponds to hypothetical protein. (**D**) Nucleotide mutations in *lacI* compared to the EcN reference genome. As above, the output tables generated by the snippy tool in Galaxy Australia were used to plot nucleotide variants identified in *lacI*. Blue dots represent SNPs, orange circles MNPs and green diamonds complex mutations

5.2.3 Comparison of colibactin transcription in EcN_AR and EcN_EO wildtype strains and *lacI* mutants

To evaluate the role of LacI in colibactin expression, the *lacI* gene was deleted from both EcN_AR and EcN_EO using the lambda red recombineering system. The modified strains were subsequently transformed with *pclbB:gfp* (reporter construct) to determine by reporter assays whether *lacI* deletion influenced colibactin transcription. Results revealed that mutation of *lacI* did not modify activity from the *clbB* promoter in any of the strains tested, with comparable activity measured in EcN_AR, EcN_EO and their respective mutants (Figure 5-7A). Growth was also unaffected following *lacI* deletion (Figure 5-7B). Streaking on MacConkey agar – which is a selective culture medium for differentiating Gram-negative, lactose-fermenting bacteria – revealed that all strains are capable of fermenting lactose, as evidenced by the appearance of pink colonies on the plate (Figure 5-7C). This is most likely explained by the fact that deletion of *lacI*, results in the absence of the repressor protein binding to the operator, and consequently the bacterial cell constitutively expresses the *lac* operon.

In a later stage of the project, EcN_AR was found to exhibit kanamycin resistance and based on this observation, it was hypothesised that EcN_AR was unexpectedly carrying a kanamycin-resistant plasmid. To establish whether this was the case, an attempt was made to purify the plasmid using a QIAprep Spin Miniprep kit (Qiagen). The resultant purified plasmid was sent to Plasmidsaurus for sequencing using the Oxford Nanopore Technologies v14 long-read sequencing system. Sequencing revealed that EcN_AR carried a kanamycin-resistant plasmid in which GFP expression had been placed under control of a *lac* promoter, and which also carried a *lacI* sequence. Comparison of the plasmid's *lacI* region with that of EcN's reference genome revealed all the previously-identified nucleotide variants, matching those represented in Figure 5-6B. This likely suggests that all the *lacI* nucleotide variants identified in EcN_AR reflect the plasmid-encoded *lacI* and not the chromosomal gene. The purified kanamycin-resistant plasmid sequence, and full alignment of its *lacI* regions with the EcN reference genome can be found in the appendix, in Figure 10-4 and 10-5 respectively.

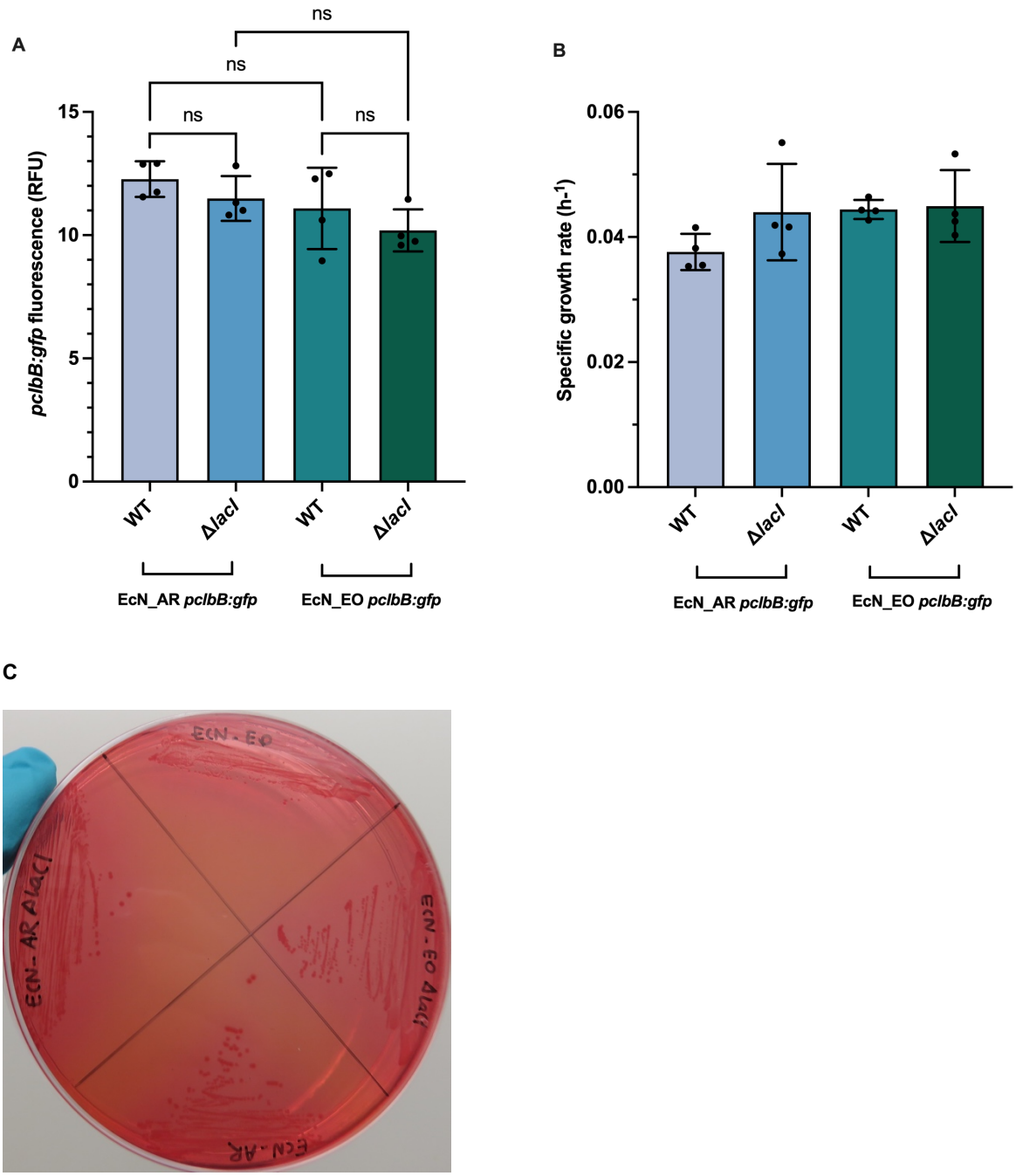


Figure 5-7 Comparison of *clbB* transcription in EcN_AR and EcN_EO and their respective *lacI* mutants, and growth assay of all four strains on MacConkey agar. (**A**) The *pclbB:gfp* reporter activity of EcN_AR and EcN_EO WT strains and their respective *lacI* mutants. EcN_AR and EcN_EO were cultured in 50 mL falcon tubes in MEM-HEPES at 37°C, 200 rpm. Fluorescence was measured hourly over a 3 to 8 h period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$ at 7 hours. Columns represent means, data points represent individual values, and error bars indicated standard deviation, calculated from four biological replicates. Statistical significance was assessed from multiple Student's t-tests with ns indicating lack of significance. (**B**) The specific growth rates of EcN_AR and EcN_EO WT strains and their respective *lacI* mutants. Strains were cultured as described above, absorbance ($OD_{600\ nm}$) was measured hourly and the specific growth rate ($\mu = \Delta \ln OD_{600nm}/\Delta t$) was calculated during exponential growth. The significance of columns, data points and error bars is as described above. (**C**) The growth of EcN_AR and EcN_EO WT strains and their respective *lacI* mutants on MacConkey agar. EcN_AR, EcN_EO and their repsective *lacI* mutants were streaked from glycerol stocks on MacConkey agar plates and incubated at 37°C in a static incubator for 12-16 h. The formation of pink colonies signifies the ability of all strains to ferment lactose.

5.3 Discussion

The main phenotypic difference observed between EcN_AR and EcN_EO was that EcN_EO elicited significantly more DNA crosslinking than EcN_AR. Genotyping of the two strains revealed a very small number of SNPs in either genome although the higher number was observed in the less genotoxic EcN_AR strain. One possible explanation for EcN_EO's ability to cause more DNA cross-linking might lie in its resistance to streptomycin. For example, a study looking at the effect of polymyxin B – a last resort antibiotic – on *pks*+ strains, showed that the cellular stress induced by polymyxin B increased colibactin production in NC101 and EcN (Patric W Sadecki *et al.*, 2021). Unsurprisingly, colibactin-producing *E. coli* that are associated to tumours in human studies have also been found to be resistant against a wide range of antibiotics (Tariq *et al.*, 2022). These studies suggest that colibactin is upregulated under stress conditions and corroborate the line of reasoning that colibactin serves as a defence mechanism against other members of the microbiota. This might also explain why more *clbB* transcription is observed in M9, which is metabolically less rich than MEM. Another potential explanation for greater genotoxicity elicited by EcN_EO might derive from the way in which the strain was isolated, cultured or stored. Discussions with members of the Oswald Group indicated that EcN_EO was isolated from the stool of a infected mice and subsequently cultured. Strains recovered following passages *in vivo* are often more adapted to the production of virulence traits. A future RNA-seq experiment comparing the transcriptomes of EcN_AR and EcN_EO in the presence and absence of streptomycin could shed some light on the relationship between antibiotic resistance and colibactin expression. In addition to EcN_EO eliciting more DNA crosslinking than EcN_AR, the observation that EcN_AR harbours a kanamycin-resistant plasmid might additionally be relevant. Carriage of such plasmids often places an additional metabolic burden on the bacterial cell, even if the plasmid is maintained at a low-copy number.

The other main genotypic difference observed between the two strains consisted in the presence of 17 nucleotide variants in the *lacI* region of EcN_AR, though these were later identified as belonging to the plasmid-encoded *lacI* and not EcN_AR's chromosomal *lacI*. In contrast, the *lacI* from EcN_EO shared complete identity with the reference genome's *lacI*-encoding sequence. To date, the relationship between the *lac* operon and the colibactin biosynthetic gene cluster

does not appear to have been investigated, but the ubiquitous presence of the *lac* operon in *E. coli* - despite the narrow timeframe in which lactose is available through breastfeeding – has sparked question about the evolutionary advantage conferred by the bacterial capacity to ferment lactose. Studies have found that in the presence of lactose, strains which metabolise this carbohydrate, have a selective advantage of up to 11%, likely facilitating early colonisation of the gut by *E. coli* for the remainder of a mammal's lifespan (Pinto *et al.*, 2021). Interestingly though, in *E. coli* the fitness advantage offered by carriage of the *lac* operon also seems to depend on the number of species present in the gut microbiota, with one study showing that the selective advantage of carriage of the *lac* operon is greater when the number of intestinal colonisers increases (Pinto *et al.*, 2021). One potential explanation for this phenomenon is that when interspecies competition increases, the ability to metabolise lactose becomes more important – though it is also true that when an intestinal environment is highly diverse, there are greater opportunities for metabolic synergism, with *E. coli* capable of consuming by-products of nutrients degraded by other microbial species, with less reliance on lactose. Taken together, this suggests that for EcN expression of the *lac* operon can be considered advantageous or redundant, but hardly deleterious, potentially explaining why production of colibactin – which is bactericidal and already poses a means of outcompeting bacterial strains – is uncoupled from lactose metabolism. However, in light of the RNA-seq results discussed in the previous chapter (Chapter 4.2.5), *clbB* appears not to be significantly downregulated by colibactin-repressing AAs. Indeed, even when D-Tyrosine and D-Serine were both used as inhibitors, *clbB* was the only gene – among those whose expression was repressed – not to be significantly downregulated. This might suggest that its regulation differs from other members of the *pks* cluster, and that *clbB* transcription might not be the best indicator of functional colibactin expression. While the *pclbB:gfp* reporter fusion construct can be confidently used to determine whether a lower level of transcription is occurring, it does not seem to correlate with the functional activity of colibactin taking place downstream. This was observed also when looking back on data from Chapters 3 and 4 where, for example, L-Isoleucine reduced *clbB* transcription in reporter assays and by RT-qPCR, but did not ameliorate the cellular senescence and DNA crosslinking derived from colibactin.

The SNPs present in the unidentified, hypothetical proteins in both EcN_AR and EcN_EO were disregarded. The SNP identified in the CysT-encoding region within EcN_AR was deemed irrelevant because it resulted in a synonymous variant. The SNP in the DMT transporter present in EcN_EO was instead further investigated, because it resulted in a missense variant encoding for serine instead of leucine. Based on the gene's length and variant location, the mutated DMT transporter was identified as encoded by *yhbE*. YhbE is an inner membrane protein, classified as a Drug/Metabolite Exporter (DME) within the DMT superfamily (Jack, Yang and Saier, 2001). Members of the DME family, have mostly been implicated in the transport of metabolites, with, for example, YdeD in *E. coli* implicated in the efflux of cysteine metabolites (Daßler *et al.*, 2000), and MadN hypothesised to be an acetate efflux pump, encoded for within the malonate utilisation operon (Berg, Hilbi and Dimroth, 1997). Studies have shown that regulation of bacterial efflux pumps and antibiotic resistance seem to be linked. For example, *E. coli* urinary tract isolates that are resistant to beta-lactams were shown to express high levels of *ynfA*, a small multidrug resistance (SMR) transporter (Sarkar, Bhattacharyya and Mandal, 2015). A comprehensive study looking at a library of 447 *E. coli* transporter knockout strains against antimicrobials showed that mutated strains almost always showed either increased resistance or increased sensitivity to antimicrobial compounds (Munro and Kell, 2022), demonstrating that bacterial membrane transporters are linked to their antimicrobial susceptibility. Another study in *E. coli* showed that cells with a higher expression of *acrAB*, the inner membrane component of the major drug efflux pump AcrAB-TolC, also display lower expression of *mutS*, the DNA mismatch repair gene (El Meouche and Dunlop, 2018). Therefore, even though expression of the DMT transporter YhbE was not tested, given the correlations between antibiotic resistance, efflux pump regulation, and promotion of mutations existing in the literature, it is unsurprising that EcN_EO, which contains an *rpsL* K43R mutation conferring resistance to streptomycin, should also display a mutation in one of its DMT transporters.

In light of these observations and the *in vitro* data showing EcN_EO capable of causing greater crosslinking and more consistent megalocytosis, EcN_EO was used for the remaining experiments and *in vivo* infections.

6 Investigation of colibactin's role on colorectal tumour formation and evaluation of D-Serine's impact on colibactin *in vivo*

I would like to acknowledge Dr. Gillian Douce for having carried out AOM IP injections, helped in monitoring mice during periods of acute sickness, and provided assistance in carrying out culling and dissections throughout all *in vivo* experiments. I would like to acknowledge Iris Floria for the schematic representation (Figure 6-1); Dr. Rebecca McHugh and Patricia Rimbi for their help in carrying out weight checks (Figure 6-3); Dr. Emily Addington for her help in carrying out weight checks, stool collection and dissections (Figure 6-5, 6-10, and 6-14); and Dr. Angelika Rupp for scoring histology samples (Figure 6-19).

6.1 Introduction

In vivo studies have been essential in investigating the relationship between *pks*+ *E. coli* and tumour development, revealing that the mechanisms by which *pks*+ *E. coli* may induce carcinogenesis depend on interaction with a variety of host stimuli. The current hypothesis is that colibactin induces the emergence of senescent host cells, which secrete growth factors that subsequently trigger malignant cell proliferation: more specifically, the expression of microRNA-20a-5p is induced, which targets SENP1, a key protein in the regulation of p53 SUMOylation (Cougnoux *et al.*, 2014). Given that p53 SUMOylation is an essential post-translational modification in eukaryotic cells, an accumulation of SUMO-conjugated p53, enhances cellular senescence (Bischof and Dejean, 2007). *In vivo* experiments have also shown that colon cancer-associated *E. coli* are able to secrete carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6), persist strongly, and promote inflammation and cell proliferation (Raisch *et al.*, 2014).

Given that carriage of *pks* has been robustly linked to tumour development (Cougnoux *et al.*, 2014) (Salesse *et al.*, 2021), (Raisch *et al.*, 2014), the focus of this chapter was to assess the capacity of the *E. coli* EcN strain to promote CRC in susceptible mice. The strain is of particular interest because it is a well-established probiotic used for the treatment of a variety of gastrointestinal disorders (Schultz and Burton, 2017).

The aims of this chapter were to (1) establish a mouse model of CRC in which to test colibactin's enhancement of tumour formation, (2) test the effect that EcN colonisation has on the formation of colorectal polyps, and (3) evaluate D-Serine's therapeutic potential against colibactin-driven colorectal tumours.

6.2 Results

6.2.1 Establishment of a mouse model suitable for assessing colibactin effects *in vivo*

To assess the effect of EcN-derived colibactin on colorectal tumours, and to try to understand the interplay between D-Serine, EcN, and polyp development, a CRC mouse model had to first be established in the lab. The methodology described by Cougnoux *et al.* – who show that *pks+ E. coli* enhance polyps – was used as an initial reference to establish a protocol, in which CRC was induced by administration of 10 mg/kg of azoxymethane (AOM), followed by two, 7-day cycles of 2% dextran sulfate sodium salt (DSS). To determine the impact of infection, mice were orally gavaged with 10^9 CFU of colibactin-producing bacteria after receiving 2 mg/mL of streptomycin for the first two days (Cougnoux *et al.*, 2014). A total of six animal experiments were carried out during this PhD project. Briefly, the first two were pilot experiments, involving small animal cohorts, that were used to assess varied DSS concentrations, determine EcN persistence, and familiarise investigators with the clinical symptoms associated to AOM/DSS models. These experiments informed protocol optimisation and are not described in the thesis. The first two animal experiments were undertaken using EcN_AR, which was subsequently substituted with EcN_EO, the more genotoxic strain. A schematic representation of the protocol used in this project's animal experiments is provided in Figure 6-1. The schematic representation was designed in BioRender (Biorender.com).

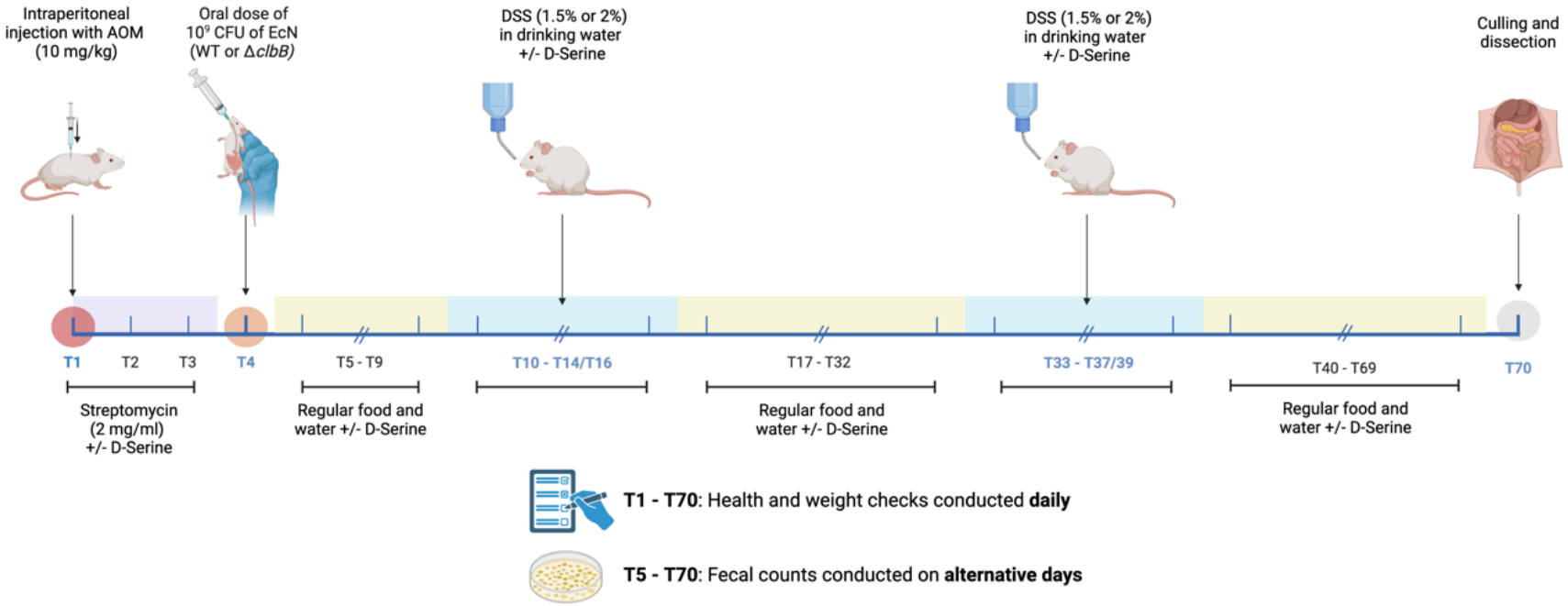


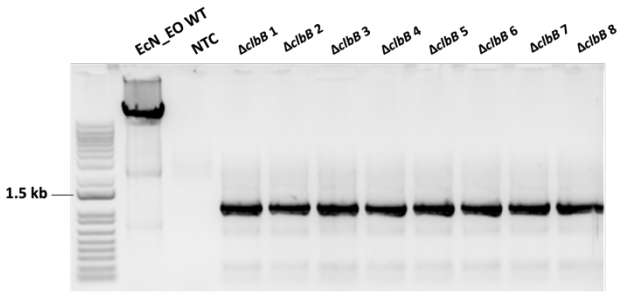
Figure 6-1 Timeline of mouse experiments. Schematic representation of the protocol used to assess EcN's effect on CRC development and to test D-Serine's modulation of colibactin *in vivo* against an AOM/DSS mouse model. On the first day, mice received an intraperitoneal (IP) injection of AOM (10 mg/kg body weight). In parallel, they were treated by streptomycin in drinking water (2 mg/ml) for three consecutive days. On the fourth day, mice belonging to infection groups received by oral gavage 10^9 CFUs of WT EcN or of the isogenic *clbB* mutant. All mice were then maintained on a regimen of regular drinking water and diet for 5 days, before being subjected to two cycles of 1.5% or 2% DSS for 5-7 days. Fresh DSS solution was made up mid-cycle for each of the two cycles, which were separated by a 14-day recovery period during which mice were administered regular drinking water. Following the second DSS cycle, mice were maintained on a regimen of regular diet and water for approximately one more month before being culled and dissected. In groups that were under AA treatment, fresh D-Serine was replaced daily except during DSS cycles. Health status and weight checks were conducted and recorded daily. Faecal counts were conducted on alternate days.

The use of antibiotics, in particular streptomycin, has been shown to reduce the diversity of the gut microbiome, allowing for more effective colonisation by Gram-negative bacteria (Barthel *et al.*, 2003). Based on previous mice models conducted in the Douce Group, it was decided that mice would be administered 2 mg/mL of streptomycin in their drinking water for three days before infection, instead of two days. On the fourth day, mice were infected with 10^9 CFU of streptomycin-resistant bacteria by oral gavage. Whereas EcN_AR, used in the first two pilot experiments, had been made resistant to streptomycin by inducing spontaneous generation of resistant colonies, EcN_EO – received from the Eric Oswald laboratory – already possessed a K43R mutation in the *rpsL* gene, conferring streptomycin resistance. This was confirmed by sequencing of *rpsL* (Figure 6-2A). As a negative control, an isogenic mutant of EcN_EO was created by knocking out *clbB* through lambda red mutagenesis. The *clbB* mutant was confirmed by both PCR and by sequencing (Figure 6-2B) and was tested for potential growth defects in both MEM-HEPES (Figure 6-2C) and M9 (Figure 6-2D). The wildtype and mutant strains exhibited almost identical growth profiles. As previously observed, EcN_EO's growth rate was higher in the presence of D-Serine when cultured in MEM-HEPES, but not when cultured in M9.

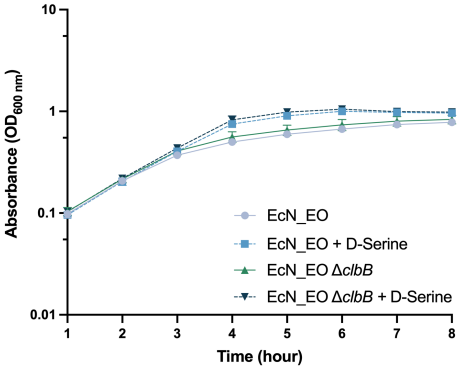
A

```
1   MATVNQLVRKPRARKVAKSNVPALEACPQKRGVCTRVYTTTPKKPNSALR 50
    ||||||||||||||||||||||||||||||||||||||||||.|||||||
1   MATVNQLVRKPRARKVAKSNVPALEACPQKRGVCTRVYTTTPRKPNSALR 50

51  KVCRVRLTNGFEVTSYIGGEGHNLQEHSVILIRGGRVKDLPGVRYHTVRG 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
51  KVCRVRLTNGFEVTSYIGGEGHNLQEHSVILIRGGRVKDLPGVRYHTVRG 100

101 ALDCSGVKDRKQARSKYGVKRPKA 124
    |||||||||||||||||||||||||
101 ALDCSGVKDRKQARSKYGVKRPKA 124
```

B



C



D

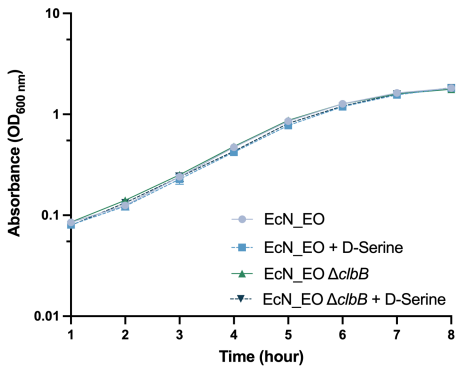


Figure 6-2 Generation and validation of the EcN strains for *in vivo* infections. (**A**) Confirmation of *rpsL* mutation in EcN_EO. The EcN_EO strain received from the Oswald laboratory was known to be streptomycin-resistant, but its resistance-conferring mutation was unknown. The *rpsL* gene was amplified, sent for Sanger sequencing, and revealed to have a K43R mutation. (**B**) Confirmation of *clbB* deletion in EcN_EO. The EcN_EO Δ*clbB* strain was generated by lambda red recombineering and checked for successful gene deletion by PCR, performed with a set of primers amplifying the 1667 bp region expected in successful mutants. The gel shows DNA amplified from wild-type EcN and 8 mutants, which were further confirmed by Sanger sequencing. (**C**) Growth curves of EcN_EO and EcN_EO Δ*clbB* in MEM-HEPES +/- 1 mM D-Serine. Strains were cultured in 50 mL falcon tubes in MEM-HEPES alone or in MEM-HEPES supplemented with 1 mM D-Serine at 37°C, 200 rpm. Absorbance ($OD_{600\ nm}$) was measured hourly for 8 h and values were plotted on a log-10 scale. Data points represent average absorbance values, dashed lines represent growth in the presence of D-Serine, and error bars indicate standard deviation, calculated from three biological replicates. (**D**) Growth curves of EcN_EO and EcN_EO Δ*clbB* in M9 +/- 1 mM D-Serine. Strains were cultured as described above, but in M9 instead of MEM-HEPES. Absorbance was measured and plotted as previously described.

6.2.2 Assessment of 1.5% DSS on colorectal cancer development in the presence of EcN_EO

It was decided that mice would be administered two, 7-day cycles of 1.5% DSS to limit unnecessary animal suffering. The aims of this first animal experiment were therefore to test (1) if a concentration of 1.5% DSS instead of 2% DSS triggered sufficient inflammation for the development of colorectal polyps, (2) how long EcN_EO Δ*clbB* and EcN_EO persist in the mouse gut, and (3) if EcN_EO caused an enhancement of colorectal tumours. A total of 18 C57BL/6 female mice between 6-8 weeks old were placed into three different groups, with two cages per group to minimise cage effects. As described, all mice were treated with 2 mg/mL of streptomycin for three days, given an intraperitoneal (IP) injection of 10 mg/kg of AOM, and were subjected to two 7-day cycles of 1.5% DSS. Mice in cages 1 and 2 received no additional treatment (referred to as AOM/DSS), those in cages 3 and 4 were infected with 10^9 CFU of EcN_EO Δ*clbB* (referred to as AOM/DSS + EcN_EO Δ*clbB*), and those in cages 5 and 6 were infected with 10^9 CFU of EcN_EO (referred to as AOM/DSS + EcN_EO). Health status and weight checks were conducted and recorded daily, whereas faecal pellets for monitoring bacterial persistence were collected on alternate days. In the first 14 days of the experiment, mice exhibited weight fluctuations trending towards a general weight gain. Interestingly, none appeared to lose weight during or following first cycle of 1.5% DSS treatment, despite observation of several clinical symptoms, such as bowel irritation, soft stool, and traces of blood in faecal pellets (Figure 6-3A). Following the second, 1.5% DSS cycle, several mice lost up to 15% of their of their body weight with average weight losses being: 8.69% in AOM/DSS cages, 6.21% in AOM/DSS + EcN_EO Δ*clbB* mice and 9.74% in AOM/DSS + EcN_EO animals (Figure 6-3B). On day 49, one cage from each treatment group was culled, but, since no colorectal tumours were observable, the experiment was extended for a further 17 days (66 days) to allow time for potential tumour development in the remaining mice. However, no macroscopic polyps were detected upon dissection in any of the animals. Subsequent intestinal histology revealed localised and generalised dysplasia in several mice, an adenoma in a mouse from group 4, and adenomas in two mice from group 5 (data not shown).

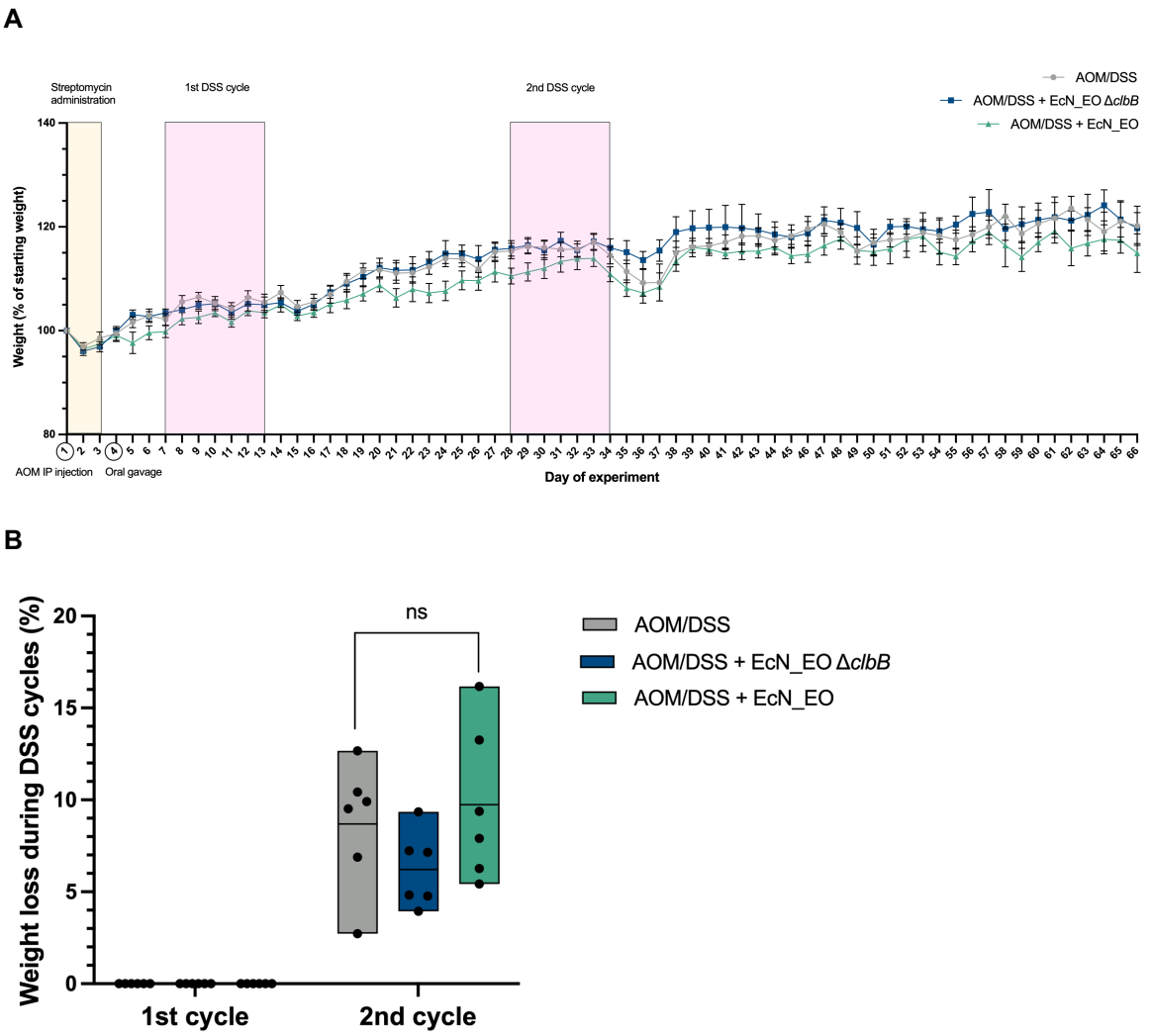


Figure 6-3 Weight changes of AOM/DSS-treated mice under different infection regimens. (**A**) Daily weights of AOM/DSS-treated mice. Following an IP injection with 10 mg/kg of AOM and administration of 2 mg/mL of streptomycin (yellow section), mice were left uninfected (grey), gavaged with 10^9 CFU/dose of EcN_EO Δ*clbB* (blue), or gavaged with 10^9 CFU/dose of EcN_EO (green). They were then treated with two, 7-day cycles of 1.5% DSS (pink sections), before being culled and dissected. Mice were monitored daily for weight loss. Each treatment group was made up of two cages containing three mice each; data points represent average daily weights per treatment group, and error bars represent standard deviation. (**B**) Weight loss following DSS cycles. Mice were weighed daily and were expected to lose weight during and following DSS cycles. Their lowest weight before recovering from colitis was subtracted from their highest weight before DSS treatment. Data points represent individual values, the line within the boxplot represents the mean. Statistical significance was calculated with an ordinary one-way ANOVA test.

Faecal pellets collected on alternate days were homogenised at a ratio of 0.1 g of stool per 1 ml of PBS, subjected to serial tenfold dilutions, and, for each sample, 20 µl of each dilution were spotted on solid agar plates containing 100 µg/ml streptomycin. The plates were allowed to dry and incubated overnight at 30°C. The number of colonies within each spot were counted and the number of CFU/ml calculated. Results showed that EcN_EO Δ*clbB* and EcN_EO exhibit almost identical levels of colonisation and persistence throughout the entire experiment. This is unsurprising given that mice are coprophages and can therefore reinfect themselves with their intestinal bacteria. The final stool samples were also checked by PCR through amplification of *clbB* to ensure that no cross-contamination between cages had occurred, and that mice were indeed colonised for the experimental duration with only the strains with which they had been initially infected. Wildtype and mutant colonies were added for comparison alongside a no template control (NTC). The PCR revealed that mice in cages 3 and 4 were still infected with EcN_EO Δ*clbB* and those in cages 5 and 6 with the wildtype strain (Figure 6-4).

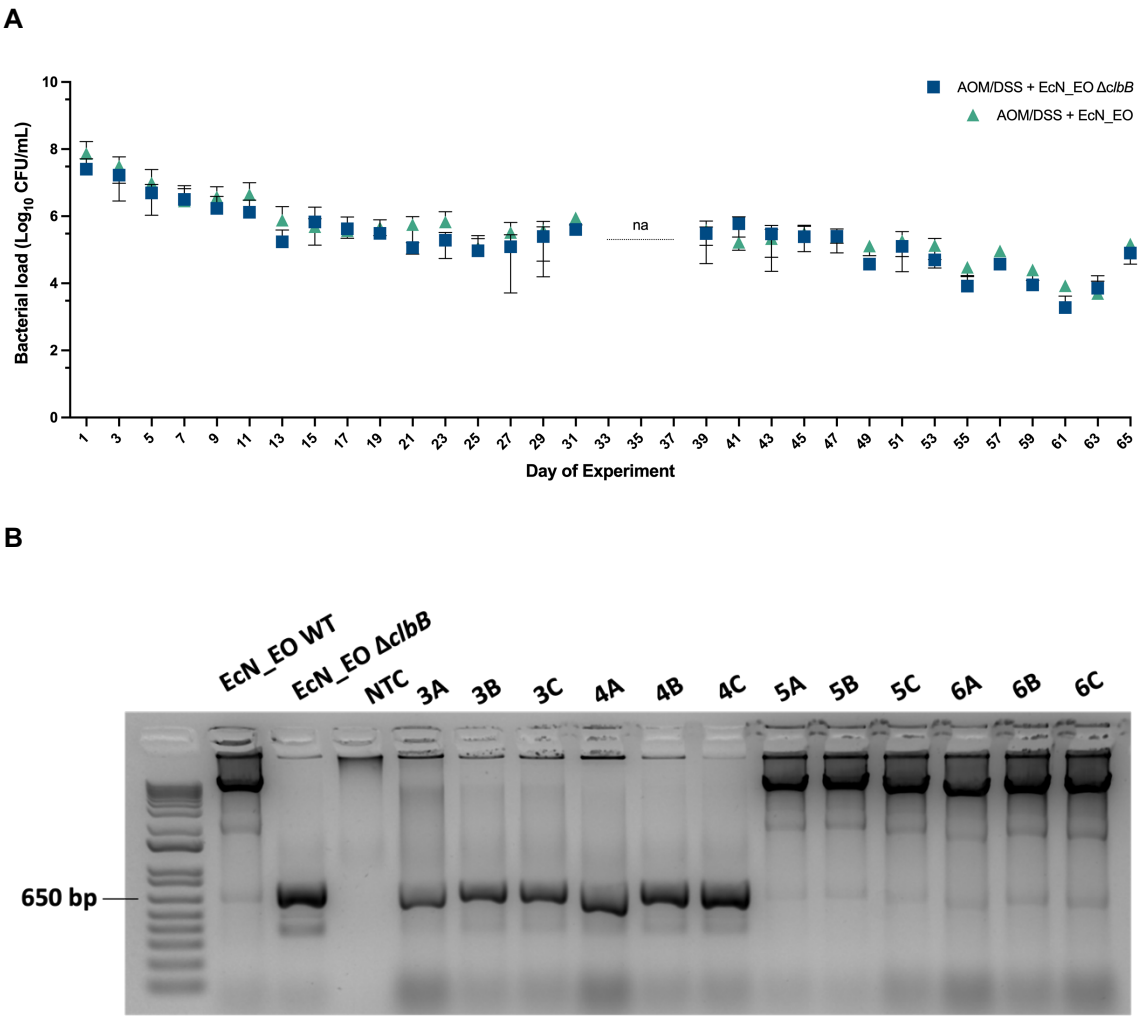


Figure 6-4 Assessment of EcN WT and Δ*clbB* colonisation in C57BL/6 mice. (**A**) Colonisation of EcN_EO Δ*clbB* and EcN_EO in an AOM/DSS background. Colonisation of the two bacterial strains was monitored on alternate days by faecal shedding for the duration of the entire experiment. Samples were homogenised at a ratio of 1 ml PBS/0.1 g of stool, serially diluted at a ratio of 1:10, and spotted on solid agar plates containing 100 µg/ml streptomycin. Plates were incubated overnight at 30°C, single colonies were counted within each spot, and multiplied by the relevant dilution factor to calculate CFU/ml. Data points represent the average bacterial load in EcN_EO Δ*clbB* – infected mice (blue) and EcN_EO – infected mice (green); error bars represent standard deviation. Missing data points are indicated by 'na' and correspond to a period of acute colitis in which bowel movement was too irregular to allow for sample collection. (**B**) Verification of EcN_EO Δ*clbB* and EcN_EO persistence. Stool samples in from infected cages were collected from mice prior to culling. Samples were homogenised and spotted as described above. Colonies from the same spot were collected with a single streak, DNA was amplified and loaded on an agarose gel for visualisation.

6.2.3 Assessment of 2% DSS on colorectal cancer development in the presence of EcN_EO and comparison of male and female C57BL/6 mice models

As the previous experiment indicated that 1.5% DSS was insufficient to trigger development of carcinoma in C57BL/6 mice, the original protocol, in which mice had been given 2% DSS, was re-evaluated. Further consultation of published literature and advice from collaborators at the Cancer Research UK (CRUK) Scotland Institute also stimulated other changes to the protocol. These included the provision of regular tap water and feed instead of sterile, bottled water and sterilised feed, the rationale being that DSS-induced colitis is enhanced when animals are exposed to non-sterile conditions (Chassaing *et al.*, 2015). In addition, male mice were tested alongside female ones, given that most published AOM/DSS models have been conducted in males. Some protocols also mentioned that provision of fresh DSS may enhance inflammation, consequently the drinking water containing 2% DSS was replaced mid-cycle for each of the two cycles. The aims of this experiment were therefore to test (1) if a concentration of 2% DSS instead of 1.5% DSS triggered sufficient inflammation for the development of colorectal polyps, and (2) if EcN_EO caused an enhancement of colorectal tumours.

A total of 18 male and 18 female C57BL/6 mice between 6-8 weeks old were respectively placed into three different groups – with two cages assigned to each group – based on their subsequent treatments. All mice were administered 2 mg/mL of streptomycin for three days, an IP injection of 10 mg/kg of AOM, and two 7-day cycles of 2% DSS. Two male and two female cages of mice received no additional treatment (referred to as AOM/DSS), two cages of each sex of mice were infected with 10^9 CFU of EcN_EO Δ*clbB* (referred to as AOM/DSS + EcN_EO Δ*clbB*), and the remaining four cages of animals were infected with 10^9 CFU of EcN_EO (referred to as AOM/DSS + EcN_EO). As observed in previous experiments, the weight of both sexes of mice dropped when placed on streptomycin before gradually recovering (Figure 6-5A and 6-5B). In contrast to the previous experiment, males and females lost considerable weight following not only the second DSS cycle, but also the first one, suggesting an earlier, more aggressive onset of colitis due to the increase in DSS concentration. During the first DSS cycle, weight losses exhibited in males were as following: 11.60% (AOM/DSS), 9.86% (AOM/DSS + EcN_EO Δ*clbB*), and 4.21% (AOM/DSS + EcN_EO); and the following in

females: 13.25% (AOM/DSS), 12.19% (AOM/DSS + EcN_EO Δ*clbB*), and 11.86% (AOM/DSS + EcN_EO). It's interesting to note the significantly lower weight loss ($p = 0.0337$) in male mice infected with EcN, compared to uninfected controls (Figure 6-5C). It seems to suggest that in this *in vivo* scenario, EcN may be exerting its probiotic, anti-inflammatory effect, over its genotoxic one. In the second DSS cycle, male mice lost similar weight across all three conditions with EcN-infected mice losing on average 11.33% of their body weight, compared to the previous 4.21% weight loss. This might be due to a reduction in EcN colonisation (observed over time), or to the fact that colorectal disease has progressed to the point that all mice were sick and consequently manifesting similar levels of colitis. Female mice in contrast lost similar percentages of weight across all conditions in both DSS cycles, but it's interesting to notice that, on average, weight losses in the second DSS cycle were lower than in the first (Figure 6-5C).

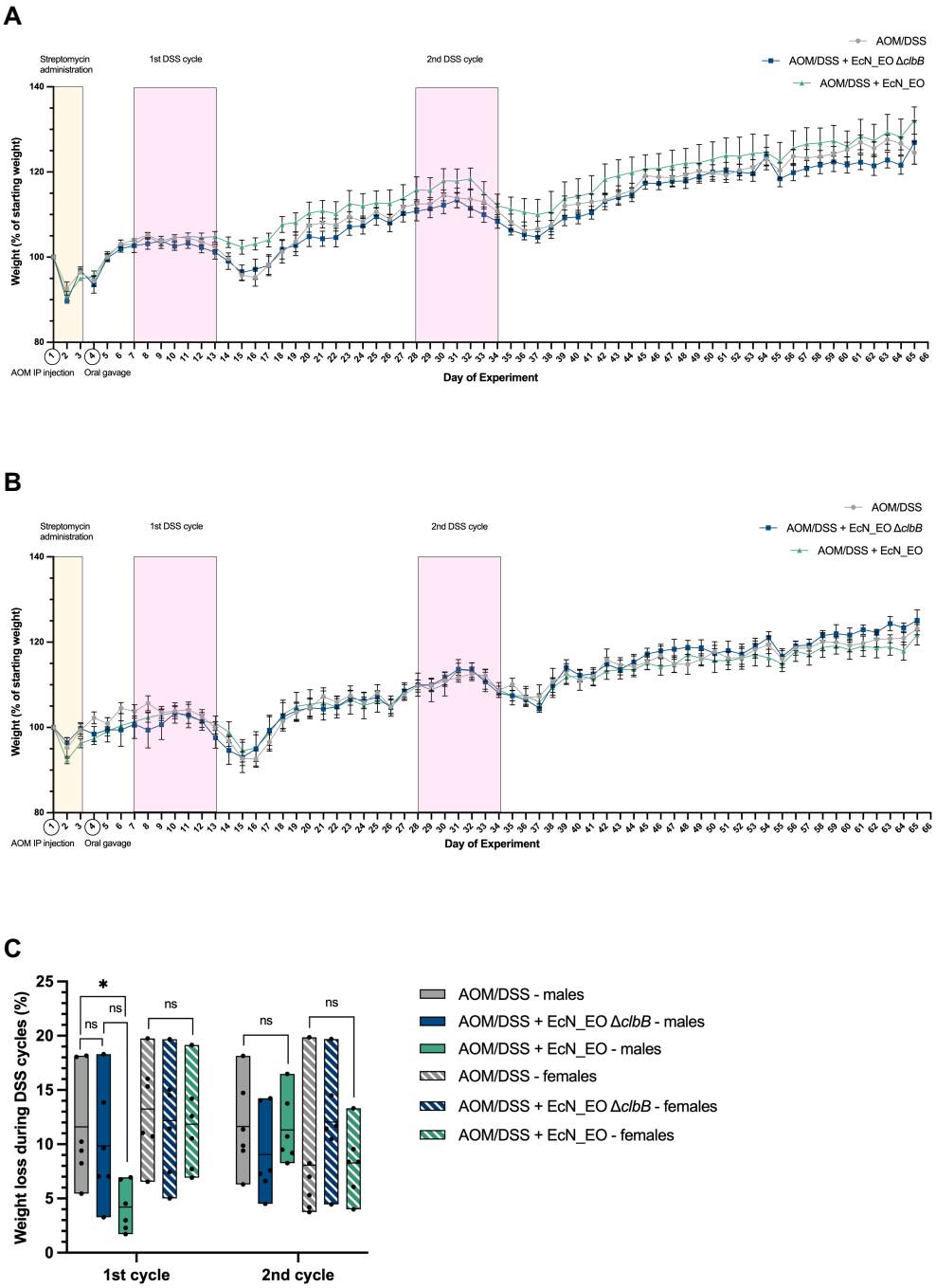


Figure 6-5 Weight changes of AOM/DSS treated males and females under different infection regimens. (**A**) Daily weights of AOM/DSS-treated males. Following an IP injection with 10 mg/kg of AOM and administration of 2 mg/mL of streptomycin (yellow section), mice were left uninfected (grey), gavaged with 10^9 CFU/dose of EcN_EO Δ*clbB* (blue), or gavaged with 10^9 CFU/dose of EcN_EO (green). They were then treated with two, 7-day cycles of 2% DSS (pink sections), before being culled and dissected. Mice were monitored daily for weight loss. Each treatment group was made up of two cages containing three mice each; data points represent average daily weights per treatment group, and error bars represent standard deviation. (**B**) Daily weights of AOM/DSS-treated females. Female mice were treated and infected as described above; data points represent average daily weights per treatment group, and error bars represent standard deviation. (**C**) Weight loss following DSS cycles. Mice were expected to lose weight during and following DSS cycles. Their lowest weight before recovering from colitis was subtracted from their highest weight before DSS treatment. Data points represent individual values, the line within the boxplots represents the mean. Statistical significance was calculated with an ordinary one-way ANOVA test and, when significance was detected within a three-way comparison, individual comparisons were performed with a Tukey's multiple comparisons test. Significance is represented by *, with $p < 0.05$.

On the scheduled cull date, mice were euthanised and dissected. Their colons were flushed with PBS, cut longitudinally, and pinned on ralwax plates to assess the presence of polyps. Almost all mice presented macroscopic, colorectal polyps, which, following fixation, were counted, measured, and assessed for their location along the colon length (Figure 6-6).

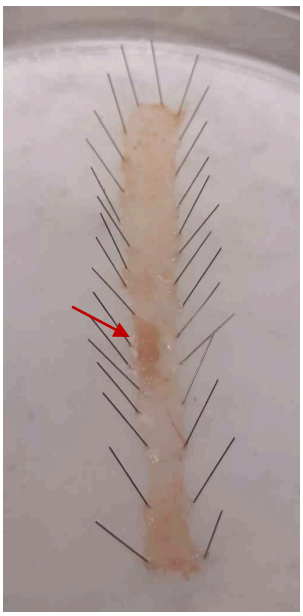


Figure 6-6 Image of dissected mouse colon containing polyps. Approximately 10 weeks into the experiment, mice were culled and dissected. Their colons were flushed with PBS, cut longitudinally, and laterally pinned onto ralwax plates. Colons were fixed with 10% buffered formalin for 24 h, before being made into Swiss-rolls for histological analyses. The red arrow points to two polyps that developed in the mid-section of the colon.

Macroscopic assessment revealed that male mice developed on average the following number of polyps: 3.85 (AOM/DSS), 2.00 (AOM/DSS + EcN_EO Δ*clbB*), and 3.50 (AOM/DSS + EcN_EO); while results in females were the following: 4.50 (AOM/DSS), 3.33 (AOM/DSS + EcN_EO Δ*clbB*), and 3.67 (AOM/DSS + EcN_EO). One male in the AOM/DSS treatment and one female in the AOM/DSS + EcN_EO Δ*clbB* failed to develop polyps (Figure 6-7A). In addition to increasing polyp numbers, colibactin has also been reported to increase polyp size (Cougnoux *et al.*, 2014), so tumours were measured to reveal the following average sizes in males: 0.3 cm (AOM/DSS), 0.18 cm (AOM/DSS + EcN_EO Δ*clbB*), and 0.25 (AOM/DSS + EcN_EO); and the following in females: 0.19 (AOM/DSS), 0.27 (AOM/DSS + EcN_EO Δ*clbB*), and 0.28 (AOM/DSS + EcN_EO). The size of polyps in AOM/DSS + EcN_EO Δ*clbB* male mice was significantly lower than in AOM/DSS controls ($p = 0.0195$); while in females the size of polyps was significantly larger in both AOM/DSS + EcN_EO Δ*clbB* ($p = 0.0166$) and AOM/DSS + EcN (0.0040) mice, compared to controls (Figure 6-7B). Overall, a total of 56 polyps were macroscopically assessed in males, while in females a total of 69 polyps were counted and measured. However it should be noted that in females, compared to males, several bumps in the tissue were suspected to be mucosa-associated lymphoid tissues (MALTs), rather than tumours. Therefore – even though all tissue protrusions were equally assessed in both sexes of mice – females were suspected to display more MALTs than males which may have led to skewing of the macroscopic polyp assessment.

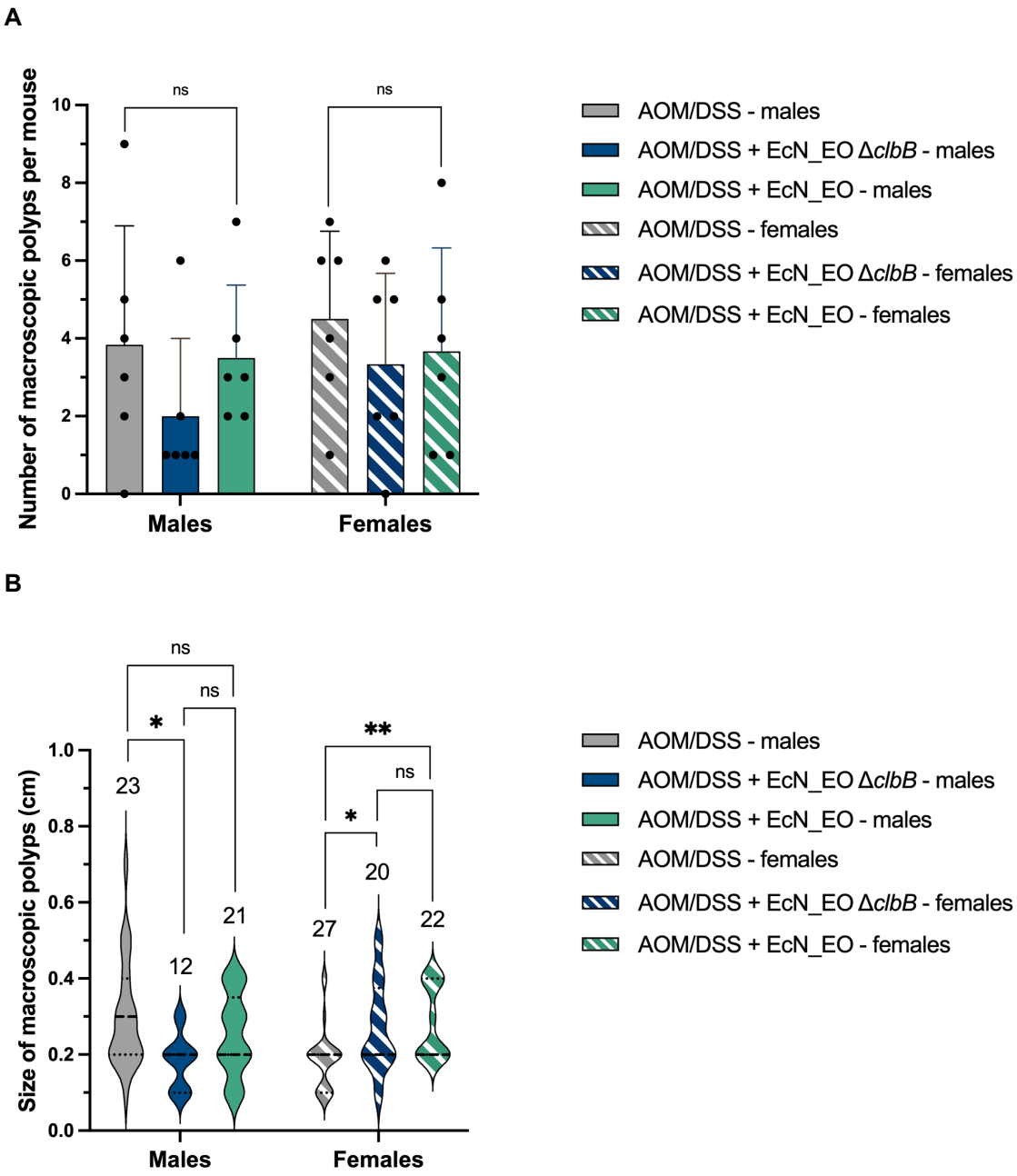


Figure 6-7 Number and size of macroscopic polyps developed in AOM/DSS-treated mice infected with EcN WT and Δ*clbB*. Formalin-fixed colons were examined for the size, number, and distribution of polyps. (**A**) Number of macroscopic polyps across different treatment groups. The total number of macroscopic tumours was counted in mice that were uninfected (gray), infected with EcN_EO Δ*clbB* (blue), and infected with EcN_EO mice (green). The values obtained in males are represented in solid colours, while the measurements obtained in females are represented with diagonal, white stripes. Data points represent individual values, bars represent average values, and error bars represent standard deviation. Statistical significance was calculated with an ordinary one-way ANOVA test. (**B**) Size of macroscopic polyps across different treatment groups. The size of macroscopic tumours was measured in mice that were uninfected (gray), infected with EcN_EO Δ*clbB* (blue), and infected with EcN_EO mice (green). The values obtained in males are represented in solid-coloured violin plots, while the measurements obtained in females are represented with diagonal, white stripes. The numbers written above violin plots represent the total number of polyps measured within that treatment group. Statistical significance was calculated with an ordinary one-way ANOVA test and, when significance was detected within a three-way comparison, individual comparisons were performed with a Tukey's multiple comparisons test. Significance is represented with * and ** respectively indicating $p < 0.05$ and $p < 0.01$.

To test whether infection with EcN_EO Δ*clbB* or EcN_EO triggered tumour development in specific regions of the intestine, macroscopic polyps were assessed for their location along the colon length. Colons were on average 9 cm long, measured from the caecum (1 cm) to the rectum (9 cm). Given that AOM/DSS is a CRC model, it was unsurprising that the majority of polyps were located towards the rectal end of the colon in both males (Figure 6-8A) and females (Figure 6-8B). While rectal polyps were also seen in EcN-infected mice, it was interesting to note that EcN-infected males and females displayed a greater incidence of polyps also located in the early-to-mid section of the gut (Figure 6-8).

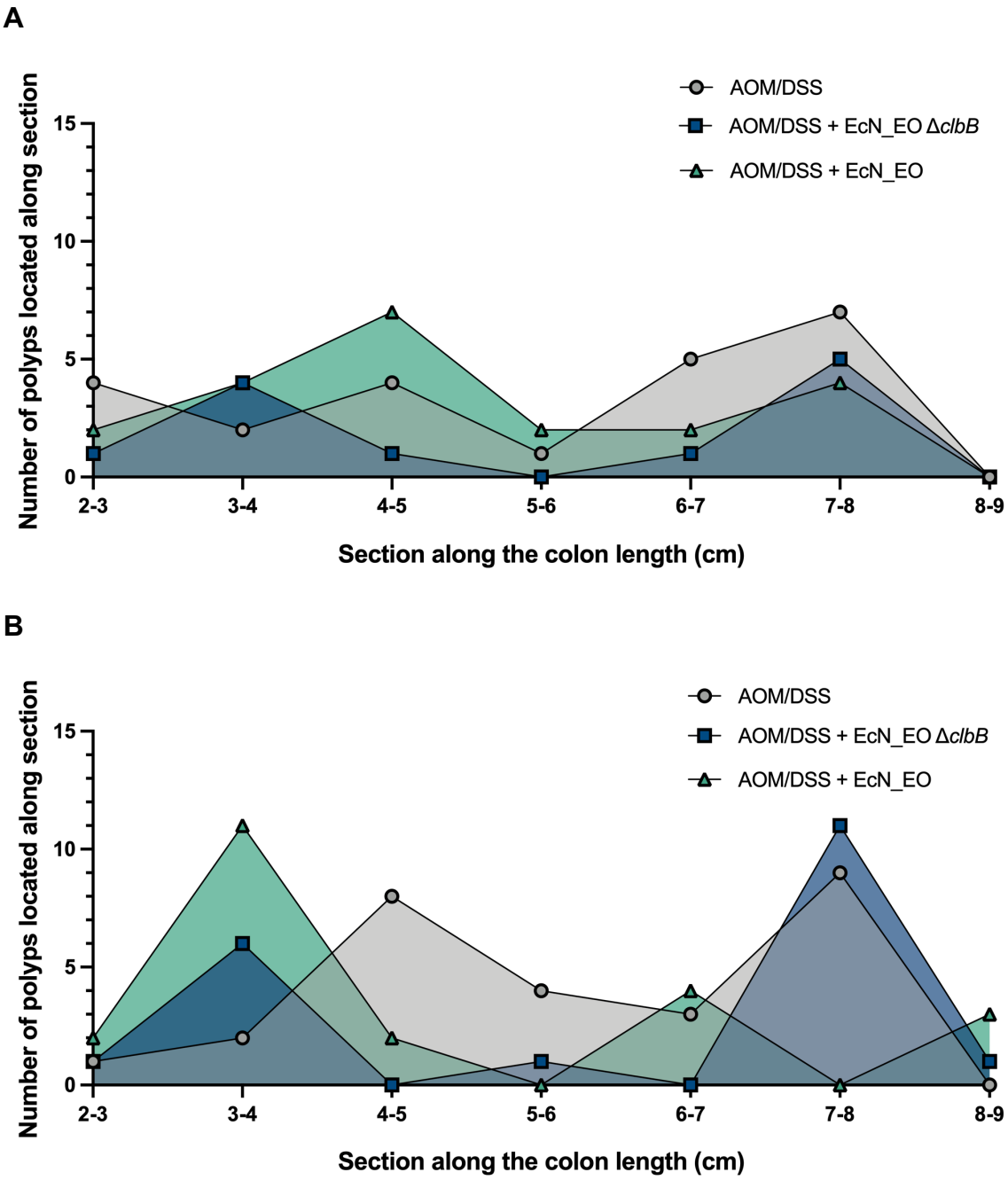


Figure 6-8 Distribution of macroscopic polyps developed in AOM/DSS mice infected with EcN WT and Δ*clbB*. Formalin-fixed colons were examined for the size, number, and distribution of polyps. (**A**) Distribution of macroscopic polyps across different treatment groups in male mice. The length of the colon was measured into 1 cm sections from the caecum to the rectum and the polyps located in each section were counted. All mice were placed under AOM/DSS treatment; some were left uninfected (gray), others were infected with EcN_EO Δ*clbB* (blue), or infected with EcN_EO mice (green). (**B**) Distribution of macroscopic polyps across different treatment groups in female mice. The number of polyps along the colon length was assessed as described above.

Following fixation with formalin, colons were made into Swiss-rolls and sent for embedding, sectioning, and H&E staining at the CRUK Scotland Institute histology facility. Microscopic assessment revealed that sections from males had, on average, the following number of polyps: 0.83 (AOM/DSS), 1.83 (AOM/DSS + EcN_EO Δ*clbB*), and 1.33 (AOM/DSS + EcN_EO); while results in females were the following: 2.17 (AOM/DSS), 2.17 (AOM/DSS + EcN_EO Δ*clbB*), and 1.33 (AOM/DSS + EcN_EO), (Figure 6-9A). In some mice polyps appeared to be absent, but – given that histological slides represent a very small section of the colon – the absence of polyps on a microscopic level does not provide conclusive information and needs to be considered alongside the previously-illustrated macroscopic data. The size of microscopic tumours was also assessed, but instead of being measured in length with a ruler, the area of the tumour was measured using a tool within the imaging software Pathomation (Pathomation.com). The following average sizes were obtained in males: 0.44 mm^2 (AOM/DSS), 0.49 mm^2 (AOM/DSS + EcN_EO Δ*clbB*), and 1.09 mm^2 (AOM/DSS + EcN_EO); and the following in females: 0.44 mm^2 (AOM/DSS), 0.59 mm^2 (AOM/DSS + EcN_EO Δ*clbB*), and 0.49 mm^2 (AOM/DSS + EcN_EO). There were no significant differences in microscopic polyp size across treatment groups in either males or females. Overall, a total of 24 polyps were microscopically assessed in males and 34 polyps in females (Figure 6-9B). This is consistent with the observation that female mice displayed a high number of polyps both at the proximal and distal end of the colon (Figure 6-8).

This experiment showed that administration of two, seven-day cycles of 2% DSS successfully induced CRC development in C57BL/6 male and female mice. The impact of infection with EcN_EO Δ*clbB* and EcN_EO on tumours development was unclear although there was some evidence that EcN_EO ameliorated colitis (Figure 6-5C), enhanced the size of polyps (Figure 6-7B and Figure 6-9B), and triggered tumour development towards the proximal end of the colon (Figure 6-8).

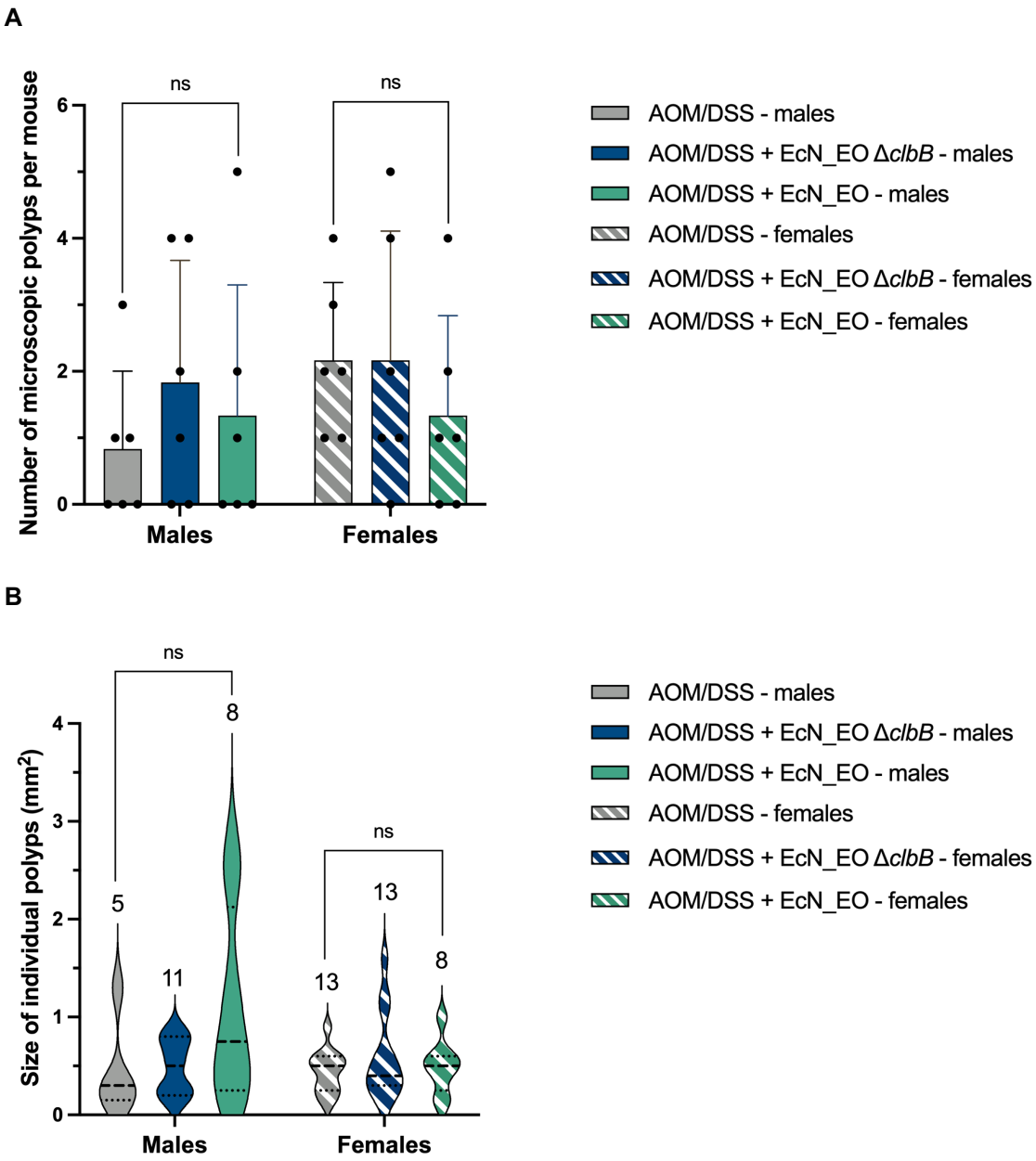


Figure 6-9 Number and size of microscopic polyps developed in AOM/DSS-treated mice infected with EcN WT and Δ*clbB*. Formalin-fixed colons were made into Swiss-rolls and sent for embedding, sectioning, and H&E staining. Scanned sections of the colons were then analysed for the number and size of the polyps. (**A**) Number of microscopic polyps across different treatment groups. The total number of microscopic tumours was counted in mice that were uninfected (gray), infected with EcN_EO Δ*clbB* (blue), and infected with EcN_EO mice (green). The values obtained in males are represented in solid colours, while the measurements obtained in females are represented with diagonal, white stripes. Data points represent individual values, bars represent average values, and error bars represent standard deviation. Statistical significance was calculated with an ordinary one-way ANOVA test. (**B**) Size of microscopic polyps across different treatment groups. The size of microscopic tumours was measured in mice that were uninfected (gray), infected with EcN_EO Δ*clbB* (blue), and infected with EcN_EO mice (green) by using an area-measuring tool on Pathomation, with all images at the same magnification. The values obtained in males are represented in solid-coloured violin plots, while the measurements obtained in females are represented with diagonal, white stripes. The numbers written above violin plots represent the total number of microscopic polyps measured within that treatment group. Statistical significance was calculated with an ordinary one-way ANOVA test.

6.2.4 Comparison of L-Serine and D-Serine palatability to C57BL/6 mice and investigation of amino acid-driven changes in the gut microbiome

The aim of these experiments was to assess whether EcN infection in mice enhanced CRCs, and if so whether treatment with D-Serine would repress the development of such tumours. However, as this would require continual exposure to the AA in the drinking water, a pilot experiment was undertaken to evaluate the impact of long-term exposure to D-Serine via this route. More specifically, the aims were to determine (1) if D-Serine was palatable to mice in their drinking water, (2) if D-Serine had an impact on DSS-elicited colitis, and (3) if addition of D-Serine altered the composition of the mouse microbiome. A total of 12 female C57BL/6 mice aged approximately 10 weeks, were divided into three cages: mice in the first cage received 2% DSS and 1 mM L-Serine, mice in the second cage received 2% DSS and 1 mM D-Serine, and mice in the third cage received 2% DSS alone. AAs were replaced daily in the drinking water, except days constituting the DSS-cycle, during which, as previously described, DSS was replenished mid-cycle, alongside AAs. Collection of weights and health checks were conducted daily (Figure 6-10A). Water consumption was also monitored by daily weighing of water bottles. As expected, mice lost considerable weight following the DSS cycle, with DSS + L-Serine mice losing an average of 14.58%, DSS + D-Serine mice losing 11.82%, and DSS mice losing on 19.13% of their body weight. Although there was no significant difference between the treatment groups, it was noted that mice whose drinking water was supplemented with D-Serine lost the least amount of weight (Figure 6-10B). No difference in the volumes of drinking water were observed across the three groups, which suggested the additional AA did not alter palatability of the water (Figure 6-10C).

To analyse the impact of AA supplementation on the microbiome of these animals, faecal samples for microbiome testing were collected before AA treatment (day 1), five days after starting the AA regimen, but before commencing the DSS cycle (day 5), and before the end of the experiment (day 13).

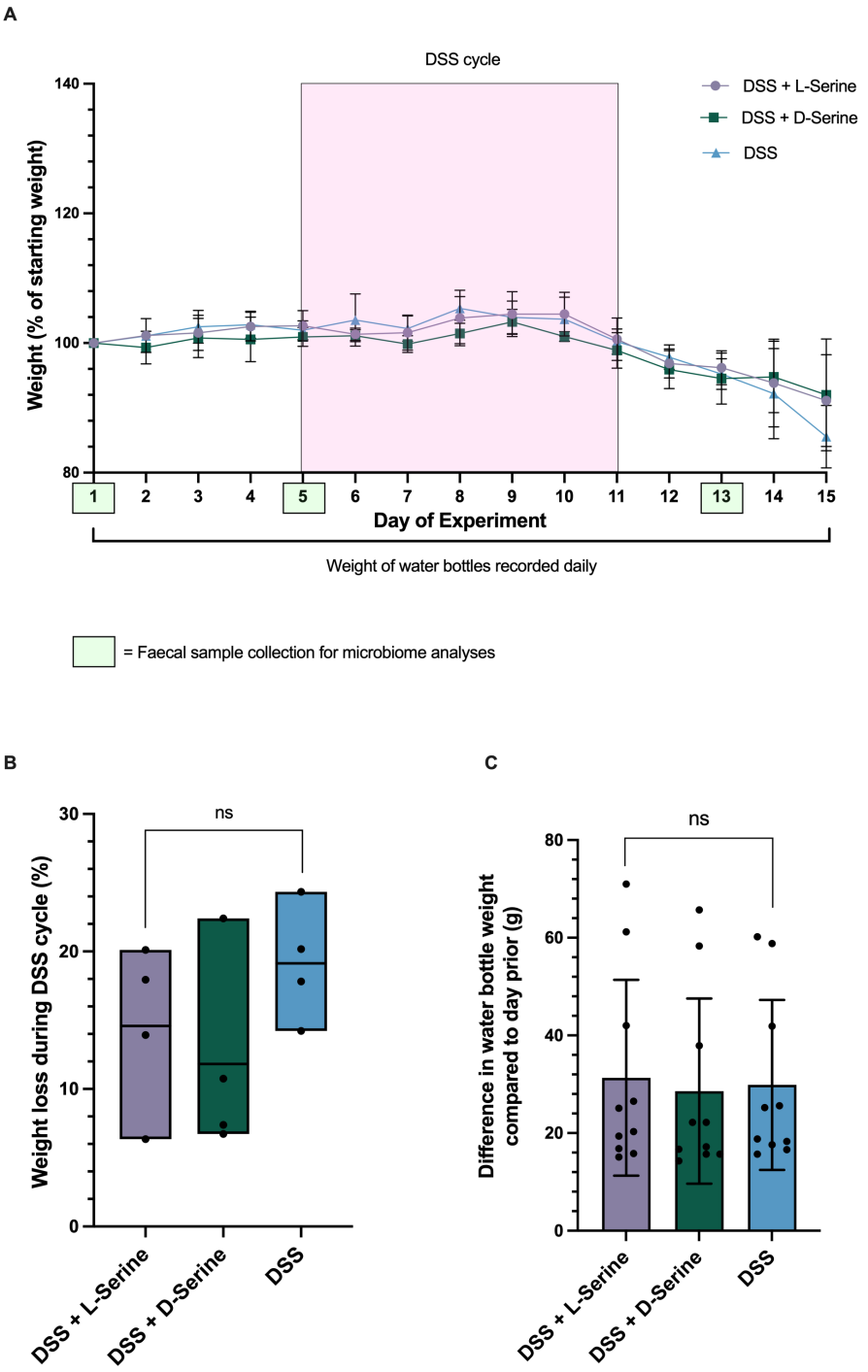


Figure 6-10 Weight changes of DSS-treated mice under different dietary regimens and assessment of AA palatability. (**A**) Daily weights of DSS-treated mice under different types of AA supplementation. Mice were placed under regular drinking water that was either supplemented with L-Serine (purple), supplemented with D-Serine (green) or not supplemented (blue). After five days, they were placed on a 7-day cycle of 2% DSS (pink section), before being culled and dissected. Mice were monitored daily for weight loss. Each treatment group was made up of four mice; data points represent average daily weights per treatment group, and error bars represent standard deviation. Green boxes represent the days in which faecal pellets were collected for microbiome analyses. (**B**) Weight loss following DSS cycles. Mice were expected to lose weight during and following DSS cycles. Their lowest weight before recovering from colitis was subtracted from their highest weight before DSS treatment. Data points represent individual values, the line within the boxplots represents the mean. Statistical significance was calculated with an ordinary one-way ANOVA test. (**C**) Changes in water bottle weight. Water bottles were weighed daily to monitor whether mice under different dietary regimens drank different amounts of water. Daily changes in water bottle weights are plotted. Data points represent individual values, bars represent the mean, error bars represent standard deviation. Statistical significance was calculated with an ordinary one-way ANOVA test.

Genomic DNA from the collected stool samples was extracted using the DNeasy PowerSoil Pro Kit (Qiagen). Samples were sequenced by Novogene Europe by 16S amplicon metagenomic sequencing. More specifically by amplification of the V3-V4 region at a sequencing depth of 30000 raw tags. Based on sequencing results, samples were assigned operational taxonomic units (OTUs), the distribution of which is represented in Venn diagrams illustrating the shared bacterial communities among the three different treatment groups. Among all treatment groups, 283 OTUs were shared at the start of the experiment. Interestingly the number of shared taxa increased to 300 OTUs when all cages were under different drinking water regimens, and instead decreased to 244 OTUs despite all mice receiving a 7-day cycle of 2% DSS. It was also interesting to note, that over the course of the experiment, mice under the D-Serine regimen increasingly shared more OTUs with mice under the DSS alone regimen. The OTUs shared between L-Serine and D-Serine mice remained similar for the most part, while those shared between L-Serine and DSS alone mice gradually decreased (Figure 6-11).

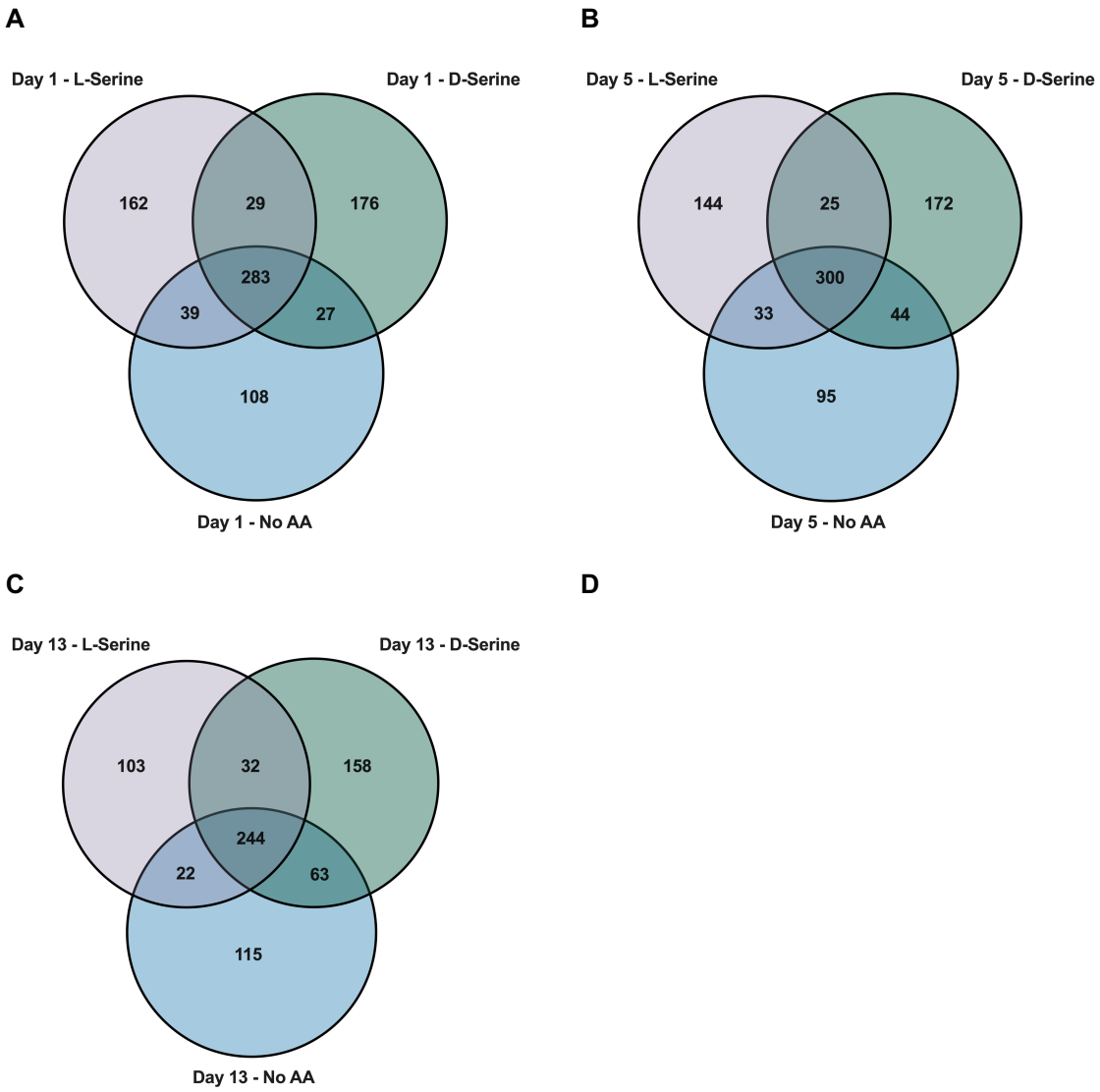


Figure 6-11 Venn diagrams illustrating the distribution of bacterial OTUs over time between different treatment groups. Based on sequencing results from 16S amplicon metagenomic sequencing, samples were assigned operational taxonomic units (OTUs). Venn diagrams represent shared OTUs between mice that were fed L-Serine (purple), mice that were fed D-Serine (green) and mice on a regular dietary regimen (blue). (**A**) Shared bacterial OTUs before mice were placed on procedure, on day 1. (**B**) Shared bacterial OTUs before mice were administered DSS, and after they were administered AAs, on day 5. (**C**) Shared bacterial OTUs following DSS treatment and following AA-enriched diets, on day 13.

A principal component analysis (PCA) plot was employed to assess similarity among the bacterial communities of each sample. Circles represent microbiome samples from the same group, collected on the same day. On day 13, all samples were clustered together and were spatially separated from the other samples. This appears to be related to DSS treatment, since sample collection on day 13 followed 2% DSS administration in mice. It's also worth noting that mice from the same treatment groups exhibited a considerable amount of variation even though their faecal samples were collected at the same time point. This indicates that, despite being housed in the same cage and being under the same dietary regimen, individual mice had a strong unique microbial composition (Figure 6-12).

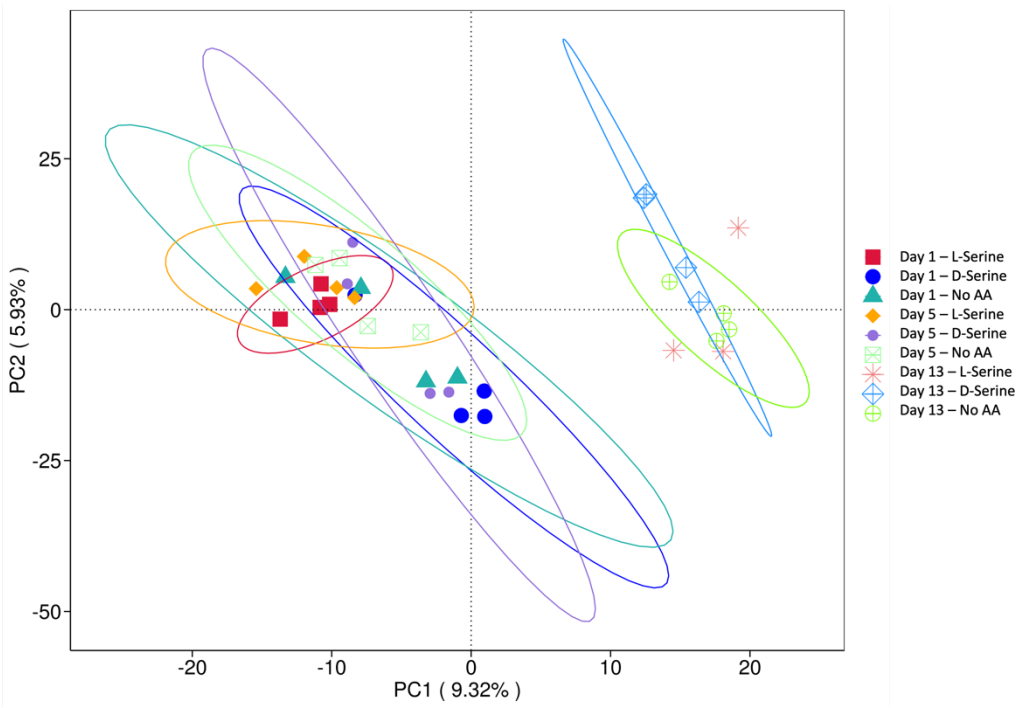


Figure 6-12 PCA results of the microbiota community. Circles represent microbiota samples from the same group, collected on the same day. Samples with comparable microbial communities are clustered together.

To compare variation of bacterial communities over time, across different treatment groups, the average relative abundance of the top ten bacterial families was evaluated. On day 1, the three most abundant families in all mice were Lachnospiraceae, Muribaculaceae, and Lactobacillaceae, though present at different levels of colonisation. For example, animals in the cage which would be subsequently treated with D-Serine, presented a larger relative abundance of Muribaculaceae and a lower relative abundance of Lachnospiraceae compared to mice in the other two treatment groups (Figure 6-13).

At the second time point, the no AA mice, which had not been exposed to a change in diet, unsurprisingly presented an almost identical microbial composition compared to their previous time point, with a small increase in Lachnospiraceae and a slight decrease in Muribaculaceae and Lactobacillaceae. It was expected that mice under the L-Serine and D-Serine supplementation regimens would exhibit a noticeable variation on day 5 compared to day 1. Instead, the microbial community of L-Serine mice remained almost unaltered, mimicking a shift similar to that of control mice: a slight increase in Lachnospiraceae and a small decrease in Muribaculaceae and Lactobacillaceae. D-Serine mice instead displayed an increase in both Lachnospiraceae and Lactobacillaceae, and a decrease in Muribaculaceae only (Figure 6-13).

The biggest shift in microbial composition, occurred following administration of 2% DSS. Mice across different treatment groups exhibited a larger relative presence of Akkermansiaceae, Erysipelotrichaceae, and Prevotellaceae. In addition, they all displayed a larger relative abundance of Bacteroidaceae, with D-Serine mice displaying this more noticeably. The different treatment groups also exhibited larger relative abundances of distinct members of the microbiome: L-Serine mice of Bifidobacteriaceae, D-Serine mice of Oscillospiraceae, and control mice of Peptostreptococcaceae (Figure 6-13).

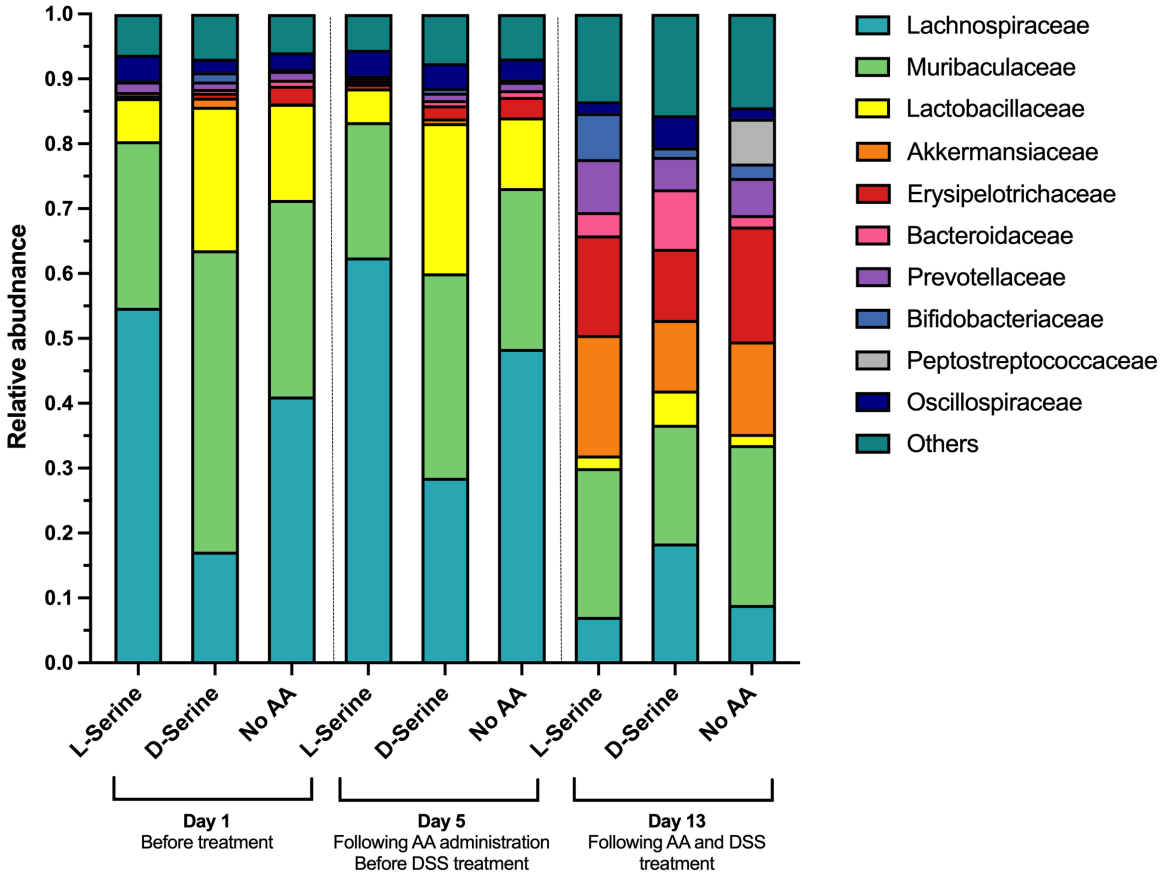


Figure 6-13 Relative abundance of the major bacterial families identified in the intestinal microbiome. A bar graph showing average relative abundances of the major bacterial families identified in the faecal microbiome of C57BL/6 mice at different time points. Different colours depict average relative levels of abundance of different microbial families in mice fed L-Serine, D-Serine, or maintained on a regular diet. Day 1 samples represent the gut microbiome before the mice were placed on procedure. Day 5 samples represent the intestinal bacterial composition after five days of regular drinking water or AA regimen. Day 13 samples represent the microbiome following DSS-treatment among all groups and AA treatment in the specified groups.

6.2.5 Assessment of colorectal cancer development in the presence of EcN_EO, and of D-Serine's effect on colibactin *in vivo*

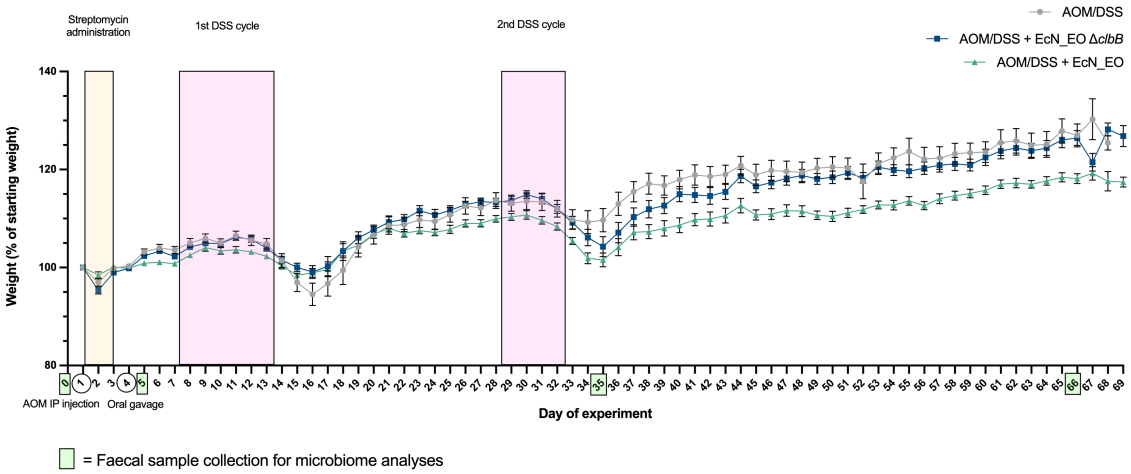
While our own work (Chapter 6.2.3) had indicated that both male and female mice appeared equally susceptible to tumour development, the decision to work with male mice was made, to allow more direct comparison with existing literature. This final experiment aimed to test (1) if EcN_EO caused an enhancement of colorectal tumours, and (2) whether addition of 1 mM D-Serine reduced the size and number of tumours. A total of 72 males between 8-10 weeks old were respectively placed into six different groups – with three cages per group and 4 mice per cage. All mice were administered 2 mg/mL of streptomycin for three days, an IP injection of 10 mg/kg of AOM, and two cycles of 2% DSS. The first DSS cycle was administered, as scheduled, for 7 days, while the second DSS cycle was reduced to 5 days due to several animals showing rapid weight loss and severe symptoms of colitis. As in previous experiments, three cages of animals received no additional treatment (referred to as AOM/DSS), three were infected with 10^9 CFU of EcN_EO Δ*clbB* (referred to as AOM/DSS + EcN_EO Δ*clbB*), and the remaining three cages were infected with 10^9 CFU of EcN_EO (referred to as AOM/DSS + EcN_EO). For each condition, there was a corresponding group that was administered 1 mM D-Serine in addition to the described treatment. A schematic representation of the experimental conditions can be found in the table below (Table 1-1).

Table 6-1 Categorisation of treatment groups in the sixth animal experiment

	Treatment		
Supplementation	2 mg/mL streptomycin AOM IP (10 mg/kg) 2 x 2% DSS Uninfected	2 mg/mL streptomycin AOM IP (10 mg/kg) 2 x 2% DSS Infected: 10^9 CFU EcN_EO Δ*clbB*	2 mg/mL streptomycin AOM IP (10 mg/kg) 2 x 2% DSS Infected: 10^9 CFU EcN_EO
No D-Serine	Cage 1 Cage 2 Cage 3	Cage 4 Cage 5 Cage 6	Cage 7 Cage 8 Cage 9
1 mM D-Serine	Cage 10 Cage 11 Cage 12	Cage 13 Cage 14 Cage 15	Cage 16 Cage 17 Cage 18

As observed previously, mice lost weight when placed on streptomycin and following both 2% DSS cycles (Figure 6-14A and 6-14B). As mentioned, the second DSS cycle was reduced to 5 days due to mice exhibiting severe weight loss. Indeed, following the first DSS cycle, five males from treatment groups not given D-Serine supplementation had to be culled due to excessive weight loss. Therefore, a more cautionary approach was taken during the second cycle, and, despite a reduction in the number of days on DSS treatment, the protocol nonetheless resulted in the development of colorectal tumours. During the first DSS cycle, the exhibited average weight losses were the following in treatment groups not supplemented with D-Serine: 14.48% (AOM/DSS), 9.03% (AOM/DSS + EcN_EO Δ*clbB*), and 8.04% (AOM/DSS + EcN_EO); and the following in the treatment groups with D-Serine: 7.39% (AOM/DSS), 8.00% (AOM/DSS + EcN_EO Δ*clbB*), and 7.96% (AOM/DSS + EcN_EO). As observed in the previous experiments, mice infected with EcN lost significantly less weight ($p = 0.0436$) compared to uninfected controls (Figure 6-15A). This seems to corroborate the existing literature in support of EcN's probiotic, anti-inflammatory effect *in vivo*. However, this was observed only in males that were not supplemented with D-Serine. During the second DSS cycle instead, average weight losses were the following in groups not supplemented with D-Serine: 6.40% (AOM/DSS), 11.30% (AOM/DSS + EcN_EO Δ*clbB*), and 9.91% (AOM/DSS + EcN_EO); and the following in the treatment groups with D-Serine: 10.43% (AOM/DSS), 9.56% (AOM/DSS + EcN_EO Δ*clbB*), and 11.35% (AOM/DSS + EcN_EO). It was also interesting to note that after the first DSS cycle, AOM/DSS mice whose drinking water was supplemented with D-Serine lost on average significantly less weight than their no D-Serine counterparts ($p = 0.0201$). Although this was not observed in any other treatment group, it suggested that D-Serine might be having an effect on the intestinal mucosa of mice, beyond its effect on the colibactin transcription of EcN. It's also interesting to note that, in the second DSS cycle, weight loss was larger in EcN-infected mice whose diet included D-Serine than in those whose diet did not include the AA (Figure 6-15B). This might suggest that the anti-inflammatory effect elicited by EcN is being reduced by D-Serine's repressive activity on the genotoxin's biosynthesis

A



B

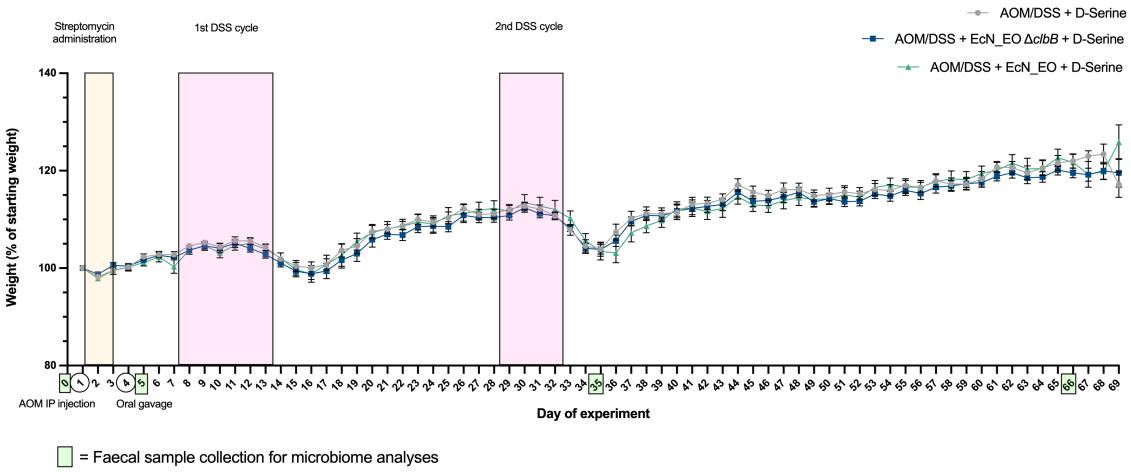


Figure 6-14 Weight changes of AOM/DSS treated mice under different infection procedures and under different dietary regimens. (**A**) Daily weights of AOM/DSS-treated males under different infection procedures and under a regular diet. Following an IP injection with 10 mg/kg of AOM and administration of 2 mg/mL of streptomycin (yellow section), mice were left uninfected (grey), gavaged with 10^9 CFU/dose of EcN_EO Δ*clbB* (blue), or gavaged with 10^9 CFU/dose of EcN_EO (green). They were then treated with two, 7-day cycles of 2% DSS (pink sections), before being culled and dissected. Mice were monitored daily for weight loss. Each treatment group was made up of three cages containing four mice each; data points represent average daily weights per treatment group, and error bars represent standard deviation. Green boxes represent the days in which faecal pellets were collected for microbiome analyses. (**B**) Daily weights of AOM/DSS-treated males under different infection procedures and under a D-Serine supplemented diet. Male mice were treated and infected as described above. They were also placed on drinking water supplemented with 1 mM of D-Serine for the whole duration of the experiment. Data points represent average daily weights per treatment group and error bars represent standard deviation. Green boxes represent the days in which faecal pellets were collected for microbiome analyses.

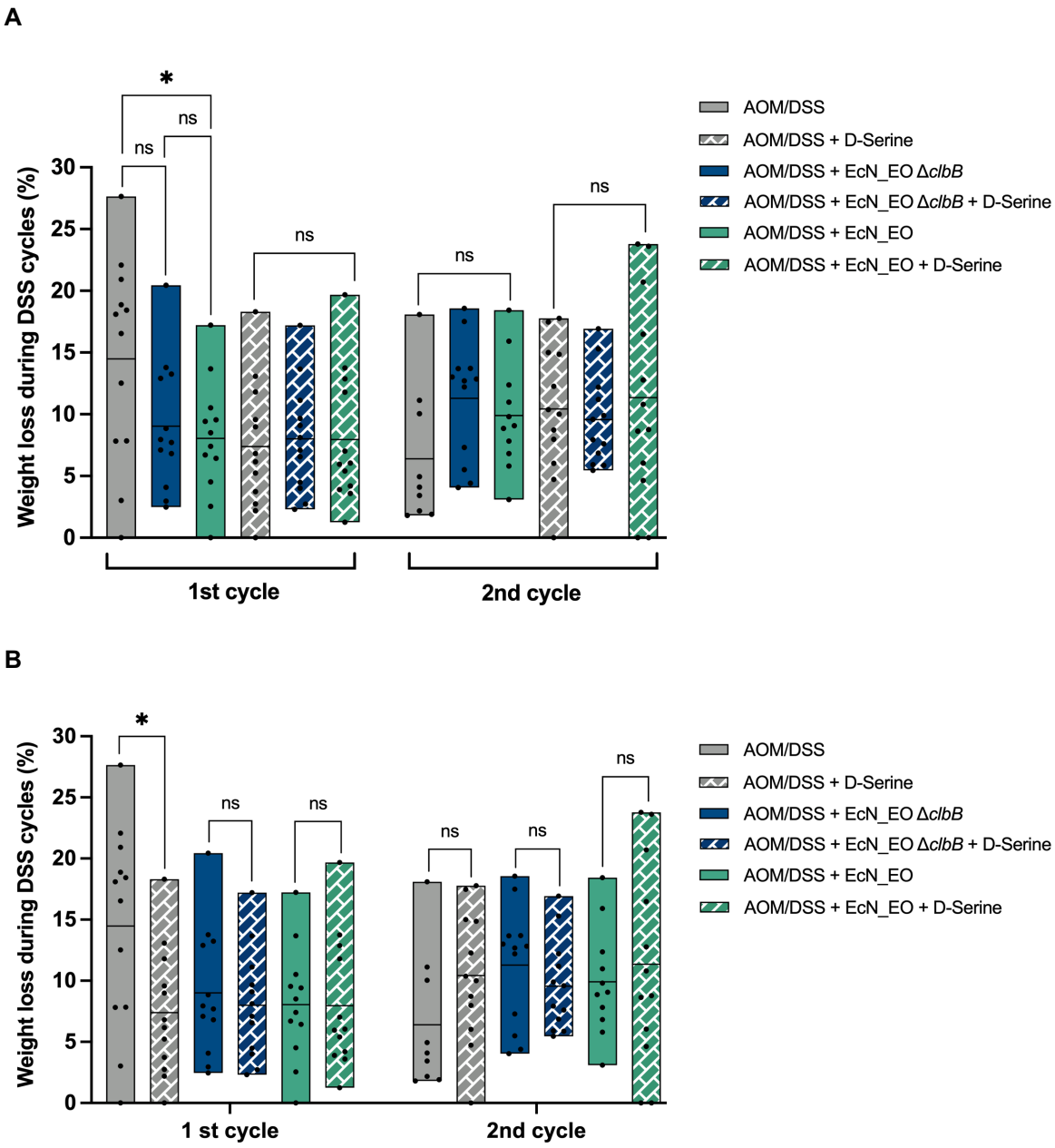


Figure 6-15 Weight loss following DSS cycles in AOM/DSS treated mice under different infection procedures and under different dietary regimens. (**A**) DSS-induced weight loss in mice under different infection procedures. Mice achieved their lowest weights following DSS-induced colitis. Their lowest weight before recovering from colitis was subtracted from their highest weight before DSS treatment. All mice were treated with AOM/DSS and, in addition, some were left uninfected (gray), some were infected with EcN_EO Δ*clbB* (blue), and others were infected with EcN_EO (green). The values obtained in mice placed on a regular diet are represented in solid colours, while the measurements obtained in D-Serine-fed mice are represented with a white pattern. Data points represent individual values, the line within the boxplots represents the mean. Statistical significance was calculated with an ordinary one-way ANOVA test and, when significance was detected within a three-way comparison, individual comparisons were performed with a Tukey's multiple comparisons test. Significance is represented with * indicating $p < 0.05$. (**B**) DSS-induced weight loss in mice under different dietary regimens. The percentage of weight loss was calculated as described above. . Data points represent individual values, the line within the boxplots represents the mean. Mice under the same infection procedure but under different dietary treatments were compared with a Student's t-test, with significance represented by *, indicating $p < 0.05$.

On dissection, tumours were macroscopically assessed as previously described. Assessment revealed that AOM/DSS mice developed, on average, 7.25 (no AA) and 5.08 (D-Serine) polyps; AOM/DSS + EcN_EO Δ*clbB* mice grew on average 9.08 (no AA) and 5.42 (D-Serine) tumours; and AOM/DSS + EcN_EO mice on average developed 4.64 (no AA) and 5.50 (D-Serine) polyps (Figure 6-16A). Tumours were also macroscopically assessed for their size. Results revealed that AOM/DSS and AOM/DSS + EcN_EO Δ*clbB* mice whose diet had been supplemented with D-Serine developed significantly smaller polyps than their no AA counterparts, with tumours that were, on average, 25.83% ($p = 0.0015$) and 24.31% ($p = 0.0003$) smaller. Among the EcN-infected males, there was no difference in polyp size between cages that had been administered D-Serine, and those that had not received the dietary supplement (Figure 6-16B). Results also showed that – while D-Serine supplemented mice showed no difference in polyp size among different treatment groups – in the no AA category, EcN-infected mice had significantly smaller polyps compared to those infected with the mutant strain ($p < 0.0001$) and those that were uninfected ($p = 0.0005$), (Figure 6-16C). This result corroborates our previous observation that EcN seems to ameliorate outcomes in AOM/DSS models.

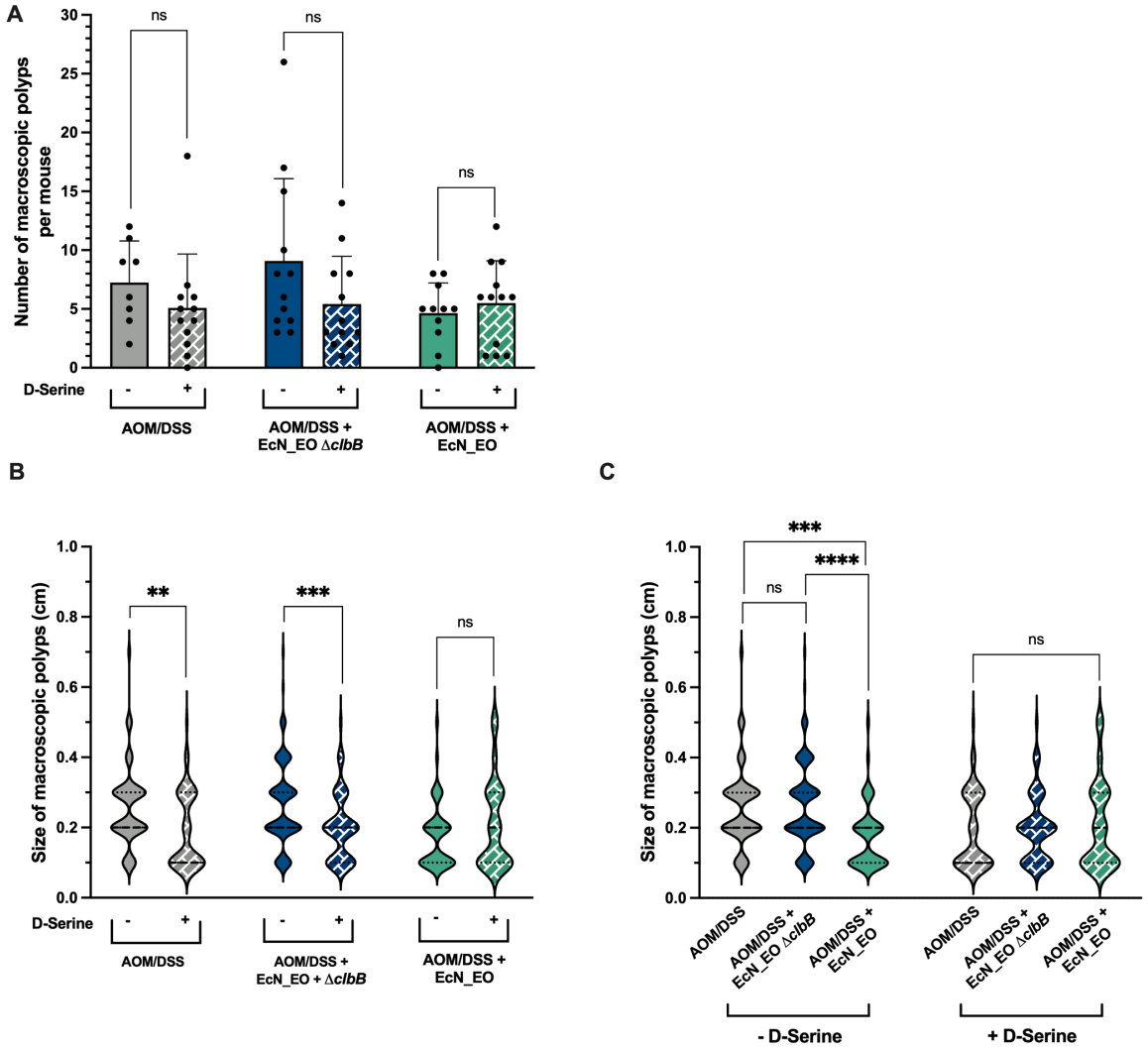


Figure 6-16 Number and size of macroscopic polyps developed in AOM/DSS-treated mice infected with EcN WT and Δ*clbB* and fed D-Serine. Formalin-fixed colons were examined for the size, number, and distribution of polyps. (**A**) Number of macroscopic polyps across different treatment groups. The total number of macroscopic tumours was counted in mice that were uninfected (gray), infected with EcN_EO Δ*clbB* (blue), and infected with EcN_EO mice (green). The values obtained in mice placed on a regular diet are represented in solid colours, while the measurements obtained in D-Serine-fed mice are represented with a white pattern. Data points represent individual values, bars represent average values, and error bars represent standard deviation. Comparisons were performed between mice placed under the same infection treatment, but under different dietary regimens. Statistical significance was calculated with multiple Student's t-test. (**B**) Size of macroscopic polyps in mice under different dietary regimens. The size of colorectal polyps was macroscopically assessed with a ruler. The values obtained in mice placed on a regular diet are represented in solid colours, while the measurements obtained in D-Serine-fed mice are represented with a white pattern. Statistical significance was calculated with Student's t-tests with ** and *** respectively indicating $p < 0.01$ and $p < 0.001$. (**C**) Size of macroscopic polyps in mice under the same dietary regimen, but under different infection procedures. Values were measured and are represented as described above. Statistical significance was calculated with an ordinary one-way ANOVA test and, when significance was detected within a three-way comparison, individual comparisons were performed with a Tukey's multiple comparisons test. Significance is represented with *** and **** respectively indicating $p < 0.001$ and $p < 0.0001$.

To determine if D-Serine supplementation altered the region of the gut in which colorectal tumours developed, their distribution along the colon was assessed. As previously observed, most polyps were located towards the lower end of the colon, both in mice that were fed D-Serine, and those that received regular drinking water. Mice infected with the *clbB* mutant appeared to develop considerably more polyps in the mid-to-low end of the colon compared to their D-Serine-fed counterparts (Figure 6-17B), even though there was no significant difference in the average number of polyps per group (Figure 6-16A). It was interesting to note the exceptionally similar profiles shared by EcN-infected mice across both treatment conditions (Figure 6-17C). It aligns with previous data showing that EcN groups have similar numbers (Figure 6-16A) and sizes (Figure 6-16B) of polyps both under D-Serine supplementation and under no AA conditions.

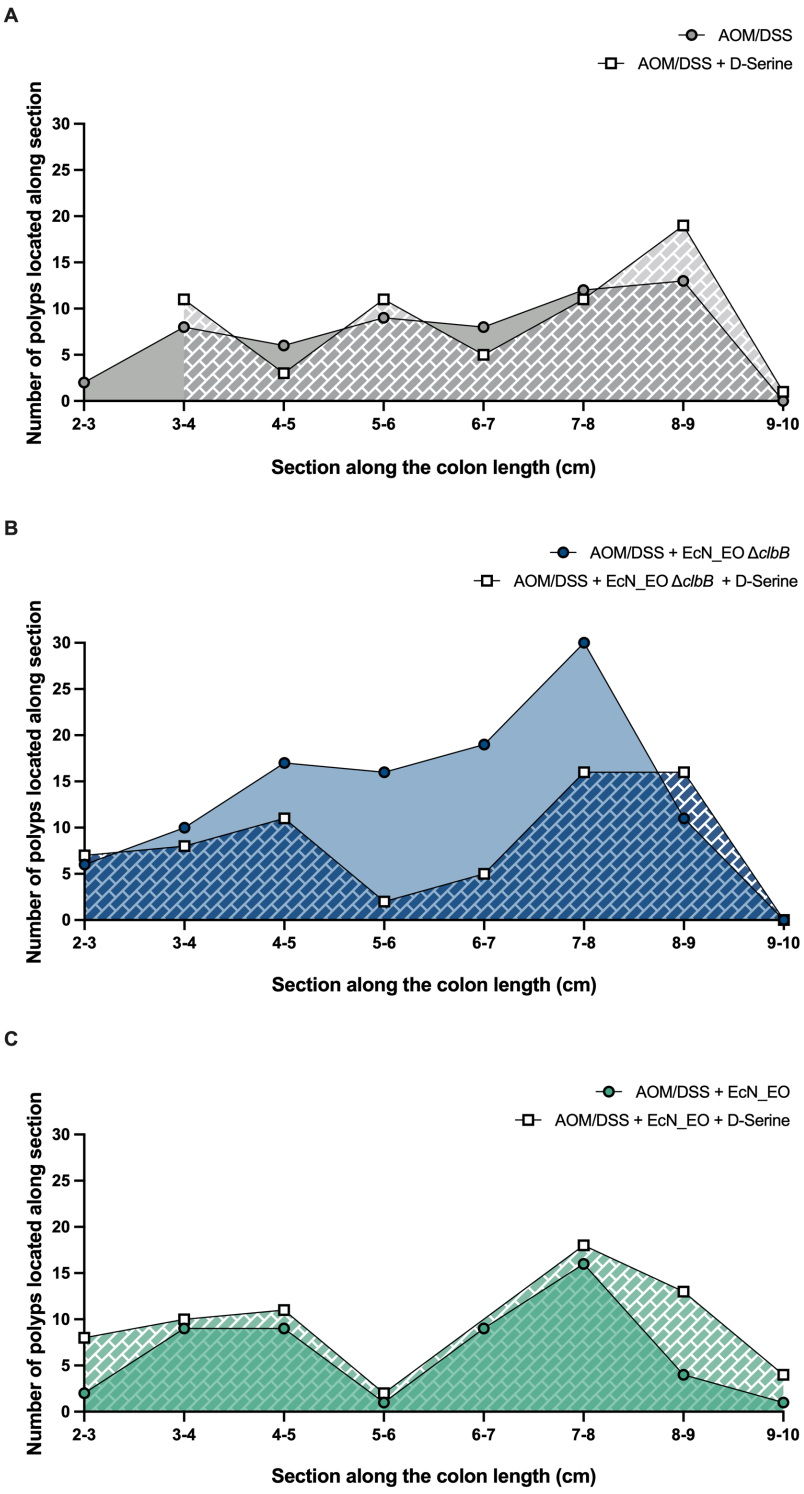


Figure 6-17 Distribution of macroscopic polyps developed in AOM/DSS-treated mice infected with EcN WT and Δ*clbB* and fed D-Serine. Formalin-fixed colons were examined for the size, number, and distribution of polyps. (**A**) Distribution of macroscopic polyps in uninfected AOM/DSS-treated mice, +/- D-Serine. The length of the colon was measured in 1 cm sections from the caecum to the rectum and the polyps located in each section were counted and plotted. The gray data set representes mice on a standard diet, while the patterned data represents mice on a D-Serine supplemented diet. (**B**) Distribution of macroscopic polyps in AOM/DSS-treated mice infected with EcN_EO Δ*clbB* +/- D-Serine. The number of polyps along the colon length was assessed as described above, the patterned data set represent mice on the D-Serine supplemented diet. (**C**) Distribution of macroscopic polyps in AOM/DSS-treated mice infected with EcN_EO +/- D-Serine. The number of polyps along the colon length was assessed as described above, the patterned data set represent mice on the D-Serine supplemented diet.

H&E stained sections of formalin-fixed colons were blindly scored for inflammation and neoplasia by a senior veterinary clinician. A full description of the grading scheme used to assess the degree of inflammation and neoplasia can be found in the appendix (page 261). The most severe microscopic lesions were also evaluated and, whereas AOM/DSS mice and EcN_EO Δ*clbB*-infected mice, developed larger and more severe lesions when not supplemented with D-Serine, EcN-infected mice, once again, displayed comparable outcomes under both treatment conditions (Figure 6-18A). Interestingly, EcN-infected mice that were treated with D-Serine had a significantly ($p = 0.0249$) higher median inflammation score of 14, compared to EcN-infected mice who were fed regular drinking water and who scored 8 (Figure 6-18B). This result supports our previous hypothesis that the anti-inflammatory effect elicited by EcN in this context, is being modified by repression of colibactin biosynthesis by D-Serine. It was also interesting to notice that on average, neoplasia scores in EcN-infected mice were lower than in uninfected or mutant-infected mice, albeit not significantly (Figure 6-18C). To determine overall inflammation scores, samples were checked for certain parameters and assigned numerical values based on the degree and distribution of each inflammation-associated feature. Parameters associated with inflammation included: epithelial changes, infiltration of neutrophils and lymphocytes, crypt hyperplasia and loss of goblet cells, distorted crypt architecture, crypt mitotic activity, and mucosal fibrosis and atrophy. To determine overall neoplasia scores, the most severe masses in each section were categorised into hyperplasia, adenoma, or carcinoma, and further classified based on the lesions' stage and severity. Numerical values in intervals of 0.5 were assigned to all graded lesions. Representative images of the scored parameters can be found in Figure 6-19. Panels A and B show a healthy colon, derived from a mouse that was housed with regular food and water for 70 days, without being subjected to any treatment. The healthy colon is identified by the consistent parallel crypt orientation and by the presence of regular folds in the upper region of the colon (Figure 6-19A and B). The representative images in panels C to N instead show a mouse from each treatment group in the sixth animal experiment: average inflammation scores were calculated for each group and the mouse whose individual score was closest to the average was selected to represent the average outcome from each treatment.

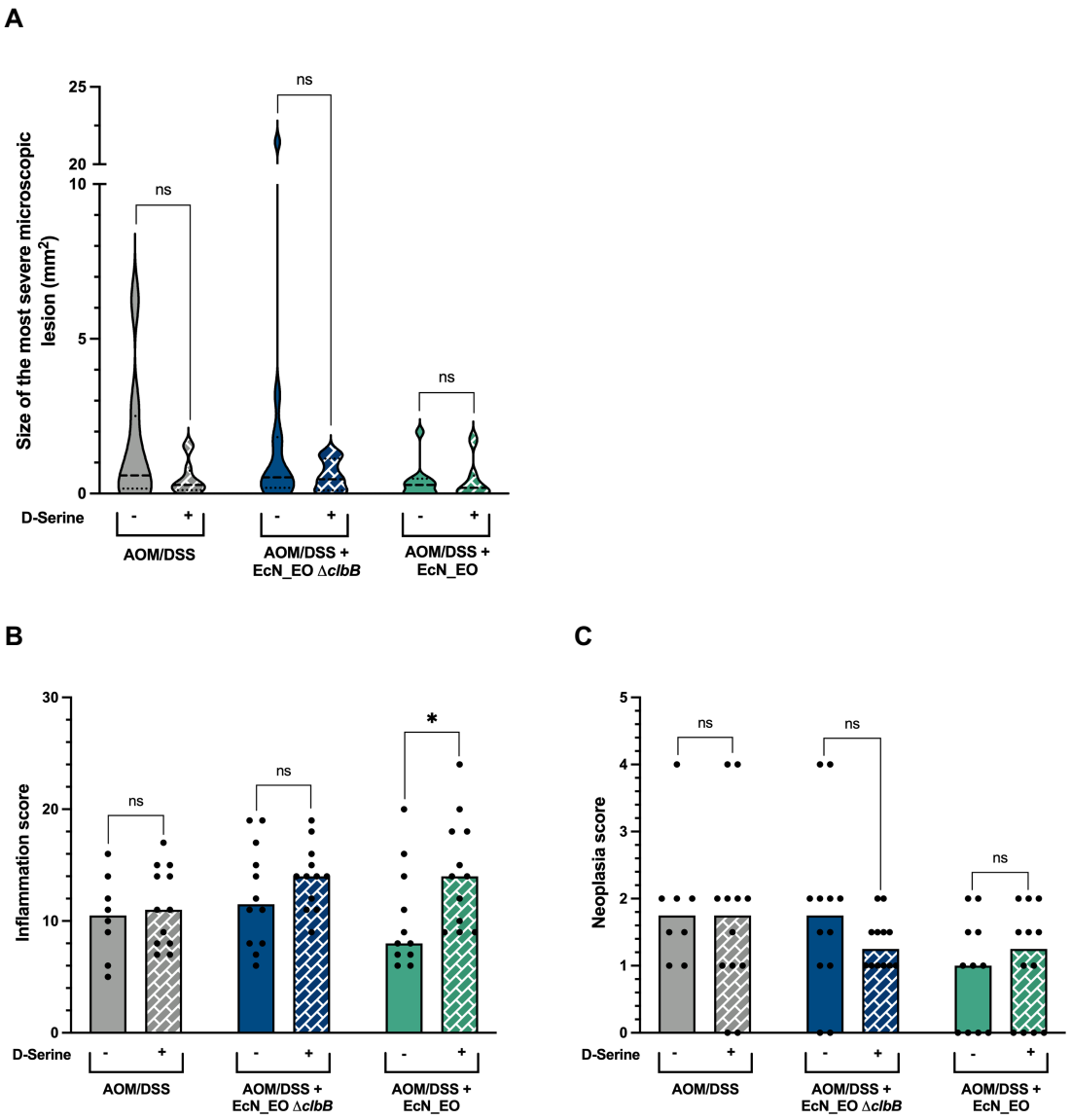
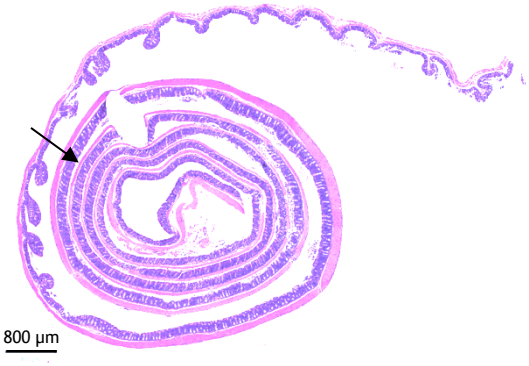
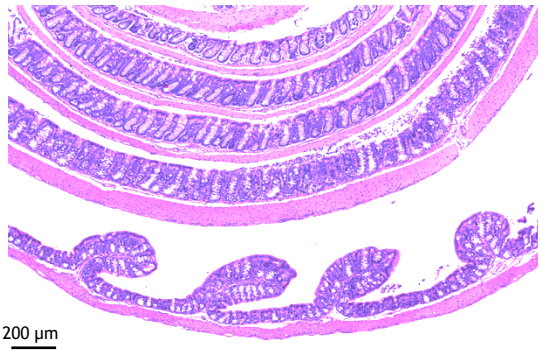


Figure 6-18 Microscopic assessment of mouse colons. Microscopic lesions and parameters of inflammation were blindly scored by a senior veterinary pathologist. (**A**) Violin plots representing the sizes of the most severe microscopic lesions detected in mouse colons. The size of the most severe miroscopic lesion was assessed in mice that, in addition to being treated with AOM/DSS, were left uninfected (gray), infected with EcN_EO Δ*clbB* (blue), or infected with EcN_EO (green). The values obtained in mice placed on a regular diet are represented in solid colours, while the measurements obtained in D-Serine-fed mice are represented with a white pattern. (**B**) Several inflammation parameters were assessed for their severity and distribution and then summed together. Data points represent individual values, bars represent median values. Comparisons were performed between mice placed under the same infection treatment, but under different dietary regimens. Statistical significance was calculated with multiple Mann-Whitney tests, with significance indicated by * indicating $p < 0.05$. (**C**) The most severe masses in each section were categorised into hyperplasia, adenoma, or carcinoma, and further classified based on the lesions' stage and severity. Data points represent individual values, bars represent median values. Comparisons were performed between mice placed under the same infection treatment, but under different dietary regimens. Statistical significance was calculated with multiple Mann-Whitney tests.

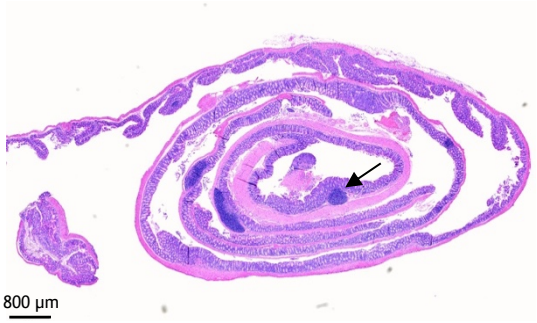
A



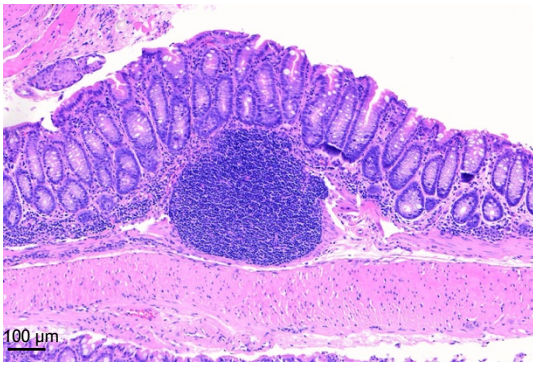
B



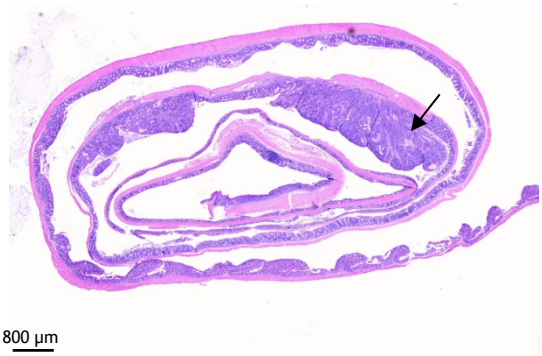
C



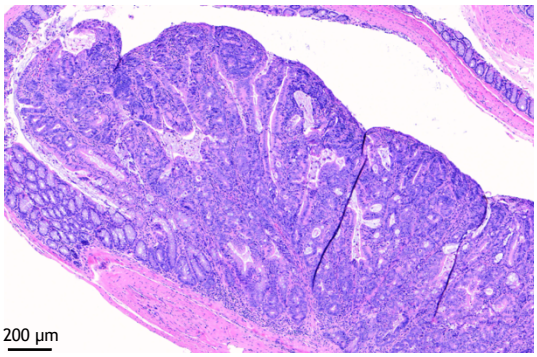
D



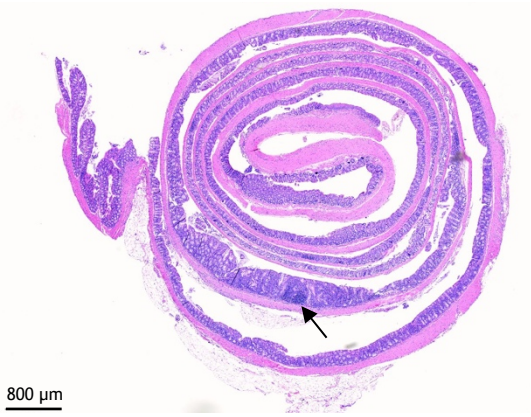
E



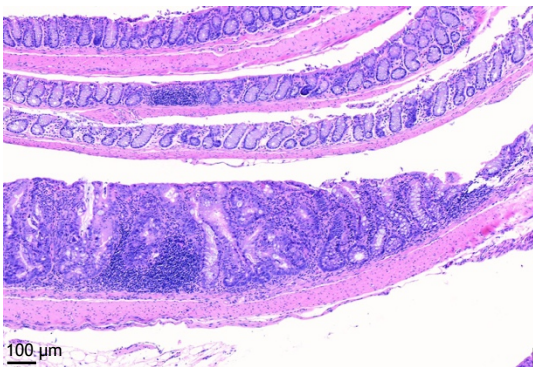
F



G



H



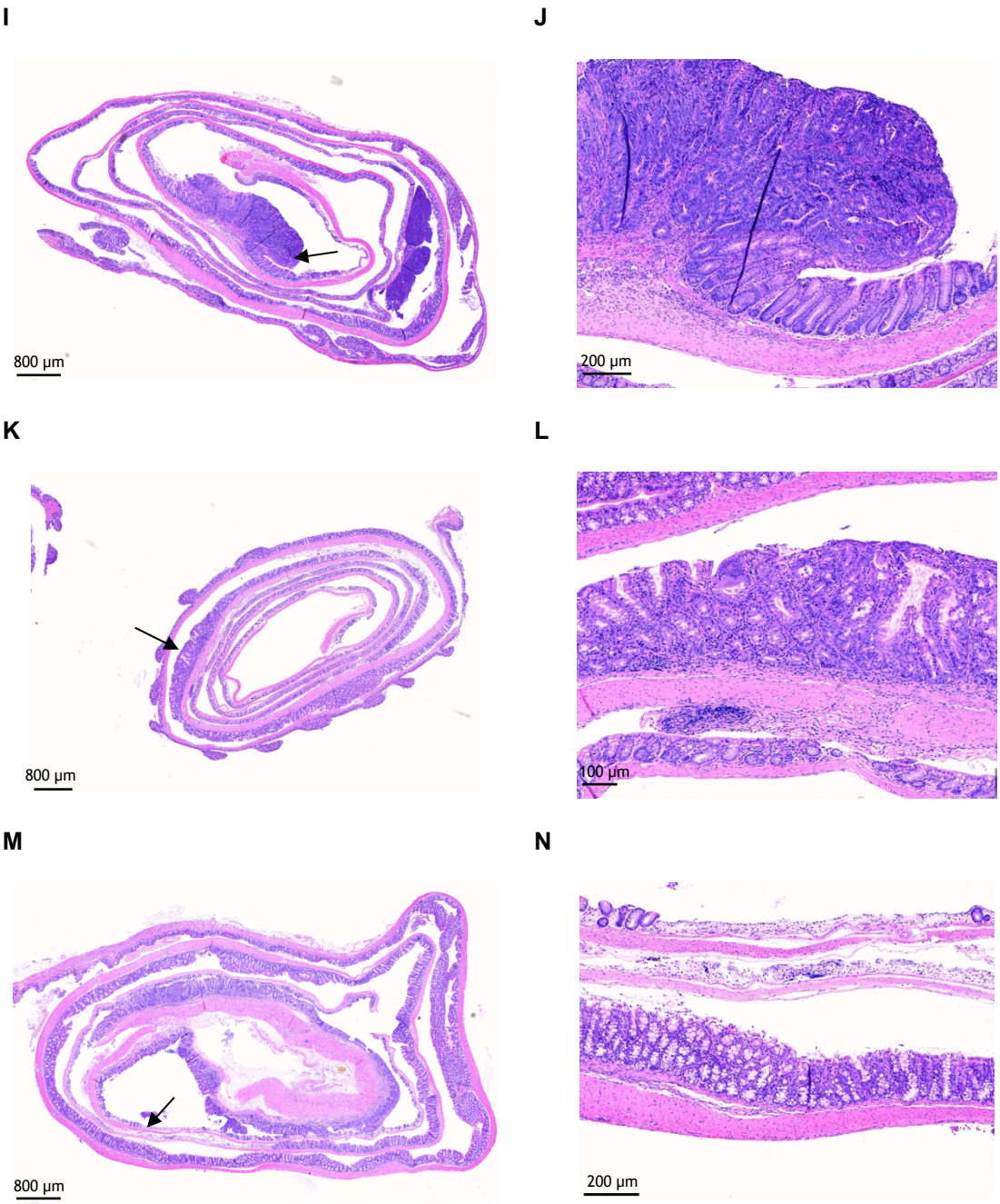
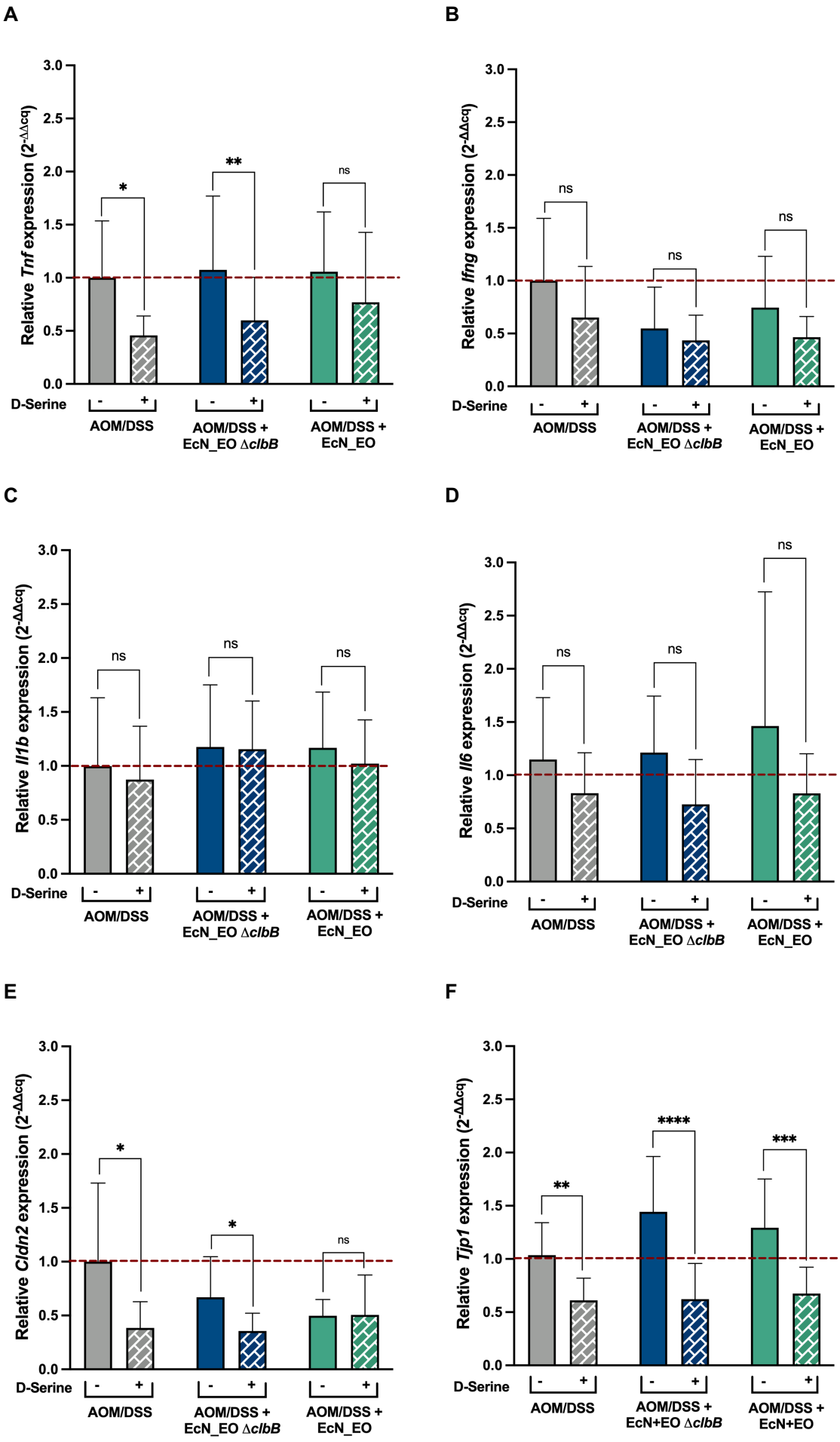


Figure 6-19 Representative histological images of the parameters scored to assess inflammation and neoplasia. (**A**) Representative image of a healthy colon and (**B**) representation of regular and healthy intestinal crypts. (**C**) Colon from a mouse treated with AOM/DSS and (**D**) a display of immune cell influx in a region of the colon. (**E**) Colon from a mouse treated with AOM/DSS and infected with EcN_EO Δ*clbB;* and (**F**) the representation of an early carcinoma in the colon. (**G**) Colon from a mouse treated with AOM/DSS and infected with EcN_EO; and (**H**) a display of immune influx and hyperplasia in the colon. (**I**) Colon from a mouse treated with AOM/DSS and fed a diet supplemented with 1 mM D-Serine; and (**J**) a representation of a carcinoma in the colon. (**K**) Colon from a mouse treated with AOM/DSS, infected with EcN_EO Δ*clbB* and fed a diet supplemented with 1 mM D-Serine; and (**L**) an image displaying one of its regions of hyperplasia. (**M**) Colon from a mouse treated with AOM/DSS, infected with EcN_EO and fed a diet supplemented with 1 mM D-Serine; and (**N**) a region of fibrosis and epithelial destruction in the colon. The black arrows in the images on the left, represent the regions in which the respective images on the right were taken. Full colons are represented with scale bars of 800 µm and zoomed in colons are represented with scale bars of 100 µm or 200 µm, depending on the size of the displayed feature.

To determine the host response to the different infection and treatment regimens, on final dissection, a 1 cm piece of colon was collected from the region immediately below the caecum. It was immediately submerged in RNAlater, placed on ice, and then transferred to -20°C. Colonic tissues were lysed in Trizol using a bead-beater; RNA was then extracted using the RNeasy Mini Kit (Qiagen) and converted to cDNA with the high capacity RT kit (Invitrogen). Several RT-qPCRs were performed to test the expression of cytokines and genes relating to gut barrier integrity. In all RT-qPCRs, expression values were assessed relative to the housekeeping ribosomal protein gene *Rps29*, and measurements were normalised to those of uninfected mice under the regular drinking water regimen, regarded as controls.

Expression of the following pro-inflammatory cytokines *Tnf*, *Ifng*, *Il1b*, and *Il6* was tested for modification of expression in the colonic tissue. Across all treatment groups, every cytokine was, on average, expressed at a lower level in D-Serine-fed mice, compared to their counterparts with no AA supplementation. This was especially noticeable with the expression of *Tnf*, that was significantly lower in uninfected ($p = 0.0123$) and mutant-infected mice ($p = 0.0055$) under the D-Serine regimen. Aligned with this result, is the significantly lower expression of the tight junction protein gene *Cldn2* in uninfected ($p = 0.0252$) and mutant-infected mice ($p = 0.0097$) that were fed D-Serine. TNFα is known to increase upregulation of Claudin-2 which in turn is linked to inflammation and to increased gut permeability (Maryam *et al.*, 2023), (Luettig *et al.*, 2015). The fact that macroscopic polyps were, on average, fewer and smaller in mice who were administered D-Serine (Figure 6-16A and B) – with the exception of EcN – seemed to suggest that D-Serine has an ameliorative impact on AOM/DSS-induced disease that is independent from its effect on colibactin. This is corroborated by the lower cytokine expression in all mice who had a D-Serine supplemented diet (Figure 6-20A – 20D). Interestingly though, the decreased tumour burden and the lower cytokine expression are not correlated to lower inflammation or neoplasia scores (Figure 6-18B and C). The other tight junction protein genes *Tjp1* and *Cdh1* were starkly, consistently expressed at significantly lower levels in D-Serine-fed mice across all treatment groups (Figure 6-20F and G). Tight junction proteins are known to be involved in the maintenance of intestinal barrier integrity so a downregulation of their gene expression is associated to a higher tumour

progression (Kumar Ram, Vairappan, 2022) (Christou *et al.*, 2017). However, this result is inconsistent with the decreased cytokine expression, which suggests lower inflammation in D-Serine fed mice. One potential explanation is that mice that were not fed D-Serine displayed a higher expression of genes involved in the maintenance of intestinal epithelial morphology due to the need to repair their gut barrier. On the other hand – as suggested by other data – the intestinal mucosa of D-Serine-fed mice appears to be less compromised, and consequently their need to upregulate genes involved in junctional integrity permeability is lower. Interestingly, expression of F11r, also involved in the maintenance of tight junctions (Lauko *et al.*, 2020), and of S1008A, which plays a role in regulating inflammation and immune responses (Mizoguchi *et al.*, 2021), were not significantly different among treatment conditions (Figure 6-20H and I).


A
Relative Tnf expression (2-ΔΔcq)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
*
**
ns
D-Serine
-
+
AOM/DSS
AOM/DSS + EcN_EO ΔclbB
AOM/DSS + EcN_EO
B
Relative Ifng expression (2-ΔΔcq)
ns
ns
ns
C
Relative Il1b expression (2-ΔΔcq)
ns
ns
ns
D
Relative Il6 expression (2-ΔΔcq)
ns
ns
ns
E
Relative Cldn2 expression (2-ΔΔcq)
*
*
ns
F
Relative Tjp1 expression (2-ΔΔcq)
**
****
***
AOM/DSS + EcN+EO ΔclbB
AOM/DSS + EcN+EO

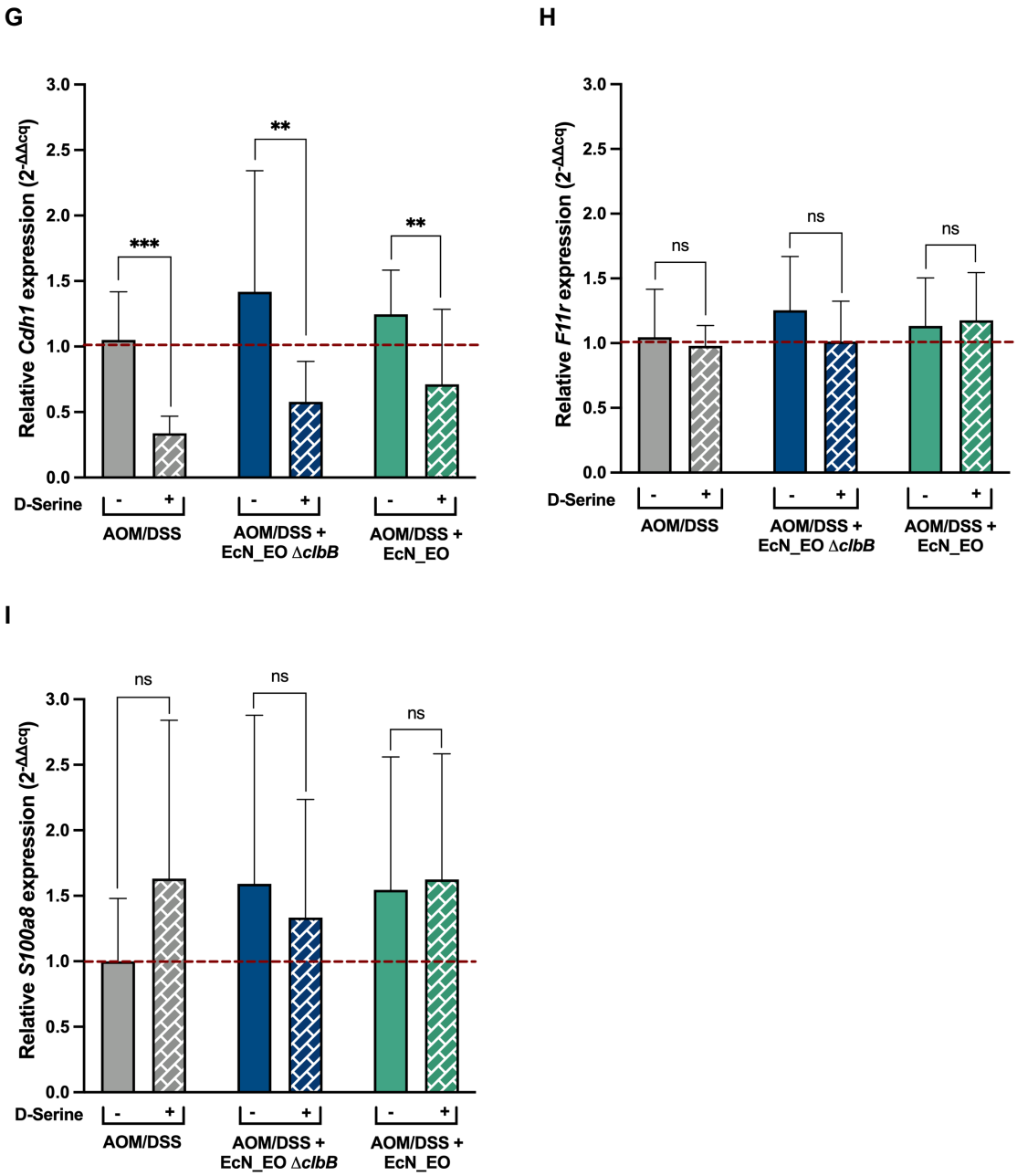


Figure 6-20 The expression of different cytokines and gut barrier genes screened by RT-qPCR. A 1 cm section close to the caecum was snipped off from each colon during dissection. The tisse was homogenised to extract RNA which was then converted to cDNA. The expressions of (A) *Tnf*, (B) *Ifng*, (C) *Il1b*, (D) *Il6*, (E) *Cldn2*, (F) *Tjp1*, (G) *Cdh1*, (H) *F11r*, (I) *S100a8* were measured relative to the housekeeping gene *Rps29*, using a control condition of AOM/DSS treatment with no D-Serine supplementation as the calibrator for baseline expression, as indicated by the dashed red line. Statistical significance was determine through Student's t-tests between the same procedure groups, but under different AA regimens. Significance is indicated by *, **, ***, and ****, respectively indicating $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$.

To compare variation of bacterial communities over time across different treatment groups, the average relative abundance of the top ten bacterial families was evaluated. This was determined as described previously by 16S amplicon metagenomic sequencing of faecal samples. This analysis revealed that there were no notable changes over time or differences in microbial compositions between D-Serine treated mice and animals on regular diets. On day 5, oral gavages with EcN_EO Δ*clbB* and EcN_EO are clearly visible in the microbial community of the relevant cages, as indicated by the family Enterobacteriaceae. In later time points, Enterobacteriaceae no longer appear among the top ten most abundant bacterial families in the mouse samples, suggesting that the colonisation of EcN_EO Δ*clbB* and EcN_EO has largely dropped. It's also interesting to note that on day 5 uninfected mice under the regular dietary regimen displayed a decrease in relative abundance of Lactobacillaceae, as did mice infected EcN_EO Δ*clbB* and EcN_EO and placed under D-Serine supplementation. All groups exhibited a larger relative abundance of Clostridiaceae, and – similarly to what was observed previously (Figure 6-13) – an increase in relative abundance of Akkermansiaceae, and Erysipelotrichaceae following DSS cycles. Interestingly, by the end of the experiment, mice infected with EcN_EO were the ones displaying the largest average relative abundance of Erysipelotrichaceae and the lowest relative abundance of Bacteroidaceae, independently from D-Serine supplementation (Figure 6-21).

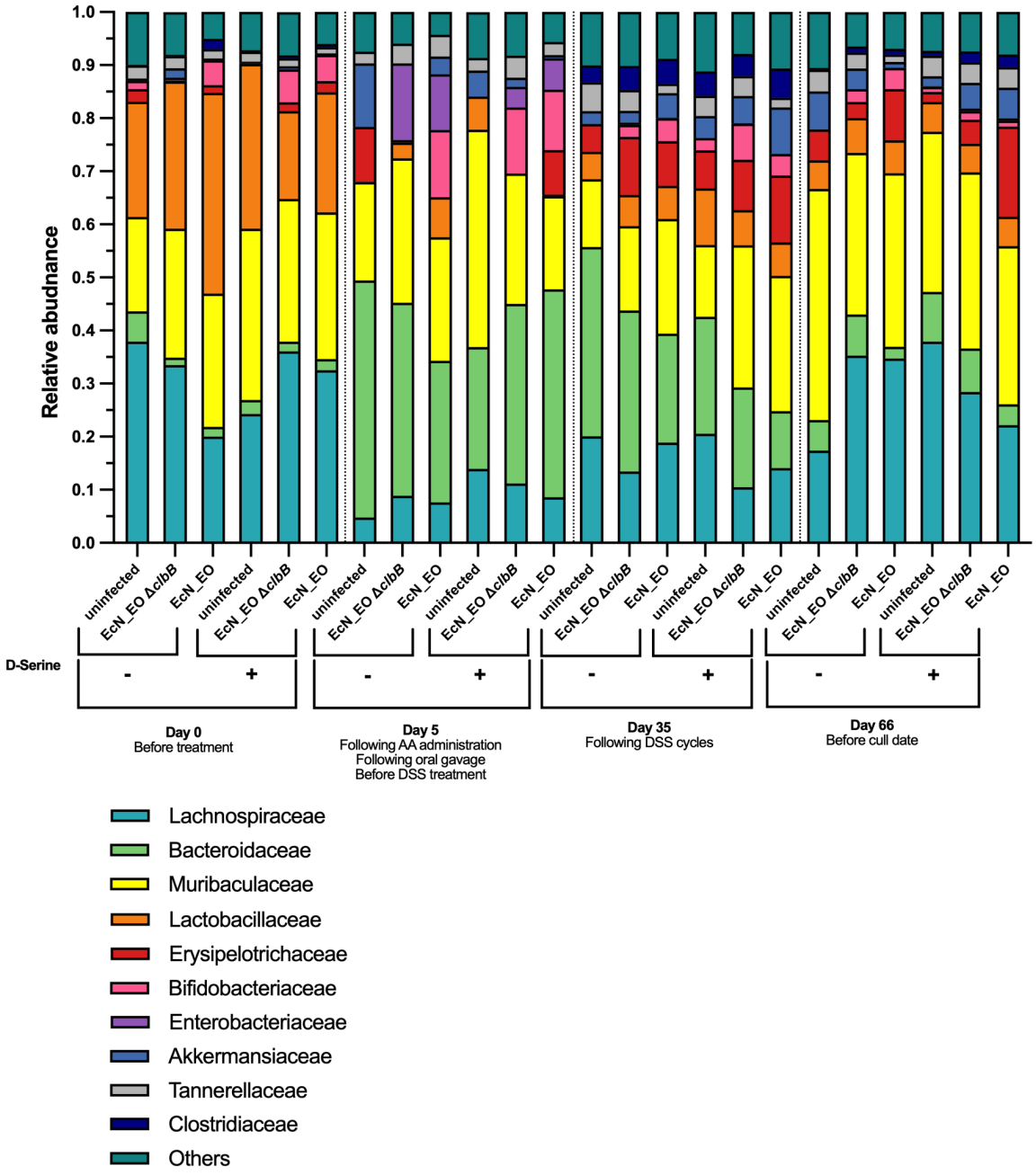


Figure 6-21 Relative abundance of the major bacterial families identified in the intestinal microbiome. A bar graph showing average relative abundances of the major bacterial families identified in the faecal microbiome of C57BL/6 mice at different time points. Different colours depict average relative levels of abundance of different microbial families in mice administered streptomycin and AOM/DSS and either left uninfected, infected with EcN_EO Δ*clbB*, or infected with EcN_EO. For each treatment category, half of the mice were also treated with 1 mM D-Serine. Day 0 samples represent the gut microbiome before the mice were placed on procedure. Day 5 samples represent the intestinal bacterial composition after being administered streptomycin, receiving an IP injection of AOM and after five days of drinking regular water or water with 1 mM D-Serine. Day 35 samples represent the microbiome following two cycles of 2% DSS treatment among all groups and a continuation of AA treatment in the specified groups. Day 66 samples represent the microbiome at the end of the experiment, shortly before the cull date.

6.3 Discussion

The AOM/DSS model is considered particularly effective at causing CRC development in healthy mice because it elicits a sequence of intestinal tissue alterations – from crypt foci, to adenomas, to carcinomas – that is highly comparable to the one observed in humans, where aberrant crypts firstly develop into adenomas and then into carcinomas, before metastasising. Another advantage of the AOM/DSS model, compared to the classical model – which consists of several, repeated, AOM injections without administration of DSS– is that it typically has a latency time of 10 rather than 30 weeks (De Robertis *et al.*, 2011). Furthermore, it drives a highly localised development of cancer in the intestine, making it a relevant model for studying the impact of the gut microbiota on CRC. However, establishing an AOM/DSS model to allow evaluation of colibactin expression by E. coli on CRC development, required significant effort and optimisation. Initially, it was postulated that the lack of polyps in the first attempt at the AOM/DSS model may reflect the choice of immunocompetent rather than immunocompromised mice. However, immunocompetent C57BL/6 mice have been extensively used in published literature describing CRC models. A more likely explanation for the lack of polyps was the decision to initially test a lower concentration of DSS to minimise animal suffering. Most published literature describing these models typically employs DSS concentrations of 2% or 3%. In addition subsequent analysis of the existing literature highlighted that colitogenic potential of DSS varies between molecular weights, but more surprisingly, between vendors and reportedly even between manufacturing lots. In particular, one published protocol recommend the use of DSS salt, colitis grade 36,000 – 50,000 Da (MP Biomedicals) over any other DSS (Chassaing *et al.*, 2015). The DSS salt employed in this chapter's experiments instead has a molecular weight of 40,000 Da and is produced by Thermo Fisher Scientific. The same publication also reported that colitis development was limited in germ-free conditions, which led to the decision to no longer treat mice with sterile food and drinking water. Symptoms of DSS-induced colitis include thinning and shortening of the colon and enlargement of the spleen. Upon dissection, colons were measured and spleens were weighed, but no significant differences were observed between treatment groups and the data is not represented. Colitis induced by DSS is known to wane

after a few weeks, hence the need for multiple DSS cycles in models of chronic intestinal inflammation

There is extensive evidence in the literature linking pks+ E. coli to CRC development. Most of these studies were conducted using clinical isolates from CRC patients, rather than a prototypical EcN strain. Indeed, the fact that EcN's probiotic nature cannot be uncoupled from its antagonistic one (Massip *et al.*, 2019) adds a layer of complexity to the study of EcN *in vivo*. EcN's dual nature – whereby on one hand it functions as a probiotic, with effective antagonistic activity against other Enterobacteriaceae, and on the other it synthesises colibactin, a putative pro-carcinogenic metabolite – is reflected in the *in vivo* investigations that have been conducted to date. Whereas some experiments have shown that EcN damages DNA *in vivo* (Nougayrède *et al.*, 2021), successfully colonises tumours (Stritzker *et al.*, 2007), proliferates in inflamed intestines (Cevallos *et al.*, 2019), and, if the microbiota and adaptive immunity are defective, can translocate across the intestinal epithelium (Gronbach *et al.*, 2010), countless other experiments have shown that EcN ameliorates intestinal inflammation and colitis (Schultz *et al.*, 2004), (Sturm *et al.*, 2005), (Rodríguez-Nogales *et al.*, 2018). Publications investigating the interplay between colibactin-producing E. coli and CRC have employed strains such as CCR20 (Cougnoux *et al.*, 2014); NC101, shown to promote invasive carcinomas (Arthur *et al.*, 2012); and 11G5, which appears to promote procarcinogen environments by impairing the response of antitumour T-cells (Lopès *et al.*, 2020) and to induce tumorigenicity, with an increase in the number and size of adenocarcinomas in infected mice (Bonnet *et al.*, 2014), (Raisch *et al.*, 2014). The fact that EcN did not enhance the number and size of polyps in the AOM/DSS model, may be due to its reported probiotic activity. However, one study showed that the expansion of EcN was observed in mice treated with 3% DSS, which increased epithelial oxygenation in the colon (Cevallos *et al.*, 2019). This study furthermore showed that this dysbiotic expansion was required for colibactin-producing E. coli to manifest their cancer-inducing activity. This information might suggest that EcN has the potential to enhance CRC, but failed to do so due to a lack of sufficient epithelial oxygenation, required to alter the host microbiota and to drive EcN's expansion in a way that triggers its carcinogenicity. Furthermore, most of these studies conducted their experiments with immunodeficient mice like $Apc^{Min/+}$, that are predisposed to

adenoma formation (Salesse *et al.*, 2021), or IL10-/- mice, that develop spontaneous colitis (Arthur *et al.*, 2012). One study reported that an increase in tumour size in response to pks+ E. coli infection in the AOM/DSS colon cancer model, was not observed when using immunocompetent mice (Dalmasso *et al.*, 2015).

When considering inflammation scores, the only significant difference in inflammation was detected in EcN-infected mice, who on average displayed a higher inflammation score when fed D-Serine (Figure 6-18B). Upon closer inspection of the parameters used to grade inflammation, it was observed that EcN_EO-infected mice under D-Serine supplementation displayed high grades and distributions of mucosal fibrosis and atrophy. These are indicators of chronic inflammation and fall in line with the reasoning that EcN's probiotic, anti-inflammatory effect was reduced by D-Serine in the early stages of infection and maintained by the chronic supplementation of the AA. Inflammation and neoplasia do not seem to be correlated, with EcN_EO-infected mice under D-Serine supplementation not displaying a larger neoplasia grade compared to their counterparts fed a regular diet. This however is unsurprising because pks+ E. coli are known to enhance tumour growth through the emergence of senescent cells, and not by increasing inflammation (Cougnoux *et al.*, 2014), though some findings also suggest that pks+ E. coli enhance inflammation and the production of reactive oxygen species in tumours (Veziant *et al.*, 2016). This suggests that, in future experiments, inflammation might not be the best parameter for assessing a potential exacerbation of disease caused by EcN, and that senescence and DNA damage might be better markers of disease severity. However, it's also worth noting that the levels of inflammation and biomarkers detected in tumours may vary from those found in the rest of the colonic tissue, and that timing also poses an issue. The fact that mice were culled over 30 days after their last DSS cycle, might suggest why some proinflammatory cytokines were not detectable and why there were no differences in the expression of S100a8, whose protein product dimerises to form calprotectin, a well-established biomarker in intestinal inflammation. (Mizoguchi *et al.*, 2021).

In addition to genetic differences, environmental factors have also been reported to influence tumour development in AOM/DSS models. For example, one study

showed that mice on a high fat/ low protein diet, developed a 50% lower tumour penetrance (Bissahoyo *et al.*, 2005). In regard to the specific role of individual AAs in tumour development, several studies have reported that L-Serine supplementation positively influences intestinal health. One study showed that in specific pathogen free mice, a diet deficient in L-Serine exacerbated DSS-driven colitis and fostered blooms of mucus-degrading bacteria (Sugihara *et al.,* 2020); while another study demonstrated that L-Serine alleviates inflammation following DSS-induced colitis (Zhang *et al.*, 2018). On the other hand, research also reported that blooms of Enterobacteriaceae are reduced when L-Serine is removed from the diet (Kitamoto *et al.*, 2020). In contrast, there is little data in the literature focussing on the effects of D-Serine on gut health. Studies focusing on kidney injuries have shown that D-Serine derived from the gut has a renoprotective effect (Nakade *et al.*, 2018), while work looking at targeting the modulatory site of the NMDA receptor for psychiatric treatments, have focused on natural compounds such as D-Serine which improves symptoms significantly (Tsai *et al., 1998)*. Research focusing on the gut has found that mice treated with 1.5% D-Serine exhibit a reduction in T-cell infiltration and a decrease in colonic inflammation, compared to mice whose diet was not supplemented or supplemented with L-Serine (Asakawa *et al.*, 2021). Research has also found that administration of a 1% D-AA mixture composed of D-Alanine, D-Tryptophan, D-Glutamic acid, D-Serine, and D-Asparagine ameliorates colitis by inhibiting the growth of proteobacteria (Umeda *et al.*, 2023).

Although results obtained in this chapter cannot be considered conclusive, there is evidence that parallel mechanisms might occur when AOM/DSS mice are infected with EcN and fed a diet supplemented with D-Serine. EcN appeared to be eliciting probiotic activity, driving down colonic inflammation, especially in the initial stages of the experiment, when EcN colonisation was at its highest. At the same time D-Serine – which we know reduces colibactin expression *in vitro* – seemed to act on EcN, reducing its probiotic activity and driving more inflammation in EcN-infected mice that were fed D-Serine, compared to those on a regular dietary regimen. It is tempting to speculate that EcN's interlinked probiotic and pro-carcinogenic natures, which cannot be uncoupled, mean that D-Serine supplementation dampens probiotic effects present in EcN, but not in other pks+ strains. A similar study, conducted with another colibactin-producing strain

instead of EcN, could help decipher whether D-Serine effectively reduces colibactin expression *in vivo*. D-Serine also appeared to ameliorate symptoms in AOM/DSS mice, independently of the infection regimen to which they were subjected. Taking this evidence together, although an exacerbation of colorectal polyps was not observed in the presence of EcN – likely due to its probiotic nature that distinguishes it from other pks+ strains (as explained above) – an effect of D-Serine on EcN was nonetheless observed, reinforcing its potential as a therapeutic candidate. The fact that D-Serine has also been tested for nephrotoxicity and appears safe at the currently studied maximal doses (Meftah, Hasegawa and Kantrowitz, 2021), can further strengthen research efforts in studying its therapeutic outcomes.

7 Investigation of the molecular mechanisms underlying colibactin's repression by D-Serine

7.1 Introduction

Investigations within the Roe Group and published literature have shown that nitrogen metabolism is involved in the expression of the T3SS in EHEC. In particular, one study showed that L-Glutamine is able to repress the T3SS in EHEC by downregulating expression of the locus of enterocyte effacement (LEE) pathogenicity island (Fang *et al.*, 2023), which contains five operons and encodes for the proteins required for an effective secretion system. The study's transcriptomics data revealed that nitrogen metabolism-relevant genes were downregulated as a consequence of L-Glutamine-induced repression of LEE. Further investigations suggested that NtrC phosphorylation – which results in a downregulation of *ler* (the LEE transcription regulator) through σ^S and PchA (known to have a positive regulatory effect on LEE) – is responsible for L-Glutamine's repression of the T3SS (Fang *et al.*, 2023), (Iyoda and Watanabe, 2004), (Riordan and Mitra, 2017). In addition to carbon, nitrogen is an essential nutrient for bacteria, acting as a precursor for most macromolecules in the bacterial cell, particularly for the synthesis of cell wall components (Brown *et al.*, 2014). Ammonia is typically the preferred nitrogen source for *E. coli* when cultured *in vitro*. They respond to nitrogen starvation by activating the nitrogen regulation (Ntr) stress response, which results in the expression of approximately 100 genes , and whose regulator, NtrC, controls approximately 2% of the whole *E. coli* genome (Zimmer *et al.*, 2000).

The Roe group has shown that D-Serine is able to repress the T3SS in EHEC (Connolly *et al.*, 2015), and given that D-Serine is also able to downregulate expression of colibactin (Hallam *et al.*, 2023), it was questioned whether the mechanism underlying D-Serine mediated downregulation of *clb* genes could also depend on nitrogen metabolism. This was hypothesised in light of the fact that D-Serine is broken down into pyruvate and ammonia and can be used as a sole carbon and nitrogen source in *E. coli* carrying a *dsdCXA* locus; as well as the fact that transcription of *clbB* was higher in M9 (which contains ammonium chloride) than in MEM-HEPES (Chapter 5.2.1). The aim of this chapter was to investigate whether NtrC plays a role in D-Serine's repression of *clb* genes.

7.2 Results

7.2.1 Deletion of the gene encoding for the nitrogen regulatory protein *glnG* in EcN and observation of the downstream effects on colibactin transcription

The investigation begun by knocking out *glnG*, which is the homolog of *ntrC* in EcN. The knock out was generated through lambda red mutagenesis and confirmed by both PCR and sequencing (Figure 7-1A). Mutant colonies appeared smaller, but growth assays revealed no growth defects in EcN Δ*glnG*. The mutant strain was transformed with *pclbB:gfp* and pACYC plasmids in order to carry out reporter assays alongside the wildtype strain, which had previously already been transformed with the reporter plasmids. Data obtained using these reporter strains revealed that transcription of *clbB* was much higher in EcN Δ*glnG* compared to the unmodified strain at all time points. The biggest difference was observed after 8 h, with the transcription of *clbB* in EcN Δ*glnG* being 2.42-times higher ($p = 0.000123$) than in the wildtype (Figure 7-1B). The growth profiles of the wildtype and mutant strains were similar, with both reaching similar phases of growth at the same time, although the wildtype strain achieved higher ODs (Figure 7-1C). Overall these results suggest that in EcN deletion of the nitrogen regulatory protein *glnG* may upregulate colibactin transcription.

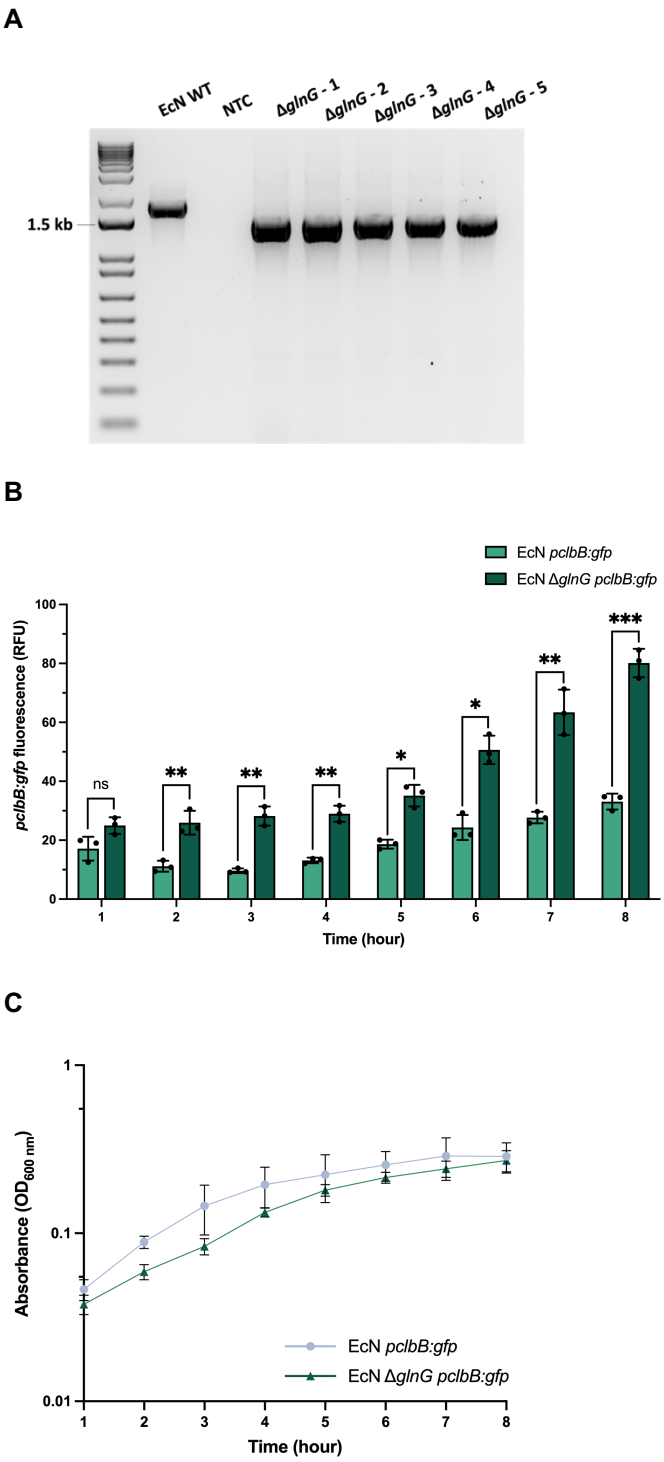


Figure 7-1 Assessment of the effect of *glnG* deletion on colibactin transcription in EcN. (**A**) Confirmation of *glnG* deletion in EcN. The EcN Δ*glnG* strain was generated by lambda red recombineering and checked for successful gene deletion by PCR, performed with a set of primers amplifying the 1514 bp region expected in successful mutants. The gel shows DNA amplified from wild-type EcN and 5 mutants, which were further confirmed by Sanger sequencing. (**B**) The *pclbB:gfp* reporter activity of EcN and EcN Δ*glnG* in MEM-HEPES. Bacterial strains were cultured in 50 mL falcon tubes, at 37°C, 200 rpm. Fluorescence was measured hourly over an 8-hour period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$. Statistical significance was assessed from three replicates with multiple Student's t-tests with *, **, *** indicating significance, $p < 0.05$, < 0.01, and < 0.001 respectively. (**C**) Growth curves of EcN and EcN Δ*glnG* in MEM-HEPES. The strains were cultured as described above. Absorbance ($OD_{600\ nm}$) was measured hourly for 8 hours and values were plotted on a log-10 scale. Data points represent average absorbance values and error bars indicate standard deviation calculated from three replicates.

7.2.2 The effects of exogenous D-Serine on colibactin transcription in EcN Δ*glnG*

It was questioned whether the enhanced expression of colibactin in the *glnG* mutant could be modified by the addition of D-Serine. Therefore, further reporter assays were performed to test *clbB* transcription in EcN Δ*glnG* in the presence of the AA. Results showed that at 8 h following commencement of the growth culture, the mutant's transcription of *clbB* was 0.58-fold lower ($p = 0.0023$) in the presence of 1 m*M* D-Serine. However, even under D-Serine supplementation, the upregulation of colibactin transcription driven by *glnG* deletion was still observable, with EcN Δ*glnG*'s transcription of *clbB* being 2.44-fold higher ($p = 0.0038$) than EcN's (Figure 7-2A). In addition, enhanced growth of EcN Δ*glnG* was observed when the growth media was supplemented with 1 m*M* D-Serine (Figure 7-2B). Overall, these results show that D-Serine is still able to downregulate colibactin expression in a Δ*glnG* background, despite deletion of the nitrogen regulatory protein upregulating colibactin transcription.

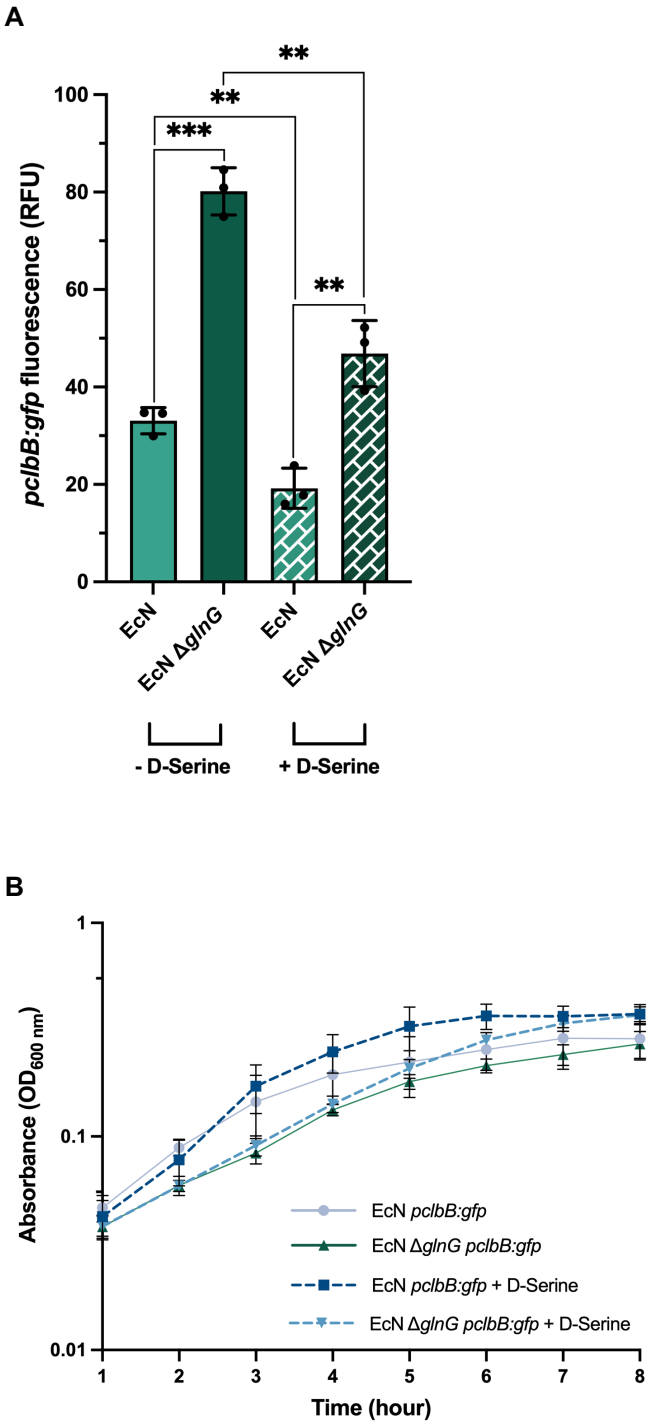


Figure 7-2 Assessment of D-Serine's effect on colibactin transcription in EcN Δ*glnG*. (**A**) The *pclbB:gfp* reporter activity of EcN and EcN Δ*glnG* in MEM-HEPES under 1 mM D-Serine supplementation. Bacterial strains were cultured in 50 mL falcon tubes, at 37°C, 200 rpm. Fluorescence was measured hourly over an 8-hour period, *clbB* promoter activity is expressed as relative fluorescence units (RFU) and is calculated as *pclbB:gfp*/$OD_{600\ nm}$ at 8 hours. Statistical significance was assessed from three replicates with multiple Student's t-tests with *, **, *** indicating significance, $p < 0.05$, < 0.01, and < 0.001 respectively. (**B**) Growth curves of EcN and EcN Δ*glnG* in MEM-HEPES alone and MEM-HEPES supplemented with 1 mM D-Serine. The strains were cultured as described above. Absorbance ($OD_{600\ nm}$) was measured hourly for 8 hours and values were plotted on a log-10 scale. Data points represent average absorbance values and error bars indicate standard deviation calculated from three replicates.

7.2.3 Validating EcN Δ*glnG*'s enhanced transcription of colibactin by RT-qPCR and by ISCLA

The observed mechanisms of colibactin upregulation by *glnG* deletion and of colibactin downregulation by D-Serine supplementation were further validated by RT-qPCR and by ISCLAs.

Cultures of EcN and EcN Δ*glnG* were grown in MEM-HEPES alone or MEM-HEPES supplemented with 1 mM of D-Serine and harvested for RNA extraction. The expression of *clbB* was assessed by RT-qPCR relative to the housekeeping gene *gapA* and measurements were normalised to a control condition of no AA supplementation in the wildtype strain (Figure 7-3A). Compared to EcN, in EcN Δ*glnG* the expression of *clbB* was 2.66-fold higher ($p = 0.0015$), but when the mutant strain was grown under D-Serine supplementation conditions the expression of *clbB* was 0.47-fold lower ($p = 0.011$).

As previously described, linearised pUC19 was incubated with EcN and EcN Δ*glnG* for 5 h in M9 alone, or M9 supplemented with 1 mM D-Serine and visualised on an alkaline agarose gel (Figure 7-3B). Quantification of DNA crosslinking was done by densitometry, using ImageJ (Rasband, 1997-2018). The percentage of cross-linked DNA was 17.07% higher ($p = 0.0722$) when pUC19 was exposed to EcN Δ*glnG* instead of EcN. The percentage of cross-linked DNA was 55.86% lower ($p = 0.0171$) when pUC19 was exposed to the mutant strain in the presence of D-Serine, compared to no AA conditions (Figure 7-3C).

Overall, these results align with previous findings, showing that in EcN deletion of the nitrogen regulatory protein *glnG* upregulates colibactin transcription and that D-Serine is able to downregulate the expression of colibactin in both the wildtype and mutant strains.

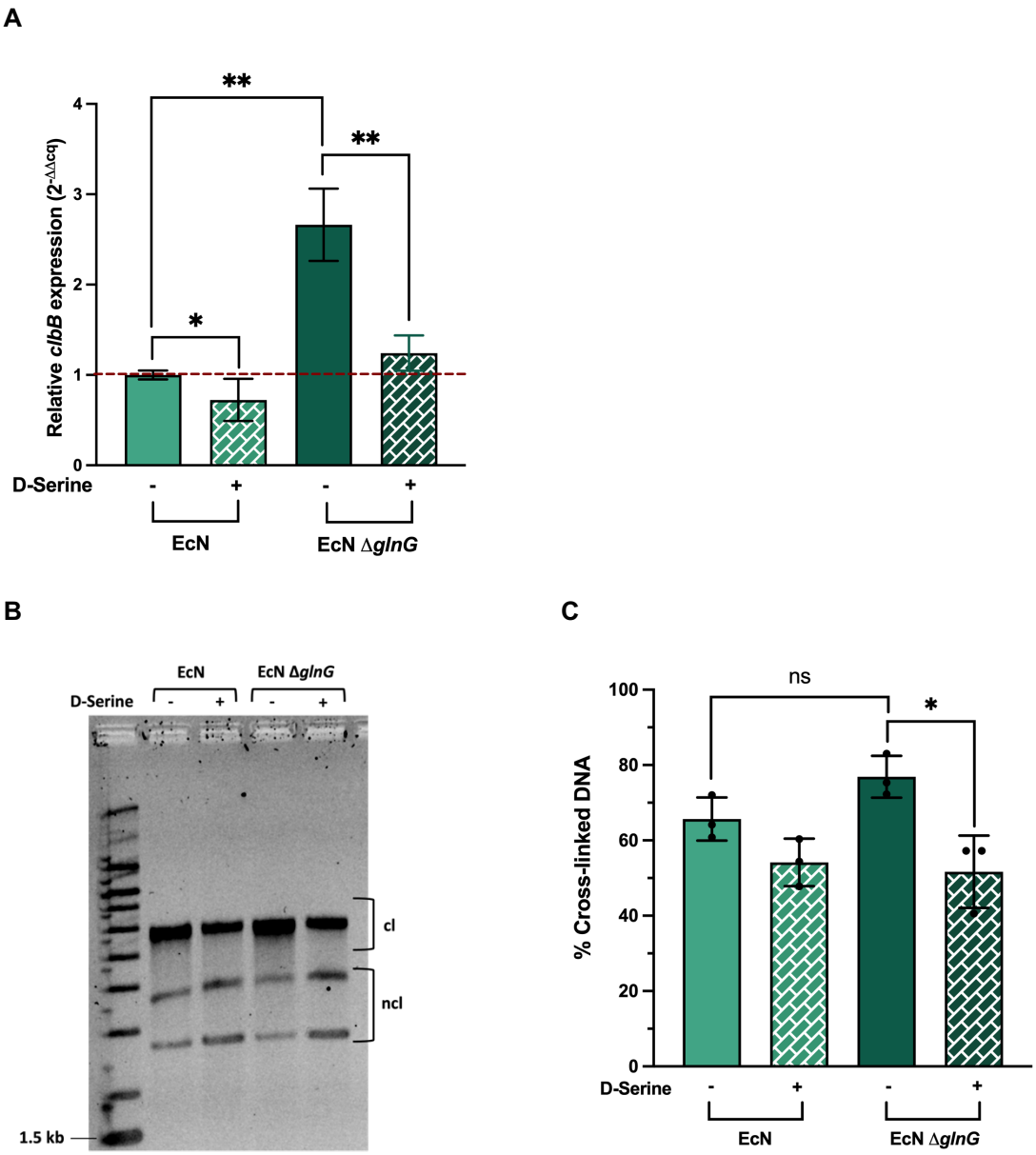


Figure 7-3 Validation of the effect of *glnG* deletion on colibactin expression in the presence and absence of D-Serine. (**A**) The expression of *clbB* in wildtype and mutant strains screened by RT-qPCR. EcN and EcN Δ*glnG* were grown in 50 mL falcon tubes at 37°C, 200 rpm for 7 h in MEM-HEPES alone or supplemented with 1 mM D-Serine. RNA was extracted from bacterial cultures and converted to cDNA. The expression of *clbB* was measured relative to the housekeeping gene *gapA*, using a control condition of no amino acid supplementation in the wildtype strain as the calibrator for baseline expression, as indicated by the dashed red line. Statistical significance was determined from four replicates through multiple Student's t-tests with * and ** indicating significance, $p < 0.05$ and $p < 0.01$ respectively. (**B**) Visualising cross-linked DNA after migration under alkaline gel electrophoresis. EcN and EcN Δ*glnG* were cultured for 5 h in M9 alone (-) or M9 supplemented with 1 mM D-Serine (+) at 37°C, 200 rpm. Linearised pUC19 was then exposed to 1.5 x 10^6 CFUs of EcN or EcN Δ*glnG* for 40 mins, in M9 alone or with 1 mM D-Serine. Plasmid DNA was then purified and crosslinking (cl) and non-crosslinking (ncl) activities were visualised by gel electrophoresis in alkaline conditions. The gel was run at 25 V for 45 mins, and then at 50 V for 150 mins; it was soaked in neutralisation buffer for 20 mins prior to image acquisition; the DNA size marker is the 1 kB DNA ladder (Invitrogen). (**C**) Percentage of cross-linked DNA relative to total DNA signal under different infection and AA regimes. Signal intensities for both crosslinked (cl) and non-crosslinked (ncl) DNA were quantified by ImageJ, and the percentage of crosslinked DNA was calculated relative to the total DNA in each lane. Columns represent means +/- standard deviation calculated from three replicates. Statistical significance was assessed by a Student's t-test with * indicating $p < 0.05$.

7.3 Discussion

Taken together, data in this chapter shows that deletion of *glnG* results in increased levels of *clbB* transcription, with knockouts still showing susceptibility to the D-Serine-driven repression of colibactin production. The fact that colibactin production is upregulated in EcN Δ*glnG* suggests an involvement of *glnG* in the complex regulation of the colibactin locus; while the fact that D-Serine continues to downregulate the genotoxin in mutant strains, seems to suggest that the mechanism of action by which D-Serine represses colibactin is unlikely to depend on *glnG*. Studies have shown that colibactin – beyond being a carcinogen in mammalian systems – is mainly a bacteriocin that has evolved to confer a competitive advantage over other microorganisms: colibactin in fact induces lytic development in bacteria that contain prophages, with this mechanism observed across multiple phage-bacteria systems, and with immunity displayed only in bacteria possessing ClbS homologues (Silpe *et al.*, 2022). The stark upregulation of colibactin in the absence of *glnG* might therefore be a response to difficulties in nitrogen assimilation that – albeit not having a significant impact on growth – risks limiting the bacteria's competitiveness and may induce a stress-response in the bacterial cell, prompting an upregulation of colibactin biosynthesis genes. Supplementation with D-Serine – which EcN can use a sole carbon and nitrogen source and towards which, as shown in Chapter 4, it has a nutritional preference – may restore a stable nutrient source and decrease EcN's need to outcompete other bacteria through colibactin production. RNA-seq results illustrated in Chapter 4 had shown that EcN, when cultured in M9 with 1 mM D-Tyrosine, exhibited significant downregulation of several genes beyond those responsible for colibactin biosynthesis. These included genes involved in nitrogen metabolism: *glnG* (nitrogen regulation protein), *arcC* (which catalyses the reaction from carbamoyl phosphate and ADP to ATP and ammonium carbamate), and *narZ* (a nitrate reductase that reduces nitrate to nitrite). Although, as described above, D-Serine's repression of colibactin does not seem to involve *glnG*, the fact that with D-Tyrosine supplementation both *clb* genes and genes involved in nitrogen metabolism were downregulated, hints at an interplay existing between colibactin production and nitrogen metabolism, making this an interesting and promising avenue to explore in order to further elucidate colibactin regulation.

8 Discussion

Globally, CRC is the third most common malignant disease and the second deadliest form of cancer. While the incidence of CRC has been decreasing in high income countries – likely linked to effective screening programmes – the global burden of CRC is expected to increase by 20-25%, reaching 2.2 million new cases by 2030 (Arnold *et al.*, 2017). It is also concerning that CRC in individuals below 50 years of age appears to be on the rise (Bailey *et al.*, 2015). The development of CRC development can arise from genetic predispositions, intestinal disorders, environmental factors like lifestyle and diet, and from oncogenic microorganisms, which are increasingly being investigated for their role in cancer aetiology. Over the past two decades, *E. coli* harbouring the *pks* island, which encodes for the synthesis of colibactin, have been strongly linked to CRC. Colibactin is a powerful genotoxic metabolite and cyclomodulin, reported to induce ICLs and DSBs, leading to cell-cycle arrest in G2/M phase and eventually to cell death (Nougayrède *et al.*, 2006). Colibactin-producing *E. coli* (CoPEC) are largely overrepresented in CRC-patients from whom *pks*+ *E. coli* have been isolated three times more frequently than from healthy individuals (Dejea *et al.*, 2018). However, in addition to being detected in pathogenic and polyp-associated bacteria, the *pks* island is also carried by commensal strains, such as Nissle 1917 (EcN), a widely administered probiotic (Schultz and Burton, 2017).

Several exogenous factors have been found to contribute to colibactin expression (Addington, Sandalli and Roe, 2024). For example, the colibactin biosynthetic enzymes ClbR and ClbA appear upregulated in iron-limited conditions, while iron-rich environments seem to repress colibactin transcription (Tronnet *et al.*, 2016). Oxygen also appears to be a regulator of colibactin production, with research demonstrating that *pks*+ *E. coli* genotoxicity is optimal in anoxic environments and repressed in oxygen-rich conditions (Bossuet *et al.*, 2023). Spermidine (Chagneau *et al.*, 2019), polymyxin B (Patric W. Sadecki *et al.*, 2021), and oligosaccharides such as inulin and galacto-oligosaccharide (Oliero *et al.*, 2021) have been found to enhance transcription of *clb* genes. Conversely, mesalamine (Tang-Fichaux *et al.*, 2021), L-Tryptophan (Bayne *et al.*, 2024), SCFAs (Zhang *et al.*, 2020), and plant-derived compounds like tannin, quercetin (Kaewkod *et al.*, 2021), and cinnamon (Kosari *et al.*, 2020), have been found to downregulate colibactin synthesis.

As research has demonstrated a strong interplay between CRC, diet, and the gut microbiome, this PhD project aimed to investigate the impact of dietary components on colibactin and subsequent downstream effects on CRC. More specifically, given the Roe Group's longstanding interest in the effect of dietary AAs on bacterial metabolism, the impact of AAs on colibactin expression was investigated. The work presented in this thesis showed that several L- and D-isomers of proteinogenic AAs can modulate transcription of colibactin, corroborating and advancing previous research in the Roe Group, demonstrating that D-Serine represses colibactin transcription, and, further, showed that D-Tyrosine also appears to repress colibactin expression. Moreover, for the first time, the data presented in this thesis shows evidence that D-Serine affects colibactin production *in vivo*, and that *glnG* deletion results in increased levels of colibactin transcription.

By definition, commensal microbes do not harm the host, whereas pathogens are described as microorganisms that cause disease. However, we now recognise a middle ground exists, whereby some pathogens are present without causing disease and conversely some commensals lead to disease under specific conditions. These organisms are increasingly being referred to as 'pathobionts'. Whether the relationship between the intestinal microbiome and CRC is causative or associative is still poorly defined. For some microbes, like *H. pylori*, evidence suggests a direct, causative relationship between the bacterium and the onset of cancer, but for others, connections are not so clear. Moreover, the causation of cancer by the host microbiome is extremely difficult to define due to the dynamic nature of the microbiome itself, which changes with diet, age, geographic location, and other environmental exposures (Scott *et al.*, 2018). This is further complicated by the fact that most studies looking at the association between the microbiome and CRC involve single-time-point sampling, and are not based on prospective, longitudinal data from large cohorts.

Research has shown that the microbiome of CRC patients differs from that of healthy individuals, with the term 'dysbiotic' often used to indicate the microbiome of patients suffering from intestinal disorders. However, not only is it unclear whether an altered microbial community is the cause of cancer or its consequence, but the definition of a 'healthy gut' – and consequently that of a

'dysbiotic gut' – is host-specific and therefore subjective. The 'alpha-bug' hypothesis suggests that some bacteria are directly pro-oncogenic and modulate host immune responses to promote CRC. Furthermore, the authors of this hypothesis suggest that alpha-bugs can remodel the intestinal microbiota to drive expansion of 'collaborators' that promote oncogenesis (Sears and Pardoll, 2011). In contrast, the 'driver-passenger' hypothesis suggests that CRC can indeed be initiated by certain bacterial 'driver' species that promote oncogenesis, but that the tumour microenvironment then favours opportunistic 'passenger' species, that over time replace the drivers (Tjalsma *et al.*, 2012). It is unclear whether colibactin-producing bacteria should be considered 'drivers' or 'passengers' (Faïs *et al.*, 2018). On one hand, *pks+ E. coli* directly damage DNA (Nougayrède *et al.*, 2006) and induce a specific mutational signature in intestinal crypts (Dziubańska-Kusibab *et al.*, 2020), suggesting a driver role. Conversely, animal models have shown that *pks+* bacteria do not spontaneously induce CRC, but rather increase the severity of disease in mice predisposed to CRC (Cougnoux *et al.*, 2014), suggesting a passenger role. Moreover, the fact that the *pks* signature has also been identified in a small subset of colonic crypts from healthy individuals strongly suggests that colibactin alone is not sufficient to trigger the onset of cancer, but is a component amongst multifactorial and combinatorial risks for CRC (Chen *et al.*, 2023).

It's also interesting to note that whereas numerous investigations have found that *pks+ E. coli* are isolated more frequently from CRC patients, there was a study carried out in Japan that showed no difference in *pks+ E. coli* colonisation between CRC patients and healthy individuals (Shimpoh *et al.*, 2017). The authors of the study hypothesise that this could be due to the Japanese population, including healthy individuals, displaying on average a high prevalence of B2 *E. coli* strains. Correspondingly, a study conducted in the US found that colibactin-producing *E.coli* are more prevalent in CRCs associated to a Western-diet (Arima *et al.*, 2022). These findings suggests that a conjunction of multiple factors may be required for *pks+* strains to exert their oncogenic activity. They also emphasise the need to explore how the microbiome impacts physiology in the context of external factors, like diet and geography. Work conducted in this thesis also showed that CoPEC strains can produce different levels of colibactin, both at a transcriptional level (Chapter 4.2.6, looking at clinical isolates), and at a

functional level (Chapter 5, comparing EcN_AR and EcN_EO). This could explain why some B2 *E. coli* strains enhance tumour development more than others and why a concurrence of multiple host and bacterial factors is needed for colibactin-derived oncogenesis.

The *clb* gene cluster, aside from synthesising colibactin, is also involved in the production of siderophores (Martin *et al.*, 2013), microcins (Massip *et al.*, 2020), and lipopetides with analgesic activity (Pérez-Berezo *et al.*, 2017). Hence, it should be considered that modulating colibactin expression may also influence the production of other metabolites. This is especially relevant in animal models where isogenic mutants of a *clb* gene are often used as controls, potentially introducing even more variables into a complex system. Indeed, one of the limitations of the work described in this thesis resulted from using EcN for *in vivo* infections. As described above, unlike other CoPEC strains, EcN is a well-established probiotic, used for the treatment of an array of intestinal disorders and known to produce microcins and analgesic peptides in addition to colibactin. The fact that EcN's probiotic activity is connected to its production of colibactin (Olier *et al.*, 2012) is likely the reason why in this thesis (Chapter 6), and in numerous publications (Schultz *et al.*, 2004), (Rodríguez-Nogales *et al.*, 2018), (Sturm *et al.*, 2005), EcN-infected mice tend to display better clinical outcomes than uninfected mice in colitis models. Work in this PhD project showed that EcN-gavaged mice who were fed D-Serine did not seem to display the probiotic effects of EcN, possibly since, by acting on colibactin, D-Serine counteracted EcN's probiotic activity. Carrying out the same *in vivo* experiments with a *pks*+ non-probiotic *E. coli* strain that is known to enhance CRC polyps (such as CCR20), could provide more clarity on whether D-Serine is truly able to downregulate colibactin production *in vivo*. From a clinical standpoint, while D-Serine supplementation could offer therapeutic benefits against CoPEC strains that risk exacerbating CRC tumours, it may instead pose a risk for patients receiving EcN probiotic treatment for intestinal disorders. In such cases, D-Serine could neutralise EcN's anti-inflammatory activity, potentially resulting in poorer clinical outcomes. The boundary between the pathogenicity and probiotic activity of EcN has not yet been deciphered and its optimal use as a probiotic remains uncertain. Further studies are of utmost importance, especially given recent efforts in bioengineering EcN for the delivery of tumour-targeting drugs (Gurbatri *et al.*, 2024). Indeed, due to

its natural ability to colonise polyps and to its reputation as a safe probiotic, the application of EcN in bacterial cancer therapy is being explored, with research showing that EcN-driven microrobots delivering chemotherapeutic drugs successfully induce tumour regression in mice, without tumour recurrence (Wu *et al.*, 2023).

In conclusion, results obtained in this PhD project contribute to what is currently understood about colibactin's endogenous and exogenous regulation and may help further elucidate how D-Serine reduces colibactin production. It is apparent that the synthesis of colibactin is tightly regulated by numerous environmental cues, serving as a reminder to pursue microbiological research beyond an anthropocentric standpoint: colibactin has primarily evolved to outcompete other bacterial strains and oncogenesis is likely just a by-product of this mechanism, making colibactin an excellent example of how bacteria simultaneously accomplish commensalism, symbiosis and pathogenesis.

9 References

Abram, K. *et al.* (2021) 'Mash-based analyses of *Escherichia coli* genomes reveal 14 distinct phylogroups', *Commun Biol,* 4(117).

Addington, E., Sandalli, S. and Roe, A.J. (2024) 'Current understandings of colibactin regulation', *Microbiology*, 170, pp. 1427.

Ahluwalia, V.K., Kumar, L.S. and Kumar, S. (2022) 'Amino Acids', *Chemistry of Natural Products*, pp. 1–65.

Anfora, A.T. *et al.* (2007) 'Roles of serine accumulation and catabolism in the colonization of the murine urinary tract by *Escherichia coli* CFT073', *Infection and Immunity*, 75(11), pp. 5298–5304.

Anfora, A.T. and Welch, R.A. (2006) 'DsdX is the second D-serine transporter in uropathogenic *Escherichia coli* clinical isolate CFT073', *Journal of Bacteriology*, 188(18), pp. 6622–6628.

Arima, K. *et al.* (2022) 'Western-Style Diet, *pks* Island-Carrying *Escherichia coli*, and Colorectal Cancer: Analyses From Two Large Prospective Cohort Studies', *Gastroenterology*, 163(4), pp. 862–874.

Arnold, M. *et al.* (2017) 'Global patterns and trends in colorectal cancer incidence and mortality', *Gut*, 66(4), pp. 683–691.

Arthur, J.C. *et al.* (2012) 'Intestinal inflammation targets cancer-inducing activity of the microbiota', *Science*, 338(6103), pp. 120–123.

Asakawa, T. *et al.* (2021) 'Oral administration of D-serine prevents the onset and progression of colitis in mice', *Journal of Gastroenterology*, 56(8), pp. 732-745.

Atasoglu, C. *et al.* (1998) 'De novo synthesis of amino acids by the ruminal bacteria *Prevotella bryantii* B14, *Selenomonas ruminantium* HD4, and *Streptococcus bovis* ES1', *Applied and Environmental Microbiology*, 64(8), pp. 2836–2843.

Auvray, F. *et al.* (2021) 'Insights into the acquisition of the *pks* island and production of colibactin in the *Escherichia coli* population', *Microbial Genomics*, 7(5).

Bailey, C.E. *et al.* (2015) 'Increasing Disparities in the Age-Related Incidences of Colon and Rectal Cancers in the United States, 1975-2010', *JAMA Surgery*, 150(1), pp. 17–22.

Balskus, E.P. (2015) 'Colibactin: understanding an elusive gut bacterial genotoxin', Natural Product Reports, 32(11), pp.1534–1540.

Barnard, A.M.L. *et al.* (2010) 'Mutations in *rpsL* that confer streptomycin resistance show pleiotropic effects on virulence and the production of a carbapenem antibiotic in *Erwinia carotovora*', Microbiology, 156(4), pp.1030–1039

Barthel, M. *et al.* (2003) 'Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar *Typhimurium* colitis model that allows analysis of both pathogen and host', *Infection and Immunity*, 71(5), pp. 2839–2858.

Baruch, M. *et al.* (2014) 'An Extracellular Bacterial Pathogen Modulates Host Metabolism to Regulate Its Own Sensing and Proliferation', Cell, 156(1-2), pp.97–108.

Bastings, J.J.A.J. *et al.* (2019) 'D-amino acids in health and disease: A focus on cancer', *Nutrients*, 11(9), pp. 1–18.

Bayne, C. *et al.* (2024) 'L-tryptophan and copper interactions linked to reduced colibactin genotoxicity in *pks+ Escherichia coli*', *mSystems* [Preprint]. Edited by J.L. Blanchard.

Beaumont, M. and Blachier, F. (2020) 'Amino acids in intestinal physiology and health', *Advances in Experimental Medicine and Biology*, 1265, pp. 1–20.

van den Belt, M. *et al.* (2023) 'CAGECAT: The CompArative GEne Cluster Analysis Toolbox for rapid search and visualisation of homologous gene clusters', *BMC Bioinformatics*, 24(181).

Bender, R.A. (2012) 'Regulation of the Histidine Utilization (Hut) System in Bacteria', *Microbiology and Molecular Biology Reviews*, 76(3), pp. 565–584.

Berg, M., Hilbi, H. and Dimroth, P. (1997) 'Sequence of a gene cluster from Malonomonas rubra encoding components of the malonate decarboxylase Na' pump and evidence for their function', *European journal of biochemistry* , 245(1), pp. 103–115.

Bian, X. *et al.* (2013) '*In Vivo* Evidence for a Prodrug Activation Mechanism during Colibactin Maturation', *ChemBioChem*, 14(10), pp. 1194–1197.

Bian, X. *et al.* (2015) 'Two more pieces of the colibactin genotoxin puzzle from *Escherichia coli* show incorporation of an unusual 1-aminocyclopropanecarboxylic acid moiety', *Chemical Science*, 6(5), pp.3154–3160.

Bingham, S.A. *et al.* (2003) 'Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): An observational study', *Lancet*, 361(9368), pp. 1496–1501.

Bischof, O. and Dejean, A. (2007) 'SUMO is growing senescent', *Cell Cycle*, 6(6), pp. 677–681.

Bissahoyo, A. *et al.* (2005) 'Azoxymethane is a genetic background-dependent colorectal tumor initiator and promoter in mice: Effects of dose, route, and diet', *Toxicological Sciences*, 88(2), pp. 340–345.

Bloom, F.R. and McFall, E. (1975) 'Isolation and characterization of D serine deaminase constitutive mutants by utilization of D serine as sole carbon or nitrogen source', *Journal of Bacteriology*, 121(3), pp. 1078–1084.

Boleij, A. *et al.* (2015) 'The *Bacteroides fragilis* Toxin Gene Is Prevalent in the Colon Mucosa of Colorectal Cancer Patients', *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 60(2), p. 208.

Bonnet, M. *et al.* (2014) 'Colonization of the human gut by *E. coli* and colorectal cancer risk', *Clinical Cancer Research*, 20(4), pp. 859–867.

Bossuet, N. *et al.* (2023) 'Oxygen concentration modulates colibactin production', *Gut Microbes*, 15(1).

Bossuet-Greif, N. *et al.* (2018) 'The colibactin genotoxin generates DNA interstrand cross-links in infected cells', *mBio*, 9(2).

Brachmann, A.O. *et al.* (2015) 'Colibactin biosynthesis and biological activity depend on the rare aminomalonyl polyketide precursor, *Chem. Commun*, 51(66), pp.13138–13141.

Brooks, W.H., Guida, W.C. and Daniel, K.G. (2011) 'The Significance of Chirality in Drug Design and Development HHS Public Access', *Curr Top Med Chem*, 11(7), pp. 760–770.

Brotherton, C.A. and Balskus, E.P. (2013) 'A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity', *Journal of the American Chemical Society*, 135(9), pp. 3359–3362.

Brown, D.R. *et al.* (2014) 'Nitrogen stress response and stringent response are coupled in *Escherichia coli*', *Nature Communications 2014 5:1*, 5(1), pp. 1–8.

Buc, E. *et al.* (2013) 'High Prevalence of Mucosa-Associated *E. coli* Producing Cyclomodulin and Genotoxin in Colon Cancer', *PLoS ONE*, 8(2), p. 56964.

Butterworth, A.S., Higgins, J.P.T. and Pharoah, P. (2006) 'Relative and absolute risk of colorectal cancer for individuals with a family history: A meta-analysis', *European Journal of Cancer*, 42(2), pp. 216–227.

Caini, S. *et al.* (2022) 'Fish Consumption and Colorectal Cancer Risk: Meta-Analysis of Prospective Epidemiological Studies and Review of Evidence from Animal Studies', *Cancers*, 14(3), pp. 1–20.

Caldara, M. *et al.* (2007) 'ArgR-dependent Repression of Arginine and Histidine Transport Genes in *Escherichia coli* K-12', *Journal of Molecular Biology*, 373(2), pp. 251–267.

Calendar, R. and Berg, P. (1967) ‘d-Tyrosyl RNA: Formation, hydrolysis and utilization for protein synthesis’, *Journal of Molecular Biology*, 26(1), pp. 39–54.

Caprioli, A. *et al.* (1983) ‘Partial Purification and Characterization of an Escherichia coli Toxic Factor That Induces Morphological Cell Alterations’, *Infection and Immunity*, 39(3), pp.1300–1306.

Celiberto, F. *et al.* (2023) ‘Fibres and Colorectal Cancer: Clinical and Molecular Evidence’, *International Journal of Molecular Sciences*, 24(17), p.13501.

Cevallos, S.A. *et al.* (2019) ‘Increased epithelial oxygenation links colitis to an expansion of tumorigenic bacteria’, *mBio*, 10(5).

Chagneau, C. V. *et al.* (2019) ‘The Polyamine Spermidine Modulates the Production of the Bacterial Genotoxin Colibactin’, *mSphere*, 4(5).

Chagneau, C. V *et al.* (2021) ‘Uropathogenic E. coli induces DNA damage in the bladder’, *PLOS Pathogens*, 17(2).

Chagneau, C. V. *et al.* (2022) ‘The pks island: a bacterial Swiss army knife? Colibactin: beyond DNA damage and cancer’, *Trends in Microbiology*, 30(12), pp. 1146–1159.

Chan, D.S.M. *et al.* (2011) ‘Red and Processed Meat and Colorectal Cancer Incidence: Meta-Analysis of Prospective Studies’, *PLoS ONE*, 6(6).

Chaoprasid, P. and Dersch, P. (2021) ‘The Cytotoxic Necrotizing Factors (CNFs)-A Family of Rho GTPase-Activating Bacterial Exotoxins’, *Toxins*, 13(12), p.901

Chassaing, B. *et al.* (2015) ‘DSS Protocol’, *Curr Protoc Immunol*, 27(4), pp. 1–19.

Chen, B. *et al.* (2023) ‘Contribution of pks+ E. coli mutations to colorectal carcinogenesis’, *Nature Communications 2023 14:1*, 14(1), pp. 1–9.

Chen, Q. *et al.* (2017) ‘Probiotic *E. coli* Nissle 1917 biofilms on silicone substrates for bacterial interference against pathogen colonization’, *Acta Biomaterialia*, 50, pp. 353–360.

Chen, S.L. *et al.* (2013) ‘Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection’, *Science translational medicine*, 5(184).

Cheng, W.T., Kantilal, H.K. and Davamani, F. (2020) ‘The mechanism of *Bacteroides fragilis* toxin contributes to colon cancer formation’, *Malaysian Journal of Medical Sciences*, 27(4), pp. 9–21.

Christou, N. *et al.* (2017) 'E-cadherin: A potential biomarker of colorectal cancer prognosis', *Oncology Letters*, 13(6), pp. 4571–4576.

Clermont, O. *et al.* (2013) 'The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups', *Environmental Microbiology Reports*, 5(1), pp. 58–65.

Clermont, O., Bonacorsi, S. and Bingen, E. (2000) 'Rapid and simple determination of the *Escherichia coli* phylogenetic group', *Applied and Environmental Microbiology*, 66(10), pp. 4555–4558.

Connolly, J.P.R. *et al.* (2015) 'The host metabolite D-serine contributes to bacterial niche specificity through gene selection', *ISME Journal*, 9(4), pp. 1039–1051.

Connolly, J.P.R. *et al.* (2016) 'A Highly Conserved Bacterial D-Serine Uptake System Links Host Metabolism and Virulence', *PLoS Pathogens*, 12(1).

Connolly, J.P.R *et al.* (2021) 'd-Serine induces distinct transcriptomes in diverse *Escherichia coli* pathotypes', *Microbiology*, 167(10).

Corteggiani, M. *et al.* (2022) 'Uncoupling the Hsp90 and DnaK chaperone activities revealed the *in vivo* relevance of their collaboration in bacteria', *PNAS*, 119(37).

Cosloy, S.D. (1973) 'D-Serine Transport System in *Escherichia coli* K-121', *Journal of Bacteriology*, 114(2), pp. 679–684.

Cotton, C.A. *et al.* (2020) 'Underground isoleucine biosynthesis pathways in *E. coli*', *eLife*, 9.

Cougnoux, A. *et al.* (2012) 'Analysis of structure-function relationships in the colibactin-maturating enzyme ClbP', *Journal of Molecular Biology*, 424(3–4), pp. 203–214.

Cougnoux, A. *et al.* (2014) 'Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype', *Gut*, 63(12), pp.1932–1942.

Cuevas-Ramos, G. *et al.* (2010) 'Escherichia coli induces DNA damage *in vivo* and triggers genomic instability in mammalian cells', *PNAS*, 107(25), pp. 11537–11542.

Dalmasso, G. *et al.* (2015) 'The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment', *Gut Microbes*, 5(5), pp. 675–680.

Daßler, T. *et al.* (2000) 'Identification of a major facilitator protein from *Escherichia coli* involved in efflux of metabolites of the cysteine pathway', *Molecular Microbiology*, 36(5), pp. 1101–1112.

Dé Ric Taieb, F., Nougayrè De, J.-P. and Oswald, E. (2011) 'Cycle Inhibiting Factors (Cifs): Cyclomodulins That Usurp the Ubiquitin-Dependent Degradation Pathway of Host Cells', *Toxins*, 3, pp. 356–368.

Defoirdt, T. et al. (2019) 'Elucidating the Hot Spot Residues of Quorum Sensing Peptidic Autoinducer PapR by Multiple Amino Acid Replacements', *Frontiers in Microbiology*, 10.

Dejea, C.M. *et al.* (2018) 'Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria', *Science*, 359(6375), pp. 592–597.

Desai, M.S. *et al.* (2016) 'A dietary fibre-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility HHS Public Access Graphical abstract', *Cell*, 167(5), pp. 1339–1353.

Dineen, S.S. *et al.* (2007) 'Repression of *Clostridium difficile* toxin gene expression by CodY', *Molecular Microbiology*, 66(1), pp. 206–219.

Dougherty, M.W. and Jobin, C. (2021) 'Shining a Light on Colibactin Biology', *Toxins*, 13(5), p. 346.

Duan, X. *et al.* (2016) 'L-Serine potentiates fluoroquinolone activity against *Escherichia coli* by enhancing endogenous reactive oxygen species production', *Journal of Antimicrobial Chemotherapy*, 71(8), pp.2192–2199.

Dubois, D. *et al.* (2011) 'ClbP Is a Prototype of a Peptidase Subgroup Involved in Biosynthesis of Nonribosomal Peptides', *Journal of Biological Chemistry*, 286(41), pp.35562–35570

Dziubańska-Kusibab, P.J. *et al.* (2020) 'Colibactin DNA-damage signature indicates mutational impact in colorectal cancer', *Nature Medicine*, 26(7), pp. 1063–1069.

Eagle, H. (1958) 'Amino acid metabolism in human cell cultures.', *Federation proceedings*, 17(4), pp. 985–986.

Escalas, N. *et al.* (2000) 'Study of the cytolethal distending toxin-induced cell cycle arrest in HeLa cells: Involvement of the CDC25 phosphatase', *Experimental Cell Research*, 257(1), pp. 206–212.

Ezzati, M. *et al.* (2002) 'Comparative Risk Assessment Collaborating Group. Selected major risk factors and global and regional burden of disease.', *Lancet (London, England)*, 360(9343), pp. 1347–1360.

Faïs, T. *et al.* (2018) 'Colibactin: More than a new bacterial toxin', *Toxins*, 10(4), p.151.

Falzone, L. *et al.* (2024) 'Benefits and concerns of probiotics: an overview of the potential genotoxicity of the colibactin-producing *Escherichia coli* Nissle 1917 strain', *Gut microbes*, 16(1), p. 2397874.

Fang, F. *et al.* (2023) 'L-glutamine protects against enterohemorrhagic *Escherichia coli* infection by inhibiting bacterial virulence and enhancing host defense concurrently', *Microbiology Spectrum*, 11(6).

Fehlbaum, P. *et al.* (2000) 'An essential amino acid induces epithelial-defensin expression', *PNAS*, 97(23), pp.12723–12728.

Fu, J. *et al.* (2022) 'Dietary Fiber Intake and Gut Microbiota in Human Health', *Microorganisms*, 10(12), pp. 1–18.

Gainza-Cirauqui, M.L. *et al.* (2013) 'Production of carcinogenic acetaldehyde by *Candida albicans* from patients with potentially malignant oral mucosal disorders', *Journal of Oral Pathology and Medicine*, 42(3), pp. 243–249.

Gandhi, M.K., Tellam, J.T. and Khanna, R. (2004) 'Epstein-Barr virus-associated Hodgkin's lymphoma', *British Journal of Haematology*, 125(3), pp. 267–281.

Gandolfi, I. *et al.* (1994) 'D-alanine in Fruit Juices: A Molecular Marker of Bacterial Activity, Heat Treatments and Shelf-life', *Journal of Food Science*, 59(1), pp. 152–154.

Garcie, C. *et al.* (2016) 'The Bacterial Stress-Responsive Hsp90 Chaperone (HtpG) Is Required for the Production of the Genotoxin Colibactin and the Siderophore Yersiniabactin in *Escherichia coli*', *The Journal of Infectious Diseases*, 214, pp. 916–940.

Garénaux, A., Caza, M. and Dozois, C.M. (2011) 'The Ins and Outs of siderophore mediated iron uptake by extra-intestinal pathogenic *Escherichia coli*', *Veterinary Microbiology*, 153(1–2), pp. 89–98.

Genchi, G. (2017) 'An overview on d-amino acids', *Amino Acids*. Springer-Verlag Wien, pp. 1521–1533.

Gerner, E.W., Bruckheimer, E. and Cohen, A. (2018) 'Cancer pharmacoprevention: Targeting polyamine metabolism to manage risk factors for colon cancer', *Journal of Biological Chemistry*, 293(48), pp. 18770–18778.

Gill, S.K. *et al.* (2021) 'Dietary fibre in gastrointestinal health and disease', *Nature Reviews*, *Gastroenterology & Hepatology*, 18(2), pp.101–116.

Grishin, D. V *et al.* (2019) 'D-amino acids in nature, agriculture and biomedicine', *All Life*, 13(1), pp.11–22.

Grisolia, V., Carlomagno, M.S. and Bruni, C.B. (1982) 'Cloning and expression of the distal portion of the histidine operon of *Escherichia coli* K-12', *Journal of Bacteriology*, 151(2), pp. 692–700.

Gronbach, K. *et al.* (2010) 'Safety of probiotic *Escherichia coli* strain Nissle 1917 depends on intestinal microbiota and adaptive immunity of the host', *Infection and Immunity*, 78(7), pp. 3036–3046.

Guerra, L. *et al.* (2011) 'The biology of the cytolethal distending toxins', *Toxins*, 3(3), pp. 172–190.

Guo, P. *et al.* (2020) 'FadA promotes DNA damage and progression of *Fusobacterium nucleatum*-induced colorectal cancer through up-regulation of chk2', *Journal of Experimental & Clinical Cancer Research : CR*, 39(1).

Gurbatri, C.R. *et al.* (2024) 'Engineering tumor-colonizing *E. coli* Nissle 1917 for detection and treatment of colorectal neoplasia', *Nature Communications 2024 15:1*, 15(1), pp. 1–13.

Haber, A. *et al.* (2017) 'L-glutamine Induces Expression of *Listeria monocytogenes* Virulence Genes', *PLOS Pathogens*, 13(1).

Hallam, J.C. *et al.* (2023) 'D-Serine reduces the expression of the cytopathic genotoxin colibactin', *Microbial Cell*, 10(3), pp.63–77.

Hancock, V., Dahl, M. and Klemm, P. (2010) 'Probiotic *Escherichia coli* strain Nissle 1917 outcompetes intestinal pathogens during biofilm formation', *Journal of Medical Microbiology*, 59(4), pp.392–399.

Hansen, C.H., Endres, R.G. and Wingreen, N.S. (2008) 'Chemotaxis in *Escherichia coli*: A Molecular Model for Robust Precise Adaptation', *PLoS Computational Biology*, 4(1).

Hashimoto, A. *et al.* (1992) 'The presence of free D-serine in rat brain', *FEBS Letters*, 296(1), pp. 33–36.

Hatakeyama, M. (2004) 'Oncogenic mechanisms of the Helicobacter pylori CagA protein', *Nature Reviews Cancer 2004 4:9*, 4(9), pp. 688–694.

Henderson, I.R. *et al.* (1999) 'Characterization of Pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*', *Infection and Immunity*, 67(11), pp. 5587–5596.

Herzer, P.J. *et al.* (1990) 'Phylogenetic Distribution of Branched RNA-Linked Multicopy Single-Stranded DNA among Natural Isolates of *Escherichia coli*', *Journal of Bacteriology*, 172(11), pp. 6175–6181.

Hochbaum, A.I. *et al.* (2011) 'Inhibitory Effects of D-Amino Acids on *Staphylococcus aureus* Biofilm Development', *Journal of Bacteriology*, 193(20), pp. 5616–5622.

Homburg, S. *et al*. (2007) 'Expression analysis of the colibactin gene cluster coding for a novel polyketide in *Escherichia coli*', *FEMS Microbiology Letters*, 275(2), pp. 255–262.

Hook, C. *et al*. (2022) 'The *Escherichia coli* Amino Acid Uptake Protein CycA: Regulation of Its Synthesis and Practical Application in L-Isoleucine Production', *Microorganisms*, 10(3).

Horiguchi, Y. (2001) '*Escherichia coli* cytotoxic necrotizing factors and *Bordetella* dermonecrotic toxin: The dermonecrosis-inducing toxins activating Rho small GTPases', *Toxicon*, 39(11), pp. 1619–1627.

Huncharek, M., Muscat, J. and Kupelnick, B. (2008) 'Colorectal Cancer Risk and Dietary Intake of Calcium, Vitamin D, and Dairy Products: A Meta-Analysis of 26,335 Cases From 60 Observational Studies', *Nutrition and Cancer*, 61(1), pp.47–69.

Ichikawa, H., Halberg, R. and Kroos, L. (1999) 'Negative regulation by the Bacillus subtilis GerE protein', *Journal of Biological Chemistry*, 274(12), pp. 8322–8327.

Iftekhar, A. *et al*. (2021) 'Genomic aberrations after short-term exposure to colibactin-producing *E. coli* transform primary colon epithelial cells', *Nature Communications*, 12(1), pp. 1–15.

Iyoda, S. and Watanabe, H. (2004) 'Positive effects of multiple pch genes on expression of the locus of enterocyte effacement genes and adherence of enterohaemorrhagic *Escherichia coli* O157 : H7 to HEp-2 cells', *Microbiology*, 150(7), pp.2357–2571.

Jack, D.L., Yang, N.M. and Saier, M.H. (2001) 'The drug/metabolite transporter superfamily', *European Journal of Biochemistry*, 268(13), pp. 3620–3639.

Jalili, V. *et al*. (2020) 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update', *Nucleic Acids Research*, 48, pp. 395–402.

Jensen, J.N.E. *et al*. (2024) 'Human Papillomavirus and Associated Cancers: A Review', *Viruses*, 16(5), pp. 1–12.

Jiang, Y. *et al*. (2019) 'Reactivity of an Unusual Amidase May Explain Colibactin's DNA Cross-Linking Activity', *Journal of the American Chemical Society*, 141(29), pp.11489–11496.

Johnson, J.R. *et al*. (2008) 'Molecular epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island', *Journal of Clinical Microbiology*, 46(12), pp. 3906–3911.

Juttukonda, L.J., Chazin, W.J. and Skaar, E.P. (2016) '*Acinetobacter baumannii* coordinates urea metabolism with metal import to resist host-mediated metal limitation', *mBio*, 7(5).

Kaewkod, T. *et al.* (2021) 'Medicinal plant extracts protect epithelial cells from infection and DNA damage caused by colibactin-producing *Escherichia coli*, and inhibit the growth of bacteria', *Journal of Applied Microbiology*, 130(3), pp. 769–785.

Kaiser, J.C. *et al.* (2016) 'The role of two branched-chain amino acid transporters in Staphylococcus aureus growth, membrane fatty acid composition and virulence', *Molecular Microbiology*, 102(5), pp. 850–864.

Khatun, M., Ray, R. and Ray, R.B. (2021) 'Hepatitis C virus associated hepatocellular carcinoma', *Advances in Cancer Research*, 149, pp. 103–142.

Kitamoto, S. *et al.* (2020) 'Dietary l-serine confers a competitive fitness advantage to Enterobacteriaceae in the inflamed gut', *Nature Microbiology*, 5(1), pp. 116–125.

Knasmüller, S. *et al.* (1997) 'Genotoxic effects of three *Fusarium* mycotoxins, fumonisin B1, moniliformin and vomitoxin in bacteria and in primary cultures of rat hepatocytes', *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 391(1–2), pp. 39–48.

Kosari, F. *et al.* (2020) 'Evaluation of cinnamon extract effects on *clbB* gene expression and biofilm formation in *Escherichia coli* strains isolated from colon cancer patients', *BMC Cancer*, 20(1).

Kostic, A.D. *et al.* (2013) '*Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor immune microenvironment', *Cell host & microbe*, 14(2), p. 207.

Kruis, W. *et al.* (2004) 'Maintaining remission of ulcerative colitis with the probiotic *Escherichia coli* Nissle 1917 is as effective as with standard mesalazine', *Gut*, 53(11), pp. 1617–1623.

Kumar Ram, A. and Vairappan, B. (2022) 'World Journal of Clinical Cases Role of zonula occludens in gastrointestinal and liver cancers, *World J Clin Cases*, 10(12), pp. 3647–3661.

Kuzela, L., Kascak, M. and Vavrecka, A. (2001) 'Induction and maintenance of remission with nonpathogenic *Escherichia coli* in patients with pouchitis', *The American Journal of Gastroenterology*, 96(11), pp. 3218–3219.

Landry, J.J.M. *et al.* (2013) 'The genomic and transcriptomic landscape of a hela cell line', *G3: Genes, Genomes, Genetics*, 3(8), pp. 1213–1224.

Lauko, A. et al. (2020) 'Junctional Adhesion Molecules in Cancer: A Paradigm for the Diverse Functions of Cell-Cell Interactions in Tumor Progression', *Cancer Research*, 80(22), pp.4878–4885.

Li, H. *et al*. (2023) 'Chirality of tyrosine controls biofilm formation via the regulation of bacterial adhesion', *Biochemical Engineering Journal*, 192(February), p. 108844.

Li, S. *et al*. (2022) 'Tumorigenic bacteria in colorectal cancer: mechanisms and treatments', *Cancer Biology & Medicine*, 19(2), p. 147.

Li, Z.-R. *et al*. (2019) 'Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage', *Nature Chemistry*, 11(10), pp.880–889.

Lin, R. *et al*. (2017) 'A review of the relationship between the gut microbiota and amino acid metabolism', *Amino Acids*, 49, pp. 2083–2090.

Liu, B. *et al*. (2020) 'Structure and genetics of *Escherichia coli* O antigens', *FEMS Microbiology Reviews*, 028, pp. 655–683.

Lopès, A. *et al*. (2020) 'Colibactin-positive *Escherichia coli* induce a procarcinogenic immune environment leading to immunotherapy resistance in colorectal cancer', *International Journal of Cancer*, 146(11), pp. 3147–3159.

Luettig, J. *et al*. (2015) 'Claudin-2 as a mediator of leaky gut barrier during intestinal inflammation', *Tissue Barriers*, 3(1-2).

Luo, C. *et al*. (2011) 'Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species', *PNAS*, 108(17), pp.7200–7205.

Ma, Y., Liu, X. and Wang, J. (2022) 'Small molecules in the big picture of gut microbiome-host cross-talk', *eBioMedicine*, 81, p. 104085.

Macia, L. *et al*. (2017) 'Intestinal Microbiota-Derived gaBa Mediates interleukin-17 expression during enterotoxigenic *Escherichia coli* infection',. *Front. Immunol*, 7, p. 685.

Mackie, R.I., Sghir, A. and Gaskins, H.R. (1999) 'Developmental microbial ecology of the neonatal gastrointestinal tract', *American Journal of Clinical Nutrition*, 69(5), pp. 1035S-1045S.

Martin, P. *et al*. (2013) 'Interplay between Siderophores and Colibactin Genotoxin Biosynthetic Pathways in *Escherichia coli*', *PLoS Pathogens*, 9(7).

Martínez-Maqueda, D., Miralles, B. and Recio, I. (2015) 'HT29 Cell Line', *The Impact of Food Bioactives on Health: In Vitro and Ex Vivo Models*, pp. 113–124.

Maryam, S. *et al*. (2023) 'Interleukins (Cytokines) as Biomarkers in Colorectal Cancer : Progression , Detection , and Monitoring', *Journal of Clinical Medicine*, 12(9), pp. 1–32.

Massip, C. *et al.* (2019) 'Deciphering the interplay between the genotoxic and probiotic activities of *Escherichia coli* Nissle 1917', *PLOS Pathogens*, 15(9).

Massip, C. *et al.* (2020) 'The synergistic triad between microcin, colibactin, and salmochelin gene clusters in uropathogenic *Escherichia coli*', *Microbes and Infection*, 22(3), pp. 144–147.

Mattiuzzi, C., Sanchis-Gomar, F. and Lippi, G. (2019) 'Concise update on colorectal cancer epidemiology', *Annals of Translational Medicine*, 7(21), pp. 609–609.

Meftah, A., Hasegawa, H. and Kantrowitz, J.T. (2021) 'D-Serine: A Cross Species Review of Safety', *Frontiers in Psychiatry*, 12, pp. 1–13.

Meng, Q. *et al.* (2021) '*Fusobacterium nucleatum* secretes amyloid-like FadA to enhance pathogenicity', *EMBO Reports*, 22(7).

El Meouche, I. and Dunlop, M.J. (2018) 'Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation', *Science*, 362(6415), pp. 686–690.

Metzler, D.E. and Snell, E.E. (1952) 'Deamination of serine. II. d-Serine dehydrase, a vitamin B6 enzyme from *Escherichia coli*', *The Journal of biological chemistry*, 198(1).

Miyamoto, T. *et al.* (2017) 'Identification and characterization of novel broad-spectrum amino acid racemases from *Escherichia coli* and *Bacillus subtilis*', *Amino Acids*, 49(11), pp.1885–1894.

Miyamoto, T. *et al.* (2018) 'Cystathionine ß-lyase is involved in D-amino acid metabolism', *Biochemical Journal*, 475(8), pp.1397–1410.

Mizoguchi, E. *et al.* (2021) 'A Review of Selected IBD Biomarkers: From Animal Models to Bedside', *Diagnostics*, 11(2).

Morgan, R.N. *et al.* (2022) 'Bacterial cyclomodulins: types and roles in carcinogenesis', *Critical Reviews in Microbiology*, 48(1), pp. 42–66.

Moritz, R.L. and Welch, R.A. (2006) 'The *Escherichia coli* argW-dsdCXA genetic island is highly variable, and E. coli K1 strains commonly possess two copies of dsdCXA', *Journal of Clinical Microbiology*, 44(11), pp. 4038–4048.

Mostafa, M.H., Sheweita, S.A. and O'Connor, P.J. (1999) 'Relationship between schistosomiasis and bladder cancer', *Clinical Microbiology Reviews*, 12(1), pp. 97–111.

Mousa, J.J. *et al.* (2016) 'MATE transport of the E. coli-derived genotoxin colibactin', *Nature Microbiology*, 1(1).

Munro, L.J. and Kell, D.B. (2022) 'Analysis of a Library of *Escherichia coli* Transporter Knockout Strains to Identify Transport Pathways of Antibiotics', *Antibiotics*, 11(8), pp. 1–15.

Murkin, A.S. and Tanner, M.E. (2002) 'Dehydroalanine-Based Inhibition of a Peptide Epimerase from Spider Venom', *The Journal of Organic Chemistry*, 67(24), pp.8389–8394.

Nagao, T. *et al.* (2018) 'L-histidine augments the oxidative damage against Gram-negative bacteria by hydrogen peroxide', *International Journal of Molecular Medicine*, 41(5), pp. 2847–2854.

Nagata, Y. *et al.* (1995) 'Free d-serine concentration in normal and Alzheimer human brain', *Brain Research Bulletin*, 38(2), pp. 181–183.

Nakade, Y. *et al.* (2018) 'Gut microbiota-derived D-serine protects against acute kidney injury', *JCI Insight*, 3(20).

Nguyen, Y. and Sperandio, V. (2012) 'Enterohemorrhagic *E. coli* (EHEC) pathogenesis', *Frontiers in Cellular and Infection Microbiology*, 2(90).

Nishikawa, T. (2011) 'Analysis of free d-serine in mammals and its biological relevance', *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 879(29), pp. 3169–3183.

Norat, T. *et al.* (2005) 'Meat, fish, and colorectal cancer risk: The European prospective investigation into cancer and nutrition', *Journal of the National Cancer Institute*, 97(12), pp. 906–916.

Norregaard-Madsen, M., McFall, E. and Valentin-Hansen, P. (1995) 'Organization and transcriptional regulation of the *Escherichia coli* K-12 D-serine tolerance locus', *Journal of Bacteriology*, 177(22), pp. 6456–6461.

Nougayrède, J.P. *et al.* (2005) 'Cyclomodulins: Bacterial effectors that modulate the eukaryotic cell cycle', *Trends in Microbiology*, 13(3), pp. 103–110.

Nougayrède, J.P. *et al.* (2006) '*Escherichia coli* induces DNA double-strand breaks in eukaryotic cells', *Science*, 313(5788), pp. 848–851.

Nougayrède, J.P. *et al.* (2021) 'A Toxic Friend: Genotoxic and Mutagenic Activity of the Probiotic Strain *Escherichia coli* Nissle 1917', *mSphere*, 6(4).

Novikova, O.D. and Solovyeva, T.F. (2009) 'Nonspecific Porins of the Outer Membrane of Gram-Negative Bacteria: Structure and Functions', *Biologicheskie Membrany*, 3(1), pp. 6–20.

Olier, M. *et al.* (2012) 'Genotoxicity of *Escherichia coli* Nissle 1917 strain cannot be dissociated from its probiotic activity', *Gut Microbes*, 3(6), pp. 501–509.

Oliero, M. *et al.* (2021) 'Oligosaccharides increase the genotoxic effect of colibactin produced by *pks+ Escherichia coli* strains', *BMC Cancer 2021 21:1*, 21(1), pp. 1–10.

Orskov, I. *et al.* (1977) 'Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*', *Bacteriological Reviews*, 41(3), pp. 667–710.

Oswald, E. *et al.* (2005) 'Bacterial toxins that modulate host cell-cycle progression', *Current Opinion in Microbiology*, 8(1), pp. 83–91.

Pannone, G. *et al.* (2014) 'The role of EBV in the pathogenesis of Burkitt's Lymphoma: an Italian hospital based survey', *Infectious Agents and Cancer*, 9(1).

Partanen, K.H. and Mroz, Z. (1999) 'Organic acids for performance enhancement in pig diets', *Nutrition Research Reviews*, 12(1), pp. 117–145.

Pavao, A. *et al.* (2022) 'Reconsidering the *in vivo* functions of Clostridial Stickland amino acid fermentations', *Anaerobe*, 76.

Pérès, S.Y. *et al.* (1997) 'A new cytolethal distending toxin (CDT) from *Escherichia coli* producing CNF2 blocks HeLa cell division in G2/M phase', *Molecular Microbiology*, 24(5), pp. 1095–1107.

Pérez-Berezo, T. *et al.* (2017) 'Identification of an analgesic lipopeptide produced by the probiotic *Escherichia coli* strain Nissle', *Nature Communications*, 8(1), pp. 1–12.

Perez-Carrasco, V. *et al.* (2021) 'Urinary Microbiome: Yin and Yang of the Urinary Tract', *Frontiers in Cellular and Infection Microbiology*, 11, pp. 1–17.

Peters, U. *et al.* (2003) 'Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme', *Lancet*, 361(9368), pp. 1491–1495.

Pinto, C. *et al.* (2021) 'The Selective Advantage of the lac Operon for *Escherichia coli* Is Conditional on Diet and Microbiota Composition', *Frontiers in Microbiology*, 12(709259).

Pleguezuelos-Manzano, C. *et al.* (2020) 'Mutational signature in colorectal cancer caused by genotoxic *pks* + *E. coli*', *Nature*, 580(7802), pp. 269–273.

Putze, J. *et al.* (2009) 'Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae', *Infection and Immunity*, 77(11), pp. 4696–4703.

Raisch, J. *et al.* (2014) 'Colon cancer-associated B2 *Escherichia coli* colonize gut mucosa and promote cell proliferation', *World Journal of Gastroenterology : WJG*, 20(21).

Reddy Kunduru, B., Nair, S.A. and Rathinavelan, T. (2015) 'EK3D: an *E. coli* K antigen 3-dimensional structure database', *Nucleic Acids Research*, 44, pp. 675–681.

Rehm, N. *et al.* (2022) 'Two Polyketides Intertwined in Complex Regulation: Posttranscriptional CsrA-Mediated Control of Colibactin and Yersiniabactin Synthesis in *Escherichia coli*', *mBio*, 13(1).

Rhee, K.Y., Parekh, B.S. and Hatfield, G.W. (1996) 'Leucine-responsive regulatory protein-DNA interactions in the leader region of the ilvGMEDA operon of *Escherichia coli*', *Journal of Biological Chemistry*, 271(43), pp. 26499–26507.

Riordan, J.T. and Mitra, A. (2017) ' Regulation of *Escherichia coli* Pathogenesis by Alternative Sigma Factor N ', *EcoSal Plus*, 7(2).

De Robertis, M. *et al.* (2011) 'The AOM/DSS murine model for the study of colon carcinogenesis: From pathways to diagnosis and therapy studies', *Journal of Carcinogenesis*, 10.

Rodríguez-Nogales, A. *et al.* (2018) 'The administration of *Escherichia coli* Nissle 1917 ameliorates development of DSS-induced colitis in mice', *Frontiers in Pharmacology*, 9(468).

Roe, A.J. *et al.* (2004) 'Co-ordinate single-cell expression of LEE4- and LEE5-encoded proteins of *Escherichia coli* O157:H7', *Molecular Microbiology*, 54(2), pp. 337–352.

Roe, A.J. *et al.* (2007) 'Analysis of the expression, regulation and export of NleA-E in *Escherichia coli* O157 : H7', *Microbiology*, 153(5), pp.1350–1360.

Roesch, P.L. *et al.* (2003) 'Uropathogenic *Escherichia coli* use D-serine deaminase to modulate infection of the murine urinary tract', *Molecular Microbiology*, 49(1), pp. 55–67.

Rumbo, C. et al. (2016) 'Assessment of antivirulence activity of several D-amino acids against *Acinetobacter baumannii* and *Pseudomonas aeruginosa*', *Journal of Antimicrobial Chemotherapy*, 71(12), pp.3473–3481.

De Rycke, J. *et al.* (1990) 'Evidence for two types of cytotoxic necrotizing factor in human and animal clinical isolates of *Escherichia coli*', *Journal of Clinical Microbiology*, 28(4), pp. 694–699.

Sabri, S., Nielsen, L.K. and Vickers, C.E. (2013) 'Molecular Control of Sucrose Utilization in *Escherichia coli* W, an Efficient Sucrose-Utilizing Strain', *Applied and Environmental Microbiology*, 79(2), pp.478–487.

Sadecki, Patric W. *et al.* (2021) 'Evolution of polymyxin resistance regulates colibactin production in *Escherichia coli*', *ACS chemical biology*, 16(7).

Sadecki, Patric W *et al.* (2021) 'Evolution of polymyxin resistance regulates colibactin production in *Escherichia coli*', *ACS Chem Biol*, 16(7), pp. 1243–1254.

Salesse, L. *et al.* (2021) 'Colibactin-producing *Escherichia coli* induce the formation of invasive carcinomas in a chronic inflammation-associated mouse model', *Cancers*, 13(9).

Sandhya Guntaka, N. *et al.* (2017) 'Structure and Functional Analysis of ClbQ, an Unusual Intermediate-Releasing Thioesterase from the Colibactin Biosynthetic Pathway', *ACS Chemical Biology*, 12(10), pp.2598–2608.

Sarkar, S.K., Bhattacharyya, A. and Mandal, S.M. (2015) 'YnfA, a SMR family efflux pump is abundant in *Escherichia coli* isolates from urinary infection', *Indian Journal of Medical Microbiology*, 33(1), pp. 139–142.

Schultz, M. *et al.* (2004) 'Preventive Effects of *Escherichia coli* Strain Nissle 1917 on Acute and Chronic Intestinal Inflammation in Two Different Murine Models of Colitis', *Clinical and Diagnostic Laboratory Immunology*, 11(2), pp. 372–378.

Schultz, M. and Burton, J.P. (2017) '*Escherichia coli* Nissle 1917', in *The Microbiota in Gastrointestinal Pathophysiology: Implications for Human Health, Prebiotics, Probiotics, and Dysbiosis*. Elsevier Inc., pp. 59–69.

Scott, A.J. et al. (2018) 'Gut microbiota International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis', *Gut*, 68(9), pp.1624–1632.

Sears, C.L. and Pardoll, D.M. (2011) 'Perspective: Alpha-Bugs, Their Microbial Partners, and the Link to Colon Cancer', *The Journal of Infectious Diseases*, 203(3), pp. 306–311.

Secher, T. *et al.* (2013) '*Escherichia coli* Producing Colibactin Triggers Premature and Transmissible Senescence in Mammalian Cells', *PLoS ONE*, 8(10), p. 77157.

Shen, C. *et al.* (2017) 'Long-distance interaction of the integrated HPV fragment with MYC gene and 8q24.22 region upregulating the allele-specific MYC expression in HeLa cells', *International Journal of Cancer*, 141(3), pp. 540–548.

Sherman, H., Chapnik, N. and Froy, O. (2006) 'Albumin and amino acids upregulate the expression of human beta-defensin 1', *Molecular Immunology*, 43(10), pp. 1617–1623.

Shimpoh, T. *et al.* (2017) 'Prevalence of *pks*-positive *Escherichia coli* in Japanese patients with or without colorectal cancer', *Gut Pathogens*, 9(1).

Shivers, R.P. and Sonenshein, A.L. (2004) 'Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids', *Molecular Microbiology*, 53(2), pp. 599–611.

Shleper, M., Kartvelishvily, E. and Wolosker, H. (2005) 'D-Serine Is the Dominant Endogenous Coagonist for NMDA Receptor Neurotoxicity in Organotypic Hippocampal Slices', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(41), pp.9413–7.

Silpe, J.E. *et al.* (2022) 'The bacterial toxin colibactin triggers prophage induction', *Nature*, 603(7900), pp. 315–320.

Smith, J.M., Smolin, D.E. and Edwin Umbarger, H. (1976) 'Polarity and the Regulation of the *ilv* Gene Cluster in *Escherichia coli* Strain K-12', MGG Molecular & General Genetics, 148(2), pp.111–124.

Sogari, A. *et al.* (2024) 'Tolerance to colibactin correlates with homologous recombination proficiency and resistance to irinotecan in colorectal cancer cells', *Cell Reports Medicine*, 5(2), p. 101376.

Solanas, C. *et al.* (2009) 'Therapeutic index of gramicidin S is strongly modulated by D-phenylalanine analogues at the ß-turn', *Journal of Medicinal Chemistry*, 52(3), pp. 664–674.

Soutourina, J. *et al.* (1999) 'Functional Characterization of the D-Tyr-tRNA Tyr Deacylase from *Escherichia coli*', *Journal of Biological Chemistry*, 274(27), pp.19109–19114.

Soutourina, O. *et al.* (2004) 'Formation of D-Tyrosyl-tRNA Tyr Accounts for the Toxicity of D-Tyrosine toward *Escherichia coli*', *Journal of Biological Chemistry*, 279, pp. 42560–42565.

Sripa, B. *et al.* (2012) 'The tumorigenic liver fluke Opisthorchis viverrini-multiple pathways to cancer', *Trends in parasitology*, 28(10), pp.395–407.

Stauffer, G. V. (2004) 'Regulation of Serine, Glycine, and One-Carbon Biosynthesis', *EcoSal Plus*, 1(1).

Stoll, T. *et al.* (2009) 'A new member of a growing toxin family - *Escherichia coli* cytotoxic necrotizing factor 3 (CNF3)', *Toxicon*, 54(6), pp. 745–753.

Stritzker, J. *et al.* (2007) 'Tumor-specific colonization, tissue distribution, and gene induction by probiotic *Escherichia coli* Nissle 1917 in live mice', *International Journal of Medical Microbiology*, 297(3), pp. 151–162.

Sturm, A. *et al.* (2005) '*Escherichia coli* Nissle 1917 distinctively modulates T-cell cycling and expansion via Toll-like receptor 2 signalling', *Infection and Immunity*, 73(3), pp. 1452–1465.

Sugihara, K. *et al.* (2020) 'The role of dietary L-Serine in the regulation of the intestinal mucus barrier during inflammation', *Gastroenterology*, 158, p. S70.

Takayama, T. *et al.* (2005) 'Esterification of *Escherichia coli* tRNAs with D-histidine and D-lysine by aminoacyl-tRNA synthetases', *Bioscience, Biotechnology and Biochemistry*, 69(5), pp. 1040–1041.

Tang, J.-W. *et al.* (2022) 'Biosynthesis and bioactivities of microbial genotoxin colibactins', *Natural Product Reports*, 39(5), pp.991–1014.

Tang-Fichaux, M. *et al.* (2020) 'The Polyphosphate Kinase of *Escherichia coli* Is Required for Full Production of the Genotoxin Colibactin', *mSphere*, 5(6).

Tang-Fichaux, M. *et al.* (2021) 'Tackling the Threat of Cancer Due to Pathobionts Producing Colibactin: Is Mesalamine the Magic Bullet?', *Toxins 2021, Vol. 13, Page 897*, 13(12), p. 897.

Tariq, H. *et al.* (2022) 'Colibactin possessing *E. coli* isolates in association with colorectal cancer and their genetic diversity among Pakistani population', *PLoS ONE*, 17(11 November), pp. 1–13.

Tenaillon, O. *et al.* (2010) 'The population genetics of commensal *Escherichia coli*', *Nature Reviews Microbiology*, 8(3), pp.207–217.

Thanikachalam, K. and Khan, G. (2019) 'Colorectal cancer and nutrition', *Nutrients*, 11(1).

Thibodeaux, C.J., Chang, W.-C. and Liu, H.-W. (2011) 'The Enzymatic Chemistry of Cyclopropane, Epoxide, and Aziridine Biosynthesis', *Chemical Reviews*, 112(3), pp.1681–1709.

Thrift, A.P. *et al.* (2015) 'Mendelian Randomization Study of Body Mass Index and Colorectal Cancer Risk', *Cancer Epidemiology Biomarkers & Prevention*, 24(7), pp.1024–1031.

Tjalsma, H. *et al.* (2012) 'A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects', *Nature Reviews Microbiology 2012 10:8*, 10(8), pp. 575–582.

Tripathi, P. *et al.* (2017) 'ClbS Is a Cyclopropane Hydrolase That Confers Colibactin Resistance', *Journal of the American Chemical Society*, 139(49), pp.17719–17722.

Tripathi, P. *et al.* (2023) 'Structural basis of the amidase ClbL central to the biosynthesis of the genotoxin colibactin', *Acta Crystallographica Section D: Structural Biology*, 79, pp. 830–836.

Troge, A. *et al.* (2012) 'More than a marine propeller - the flagellum of the probiotic *Escherichia coli* strain Nissle 1917 is the major adhesin mediating binding to human mucus', *International Journal of Medical Microbiology*, 302(7–8), pp. 304–314.

Trommelen, J., Tom, D. and van Loon, L.J. (2021) 'Gut amino acid absorption in humans: Concepts and relevance for postprandial metabolism', *Clinical Nutrition Open Science*, 36, pp.43–55.

Tronnet, S. *et al.* (2016) 'Iron Homeostasis Regulates the Genotoxicity of *Escherichia coli* That Produces Colibactin', *Infection and Immunity*, 84(12), p. 3358.

Tsai, G. *et al.* (1998) 'D-serine added to antipsychotics for the treatment of schizophrenia', *Biological Psychiatry*, 44(11), pp. 1081–1089.

Turnbaugh, P.J. *et al.* (2007) 'The Human Microbiome Project', *Nature*, pp. 804–810.

Umeda, S. *et al.* (2023) 'D-amino Acids Ameliorate Experimental Colitis and Cholangitis by Inhibiting Growth of Proteobacteria: Potential Therapeutic Role in Inflammatory Bowel Disease', *Cellular and Molecular Gastroenterology and Hepatology*, 16(6), pp.1011–1031.

Veziant, J. *et al.* (2016) 'Association of colorectal cancer with pathogenic *Escherichia coli*: Focus on mechanisms using optical imaging', *World Journal of Clinical Oncology*, 7(3), pp. 293–301.

Vizcaino, M.I. *et al.* (2014) 'Comparative Metabolomics and Structural Characterizations Illuminate Colibactin Pathway-Dependent Small Molecules', *Journal of the American Chemical Society*, 136(26), pp.9244–9247.

Vizcaino, M.I. and Crawford, J.M. (2015) 'The colibactin warhead crosslinks DNA', *Nature Chemistry*, 7(5), pp. 411–417.

Wallenstein, A. *et al.* (2020) 'ClbR Is the Key Transcriptional Activator of Colibactin Gene Expression in *Escherichia coli*', *mSphere*, 5(4).

Wami, H. *et al.* (2021) 'Insights into evolution and coexistence of the colibactin-and yersiniabactin secondary metabolite determinants in enterobacterial populations', *Microbial Genomics*, 7(6).

Wang, C. *et al.* (2020) 'Improving l-serine formation by *Escherichia coli* by reduced uptake of produced l-serine', *Microbial Cell Factories*, 19(1), pp. 1–10.

Wang, L. *et al.* (2003) 'Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene', *Journal of Bacteriology*, 185(9), pp. 2936–2943.

Wehkamp, J. *et al.* (2004) 'NF-κB- and AP-1-mediated induction of human beta defensin-2 in intestinal epithelial cells by *Escherichia coli* Nissle 1917: A novel effect of a probiotic bacterium', *Infection and Immunity*, 72(10), pp. 5750–5758.

van der Wielen, N., Moughan, P.J. and Mensink, M. (2017) 'Amino Acid Absorption in the Large Intestine of Humans and Porcine Models', *The Journal of Nutrition*, 147(8), pp. 1493–1498.

Wilson, M.R. *et al.* (2019) 'The human gut bacterial genotoxin colibactin alkylates DNA', *Science*, 363(6428).

Wolosker, H. *et al.* (2008) 'D-amino acids in the brain: D-serine in neurotransmission and neurodegeneration', *FEBS Journal*, 275(14), pp. 3514–3526.

Wong, J.J. *et al.* (2022) '*Escherichia coli* BarA-UvrY regulates the *pks* island and kills Staphylococci via the genotoxin colibactin during interspecies competition', *PLoS Pathogens*, 18(9), pp. 1–26.

World Health Organisation International Agency for Research on Cancer (1994) 'Schistosomes , liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans', *IARC Monogr Eval Carcinog Risks Hum.*, 61, pp. 1–241.

Wu, D. *et al.* (2023) '*Escherichia coli* Nissle 1917-driven microrobots for effective tumor targeted drug delivery and tumor regression', *Acta Biomaterialia*, 169, pp. 477–488.

Xue, M. *et al.* (2019) 'Structure elucidation of colibactin and its DNA cross-links', *Science*, 365(6457).

Yu, C. *et al.* (2016) 'Inhibition of biofilm formation by d-tyrosine: Effect of bacterial type and d-tyrosine concentration', *Water Research*, 92, pp. 173–179.

Yu, D., Banting, G. and Neumann, N.F. (2021) 'A review of the taxonomy, genetics, and biology of the genus *Escherichia* and the type species *Escherichia coli*', *Canadian Journal of Microbiology*, 67(8).

Zeng, W. *et al.* (2023) 'Enteral nutrition promotes the remission of colitis by gut bacteria-mediated histidine biosynthesis', *EBioMedicine*, 100, pp.104959–104959.

Zha, L. *et al.* (2016) 'Characterization of Polyketide Synthase Machinery from the pks Island Facilitates Isolation of a Candidate Precolibactin', *ACS chemical biology*, 11(5), pp.1287–1295.

Zha, L. *et al.* (2017) 'Colibactin assembly line enzymes use S-adenosylmethionine to build a cyclopropane ring', *Nat Chem Biol*, 13(10), pp. 1063–1065.

Zhang, H. *et al.* (2018) 'Serine alleviates dextran sulfate sodium-induced colitis and regulates the gut microbiota in mice', *Frontiers in Microbiology*, 9, pp. 1–10.

Zhang, S. *et al*. (2020) 'Short Chain Fatty Acids Modulate the Growth and Virulence of Pathosymbiont *Escherichia coli* and Host Response', *Antibiotics*, 9(8), p.462.

Zhang, X.X. *et al*. (2012) 'Variation in transport explains polymorphism of histidine and urocanate utilization in a natural *Pseudomonas* population', *Environmental Microbiology*, 14(8), pp. 1941–1951.

Zhou, J. *et al*. (2016) 'L-Serine enables reducing the virulence of *Acinetobacter baumannii* and modulating the SIRT1 pathway to eliminate the pathogen', *Microbiology Spectrum*, 12(3).

Zhou, Q. *et al*. (2019) 'Risk of Colorectal Cancer in Ulcerative Colitis Patients: A Systematic Review and Meta-Analysis', *Gastroenterology Research and Practice*, 2019, pp.1–11.

Zimmer, D.P. *et al*. (2000) 'Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: Scavenging as a defense against nitrogen limitation', *PNAS*, 97(26).

10 Appendix

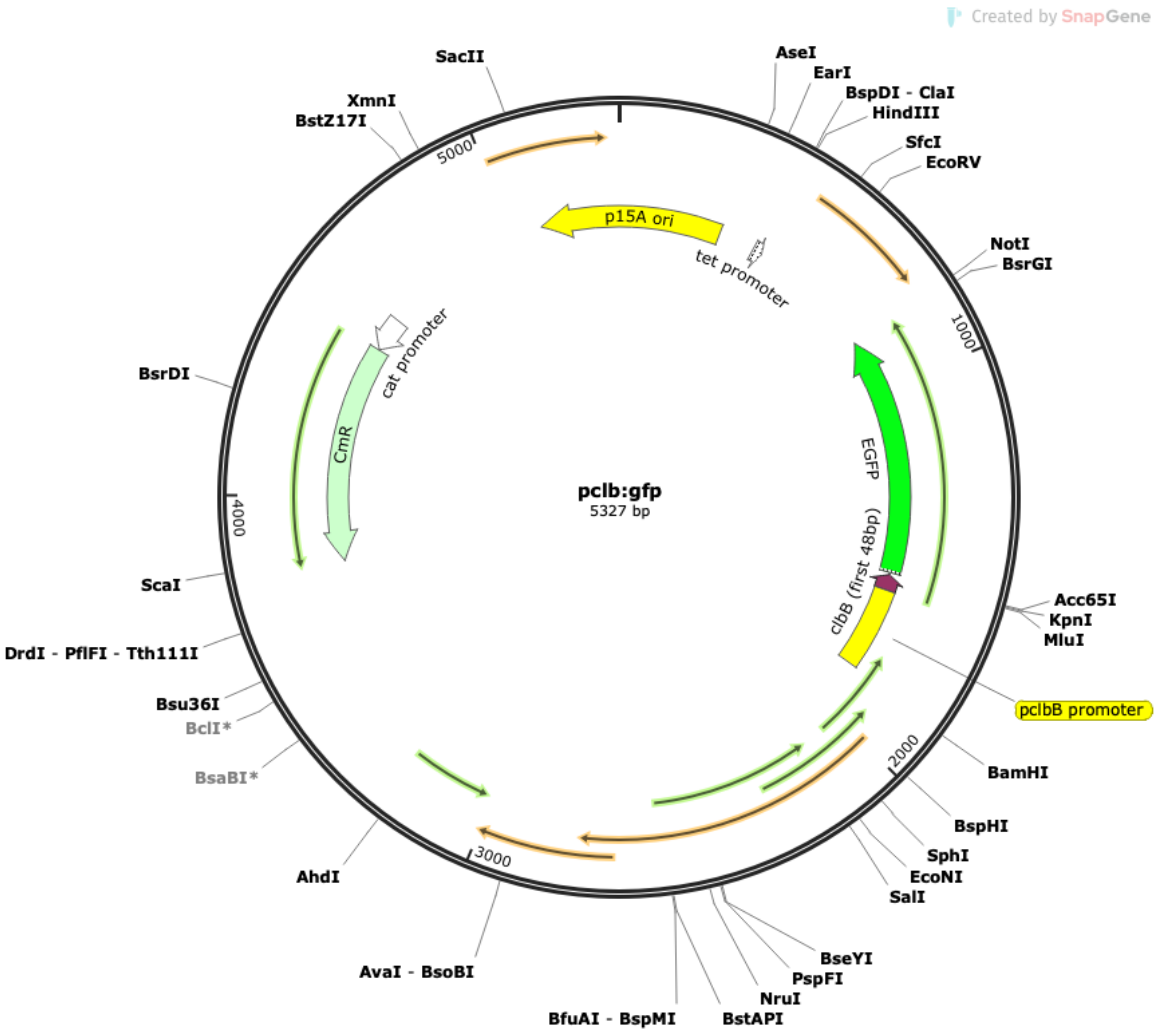


Figure 10-1 Map of reporter fusion plasmid *pclbB:gfp*. Expression of *clbB* was measured indirectly through the fluorescence emitted by the *clbB* promoter:GFP reporter fusion construct. It was transformed by electroporation into bacterial strains and maintained by chloramphenicol supplementation.

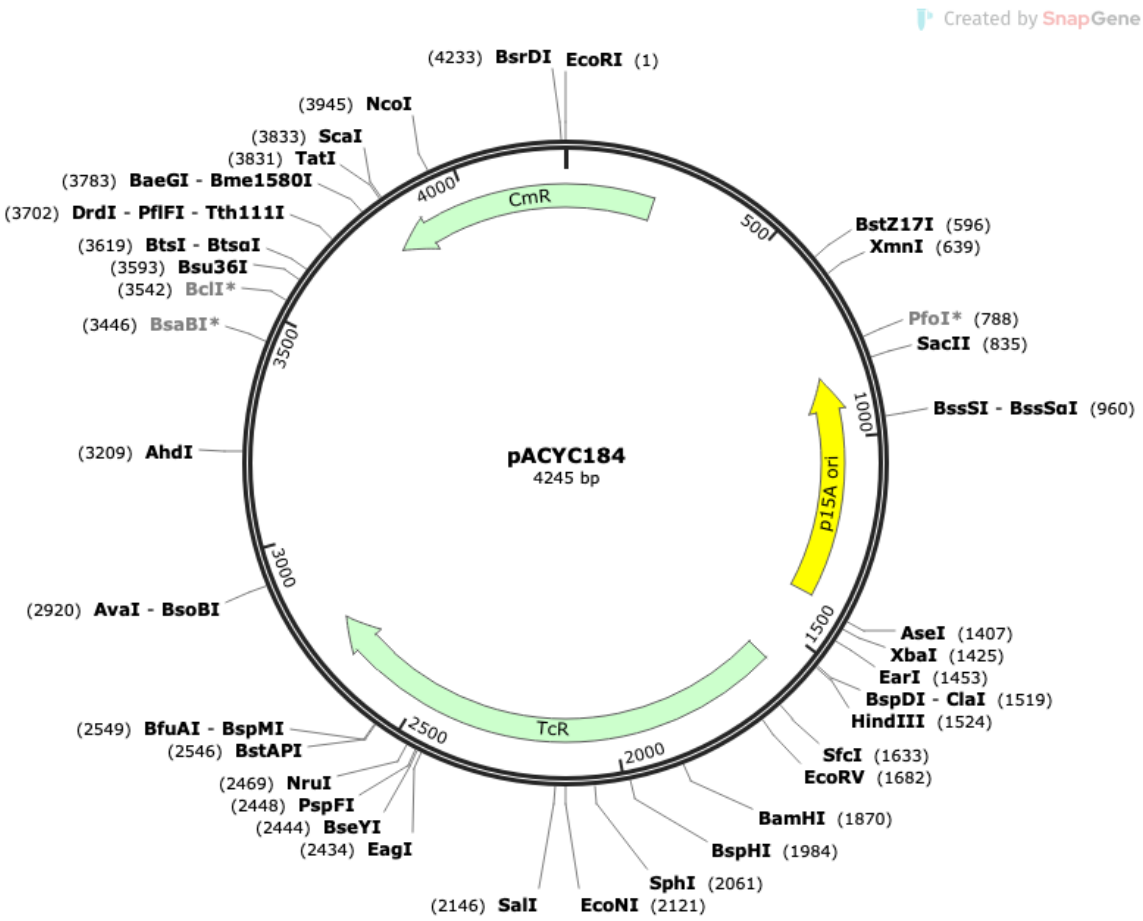


Figure 10-2 Map of empty vector plasmid pACYC184. The plasmid used to measure background fluorescence during transcriptional reporter assays. It was transformed by electroporation into bacterial strains and maintained by chloramphenicol supplementation.

Table 10-1 Genes with significant differential expression in EcN_EO cultured in M9 + 1 mM D-Serine as identified by RNA-seq

Gene name	Gene product	M9 vs. M9 +D-Serine Log fold change	-Log_{10} FDR-corrected p-value
aldB_2	aldehyde dehydrogenase AldB	-0.844389906	2.011351875
arcC_3	carbamate kinase	-0.858093159	1.838911069
betB	betaine-aldehyde dehydrogenase	-0.621623375	2.036117771
betI	transcriptional regulator BetI	-0.592627205	1.905573725
btsT	pyruvate/proton symporter BtsT	-3.217847321	33.194771090
cheB	protein-glutamate methylesterase/protein glutamine deamidase	0.707622263	3.001261284
cheR	protein-glutamate O-methyltransferase CheR	0.629191574	1.998575547
cheW	chemotaxis protein CheW	0.713296457	3.700928740
cheY	chemotaxis response regulator CheY	0.706402212	3.193480866
cheZ	protein phosphatase CheZ	0.673431181	4.009084245
crfC	clamp-binding protein CrfC	0.606636890	1.615359134
eptA	phosphoethanolamine transferase EptA	0.655622837	1.881179187
fadA	acetyl-CoA C-acyltransferase FadA	0.608765360	1.746279768
flgK	flagellar hook-associated protein FlgK	0.913040608	5.968591536
flgL	flagellar hook-associated protein FlgL	0.820921051	5.500038134
flgN	flagella biosynthesis chaperone FlgN	0.647715061	3.137808969
fliA	RNA polymerase sigma factor FliA	0.760426122	4.361111575
fliD	flagellar filament capping protein FliD	0.826199819	3.700928740
fliS	flagellar export chaperone FliS	0.857404293	3.874844170
fliT	flagella biosynthesis regulatory protein FliT	0.702981625	1.469853563
fliZ_1	flagella biosynthesis regulatory protein FliZ	0.755582119	1.615359134
fliZ_3	flagella biosynthesis regulatory protein FliZ	0.748606604	1.525453449

pdeH	cyclic-guanylate-specific phosphodiesterase	1.044566279	2.011351875
queA_2	tRNA preQ1(34) S-adenosylmethionine ribosyltransferase-isomerase QueA	2.731373153	2.036117771
rdcB	LysR family transcriptional regulator	0.826606564	5.492548939
suhB	inositol-1-monophosphatase	0.854536897	2.052566278
tar	methyl-accepting chemotaxis protein II	0.727265743	2.945004138
tsr	methyl-accepting chemotaxis protein	0.742749659	3.050854048
uraA	uracil permease	1.027562195	1.545918734
ycgR	flagellar brake protein	0.872787688	3.700928740
yidD	membrane protein insertion efficiency factor YidD	0.756601859	3.137808969
yjiX	YbdD/YjiX family protein	-1.120887569	4.429340330
zraP	zinc resistance sensor/chaperone ZraP	-0.844723654	2.389765825

Table 10-2Genes with significant differential expression in EcN_EO cultured in M9 + 1 mM D-Tyrosine as identified by RNA-seq

Gene name	Gene product	M9 vs. M9 + D-Tyrosine Log fold change	-Log10 FDR-corrected p-value
arcA_1	arginine deiminase	-1.117665944	1.483957914
arcC_3	carbamate kinase	-1.104561416	1.548889501
aroH	3-deoxy-7-phosphoheptulonate synthase AroH	0.610371732	1.609210171
btsT	pyruvate/proton symporter BtsT	1.780885544	3.260032303
cbl	HTH-type transcriptional regulator Cbl	-0.867768561	1.658042127
clbD	colibactin biosynthesis dehydrogenase ClbD	-0.932417901	1.971749662
clbE	colibactin biosynthesis aminomalonyl-acyl carrier protein ClbE	-0.935837589	1.497293611
clbF	colibactin biosynthesis dehydrogenase ClbF	-1.076832432	2.419759492
clbG	colibactin biosynthesis acyltransferase ClbG	-1.134651156	2.825068406
clbH	colibactin non-ribosomal peptide synthetase ClbH	-1.275569609	4.478338985
clbI	colibactin polyketide synthase ClbI	-1.247791719	2.332360294
clbJ	colibactin non-ribosomal peptide synthetase ClbJ	-1.277110966	3.532836034
clbK	colibactin hybrid non-ribosomal peptide synthetase/type I polyketide synthase ClbK	-1.223711155	4.286761538
clbL	colibactin biosynthesis amidase ClbL	-1.197452195	2.384997385
clbM	precolibactin export MATE transporter ClbM	-1.013075415	1.851451096
clbN	colibactin non-ribosomal peptide synthetase ClbN	-1.113218640	3.708631150
clbO	colibactin polyketide synthase ClbO	-1.009380985	2.277119389
clbP	precolibactin peptidase ClbP	-0.965926232	2.150949309
clbQ	colibactin biosynthesis thioesterase ClbQ	-0.974561517	2.199764211
csgA	curli major subunit CsgA	-2.115167657	17.928117993
csgB	curli minor subunit CsgB	-2.478107778	25.265760396
cysS	cysteine--tRNA ligase	0.752884165	2.009217308
dcuB	anaerobic C4-dicarboxylate transporter DcuB	1.400303025	3.142064735
dsdA	D-serine ammonia-lyase	0.802462549	1.961383172

dusC_1	tRNA dihydrouridine(16) synthase DusC	-0.825407373	1.600899841
efeO	iron uptake system protein EfeO	0.616443873	1.521933233
entA	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase EntA	0.682412099	2.091997799
entB	enterobactin biosynthesis bifunctional isochorismatase/aryl carrier protein EntB	0.671954080	1.909200937
entH	proofreading thioesterase EntH	0.732953309	2.352910571
eptA	phosphoethanolamine transferase EptA	0.734791243	2.620694482
frdC_2	fumarate reductase subunit FrdC	0.951489055	1.466460873
ftnA	non-heme ferritin	0.745247172	1.558618249
gadA	glutamate decarboxylase	-1.000106664	2.248336054
gadE	acid resistance transcriptional activator GadE	-0.869550493	2.286593468
glnG	nitrogen regulation protein NR(I)	-0.637711332	1.968448884
hdeA	acid-activated periplasmic chaperone HdeA	-1.461753310	2.609418121
hdeB	acid-activated periplasmic chaperone HdeB	-1.476597855	5.436637591
hdeD	acid-resistance protein HdeD	-1.065179334	1.854835852
ihfB	integration host factor subunit beta	0.643880505	1.550482101
ilvB	acetolactate synthase large subunit	-1.376916200	6.079354999
ilvG	acetolactate synthase 2 catalytic subunit	-0.820779286	1.521933233
kfiC_1	K5 polysaccharide biosynthesis glycosyltransferase KfiC	1.222479070	2.277119389
kfiD	K5 polysaccharide biosynthesis UDP-glucose dehydrogenase KfiD	1.136864125	1.558618249
lacZ	beta-galactosidase	0.826561423	2.091997799
leuA	2-isopropylmalate synthase	-1.631966415	5.540004744
leuB	3-isopropylmalate dehydrogenase	-1.050976963	6.050805226
lysP	lysine-specific permease	0.674094415	1.828455318
malQ	4-alpha-glucanotransferase	-0.664841283	1.600899841
maoP	macrodomain Ori organization protein MaoP	0.761187318	2.248336054
mdtE	multidrug transporter subunit MdtE	-0.825169433	1.410895958

mipA	scaffolding protein MipA	1.399099717	7.051587034
mnmA	tRNA 2-thiouridine(34) synthase MnmA	0.671866317	1.851451096
mtr	tryptophan permease	1.614665386	9.584025589
nagE	PTS N-acetyl glucosamine transporter subunit IIABC	0.605741148	1.841940730
narZ_1	nitrate reductase Z subunit alpha	-0.605168446	1.521933233
nrdA	class 1a ribonucleoside-diphosphate reductase subunit alpha	0.684884455	1.600899841
nuoM	NADH-quinone oxidoreductase subunit M	0.595461560	1.458773633
ompA	porin OmpA	1.413978031	4.526221165
ompC_1	porin OmpC	1.545349162	4.888401475
ompC_2	porin OmpC	1.234128298	2.602233744
ompF	porin OmpF	2.237722764	5.869666232
ompT	omptin family outer membrane protease OmpT	1.527890638	2.942714356
ompW	outer membrane protein OmpW	2.117543020	7.051587034
ompX	outer membrane protein OmpX	1.007917353	2.091997799
pagP	lipid IV(A) palmitoyltransferase PagP	-0.846656875	2.245345931
pheP	phenylalanine transporter	0.821830701	1.851194360
pheS	phenylalanine--tRNA ligase subunit alpha	1.239033789	2.006387442
pheT	phenylalanine--tRNA ligase subunit beta	1.101445715	1.600899841
potD	spermidine/putrescine ABC transporter substrate-binding protein PotD	1.180036097	2.823619308
proY	proline-specific permease ProY	0.771647771	2.705313376
queA_2	tRNA preQ1(34) S-adenosylmethionine ribosyltransferase-isomerase QueA	3.101691447	2.006387442
sad	succinate-semialdehyde dehydrogenase	0.762587091	1.484161195
sbp	sulfate/thiosulfate ABC transporter substrate-binding protein Sbp	-0.973207744	1.721354095
speD	adenosylmethionine decarboxylase	0.694709310	2.091997799

ssuB_1	aliphatic sulfonates ABC transporter ATP-binding protein	-0.873957637	2.007269796
suhB	inositol-1-monophosphatase	1.325335908	1.743679452
tdh	L-threonine 3-dehydrogenase	0.638796718	1.681292825
tehB	tellurite resistance methyltransferase TehB	0.947801919	2.233810307
thiF	thiazole biosynthesis adenylyltransferase ThiF	-0.882577025	1.504753998
thiS	sulfur carrier protein ThiS	-0.762865513	2.823619308
trhO	rhodanese-related sulfurtransferase	0.612341190	1.609210171
trpB	tryptophan synthase subunit beta	1.065474120	1.375293243
trpCF	bifunctional indole-3-glycerol-phosphate synthase TrpC/phosphoribosylanthranilate isomerase TrpF	1.526610349	2.007093182
trpD	bifunctional anthranilate synthase glutamate amidotransferase component TrpG/anthranilate phosphoribosyltransferase TrpD	1.730836179	2.332360294
trpE	anthranilate synthase component I	1.665919515	2.781464495
tuf_2	elongation factor Tu	0.882162981	1.344891015
upp	uracil phosphoribosyltransferase	0.883688050	1.506095711
uraA	uracil permease	0.929477612	1.437869468
ycjY	alpha/beta hydrolase	0.871120902	2.060830320
ydiY	YdiY family protein	1.161601548	4.421360790
yeaY	Slp family lipoprotein YeaY	0.801975595	2.823619308
ygaC	DUF2002 family protein	0.908592673	2.091997799
yhaV_3	type II toxin-antitoxin system ribonuclease toxin YhaV	0.792929680	1.900234897
yidD	membrane protein insertion efficiency factor YidD	0.819232940	2.194635093

Table 10-3 Genes with significant differential expression in EcN_EO cultured in M9 + 1 mM D-Serine + 1 mM D-Tyrosine as identified by RNA-seq

Gene name	Gene product	M9 vs. M9 + D-Serine + D-Tyrosine Log fold change	Negative log10 FDR p-value
ada_1	bifunctional DNA-binding transcriptional regulator/O6-methylguanine-DNA methyltransferase Ada	-0.669046796	1.361453954
adk	adenylate kinase	0.786185461	2.009794385
ais	lipopolysaccharide core heptose(II)-phosphate phosphatase Ais	0.899522692	2.335734200
aldB_2	aldehyde dehydrogenase AldB	-1.033930037	2.140981856
alkB_1	DNA oxidative demethylase AlkB	-0.787308694	1.545547236
ansB	L-asparaginase 2	0.687974168	1.680303737
arcA_1	arginine deiminase	-1.904419309	2.877784122
arcC_3	carbamate kinase	-1.832866394	2.252976823
argG_4	argininosuccinate synthase	0.994088811	2.335734200
aroH	3-deoxy-7-phosphoheptulonate synthase AroH	1.152839093	4.618163200
aroP	aromatic amino acid transporter AroP	1.362468250	1.819237160
artJ_1	arginine ABC transporter substrate-binding protein ArtJ	0.762239558	1.645089522
artM	arginine ABC transporter permease ArtM	0.733299185	1.751884068
asnB	asparagine synthase B	1.171024692	3.854803594
aspC	aspartate transaminase	0.944788632	1.606364177
atpC	F0F1 ATP synthase subunit epsilon	0.839682899	1.673012054
bioF	8-amino-7-oxononanoate synthase	-0.640852338	2.081497860
carA	glutamine-hydrolyzing carbamoyl-phosphate synthase small subunit	0.693602410	2.400116928
carB	carbamoyl-phosphate synthase large subunit	0.646750613	1.720197764
cbl	HTH-type transcriptional regulator Cbl	-0.896123404	1.680303737
cdh	CDP-diacylglycerol diphosphatase	-0.943550264	2.009794385
cheB	protein-glutamate methylesterase/protein glutamine deamidase	0.629769364	1.818142366
cheW	chemotaxis protein CheW	0.802175850	2.679645967

clbC	colibactin polyketide synthase ClbC	-1.314194423	2.667763585
clbD	colibactin biosynthesis dehydrogenase ClbD	-1.534871507	2.819300799
clbE	colibactin biosynthesis aminomalonyl-acyl carrier protein ClbE	-1.546212092	2.877784122
clbF	colibactin biosynthesis dehydrogenase ClbF	-1.677194256	3.368759220
clbG	colibactin biosynthesis acyltransferase ClbG	-1.800696574	4.015067834
clbH	colibactin non-ribosomal peptide synthetase ClbH	-1.971000894	4.618163200
clbI	colibactin polyketide synthase ClbI	-1.935538464	3.191586449
clbJ	colibactin non-ribosomal peptide synthetase ClbJ	-2.054596587	4.150949309
clbK	colibactin hybrid non-ribosomal peptide synthetase/type I polyketide synthase ClbK	-1.994524178	4.744245213
clbL	colibactin biosynthesis amidase ClbL	-1.967389492	4.150949309
clbM	precolibactin export MATE transporter ClbM	-1.795788583	4.861697302
clbN	colibactin non-ribosomal peptide synthetase ClbN	-1.971495051	4.763210901
clbO	colibactin polyketide synthase ClbO	-1.749790931	4.582030358
clbP	precolibactin peptidase ClbP	-1.511705565	4.015067834
clbQ	colibactin biosynthesis thioesterase ClbQ	-1.539551565	3.736363931
clbS	colibactin self-protection protein ClbS	-0.883087927	1.378160560
corA	magnesium/cobalt transporter CorA	0.634234528	1.533220931
cpsB_1	mannose-1-phosphate guanyltransferase	1.062784583	2.307505592
cpxP	cell-envelope stress modulator CpxP	-0.705470340	2.785686103
csgA	curli major subunit CsgA	-2.141421728	14.946537395
csgB	curli minor subunit CsgB	-2.299328735	13.635449005
cvpA	colicin V production protein	0.907497042	1.906514541
cydX	colicin V production protein	0.752962884	1.502181121
cysC	adenylyl-sulfate kinase	-1.099746625	1.502181121
cysS	cysteine--tRNA ligase	1.242298213	5.659952682
dcp	peptidyl-dipeptidase Dcp	0.597717268	1.348433415
dcuB	anaerobic C4-dicarboxylate transporter DcuB	1.300260791	3.141823862

dmsB_2	DMSO reductase like molybdoenzyme iron-sulfur subunit YnfG	1.527168192	3.834755674
dpaA	peptidoglycan meso-diaminopimelic acid protein amidase	0.830970821	2.009394789
dsdA	D-serine ammonia-lyase	0.884251782	2.999132278
dusC_1	tRNA dihydrouridine(16) synthase DusC	-0.981253400	1.926740357
efeO	iron uptake system protein EfeO	0.672346303	1.668254242
entA	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase EntA	0.800141755	2.374687549
entB	enterobactin biosynthesis bifunctional isochorismatase/aryl carrier protein EntB	0.906985067	1.702418245
entD	enterobactin synthase subunit EntD	0.585187724	1.430675084
entE	(2,3-dihydroxybenzoyl)adenylate synthase EntE	0.709765075	1.884158508
entH	proofreading thioesterase EntH	0.754231050	1.763380213
eptA	phosphoethanolamine transferase EptA	1.116776919	6.082494490
eptC	phosphoethanolamine transferase CptA	1.061156522	2.394048842
evgS	acid-sensing system histidine kinase EvgS	-0.789624615	2.877784122
fabI	enoyl-ACP reductase FabI	0.846681584	1.409496613
fiu	catecholate siderophore receptor Fiu	0.712281940	1.734973950
fkpA	FKBP-type peptidyl-prolyl cis-trans isomerase	0.711166070	1.517001369
fldA	flavodoxin FldA	0.955514747	3.323397830
flgB	flagellar basal body rod protein FlgB	1.679148171	2.052664324
flgC	flagellar basal body rod protein FlgC	1.753335459	1.784924925
flgD	flagellar hook assembly protein FlgD	1.050146862	2.020861530
flgE	flagellar hook protein FlgE	1.003528621	3.191586449
flgG	flagellar basal-body rod protein FlgF	1.163847951	2.964570262
flgI	flagellar basal-body rod protein FlgG	0.996717174	2.335734200
flgK	flagellar hook-associated protein FlgK	1.113420728	4.564792897
flgL	flagellar hook-associated protein FlgL	1.106415066	3.877129077

fliS	flagellar export chaperone FliS	0.615816042	2.376647318
fliZ_1	flagella biosynthesis regulatory protein FliZ	1.120310175	3.410608977
fliZ_2	flagella biosynthesis regulatory protein FliZ	0.994052513	2.575935475
fliZ_3	flagella biosynthesis regulatory protein FliZ	1.065796881	3.080712659
frdA	fumarate reductase (quinol) flavoprotein subunit	0.721039188	1.525796514
frdB_2	fumarate reductase iron-sulfur protein	0.856537737	2.480172006
frdC_1	fumarate reductase subunit FrdC	0.797476878	1.684923601
frdC_2	fumarate reductase subunit FrdC	0.954981420	2.446116973
fsr	fosmidomycin MFS transporter	1.406180804	3.965772739
ftnA	non-heme ferritin	1.124093862	2.980468315
gabD	NADP-dependent succinate-semialdehyde dehydrogenase	-0.721619300	1.422437909
gabT	4-aminobutyrate--2-oxoglutarate transaminase	-0.626508019	1.372931337
gadA	glutamate decarboxylase	-1.264992538	3.048954052
gadB	glutamate decarboxylase	-1.765561657	2.005330978
gadC	acid resistance gamma-aminobutyrate antiporter GadC	-1.435509903	1.734973950
gadE	acid resistance transcriptional activator GadE	-1.118470660	2.376647318
glmS	glutamine--fructose-6-phosphate transaminase (isomerizing)	0.847987770	1.511999470
gloA	lactoylglutathione lyase	1.094276651	4.841336019
glyA	serine hydroxymethyltransferase	0.869069971	1.346219130
gnsA	addiction module toxin, GnsA/GnsB family	0.888773958	1.595161161
gpt	xanthine phosphoribosyltransferase	1.029555938	2.727462223
greA	transcription elongation factor GreA	0.752902585	1.415957203
gspD_2	type II secretion system secretin GspD	-0.638338595	1.680303737
hdeA	acid-activated periplasmic chaperone HdeA	-1.796169185	2.546070408
hdeB	acid-activated periplasmic chaperone HdeB	-1.726695825	3.323397830
hdeD	acid-resistance protein HdeD	-1.539875996	2.856672870
hemW	radical SAM family heme chaperone HemW	0.669235886	1.519959362

hisIE	bifunctional phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP diphosphatase HisIE	1.068631855	1.420220498
hisJ	histidine ABC transporter substrate-binding protein HisJ	0.620131483	1.396287186
hupB	nucleoid-associated protein HU-beta	0.710006934	1.483946818
hybA	hydrogenase 2 operon protein HybA	1.007031198	1.680303737
ihfB	integration host factor subunit beta	1.461327144	3.666955970
ilvB	acetolactate synthase large subunit	-1.461767597	6.154220033
ilvC	ketol-acid reductoisomerase	1.768425912	4.607127255
ilvG	acetolactate synthase 2 catalytic subunit	-0.898785951	1.832589033
iss_1	increased serum survival lipoprotein Iss	1.538559881	1.727567078
iss_2	increased serum survival lipoprotein Iss	1.743793837	2.672641066
katG_1	catalase/peroxidase HPI	0.614918292	1.680303737
kefF	glutathione-regulated potassium-efflux system oxidoreductase KefF	0.824145772	1.458531696
kfiA_1	K5 polysaccharide biosynthesis alpha-UDP-GlcNAc glycosyltransferase KfiA	0.862710815	1.886586011
kfiC_1	K5 polysaccharide biosynthesis glycosyltransferase KfiC	1.550710783	4.029978734
kfiD	K5 polysaccharide biosynthesis UDP-glucose dehydrogenase KfiD	1.452064835	2.951946827
lacZ	beta-galactosidase	0.694094617	1.479776305
leuA	2-isopropylmalate synthase	-1.710403544	5.579219380
leuB	3-isopropylmalate dehydrogenase	-0.868047921	2.101000729
livJ	branched chain amino acid ABC transporter substrate-binding protein LivJ	0.801092636	1.478254866
malK	maltose/maltodextrin ABC transporter ATP-binding protein MalK	-1.457603406	1.852754679
malQ	4-alpha-glucanotransferase	-0.684499031	2.346401618
maoP	macrodomain Ori organization protein MaoP	0.824360946	3.442492798
marA	MDR efflux pump AcrAB transcriptional activator MarA	0.673105951	1.376553031
mdtE	multidrug transporter subunit MdtE	-0.939715036	2.300248968

mdtF	multidrug efflux pump RND permease MdtF	-1.039058675	2.559720787
mdtM	multidrug efflux MFS transporter MdtM	1.042562968	2.009794385
metA_3	homoserine O-succinyltransferase	-0.672434932	1.536246636
metB	cystathionine gamma-synthase	-0.760432984	3.049197177
metF	methylenetetrahydrofolate reductase	-0.800196758	1.567033592
metL	bifunctional aspartate kinase/homoserine dehydrogenase II	-0.753787306	1.734973950
mipA	scaffolding protein MipA	1.970188749	10.399571674
mnmA	tRNA 2-thiouridine(34) synthase MnmA	0.786782339	2.121133663
modF	molybdate ABC transporter ATP-binding protein ModF	0.803661383	3.271483895
mtnN	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	0.700268816	1.595161161
mtr	tryptophan permease	1.401578889	7.740406121
nadE	PTS N-acetyl glucosamine transporter subunit IIABC	0.774275654	1.720887429
nagB	glucosamine-6-phosphate deaminase	0.681618950	1.533220931
nagE	PTS N-acetyl glucosamine transporter subunit IIABC	0.942059835	3.896538378
narI_1	respiratory nitrate reductase subunit gamma	0.849136979	1.345378033
narZ_1	nitrate reductase Z subunit alpha	-0.803070925	2.595508382
ndk	nucleoside-diphosphate kinase	0.812275957	2.643782866
nfsB	oxygen-insensitive NAD(P)H nitroreductase	0.685518153	2.324221658
nrdA	class 1a ribonucleoside-diphosphate reductase subunit alpha	0.885418557	2.027426919
nuoA	NADH-quinone oxidoreductase subunit NuoA	0.650123550	2.559720787
nuoE	NADH-quinone oxidoreductase subunit NuoE	0.809917971	3.228265575
nuoF	NADH-quinone oxidoreductase subunit NuoF	0.647584177	1.681452596
nuoI	NADH-quinone oxidoreductase subunit NuoI	0.837073703	3.232918379
nuoM	NADH-quinone oxidoreductase subunit M	0.766676440	2.484391051
nuoN	NADH-quinone oxidoreductase subunit NuoN	0.628111793	1.794634284

ompA	porin OmpA	1.855948014	5.280006174
ompC_1	porin OmpC	2.404167609	7.452964100
ompC_2	porin OmpC	1.464233480	3.200590520
ompF	porin OmpF	2.808427070	8.445268623
ompT	omptin family outer membrane protease OmpT	2.279327977	7.736363931
ompW	outer membrane protein OmpW	2.710612920	10.399571674
ompX	outer membrane protein OmpX	1.330008642	2.903089987
oppF	murein tripeptide/oligopeptide ABC transporter ATP-binding protein OppF	0.594788748	1.322113768
orn	oligoribonuclease	0.620580500	1.470132733
osmB	osmotically-inducible lipoprotein OsmB	-0.917288024	4.003225729
otsB	trehalose-phosphatase	-0.620344738	1.447171173
pagP	lipid IV(A) palmitoyltransferase PagP	-1.191726237	4.150949309
pdeH	cyclic-guanylate-specific phosphodiesterase	0.821492806	1.794634284
pdxK	pyridoxine/pyridoxal/pyridoxamine kinase	0.978824875	2.101878424
pgaD	poly-beta-1,6-N-acetyl-D-glucosamine biosynthesis protein PgaD	0.876035502	1.376166712
pgl	6-phosphogluconolactonase	0.904867814	3.820448209
pheA	bifunctional chorismate mutase/prephenate dehydratase	1.190368144	3.046820399
pheP	phenylalanine transporter	1.042295367	3.854803594
pheS	phenylalanine--tRNA ligase subunit alpha	1.909077767	5.135963817
pheT	phenylalanine--tRNA ligase subunit beta	1.603899029	4.141342516
phoA	alkaline phosphatase	-0.668263223	1.702418245
pmrB	two-component system sensor histidine kinase PmrB	0.640448533	2.689094371
potD	spermidine/putrescine ABC transporter substrate-binding protein PotD	1.675259836	6.082494490
potI	putrescine ABC transporter permease PotI	-0.672764230	1.888307345
ppiA	peptidylprolyl isomerase A	0.829164905	2.121133663
ppiC	peptidylprolyl isomerase PpiC	0.688848778	1.315975175
ppnP	pyrimidine/purine nucleoside phosphorylase	0.867539436	3.459795700
priB	primosomal replication protein N	1.000374873	1.494553036

prmB	50S ribosomal protein L3 N(5)-glutamine methyltransferase	0.840403855	3.207748428
proS	proline--tRNA ligase	0.681028024	1.430675084
proY	proline-specific permease ProY	0.772423323	1.680303737
pspA	phage shock protein PspA	-0.979345230	2.347173697
pspB	envelope stress response membrane protein PspB	-0.649735127	1.415166164
pspD	phage shock protein PspD	-0.685664841	1.481341319
purE	5-(carboxyamino)imidazole ribonucleotide mutase	1.107346598	1.428150900
purH	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	0.835757207	1.521689749
purL	phosphoribosylformylglycinamidine synthase	1.015761851	2.140981856
purM	phosphoribosylformylglycinamidine cyclo-ligase	0.915306852	1.416540735
purR	HTH-type transcriptional repressor PurR	0.746898791	1.928086795
purT	formate-dependent phosphoribosylglycinamide formyltransferase	1.293953420	2.517841305
putA	trifunctional transcriptional regulator/proline dehydrogenase/L-glutamate gamma-semialdehyde dehydrogenase	0.794299924	2.077274542
queA_2	tRNA preQ1(34) S-adenosylmethionine ribosyltransferase-isomerase QueA	3.380655216	4.550059011
rcnB	Ni(II)/Co(II) efflux transporter accessory subunit RcnB	1.001258050	2.597566654
rdgB	XTP/dITP diphosphatase	0.695124631	1.474839345
ribB	3,4-dihydroxy-2-butanone-4-phosphate synthase	-0.827251079	1.474839345
rimO	30S ribosomal protein S12 methylthiotransferase RimO	0.618660638	1.315390108
rnb	exoribonuclease II	-0.926326657	1.734973950
rpmG	50S ribosomal protein L33	0.884019817	2.646083769
rpsF	30S ribosomal protein S6	0.923021090	1.312151746
rpsU	30S ribosomal protein S21	0.614399030	1.422437909
rspB	30S ribosomal protein S21	-1.271189367	1.461815744
rsuA	16S rRNA pseudouridine(516) synthase RsuA	0.710589732	1.993177110
sad	succinate-semialdehyde dehydrogenase	0.805023124	2.140981856

sbp	sulfate/thiosulfate ABC transporter substrate-binding protein Sbp	-0.885745583	1.595944883
sodB	superoxide dismutase [Fe]	0.822418435	1.794731683
speC	ornithine decarboxylase	0.662622962	1.416540735
speD	adenosylmethionine decarboxylase	0.853968476	2.009394789
speE	polyamine aminopropyltransferase	0.666284742	1.478254866
srmB	ATP-dependent RNA helicase SrmB	0.665422157	1.409496613
ssuA	aliphatic sulfonate ABC transporter substrate-binding protein SsuA	-0.868988834	2.062231433
ssuB_1	aliphatic sulfonates ABC transporter ATP-binding protein	-0.968788547	3.636388020
ssuC	aliphatic sulfonate ABC transporter permease SsuC	-1.145889179	3.570247720
ssuD	FMNH2-dependent alkanesulfonate monooxygenase	-0.968127672	1.884158508
suhB	inositol-1-monophosphatase	1.595061121	4.254144805
symE	endoribonuclease SymE	-0.642649307	1.618666608
tauA	taurine ABC transporter substrate-binding protein	-1.084872586	7.098759698
tauB	taurine ABC transporter ATP-binding subunit	-1.273113687	3.100288905
tauC	taurine ABC transporter permease TauC	-1.300695976	2.663140179
tauD	taurine dioxygenase	-1.103419207	3.765989182
tdh	L-threonine 3-dehydrogenase	0.656691507	2.005330978
tehB	tellurite resistance methyltransferase TehB	1.192054245	3.323397830
thiC	phosphomethylpyrimidine synthase ThiC	-0.844830475	1.419816106
thiE	thiamine phosphate synthase	-0.794299210	1.536246636
thiF	thiazole biosynthesis adenylyltransferase ThiF	-1.020220448	2.400116928
thiG	thiazole synthase	-0.881492524	1.702418245
thiS	sulfur carrier protein ThiS	-0.936844181	2.913640169
trhO	rhodanese-related sulfurtransferase	0.771777463	2.324221658
trpA	tryptophan synthase subunit alpha	0.999897993	1.926740357
trpB	tryptophan synthase subunit beta	1.335098514	3.323397830
trpCF	bifunctional indole-3-glycerol-phosphate synthase TrpC/phosphoribosylanthranilate isomerase TrpF	1.499128701	3.323397830
trpD	bifunctional anthranilate synthase glutamate amidotransferase	1.773479408	3.764976841

	component TrpG/anthranilate phosphoribosyltransferase TrpD		
trpE	anthranilate synthase component I	1.539932644	3.048954052
trpS	tryptophan--tRNA ligase	0.812452018	1.593701995
trxB	thioredoxin-disulfide reductase	0.668243286	1.521533023
tsr	methyl-accepting chemotaxis protein	0.672543242	1.941033718
tssG	type VI secretion system baseplate subunit TssG	-0.591310123	1.372931337
tsx	nucleoside-specific channel-forming protein Tsx	1.229509521	4.564792897
tuf_2	elongation factor Tu	1.337268046	3.228265575
tyrS	tyrosine--tRNA ligase	1.092349743	2.121133663
ubiU	peptidase U32 family protein	1.172590011	1.994992927
ung	uracil-DNA glycosylase	0.598594089	1.928086795
upp	uracil phosphoribosyltransferase	1.255236157	4.010016542
uraA	uracil permease	1.030290170	1.884158508
ushA	bifunctional UDP-sugar hydrolase/5'-nucleotidase	0.588814172	1.458531696
xthA	exodeoxyribonuclease III	0.684464451	1.430675084
yaaY	DUF2575 family protein	-0.614440220	1.587325110
yahK	NADPH-dependent aldehyde reductase YahK	-0.653976859	1.720079560
yajG	lipoprotein	0.595199759	1.521533023
ybaA	DUF1428 domain-containing protein	-1.023097103	1.680303737
ybaB	YbaB/EbfC family nucleoid-associated protein	-0.620045358	1.928086795
ybcI	metal-dependent hydrolase	1.074474943	1.582072369
ybhA	bifunctional pyridoxal phosphate/fructose-1,6-bisphosphate phosphatase	0.921184695	1.926740357
ybtE	yersiniabactin biosynthesis salycil-AMP ligase YbtE	-1.095507996	1.595161161
ybtT	yersiniabactin biosynthesis thioesterase YbtT	-1.101877510	1.538264430
ybtU	yersiniabactin biosynthesis oxidoreductase YbtU	-1.142807305	1.686826136
ycgR	flagellar brake protein	0.645833641	1.668254242
yciI	YciI family protein	0.600362165	2.009794385
yciW	CMD domain-containing protein	-1.207365050	2.877784122
ycjY	alpha/beta hydrolase	1.214130315	3.365019200
yddA	ABC transporter ATP-binding protein/permease	-0.760006816	2.346401618
ydgT	transcription modulator YdgT	1.049710970	2.009794385
ydiY	YdiY family protein	1.496700900	4.618163200
yebV	YebV family protein	-1.000987640	3.965772739

yecJ	DUF2766 family protein YecJ	0.802305689	2.404614019
yecU	protein YecU	-0.913896524	1.534164405
yedF	sulfurtransferase-like selenium metabolism protein YedF	1.170735204	2.347173697
yegQ	tRNA 5-hydroxyuridine modification protein YegQ	0.984446740	2.209081805
yeiP	elongation factor P-like protein YeiP	1.130414677	3.703334810
yejK	nucleoid-associated protein YejK	0.762059515	1.512435195
yfdX	YfdX family protein	0.909064500	1.325858352
ygaC	DUF2002 family protein	1.054268952	3.541664374
ygbE	DUF3561 family protein	-1.173653366	2.480172006
ygiQ	YgiQ family radical SAM protein	0.650572267	1.416540735
yhaV_3	type II toxin-antitoxin system ribonuclease toxin YhaV	0.690260596	1.706686642
yhaV_4	type II toxin-antitoxin system ribonuclease toxin YhaV	-2.075780719	2.130240405
yidD	membrane protein insertion efficiency factor YidD	0.813517604	1.595161161
yieE_1	DNA-binding transcriptional regulator YeiE	0.790431566	1.680303737
yifK	amino acid permease	1.037768210	2.437588167
yihD	YihD family protein	0.855405531	1.402920185
yiiX	YiiX family permuted papain-like enzyme	-0.908966509	2.209081805
yijD	YijD family membrane protein	0.626338394	1.315390108
yjbE	exopolysaccharide production protein YjbE	-0.756925686	2.010550182
yjbH	YjbH domain-containing protein	-0.742396494	1.512435195
yjfY	DUF1471 family protein YjfY	-0.797778787	1.338188850
yjiM	double-cubane-cluster-containing anaerobic reductase	0.861255763	2.319483191
yjiX	YbdD/YjiX family protein	-0.871289527	1.702418245
ylaC	YlaC family protein	0.832550422	2.209785230
ynjB	ABC transporter substrate-binding protein	0.585127004	1.355433225
ynjH	YnjH family protein	0.792420646	2.877784122
yoaC	DUF1889 family protein	1.000587877	2.546070408
yodC	DUF2158 domain-containing protein	-0.876564252	1.419816106
yohF	SDR family oxidoreductase	-0.617526216	1.325858352
yrbL	PhoP regulatory network protein YrbL	-0.598198545	1.994992927
zntA	Zn(II)/Cd(II)/Pb(II) translocating P-type ATPase ZntA	0.968760385	1.680303737

Figure 10-3 Summary of the workflow used in Galaxy Australia. The tool ‘snippy’ was used to identify nucleotide mutations compared to the EcN reference genome (NCBI accession number NZ_CP007799.1), and the tools ‘SPAdes’, ‘Prokka’ and ‘JBrowse’ were used to respectively assemble, annotate, and visualise the sequenced genomes.

Table 10-4 Summary table of the variants detected in EcN_AR

Chromosome	Position	Type	Reference	Alternative	Evidence	Strand	Nucleotide position	AA position	Effect	Locus tag	Gene	Product
CP058217	1450549	snp	T	C	C:36 T:0	+	276/ 834	92/277	synonymous_variant c.276T> C p.Gly92Gly	HW372_ 06875	*cysT*	sulfate/thiosulfate ABC transporter permease CysT
CP058217	2594737	snp	C	T	T:44 C:0	-	637/ 756	213/251	missense_variant c.637G> A p.Asp213Asn	HW372_ 12215		hypothetical protein
CP058217	3761718	snp	C	T	T:2836 C:59	+	16/ 1083	6/360	synonymous_variant c.16C> T p.Leu6Leu	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3761784	snp	T	A	A:5236 T:48	+	82/ 1083	28/360	missense_variant c.82T> A p.Cys28Ser	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3761970	snp	T	C	C:6322 T:35	+	268/ 1083	90/360	synonymous_variant c.268T> C p.Leu90Leu	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3761999	snp	T	A	A:5834 T:39	+	297/ 1083	99/360	synonymous_variant c.297T> A p.Val99Val	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762065	snp	A	G	G:6284 A:39	+	363/ 1083	121/360	synonymous_variant c.363A> G p.Gly121Gly	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762123	snp	G	A	A:6989 G:42	+	421/ 1083	141/360	missense_variant c.421G> A p.Ala141Thr	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762164	snp	T	A	A:7254 T:44	+	462/ 1083	154/360	synonymous_variant c.462T> A p.Thr154Thr	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762221	snp	C	T	T:7686 C:51	+	519/ 1083	173/360	synonymous_variant c.519C> T p.His173His	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762260	snp	A	G	G:7772 A:46	+	558/ 1083	186/360	synonymous_variant c.558A> G p.Ala186Ala	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762299	snp	G	T	T:8098 G:53	+	597/ 1083	199/360	synonymous_variant c.597G> T p.Ala199Ala	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762320	snp	A	T	T:7960 A:51	+	618/ 1083	206/360	synonymous_variant c.618A> T p.Thr206Thr	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762342	complex	ACA	GCG	GCG:7583 ACA:48	+	640/ 1083	214/360	missense_variant c.640_642delACAinsGCG p.Thr214Ala	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762401	snp	A	G	G:7285 A:45	+	699/ 1083	233/360	synonymous_variant c.699A> G p.Leu233Leu	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762452	snp	A	G	G:6292 A:40	+	750/ 1083	250/360	synonymous_variant c.750A> G p.Ala250Ala	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762542	snp	G	A	A:5304 G:36	+	840/ 1083	280/360	synonymous_variant c.840G> A p.Ser280Ser	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762692	mnp	TC	GG	GG:2776 TC:41	+	990/ 1083	330/360	missense_variant c.990_991delTCinsGG p.Pro331Ala	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI
CP058217	3762758	snp	A	G	G:1399 A:51	+	1056/ 1083	352/360	synonymous_variant c.1056A> G p.Gln352Gln	HW372_ 17935	*lacI*	DNA-binding transcriptional repressor LacI

Table 10-5 Summary table of the variants detected in EcN_EO

Chromosome	Product	Type	Reference	Alternative	Evidence	Strand	Nucleotide position	AA position	Effect	Locus tag	Gene	Product
CP058217	430534	snp	A	G	G:53 A:0	+	128/ 375	43/124	missense_variant c.128A>G p.Lys43Arg	HW372_ 01980	*rpsL*	30S ribosomal protein S12
CP058217	570946	snp	T	C	C:42 T:0	+	578/ 966	193/321	missense_variant c.578T>C p.Leu193Ser	HW372_ 02765		DMT family transporter
CP058217	2594737	snp	C	T	T:62 C:0	-	657/ 756	213/251	missense_variant c.637G>A p.Asp213Asn	HW372_ 12215		hypothetical protein

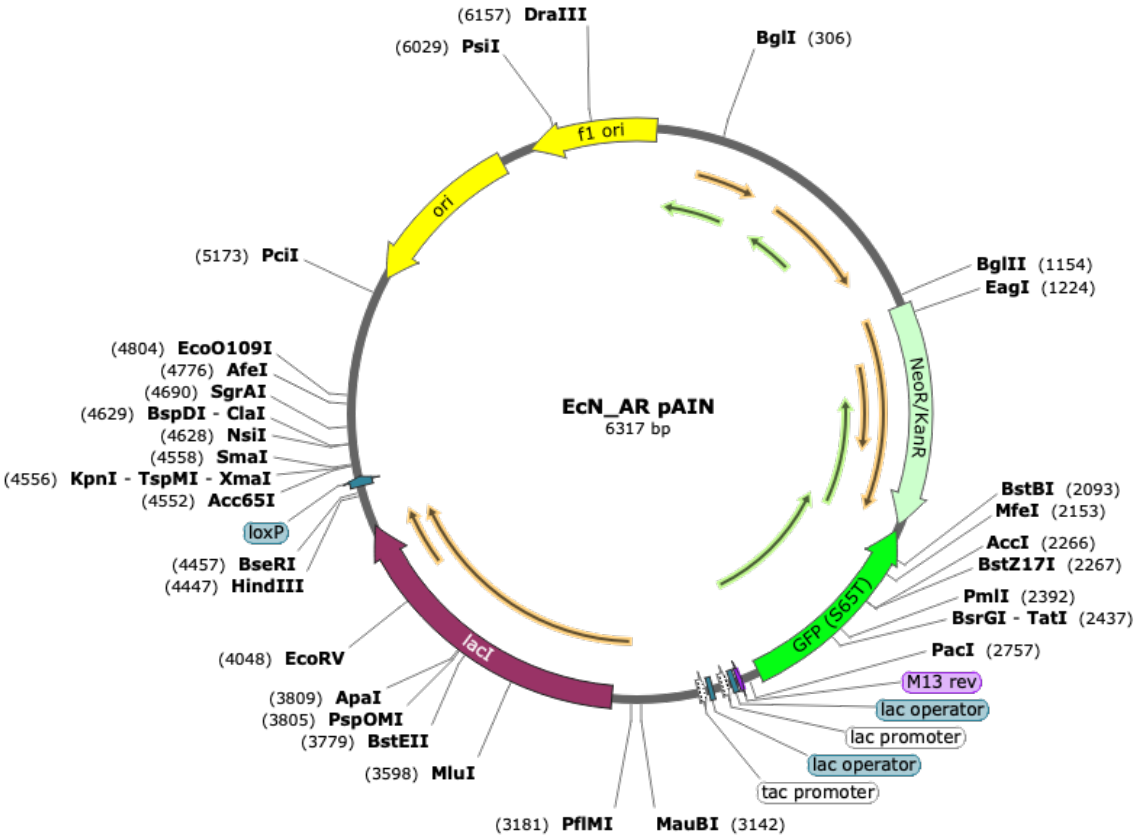


Figure 10-4 Map of the plasmid extracted from EcN_AR conferring kanamycin resistance. EcN_AR was found to exhibit kanamycin resistance and hypothesised to unexpectedly carry a kanamycin-resistant plasmid. Plasmid purification and sequencing revealed that EcN_AR indeed carried a kanamycin-resistant plasmid in which GFP expression had been placed under control of a *lac* promoter, and which also carried a *lacI* sequence.

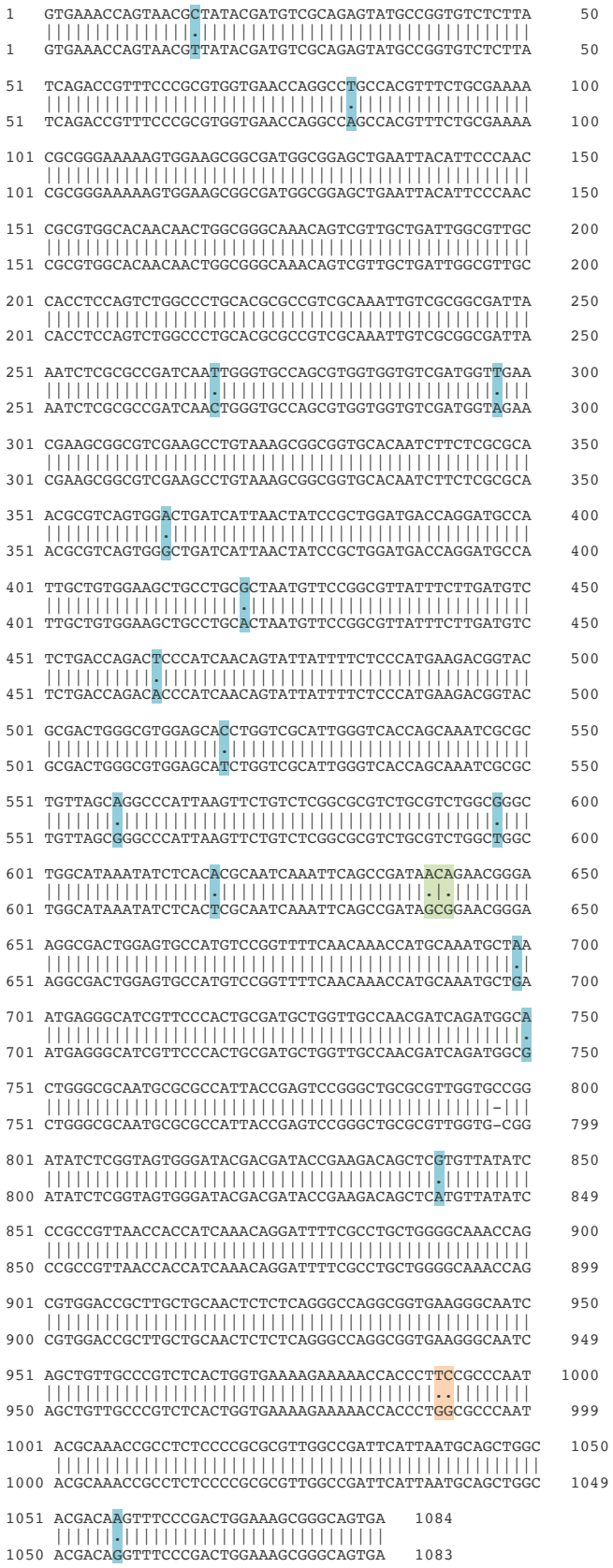


Figure 10-5 Alignment of the *lacI* gene in the EcN reference genome and of the *lacI* sequence on the EcN_AR pAIN plasmid. Alignment of the plasmid's *lacI* region with that of EcN's reference genome revealed all the previously-identified nucleotide variants, matching those represented in Figure 5-6B. SNPs are highlighted in blue, complex mutations in green, and mnps in orange (Chapter 5).

Mouse histology grading scheme

Inflammation (scoring over entire roll)

A. Distribution (to multiply individual scores with)

Absent – 0
Focal – 1
Multifocal – 2
Multifocal to coalescing – 3
Diffuse – 4

B. Parameters (each with severity scoring)

a. Epithelial changes:
Normal - 0
Mild epithelial injury – 1
Moderate epithelial injury – 2
Marked/severe epithelial injury – 3

b. **Inflammatory infiltrates**
Normal – 0
Mild increase – 1
Moderate increase – 2
Marked increase – 3

c. **Epithelial infiltrates by neutrophils**:
None – 0
Intra-epithelial – 1
Small numbers in crypts – 2
Crypt abscesses – 3

d. **Crypt hyperplasia/goblet cell depletion**:
Normal – 0
Mild – 1
Moderate – 2
Marked – 3

e. **Distorted crypt architecture**:
Normal – 0
Mild – 1
Moderate – 2
Marked - 3

f. **Crypt mitotic activity (has not distribution parameter)**:
Lower third – 0
Mild mid third – 1
Moderate mid third – 2
Upper third - 3

g. **Mucosal fibrosis and atrophy**:
Normal – 0
Mild - 1
Moderate – 2
Marked – 3

Neoplasia (according to mouse consensus; scoring of most severe lesion)

Process
GIN (dysplasia, hyperplasia, adenoma – < 1 mm grossly) – 1
Hyperplasia – 1
GIN/adenoma (just about grossly appreciable mass) – 1.5
Adenoma arising in hyperplasia - 1.5
Adenoma (benign circumscribed neoplasm) - low grade - 2
Adenoma - early high grade – 2.5
Adenoma – high grade - 3
(Early) Carcinoma – 4
Carcinoma (invasion through muscularis mucosae) – 5

11 Publications

microbial cell

Research Article

www.microbialcell.com

D-Serine reduces the expression of the cytopathic genotoxin colibactin

Jennifer C. Hallam[1,#], Sofia Sandalli[1,#], Iris Floria[1], Natasha C. A. Turner[1], Min Tang-Fichaux[2], Eric Oswald[2,3], Nicky O'Boyle[1,4,*] and Andrew J. Roe[1,*]

[1] School of Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, United Kingdom.
[2] IRSD, INSERM, INRAE, Université de Toulouse, ENVT, Toulouse, France.
[3] CHU Toulouse, Hôpital Purpan, Service de Bactériologie-Hygiène, Toulouse, France.
[4] School of Microbiology, University College Cork, National University of Ireland, Cork, Ireland.
[#] JCH and SS contributed equally to this work.
* Corresponding Authors:
Andrew J. Roe, School of Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, United Kingdom; E-mail: Andrew.roe@glasgow.ac.uk
Nicky O'Boyle, School of Microbiology, University College Cork, National University of Ireland, Cork, Ireland; E-mail: NOboyle@ucc.ie

ABSTRACT **Some *Escherichia coli* strains harbour the *pks* island, a 54 kb genomic island encoding the biosynthesis genes for a genotoxic compound named colibactin. In eukaryotic cells, colibactin can induce DNA damage, cell cycle arrest and chromosomal instability. Production of colibactin has been implicated in the development of colorectal cancer (CRC). In this study, we demonstrate the inhibitory effect of D-Serine on the expression of the *pks* island in both prototypic and clinically-associated colibactin-producing strains and determine the implications for cytopathic effects on host cells. We also tested a comprehensive panel of proteinogenic L-amino acids and corresponding D-enantiomers for their ability to modulate *clbB* transcription. Whilst several D-amino acids exhibited the ability to inhibit expression of *clbB*, D-Serine exerted the strongest repressing activity (>3.8-fold) and thus, we focussed additional experiments on D-Serine. To investigate the cellular effect, we investigated if repression of colibactin by D-Serine could reduce the cytopathic responses normally observed during infection of HeLa cells with *pks*$^+$ strains. Levels of γ-H2AX (a marker of DNA double strand breaks) were reduced 2.75-fold in cells infected with D-Serine treatment. Moreover, exposure of *pks*$^+$ *E. coli* to D-Serine during infection caused a reduction in cellular senescence that was observable at 72 h post infection. The recent finding of an association between *pks*-carrying commensal *E. coli* and CRC, highlights the necessity for the development of colibactin targeting therapeutics. Here we show that D-Serine can reduce expression of colibactin, and inhibit downstream cellular cytopathy, illuminating its potential to prevent colibactin-associated disease.**

doi: *10.15698/mic2023.03.793*
Received *originally: 15.12.2022;*
in revised form: 08.02.2023,
Accepted *09.02.2023,*
Published *06.03.2023.*

Keywords*: E. coli, colibactin, expression, genotoxin, cell-cycle.*

Abbreviations:
CRC – colorectal cancer,
DSB – double strand break,
EHEC – enterohaemorrhagic E. coli,
ICL – interstrand cross-link,
LEE - locus of enterocyte effacement,
UPEC - uropathogenic E. coli.

INTRODUCTION

The gastrointestinal tract is a complex environment, where the composition and abundance of metabolites and signalling molecules can be affected by the host diet and physiology, the microbiota, and invading pathogens [1]. How bacteria sense and respond to these signals is crucial for their survival and successful colonization within the host. Most of the colonizers are commensal organisms which live mutualistically, however, certain pathogenic strains can outcompete resident bacteria by utilizing host metabolites. Indeed, competition in this environment goes beyond contending for food, as metabolites have also been shown to influence gene expression and drive evolutionary change amongst bacteria residing in the gut [2].

The composition of the human diet greatly influences the availability of metabolites within the gastrointestinal environment, as many indigestible components provide substrates for the resident microflora. An often-overlooked group of metabolites are the D-amino acids, of which the average dietary consumption is estimated at 100 mg/d [3].

D-amino acids are found naturally in foods such as fruits and vegetables [4], however, higher concentrations are more commonly associated with fermented and processed foods, including aged cheeses and breakfast cereals [3]. D-amino acids are also an important source of energy for the resident microflora. Indeed, bacteria can utilize these for essential processes including supporting growth, regulating spore germination, and in cell wall synthesis, where D-Alanine and D-Glutamate are routinely found as components of the peptidoglycan sacculus [5]. Furthermore, D-amino acids have been found to have a profound effect on gene expression and studies have shown they can manipulate the expression of virulence genes in diverse bacteria [6–8]. The concentrations of these D-amino acids are often dependent on the site within the host. For example, D-Serine is found at 1 µM in the gut and at 1 mM in the urine [9], a substantial difference that we believe could play a role in regulating niche-specificity [10]. In order to metabolize these molecules, bacteria must possess specialized catabolic enzymes [11–16]. In the case of D-Serine, certain bacteria encode a specialized tolerance locus, enabling them to exploit D-Serine as a carbon source, facilitating colonization at nutrient deficient extraintestinal sites. The uropathogenic *Escherichia coli* (UPEC), commonly possess this capability. UPEC strains can traverse through the gastrointestinal tract and colonize the bladder, where carbon sources are scarce, but D-Serine concentrations are high [17, 18]. UPEC, and many other *E. coli* that belong to the B2 phylogroup, commonly encode a D-Serine tolerance locus, *dsdCXA*. This locus encodes a D-Serine deaminase (DsdA), a D-Serine inner membrane transporter (DsdX) and an essential LysR-type transcriptional regulator (LTTR) (DsdC) that regulates the system [16]. Indeed, the role of *dsdCXA* is to prevent UPEC strains from succumbing to inhibitory concentrations of D-Serine in the urinary tract, by converting the substrate to ammonia and pyruvate [9, 16, 19]. While there have been conflicting reports surrounding the role of D-Serine in regulating virulence in UPEC [9, 20], recent transcriptome analysis has demonstrated that D-Serine exhibits the ability to modulate the expression of genes beyond the *dsdCXA* locus in both *dsd*- and *dsd*+ strains [21].

Recently, we described the transcriptional response to D-Serine in strains from three distinct *E. coli* pathotypes; enterohaemorrhagic *E. coli* (EHEC), neonatal meningitis associated *E. coli* (NMEC) and UPEC [21]. Strikingly, the results revealed a unique transcriptional profile for each pathotype with not a single differentially expressed gene shared between all three strains. Interestingly, the transcriptional response in UPEC highlighted differential expression in a cluster of genes involved in a non-ribosomal peptide synthetase pathway. This peptide synthetase pathway, termed the *pks* island, was first described in 2006, where *E. coli* encoding this 54 kb genomic island were implicated with inflicting a genotoxic insult on eukaryotic cells [22]. The *pks* island is comprised of 19 genes which encode the machinery for biosynthesis and transportation of colibactin, the peptide-polyketide hybrid compound responsible for exerting genotoxic activity [23]. The *pks* island was first identified in extraintestinal *E. coli* (ExPEC) strains, with colibactin described as a *bona fide* virulence factor in several studies [24–27]. Carriage of the *pks* island has also been described in around 34% of commensal *E. coli* strains belonging to the B2 phylogenetic group [23, 27, 28], and in other members of the *Enterobacteriaceae* including *Citrobacter koseri*, *Klebsiella pneumoniae* and *Enterobacter aerogenes* [23].

Recently, there has been a dramatic increase in interest in colibactin research, largely due to the identification of an association between colibactin activity and the development of colorectal cancer (CRC). This association arises from the ability of colibactin to cause DNA double-strand breaks (DSBs), DNA crosslinks and chromosome instability in eukaryotic cells [29–32]. *E. coli* harbouring the *pks* island have been isolated from biopsy specimens of CRC patients and have been found to be capable of establishing persistent colonization, inducing inflammation, and triggering tumour growth in these tissues [33–35]. Indeed, recent whole genome analysis of *pks*$^+$ infected organoids revealed a distinct mutational signature in adenine rich motifs. This mutational signature was also observed in CRC tissues, and it was confirmed that the mutations were the result of colibactin exposure [36, 37]. CRC is the third most frequently diagnosed cancer [38] and attributes to around 610,000 deaths per year worldwide [35]. Therefore, colibactin-producing *E. coli* represent an urgent public health matter. Here, we compared the effects of a comprehensive panel of naturally-occurring amino acids – comprising the 20 proteinogenic L-amino acids and their corresponding D-enantiomers – on the expression of colibactin synthesis genes. Consistent with our recent study [21], we found that exposure to D-Serine induced a significant downregulation of the *pks* gene cluster. Further, we tested the effects of D-Serine during infection of HeLa cells and found that treatment of *pks* harbouring *E. coli* resulted in a dampening of the genotoxic effect exerted upon eukaryotic cells. Considering this research, we propose D-Serine has potential to be used as a novel therapeutic to control expression of colibactin in *pks*-encoding *E. coli* strains.

RESULTS

Colibactin biosynthesis is downregulated in response to D-serine

The host metabolite D-Serine has been shown to selectively affect the expression of virulence factors in *E. coli* pathotypes that do not possess the intact *dsdCXA* locus [10]. However, little is known of the effects on gene expression in pathotypes found in D-Serine rich environments such as the bladder. Therefore, this prompted an investigation into how exposure to D-Serine implicated gene expression in *E. coli* that encoded the complete *dsdCXA* locus. Previous work examining the transcriptome of CFT073 exposed to D-Serine revealed many significant differentially expressed genes throughout the genome, but of particular interest was the downregulation of several genes encoding proteins involved in the synthesis of the genotoxin colibactin [21]. Comparison of read data from D-Serine treated (red)

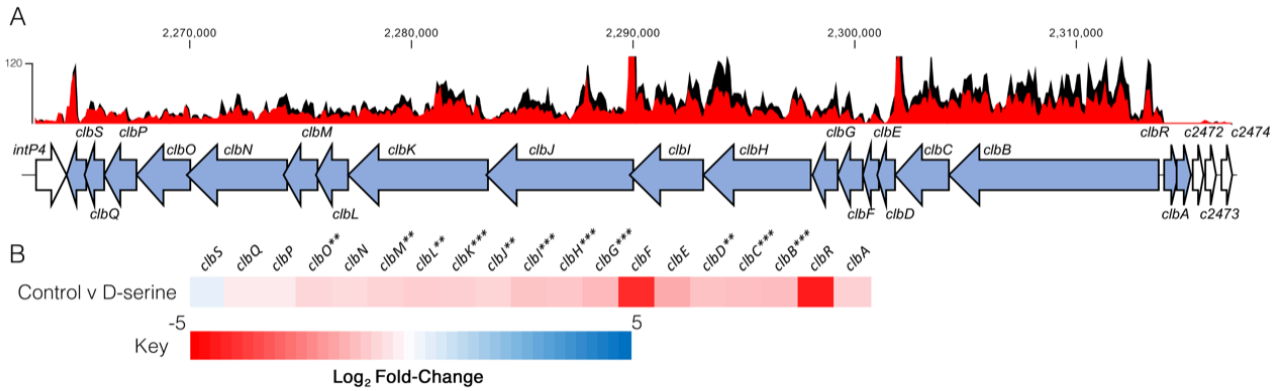


FIGURE 1: Transcriptome analysis of CFT073 reveals downregulation of the colibactin biosynthesis operon in response to D-Serine. Using read data from our previous transcriptomic study [21], we constructed track maps of the *clb* locus in CFT073 in the presence and absence of D-Serine. (**A**) Read density in the colibactin biosynthesis locus for representative samples of the untreated control (black) and D-Serine treated (red) CFT073 is indicated. Read tracks were normalized, exported from CLC Genomics Workbench and overlayed. Genomic coordinates are displayed above the read tracks and the corresponding genes within the colibactin biosynthesis operon (blue) beneath. (**B**) Heat map indicating the EdgeR calculated $\log_2$ relative fold changes for each gene in the colibactin biosynthesis locus with corresponding colour key beneath. False discovery rate-corrected *P* values are indicated with significantly differentially expressed genes with *, ** and *** indicating $P < 0.05$, 0.01 and 0.001, respectively**.**

and untreated control (black) CFT073, indicate that treatment results in a downshift in expression of several colibactin synthesis genes (**Fig. 1A**). The $\log_2$-transformed fold changes observed are indicated in **Fig. 1B**. The most significant reductions were identified for genes *clbB, I, G, C, H* and *K* respectively. Full details for fold changes and *P*-values can be found in Table S1. Interestingly, these genes belong to four (*clbB*, *clbC-G*, *clbH* and *clbI-N*) of seven putative transcriptional units of the *clb* locus [39] and encode enzymes essential for production of precolibactin, the precursor to cytotoxic colibactin [40], suggesting D-Serine can interfere with the manufacture of active colibactin in CFT073.

Expression of the genotoxin colibactin is affected by D-and L-amino acids in both CFT073 and Nissle 1917

We next investigated if the downregulation of the colibactin genes was a unique property of D-Serine, by comparing a comprehensive panel of L- and D-amino acids under the same growth conditions. Real-time quantitative polymerase chain reaction (RT-qPCR) was used to determine expression of the *pks* encoded gene *clbB*; chosen as it encodes a hybrid non-ribosomal peptide synthetase/type I polyketide synthase (NRP-PKS) that is integral to colibactin synthesis [41]. Exposing CFT073 to L-Aspartic acid, L-Isoleucine, and L-Selenocysteine led to the most significant decrease in *clbB* expression; reductions of 7.52, 8.84, and 6.17-fold were observed, respectively (**Fig. 2A**). In addition, expression of *clbB* was significantly reduced in nine of 20 D-amino acids tested, with the most significant observations recorded for expression of *clbB* in D-Cysteine (0.11) and D-Serine (0.32), $P = 0.009$ and 0.006, respectively (**Fig. 2B**). Interestingly, of the seven L-amino acids showing a significant reduction, Isoleucine was the only amino acid to show repression of *clbB* with both L- and D-enantiomers, highlighting the distinction in responses to L- and D-amino acids. Next, we selected some of the amino acids which displayed a significant effect on *clbB* expression in CFT073, and tested them with the commensal *E. coli* strain, Nissle 1917. This strain was chosen based on its commensal origin and clinical significance as a probiotic [42, 43]. Furthermore, it has successfully been used in the characterization of the colibactin-associated phenotype observed in HeLa cells and recently the genotoxicity of Nissle 1917 was demonstrated in the gut lumen of mice [44, 45], while CFT073 is precluded from such analysis as the cytopathic effects of colibactin are often confounded by haemolysin activity [46]. In response to the L- and D- amino acids tested (**Fig. 2C**), significant downregulation of *clbB* in Nissle 1917 was observed for seven of 17 amino acids tested. Interestingly, L-Selenocysteine yielded the largest decrease, with *clbB* expression reduced 5.64-fold compared to the control ($P = 0.029$). However, it should be noted that growth of both CFT073 and Nissle 1917 was strongly inhibited in the presence of L-Selenocysteine (Fig. S1), likely accompanied by stress-associated transcriptomic perturbations that may have indirectly altered *clbB* expression that is not observed with other amino acids. The second most significant decrease was observed for D-Serine where a 3.81-fold reduction was observed, ($P = 0.036$). Relative expression and *P*-values for all amino acids tested can be found in Table S2 for CFT073, and S3 for Nissle 1917. Our RT-qPCR results corroborated the findings of our previous RNA-Seq study [21] and confirmed that D-Serine repressed expression of the colibactin synthesis genes in CFT073.

Expression of *clbB* was further assessed by using a *gfp* reporter assay on bacteria grown in MEM-HEPES media. This was important to test because subsequent assays assessing DNA damage mediated by colibactin in eukaryotic cells would require growth in tissue culture media. Expression of *clbB* reduced by 2.00 and 1.75-fold in CFT073 and Nissle 1917 respectively (**Fig. 2D**), when cultured in MEM-

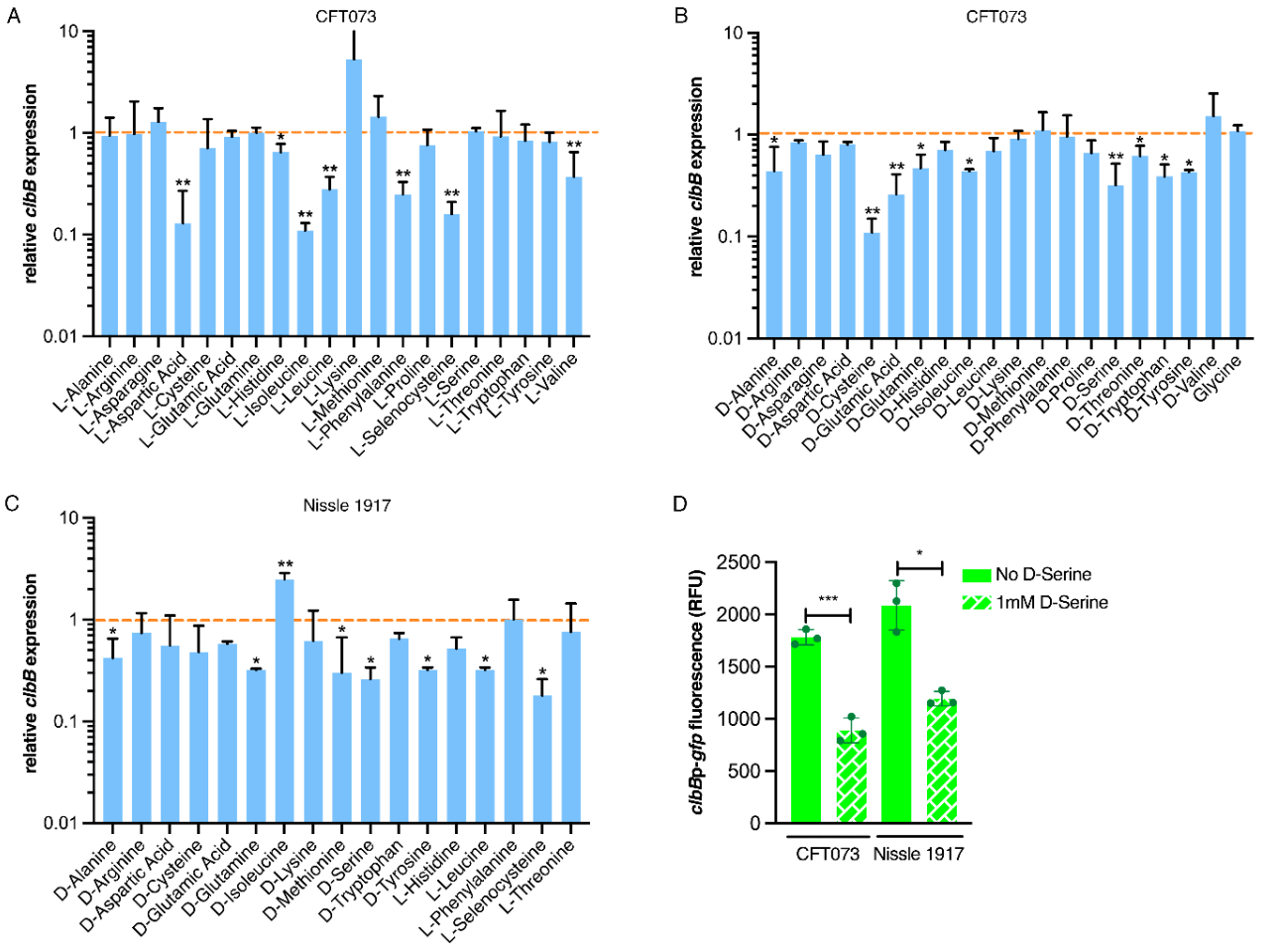


FIGURE 2: Expression of colibactin is modulated by L- and D-amino acids. Relative *clbB* expression was measured by RT-qPCR. CFT073 was grown in M9 minimal media supplemented with amino acids to a final concentration of 1 mM for 5 h. (**A**) Shows relative expression in the presence of L-amino acids and (**B**) indicates expression in the presence of D-amino acids. The orange dashed line indicates baseline expression with bars above and below this line representing up and downregulation, respectively. Statistical significance was determined from three biological replicates using an unpaired Student's *t*-test with, * and ** indicating significance, $P < 0.05$ and 0.01, respectively. (**C**) Amino acids significantly affecting *clbB* expression in CFT073 were tested in Nissle 1917 under the same growth conditions as discussed above. (**D**) *clbBp*:gfp reporter activity in MEM-HEPES in the presence and absence of 1 mM D-serine. Bacteria were sampled at 4 h post-inoculation, *clbB* expression was measured as GFP/OD_{600}. Columns represent mean +/- standard error of the means (SEM) with individual experimental observations indicated by data points. Statistical significance was assessed using a paired Student's *t*-test with *, ***, denoting $P < 0.05$ and 0.0001, respectively.

HEPES. Moreover, to test a wider panel of clinically relevant strains, we used a group of nine *E. coli* phylogroup B2 isolates obtained from bloodstream infections of Scottish patients [47]. Genome data was accessed from the European Nucleotide Archive (PRJEB12513) to demonstrate the nine isolates carried complete *pks* clusters and expression of colibactin was assessed using the *clbBp*:gfp reporter. The assay showed significant (P <0.05) repression of *clbB* transcription when 1 mM D-serine was added to the media in every isolate tested (Fig. S2). These data, showing repression of colibactin by D-serine in both prototypic and clinical isolates, suggest a common mechanism underpinning this phenotype that is widely held across *pks* carrying *E. coli.*

D-Serine-induced repression of colibactin reduces DNA damage

Colibactin exerts a genotoxic effect in infected eukaryotic cells through the formation of DNA interstrand cross-links (ICLs) and DSBs [29]. As exposure to D-Serine caused a downregulation in the colibactin synthesis genes, we hypothesized that D-Serine would be capable of limiting the formation of DNA ICLs. To investigate this, we exposed linear DNA to live *pks*$^+$ *E. coli* cultured in M9 minimal media alone or supplemented with 1 mM D-Serine and assessed colibactin-associated cross-linking activity. Upon exposure of linear DNA to Nissle 1917 and CFT073, strong cross-linking activity was apparent, as observed by an increase in molecular weight consistent with DNA duplex formation. However, upon the addition of D-Serine, a marked reduc-

tion in cross-linking activity was observed (**Fig. 3A**). Quantification of DNA ICLs by densitometry revealed that the addition of D-Serine reduced the cross-linking of DNA by 2.83 and 1.39-fold, by Nissle 1917 and CFT073, respectively (**Fig. 3B**). The formation of colibactin-associated ICLs in eukaryotic cells activates the ICL repair response pathway and results in the production of ICL-dependent DNA DSBs. Therefore, as exposure to D-Serine reduced cross-linking activity in linear DNA, we hypothesized that treatment with D-Serine could also lead to a reduction of DNA DSBs in HeLa cells infected with *pks*$^{+}$ *E. coli*. To examine DSBs, we infected HeLa cells with Nissle 1917 and measured levels of γ-H2AX, a variant of the histone family H2A, that becomes phosphorylated in response to DNA DSBs. Cell lysates were extracted 4 h post-infection and detection of γ-H2AX was determined by immunoblotting (**Fig. 3C**). To quantify the level of H2AX phosphorylation, normalized γ-H2AX signal intensities from three experiments were compared (**Fig. 3D**). Nissle 1917 infected cells exhibited a 6.41-fold (P = 0.006) increase in normalized γ-H2AX signal compared with uninfected cells. Inclusion of D-Serine during infection with Nissle 1917 resulted in a 2.75-fold reduction in γ-H2AX

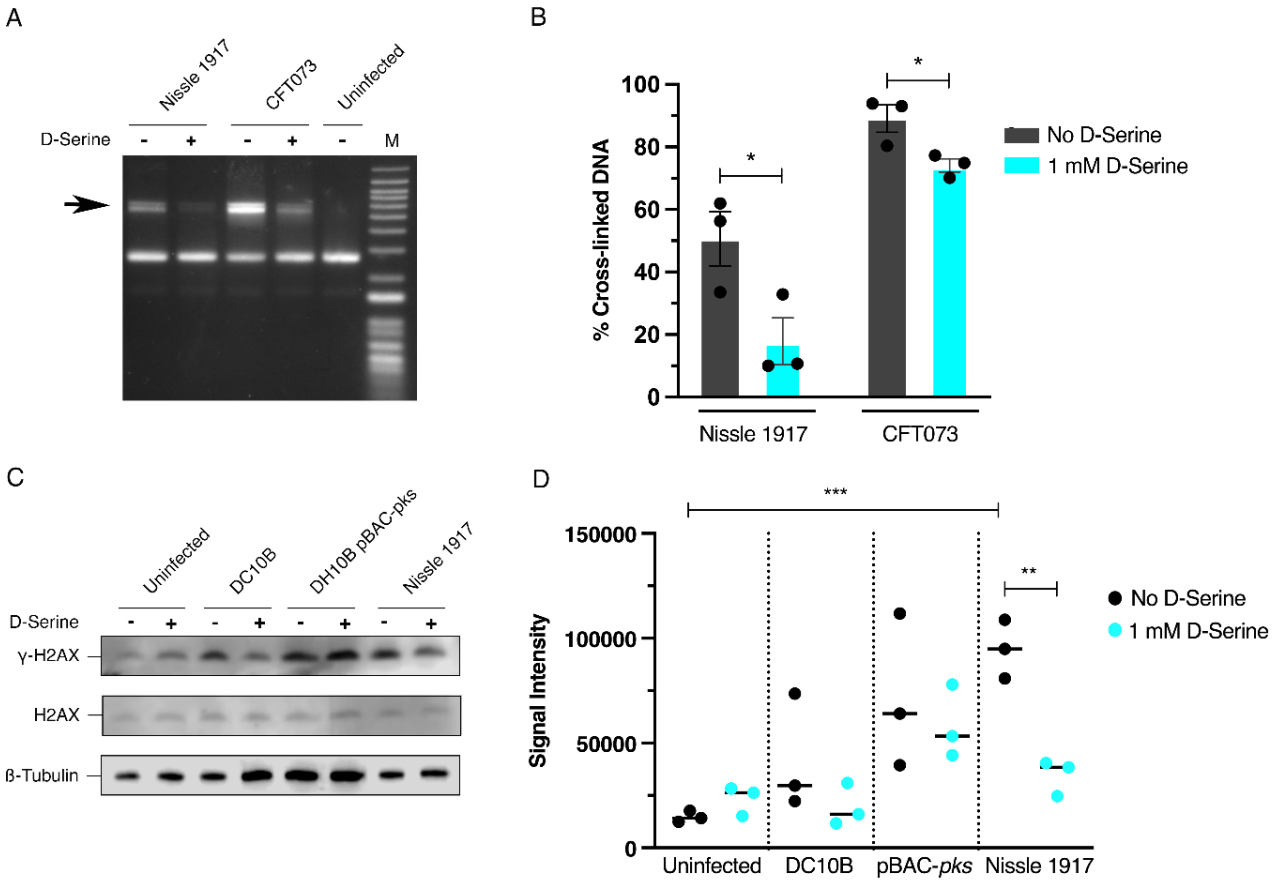


FIGURE 3: Repression of colibactin by D-Serine reduces cross-link formation, and DNA damage in Hela cell infection. (**A and B**) Nissle 1917 and CFT073 were cultured for 5 h in M9 minimal media alone (-) or in media supplemented with 1 mM D-Serine (+) before 1.5 × 10^6 CFU was exposed to linearized plasmid DNA for 40 min. The DNA was extracted, and cross-linking activity was determined by electrophoresis in denaturing conditions. (**A**) DNA cross-link formation of linearized plasmid DNA exposed to Nissle 1917 and CFT073 was visualized after migration under alkaline denaturing conditions. M, DNA size marker (1 kB plus DNA ladder, Invitrogen). (**B**) The percentage of the DNA signal in the upper, cross-linked band (indicated by the arrow in panel A) relative to the total DNA signal. Signal intensities were quantified using ImageJ for three independent experiments and statistical significance was assessed by unpaired Student's *t*-test with, * indicating P = < 0.01. (**C and D**) HeLa cells were infected for 4 h with live *pks*$^{+}$ and *pks*$^{-}$ *E. coli* with a multiplicity of infection (MOI) of 400 bacteria per cell or left uninfected. Infections were performed in wells containing MEM-HEPES alone (-) or with media supplemented with 1 mM D-Serine (+). (**C**) Immunoblot analysis of cell lysates extracted 4 h post infection. Phosphorylated histone (γ-H2AX) was used as an indicator of double stranded DNA breaks and total histone (H2AX) was used as an internal control. B-Tubulin was used as a loading control for cell lysates. DH10B pBAC-*pks* and DC10B were used as positive and negative controls, respectively. (**D**) Signal intensities of bands were measured using LI-COR Image Studio software. γ-H2AX signals were corrected to account for any variation in loading using β-Tubulin signal intensity. Experimental signal was normalized so that the mean signal intensity of the eight samples was equivalent for each experiment. The experiment was carried out in triplicate. Columns represent mean +/- SEM with individual experimental observations indicated by data points. Statistical significance was assessed by unpaired Student's *t*-test with, ** and *** indicating P < 0.01 and 0.001, respectively.

signal intensity compared to cells infected with untreated Nissle 1917 (P = 0.003). Interestingly, D-Serine did not reduce γ-H2AX signal intensity upon infection with DH10B pBAC-*pks*, suggesting that D-Serine activity requires factors specific to the natural colibactin-carrying isolates described here. To confirm that D-Serine acted specifically on colibactin, DC10B was employed as a *pks*⁻ *E. coli* strain. γ-H2AX signal intensity indicated that levels of phosphorylation were similar to that of uninfected cells and the addition of D-Serine did not significantly change levels of phosphorylation. Taken together, these results indicate that D-Serine induced repression of natively encoded colibactin during infection of HeLa cells and led to a reduction in the formation of DSBs as measured by decreased phosphorylation of H2AX.

DSB lesion formation is reduced upon exposure *of pks*⁺ *E. coli* to D-Serine

In response to DNA DSBs, factors involved in the DNA damage response (DDR), including γ-H2AX, accumulate at sites of damage temporarily [48]. These so-called nuclear foci are formed by γ-H2AX which can spread over megabases along the DNA flanking the breakage site [49] and can be visualized using labelled antibodies under fluorescence microscopy [50]. Therefore, we investigated the effects of D-Serine on the formation of DNA DSBs at the subcellular level by visualizing nuclear foci positive for γ-H2AX [51]. HeLa cells were infected with *pks*⁺ *E. coli* and then incubated with anti-γ-H2AX monoclonal antibody, labelled with an Alexa Fluor 555 secondary antibody, and visualised by confocal microscopy. We observed intense punctate staining of γ-H2AX after infecting cells with untreated Nissle 1917 indicative of the induction of DNA DSBs. Exposure to D-Serine led to the formation of fewer nuclear lesions/foci with nuclei resembling those of uninfected cells. The number of foci remained unchanged in cells infected with DH10B pBAC-*pks* both in the presence and absence of D-Serine (**Fig. 4A**). To investigate the heterogeneity in the response of individual HeLa cells to D-Serine during infection γ-H2AX phosphorylation was measured by flow cytometry. This revealed that D-Serine reduced the number of γ-H2AX positive cells by 2.5-fold (54.6% vs 21.1% of cells γ-H2AX +ve). However, exposure to D-Serine did not fully return levels of γ-H2AX fluorescence to that of the uninfected cells (**Fig. 4B**).

Exposure to D-Serine reduces colibactin-associated cellular senescence

Cellular senescence has been described as an irreversible state of cell cycle arrest, often in response to DNA damage [52]. This phenomenon is associated with colibactin and is observed when mammalian cells are infected with live *pks*⁺ *E. coli* [22]. Exposure to D-Serine led to a reduction in H2AX phosphorylation, demonstrating that colibactin-associated DNA damage was reduced by D-Serine (**Fig. 4**). Therefore, we aimed to show that exposure to D-Serine would also result in a reduction in downstream cellular senescence. HeLa cells were infected with live *pks*⁺ *E. coli* for 4 h, before cells were treated with gentamicin for 72 h. Cell morphology was observed by staining actin filaments with Phalloidin-Alexa Fluor 555 and using fluorescence microscopy for visualization (**Fig. 5A**). Colibactin-producing *E. coli* have been distinguished by their ability to induce megalocytosis in cultured eukaryotic cells, a phenotype associated with senescence, characterized by progressive enlargement of the cell body, and the nucleus, and the abolishment of mitosis [23]. In the absence of D-Serine, cells infected with Nissle 1917 displayed the characteristic cytopathic effect,

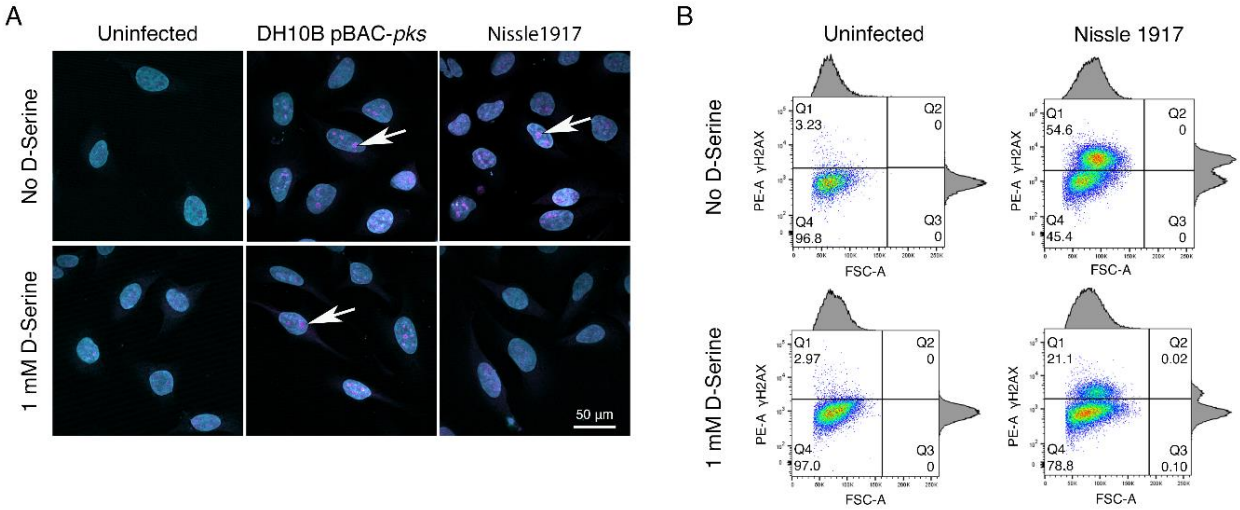


FIGURE 4: D-Serine reduces nuclear foci observed in HeLa cells. HeLa cells were infected for 4 h with *E. coli* Nissle 1917 or DH10B hosting BAC-*pks* (MOI = 400). Infections were performed with and without the addition of 1 mM D-Serine to the growth media. At 8 h post infection, cells were washed, fixed, and stained with anti-γ-H2AX antibody. (**A**) Cells were examined by confocal microscopy for DNA in cyan and phosphorylated histone H2AX protein in magenta. Images of uninfected, and *pks*⁺ infected cells are shown, scale bar = 50 µm. (**B**) Intracellular levels of phosphorylated histone H2AX were measured by flow cytometry 8 h after infection. Dot plots reveal the percentage of viable cells fluorescing in the γ-H2AX channel, 100k events were analysed for each sample.

with cellular and nuclear enlargement being apparent (**Fig. 5**). However, upon exposure to D-Serine, cells infected with Nissle 1917 had a cell morphology that was markedly more like that of the uninfected control. Consistent with our findings described earlier (**Fig. 3**), D-Serine treatment did not result in a decrease in the senescence-associated mor-

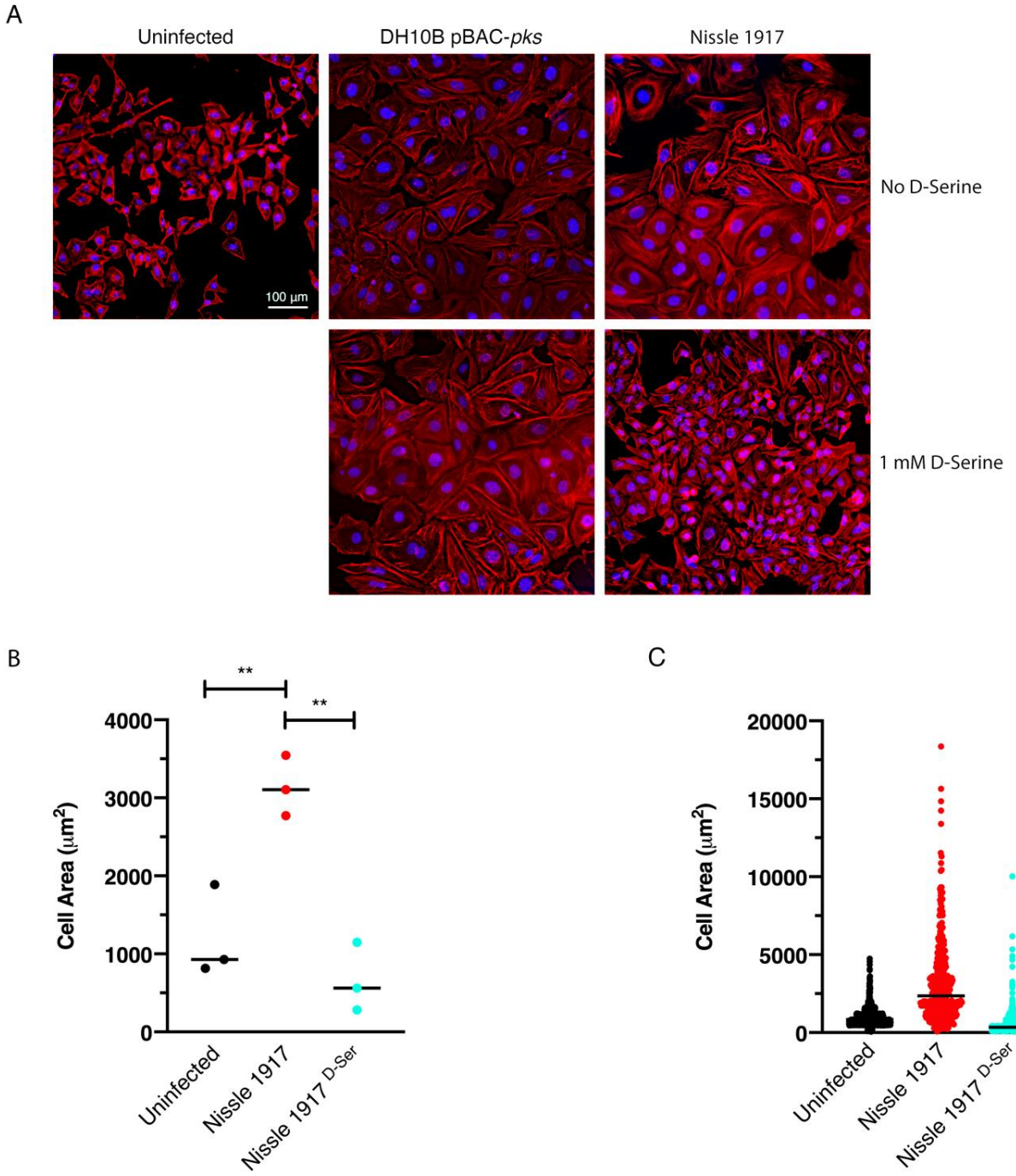


FIGURE 5: Exposure to D-Serine reduces the colibactin associated cellular senescence. HeLa cells were infected for 4 h with *E. coli* Nissle 1917 or DH10B hosting BAC-*pks* (MOI = 400). Infections were performed with and without the addition of 1 mM D-Serine to the growth media. At 8 h after infection, cells were washed and incubated for 72 h to allow for the megacell phenotype to develop. (**A**) HeLa cell morphology was observed by wide field fluorescence. Actin cytoskeleton was stained with Phalloidin in red and DNA was counterstained with DAPI in blue at 72 h post infection. Scale bar = 100 µm. (**B**) CellProfiler software was employed to measure the area of the HeLa cells shown in A using images acquired at 10X magnification. 100 cells were measured per sample. Columns represent the mean cell area measured with individual experimental observations indicated by data points for each infection condition. Measurements were acquired from images taken from three independent experiments and statistical significance was assessed by unpaired Student's *t*-test with, ** indicating $P < 0.01$. (**C**) Individual cell area measurements were recorded across triplicate experiments. Black lines indicate the mean.

phological alterations observed in DH10B pBAC-*pks*. The mean cell area was determined for three independent replicates using image analysis software, CellProfiler (**Fig. 5B**). Cell enlargement increased 2.60-fold (from 1209.52 to 3140.90 µm) during infection with Nissle 1917, however, upon addition of D-Serine, the mean cell area decreased to 666.62 µm ($P < 0.01$). Individual cell area measurements indicate the variable extent of cell enlargement observed across the three experiments (**Fig. 5C**). Overall, these results demonstrated that reduction of colibactin expression through exposure to D-Serine reduced both the acute and more long-term colibactin-related cytopathic effects of infection with Nissle 1917.

The D-Serine metabolism locus is not essential for repression of colibactin in *pks*⁺ *E. coli* in response to D-Serine

DsdC is a well-studied D-Serine responsive transcriptional regulator that is required for catabolism of D-Serine via activation of *dsdXA* [16]. Our recent work has shown that DsdC does not only regulate the *dsdCXA* operon and is in fact capable of binding of numerous areas of the genome [53]. We hypothesized that the repression of the *pks* locus by D-Serine might be mediated by DsdC. Therefore, *dsdC* was deleted in Nissle 1917 and the effect on colibactin production with and without D-Serine was assessed. First, we compared transcriptomes from a previous study [21], which revealed that exposure to D-Serine triggered similar reductions in *clb* gene expression in both wild type (WT) and Δ*dsdC* CFT073 strains (**Fig. 6A**). To assess the effects on genotoxic activity, HeLa cells were infected with Nissle Δ*dsdC* as described previously. Inclusion of D-Serine resulted in a 6.34-fold reduction in *clbB* expression compared to cells infected with Nissle Δ*dsdC* alone ($P = 0.0145$; **Fig. 6B** and **C**). Furthermore, HeLa cells infected with Nissle Δ*dsdC* in the presence of D-Serine were protected from genotoxicity as lower levels of phosphorylated H2AX were detected, indicating that the DNA repair response was reduced in both WT and Δ*dsdC* genetic backgrounds. Taken together, these data demonstrate that D-Serine-associated down-regulation of *pks* encoded genes occurs independently of DsdC.

DISCUSSION

Sensing and responding to environmental cues and signalling molecules is crucial for bacterial survival. Indeed, modulating gene expression enables bacteria to adapt and persist in changing environments; a trait that is essential for certain *E. coli* pathotypes that can colonize multiple sites of the human host. For instance, UPEC strains can survive in both the gastrointestinal tract and the bladder. Strikingly, previous work has shown that UPEC strains extensively carry the *dsdCXA* locus [10], enabling these strains to metabolize D-Serine in the nutrient deficient bladder. Furthermore, metabolizing D-Serine has been shown to confer a fitness advantage in strains that infect this site [9]. Indeed, there has been a growing appreciation

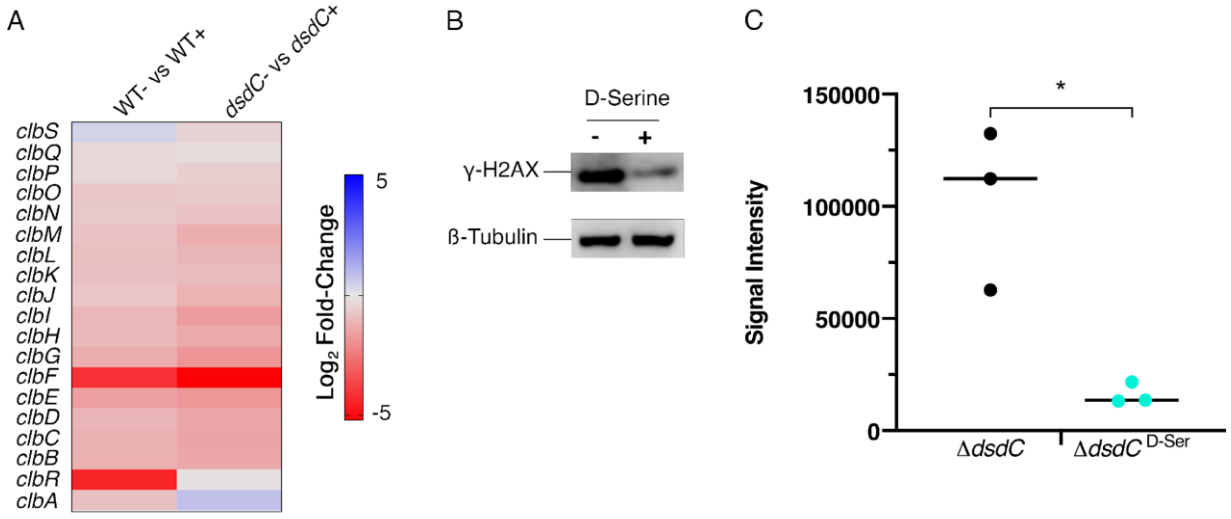


FIGURE 6: Deletion of *dsdC* does not affect D-Serine associated repression of colibactin. Generation of an isogenic mutant, Nissle Δ*dsdC,* revealed that D-Serine-associated repression of colibactin activity occurred independently of the D-Serine tolerance locus in *pks*⁺ *E. coli* strains. (**A**) Heat map showing the EdgeR calculated $\log_2$ relative fold changes for each gene in the *pks* island with corresponding colour key adjacent. False discovery rate-corrected *P* values can be found in Table S1. (**B**) The genotoxic activity of Nissle Δ*dsdC* was assessed by infecting HeLa cells as discussed above. Proteins were extracted and the level of H2AX phosphorylation was determined. Immunoblot analysis of cell lysates extracted 4 h post-infection is shown. Anti-γ-H2AX antibody was used as an indicator of double stranded DNA breaks and β-Tubulin was used as a loading control for cell lysates. (**C**) Signal intensities of bands were measured as described in the methods using LI-COR Image Studio. γ-H2AX signals were corrected to account for any variation in loading using β-Tubulin signal intensity. Experimental signal was normalized so that the mean signal intensity of the samples was equivalent for each experiment. The experiment was carried out in triplicate. Columns represent mean +/- SEM with individual experimental observations indicated by data points. Statistical significance was assessed by unpaired Student's *t*-test with, * indicating $P < 0.05$.

for the important role of amino acids as carbon sources; however, there has also been an emphasis on investigating the regulatory functions of certain amino acids. Investigations into *E. coli* biofilm formation revealed the spatiotemporal regulation of L-Alanine metabolism is essential for cell viability and growth of colonies [54] whereas catalysis of the amino acid L-Tryptophan was implicated with the inhibition of biofilm formation [55]. In the mammalian host, sensing metabolites can modulate the expression of essential virulence genes. In response to L-Arginine in the gut, EHEC upregulate locus of enterocyte effacement (LEE)-encoded virulence genes to facilitate site-specific colonization of the host [56]. Exposure to D-Serine has been implicated in the downregulation of virulence genes in this *E. coli* pathotype, with our previous work demonstrating that D-Serine represses the LEE-encoded type three secretion system [10]. While our work has shown that D-Serine is present in trace concentrations in the gut [10], approximately 1000-fold lower than that of the bladder [9], the production of D-Serine by members of the gut microbiome has been reported [57]. As a result, the *E. coli* strains residing in the gut may encounter localized micro-niches rich in this metabolite, raising the possibility that it functions as a niche-specific regulator of diverse virulence genes in pathogenic *E. coli*. In this study we observed that, in general, D-amino acids exerted an enhanced ability to regulate expression of the genes involved in colibactin production over their L-amino acid enantiomers. We demonstrated that treatment with several D-amino acids, most notably D-Cysteine and D-Serine, resulted in downregulation of the *pks* genomic island in colibactin producing *E. coli.* Interestingly, analysis of gene expression in CFT073, revealed that L-Cysteine and L-Serine, did not significantly repress *clbB*. Therefore, while L-amino acids are favoured in nature [3], D-amino acids can have distinct effects in regulating gene expression and the mechanisms underpinning these effects warrant further investigation.

Colibactin research has been invigorated after a recent breakthrough study revealed that exposure to colibactin caused a specific mutational signature that linked *pks*$^+$ harbouring *E. coli* to CRC tumours [36]. Thus, researchers have endeavoured to identify compounds that will inhibit colibactin and prevent the pro-tumorigenic effects. Cougnoux *et al*., (2016) described the use of boron-based compounds that functioned as enzyme competitors and inhibited the activity of ClbP, the serine peptidase involved in colibactin maturation [58]. Furthermore, the use of these compounds prevented the genotoxic and tumorigenic activity of colibactin on epithelial cells and in a CRC mouse model [58]. Similarly, mesalamine – an anti-inflammatory drug that is used for treating inflammatory bowel disease (IBD) and is associated with reduced risk of CRC in IBD patients [59] – has been shown to reduce *clbB* expression and hence to inhibit production of colibactin [60]. The increasing prevalence of drug resistance has led to the development of new and natural antimicrobial agents, some of which have interesting biological activities beyond their intended use as bactericidal agents. For example, cinnamon and its essential oil (cinnamaldehyde) have been studied for their antibacterial properties [61]. Interestingly, recent research showed that treatment with these compounds induced downregulation of the *pks*-encoded *clbB* gene *in E. coli* strains isolated from patients with CRC [62]. In addition, tannin, a compound extracted from medicinal plants, was also shown to repress transcription of colibactin and prevent the associated genotoxic activity of colibactin producing *E. coli* [63]. In this study we have highlighted the important role of naturally available amino acids, by identifying D-Serine, as a potent repressor of colibactin. We showed that exposure to D-Serine prevented colibactin-associated cytopathic effects in eukaryotic cells and that treatment with 1 mM D-Serine was sufficient to induce prolonged protection, with signs of cellular senescence remaining absent after 72 h. This suggests that D-Serine could have prophylactic potential, providing the host with long-lived protection against colibactin production by commensal *E. coli* residing in the gastrointestinal tract. Such a treatment would particularly benefit high-risk patients, such as individuals with IBD, where the prevalence of *E. coli* belonging to the B2 phylogroup is high and the incidence of developing CRC is significantly greater [64, 65]. While our results indicate promising potential for D-Serine as a therapeutic, it should be noted that colibactin has been detected in urine (an environment rich in D-Serine [9]) from individuals with *pks*$^+$ urinary tract infections [66]. In the same study, colibactin-induced DNA damage was observed in a murine model of cystitis. Levels of D-Serine can vary across individuals and be affected by diet, hence further work will be needed to understand the potential of D-Serine to repress colibactin in complex host environments.

Bacterial gene regulatory networks tightly govern the expression of genes in response to chemical or environmental stimuli; however, current understanding on the regulation of the colibactin gene locus is limited. Homburg *et al.,* (2007) sought to elucidate the transcriptional organization and regulation of the colibactin genes using Nissle 1917 as a model [39]. Their results showed that the *clb* locus could be divided in to at least seven transcriptional units, of which four were found to be transcribed polycistronically. The polycistronic expression of *clbR*/*clbA* indicated a potential regulatory function exerted by the *clbR* encoded LuxR-like regulatory protein on *clbA.* Indeed, the expression of *clbA* is crucial for colibactin production, as it encodes a phosphopantetheinyl transferase that is responsible for the post-translational activation of the PKS and NRPS proteins of the colibactin biosynthesis pathway [39]. Recently, *clbR* was identified to encode the key transcriptional activator of the *clb* genes, and expression of this gene was found to directly correlate with the function and production of colibactin in *E. coli pks*$^+$ strain M1/5 [67]. Different carbon sources have been found to influence *clbRA* transcript levels, with *clbA* and *clbR* upregulated in the presence of glucose and glycerol compared with pyruvate and acetate, where expression levels remained at an intermediate level [39]. Increased expression of *clbA* was found to be associated with exposure to nondigestible oligosaccharides commonly found in prebiotics such as

lactose and raffinose [68], inulin and galacto-oligosaccharides [69]. Inulin also increased expression of *clbB*, *clbQ* and *clbR* [69]. These data highlight that the diet may play a significant role in controlling the production of colibactin, with certain foods potentially creating a pro-tumorigenic environment. Interestingly, iron homeostasis has also been implicated with the production of colibactin, with *clbA* downregulated and colibactin-associated megalocytosis reduced upon exposure to $FeCl_3$ [70]. Furthermore, supplementation of iron into media containing colibactin-inducing oligosaccharides resulted in the abrogation of *clbA* induction [69]. The production of colibactin was shown to be regulated by the ferric uptake regulator (Fur) by Tronnet *et al.*, (2016). The study revealed that Fur positively regulated *clbA* expression, and that transcription was initiated by direct binding of Fur to the *clbA* promoter [71]. However, iron-dependent transcription of *clbA* was found to be independent of *clbR* [71], suggesting the involvement of a second *clbA* promoter that may be activated upon specific iron conditions. Iron availability is tightly controlled in the host, and Gram-negative bacteria have evolved several iron uptake systems to overcome this limitation, including the secretion of iron sequestering siderophores [72]. Intriguingly, *clbA* has been shown to have diverse functionality, contributing to the synthesis of both colibactin and the yersiniabactin siderophore [73]. Taken together, these reports suggest that complex regulatory networks are involved in the production of colibactin and that transcription of *clb* genes may be activated under specific nutritional conditions which could serve as a fitness advantage to *pks*-harbouring bacteria.

D-Serine exposure has been demonstrated to induce the activation of the LTTR DsdC, whose main function has been described to be the activation of *dsdXA* transcription [16] in response to D-Serine. This cluster facilitates uptake and catabolism of D-Serine, respectively. As DsdC is the primary transcriptional regulator induced in response to D-Serine, we investigated whether the repression of *clb* genes was mediated by DsdC. Our data indicate that repression of colibactin by D-Serine is not directly facilitated by DsdC, as similar responses to D-Serine were observed in both WT and Nissle Δ*dsdC* both in terms of repression of *clb* gene expression and cytopathic responses. These findings were unexpected. Interestingly, Wallenstein *et al.*, (2020) recently demonstrated that the *pks*-encoded *clbR* gene is the transcriptional activator of the colibactin genes [67]. Like DsdC, ClbR has been described as a LTTR with a helix-turn-helix DNA-binding motif that interacts with the *clbR-clbB* intergenic region, suggesting it is involved in the regulation of the *pks* island [39, 67]. LTTR regulators are thought to be activated by a small molecule co-inducer [74]. Therefore, it is also plausible to consider that D-Serine could enter the cell by an alternative transport system and act as a co-inducer of ClbR, Fur, or indeed another elusive LTTR. Indeed, D-Serine can enter the cell via CycA [19, 75] thereby potentially explaining why exposure to D-Serine reduced *clbB* expression in both Nissle Δ*dsdC* and in the WT. The precise regulatory mechanism by which D-Serine causes a reduction in colibactin gene expression remains unsolved and will be a focus of future work in our group.

In conclusion, in this study we have identified a D-amino acid that has a strong regulatory effect on the genes encoded in the *pks* genomic island. Exposure to D-Serine in *pks*$^+$ *E. coli* results in downregulation of the *clb* genes and consequently, reduced DNA cross-linking and a reduction in the phenotypic responses associated with colibactin-induced DNA damage in cultured eukaryotic cells. Furthermore, deletion of *dsdC* revealed that D-Serine-induced inhibition of colibactin is not mediated by the D-Serine metabolism locus regulator, DsdC. Understanding the physiological implications *in vivo* will be key in further exploring the prophylactic potential of D-Serine.

MATERIALS AND METHODS

Bacterial strains, plasmids and cultures

The bacterial strains, plasmids and oligonucleotides used in this study are listed in Table S4 and Table S5. Bacteria were routinely grown at 37°C in Luria broth (LB [Miller's recipe]) before diluting 1/100 into the appropriate medium for experiments or growth analysis. Chloramphenicol was used when appropriate at a concentration of 25 µg/ml. All preparations of M9 minimal medium (Sigma Aldrich; cat# M6030) were supplemented with 0.4% (*w/v*) glucose unless otherwise stated. For HeLa cell infection experiments, bacteria grown overnight were inoculated in prewarmed MEM-HEPES (Sigma Aldrich; cat# M7278) -/+ 1 mM D-Serine and incubated at 37°C, 200 RPM for 4.5 h. All growth media, antibiotics and chemicals were purchased from Sigma Aldrich unless stated otherwise.

Assessment of DNA crosslinking activity

The assay was performed as previously described [29]. Briefly, linearized plasmid DNA was generated by digestion of pUC19 plasmid with BamHI (New England Biolabs). For bacteria-DNA interactions, bacteria were inoculated 1:20 from overnight cultures into 10 mL M9 Minimal Media alone or supplemented with 1 mM D-Serine and grown for 1.5 h. After reaching an $OD_{600\ nm}$ of ~0.6, 1 × 10^6 CFU were inoculated in 100 µL M9 Minimal media alone or supplemented with 1 mM D-Serine and incubated statically at 37°C for 5 h. Following cultivation, cells were harvested by centrifugation and the media was removed. Cells were resuspended in sterilised nuclease-free water. Then, a mixture of 450 ng of linearized DNA and 1 mM EDTA was added and samples were incubated for a further 40 min. Bacteria were pelleted by centrifuging samples at 5,000 x g for 5 min, then the DNA present in the supernatants was purified using the PCR Purification kit (Qiagen) according to the manufacturer's instructions. A denaturing 1% agarose gel was prepared in a 100 mM NaCl and 2 mM EDTA solution (pH 8.0), then the gel was soaked overnight in an alkaline running buffer solution (40 mM NaOH and 1 mM EDTA, pH ~12). For each sample, 100 ng of DNA was loaded on to the denaturing agarose gel and then the gel was run at 1 V/cm for 45 min and then for 2 h at 2 V/cm. The gel was neutralized in a 100 mM Tris pH 7.4 buffer solution containing 150 mM NaCl, that was frequently changed, for a total of 45 min. The gel was stained with GelRed and DNA was revealed with UV exposure using the ChemiDoc Imaging System (Bio-Rad).

HeLa cell culture, infection and examination of cellular senescence

For bacterial infections, overnight LB cultures of bacteria were cultured in prewarmed MEM-HEPES -/+ 1 mM D-Serine and cultured for 4.5 h at 37˚C with 200 RPM agitation. Bacterial suspensions were diluted to $OD_{600\ nm}$ of 0.1, before serially diluting and spot plating on LB plates to confirm appropriate cell density. HeLa cells were routinely cultured in DMEM (ThermoFisher Scientific; cat# 61965026) with 10% (*v/v*) foetal calf serum (FCS) at 37˚C, in a 5% CO_2 incubator and were maintained by serial passage. For infection experiments, 4 x 10^4 cells/well were seeded on 13 mm glass coverslips pre-coated with collagen (Millipore; cat# 08-115) as per the manufacturer's instructions. After 24 h and immediately prior to infection, the HeLa cells were washed with DPBS (ThermoFisher Scientific, cat# 14190086) and medium was replaced with MEM-HEPES -/+ 1 mM D-Serine. Bacteria were added to each coverslip at a multiplicity of infection of 400 (200 µl of 0.1 $OD_{600\ nm}$ suspension) and infected for 4 h. The cells were washed twice with DPBS 4 h after inoculation, then cells were replenished with DMEM containing 10% (v/v) FCS and 50 µg/ml gentamicin (Sigma Aldrich, cat# G1397) and incubated for 72 h at 37ºC, 5% CO_2. Next, the cells were fixed in 4% (*w/v*) paraformaldehyde at room temperature for 15 min. The cells were then permeabilized with 0.1% (*v/v*) Triton X-100 in DPBS for 5 min. After two washes, each coverslip was stained with 0.2 U Phalloidin- Alexa Fluor 555 (Invitrogen, cat# A34055) for 1 h at room temperature. The cells were washed twice with DPBS before the coverslips were mounted on to a glass slide with 4 µl Vectashield with DAPI (Vector Laboratories, cat# H-1200) and sealed with clear nail polish. Images were acquired using a Zeiss AxioImager M1 and images were processed by deconvolution using Zen 2.3 Pro software (Zeiss). The area of each cell was measured using a pipeline developed on CellProfiler [76]. Briefly, sample images were acquired at 10X magnification to allow for >100 cells to be captured per image. Image files were uploaded to the CellProfiler workspace and analysis was performed for images taken from three replicate experiments.

Immunofluorescence analysis of H2AX phosphorylation

HeLa cells were seeded on 13 mm coverslips and infected as described above. Following 4 hours infection cells were washed twice with DPBS and fixed in 4% (*w/v*) paraformaldehyde for 15 min at room temperature. The cells were permeabilized with 0.1% Triton X-100 and then blocked with 1X Phosphate-Buffered Saline, 0.1% Tween® 20 (PBST) + 10% normal goat serum (Sigma Aldrich, cat# NS02L) for 1 h at room temperature. Next cells were incubated with rabbit monoclonal anti γ-H2AX antibodies (Cell Signalling, cat# 5438S) diluted 1:100 in blocking solution and incubated for 1 h at room temperature. The tissues were washed three times with DPBS, then a fluorescent secondary antibody, Alexa Fluor 555 Goat anti-rabbit IgG (Invitrogen, cat# A32732) diluted 1:400 in blocking solution was applied and incubated for 1 h in the dark at room temperature. Following incubation, tissues were washed three times with DPBS, then coverslips were mounted to a glass slide with 4 µl Vectashield with DAPI and sealed with clear nail polish. Nuclear foci were visualized using a Zeiss LSM 880 confocal microscope (Zeiss).

Flow cytometry analysis of H2AX phosphorylation

Twenty-four well tissue culture plates were collagen-coated, seeded with 10^6 cells/well and infected as described above. The cells were collected by trypsinization 4 h post-infection and washed in DPBS. The cells were then collected by centrifugation and resuspended in a live/dead stain (eFluor 780 [eBioscience, cat# 65-0865-14]) and incubated for 20 min on ice. The cells were washed in excess Stain Buffer (BD Biosciences, cat# 55456) and FC receptor block (DPBS + 10% FCS) was applied before incubating for a further 20 min on ice. Next, the cells were fixed with Cytofix (BD Biosciences, cat# 554655) for 15 min at 37˚C and permeabilized with Perm Buffer (BD Biosciences, cat# 558050) for 30 min on ice, before purified mouse anti γ-H2AX (BD Biosciences, cat# 560443) diluted 1:200 in BD Stain Buffer was applied and cells were incubated for 1 h at room temperature. A multichromatic-conjugated secondary antibody, goat anti-mouse IgG (BD Biosciences, cat# 550589) diluted 1:1000 in Stain Buffer was applied and cells were incubated for 1 h at room temperature, keeping samples protected from the light. Cells were eventually suspended in BD Stain Buffer and filtered with a 70 µm filter. Cells were analysed using the BD FACSAria (BD) and the data was analysed using FloJo software. Analysis of the stained cell populations was performed by gating on single, live cells.

Western Blot analysis of H2AX phosphorylation

HeLa cells were infected with *E. coli* Nissle 1917 (MOI = 400) for 4 h then treated with gentamicin for a further 4 h. The tissues were washed and lysed directly in the cell culture well by applying 100 µl of 1X SDS sample buffer and incubating for 5 min then wells were scraped to release attached cells, cell lysates were immediately stored on ice. The cell lysates were heated for 10 min at 90°C, and aliquots were stored at −20°C. Proteins were separated on 4-12% Bis Tris Gel by SDS-PAGE (Invitrogen, cat# NP0321) and transferred to a nitrocellulose membrane (FisherScientific, cat# 88018). Blocking was performed using 5% skimmed milk powder for 1 h in PBST. The membrane was then incubated with anti-γ H2AX primary antibody (Cell Signalling Technologies, cat# 5438S) diluted 1:100 in 5% BSA-PBST and incubated overnight at 4˚C. The membrane was then washed three times with PBST for 10 min before being incubated for 1 h with anti-rabbit horseradish peroxidase (HRP)-conjugated secondary antibody (Invitrogen, cat# 656120) diluted 1:2500 in PBST. The membrane was again washed three times with PBST for 10 min. Bound secondary HRP-labelled antibodies were revealed with SuperSignal™ West Femto maximum sensitivity substrate (ThermoFisher Scientific, cat# 34096) and analysed with the C-DiGit® blot scanner (LI-COR). Membranes were stripped with mild stripping buffer and incubated with the primary antibody H2AX (Cell Signalling Technologies, cat# 25955) and detected with the secondary antibody as described above. To control for sample loading, membranes were probed for β-Tubulin (Abcam, cat# ab6046). Proteins were quantified with Image Studio Lite (Licor) and normalized in relation to the β-Tubulin level.

RNA Extraction

For screening of amino acids capable of modulating expression of *clbB*, bacteria were cultured in M9 Minimal Media supplemented with amino acids at 1 mM final concentration for 5 h. The cells were harvested by centrifugation before resuspen-

sion in two volumes of RNAprotect Bacteria Reagent (Qiagen, cat# 76506). After a 5 min incubation at room temperature, cells were harvested, and RNA extractions were carried out using PureLink RNA Mini Kit (Thermofisher Scientific, cat# 12183018A) according to manufacturer's instructions. Contaminating DNA was removed by TurboDNase (ThermoFisher Scientific, cat# AM2238) treatment, followed by extraction in phenol-chloroform-isoamyl alcohol (Sigma Aldrich, cat# P2069) and ethanol precipitation at -80˚C overnight. RNA was collected by centrifugation before washing with 70% ethanol and resuspending in nuclease-free water.

Synthesis of cDNA and RT-qPCR

Ten nanograms of DNA-free total RNA, extracted as described above, was used as a template to prepare 10 µl cDNA using LunaScript®RT SuperMix Kit (New England Biolabs, cat# E3010L) according to manufacturer's instructions. Luna® Universal qPCR Master mix (New England Biolabs, cat# M3003L) was employed for RT-qPCR according to manufacturer's recommendations (Initial denaturation 95°C, 1 min; denaturation 95°C, 15 sec; extension 60°C, 30 sec; 39 cycles). Technical duplicate 20 µl reactions were carried out with 1 µl volumes of cDNA as template and triplicate biological samples of cDNA being analysed. Expression relative to the untreated control was calculated as $2^{-\Delta\Delta ct}$ with [77] *gapA* amplification being employed as a housekeeping control.

Assessment of *clbB* expression using promoter-*gfp* fusion reporter

The promoter region upstream of *clbB* was amplified using the P*clbB*-Fw and P*clbB*-Rev primers Table S5, before digesting with BamHI/KpnI and cloning in frame upstream of green fluorescent protein (*gfp*) in pAJR70 [78]. Bacteria were cultured in MEM-HEPES for 4 h whilst recording $OD_{600\ nm}$, and fluorescence using a BMG Fluostar plate reader. Relative fluorescence was determined by subtracting the fluorescence intensity of an empty vector control from each sample and dividing the corrected fluorescence intensity by $OD_{600\ nm}$. Data presented are from three replicate experiments.

Determination of growth rates

To determine the specific growth rate (SGR) of strains, bacteria were cultured in MEM-HEPES or M9 minimal media for up to 8 h whilst recording $OD_{600\ nm}$. Exponential growth was observed between two and five hours for all the strains and media tested. SGR (μ = ΔlnOD 600 nm/Δt) was calculated - $\Delta\ln OD_{600\ nm}$ is the change in natural logarithm of $OD6_{00\ nm}$ and Δt is the change in time (hours) between two and five hours. $OD_{600\ nm}$ was recorded in an Eppendorf D30 Biophotometer. At least three replicate growth curves were constructed for determination of SGR of any given strain.

Construction of isogenic Δ*dsdC* Nissle 1917 mutant strain

A Nissle 1917 mutant lacking *dsdC* was constructed using standard procedures [79]. Briefly, Nissle 1917 WT was transformed with pKD46. A single colony was cultured at 37ºC in LB supplemented with 100 µg/ml ampicillin and 100 mM L-Arabinose to an $OD_{600\ nm}$ of 0.4. The cells were then washed and resuspended three times with ice-cold distilled water. A linear deletion fragment was prepared by amplifying the chloramphenicol resistance cassette from pKD3 with oligonucleotides bearing 50 bp 5'-end flanking regions homolgous to the 50 bp regions immediately upstream and downstream of *dsdC*, Table S5. One microgram of PCR product (phenol-chloroform extracted, before ethanol precipitation and resuspension in 10 µl nuclease-free water) was electroporated at 2500 V into an aliquot of competent Nissle 1917 cells using an Eppendorf Eporator. Insertional mutants grown under chloramphenicol selection were verified by PCR using check primers, Table S5. Resistance cassettes were removed by expression of FLP-recombinase under transient temperature shift to 42ºC after transformation of insertional mutants with pCP20. Excision of the resistance cassette was confirmed by PCR.

Statistical analysis

Statistical significance was assessed by unpaired two-tailed Student's *t*-test or ordinary one- way ANOVA, as indicated. Tests were carried out using GraphPad Prism 8.

ACKNOWLEDGMENTS

We thank Diane Vaughan and Alana Hamilton (Flow Core Facility, University of Glasgow, United Kingdom) for their assistance with processing samples for flow cytometry and Dr Leandro Lemgruber Soares and Susan Baillie (Glasgow Imaging Facility, University of Glasgow, United Kingdom) for their assistance with the acquisition of confocal images. We are very grateful to Dr Michael Ormsby (University of Stirling, United Kingdom) and Dr James Connolly (Newcastle University, United Kingdom) for their insightful appraisal of our manuscript. This work is supported by the Biotechnology and Biological Sciences Research Council, grant numbers BB/M029646/1 and BB/R006539/1 and by the Wellcome Trust Integrative Infection Biology PhD Programme at the University of Glasgow, grant number 218518/Z/19/Z. The work performed in Toulouse, France, was funded by the French National Agency for Research (ANR), grant number UTI-TOUL ANR-17-CE35-0010 and ANR-19-AMRB-0008.

SUPPLEMENTAL MATERIAL

All supplemental data for this article are available online at www.microbialcell.com.

CONFLICT OF INTEREST

All authors declare that they have no conflicts of interest.

COPYRIGHT

Please cite this article as: Jennifer C. Hallam, Sofia Sandalli, Iris Floria, Natasha C. A. Turner, Min Tang-Fichaux, Eric Oswald, Nicky O'Boyle and Andrew J. Roe (**2023**). D-Serine reduces the expression of the cytopathic genotoxin colibactin. **Microbial Cell** 10(3): 63-77. doi: 10.15698/mic2023.03.793

REFERENCES

1. Bäumler AJ, and Sperandio V (**2016**). Interactions between the microbiota and pathogenic bacteria in the gut. **Nature** 535(7610): 85–93. doi: 10.1038/nature18849

2. O'Boyle N, Connolly JPR, Tucker NP, and Roe AJ (**2020**). Genomic plasticity of pathogenic *Escherichia coli* mediates D -serine tolerance via multiple adaptive mechanisms. **Proc Natl Acad Sci USA** 117(36): 22484–22493. doi: 10.1073/pnas.2004977117

3. Marcone GL, Rosini E, Crespi E, and Pollegioni L (**2020**). D-amino acids in foods. **Appl Microbiol Biotechnol** 104(2): 555–574. doi: 10.1007/s00253-019-10264-9

4. Brückner H, and Westhauser T (**2003**). Chromatographic determination of L- and D-amino acids in plants. **Amino Acids** 24(1–2): 43–55. doi: 10.1007/s00726-002-0322-8

5. Cava F, Lam H, de Pedro MA, and Waldor MK (**2011**). Emerging knowledge of regulatory roles of D-amino acids in bacteria. **Cell Mol Life Sci** 68(5): 817–831. doi: 10.1007/s00018-010-0571-8

6. Lam H, Oh D-C, Cava F, Takacs CN, Clardy J, de Pedro MA, and Waldor MK (**2009**). D-amino acids govern stationary phase cell wall remodeling in bacteria. **Science** 325(5947): 1552–1555. doi: 10.1126/science.1178123

7. Elgamoudi BA, Taha T, and Korolik V (**2020**). Inhibition of Campylobacter jejuni Biofilm Formation by D-Amino Acids. **Antibiotics** 9(11): 836. doi: 10.3390/antibiotics9110836

8. Brauer AL, White AN, Learman BS, Johnson AO, and Armbruster CE (**2019**). d-Serine Degradation by Proteus mirabilis Contributes to Fitness during Single-Species and Polymicrobial Catheter-Associated Urinary Tract Infection. **mSphere** 4(1): e00020-19. doi: 10.1128/mSphere.00020-19

9. Anfora AT, Haugen BJ, Roesch P, Redford P, and Welch RA (**2007**). Roles of serine accumulation and catabolism in the colonization of the murine urinary tract by Escherichia coli CFT073. **Infect Immun** 75(11): 5298–5304. doi: 10.1128/IAI.00652-07

10. Connolly JPR, Goldstone RJ, Burgess K, Cogdell RJ, Beatson SA, Vollmer W, Smith DGE, and Roe AJ (**2015**). The host metabolite D-serine contributes to bacterial niche specificity through gene selection. **ISME J** 9(4): 1039–1051. doi: 10.1038/ismej.2014.242

11. METZLER DE, and SNELL EE (**1952**). Deamination of serine. II. D-Serine dehydrase, a vitamin B6 enzyme from Escherichia coli. **J Biol Chem** 198(1): 363–373. doi: 10.1016/s0021-9258(18)55590-3

12. Tanigawa M, Shinohara T, Saito M, Nishimura K, Hasegawa Y, Wakabayashi S, Ishizuka M, and Nagata Y (**2010**). D-Amino acid dehydrogenase from Helicobacter pylori NCTC 11637. **Amino Acids** 38(1): 247–255. doi: 10.1007/s00726-009-0240-0

13. Chang YF, and Adams E (**1974**). D-lysine catabolic pathway in Pseudomonas putida: interrelations with L-lysine catabolism. **J Bacteriol** 117(2): 753–764. doi: 10.1128/jb.117.2.753-764.1974

14. Pioli D, Venables WA, and Franklin FC (**1976**). D-Alanine dehydrogenase. Its role in the utilisation of alanine isomers as growth substrates by Pseudomonas aeruginosa PA01. **Arch Microbiol** 110(23): 287–293. doi: 10.1007/BF00690240

15. Kubota T, Kobayashi T, Nunoura T, Maruyama F, and Deguchi S (**2016**). Enantioselective Utilization of D-Amino Acids by Deep-Sea Microorganisms. **Front Microbiol** 7: 511. doi: 10.3389/fmicb.2016.00511

16. Nørregaard-Madsen M, McFall E, and Valentin-Hansen P (**1995**). Organization and transcriptional regulation of the Escherichia coli K-12 D-serine tolerance locus. **J Bacteriol** 177(22): 6456–6461. doi: 10.1128/jb.177.22.6456-6461.1995

17. Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME, Hultgren SJ, and Gordon JI (**2013**). Genomic diversity and fitness of E. coli strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. **Sci Transl Med** 5(184): 184ra60. doi: 10.1126/scitranslmed.3005497

18. Roesch PL, Redford P, Batchelet S, Moritz RL, Pellett S, Haugen BJ, Blattner FR, and Welch RA (**2003**). Uropathogenic Escherichia coli use d-serine deaminase to modulate infection of the murine urinary tract. **Mol Microbiol** 49(1): 55–67. doi: 10.1046/j.1365-2958.2003.03543.x

19. Cosloy SD (**1973**). D-serine transport system in Escherichia coli K-12. **J Bacteriol** 114(2): 679–684. doi: 10.1128/jb.114.2.679-684.1973

20. Haugen BJ, Pellett S, Redford P, Hamilton HL, Roesch PL, and Welch RA (**2007**). In Vivo Gene Expression Analysis Identifies Genes Required for Enhanced Colonization of the Mouse Urinary Tract by Uropathogenic Escherichia coli Strain CFT073 dsdA. **Infect Immun** 75(1): 278–289. doi: 10.1128/IAI.01319-06

21. Connolly JPR, Turner NCA, Hallam JC, Rimbi PT, Flett T, McCormack MJ, Roe AJ, and O'Boyle N (**2021**). d-Serine induces distinct transcriptomes in diverse Escherichia coli pathotypes. **Microbiology** 167(10). doi: 10.1099/mic.0.001093

22. Nougayrède J-P, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, and Oswald E (**2006**). Escherichia coli induces DNA double-strand breaks in eukaryotic cells. **Science** 313(5788): 848–851. doi: 10.1126/science.1127059

23. Putze J, Hennequin C, Nougayrède J-P, Zhang W, Homburg S, Karch H, Bringer M-A, Fayolle C, Carniel E, Rabsch W, Oelschlaeger TA, Oswald E, Forestier C, Hacker J, and Dobrindt U (**2009**). Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. **Infect Immun** 77(11): 4696–4703. doi: 10.1128/IAI.00522-09

24. Lu M-C, Chen Y-T, Chiang M-K, Wang Y-C, Hsiao P-Y, Huang Y-J, Lin C-T, Cheng C-C, Liang C-L, and Lai Y-C (**2017**). Colibactin Contributes to the Hypervirulence of pks+ K1 CC23 Klebsiella pneumoniae in Mouse Meningitis Infections. **Front Cell Infect Microbiol** 7: 103. doi: 10.3389/fcimb.2017.00103

25. Marcq I, Martin P, Payros D, Cuevas-Ramos G, Boury M, Watrin C, Nougayrède J-P, Olier M, and Oswald E (**2014**). The genotoxin colibactin exacerbates lymphopenia and decreases survival rate in mice infected with septicemic Escherichia coli. **J Infect Dis** 210(2): 285–294. doi: 10.1093/infdis/jiu071

26. McCarthy AJ, Martin P, Cloup E, Stabler RA, Oswald E, and Taylor PW (**2015**). The Genotoxin Colibactin Is a Determinant of Virulence in Escherichia coli K1 Experimental Neonatal Systemic Infection. **Infect Immun** 83(9): 3704–3711. doi: 10.1128/IAI.00716-15

27. Johnson JR, Johnston B, Kuskowski MA, Nougayrede J-P, and Oswald E (**2008**). Molecular epidemiology and phylogenetic distribution of the Escherichia coli pks genomic island. **J Clin Microbiol** 46(12): 3906–3911. doi: 10.1128/JCM.00949-08

28. Auvray F, Perrat A, Arimizu Y, Chagneau CV, Bossuet-Greif N, Massip C, Brugère H, Nougayrède J-P, Hayashi T, Branchu P, Ogura Y, and Oswald E (**2021**). Insights into the acquisition of the pks island and production of colibactin in the Escherichia coli population. **Microb Genom** 7(5): 000579. doi: 10.1099/mgen.0.000579

29. Bossuet-Greif N, Vignard J, Taieb F, Mirey G, Dubois D, Petit C, Oswald E, and Nougayrède J-P (**2018**). The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. **mBio** 9(2): e02393-17. doi: 10.1128/mBio.02393-17

30. Cuevas-Ramos G, Petit CR, Marcq I, Boury M, Oswald E, and Nougayrède J-P (**2010**). Escherichia coli induces DNA damage in vivo and

triggers genomic instability in mammalian cells. **Proc Natl Acad Sci U S A** 107(25): 11537–11542. doi: 10.1073/pnas.1001261107

31. Vizcaino MI, Engel P, Trautman E, and Crawford JM (**2014**). Comparative Metabolomics and Structural Characterizations Illuminate Colibactin Pathway-Dependent Small Molecules. **J Am Chem Soc** 136(26): 9244–9247. doi: 10.1021/ja503450q

32. Xue M, Kim CS, Healy AR, Wernke KM, Wang Z, Frischling MC, Shine EE, Wang W, Herzon SB, and Crawford JM (**2019**). Structure elucidation of colibactin and its DNA cross-links. **Science** 365(6457): eaax2685. doi: 10.1126/science.aax2685

33. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell BJ, Abujamel T, Dogan B, Rogers AB, Rhodes JM, Stintzi A, Simpson KW, Hansen JJ, Keku TO, Fodor AA, and Jobin C (**2012**). Intestinal inflammation targets cancer-inducing activity of the microbiota. **Science** 338(6103): 120–123. doi: 10.1126/science.1224820

34. Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, Pezet D, and Bonnet R (**2013**). High Prevalence of Mucosa-Associated E. coli Producing Cyclomodulin and Genotoxin in Colon Cancer. **PLOS ONE** 8(2): e56964. doi: 10.1371/journal.pone.0056964

35. Raisch J, Buc E, Bonnet M, Sauvanet P, Vazeille E, de Vallée A, Déchelotte P, Darcha C, Pezet D, Bonnet R, Bringer M-A, and Darfeuille-Michaud A (**2014**). Colon cancer-associated B2 Escherichia coli colonize gut mucosa and promote cell proliferation. **World J Gastroenterol** 20(21): 6560–6572. doi: 10.3748/wjg.v20.i21.6560

36. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, Gurjao C, Manders F, Dalmasso G, Stege PB, Paganelli FL, Geurts MH, Beumer J, Mizutani T, Miao Y, van der Linden R, van der Elst S, Genomics England Research Consortium, Garcia KC, Top J, Willems RJL, Giannakis M, Bonnet R, Quirke P, Meyerson M, Cuppen E, van Boxtel R, and Clevers H (**2020**). Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. **Nature** 580(7802): 269–273. doi: 10.1038/s41586-020-2080-8

37. Dziubańska-Kusibab PJ, Berger H, Battistini F, Bouwman BAM, Iftekhar A, Katainen R, Cajuso T, Crosetto N, Orozco M, Aaltonen LA, and Meyer TF (**2020**). Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. **Nat Med** 26(7): 1063–1069. doi: 10.1038/s41591-020-0908-2

38. Siegel R, Ward E, Brawley O, and Jemal A (**2011**). Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. **CA Cancer J Clin** 61(4): 212–236. doi: 10.3322/caac.20121

39. Homburg S, Oswald E, Hacker J, and Dobrindt U (**2007**). Expression analysis of the colibactin gene cluster coding for a novel polyketide in Escherichia coli. **FEMS Microbiol Lett** 275(2): 255–262. doi: 10.1111/j.1574-6968.2007.00889.x

40. Zha L, Wilson MR, Brotherton CA, and Balskus EP (**2016**). Characterization of Polyketide Synthase Machinery from the pks Island Facilitates Isolation of a Candidate Precolibactin. **ACS Chem Biol** 11(5): 1287–1295. doi: 10.1021/acschembio.6b00014

41. Brotherton CA, and Balskus EP (**2013**). A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. **J Am Chem Soc** 135(9): 3359–3362. doi: 10.1021/ja312154m

42. Wassenaar TM (**2016**). Insights from 100 Years of Research with Probiotic E. Coli. **Eur J Microbiol Immunol** 6(3): 147–161. doi: 10.1556/1886.2016.00029

43. Olier M, Marcq I, Salvador-Cartier C, Secher T, Dobrindt U, Boury M, Bacquié V, Pénary M, Gaultier E, Nougayrède J-P, Fioramonti J, and Oswald E (**2012**). Genotoxicity of Escherichia coli Nissle 1917 strain cannot be dissociated from its probiotic activity. **Gut Microbes** 3(6): 501–509. doi: 10.4161/gmic.21737

44. Massip C, Branchu P, Bossuet-Greif N, Chagneau CV, Gaillard D, Martin P, Boury M, Sécher T, Dubois D, Nougayrède J-P, and Oswald E (**2019**). Deciphering the interplay between the genotoxic and probiotic activities of Escherichia coli Nissle 1917. **PLOS Pathogens** 15(9): e1008029. doi: 10.1371/journal.ppat.1008029

45. Nougayrède J-P, Chagneau CV, Motta J-P, Bossuet-Greif N, Belloy M, Taieb F, Gratadoux J-J, Thomas M, Langella P, and Oswald E A Toxic Friend: Genotoxic and Mutagenic Activity of the Probiotic Strain Escherichia coli Nissle 1917. **mSphere** 6(4): e00624-21. doi: 10.1128/mSphere.00624-21

46. Wiles TJ, and Mulvey MA (**2013**). The RTX pore-forming toxin α-hemolysin of uropathogenic Escherichia coli: progress and perspectives. **Future Microbiol** 8: 73–84. doi: 10.2217/fmb.12.131

47. Goswami C, Fox S, Holden M, Connor M, Leanord A, and Evans TJ (**2018**). Genetic analysis of invasive Escherichia coli in Scotland reveals determinants of healthcare-associated versus community-acquired infections. **Microb Genom** 4(6): e000190. doi: 10.1099/mgen.0.000190

48. Giunta S, Belotserkovskaya R, and Jackson SP (**2010**). DNA damage signaling in response to double-strand breaks during mitosis. **J Cell Biol** 190(2): 197–207. doi: 10.1083/jcb.200911156

49. Iacovoni JS, Caron P, Lassadi I, Nicolas E, Massip L, Trouche D, and Legube G (**2010**). High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. **EMBO J** 29(8): 1446–1457. doi: 10.1038/emboj.2010.38

50. Vítor AC, Huertas P, Legube G, and de Almeida SF (**2020**). Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox. **Front Mol Biosci** 7: 24. doi: 10.3389/fmolb.2020.00024

51. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, and Bonner WM (**1998**). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. **J Biol Chem** 273(10): 5858–5868. doi: 10.1074/jbc.273.10.5858

52. Hayflick L (**1965**). THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. **Exp Cell Res** 37: 614–636. doi: 10.1016/0014-4827(65)90211-9

53. Connolly JPR, Turner NCA, Serrano E, Rimbi PT, Browning DF, O'Boyle N, and Roe AJ (**2022**). Control of resistance against bacteriophage killing by a metabolic regulator in meningitis-associated Escherichia coli. **Proc Natl Acad Sci U S A** 119(45): e2210299119. doi: 10.1073/pnas.2210299119

54. Díaz-Pascual F, Lempp M, Nosho K, Jeckel H, Jo JK, Neuhaus K, Hartmann R, Jelli E, Hansen MF, Price-Whelan A, Dietrich LEP, Link H, and Drescher K (**2021**). Spatial alanine metabolism determines local growth dynamics of Escherichia coli colonies. **Elife** 10: e70794. doi: 10.7554/eLife.70794

55. Shimazaki J, Furukawa S, Ogihara H, and Morinaga Y (**2012**). L-Tryptophan prevents Escherichia coli biofilm formation and triggers biofilm degradation. **Biochem Biophys Res Commun** 419(4): 715–718. doi: 10.1016/j.bbrc.2012.02.085

56. Menezes-Garcia Z, Kumar A, Zhu W, Winter SE, and Sperandio V (**2020**). l-Arginine sensing regulates virulence gene expression and disease progression in enteric pathogens. **Proc Natl Acad Sci U S A** 117(22): 12387–12393. doi: 10.1073/pnas.1919683117

57. Matsumoto M, Kunisawa A, Hattori T, Kawana S, Kitada Y, Tamada H, Kawano S, Hayakawa Y, Iida J, and Fukusaki E (**2018**). Free D-amino acids produced by commensal bacteria in the colonic lumen. **Sci Rep** 8: 17915. doi: 10.1038/s41598-018-36244-z

58. Cougnoux A, Delmas J, Gibold L, Faïs T, Romagnoli C, Robin F, Cuevas-Ramos G, Oswald E, Darfeuille-Michaud A, Prati F, Dalmasso G, and Bonnet R (**2016**). Small-molecule inhibitors prevent the genotoxic

and protumoural effects induced by colibactin-producing bacteria. **Gut** 65(2): 278–285. doi: 10.1136/gutjnl-2014-307241

59. Dahl J-U, Gray MJ, Bazopoulou D, Beaufay F, Lempart J, Koenigsknecht MJ, Wang Y, Baker JR, Hasler WL, Young VB, Sun D, and Jakob U (**2017**). The anti-inflammatory drug mesalamine targets bacterial polyphosphate accumulation. **Nat Microbiol** 2: 16267. doi: 10.1038/nmicrobiol.2016.267

60. Tang-Fichaux M, Chagneau CV, Bossuet-Greif N, Nougayrède J-P, Oswald É, and Branchu P (**2020**). The Polyphosphate Kinase of Escherichia coli Is Required for Full Production of the Genotoxin Colibactin. **mSphere** 5(6): e01195-20. doi: 10.1128/mSphere.01195-20

61. Singh G, Maurya S, DeLampasona MP, and Catalan CAN (**2007**). A comparison of chemical, antioxidant and antimicrobial studies of cinnamon leaf and bark volatile oils, oleoresins and their constituents. **Food Chem Toxicol** 45(9): 1650–1661. doi: 10.1016/j.fct.2007.02.031

62. Kosari F, Taheri M, Moradi A, Hakimi Alni R, and Alikhani MY (**2020**). Evaluation of cinnamon extract effects on clbB gene expression and biofilm formation in Escherichia coli strains isolated from colon cancer patients. **BMC Cancer** 20(1): 267. doi: 10.1186/s12885-020-06736-1

63. Kaewkod T, Tobe R, Tragoolpua Y, and Mihara H (**2021**). Medicinal plant extracts protect epithelial cells from infection and DNA damage caused by colibactin-producing Escherichia coli, and inhibit the growth of bacteria. **J Appl Microbiol** 130(3): 769–785. doi: 10.1111/jam.14817

64. Bernstein CN, Blanchard JF, Kliewer E, and Wajda A (**2001**). Cancer risk in patients with inflammatory bowel disease: a population-based study. **Cancer** 91(4): 854–862. doi: 10.1002/1097-0142(20010215)91:4<854::aid-cncr1073>3.0.co;2-z

65. Kotlowski R, Bernstein CN, Sepehri S, and Krause DO (**2007**). High prevalence of Escherichia coli belonging to the B2+D phylogenetic group in inflammatory bowel disease. **Gut** 56(5): 669–675. doi: 10.1136/gut.2006.099796

66. Chagneau CV, Massip C, Bossuet-Greif N, Fremez C, Motta J-P, Shima A, Besson C, Faouder PL, Cénac N, Roth M-P, Coppin H, Fontanié M, Martin P, Nougayrède J-P, and Oswald E (**2021**). Uropathogenic E. coli induces DNA damage in the bladder. **PLOS Pathogens** 17(2): e1009310. doi: 10.1371/journal.ppat.1009310

67. Wallenstein A, Rehm N, Brinkmann M, Selle M, Bossuet-Greif N, Sauer D, Bunk B, Spröer C, Wami HT, Homburg S, von Bünau R, König S, Nougayrède J-P, Overmann J, Oswald E, Müller R, and Dobrindt U (**2020**). ClbR Is the Key Transcriptional Activator of Colibactin Gene Expression in Escherichia coli. **mSphere** 5(4): e00591-20. doi: 10.1128/mSphere.00591-20

68. Lim CC, Ferguson LR, and Tannock GW (**2005**). Dietary fibres as "prebiotics": implications for colorectal cancer. **Mol Nutr Food Res** 49(6): 609–619. doi: 10.1002/mnfr.200500015

69. Oliero M, Calvé A, Fragoso G, Cuisiniere T, Hajjar R, Dobrindt U, and Santos MM (**2021**). Oligosaccharides increase the genotoxic effect of colibactin produced by pks+ Escherichia coli strains. **BMC Cancer** 21(1): 172. doi: 10.1186/s12885-021-07876-8

70. Tronnet S, Garcie C, Brachmann AO, Piel J, Oswald E, and Martin P (**2017**). High iron supply inhibits the synthesis of the genotoxin colibactin by pathogenic Escherichia coli through a non-canonical Fur/RyhB-mediated pathway. **Pathog Dis** 75(5). doi: 10.1093/femspd/ftx066

71. Tronnet S, Garcie C, Rehm N, Dobrindt U, Oswald E, and Martin P (**2016**). Iron Homeostasis Regulates the Genotoxicity of Escherichia coli That Produces Colibactin. **Infect Immun** 84(12): 3358–3368. doi: 10.1128/IAI.00659-16

72. Martin P, Tronnet S, Garcie C, and Oswald E (**2017**). Interplay between siderophores and colibactin genotoxin in Escherichia coli. **IUBMB Life** 69(6): 435–441. doi: 10.1002/iub.1612

73. Martin P, Marcq I, Magistro G, Penary M, Garcie C, Payros D, Boury M, Olier M, Nougayrède J-P, Audebert M, Chalut C, Schubert S, and Oswald E (**2013**). Interplay between Siderophores and Colibactin Genotoxin Biosynthetic Pathways in Escherichia coli. **PLOS Pathogens** 9(7): e1003437. doi: 10.1371/journal.ppat.1003437

74. Maddocks SE, and Oyston PCF (**2008**). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. **Microbiology** 154(Pt 12): 3609–3623. doi: 10.1099/mic.0.2008/022772-0

75. Anfora AT, and Welch RA (**2006**). DsdX is the second D-serine transporter in uropathogenic Escherichia coli clinical isolate CFT073. **J Bacteriol** 188(18): 6622–6628. doi: 10.1128/JB.00634-06

76. McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, Doan M, Ding L, Rafelski SM, Thirstrup D, Wiegraebe W, Singh S, Becker T, Caicedo JC, and Carpenter AE (**2018**). CellProfiler 3.0: Next-generation image processing for biology. **PLoS Biol** 16(7): e2005970. doi: 10.1371/journal.pbio.2005970

77. Livak KJ, and Schmittgen TD (**2001**). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. **Methods** 25(4): 402–408. doi: 10.1006/meth.2001.1262

78. Roe AJ, Yull H, Naylor SW, Woodward MJ, Smith DGE, and Gally DL (**2003**). Heterogeneous Surface Expression of EspA Translocon Filaments by Escherichia coli O157:H7 Is Controlled at the Posttranscriptional Level. **Infect Immun** 71(10): 5900–5909. doi: 10.1128/IAI.71.10.5900-5909.2003

79. Thomason LC, Sawitzke JA, Li X, Costantino N, and Court DL (**2014**). Recombineering: genetic engineering in bacteria using homologous recombination. **Curr Protoc Mol Biol** 106: 1.16.1-1.16.39. doi: 10.1002/0471142727.mb0116s106

80. Mobley HL, Green DM, Trifillis AL, Johnson DE, Chippendale GR, Lockatell CV, Jones BD, and Warren JW (**1990**). Pyelonephritogenic Escherichia coli and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. **Infect Immun** 58(5): 1281–1289. doi: 10.1128/iai.58.5.1281-1289.1990

81. Nougayrède J-P, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, and Oswald E (**2006**). Escherichia coli induces DNA double-strand breaks in eukaryotic cells. **Science** 313(5788): 848–851. doi: 10.1126/science.1127059

82. Monk IR, Shah IM, Xu M, Tan M-W, and Foster TJ (**2012**). Transforming the Untransformable: Application of Direct Transformation To Manipulate Genetically Staphylococcus aureus and Staphylococcus epidermidis. **mBio** 3(2): e00277-11. doi: 10.1128/mBio.00277-11

83. Datsenko KA, and Wanner BL (**2000**). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. **Proc Natl Acad Sci U S A** 97(12): 6640–6645. doi: 10.1073/pnas.120163297