UNDERSTANDING THE EFFECTS OF INDOLE ON THE INDUCED ACID STRESS RESPONSE IN *ESCHERICHIA COLI*

by

Manpreet Kaur

A thesis submitted to The University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
April 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

E. coli has five acid resistance systems, referred to as AR (AR1 - AR5) systems, which protect cells against low pH. The AR2 system is the most effective system of all AR systems. The AR2 system is activated under two conditions. First, when exponentially growing cells in minimal media are exposed to mildly acidic pH of 5.5 - 5.7 for 30-60 minutes. Second, when stationary phase cells are acid stressed at low pH in the presence of amino acid glutamate. We have previously shown in our lab that the presence of signalling molecule indole at a concentration of 1 mM in the induction medium inhibits the activation of AR2 response in exponentially growing laboratory K-12 strains MG1655 and BW25113. This study shows that the inhibitory effect of indole on the induction of AR2 response in exponentially growing cells is also conserved in pathogenic strain O157:H7 Sakai. Furthermore, the AR2 response can also be highly and significantly induced through incubation at mildly acidic pH in early stationary phase O157:H7 Sakai, but not MG1655, BW25113 or commensal strain Nissle 1917. The induction of AR2 response in early stationary phase O157:H7 Sakai cells, however, cannot be blocked with indole. The regulation of AR2 response has been extensively studied, but mostly in K-12 strain MG1655. Whether these regulatory pathways are conserved in pathogenic and commensal strains is not known. In this study, the Transposon Directed Insertion-site Sequencing (TraDIS) was used to compare the genes of MG1655 and O157:H7 Sakai which are important for resistance to low pH. The frequency of inserts in several genes, which are known to be required for the optimal fitness of cells at low pH, reduced at pH 2.5 but not at pH 7, suggesting that TraDIS can be used to successfully study the effects of low pH and the inhibitory impact of indole on induced acid resistance in *E. coli*. Comparison between the acid stress responses of MG1655 and O157:H7 Sakai using GSEA suggested that same KEGG metabolic pathways determine the fitness of cells at low pH, although subtle differences are present between the responses of two strains. TraDIS results in MG1655 also showed an interesting connection between indole and the function of Fis. The mechanism behind the effect, however, remains to be elucidated.

I dedicate this thesis to my parents.

Mum and Dad, Thanks for everything!

Acknowledgements

I would like to thank BBSRC for funding this project and Professor Chris Thomas for his support. Without him, I would not have been able to continue with my PhD.

To Pete, you have been the best supervisor. I have learnt so much about Science and people management from you. Thank you for having my back during the toughest time of this journey and accepting me in your lab. You are the most supportive and respectful supervisor a PhD student could wish for.

To Mat, thank you for putting up with all the crying and nagging. Your continued support, both scientific and emotional, helped me to conclude this project on time. Thank you for teaching me TraDIS, and an even bigger thank you for teaching me how to be patient in the lab.

To all the members of T101, especially Max, Rob, Chen, Sravanthi, Tom, Georgia, Gabriella, and Dema, your company made this journey less hectic and more enjoyable.

To my parents, thank you for everything! Dad, your lessons about hard work, discipline and honesty have helped me to reach this far in life. Mum, your belief in God and lessons on spirituality motivate me to be a better person every day. Daman, my brother, your presence grounds me and brings a sense of peace and calm.

To all my teachers who taught me in School, College, and University, you laid the foundation of my scientific career. I owe this success to you all!

Table of Contents

Chapter 1 Introduction 1

1.1 *E. coli* as a model organism and strains used in this study..... 2

1.1.1 *E. coli* as a model organism: advantages and limitations 2

1.1.2 Strains used in this study.....3

1.2 The effect of low pH on bacterial cells 6

***1.3* Mechanisms of acid resistance in *E. coli* 8**

1.3.1 Outer membrane.....8

1.3.2 Periplasm.....9

1.3.3 Inner membrane.....9

1.3.4 Cytoplasm.....11

1.3.5 Chemotaxis.....15

1.4 Glutamate dependent AR2 acid resistance system 16

1.4.1 Detection of low pH by EvgS sensor kinase.....16

1.4.2 Regulatory network of AR2 response.....20

1.5 Comparison of induction of AR2 response among different *E. coli* strains..... 29

1.6 Indole signalling and its effect on acid stress 32

1.6.1 Persistent and pulse indole signalling 33

1.6.2 Effect of indole on acid stress.....34

1.6.3 Potential routes for indole transport 36

1.7 Transposon-insertion sequencing as a method to identify essential and conditionally essential genes 36

1.7.1 TIS workflow.....37

1.7.2 Identification of conditionally essential genes.....37

1.7.3 Types of TIS methods 41

1.7.4 TraDIS-Xpress to assay essential genes 41

1.8 Aims and Objectives of this study 42

Chapter 2 Materials and Methods 45

2.1 Bacterial strains, plasmids, and oligonucleotides 46

2.2 Media preparation 52

2.2.1 M9K minimal medium (M9Ksupp medium) 52

2.2.2 Lysogeny Broth (LB broth) 53

2.2.3 2xTY Broth 53

2.2.4 pH measurement 53

2.3 Growth conditions 53

2.4 Growth curve 54

2.5 Acid survival assays 54

2.6 Induction and Indole inhibition assays 55

2.7 Promoter-probe assays 59

2.7.1 Luciferase assay 59

2.7.2 β-galactosidase assay 60

2.8 Gene knockouts 62

2.8.1 P1 phage transduction 62

2.8.2 Datsenko and Wanner method 63

2.9 Molecular biology techniques 64

2.9.1 Plasmid mini-prep 64

2.9.2 Site-directed mutagenesis 65

2.9.3 Electrocompetent cells 65

2.9.4 Electroporation 65

2.9.5 Polymerase Chain Reaction 66

2.9.6 Agarose gel electrophoresis 66
2.9.7 Gel Extraction of DNA fragments 68
2.9.8 PCR Purification 68
2.9.9 DNA quantification 68
2.10 Sequencing 68
2.10.1 Sanger sequencing 68
2.10.2 Whole genome sequencing 69
2.11 Preparation of O157:H7 Sakai TraDIS-Xpress library 69
2.11.1 Transposome preparation 69
2.11.2 Preparation of electrocompetent cells and electroporation 69
2.11.3 DNA extraction and fragmentation for preparation of TraDIS sequencing library 72
2.11.4 TraDIS library preparation for sequencing 72
2.11.5 Preparation of SPRI beads 74
2.11.6 TraDIS sequencing and data analysis 76
2.12 Optimisation for TraDIS experiments 79
2.12.1 Competition experiments 79
2.12.2 Scaled up induction and indole inhibition assays with outgrowth 80
2.13 TraDIS experiments 81
2.14 Analysis of sequencing data from TraDIS experiments 83
2.14.1 Genome alignment of reads 83
2.14.2 Comparison between ITL and conditions using BioTraDIS 84
2.14.3 Gene set enrichment analysis (GSEA) 84
2.15 TraDIS validation experiments 85
2.15.1 Competition experiments 85
2.15.2 Survival assays 85
2.16 Statistical analysis 85

Chapter 3 Promoter probe assays confirm the role of EvgS and the impact of indole on induction of the AR2 response 87

3.1 Introduction 88

3.2 Results 89

3.2.1 Induction of AR2 promoters and inhibition with indole using luciferase assay 89

3.2.2 Indole inhibits the AR2 system possibly through EvgS 94

3.2.3 Modelling of a possible indole binding site within EvgS 98

3.3 Discussion 101

Chapter 4 Comparison of acid resistance and the impact of indole on induction of AR2 response among E. coli strains 105

4.1 Introduction 106

4.2 Results 108

4.2.1 Optimisation of conditions for AR2 induction and inhibition with indole 108

4.2.2 Comparison of survival at pH 2.5 in MG1655 versus MG1655 *ΔrpoS* 110

4.2.3 Induction of acid resistance and effect of indole in exponential phase cells 114

4.2.4 Induction of acid resistance and effect of indole in early stationary phase cells 117

4.2.5 Induction of acid resistance and effect of indole in stationary phase cells 120

4.2.6 Induction of acid resistance and effect of indole in BW25113 grown to exponential, early stationary and stationary phases 121

4.2.7 Induction of acid resistance and effect of indole in O157:H7 Sakai grown to exponential, early stationary and stationary phases 126

4.2.8 Induction of acid resistance and effect of indole in Nissle 1917 grown to exponential, early stationary and stationary phases 130

4.2.9 Comparison of acid resistance profiles among MG1655, BW25113, O157:H7 Sakai, and Nissle 1917 *ΔbcsA* 134

4.3 Discussion 138

Chapter 5 Using Tradis to study the induction of the acid resistance response, and its inhibition with indole, in MG1655 143

5.1 Introduction 144

5.2 Results 146

5.2.1 Optimisation of conditions for TraDIS experiments in *E. coli* K12 MG1655 146

5.2.2 TraDIS experiment results in MG1655 158

5.2.3 Validation of TraDIS experiment results 198

5.3 Discussion 204

Chapter 6 Using Tradis to study the induction of acid resistance response, and its inhibition with indole, in O157:H7 Sakai 207

6.1 Introduction 208

6.2 Results 209

6.2.1 Preparation of TraDIS-Xpress library in O157:H7 Sakai 209

6.2.2 Assessment of essential genes in O157:H7 Sakai TraDIS-Xpress library 211

6.2.3 Pangenome analysis of K-12 strains and O157:H7 Sakai 214

6.2.4 Optimisation of conditions for TraDIS experiments in O157:H7 Sakai 217

6.2.5 Results of TraDIS experiments 221

6.2.6 Comparison between MG1655 and O157:H7 Sakai TraDIS results using GSEA 253

6.3 Discussion 258

Chapter 7 Discussion 260

7.1 Summary of key results 261

7.2 Limitations of the study 263

7.3 Future work 265

Appendices 267

8.1 Whole genome sequencing results of MG1655, BW25113, O157:H7 Sakai, and Nissle 1917 267

8.1.1 Predicted mutations in MG1655 267

8.1.2 Predicted mutations in BW25113 269

8.1.3 Predicted mutations in O157:H7 Sakai 269

8.1.4 Nissle 1917 271

8.1.5 Nissle 1917 *ΔbcsA* 271

8.2 Results of TraDIS experiments in MG1655 271

8.2.1 LogFC, q value and logCPM of all genes in each condition relative to ITL 271

8.2.2 Reproducibility among replicates in each experimental condition 271

8.2.3 Genes with depleted insertions 276

8.2.4 Genes with enriched insertions 279

8.2.5 Comparison of LogFC and q value of genes known to have a role in acid stress, among four conditions 283

8.2.6 List of essential, non-essential and unclear genes in MG1655 283

8.3 Results of TraDIS experiments in O157:H7 Sakai 283

8.3.1 LogFC, q value and logCPM of all genes in each condition relative to ITL 283

8.3.2 Reproducibility among replicates in each experimental condition 283

8.3.3 Genes with depleted insertions 288

8.3.4 Genes with enriched insertions 290

8.3.5 Comparison of LogFC and q value of genes known to have a role in acid stress, among four conditions 293

8.3.6 List of essential, non-essential and unclear genes in O157:H7 Sakai 293

References 294

List of Figures

Figure 1.1: The position of strains used in this study (MG1655, BW25113, O156:H7 Sakai, and Nissle 1917) shown in the phylogenetic tree relative to other *E. coli* strains. **5**

Figure 1.2: Overview of amino acid dependent acid resistance (AR) systems of *E. coli*. **14**

Figure 1.3: Schematic of EvgS domain organisation and homology model of Venus flytrap (VFT) domains. **17**

Figure 1.4: Overview of the regulatory network of AR2 acid resistance response in *E. coli*. **21**

Figure 1.5: Genomic location of *gadBC* operon and acid fitness island (AFI) in *E. coli* K-12 strain MG1655 (NC_ 000913.3). **24**

Figure 1.6: Workflow of Transposon-insertion sequencing (TIS) method for preparation of a transposon mutant library and analysis of insertion sites. **38**

Figure 1.7: Identification of conditionally essential genes using Transposon-insertion sequencing (TIS) method. **39**

Figure 1.8: Genes ranked according to their log_2FC obtained after comparison between the control and conditional TraDIS libraries. **40**

Figure 1.9: TraDIS-Xpress to assay conditional importance of essential genes. **43**

Figure 2.1: Dilutions plated out using tilt plate method in acid survival assay. **56**

Figure 2.2: Workflow of Induction and Indole inhibition assays. **57**

Figure 2.3: Map of Tnp001, a mini-Tn5 transposon. **70**

Figure 2.4: Steps involved in the preparation of a TraDIS-Xpress library for Illumina sequencing. **73**

Figure 3.1: Growth profiles and fold induction of AR2 promoters measured using luciferase promoter-probe assay at pH 7 and 5.5 with and without indole. **93**

Figure 3.2: The effect of indole on the constitutive activity of EvgS-S600I. **95**

Figure 3.3: Competition of MG1655 and MG1655 Δ*evgS* against KH001 under AR2 induction conditions in the presence of 1 mM indole and the same volume of ethanol. **97**

Figure 3.4: Model showing indole (purple) bound within the EvgS PAS domain. **100**

Figure 3.5: Measurement of β-galactosidase activity from *ydeP* promoter to test the potential binding site of indole within EvgS PAS domain. **102**

Figure 4.1: Survival of MG1655 at pH 2.5 in exponential, early stationary and stationary phases using different incubations times and plating methods. **109**

Figure 4.2: Comparison of innate acid resistance of MG1655 with MG1655 Δ*rpoS* cultured and acid stressed at pH 2.5 in M9Ksupp. **112**

Figure 4.3: Comparison of innate acid resistance of MG1655 with MG1655 *ΔrpoS* cultured and acid stressed at pH 2.5 in LB. **113**

Figure 4.4: Comparison of innate acid resistance of MG1655 and MG1655 *ΔrpoS* cultured and acid stressed at pH 2.5 in M9Ksupp and LB. **115**

Figure 4.5: Induction of acid resistance and inhibition with indole in exponentially growing MG1655. **116**

Figure 4.6: Induction of acid resistance and inhibition with indole in MG1655 grown to early stationary phase. **119**

Figure 4.7: Induction of acid resistance and inhibition with indole in MG1655 grown to stationary phase. **122**

Figure 4.8: Induction of acid resistance and inhibition with indole in BW25113 in exponential, early stationary and stationary growth phases. **125**

Figure 4.9: Induction of acid resistance and inhibition with indole in O157:H7 Sakai in exponential, early stationary and stationary growth phases. **128**

Figure 4.10: Induction of acid resistance and inhibition with indole in Nissle 1917 *ΔbcsA* in exponential, early stationary and stationary growth phases. **132**

Figure 4.11: Comparison of survival among MG1655, BW25113, O157:H7 Sakai and Nissle 1917 *ΔbcsA* after being acid stressed at pH 2.5 with and without induction at pH 5.5. **135**

Figure 4.12: Comparison of survival of Nissle 1917 *ΔbcsA* in the presence and absence of constitutively active EvgS protein from MG1655, after being acid stressed at pH 2.5 with and without prior induction at pH 5.5. **137**

Figure 5.1: Competition of MG1655 or MG1655 *ΔevgS* against KH001 under AR2 induction conditions in the absence or presence of 1 mM indole, followed by different acid shock durations at pH 2.5. **149**

Figure 5.2: Competition of MG1655 Δ*evgS* against KH001 under AR2 induction conditions in the absence and presence of 1 mM indole, followed by outgrowth. **152**

Figure 5.3: MG1655 induced in the presence of 1 mM indole, acid shocked at pH 2.5 and outgrown in the presence and absence of indole for 3 hours. **156**

Figure 5.4: Workflow of TraDIS library preparation and comparison between control and contiditional TraDIS libraries. **160**

Figure 5.5: Survival of MG1655 transposon library induced either in the presence of 1 mM indole or the same amount of ethanol followed by exposure to pH 2.5 or pH 7. **163**

Figure 5.6: Comparison of total number of read counts (A) and unique insertion sites (UIS) (B) among initial TraDIS library (ITL) and the TraDIS libraries prepared under different conditions. **165**

Figure 5.7: Clustering among replicates of same and different conditions visualised through principal component analysis done on read counts per gene normalised by sequencing depth. **166**

Figure 5.8: Comparison between RPM of two TraDIS libraries prepared from two separate biological replicates. **168**

Figure 5.9: Comparison between insertion index scores of two TraDIS libraries prepared from two separate biological replicates. **169**

Figure 5.10: Volcano plots showing the $\log_2$ fold-change and the statistical significance corrected for multiple comparisons (q value) of the mutants containing transposon insertions within different genes. **171**

Figure 5.11: Venn diagrams showing genes where the number of insertions had either decreased (depleted insertions) (A), or increased (enriched insertions) (B), in the conditional TraDIS libraries as compared to the ITL. **183**

Figure 5.12: MG1655 *ΔmrcB* (A), MG1655 *ΔgadE,* MG1655 *ΔmrcA,* and MG1655 *ΔnlpI* (B) deletion mutants prepared using P1 transductions. **199**

Figure 5.13: Validation of TraDIS experiment results using competition and survival assays. **201**

Figure 6.1: The position of unique transposon insertion sites and reproducibility between two biological replicates of O157:H7 Sakai TraDIS sequencing libraries. **212**

Figure 6.2: Distribution and frequency of insertion index score of genes from O157:H7 Sakai TraDIS-Xpress library. **213**

Figure 6.3: Pangenome analysis performed on *E. coli* strains. **215**

Figure 6.4: Survival of O157:H7 Sakai under AR2 induction conditions in the presence of 1 mM indole or the same volume of ethanol, followed by acid stress at pH 2.5 for different durations. **218**

Figure 6.5: Survival of O157:H7 Sakai under AR2 induction conditions in the presence of 1 mM indole or the same volume of ethanol, followed by outgrowth. **220**

Figure 6.6: Survival of O157:H7 Sakai transposon library induced either in the presence of 1 mM indole or the same volume of ethanol followed by exposure to pH 2.5 or pH 7. **223**

Figure 6.7: Comparison of total number of read counts (A) and unique insertion sites (UIS) (B) among initial TraDIS library (ITL) and the TraDIS libraries prepared under different conditions. **225**

Figure 6.8: Clustering among replicates of same and different conditions visualised through principal component analysis done on read counts per gene normalised by sequencing depth. **226**

Figure 6.9: Comparison between RPM of two TraDIS libraries prepared from two separate biological replicates. **228**

Figure 6.10: Comparison between insertion index scores of two TraDIS libraries prepared from two separate biological replicates. **229**

Figure 6.11: Volcano plots showing the log fold-change and the statistical significance corrected for multiple comparisons (q value) of the mutants containing transposon insertions within different genes. **232**

Figure 6.12: Venn diagrams showing genes where the number of insertions had either decreased (depleted insertions) (A), or increased (enriched insertions) (B), in the conditional TraDIS libraries as compared to the ITL. **244**

List of Tables

Table 1.1: Examples of studies which have compared the induction of AR2 response in different *E. coli* strains in exponential and stationary growth phases. **30**

Table 2.1: Strains used in this study. **47**

Table 2.2: Plasmids used in this study. **49**

Table 2.3: Oligonucleotides used in this study. **50**

Table 2.4: Composition of a 50 µL PCR reaction mixture. **67**

Table 2.5: Thermocycler conditions routinely used for PCR reactions. **67**

Table 2.6: Reagents used in the first and second PCR reactions during TraDIS-Xpress library preparation. **75**

Table 2.7: Thermocycler conditions used for the first and second PCR reactions during TraDIS-Xpress library preparation. **75**

Table 2.8: The conditions under which TraDIS experiments were performed with MG1655 and O157:H7 Sakai. **82**

Table 5.1: Comparison of log_2FC and q values of mutants containing transposon within the AR2 genes in four conditions under which TraDIS experiments were performed. **175**

Table 5.2: Comparison of log_2FC and q values of mutants containing transposon within the genes which have been tested for their role in acid stress, in four conditions under which TraDIS experiments were performed. **180**

Table 5.3: Comparison of log_2FC and q values of transposon mutants whose frequency reduced (depleted insertions) under at least one TraDIS experimental condition compared to the ITL. **185**

Table 5.4: Comparison of log_2FC and q values of transposon mutants whose frequency increased (enriched insertions) under at least one TraDIS experimental condition compared to the ITL. **186**

Table 5.5: KEGG metabolic pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions relative to the ITL. **194**

Table 5.6: Transcriptional regulatory pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions relative to the ITL. **195**

Table 5.7: A summary of results from TraDIS experiments done with MG1655 transposon library and the validation experiments (competition and survival) done with MG1655 single gene mutants. **205**

Table 6.1: The number of genes designated as essential, non-essential, or unclear in three *E. coli* strains, MG1655 (Milner, 2021), BW25113 (Goodall *et al.,* 2018) and O157:H7 Sakai (this study) determined using TraDIS. **216**

Table 6.2: The number of essential genes which are present in all three genomes (core genome) or absent in at least one genome (accessory genome). **216**

Table 6.3: Comparison of log_2FC and q values of mutants containing transposon within the AR2 genes in four conditions under which TraDIS experiments were performed. **235**

Table 6.4: Comparison of log_2FC and q values of mutants containing transposon within the genes which have been tested for their role in acid stress, in four conditions under which TraDIS experiments were performed. **240**

Table 6.5: Comparison of log_2FC and q values of transposon mutants whose frequency reduced (depleted insertions) under at least one TraDIS experimental condition compared to the ITL. **246**

Table 6.6: Comparison of log_2FC and q values of transposon mutants whose frequency increased (enriched insertions) under at least one TraDIS experimental condition compared to the ITL. **247**

Table 6.7: KEGG metabolic pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions relative to the ITL. **250**

Table 6.8: Transcriptional regulatory pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions relative to the ITL. **251**

Table 6.9: KEGG metabolic pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions and at least one of the strains, relative to the ITL. **255**

Table 6.10: Transcriptional regulatory pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions and at least one of the strains, relative to the ITL. **256**

Chapter 1

Introduction

1.1 *E. coli* as a model organism and strains used in this study

1.1.1 *E. coli* as a model organism: advantages and limitations

E. coli is a Gram-negative bacterium and belongs to Enterobacteriaceae family. It is a facultative aerobe, meaning it can grow with and without oxygen. Its natural environment includes all warm-blooded animals, that is, birds and mammals, where it co-exists with members of 200 - 1,000 other bacterial species and constitutes 0.1-5% of the community (Tenaillon *et al.*, 2010). Outside the mammalian host, it can also be found in soil, water, and food.

E. coli was discovered in 1884 by a German microbiologist Theodor Escherich (Escherich, 1988). Since then, it has been used in many foundational studies in the fields of molecular biology, bacterial physiology, genetics, pharmaceuticals, evolution, genetic engineering and biotechnology, making it one of the most well understood organisms (Blount, 2015). The ability of *E. coli* to grow quickly and on many different nutrients, ease of handling due to simple growth requirements, availability of many non-pathogenic and versatile strains, and availability of tools for its genetic manipulation are some of the reasons for its extensive use in research.

The pangenome of *E. coli* has been extensively studied. A pangenome of a species refers to all the genes which are present in all members of that species. It consists of a core genome and an accessory genome. A core genome is defined as the set of genes which are present in all members of a species whereas an accessory genome is defined as the set of genes which are absent in at least one member of the species. Pangenome analysis done on 1,324 *E. coli* genomes has revealed that ~25,000 genes are present in all *E. coli* genomes combined (Tantoso *et al.*, 2022). Yet, only ~3,000 (12%) of these genes are present in ≥95% of the genomes. The remaining 22,000 genes are part of the accessory genome. Due to the

genetic variation present among *E. coli* genomes, it is worth thinking whether the results obtained with laboratory *E. coli* strains such as K-12 represent what happens in other real-world *E. coli* strains.

It is a common practice in research to use only laboratory strains such as K-12 for most of the studies, which have evolved for generations to survive in the lab environment (Ruiz and Silhavy, 2022). These strains are usually studied under conditions which are much different to their natural habitats and without the selection pressures which they face in the wild (Hobman, Penn and Pallen, 2007). Whole genomes sequencing done on two ancestral K-12 strains, WG1 and EMG2, has shown that the K-12 strains currently being used in the labs have undergone extensive laboratory-based evolution, suggesting that *E. coli* K-12 may no longer be a suitable model organism (Browning, Hobman and Busby, 2023).

1.1.2 Strains used in this study

In this study, the acid resistance response of four *E. coli* strains including two laboratory K-12 strains MG1655 and BW25113, one pathogenic strain O157:H7 Sakai, and one commensal strain Nissle 1917 was studied. *E. coli* is a neutralophilic bacterium and maintains its cytoplasmic pH around 7.2. However, being an enteric organism, it is routinely exposed to strong acid stress on its passage through stomach where the pH ranges between 1.5 and 3.5 (Lund, Tramonti and De Biase, 2014). Other environments where *E. coli* experiences acid stress include macrophage phagosomes, dental plaque, fermented foods, proximal gut due to presence of short chain fatty acids and acidic soils. Due to this, the acid resistance mechanisms of *E. coli* have been extensively studied, especially in K-12 strain MG1655, which are discussed in later sections (1.2 and 1.3).

The original K-12 strain was isolated in 1922 from the stool sample of a patient with diphtheria. MG1655 is a 2-step descendent of this original strain because it has been cured of its lambda prophage and F plasmid (Figure 1.1) (Riley *et al.*, 2006).

The second strain, BW25113, is another commonly used laboratory K-12 strain. It was created in the laboratory of Barry L. Wanner and is the parent strain used in the construction of Keio collection, which consists of 4,000 single-gene deletion mutants (Grenier *et al.*, 2014).

The third strain, O157:H7 Sakai, is an enterohemorrhagic strain (EHEC). It produces a Shiga-like toxin which causes bloody diarrhoea. This strain was isolated from the 1996 outbreak in Sakai city of Japan, which occurred due to consumption of raw radish sprouts by school children (Hayashi *et al.*, 2001). The strain contains two plasmids, pOSAK1 and pO157, as well as two shiga-toxin encoding genes *stx1* and *stx2.* The *stx* genes were later removed, making this strain safer to work with in the lab. The pathogenic *E. coli* associated with intestinal infections have been classified into seven groups called pathotypes depending on where and how they cause disease in the host. These pathotypes are: enterohemorrhagic (EHEC), enteropathogenic (EPEC), enterotoxigenic (ETEC), enteroinvasive (EIEC), enteroaggregative (EAEC), diffusely-adherent (DAEC), and adherent-invasive (AIEC) *E. coli* (reviewed in Robins-Browne *et al.*, 2016). Out of all these, EHEC O157:H7 has been associated with most outbreaks, which results in many hospitalisations and deaths, costing millions of dollars in healthcare (Frenzen, Drake and Angulo, 2005; Rangel *et al.*, 2005). O157:H7 is a part of the normal gut microbiome of cattle and other livestock. However, it can be transferred to humans through consumption of uncooked meat or fruits and vegetables contaminated through water and manure. Therefore, how well O157:H7 survives the acid stress in stomach to reach the small intestine to cause disease is an important question for health and economic reasons.

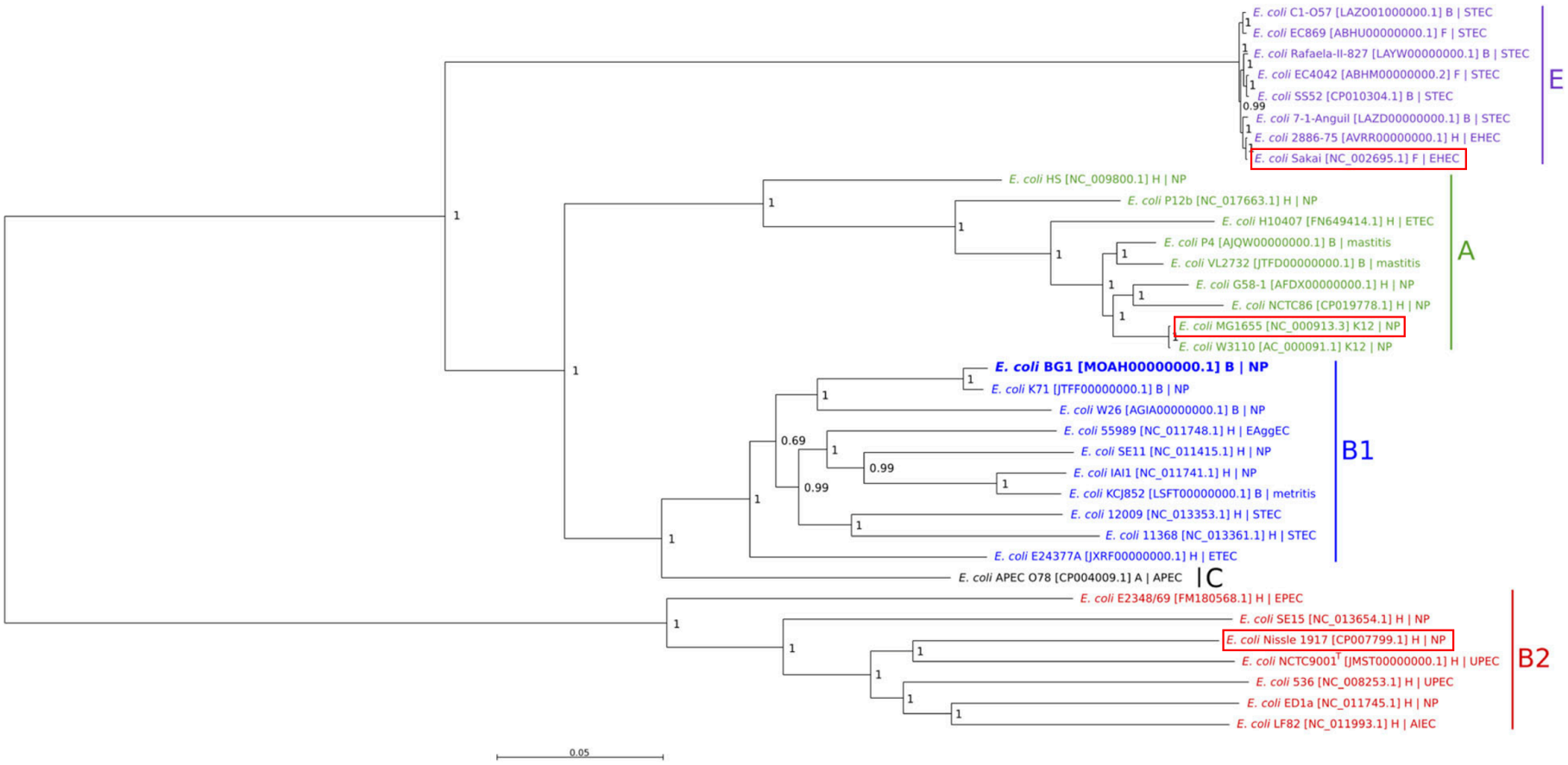


Figure 1.1: The position of strains used in this study (MG1655, BW25113, O156:H7 Sakai, and Nissle 1917) shown in the phylogenetic tree relative to other *E. coli* strains.

E. coli strains within five different phylogroups (A, B1, B2, C and E) are shown. The strains have been identified as H (Human), B (Bovine), A (Avian), F (Food) or K-12 (Laboratory). Furthermore, the clinical or non-pathogenic (NP) characteristics of strains are also mentioned. MG1655, O157:H7 Sakai and Nissle 1917 have been shown in red boxes. BW25113, which is not shown in the phylogenetic tree, belongs to phylogroup A. (Adapted from Segura *et al.* (2017); Figure 2)

The fourth strain, Nissle 1917, was isolated by Alfred Nissle in 1917 from the stool of a German soldier in First World War (Sonnenborn, 2016). This soldier did not develop any intestinal infection during Shigella outbreak which infected other soldiers. After isolating, Alfred Nissle showed that this strain displays antagonistic activity against pathogenic enterobacteria, and therefore used it to treat patients. In 1980s, pharmaceutical company Ardeypharm acquired the rights to sell this strain as probiotic capsules by the name of Mutaflor.

1.2 The effect of low pH on bacterial cells

Bacterial cells can be exposed to strong and weak acids. In an aqueous solution, an acid can dissociate to form its conjugate base and a proton, which for HCl can be written as follows:

$$\underset{\text{Acid}}{HCl} + \underset{\text{Base}}{H_2O} \rightleftharpoons \underset{\text{Conjugate base}}{Cl^-} + \underset{\text{Conjugate acid}}{H_3O^+} \qquad \text{(equation 1)}$$

When dissolved in water, strong acids such as HCl dissociate completely whereas weak acids such as acetic acid dissociate only partially. The strength of an acid is indicated by its acid dissociation constant, K_a. When the K_a is large (>>1), an acid dissociates completely and when the K_a is small (<), the dissociation is partial. K_a values can range from very small to very large, therefore pK_a is used which is calculated as follows:

$$pK_a = -\log_{10} K_a$$

Smaller the pK_a value, the stronger an acid is and more dissociated it is when dissolved in water. Strong acids have pK_a values that range between -15 and 1, and weak acids have pK_a values that range between 2 and 50. When the pH equals the pK_a of an acid, half of the acid molecules are protonated and half are deprotonated. When the pH is above the pK_a of an

acid, more of the acid molecules are deprotonated, that is, the equilibrium of equation 1 shifts to the right and vice versa.

The effect of weak acids on bacterial cells is more complex than the effect of strong acids. Ionised weak acids can associate with the protons in the periplasm and carry them into the cytoplasm when they pass through the inner membrane. This can collapse the proton gradient across the inner membrane affecting ATP synthesis through F_1F_o-ATPase. The unionised weak acids can cross the inner membrane and dissociate in the cytoplasm reducing the cytoplasmic pH due to release of protons. The weak acids that remain unionised in the cytoplasm may also inhibit cell growth (Salmond, Kroll and Booth, 1984).

When cells are exposed to strong acids, protons can cross the outer membrane through passive diffusion and due to the presence of a large number of porins. The pH in periplasm therefore drops rapidly to the same levels as the external medium (Wilks and Slonczewski, 2007). The cytoplasmic membrane, however, is less porous and reduces the flow of protons into the cytoplasm. But, some protons can enter the cytoplasm through protein transporters, transient water chains and damaged membranes (Deamer, 1987; Foster, 2004). If the pH of external medium is moderately low, cells show a transient drop in cytoplasmic pH followed by a rapid recovery in less than 4 minutes (Wilks and Slonczewski, 2007). This is likely because of the intrinsic buffering capacity of the cytoplasm due to the presence of cellular components such as amino acids, polyamines, and polyphosphates which can sequester protons. Changes in ion flux could be another reason for rapid recovery in cytoplasmic pH. The rapid recovery in internal pH is not possible when cells are exposed to extreme acid stress. Under these conditions, more effective acid stress resistance mechanisms discussed in section 1.4 become crucial and even when these mechanisms are in operation, the cytoplasmic pH can fall to <5 (Richard and Foster, 2004).

Low cytoplasmic pH can have several deleterious effects on the cells (reviewed in Lund, Tramonti and De Biase, 2014). If pH falls below the optimum pH of an enzyme, it will lead to reduced enzymatic activity affecting essential cellular processes. Furthermore, low pH can lead to protein unfolding in cytoplasm and periplasm, and cause damage to outer and inner membranes and DNA. Extreme acid stress results in DNA strand breaks and loss of purines (Lindahl and Nyberg, 1972).

1.3 Mechanisms of acid resistance in *E. coli*

As discussed above, *E. coli* is well equipped to deal with acid stress. This is reflected by their low infectious dose (< 10^2 cells) (Lin *et al.*, 1996). Almost all cell components such as cytoplasm, inner membrane, periplasm and outer membrane confer resistance against low pH, which will be discussed in detail in this section. This topic has also been reviewed in great detail by Kanjee and Houry (2013); Lund, Tramonti and De Biase (2014); Guan and Liu (2020); Li *et al.* (2020); and Schwarz *et al.* (2022).

1.3.1 Outer membrane

The outer membrane is the first cellular component to encounter any external stress including low pH. The outer membrane is an asymmetric bilayer whereby the inner leaflet is made up of phospholipids and the outer leaflet is constituted by lipopolysaccharide (LPS). LPS is highly negatively charged due to the presence of phosphates. Under acid stress, bacterial cells, such as *Salmonella,* neutralise these negative charges by masking them with positively charged divalent cations such as Mg^{2+} which reduces the influx of protons into the cells through passive diffusion (Groisman *et al.*, 2013). Another route for protons to gain entry inside the cells is through outer membrane porins. However, *E. coli* can block these porins

using molecules such as polyphosphate, spermine and cadaverine, or reduce their pore size inhibiting the proton influx (Samartzidou and Delcour, 1999; Iyer *et al.*, 2000).

1.3.2 Periplasm

The periplasm has lower buffering capacity as compared to the cytoplasm due to presence of less cellular components which can sequester protons. The proteins in periplasm, therefore, can easily get denatured due to low external pH. This is prevented by two chaperone proteins HdeA and HdeB present in the periplasm. Each protein gets activated at a different pH, with HdeA active at pH 1-3, and HdeB at pH 3-5 (Dahl *et al.*, 2015). At neutral pH, both HdeA and HdeB exist as dimers and dissociate into monomers at low pH. They prevent the aggregation of periplasmic proteins during acid stress and their client proteins also include other periplasmic chaperones such as SurA which assist HdeA and HdeB in recovering other client proteins from damage endured during acid stress (Zhang *et al.*, 2011). Furthermore, peptidoglycan also protects the cells as mutants which have defects in peptidoglycan structures, such as PBP1b (DD-carboxypeptidase), MltA (lytic transglycosylase) and MepS (endopeptidase), grown in acidic medium show reduced survival (Mueller *et al.*, 2019). The authors have suggested that the possible reasons for this observation could include pH-dependent changes in protein structure, substrate binding, catalysis or reduced affinity for other protein partners.

1.3.3 Inner membrane

Inner membrane does not contain porins, therefore membrane phospholipids present as the main barrier to protons. Membrane phospholipids consist of three components with phosphate and glycerol constituting the hydrophilic head region and fatty acids form the

hydrophobic tail. Although inner membrane has a reduced fluidity relative to the outer membrane, cells can further decrease the membrane permeability to reduce proton influx by increasing the concentration of saturated fatty acids over unsaturated, and by converting unsaturated fatty acids to cyclopropane fatty acids (CFAs) using CFA synthase (Chang and Cronan, 1999; Yuk and Marshall, 2004).

Under aerobic growth conditions and mild acid stress at pH 5-5.7, the components of electron transport chain are upregulated (Maurer *et al.*, 2005; Slonczewski *et al.*, 2009). The electron transport chain generates proton motive force by coupling metabolic redox reactions with proton export. The upregulation of these components during acid stress, therefore, might increase the rate of proton export. Mild acid stress also leads to upregulation of genes involved in transport and metabolism of secondary carbon sources which produce less acids during metabolism than glucose (Maurer *et al.*, 2005; Hayes *et al.*, 2006). Under anaerobic growth conditions and mild acid stress, the induction of FHL complex results in conversion of protons present in the cytoplasm to hydrogen gas, helping to raise cytoplasmic pH (Hayes *et al.*, 2006; Noguchi *et al.*, 2010).

Cells maintain a negative membrane potential inside the cells at neutral pH. However, under extreme acid stress, it is reversed due to the accumulation of positively charged compounds produced during decarboxylation reactions as well as increased influx of protons inside the cells (Richard and Foster, 2004). A positive membrane potential inside the cells could help by reducing the further influx of protons. Once the acid stress is over, ClC which is a H^+/Cl^- antiporter, helps to correct the membrane potential by expelling protons out of the cell in exchange for chloride ions (Iyer *et al.*, 2002; Gut *et al.*, 2006).

F_1F_o-ATPase is also thought to reduce acid stress. During respiration, the electron transport chain exports the protons into the periplasmic space generating proton motive

force. F_1F_o-ATPase couples the energy released due to the flow of protons from periplasmic space into the cells with ATP synthesis. The F_o component of the protein complex is present within the cytoplasmic membrane and is involved in proton translocation. The F1 part is present within the cytoplasm where it catalyses the conversion of ADP and inorganic phosphate to produce ATP. This reaction can also be reversed, ATP can be hydrolysed, and the energy released can be used to expel protons out of the cytoplasm. The mechanism by which F_1F_o-ATPase helps to raise cytoplasmic pH is unclear. But it has been shown that mutants deficient in F_1F_o-ATPase have a lower survival and internal pH than the wild-type cells even if other acid resistance systems such as AR2 or AR3 (discussed in more detail in section 1.3.4.2) were active in the cells (Sun *et al.*, 2012b). There are two possibilities by which F_1F_o-ATPase could be improving the survival of cells under acid stress. It could either expel the protons which have entered the cytoplasm into the periplasmic space or it could help by producing ATP which can then be used by other cell components to repair damage incurred during acid stress. Indeed, it has been shown that the deletion of genes which are involved in ADP synthesis reduces the acid resistance of *E. coli* (Sun *et al.*, 2011). Furthermore, these mutants experienced a higher DNA damage as compared to the wild-type cells. These results suggest that ATP-dependent metabolic processes are necessary to improve the survival of *E. coli* under extreme acid stress even if cells are not actively growing during this time.

1.3.4 Cytoplasm

The cytoplasm can reduce the internal pH through its intrinsic buffering capacity as mentioned in section 1.3. Additionally, protein chaperones such as GroE, GroL, DnaJ, DnaK, and Hsp31, as well as DNA binding protein Dps protect proteins and DNA against damage from low pH (Choi, Baumler and Kaspar, 2000; Houry, 2001; Mujacic and Baneyx, 2007). The

deaminase and acid resistance (AR) systems present in the cytoplasm also provide resistance against low pH. These are described in more detail in the sections below.

1.3.4.1 Deaminase systems

Deaminases are catabolic enzymes which break down a substrate to release ammonia (NH_3). *E. coli* has two deaminases: glutaminase GlsA, and adenosine deaminase Add (Sun *et al.*, 2012a; Lu *et al.*, 2013). GlsA catabolises L-glutamine into NH_3 and glutamate (Lu *et al.*, 2013). Glutamate then eventually serves as the substrate for GadA and GadB decarboxylase enzymes of AR2 system (section 1.4.2.1). Add, the second deaminase, catalyses the breakdown of adenosine into inosine and NH_3 (Sun *et al.*, 2012a). The reaction involving adenosine consumes a proton, as shown below, which helps to raise the cytoplasmic pH.

$$\text{Adenosine} + H^+ \rightarrow \text{Inosine} + NH_4^+$$

Furthermore, NH_3 produced in both the above reactions associates with a proton to produce ammonium ions (NH_4^+). This also helps to increase the cytoplasmic pH due to consumption of protons.

1.3.4.2 Acid resistance (AR) systems

E. coli has five different acid resistance (AR) systems which are activated during different growth stages and pH ranges. All AR systems (AR2 - AR5) except for AR1 are reliant on presence of a specific amino acid in growth medium. The amino acid dependent AR systems function using a decarboxylase and an antiporter. The evidence that bacterial decarboxylases are amino acid-inducible was first presented in 1940s by Ernest Gale (Gale, 1940). Since then, these systems have been extensively studied for their ability to maintain cytoplasmic pH and the regulation involved in their activation. Within each amino acid-

inducible AR system, the amino acid is imported using the antiporter and metabolised using a pyridoxal-5'-phosphate (PLP)-dependent decarboxylase (Figure 1.2). The consumption of a proton during this reaction coupled with the production of a more alkaline product than the substrate raises the cytoplasmic pH. The protonated product is eventually exported out of the cell by the antiporter.

The AR1 system is the most poorly understood AR system out of all five. This system gets activated when stationary phase cells cultured in complex medium buffered at pH 5.5 are acid stressed in either minimal or rich medium at low pH (Lin *et al.*, 1996; Castanie-Cornet *et al.*, 1999). Furthermore, presence of glucose in growth medium represses AR1. The activation of AR1 is regulated by the stationary phase sigma factor RpoS and cyclic adenosine monophosphate (cAMP) receptor protein (CRP). The function of F_1F_o-ATPase also provides protection during AR1 activation (Richard and Foster, 2004).

The AR2 system gets activated when exponentially growing cells are exposed to mildly acidic conditions of pH 5.5 - 5.7, in minimal medium supplemented with glutamate (Ma, Masuda and Foster, 2004; Burton *et al.*, 2010; Eguchi and Utsumi, 2014). Additionally, it also gets activated in stationary phase, anaerobic growth conditions and under salt stress (De Biase *et al.*, 1999). The AR2 system is most effective at dealing with acid stress among all AR systems and it is vital for survival at pH levels of 2.5 and below (Lin *et al.*, 1996). The AR2 response consists of two isoforms of glutamate decarboxylase, GadA and GadB, and a glutamate/γ-amino butyric acid (GABA) antiporter GadC. The regulation of AR2 response is highly complex and is discussed in more detail in section 1.4.

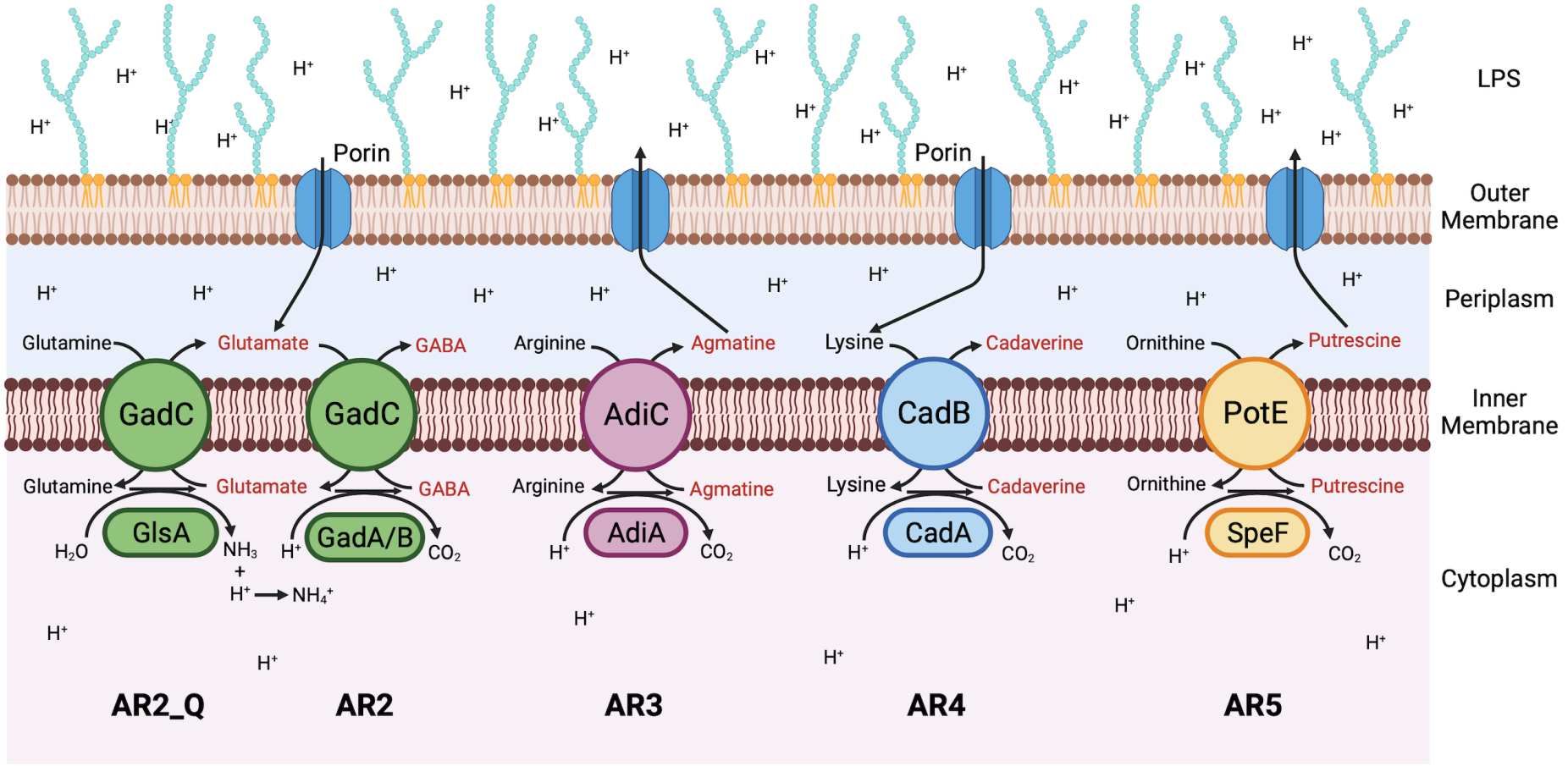


Figure 1.2: Overview of amino acid dependent acid resistance (AR) systems of *E. coli*.
The deaminase GlsA, amino acid decarboxylases (GadA/B, AdiA, CadA and SpeF), antiporters (GadC, AdiC, CadB and PotE), along with the substrate consumed and the product produced by each AR system is shown. (Adapted from Kanjee and Houry (2013); Figure 3a. Created with BioRender.com.)

The AR3 system is induced when cells cultured in rich medium under anaerobic conditions are exposed to low pH in the presence of arginine (Lin *et al.*, 1996; Gong, Richard and Foster, 2003). The AR3 components include arginine decarboxylase AdiA, and arginine/agmatine antiporter AdiC.

The AR4 system gets activated when cells are exposed to low pH in the presence of lysine (Watson *et al.*, 1992; Dell, Neely and Olson, 1994). The AR4 components include lysine decarboxylase CadA and lysine/cadaverine antiporter CadB.

The AR5 system is the only AR system which utilises a non-proteinogenic amino acid. It gets activated under anaerobic conditions and provides protection against low pH in the presence of high ornithine levels (Applebaum, Dunlap and Morris, 1977; Kashiwagi *et al.*, 1992). The components of AR5 system include ornithine decarboxylase SpeF and ornithine/putrescine antiporter PotE.

1.3.5 Chemotaxis

Chemotaxis is a phenomenon whereby cells move in response to a chemical stimulus. Bacterial cells can physically escape low or high pH environments and move to environments with neutral pH. *E. coli* has two methyl-accepting chemotaxis proteins (MCPs), the aspartate binding receptor Tar, and the serine binding receptor Tsr to achieve this goal (Repaske and Adler, 1981; Krikos *et al.*, 1985; Khan *et al.*, 1995). Under Tar activation, cells move towards low pH and repel high pH, whereas Tsr produces an opposing motion. Furthermore, upregulation in flagellar genes is observed at low pH, suggesting that bacteria might use their flagella to increase their movement during acid stress (Maurer *et al.*, 2005).

1.4 Glutamate dependent AR2 acid resistance system

1.4.1 Detection of low pH by EvgS sensor kinase

Bacteria use two-component signal transduction systems (TCSs) to sense changes in their environment and in return adjust cellular processes for adaptation. TCSs consist of a sensor kinase which senses the environmental signal and a response regulator which changes the expression of target genes allowing the cell to respond to the signal. EvgS is a part of the TCS where EvgS is the sensor kinase and EvgA is the response regulator. This system is homologous to BvgS/BvgA of *Bordetella pertussis* (the causative agent of whooping cough) and KvgS/KvgA of *Klebsiella pneumoniae* (Lai *et al.*, 2003; Herrou *et al.*, 2010; Sen *et al.*, 2017). The function of EvgS is crucial for the activation of AR2 acid resistance response in exponentially growing *E. coli* (Masuda and Church, 2003; Itou, Eguchi and Utsumi, 2009). Several conditions are necessary for maximum activation of EvgS. First, the cells must be exposed to a moderate pH of between 5.5 and 5.7 for 30-60 minutes in minimal medium (Burton *et al.*, 2010; Eguchi and Utsumi, 2014; Johnson *et al.*, 2014). Second, a high concentration (≥ 150 mM) of alkali metal ions such as Na^+ and K^+ should be present in the induction medium (Eguchi and Utsumi, 2014). Third, the activation only occurs under aerobic growth conditions and not anaerobic growth conditions (Inada *et al.*, 2021). Fourth, the presence of oxidised ubiquinone, which is an electron acceptor and a part of the electron transport chain, is necessary (Inada *et al.*, 2021).

Under non-activating conditions, EvgS forms a tight dimer which is composed of a periplasmic domain, a transmembrane domain and a cytoplasmic domain (Figure 1.3) (Johnson *et al.*, 2014; Sen *et al.*, 2017). The periplasmic domain consists of two venus flytrap domains, each of which further contains two subdomains. The venus flytrap domains are

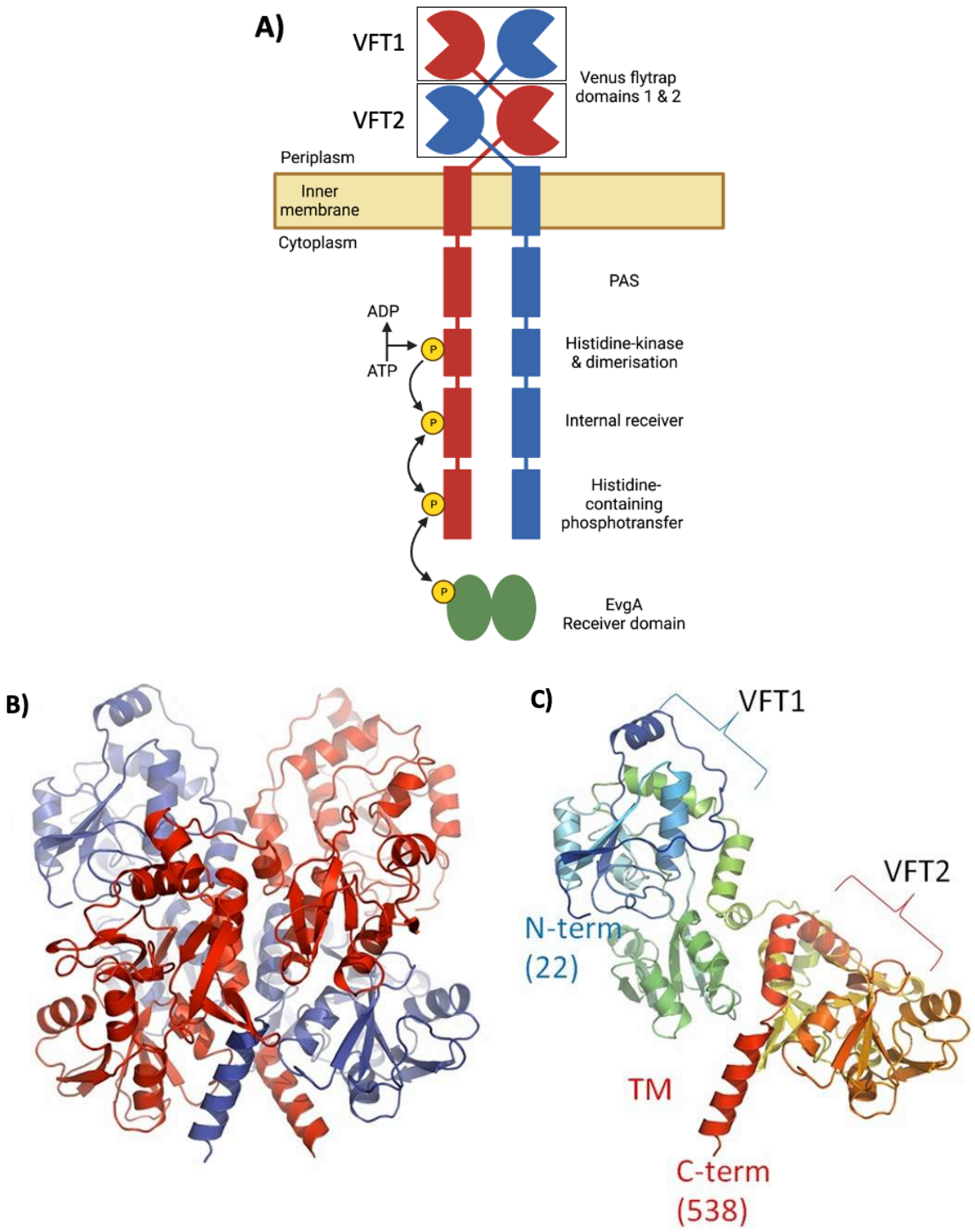


Figure 1.3: Schematic of EvgS domain organisation and homology model of Venus flytrap (VFT) domains.

(A) EvgS domain organisation including the internal phosphorelay involved in EvgS activation are shown. (B) Homology model of EvgS periplasmic domain which is composed of two VFT domains, predicted using BvgS structure as the primary model template. (C) A single protomer showing the VFT domain organisation from residues 22 to 538 of the protein. (Figure 1.2A: Adapted from Sen *et al.* (2017); Figure 1. Created with Biorender.com. Figures 1.2B & C: Taken from Sen *et al.* (2017); Figures 1B & C respectively.)

linked to the cytoplasmic domain through a flexible linker which traverses the inner membrane. The cytoplasmic domain consists of a PAS domain, a histidine-kinase and dimerisation domain, an internal receiver domain, and a histidine-containing phosphotransfer domain.

EvgS is one of the five unorthodox kinases in *E. coli* (Lund, Tramonti and De Biase, 2014). On receiving the activating signal, the conventional sensor kinases autophosphorylate at an internal histidine residue and then transfer the phosphate to an aspartate residue in the cognate response regulator. EvgS, however, has an internal phosphorelay whereby the phosphate is transferred from histidine (His721) to an internal aspartate (Asp1009), and then to a second histidine (His1137) before finally being transferred to an arpartate (Asp52) in EvgA response regulator. The internal phosphorelay along with the fact that EvgS activation only occurs within a narrow pH range, might make EvgS more tolerant to noise (Ma, Masuda and Foster, 2004; Burton *et al.*, 2010; Eguchi and Utsumi, 2014).

It is not clear how EvgS senses low pH. Whether it directly detects the change in pH or some other property of the cell which changes at low pH is not known. However, studies have shown that several EvgS domains might play a role in its activation. The periplasmic domain, composed of venus flytrap domains, is necessary for EvgS activation and a key role is played by Leucine 152 residue within this domain (Eguchi and Utsumi, 2014). The periplasmic domain is needed to sense alkali metal ions whereas the cytoplasmic PAS domain is required to detect low pH. Several mutations within the PAS domain have been identified which lead to a constitutively active form of EvgS protein that provides improved acid resistance even at neutral pH (Kato *et al.*, 2000; Johnson *et al.*, 2014). This shows that PAS domain plays a key role in EvgS activation. The PAS domain has also been suggested to act as a redox switch that activates EvgS under aerobic growth conditions, in the presence of oxidised ubiquinone

(Inada *et al.*, 2021). Under anaerobic growth conditions and in the presence of reduced ubiquinone, however, EvgS gets inactivated. C671 and C683 residues are thought to be involved in this redox switch because C671A and C683A mutations activate EvgS even under anaerobic conditions (Inada *et al.*, 2021). The remaining three domains: histidine-kinase and dimerisation, internal receiver, and histidine-containing phosphotransfer, are involved in internal phosphorelay.

A model has been proposed for the activation of EvgS through low pH (Johnson *et al.*, 2014). According to this model, the inactive EvgS exists as a tight dimer and upon activation, transitions into a weakly-bound dimer. Under low pH, the signal is perceived by the periplasmic domain and transmitted to the PAS domain. As a result, structural rearrangements initiate in the PAS domain allowing the tightly-bound inactive EvgS dimer to transition into a weakly-bound active dimer. The mutations found in PAS domain which lead to constitutive activation of EvgS, might mimic these structural rearrangements causing EvgS to get activated.

Whole genome sequencing has shown that *evgS* is conserved in all *E. coli* isolates that have been sequenced to date. However, due to natural variation present in the sequence of *evgS*, differences exist in the ability of different isolates to sense low pH, with some strains being completely non-response to low pH signal (Roggiani, Yadavalli and Goulian, 2017). In contrast, the activation of EvgA response regulator improves the acid resistance of all strains, suggesting that EvgAS TCS is the primary detector of low pH in exponential phase. This finding also suggest that other input signals may exist for EvgS to respond to low pH.

1.4.2 Regulatory network of AR2 response

1.4.2.1 Genes under direct regulation of EvgA response regulator

After getting phosphorylated by EvgS, the response regulator EvgA activates the AR2 system by modulating the expression of several genes directly and indirectly. The regulatory network of AR2 response is highly complex and involves several proteins and small noncoding RNAs, as summarised in Figure 1.4.

EvgA directly upregulates the expression of *ydeP, ydeO* and *gadE* (Masuda and Church, 2003; Ma, Masuda and Foster, 2004; Itou, Eguchi and Utsumi, 2009). The role of YdeP is unclear but it is required for acid resistance (Masuda and Church, 2002). YdeO, on the other hand, activates the expression of GadE, which further activates its own expression (Masuda and Church, 2003; Sayed, Odom and Foster, 2007). The EvgA- and YdeO-mediated GadE activation only occurs in exponential phase in minimal media but not in stationary phase (Ma, Masuda and Foster, 2004). This suggests that distinct regulatory pathways control the activation of AR2 response in different growth phases.

GadE is the central regulator of AR2 response. It regulates the expression of several AR2 genes including *gadA* and *gadB* which encode glutamate decarboxylases, and *gadC* which encodes the glutamate/GABA antiporter (Ma *et al.*, 2003). *gadA, gadB,* and *gadC* are the core and structural components of AR2 response. However, other *gad* genes such as *gadE, gadX,* and *gadW* play a critical part in regulation of AR2 response. The role and regulation of *gad* genes in AR2 response will be discussed in more detail later in section 1.4.2.3.

Another important AR2 regulator is RcsB. RcsB is a part of the RcsCDB phosphorelay, where RcsC is the sensor kinase and RcsB is the response regulator. This system is conserved in members of Enterobacteriaceae and is involved in several processes such as envelope

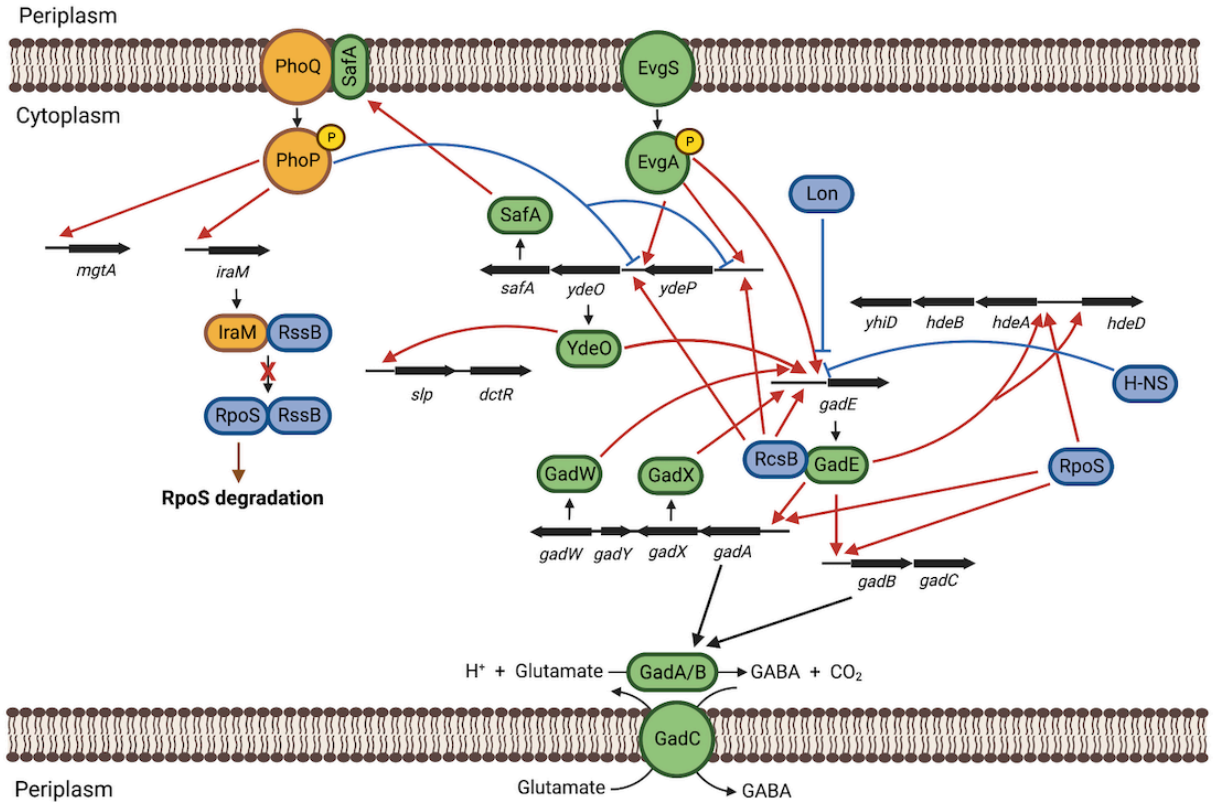


Figure 1.4: Overview of the regulatory network of AR2 acid resistance response in *E. coli*.
The induction of AR2 response in exponential phase is mainly regulated by the EvgAS two-component system (TCS). The EvgA response regulator controls the expression of multiple AR2 genes including *gadE*, which encodes the central regulator of AR2. GadE dimerises with RcsB to control the expression of several AR2 genes. Some of the proteins encoded by the AR2 system such as YdeO, GadX, GadW, and GadE along with other regulators such as Lon, RpoS, and H-NS control *gadE* expression as well. The glutamate decarboxylases GadA and GadB convert glutamate to γ - amino butyric acid (GABA). GadC is the antiporter which imports glutamate in exchange for GABA. *hdeA* and *hdeB* encode two separate periplasmic chaperones which prevent protein aggregation during acid stress. SafA mediates cross-talk between EvgAS and PhoPQ TCSs. PhoPQ has both activating and repressing effects on AR2 response. It inhibits the expression of *ydeO-safA* and *ydeP* promoters but increases cellular RpoS levels, through IraM, which in turn increases *gadE* expression. (Adapted from Schwarz *et al.* (2022); Figure 3. Created with BioRender.com.)

stress, capsule biosynthesis, virulence and formation of biofilms (reviewed in Wall, Majdalani and Gottesman, 2018). The AR2 regulation by RcsB, however, is independent of phosphorelay. Here, RcsB dimerises with GadE and possibly with EvgA to control the expression of both GadE-dependent and GadE-independent promoters (Castanié-Cornet *et al.*, 2010; Burton *et al.*, 2010; Johnson *et al.*, 2011).

1.4.2.2 Crosstalk between EvgAS and PhoPQ two-component systems

EvgA also activates the expression of *safA*. SafA, also known as b1500, is a small protein present downstream of *ydeO*. SafA, through its periplasmic domain, binds directly to the periplasmic domain of sensor kinase PhoQ (Eguchi *et al.*, 2007, 2011, 2012). This binding leads to PhoQ activation, even in the presence of high Mg^{2+} concentration, which usually inhibits PhoPQ. Upon activation, PhoQ sensor kinase phosphorylates the response regulator PhoP, which has both activating and repressing effects on AR2 response. The repressing effects of PhoP are exerted through direct inhibition of *ydeO-safA* and *ydeP* promoters (Burton *et al.*, 2010). PhoP improves the acid resistance of cells through AR2 by indirectly increasing the levels of stationary phase sigma factor RpoS.

RpoS improves the survival of cells under multiple stresses. For example, it enhances cell survival during nutrient starvation, oxidative stress, osmotic stress and acid stress (reviewed in Battesti, Majdalani and Gottesman, 2011; Schellhorn, 2020). RpoS levels in cells remain low during exponential phase but increase several fold upon entry into stationary phase (Schweder *et al.*, 1996; Becker, Klauck and Hengge-Aronis, 1999). One of the genes activated by PhoP is *iraM* (Eguchi *et al.*, 2011). IraM is an anti-adapter protein which binds to the adapter protein RssB. One of the reasons behind reduced RpoS levels in exponential phase is because RssB binds RpoS and directs it to ClpXP protease for its degradation (Becker, Klauck

and Hengge-Aronis, 1999). Through its binding with RssB, IraM stops it from associating with RpoS preventing RpoS from getting proteolysed. As a result, cellular RpoS levels increase when *iraM* is activated. RpoS then improves the AR2 activation by increasing the expression of *gad* genes, and periplasmic chaperone-encoding genes *hdeA* and *hdeB,* through regulating cellular GadE levels (Burton *et al.,* 2010).

1.4.2.3 Role and regulation of gad genes in AR2 response

gad genes, particularly *gadA, gadB* and *gadC,* form the core components of AR2 response in both exponential and stationary phases. Their function has been extensively reviewed in De Biase and Pennacchietti (2012). In *E. coli*, two isoforms of glutamate decarboxylase, GadA and GadB, are present. Both decarboxylases form hexameric strutures. Sequence analysis has shown that *gadA* and *gadB* are 98% identical at nucleotide level (Smith *et al.*, 1992). Functional analysis performed on GadB has shown that this enzyme resides in cytoplasm at neutral pH but when cytoplasmic pH falls below 5.6, it is recruited to the inner membrane (Capitani *et al.*, 2003). Consistent with this finding, an increase in pH from 5.5 to 6.5, leads to blockage of the GadB active site by its C-terminal tail, rendering the protein inactive (Gut *et al.*, 2006).

gadA is encoded on a ~14 kb genomic region termed as the acid fitness island (AFI) which is conserved in *E. coli* (Figure 1.5). In *E. coli* K-12, a total of 13 genes are present in AFI which encode 12 proteins and 1 small RNA. Some of these genes include other *gad* genes such as *gadE, gadX, gadW,* and periplasmic chaperones *hdeA* and *hdeB*. *gadA* and *gadX* are part of the same operon and can be transcribed together or individually (Tramonti *et al.*, 2002). *gadB* and *gadC,* on the other hand, are present outside the AFI. Both genes are transcribed together as a single transcript (De Biase *et al.*, 1999).

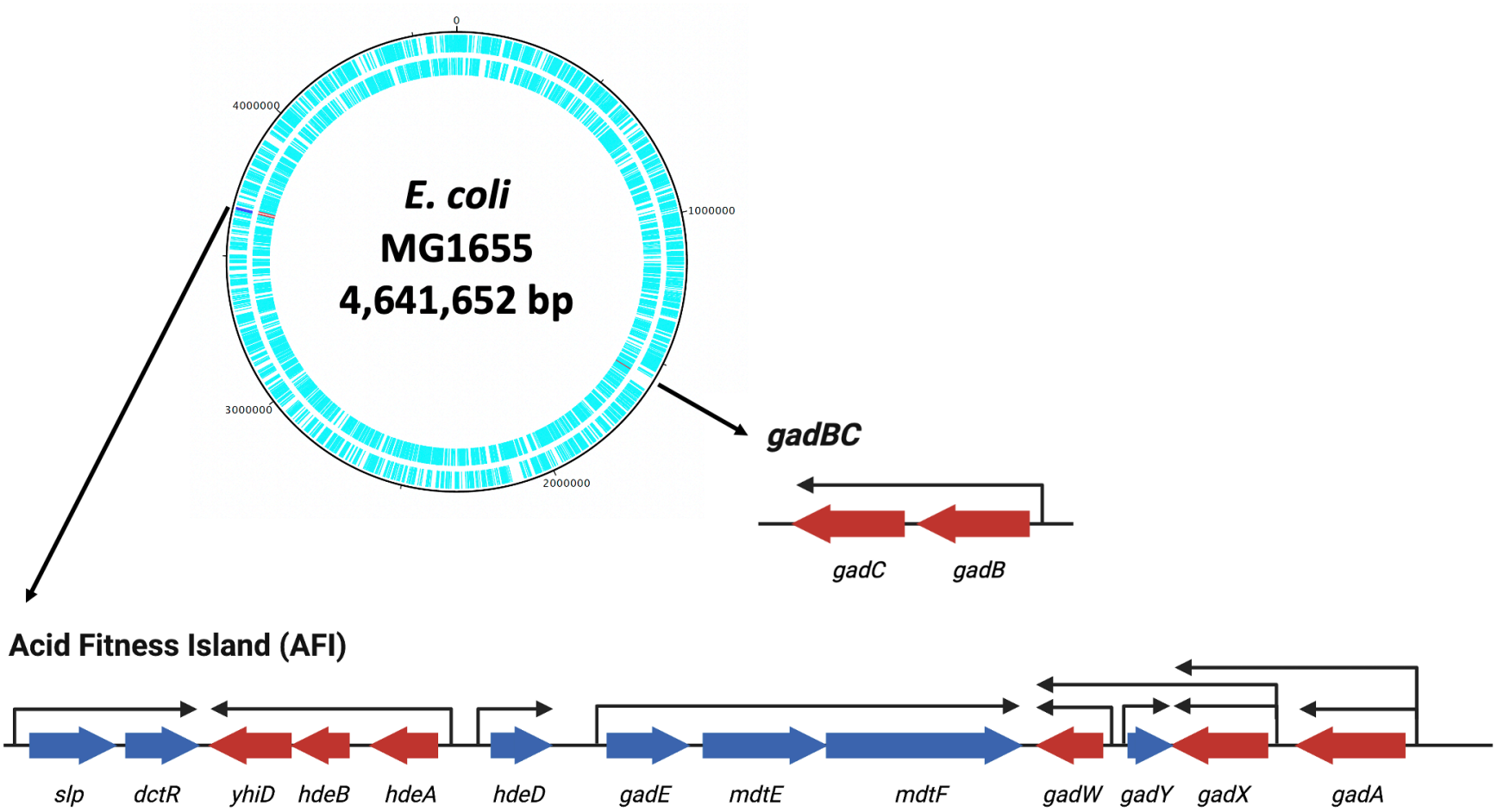


Figure 1.5: Genomic location of *gadBC* operon and acid fitness island (AFI) in *E. coli* K-12 strain MG1655 (NC_ 000913.3).
The numbers in outermost track in genome map correspond to base pairs. The next two blue tracks correspond to sense and antisense CDS respectively. The location of *gadBC* operon and AFI is highlighted in red (genes present on antisense CDS) and blue (genes present on sense CDS) colours. Bent arrows show the genes that are transcribed in the same operon. (Adapted from De Biase and Pennacchietti (2012); Figure 1B. Created with BioRender.com.)

GadC, the AR2 antiporter, imports glutamate in exchange for GABA. Furthermore, GadC also imports glutamine which is converted into glutamate by glutaminase GlsA. As mentioned before, this reaction is coupled with release of ammonia which eventually gets converted into ammonium ions, increasing the cytoplasmic pH (Lu *et al.*, 2013). In-vitro study has shown that GadC only functions when pH levels are 6 and below (Ma *et al.*, 2012). A C-terminal plug blocks this transport channel when its activity is not needed by the cells. At low pH, the C-terminal plug detaches and becomes disordered (del Alamo *et al.*, 2022). GadC is present in an equilibrium between inward-open and outward-open conformations. Furthermore, glutamate but not GABA binding opens an extracellular thin gate without shifting the equilibrium between two conformations.

The AR2 antiporter, GadC, shows substrate selectivity which is based on the protonation state of the compounds to be transported. Depending on the pH, glutamine can exist as Gln^{0} or Gln^{+1}, glutamate exists as $Glu^{0/-1/+1}$ and GABA exists as $GABA^{0/+1}$. GadC recognises the net negative charge on its substrates and only imports glutamine and glutamate in Gln^{0} and $Glu^{0/-1}$ states respectively, and exports GABA in $GABA^{+1}$ state only (Ma, Lu and Shi, 2013; Tsai, McCarthy and Miller, 2013). This helps to increase the cytoplasmic pH by removing positively charged molecules ($GABA^{+1}$) from the cell and by only allowing neutral (Gln^{0} and Glu^{0}) and negatively charged molecules (Glu^{-1}) to enter the cell.

Both glutamate and GABA possess a carboxyl group which has a pK_a of 4.1 for glutamate and 4 for GABA (Richard and Foster, 2004). When the cytoplasmic pH of the cells is above the pK_a of glutamate and GABA, the proton in the carboxyl groups dissociates. However, when the cytoplasmic pH reaches closer to pK_a, the proton is sequestered back by the carboxyl groups. This buffering capacity is absent in arginine, lysine and their

decarboxylation products, which may help explain why AR2 is more effective at dealing with low pH than other AR systems.

GadE/RcsB heterodimer activates the expression of *gadA* and *gadBC* regardless of the growth phase and medium (Ma *et al.*, 2003). GadE/RcsB heterodimer binds to a 20 bp sequence termed as GAD box present in the regulatory regions of both loci (Ma *et al.*, 2003). Furthermore, RcsB can also bind to an RcsB box which is present directly upstream from the -10 element of the *gadA* promoter, leading to *gadA* repression (Castanié-Cornet *et al.*, 2010). In stationary phase, global transcriptional regulator H-NS represses the expression of *gad* genes (*gadA, gadB*, *gadC* and *gadX*) which is relieved by RpoS (Tramonti *et al.*, 2002; Giangrossi *et al.*, 2005).

The expression of GadE to induce AR2 response is controlled by several regulators which might help to quickly increase GadE levels when cells experience acid stress. 798 bp intergenic region present upstream of *gadE* contains three promoters which encode three different transcripts starting at -124, -324/-317, and -566 bp from *gadE* start codon (Sayed and Foster, 2009). Different regulators have been shown to bind different promoters to regulate *gadE* expression. As discussed previously, *gadE* expression in exponential phase is activated directly by EvgA and YdeO. Two AraC-family regulators, GadX and GadW activate *gadE* expression in both exponential and stationary phase, by binding to control region between -798 and -360 bps (Sayed, Odom and Foster, 2007; Tramonti, De Canio and De Biase, 2008). GadE can also activate its own expression by binding to three GAD boxes present between -233 and -42 bps, and overexpression of GadE can overcome the need for GadX and GadW to regulate *gadE* expression (Ma *et al.*, 2003; Sayed, Odom and Foster, 2007).

GadE levels are kept low in cells through constitutive degradation by Lon protease and only increase during AR2 induction conditions (Heuveling, Possling and Hengge, 2008). This

effect continues even when cells are exposed to moderately low pH and when they enter stationary phase. These results suggest that there would be a fitness cost to cells in keeping the AR2 system active when cells are not under acid stress. Indeed, we have observed that EvgS^{c} mutants where the EvgS, and therefore the AR2 response, is constitutively on show slower growth rate than the wild type cells at pH 7. Constant GadE proteolysis might ensure that the AR2 system shuts down as soon as cells return to neutral pH, and that the system is not overexpressed during stationary phase.

1.4.2.4 Role of small noncoding RNAs in AR2 regulation

Several small noncoding RNAs have been suggested to be involved in AR2 regulation. GadY is a small RNA which is encoded on AFI. *gadY* is present directly downstream of *gadX* but on the opposite coding strand. GadY anneals in the intergenic region of *gadXW* transcript allowing it to be cleaved by endoribonuclease RNase III (Opdyke *et al.*, 2011). The resulting two transcripts are then translated to produce GadX and GadW, both of which increase *gadE* expression (Tramonti, De Canio and De Biase, 2008).

Another small RNA, gcvB, has been shown to improve the survival of cells acid stressed at pH 2 by upregulating RpoS expression (Jin *et al.*, 2009). *gcvB* mutants show poor survival at pH 2 but can be rescued by expressing GcvB from a vector plasmid. However, expressing GcvB from a plasmid does not improve the survival of *rpoS* mutants which also show poor survival at pH 2.

Two more small RNAs, DsrA and RprA, improve the acid resistance of cells possibly through regulating cellular RpoS levels. The expression of both these RNAs is induced at pH 5 in exponentially growing MG1655 (Bak *et al.*, 2014). Furthermore, overexpression of either of these RNAs in exponentially growing cells acid stressed at pH 2 in the presence of glutamate

improves the survival of cells. This effect is diminished when glutamate is not added to the pH 2 medium. These findings suggest that these RNAs play a role in regulating the AR2 response of *E. coli* in exponential phase. Both these RNAs have been shown to improve RpoS translation which could be the possible mechanism behind their role in improving the acid resistance of cells (Majdalani, Hernandez and Gottesman, 2002).

1.4.2.5 Potential role of multidrug efflux pumps in AR2 response

The direct role of multidrug efflux pumps in acid resistance has not been reported yet. However, the expression of several genes that encode multidrug efflux pumps is activated by components of AR2 response. For example, EvgA activates the expression of multidrug efflux pump genes *emrK and emrY* (Kato *et al.*, 2000). Furthermore, it also regulates the expression of *tolC*, which encodes TolC, an outer membrane channel that forms component of several multidrug efflux pumps (Eguchi *et al.*, 2003). *mdtE* and *mdtF* encode separate components of MdtEF multidrug efflux pump. These genes are encoded on AFI, and their expression is induced in stationary phase through GadX (Kobayashi *et al.*, 2006; Nishino, Senda and Yamaguchi, 2008).

It has been shown that TolC is required to improve the survival of stationary phase cells acid stressed at pH 2.5 (Deininger *et al.*, 2011). The deletion of *tolC* alone has a bigger effect on the survival of cells than the combined effect of 8 genes, which encode components of different efflux pumps. The authors suggest that TolC might improve the fitness of cells by removing the metabolic products produced at low pH from periplasm. This function of TolC may or may not require its interaction with inner membrane efflux pumps since deletion of multiple efflux pumps has a smaller effect on cell survival than deletion of *tolC* alone. However, the reduced survival of cells in absence of these efflux pumps as compared to the

parent strain suggests that efflux pumps may also assist in removal of metabolic products during low pH.

1.5 Comparison of induction of AR2 response among different *E. coli* strains

The regulation of AR2 response described in section 1.4.2, has mostly been studied using MG1655 strain. However, studies have been done to compare whether the AR2 response can be induced in different laboratory, pathogenic and commensal *E. coli* strains, and these are summarised in Table 1.1. These studies have been done in two conditions that activate the AR2 response: (a) exponential phase cells incubated at mildly acidic pH, and (b) stationary phase cells acid stressed at low pH in the presence of glutamate. Majority of the studies have compared different pathogenic isolates, with a particular focus on O157:H7 (Diez-Gonzalez and Karaibrahimoglu, 2004; Bhagwat *et al.*, 2005; Large, Walk and Whittam, 2005). This is understandable since O157:H7 is a causative agent of many outbreaks.

Two studies have compared strains from all three groups, that is, laboratory, pathogenic and commensal (Lin *et al.*, 1996; Roggiani, Yadavalli and Goulian, 2017). The study by Roggiani, Yadavalli and Goulian (2017) was done on exponential phase *E. coli*. Two of the strains used in this study, MG1655 and Nissle 1917, are also included in our study. The pathogenic strain O157:H7 used by them is also included in our study. However, the two O157:H7 strains, EDL933 in Roggiani, Yadavalli and Goulian (2017), and Sakai in our study, were isolated from different sources and outbreaks. Furthermore, the conditions used in our study are also different from the ones used in Roggiani, Yadavalli and Goulian (2017). In our study, the strains were cultured at neutral pH as opposed to pH 5.5. We then induce our strains at pH 5.5 for 30 minutes, followed by acid stress at pH 2.5 for 2 h. We perform all three steps: growth, induction at pH 5.5 and acid stress at pH 2.5, in M9K minimal media. In

Table 1.1: Examples of studies which have compared the induction of AR2 response in different *E. coli* strains in exponential and stationary growth phases.

Strains used	Experimental condition	Citation
Lab strains: MG1655, BW25113	**Exponential phase induction of AR2 response.** Cultures were grown at 37°C in M9 minimal medium supplemented with 0.2% casamino acids, 0.2% glucose, 100 mM KCl, and 50 mM MOPS and MES buffer. At OD_{600} of ~0.2, the cultures were acid stressed in M9 minimal medium at pH 2.5 for 2 hours with and without being induced at pH 5.5 for 30 minutes.	Boon *et al.* (2020)
Lab strain: MG1655 **Commensal strains:** HS, Nissle 1917, MP1 **Pathogenic strains:** H10407, TW10509 (enterotoxigenic), EDL933 (enterohemorragic), E2348/69 (Enteropathogenic), CFT073 and UTI89 (uropathogenic)	**Exponential phase induction of AR2 response.** Cultures were grown at 37°C in minimal A medium buffered at pH 5.7 with 100 mM MES. At OD_{600} of ~0.2, the cultures were acid shocked in LB at pH 2.5 for 1 h.	Roggiani, Yadavalli and Goulian (2017)
Lab strain: MG1655 **Pathogenic strain:** O157:H7 Sakai	**Exponential phase induction of AR2 response.** Cultures were grown at 37°C in M9 minimal medium supplemented with 0.2% casamino acids and 22.2 mM glucose, buffered with MOPS and MES. At OD_{600} of ~0.2, the cultures were incubated at pH 5.7 or pH 7 for 70 minutes before acid challenge at pH 2.4 for 2 h.	Johnson *et al.* (2011)
Pathogenic strains: 82 pathogenic isolates with different serotypes obtained from human, bovine and other sources	**Stationary phase induction of AR2 response.** Cultures were grown at 37°C for 20-22 h (OD_{600} of 3.5 or higher) either in minimal E medium containing 0.4%	Bhagwat *et al.* (2005)

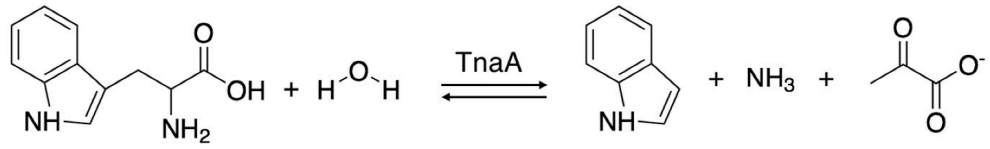
	glucose (EG) at pH 7 or in LB broth buffered with 100 mM MES at pH 5.5. Next day, cultures were diluted 1:200 into EG medium supplemented with and without 1.5 mM glutamate, at pH 2 for 1 h at 37°C.	
Pathogenic strains: 18 O157:H7 patient isolates, 18 O26:H11 patient isolates, 4 O111:H8 patient isolates, 14 O121:H19 patient isolates	**Stationary phase induction of AR2 response.** Cultures were grown overnight for 22 h at 37°C in LB broth supplemented with 0.4% glucose at pH 5. Next day, the cultures were diluted into EG medium supplemented with and without 5.7 mM glutamate, at pH 2 for 2 and 6 h at 37°C.	Large, Walk and Whittam (2005)
Pathogenic strains: 7 O157:H7 isolates obtained from human, cattle and meat sources	**Stationary phase induction of AR2 response.** Cultures were grown to stationary phase in four different media: LB at pH 7, LB supplemented with 0.4% glucose at pH 5, anaerobic basal medium (ABM) supplemented with 0.5 g L^{-1} glucose at pH 7, and ABM supplemented with 4 g L^{-1} glucose at pH 4.5. Next day, cultures were diluted 1:1000 into EG medium supplemented with and without 0.005-5.7 mmol L^{-1} glutamate, at pH 2 for 1 h at 37°C.	Diez-Gonzalez and Karaibrahimoglu (2004)
Lab strains: K-12, MC4100 **Commensal strains:** 3 patient isolates **Pathogenic strains:** O157:H7 (ATCC 43895), 10 patient isolates	**Stationary phase induction of AR2 response.** Cultures were grown overnight (18-20 h) in LB broth supplemented with 0.4% glucose at pH 5. Next day, the cultures were diluted (1:1,000) into EG medium supplemented with and without 5.7 mM glutamate, at pH 2 for 2 h at 37°C.	Lin *et al.* (1996)

contrast, Roggiani, Yadavalli and Goulian (2017) grow their strains in minimal media but acid stress them in LB and only for an hour. All the studies in Table 1.1 have either been done in exponential or stationary phase. In our study, we are testing the induction of AR2 response in three growth phases, that is, exponential, early stationary and stationary phases.

To understand the molecular mechanisms of AR2 response, high-throughput microarray studies have been done. These studies were done in exponentially growing MG1655 by either exposing the cells to pH 5.5 or through overexpression of EvgA, and led to important findings such as the role of GadE and EvgA-YdeO activation circuit and the presence of AFI in *E. coli* (Tucker, Tucker and Conway, 2002; Masuda and Church, 2002). However, neither of these studies included a pathogenic or commensal strain to determine if the molecular mechanisms of AR2 response are conserved across laboratory and real-world strains. To answer this question, in our study, we have examined and compared the acid stress response of a laboratory K-12 strain MG1655 and a pathogenic strain O157:H7 Sakai using a high-throughput technique called TraDIS (more details on TraDIS can be found in section 1.7).

1.6 Indole signalling and its effect on acid stress

Indole is an aromatic heterocyclic compound produced by both Gram-positive and Gram-negative bacterial species. Its production requires the tryptophanase enzyme (encoded by *tnaA*) which catalyses the deamination reaction during which the amine ($-NH_2$) group of tryptophan is removed and it is converted to indole, pyruvate and ammonia. (Li and Young, 2013). The reaction is shown below.



In *E. coli* cultures, the concentration of indole remains low during exponential phase and increases during stationary phase (Chant and Summers, 2007). Depending on the amount of tryptophan present in the medium, the supernatant indole concentration can reach between 0.5 and 1 mM in stationary phase LB cultures (Li and Young, 2013).

1.6.1 Persistent and pulse indole signalling

Indole is a signalling molecule which is involved in intraspecies, interspecies and interkingdom signalling (reviewed in Lee, Wood and Lee, 2015; Tomberlin *et al.*, 2017). For example, among non-indole producing bacteria, indole has been shown to increase antibiotic tolerance in *Salmonella enterica* serovar Typhimurium and *Pseudomonas aeruginosa*. Indole also inhibits biofilm formation in *Staphylococcus aureus* and reduces biofilm formation in *Candida albicans*. Indole produced by soil-borne bacterial species acts as a potent plant growth modulator. Insect behaviour, such as oviposition in mosquitoes and butterflies, is also affected by indole. Two types of indole signalling are known: persistent and pulse signalling. In persistent signalling, low levels of indole over a long period of time are produced. In pulse signalling, however, transient high levels of indole are produced. Persistent indole signalling affects gene expression during various stress responses such as antibiotic stress, pH stress, and heat stress (reviewed in Zarkan *et al.*, 2020). Furthermore, it is also involved in the formation of biofilms and persister cells, as well as virulence. Indole pulse signalling, on the

other hand, regulates cytoplasmic pH and improves survival during nutrient starvation in stationary phase cells (Zarkan *et al.*, 2019).

During transition from exponential phase to stationary phase, cells produce a sudden high level of indole after which the intracellular indole levels can reach up to 60 mM (Gaimster *et al.*, 2014). The high indole pulse inhibits cell growth and division causing wild-type cells to enter stationary phase earlier than indole non-producing mutants (*tnaA* mutants). As a result, the indole producing wild-type cells have access to more resources in stationary phase than *tnaA* mutants, which improves their long-term survival during starvation period.

Indole pulse has been shown to maintain the cytoplasmic pH of wild-type cells at 7.2 (Zarkan *et al.*, 2019). On the other hand, *tnaA* mutants do not experience indole pulse and their cytoplasmic pH is 7.8. Cells that experience indole pulse during preceding stationary phase entry retain memory of this pulse allowing them to maintain their cytoplasmic pH. The possible mechanism responsible for this pH regulation might be the ability of indole to act as an ionophore because the cytoplasmic pH of *tnaA* mutants can be changed from 7.8 to 7.2 by applying a pulse of synthetic proton ionophores such as CCCP or DNP.

1.6.2 Effect of indole on acid stress

To date, four studies have looked at the effect of indole on acid stress. A study by Hirakawa *et al.* (2010) showed that addition of 1-2 mM of indole to *E.coli* MC4100 culture grown in LB to stationary phase increases the acid resistance of cells by 3.6-6.3 times when cells are acid shocked at pH 3.5 for 1.5 h. The authors further showed that indole increases the expression of AR2 genes such as *gadA, gadB, gadC*, and *gadE,* which could be the potential reason for increased acid resistance of cells.

Another study showed that addition of 2 mM indole to BW25113 cultured to mid-exponential phase in LB, reduces the acid resistance of cells at pH 2.5 (Lee *et al.*, 2007). A third study has shown that addition of 50 µM indole to MG1655 cultured to OD_{600} of 0.5 in LB, inhibits the GAD decarboxylase activity of cells, visualised through induction at pH 5.5 for 30 minutes (Kanda *et al.*, 2020). Finally, a recent study from our lab has shown that when MG1655 and BW25113 cultured to mid-exponential phase in M9K minimal medium are induced at pH 5.5 in the presence of 1 mM indole, the induction of the AR2 response is blocked, reducing the survival of cells at pH 2.5 (Boon *et al.*, 2020).

All three studies done in exponential phase suggest that indole reduces the acid resistance of cells (Lee *et al.*, 2007; Boon *et al.*, 2020). Furthermore, two of these studies suggest that this effect is due to the inhibitory effect of indole on AR2 response (Boon *et al.*, 2020; Kanda *et al.*, 2020). However, study by Hirakawa *et al.* (2010) suggests that indole increases the acid resistance of cells in stationary phase, possibly through AR2 response. The contradictory findings of these studies could be attributed to the differences in the growth phases in which these studies were done. It is possible that indole can have different effects on the acid resistance of cells in different growth phases. The study by Hirakawa *et al.* (2010) was done at OD_{600} of 0.8, a point at which cells transition from exponential to stationary phase. This is the point where cells experience indole pulse. So the differences seen between studies done in exponential and stationary phases could be due to different types of indole signalling, with persistent signalling occurring in exponential phase and pulse signalling in the stationary phase.

The other possible reason for the differences in the findings of these studies could be due to whether indole is affecting the innate acid resistance of cells or the induced acid resistance of cells. The study by Boon *et al.* (2020) and Kanda *et al.* (2020) examined the

impact of indole on the induction of AR2 acid stress response through incubation at pH 5.5. On the other hand, the study by Hirakawa *et al.* (2010) and Lee *et al.* (2007) did not include the pH 5.5 incubation step.

1.6.3 Potential routes for indole transport

Despite being involved in various aspects of bacterial physiology, it is not clear how indole enters the cells and what is its primary target inside the cell. There is evidence that Mtr and AcrEF transporters could be involved in indole transport (Yanofsky, Horn and Gollnick, 1991; Kawamura-Sato *et al.*, 1999). However, other studies have shown that indole can diffuse freely through lipid membranes (Bean, Shepherd and Chan, 1968; Piñero-Fernandez *et al.*, 2011). So there is a possibility that indole is transported into the cells through both protein transport and passive diffusion through lipid membranes.

1.7 Transposon-insertion sequencing as a method to identify essential and conditionally essential genes

Transposon-insertion sequencing (TIS) is a method which combines transposon mutagenesis with next-generation sequencing to identify essential and conditionally essential genes. Essential genes can be defined as the genes whose function is absolutely necessary for the growth of cells under all conditions. On the other hand, conditionally essential genes are those which are not absolutely necessary for growth of cells, but their function is needed for optimal fitness of cells under a particular condition.

1.7.1 TIS workflow

TIS workflow involves introducing a transposon insertion at a unique site within each cell's genome. This is followed by selection of transformants on an appropriate selective medium which is mediated by the presence of an antibiotic resistance cassette in the transposon (Figure 1.6). The transformants are then pooled together followed by genomic DNA extraction and fragmentation. Finally, the regions flanking the transposon and genomic DNA are amplified and sequenced using next-generation sequencing. By mapping the sequencing reads to the reference genome, the location and frequency of the transposon insertions can be determined. The essential genes are identified as the ones that either do not tolerate any or a very small number of insertions.

1.7.2 Identification of conditionally essential genes

To identify conditionally essential genes, the transposon library is grown under a selective condition followed by preparation of a sequencing library referred to as the conditional library (Figure 1.7). The control sequencing library is prepared either from the transposon library as it is, or by growing the transposon library under the same experimental conditions as the conditional library but without the selection step. The frequency of mutants containing inserts in all non-essential genes, intergenic regions and regulatory regions such as promoter sequences, is compared between the control and conditional libraries.

Three types of genomic features can be identified after comparison between the control and the conditional sequencing libraries (Figure 1.8). (a) Genomic features containing transposon insertions whose frequency decreases in the conditional library relative to the control library. These genomic features are likely to be needed for optimal fitness of cells under a selective condition and are considered conditionally essential. (b) Genomic features

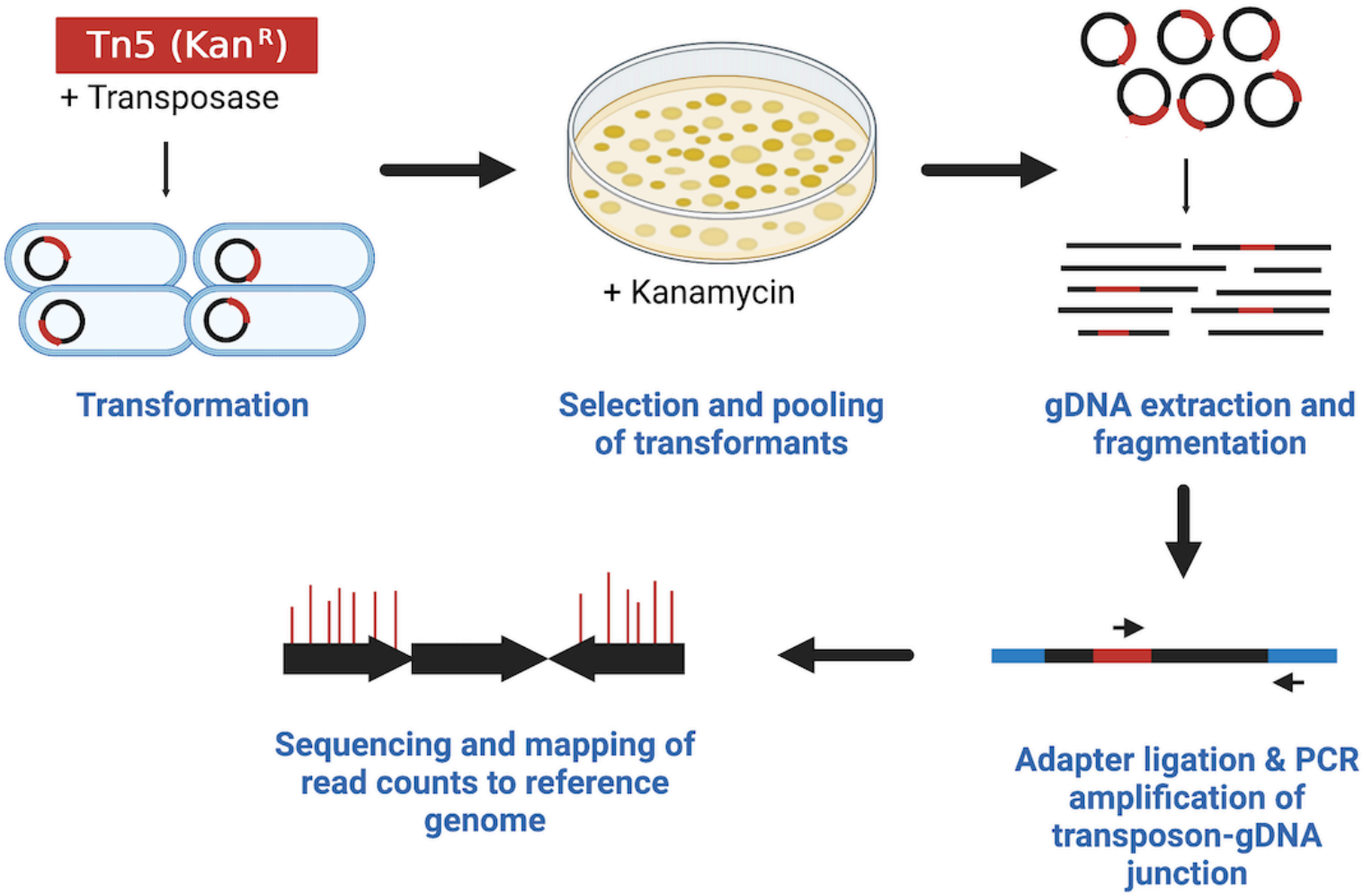
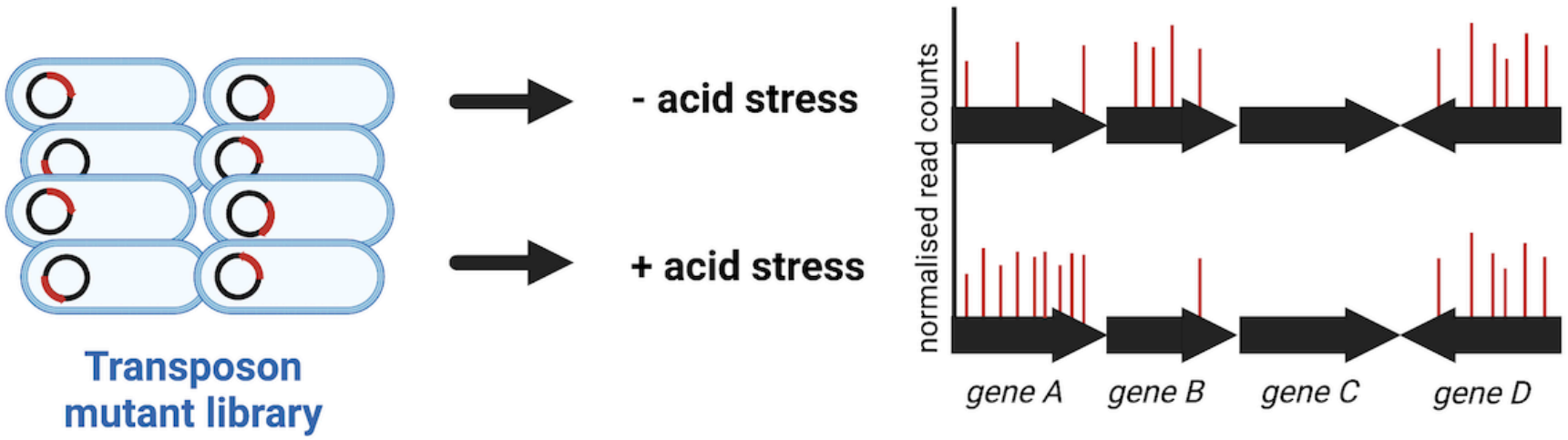


Figure 1.6: Workflow of Transposon-insertion sequencing (TIS) method for preparation of a transposon mutant library and analysis of insertion sites.
(Adapted from Yasir *et al.* (2020); Figure S1. Created with BioRender.com.)



gene A: Gene function reduces the fitness of cells under acid stress
gene B: Gene function is needed for optimal fitness of cells under acid stress
gene C: Essential gene
gene D: Gene function does not impact the fitness of cells under acid stress

Figure 1.7: Identification of conditionally essential genes using Transposon-insertion sequencing (TIS) method.
(Adapted from Cain *et al.* (2020); Figure 1B. Created with BioRender.com.)

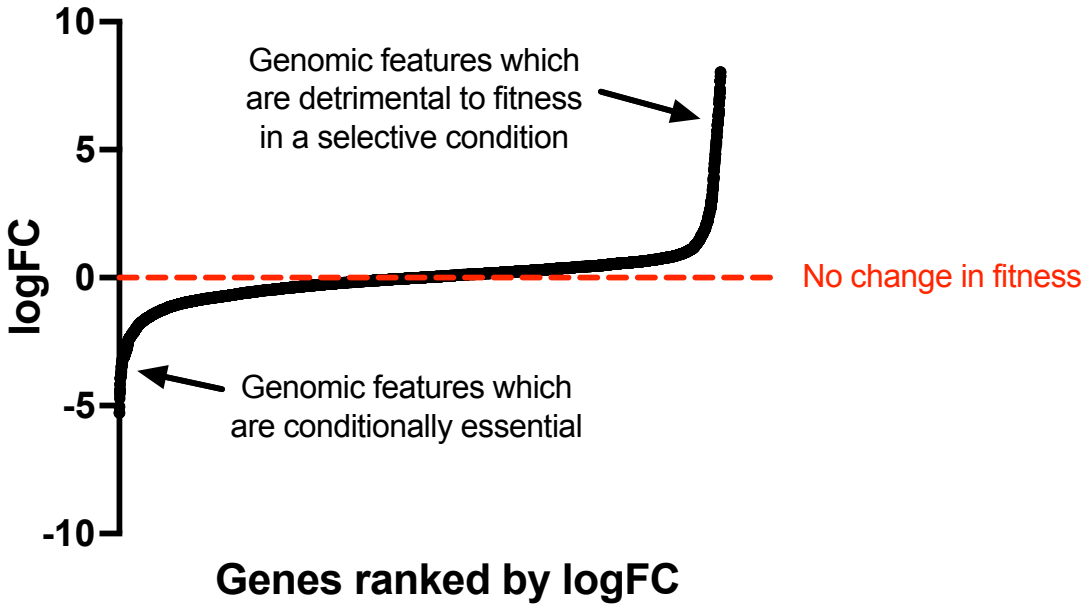


Figure 1.8: Genes ranked according to their log_2FC obtained after comparison between the control and conditional TraDIS libraries.
The comparison between the control and conditional TraDIS sequencing libraries provides a log_2FC for each gene which shows whether the frequency of insertions within a gene has increased or decreased in the conditional TraDIS library. Insertions within genomic features drop if those genomic features are conditionally essential and insertions within genomic features increase if those genomic features are detrimental to fitness in the selective condition. (Adapted from Milner (2021); Figure 5.6.1)

containing transposon insertions whose frequency increases in the conditional library relative to the control library. These features are likely to reduce the fitness of cells under the given selective condition. (c) Genomic features containing transposon insertions whose frequency remains the same in the conditional library relative to the control library. These features do not affect the fitness of cells under a selective condition.

1.7.3 Types of TIS methods

Four types of TIS methods are currently used, which are reviewed in van Opijnen and Camilli (2013). These are: transposon sequencing (Tn-Seq), transposon-directed insertion site sequencing (TraDIS), insertion sequencing (INSeq), and high-throughput insertion tracking by deep sequencing (HITS). The basic workflow behind all TIS methods is same. However, variations exist because of the different transposons that are used to make the transposon libraries and the way DNA is fragmented to prepare the sequencing libraries.

Tn-Seq and INSeq use *mariner* transposon which gets inserted in TA dinucleotide sites (Goodman *et al.*, 2009; van Opijnen, Bodi and Camilli, 2009). Furthermore, the DNA in these two methods is fragmented using restriction digestion to achieve short uniform reads, which reduces the PCR amplification bias. On the other hand, TraDIS and HITS usually use Tn5 transposon which does not have an insertion site bias (Gawronski *et al.*, 2009; Langridge *et al.*, 2009). Additionally, DNA in these two methods is sheared using sonication to achieve long reads which improves their genome mapping.

1.7.4 TraDIS-Xpress to assay essential genes

One of the drawbacks of TIS methods is that the function of essential genes in selective conditions cannot be determined since essential genes either do not tolerate any or a very

small number of insertions. To overcome this, recently a method called TraDIS-Xpress has been developed (Yasir *et al.*, 2020). In this method, the transposon library is prepared using a mini-Tn5 transposon carrying an outward-facing *tac* promoter present downstream of the kanamycin resistance gene. Through addition of IPTG (Isopropyl ß-D-1-thiogalactopyranoside) to the growth medium, the *tac* promoter can be induced allowing genes, including essential genes, present downstream of the transposon to be either overexpressed or repressed (Figure 1.9). If the *tac* promoter is in the same orientation as the gene, then this leads to the overexpression of the gene. However, if the *tac* promoter is in the opposite orientation to the gene, then the anti-sense mRNA produced through *tac* promoter binds to the sense mRNA leading to gene repression. By comparing the frequency of insertions, which affect the expression of essential genes, in the conditional and control libraries, the importance of essential genes in a selective condition can be determined.

1.8 Aims and Objectives of this study

In our lab, we have previously shown that the signalling molecule indole inhibits the induction of AR2 response in exponentially growing laboratory K-12 strains MG1655 and BW25113 (Boon *et al.*, 2020).

(a) It is not clear whether the inhibitory effect of indole on the induction of acid stress response is conserved in other *E. coli* strains. Furthermore, whether the AR2 acid stress response, through incubation at mildly acidic pH, can be induced in growth stages outside exponential phase has not been tested to date. If the AR2 can be induced outside exponential phase through incubation at mildly acidic pH, whether it can be blocked with indole is also unknown. Therefore, the first objective of this study was to test whether indole can block the

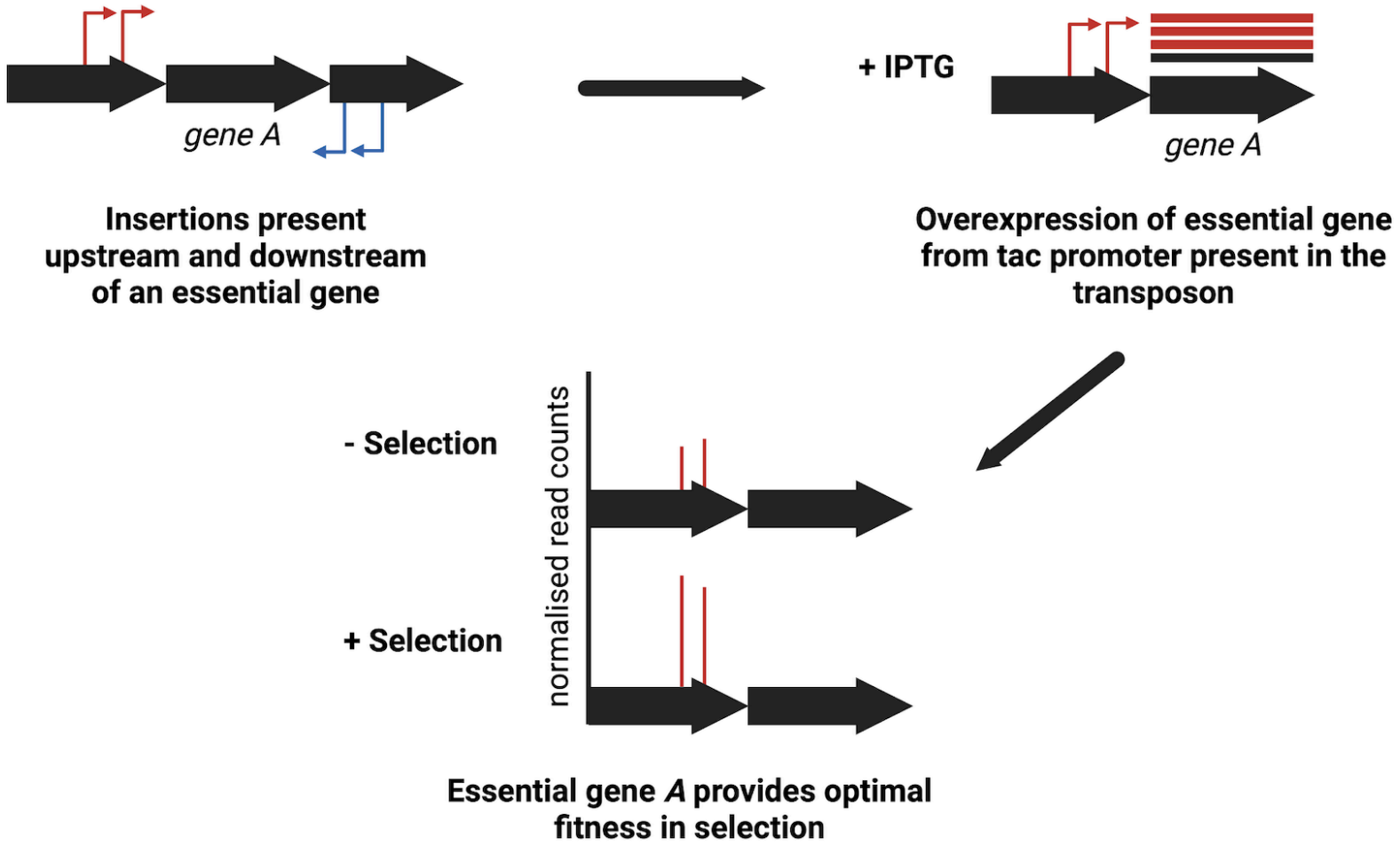


Figure 1.9: TraDIS-Xpress to assay conditional importance of essential genes.
(Adapted from Cain *et al.* (2020); Figure 3B. Created with BioRender.com.)

induction of AR2 response in exponentially growing pathogenic (O157:H7 Sakai) and commensal (Nissle 1917) strains. The second objective of this study was to test whether the AR2 response can be induced and subsequently blocked, in early stationary and stationary growth phases in MG1655, BW25113, O157:H7 Sakai and Nissle 1917.

(b) To date, no global analysis of the contribution of genes to fitness at low pH has been done for *E. coli*. Such an analysis could enable the identification of genes with an important role in acid resistance and the acid stress response that have not yet been studied. It could also allow a comparison of this response between laboratory and pathogenic strains of *E. coli*, to see whether the same networks of genes are likely to be important in resistance. For this reason, our third objective is to use TraDIS to examine and compare the effect of low pH between MG1655 and O157:H7 Sakai.

(c) TraDIS allows the conditional essentiality of non-essential genes to be determined in a selective condition. However, in order to determine the conditional essentiality of essential genes, TraDIS-Xpress library needs to be used. The fourth objective of this study was to prepare a TraDIS-Xpress O157:H7 Sakai library and determine its essential and non-essential genes. The TraDIS experiments mentioned in (b) were performed with this library and the MG1655 transposon library (normal TraDIS library, not TraDIS-Xpress). Due to time constraints the conditional essentiality of essential genes was not examined in this study. However, the TraDIS-Xpress library was prepared so that it could be done later.

(d) So far it is known that indole inhibits the induction of AR2 response in exponential phase. However, the molecular mechanism behind the interaction of indole with acid stress response is unknown. Therefore, the fifth and last objective of this study was to use TraDIS to examine the effect of indole on induced acid stress response in MG1655 and O157:H7 Sakai.

Chapter 2

Materials and Methods

2.1 Bacterial strains, plasmids, and oligonucleotides

The bacterial strains, plasmids and oligonucleotides used in this study are listed in Tables 2.1, 2.2, and 2.3 respectively.

MG1655, BW25113, O157:H7 Sakai, Nissle 1917, and Nissle 1917 *ΔbcsA* were whole genome sequenced and aligned to their reference genomes obtained from NCBI database: MG1655 (NC_000913.3; Riley *et al.,* 2006), BW25113 (NZ_CP009273.1; Grenier *et al.,* 2014), O157:H7 Sakai (NC_002695.2; Hayashi *et al.,* 2001), and Nissle 1917 (NZ_CP007799.1; Reister *et al.,* 2014). Additional mutations were found in all strains especially in Nissle 1917 which was considerably different than its reference genome. These mutations are outlined in Appendix 8.1.

Oligonucleotides were sourced from Eurogentec, at a concentration of 100 µM dissolved in milli-q water, and were diluted in-house to 10 µM using nuclease-free water (Invitrogen) before use.

Table 2.1: Strains used in this study.

Strain	Description/Genotype	Reference
MG1655	F^-, λ^-, $ilvG^-$, rfb-50 rph-1	Blattner *et al.* (1997)
MG1655 *ΔrpoS::kan^R*	Chromosomal r*poS* gene is inactivated by insertion of kanamycin resistance gene	Milner (2021)
MG1655 *Δevg::cam^R*	Chromosomal *evgS* gene is inactivated by insertion of chloramphenicol resistance gene	P1 MG1655 *ΔevgS::cam^R ydeP-lacZ::kan^R* (Eguchi and Utsumi, 2014) X MG1655
MG1655 *ydeP-lacZ::kan^R*	Chromosomal *ydeP* gene is fused with the *lacZ* gene, and the fusion is marked with kanamycin resistance gene	P1 MG1655 *ΔevgS::cam^R ydeP-lacZ::kan^R* (Eguchi and Utsumi, 2014) X MG1655
MG1655 *Δevg::cam^R ydeP-lacZ::kan^R*	Strain contains both *evgS* gene deletion and *ydeP-lacZ* fusion on the chromosome marked with antibiotic resistance genes as described above	P1 MG1655 *ΔevgS::cam^R ydeP-lacZ::kan^R* (Eguchi and Utsumi, 2014) X MG1655 *ydeP-lacZ::kan^R*
MG1655 *ΔmrcA::tet^R*	Chromosomal *mrcA* gene is inactivated by insertion of tetracycline resistance gene	P1 lysate was gifted by Manuel Banzhaf
MG1655 *ΔmrcB::tet^R*	Chromosomal *mrcB* gene is inactivated by insertion of tetracycline resistance gene	P1 lysate was gifted by Manuel Banzhaf
MG1655 *ΔnlpI::kan^R*	Chromosomal *nlpI* gene is inactivated by insertion of kanamycin resistance gene	P1 lysate was gifted by Manuel Banzhaf
MG1655 *ΔgadE::kan^R*	Chromosomal *gadE* gene is inactivated by insertion of kanamycin resistance gene	P1 MG1655 *ΔgadE::kan^R* (Milner, 2021) X MG1655

KH001	*lacZ*$^{-}$ derivative of MG1655 whereby the *lac* promoter has been replaced with pKH3 cloning site	Gifted by Jack Bryant
BW25113	lacI^{+}, Δ(araD-araB)567, Δ(rhaD-rhaB)568, ΔlacZ4787 (::rrnB-3) hsdR514, rph-1	Grenier *et al.* (2014)
BW25113 Δ*evgS::cam*^{R}	Chromosomal *evgS* gene is inactivated by insertion of chloramphenicol resistance gene	P1 MG1655 Δ*evgS::cam*^{R} *ydeP-lacZ::kan*^{R} (Eguchi and Utsumi, 2014) X BW25113
O157:H7 Sakai	pO157, pOSAK1, s*tx1*$^{-}$, s*tx2*$^{-}$	Hayashi *et al.* (2001)
Nissle 1917	wild-type	Monteiro *et al.* (2009)
Nissle 1917 Δ*bcsA::cam*^{R}	Chromosomal *bcsA* gene is inactivated by insertion of chloramphenicol resistance gene	Monteiro *et al.* (2009)

Table 2.2: Plasmids used in this study.

Plasmid	Description	Reference
pLux-xxx	Family of plasmids whereby Luciferase reporter is driven by a different gene promoter (xxx- *pydeP, pgadE, phdeA,* and *pacp*) in each plasmid	Burton *et al.* (2010)
pBAD24	Vector plasmid used to express proteins from pBAD promoter of araBAD operon; gene expression induced through addition of L-arabinose; encodes ampicillin resistance	Guzman *et al.* (1995)
pBAD24-EvgS	His-tagged EvgS expressed from pBAD promoter in pBAD24 plasmid	Eguchi and Utsumi (2014)
pBAD24-EvgS.S600I	Serine 600 in His-tagged EvgS, in pBAD24 plasmid, is mutated to isoleucine (S600I). This mutation results in EvgS protein which is constitutively active even at pH 7	Johnson *et al.* (2014)
pBAD24-EvgS.D625A	Aspartate 625 in His-tagged EvgS, in pBAD24 plasmid, is mutated to alanine (D625A)	This study
pKD4	Plasmid used for making gene knockouts; contains FRT-flanked kanamycin resistance gene used to select mutants	Datsenko and Wanner (2000)
pKD46	Red recombinase expression plasmid; encodes ampicillin resistance	Datsenko and Wanner (2000)

Table 2.3: Oligonucleotides used in this study.

Primer	Sequence (5' - 3')	Description/ Reference
a1874c_F	GAAAAAACATCTTTAAAGGGTGAGGCACTGTTTTCTAATG GTAACATTG	To mutate codon 1874 from A to C introducing D625A mutation in EvgS PAS domain of MG1655 in pBADEvgS plasmid
a1874c_R	CAATGTTACCATTAGAAAACAGTGCCTCACCCTTTAAAGAT GTTTTTTC	
c1875a_F	CATCTTTAAAGGGTGATTCACTGTTTTCTAATGGTAACATT GCATTTT	To mutate codon 1875 from C to A introducing D625E mutation in EvgS PAS domain of MG1655 in pBADEvgS plasmid
c1875a_R	AAAATGCAATGTTACCATTAGAAAACAGTGAATCACCCTT TAAAGATG	
rpoS_F	TATCGCAGGCAGCAAAGGACAGG	To confirm *rpoS* deletion in MG1655 (Milner, 2021)
rpoS_R	GGTGCGTATGGGCGGTAATTTACC	
evgS_F	CGCTATGAAGTATCACAAAATTGGC	To confirm *evgS* deletion in MG1655
evgS_R	CGTAAATGTTGTTCCTATGCCTGGATG	
ydeP-F1	AGTGGTCATTTACCCTATGG	To confirm *ydeP-lacZ* fusion in MG1655 (Eguchi and Utsumi, 2014)
lac	GACCATTTTCAATCCGCA	
Ec_bcsA5	TAGGTGAATATCGCAGTGATGC	To confirm *bcsA* deletion in Nissle 1917 (Monteiro *et al.*, 2009)
Ec_bcsA6	CATTGATCAGTGGTTGCGTTGC	
mrcA_F	CTGACGGGTATATCAATGCG	To confirm *mrcA* deletion in MG1655
mrcA_R	ATGCGTGAAGTTTGCCTG	
mcrB_F	CAGAAGAACAGAAAATCGGG	To confirm *mrcB* deletion in MG1655
mrcB_R	CAATCTGAGCGTTCCTG	
nlpI_F	TCCGTCTGAGCATTAAAGAAGC	To confirm *nlpI* deletion in MG1655
nlpI_R	CACCCGTTAAGGTGATGGC	
gadE_F	GCGTGATGGATAAATCTGAAG	To confirm *gadE* deletion in MG1655
gadE_R	GTTCTGCCAACAGTTCCTG	
Tnp001	CTGTCTCTTATACACATCT	To amplify the transposon from inverted repeat ends from BW25113 TraDIS-Xpress library
Tnp001_F	AGACAACCATCATCGTCG	To confirm the sequence of Tnp100 transposon
Tnp001_R	CACCTAAACGGACTGGC	
TKK_F	ACCTGCAGGCATGCAAGCTTCAGG	Forward primer for PCR1 during TraDIS library preparation
TXP_F	TGTTTTCACGAGCACTGCAGG	Forward primer for PCR1 during TraDIS-Xpress library preparation
TKK_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATC	Reverse primer for PCR1 in TraDIS and TraDIS-Xpress library preparation

TKK_6.1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACGAGCTTCAGGGTTGAGATGTGTA	Inline barcode primers for PCR2 during TraDIS library preparation. Barcode sequences within each primer are underlined.
TKK_7.4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCTAGAGCTTCAGGGTTGAGATGTGTA	
TKK_8.3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGCATGAGCTTCAGGGTTGAGATGTGTA	
TKK_9.2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATCGATCGAAGCTTCAGGGTTGAGATGTGTA	
TXP_6.1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACGCATGCCAGGGTTGAGATGTGTA	Inline barcode primers for PCR2 during TraDIS-Xpress library preparation. Barcode sequences within each primer are underlined.
TXP_7.4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCTAGCATGCCAGGGTTGAGATGTGTA	

2.2 Media preparation

2.2.1 M9K minimal medium (M9Ksupp medium)

Each Litre of M9K contained the following:

Compound	Formula	Amount/Concentration
di-sodium hydrogen orthophosphate (anhydrous) (Fisher Scientific)	Na_2HPO_4	6 g
Potassium dihydrogen orthophosphate (Fisher Scientific)	KH_2PO_4	3 g
Sodium chloride (Fisher Scientific)	NaCl	0.5 g
Ammonium chloride (Sigma-Aldrich)	NH_4Cl	1 g
Potassium chloride (VWR International)	KCl	7.5 g
3-morpholinopropane-1-sulfonic acid (MOPS) (Fisher Scientific)	$C_7H_{15}NO_4S$	50 mM
2-(N-morpholino)ethanesulfonic acid hydrate (MES) (Fisher Scientific)	$C_6H_{13}NO_4S \cdot xH_2O$	50 mM

M9 minimal medium was supplemented with 100mM KCl because the addition of alkali metal ions leads to better induction of promoters involved in AR2 acid resistance response (Eguchi and Utsumi, 2014). All M9K salts including KCl were dissolved in 900 mL of distilled water. The pH of the medium was adjusted (to pH of 7/5.5/2.5) through addition of either 1M HCl or 1M NaOH before making it to a final volume of 1 L. Sterilisation was performed at 121 °C, 15 psi, for 20 minutes. The medium was supplemented with following: 20% (w/v) D-glucose (anhydrous) (Fisher Scientific), 20% (w/v) casamino acids (Fisher Scientific), 1M magnesium sulphate anhydrous ($MgSO_4$) (Honeywell), and 100mM calcium chloride ($CaCl_2$) (Sigma-Aldrich). All supplements were made in distilled water and filter sterilised. These were added to the medium on the day of the experiment. Every 100 mL of

sterile M9K contained: 0.2% glucose, 0.2% casamino acids, 1mM magnesium sulphate and 0.1mM calcium chloride.

2.2.2 Lysogeny Broth (LB broth)

LB broth was prepared by dissolving 10 g of tryptone (Scientific Laboratory Supplies), 5 g of yeast extract (Sigma-Aldrich), and 10 g of sodium chloride (Fisher Scientific) in a final volume of 1 L of distilled water. When required, the pH of the media was adjusted to 2.5 before autoclaving.

2.2.3 2xTY Broth

2x TY broth was prepared by dissolving 16 g of tryptone, 10 g of yeast extract, and 5 g of sodium chloride in a final volume of 1 L of distilled water. The pH of the media was adjusted to 6.8 $\pm$ 0.2 before autoclaving.

2.2.4 pH measurement

The pH was measured with Jenway 3510 pH meter using a Thermo Scientific Orion 815600 ROS combination probe. pH calibration was performed using Thermo Scientific Orion pH buffers- pH 7 and 4.01. In few cases, smaller volumes were tested using a pocket size pH meter- CAMLAB pHBoy-P2.

2.3 Growth conditions

The strains were retrieved from glycerol stocks through streaking out on an LB agar (LB + 1.5% (w/v) agar; pH 7) plate supplemented with appropriate antibiotic. The plates were

incubated at 37 °C overnight. The overnight cultures were always set up by inoculating 5 mL of LB broth (supplemented with antibiotic if required) in a 30 mL plastic universal container with a single colony. The cultures were incubated overnight at 37 °C while shaking at 180 rpm. The antibiotic concentrations used were as follows: ampicillin 100 µg/mL, chloramphenicol 30 µg/mL and kanamycin 50 µg/mL. Fresh 15% glycerol stocks in LB were prepared for all strains and stored at -80 °C. Strains were streaked out on a fresh plate from glycerol stocks every two weeks for all phenotypic assays.

2.4 Growth curve

50 mL of M9Ksupp pH 7 medium was inoculated in a 250 mL shaker flask with the overnight culture to a starting optical density (measured at 600 nm- OD_{600}) of 0.05. The flask was placed in a 37 °C shaking (at 180 rpm) incubator and an OD_{600} reading was taken every hour in the spectrophotometer for a total of 8 hours. The experiments were done in triplicates. The replicates for this as well as all other experiments in this study were biological.

2.5 Acid survival assays

2 x 30 mL plastic universals containing 5 mL of M9Ksupp pH 7 medium were inoculated with the overnight culture to a starting OD_{600} of 0.05. The cultures were grown to required OD_{600} in 37 °C shaking incubator. 100 µL of culture was taken from first universal and added to 900 µL of M9Ksupp pH 7, while 100 µL of culture from second universal was added to 900 µL of pH 2.5 medium, in 1.5 mL microcentrifuge tubes. 10- fold serial dilutions were prepared from both cultures. pH 7 dilutions were plated out straightaway whereas pH 2.5 dilutions were incubated at room temperature for 2 hours (unless stated otherwise) before being

plated out. The plates were incubated at 37 °C overnight and colonies were counted the next day. Experiments were done in at least triplicates.

The tilt plate method was used for plating, unless stated otherwise (Figure 2.1). 10 cm x 10 cm square petri dishes were used. 8 µL of culture from appropriate dilutions was placed on top of the plate and the plate was tilted vertically to allow the culture drop to run all the way to the bottom of the plate. If spreading method was used for plating, then standard round petri dishes were used and 100 µL of culture was spread plated from each dilution.

Colonies were counted from the sample which had the highest countable number of colonies. Percentage survival was calculated by dividing the number of colonies in pH 2.5 after 2 h by the number of colonies in pH 7 at 0 h.

2.6 Induction and Indole inhibition assays

These assays were performed in three different growth stages- OD_{600} of 0.2, 0.9 and 2.0, as described below. The workflow of these experiments has been shown in Figure 2.2.

2.6.1.1 OD_{600} of 0.2

Indole stock was prepared in 100% ethanol at 100 mM. The final indole concentration was 1 mM. Fresh indole stock was prepared every month.

5 mL of M9Ksupp pH 7 medium was inoculated with the overnight culture to a starting OD_{600} of 0.01. The culture was grown in 37 °C shaking incubator. At OD_{600} of 0.2, 1 mL of culture was taken from each universal and added to a 1.5 mL microcentrifuge tube. The culture was pelleted down at 17,000 *g* for 2 mins, supernatant was removed, and the pellet was washed once with 1 mL of pre-warmed (at 37 °C) M9Ksupp pH 7 medium before being resuspended in one of

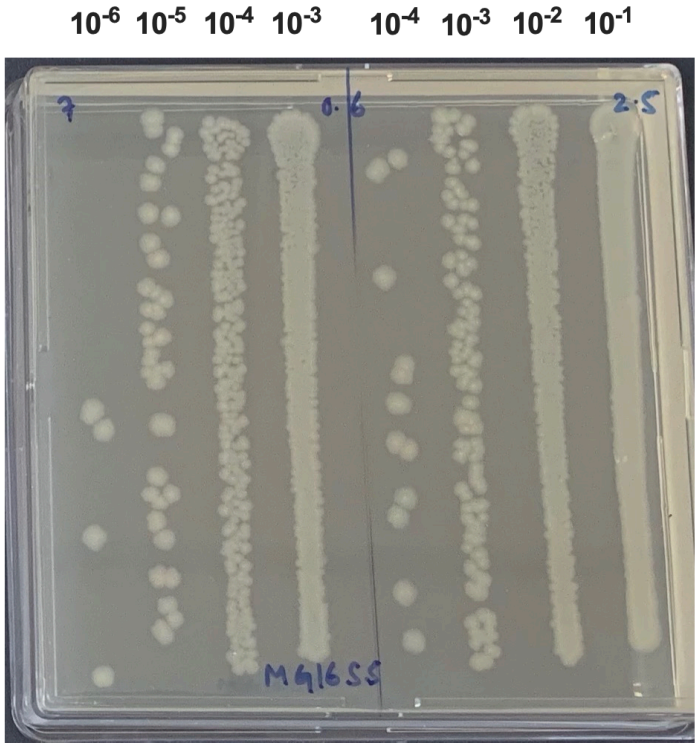


Figure 2.1: Dilutions plated out using tilt plate method in acid survival assay.
CFU/mL was estimated from the colonies counted from 10^{-5} dilution on the left and 10^{-3} dilution on the right.

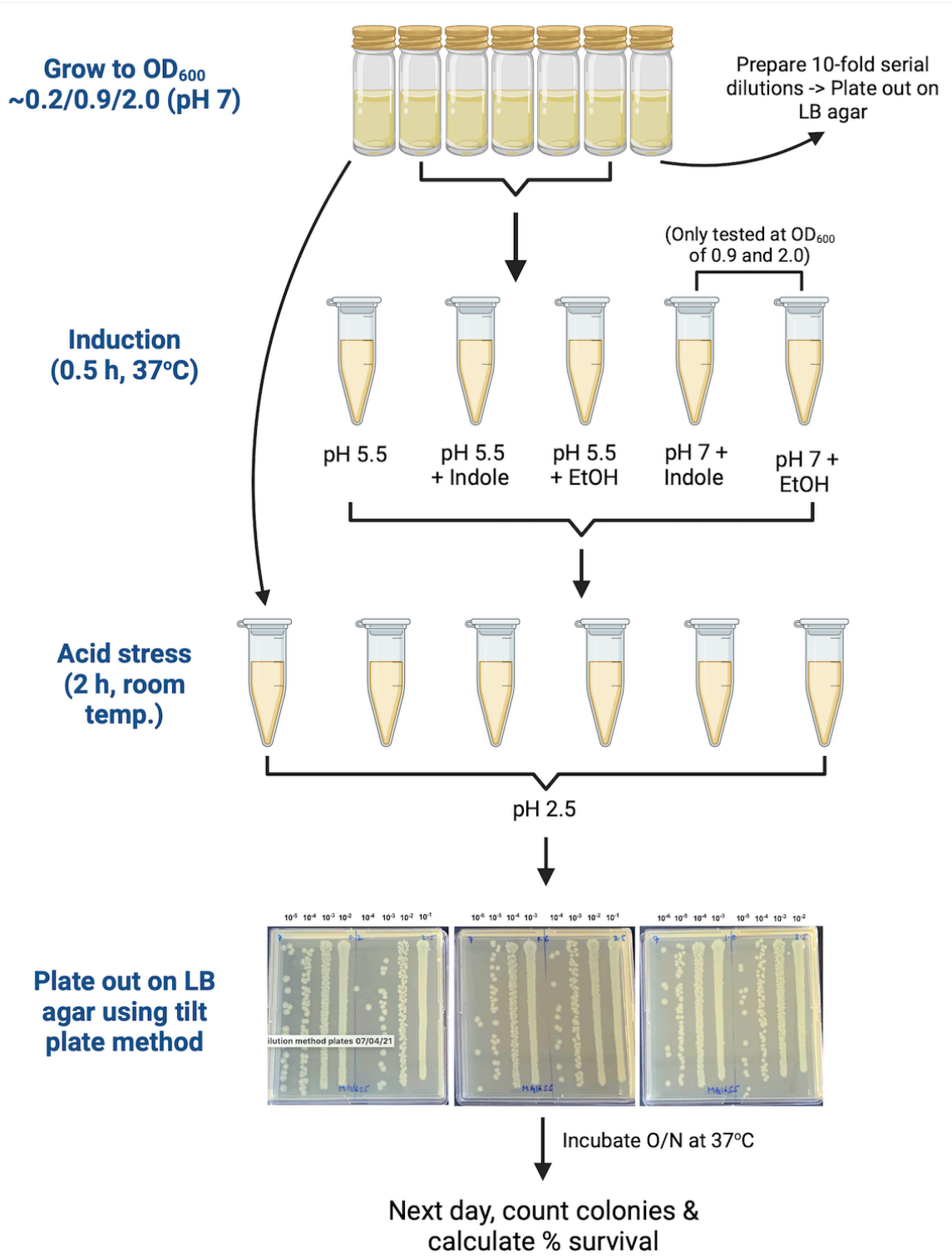


Figure 2.2: Workflow of Induction and Indole inhibition assays.

the following: 1 mL pH 7 M9Ksupp, 1 mL pH 2.5 M9Ksupp, 1 mL pH 5.5 M9Ksupp, 990 µL pH 5.5 M9Ksupp + 10 µL of 100 mM Indole dissolved in Ethanol, 990 µL pH 5.5 M9Ksupp + 10 µL of 100% Ethanol. A separate culture was set up for each condition to be tested. Depending on the medium used, samples were then treated as follows.

pH 7 and pH 2.5 M9Ksupp: 10-fold serial dilutions were made. pH 7 dilutions were plated out straightaway, and provided a baseline measure of cell numbers for all other conditions to be compared against. pH 2.5 dilutions were incubated at room temperature for 2 h before being plated out. This provided an estimate of survival at pH 2.5 without prior induction of the acid stress response.

pH 5.5, pH 5.5 + Indole, pH 5.5 + Ethanol: Cultures were incubated at 37 °C for 0.5 h, after which they were centrifuged down and resuspended in 1 mL of M9Ksupp pH 2.5 medium. 10-fold serial dilutions were prepared and incubated at room temperature for 2 h before being plated out. This provided an estimate of survival at pH 2.5 with induction of the acid stress response, and the impact of indole on the induction of the acid stress response.

The plates were incubated at 37 °C overnight and colonies were counted the next day. All plating was done using tilt plate method and percentage survival was calculated as described in section 2.5. Experiments were done in at least triplicates.

2.6.1.2 OD_{600} of 0.9

Indole stock was prepared in dimethyl sulfoxide (DMSO) at 100 mM. The final indole concentration was 1 mM.

5 mL of M9Ksupp pH 7 medium was inoculated with the overnight culture to a starting OD_{600} of 0.05. The culture was grown in 37 °C shaking incubator. At OD_{600} of 0.9, 1 mL of culture was taken from each universal, which was processed and resuspended in one of the conditions as

described previously in section 2.6.1.1. Additionally, cultures were also resuspended in two more conditions: 990 µL pH 7 M9Ksupp + 10 µL of 100 mM Indole dissolved in DMSO, and 990 µL pH 7 M9Ksupp + 10 µL of DMSO, to estimate the impact of indole on the survival of cells exposed to pH 2.5 without prior induction of the acid stress response. Depending on the medium used, samples were then treated as follows.

pH 7 and pH 2.5 M9Ksupp: 10-fold serial dilutions were made. pH 7 dilutions were plated out straightaway whereas pH 2.5 dilutions were incubated at room temperature for 2 h before being plated out.

Cultures in remaining conditions were incubated 37 °C for 0.5 h before being acid shocked and plated out as described in section 2.6.1.1.

Experiments in MG1655 were also performed with indole stock prepared in 100% ethanol.

2.6.1.3 OD_{600} of 2.0

Indole stock was prepared in DMSO at 100 mM and final indole concentration was 1 mM. The cultures were grown to an OD_{600} of 2.0 with a starting OD_{600} of 0.05. The rest of the experiment was performed as described in section 2.6.1.2.

2.7 Promoter-probe assays

2.7.1 Luciferase assay

5 mL of M9Ksupp pH 7 medium was inoculated with the overnight culture to a starting OD_{600} of 0.01. The culture was grown in 37 °C shaking incubator to mid-exponential phase (OD_{600} of 0.2-0.4). 1 mL of culture was taken from each universal and added to a 1.5 mL microcentrifuge tube. The culture was pelleted down at 17,000G for 2 mins, supernatant was

removed, and the pellet was washed once with 1 mL of pre-warmed (at 37 °C) M9Ksupp pH 7 medium before being resuspended in one of the the following: 1 mL pH 7 M9Ksupp, 1 mL pH 5.5 M9Ksupp, 990 µL pH 5.5 M9Ksupp + 10 µL of 100 mM Indole dissolved in Ethanol. A separate culture was set up for each condition to be tested. 200 µL of culture from each condition was transferred to a clear bottom, black-walled microtitre plate, and OD_{600} and luminescence readings were taken every 10 minutes for 3 hours. Samples were always kept 2 wells apart to prevent bleed-through of luminescence and wells at the edges of the plate were avoided. OD_{600} and luminescence readings were also recorded from an empty plate, the average of which was measured and subtracted from the sample readings. Luminescence readings were normalised against optical density and fold induction was calculated by dividing the normalised Luminescence at either pH 5.5 or pH 5.5 + Indole by the normalised Luminescence at pH 7. Experiments were done in at least triplicates.

2.7.2 β-galactosidase assay

The assay was performed with MG1655 Δ*evgS ydeP-lacZ* strain whereby different versions of EvgS were expressed from araBAD promoter of the arabinose operon, in the pBAD24 plasmid transformed into the strain. The overnight cultures were set up in plastic universal containers with 5 mL of LB supplemented with ampicillin to maintain the plasmid (pBAD24 plasmid contains Amp^R marker) and 0.02% (w/v) L-arabinose to induce the expression of EvgS from araBAD promoter, inoculated with a single colony. Next day, 5 mL of M9Ksupp pH 7 medium supplemented with 0.02% arabinose was inoculated with the overnight culture to a starting OD_{600} of 0.01. A separate culture was set up for each condition to be tested. The cultures were grown at 37 °C while shaking to OD_{600} of 0.2. 1 mL of culture was pelleted down at 17,000 *g* for 2 mins, washed once with 1 mL of pre-warmed (at 37 °C)

M9Ksupp pH 7 medium before being resuspended in one of the the following: 1 mL pH 7 M9Ksupp, 1 mL pH 5.5 M9Ksupp, 990 µL pH 5.5 M9Ksupp + 10 µL of 100 mM Indole dissolved in Ethanol, and 990 µL pH 5.5 M9Ksupp + 10 µL of 100% Ethanol (M9Ksupp media at pH 7 and 5.5 used for resuspending the pellets was supplemented with 0.02% arabinose). The cultures were incubated at 37 °C for 30 minutes after which their OD_{600} was measured by making 1 in 2 dilution. To measure the β-galactosidase activity, 100 µL of culture was added to the permeabilisation solution (20 µL of 0.1% (w/v) sodium dodecyl sulphate, 900 µL of Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$; pH adjusted to 7 using 1 M NaOH; supplemented with 35 µL of β-mercaptoethanol per 10 mL of buffer before use), 40 µL of chloroform ($CHCl_3$)). Cultures were vortexed for 1 minute to lyse the cells. 200 µL of ortho-Nitrophenyl-β-galactoside (ONPG, made by resuspending 40 mg of ONPG salt (Sigma-Aldrich) in 10 mL of Z-buffer; supplemented with 35 µL of β-mercaptoethanol before use) was added, the reaction was incubated at 30 °C and the timer was started. The reaction was stopped when yellow colour developed (between 10-30 mins) by adding 500 µL of 1 M sodium carbonate. Cultures were spun at 14,000 *g* for 5 minutes. 1 mL solution was removed without disturbing the pellet and the chloroform ring, to measure the OD_{420} of the solution using Z-buffer as a blank. Experiments were done in at least triplicates. Miller units were calculated using the following equation:

$$\text{Miller Units} = \frac{OD420}{OD600 \times \ Culture\ vol.(mL) \ \times \ Time\ of\ incubation\ (minutes)} \times 1000$$

2.8 Gene knockouts

2.8.1 P1 phage transduction

P1 phage transductions are used to move mutations of interest such as genes replaced with antibiotic resistance cassette, and chromosomal promoter-reporter fusions such as ydeP-lacZ from one strain to another. It is a two-step protocol which requires the preparation of a liquid P1 lysate whereby the donor strain containing the mutation of interest is infected with P1 bacteriophage, followed by transduction whereby the donor DNA is transferred to the recipient strain by infecting it with P1 phage. P1 phage transduction was done according to the protocol described in (Thomason, Costantino and Court, 2007). To prepare the liquid P1 lysate, 5 mL of LB broth was inoculated with a single colony of the donor strain containing the mutation of interest. The culture was incubated overnight in a 37 °C shaking incubator. Next day, the overnight donor culture was diluted 1:100 into 5 mL of LB broth supplemented with 0.2% glucose and 5 mM $CaCl_2$ in a 50 mL flask. The culture was grown for 30-45 mins at 37 °C while shaking, after which 100 µL of a recently prepared P1 *vir* stock was added to infect the donor strain. The culture was incubated for further 3 h while shaking until cells were lysed and cell debris appeared at the bottom of the flask. A few of drops of chloroform ($CHCl_3$) were added to the culture and shaking was continued for 5 more minutes to lyse the culture completely. The supernatant was transferred to 2 x 2mL microcentrifuge tubes and pelleted down at 17,000 *g* for 3mins. After centrifugation, the supernatant was poured into a 5 mL glass bijou bottle, labelled with the strain name on which the lysate was prepared along with the date and stored at 4 °C.

To perform P1 transduction, 5 mL of LB broth was inoculated with a single colony of the recipient strain into which the mutation of interest needs to be transferred. The culture

was incubated overnight in a 37 °C shaking incubator. Next day, 1.5 of the overnight recipient culture in a 2 mL microcentrifuge tube was pelleted down for 2 mins. The supernatant was removed, and the cells were resuspended in 0.75 mL of sterile P1 salts solution (10 mM $CaCl_2$/5 mM $MgSO_4$). 100 µL of the cells and P1 salts solution mixture was mixed with different amounts of lysate such as 0, 1, 10, and 100 µL in separate 5 mL glass bijou bottles. The phage was allowed to adsorb to the cells for 30 mins at 37 °C while shaking, after which 1 mL of LB broth supplemented with 200 µL of 1 M sodium citrate was added. Sodium citrate chelates the calcium ions and prevents the secondary phage infection. The cells were recovered at 37 °C while shaking for an hour. The cultures were added to 1.5 mL microcentrifuge tubes and pelleted down for 2 mins. The supernatant was removed, cells were resuspended in 100 µL of LB broth and plated on LB agar plates containing 5 mM sodium citrate and the appropriate antibiotic. The P1 lysate was also plated on LB agar plate supplemented with sodium citrate and appropriate antibiotic to check that no live donor cells were present in the P1 lysate. The plates were incubated overnight at 37 °C and colonies were counted next day. The colonies from the plate containing the least amount of P1 lysate were picked and streaked to achieve single colonies. The plating process to achieve single colonies was repeated twice on LB agar plates supplemented with sodium citrate to remove any residual phage.

2.8.2 Datsenko and Wanner method

This method was used to make chromosomal gene deletions and replace the deleted gene with kanamycin resistance cassette. The method has been described in Datsenko and Wanner (2000). A PCR product was generated using 56-70 nt-long primers whereby 20 nt in each primer anneal to the FRT regions on either side of the kanamycin resistance cassette in pKD4 plasmid (PCR template) and the remaining 36-50 nt are homologous to the regions

adjacent to the gene to be deleted in target strain. The PCR product was gel extracted, PCR purified and quantified with nanodrop.

The target strain was made electrocompetent and transformed with temperature-sensitive pKD46 plasmid (stable at 30 °C) according to the protocols in sections 2.9.3 and 2.9.4. pKD46 encodes three genes (*γ, β, exo*) of the lambda Red system which can be expressed from an arabinose-inducible promoter thereby allowing the PCR product to be integrated into the host chromosome through homologous recombination. The transformed cells were selected on LB agar plates supplemented with ampicillin (pKD46 encodes Amp^R). pKD46-transformed cells were then cultured in LB supplemented with ampicillin and 0.2% (w/v) arabinose and made electrocompetent. 50 µL of competent cells were transformed with ~100 ng of PCR product. Half of the cells (500 µL) were spread on LB agar plates supplemented with kanamycin, and incubated overnight at 37 °C. If no colonies were present next day, then the remaining 500 µL of transformed cells which were left to recover on the bench overnight were spread on LB agar plates containing kanamycin and incubated overnight at 37 °C.

2.9 Molecular biology techniques

Protocols performed with kits were done following the manufacturer's instructions unless otherwise stated.

2.9.1 Plasmid mini-prep

Plasmid DNA was isolated from bacterial cells using QIAprep spin Miniprep Kit (Qiagen).

2.9.2 Site-directed mutagenesis

Site-directed mutagenesis of plasmid DNA was performed using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent). The mutations in the plasmids were confirmed through Sanger sequencing from Source Bioscience.

2.9.3 Electrocompetent cells

50 mL of LB in a 250 mL shaker flask was inoculated with 100 µL (1:500 dilution) of overnight culture and grown at required temperature (usually 37 °C; 30 °C in case of cells containing temperature-sensitive vector) in a shaking incubator. Once the culture reached an OD_{600} between 0.4 and 0.6, it was transferred to a 50 mL falcon tube after which it was always kept on ice. Cells were pelleted down at 5,000 rpm for 10 minutes at 4 °C. The supernatant was removed, and pellet was washed three times in ice-cold 10% (v/v) glycerol by reducing volume each time starting with 25 mL and reducing it to 10 mL followed by 5 mL. Cells were finally resuspended in 250 µL of 10% glycerol to prepare 200x concentrated cells than the initial culture. 50 µL aliquots were prepared in 1.5 mL microcentrifuge tubes and stored at -80 °C.

2.9.4 Electroporation

50 µL of electrocompetent cells were thawed and mixed with 1-3 µL of DNA (<100 ng) and incubated on ice for 10 mins. The cells + DNA mixture was transferred to a chilled 1 mm electroporation cuvette and pulsed at 2.2 KV in an electroporator (Eppendorf). Cells were recovered in a 1.5 mL microcentrifuge tube containing either 1 mL of LB or SOC medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) (Invitrogen). Recovered cells were incubated at required temperature (37/30

°C) while shaking for an hour before being plated out on LB agar plates containing appropriate antibiotic. The plates were incubated at 37 °C overnight.

2.9.5 Polymerase Chain Reaction

PCR reactions used to confirm gene knockouts were done using Taq polymerase whereas PCR reactions performed for cloning purposes were done using high-fidelity Q5 Polymerase (NEB).

For colony PCR, a single colony was resuspended in 20 µL of nuclease-free water (Invitrogen) and boiled at 99 °C for 10 mins. While using liquid culture, 10 µL of overnight culture was boiled at 99 °C for 10 mins and diluted in 90 µL of nuclease-free water. The composition of a 50 µL PCR reaction mixture is given in Table 2.4, and the thermocycler conditions routinely used are shown in Table 2.5.

2.9.6 Agarose gel electrophoresis

DNA fragments in PCR reactions were separated using gel electrophoresis. 1% (w/v) agarose gel was prepared in 1x TAE buffer (50x TAE: 2 M Tris base, 1 M acetic acid, 0.05 M EDTA in water) and 6 µL of SYBR™ Safe DNA Gel stain (Invitrogen) per 100 mL of agarose. Hyperladder™ 100 bp and Hyperladder™ 1 Kb (Bioline) were used as DNA reference markers. PCR reactions with 2x MyTaq™ Red Mix were run without a loading dye whereas those with Q5® High-Fidelity 2x Master Mix were mixed with Gel Loading Dye, Purple (6X) (New England BioLabs). The gel, which had a size of 15 × 10 cm, was placed in a tank containing 1x TAE buffer and run at 120 V for 30-45 mins. Gels were visualised and imaged in Gel Doc light box (Bio-Rad).

Table 2.4: Composition of a 50 µL PCR reaction mixture.

Reagent	Amount
2x MyTaq™ Red Mix (Bioline)/ Q5® High-Fidelity 2x Master Mix (New England BioLabs)	25 µL
10 µM of Primer 1	2 µL
10 µM of Primer 2	2 µL
DNA template	2 µL
Nuclease-free water (Invitrogen)	19 µL

Table 2.5: Thermocycler conditions routinely used for PCR reactions.

PCR Step	Temperature	Time	Number of Cycles
Initial Denaturation	95 °C	5 min	1
Denaturation	95 °C	30 s	30
Annealing	3 °C below the T_m of primers	30 s	
Extension	72 °C	30 s/Kb of DNA	
Final Extension	72 °C	7 min	1
Hold	4 °C	∞	-

2.9.7 Gel Extraction of DNA fragments

Agarose gel was placed under the blue light LED illuminator (Geneflow) to assist with the visualisation of DNA bands while being excised out of the gel using a blade. The bands were then transferred to 1.5 mL microcentrifuge tubes and extracted using QIAquick Gel Extraction kit (Qiagen).

2.9.8 PCR Purification

Gel extracted DNA fragments and PCR reactions were purified using QIAquick PCR purification kit (Qiagen).

2.9.9 DNA quantification

2.9.9.1 Qubit

DNA was quantified using Qubit™ dsDNA HS Assay kit (Invitrogen) using 1 μL of sample.

2.9.9.2 Nanodrop

DNA was quantified with Thermo Scientific NanoDrop™ ND-1000 Spectrophotometer using 1 μL of sample.

2.10 Sequencing

2.10.1 Sanger sequencing

PCR products and plasmid regions were Sanger sequenced by Source Bioscience.

Sample and oligonucleotide preparation was done by following the instructions provided on their website (https://www.sourcebioscience.com).

2.10.2 Whole genome sequencing

Strains were whole genome sequenced by Microbes NG using 250 bp paired-end reads to 30X coverage. The sequencing reads were aligned to the reference genomes using breseq, version 0.31.1 (Deatherage and Barrick, 2014).

2.11 Preparation of O157:H7 Sakai TraDIS-Xpress library

2.11.1 Transposome preparation

Tnp001 is a mini-Tn5 transposon that carries an outward-facing IPTG-inducible *tac* promoter present downstream of the kanamycin resistance gene, which was used to prepare the O157:H7 Sakai TraDIS-Xpress mutant library (Figure 2.3). The transposon DNA was PCR-amplified using Q5 polymerase from BW25113 TraDIS-Xpress library using oligonucleotides mentioned in Table 2.3 (Yasir *et al.*, 2020). After gel electrophoresis, the PCR product was gel extracted, purified, and quantified using Qubit. The sequence of the transposon DNA was confirmed through Sanger sequencing (oligonucleotides given in Table 2.3). The transposome was prepared by mixing the purified transposon DNA with EZ-Tn5 transposase according to the supplier's instructions (Lucigen).

2.11.2 Preparation of electrocompetent cells and electroporation

This protocol is different to the one described in section 2.9.3. Here, a different broth as we all a higher volume of the broth was used to culture the cells. Additionally, the cells

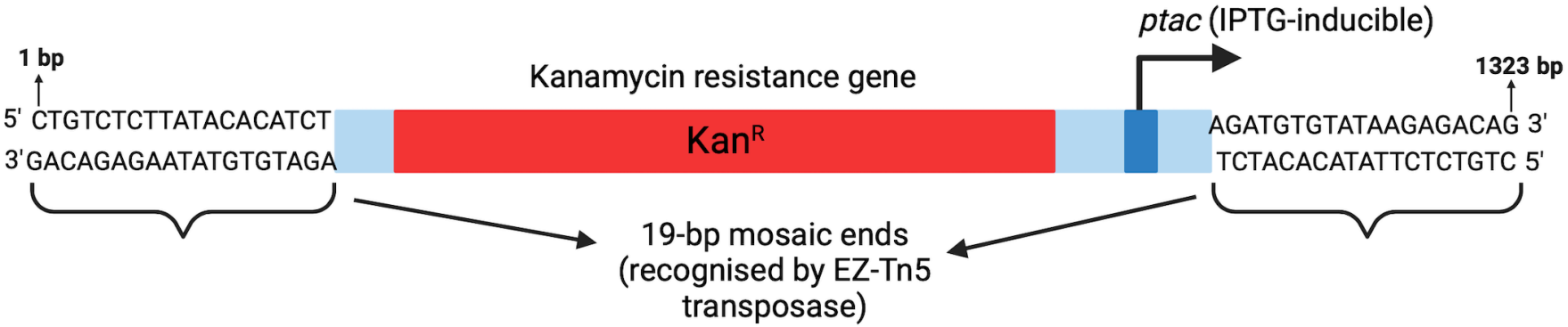


Figure 2.3: Map of Tnp001, a mini-Tn5 transposon.

were cultured to a different OD_{600}. All tips, stripettes, falcon tubes, and eppendorfs used in this protocol were also pre-chilled at -20 °C. Brief description of the protocol is given below.

800 mL of 2xTY broth in 2 L flask was inoculated with 7 mL of overnight culture of O157:H7 Sakai and grown to an OD_{600} of 0.3-0.4 at 37 °C while shaking. The culture was transferred to pre-chilled 50 mL falcon tubes, incubated on ice for 30 minutes (which was not done for the protocol given in section 2.9.3), washed 5 times with ice-cold 10% glycerol and finally resuspended in 1 mL of ice-cold 10% glycerol to prepare 800x concentrated cells than the initial culture. The resuspended cells were divided into 60 µL aliquots.

The protocol for electroporation here is also different to the one described in section 2.9.4. Here, the volume of transposome used was much smaller than the volume of DNA used previously. Furthermore, the cells + DNA mixture was incubated on ice for longer before being electroporated. The electroporation was done in 2 mm cuvettes as opposed to 1 mm cuvettes used previously. The brief description of the protocol is as follows. 0.2 µL of transposome was mixed with 60 µL of electrocompetent cells and incubated on ice for 30 minutes. The cells + DNA mixture was transferred to a chilled 2 mm electroporation cuvette and pulsed at 2.2 KV in an electroporator. Cells were recovered in a 1.5 mL microcentrifuge tube containing 1 mL of SOC medium. Recovered cells were incubated at 37 °C while shaking for an hour, diluted further in 1 mL of SOC medium and plated equally (400 µL) on five square (10 cm x 10 cm) LB agar plates supplemented with 50 µg/mL of kanamycin (minimum inhibitory concentration (MIC) of O157:H7 Sakai). MIC of O157:H7 Sakai was determined by plating on successively higher levels of kanamycin and looking for the level at which no survivors were seen. The plates were incubated at 37 °C overnight. The following day, 1 mL of 2xTY was poured on each plate, colonies were scraped off using L-shaped glass spreader and added to a 5 mL eppendorf. Approximately 1,272 colonies were present on each plate. Colonies from all

transformations were finally pooled into a 50 mL falcon tube with 10-15% glycerol and stored at -80 °C.

2.11.3 DNA extraction and fragmentation for preparation of TraDIS sequencing library

The process of TraDIS-Xpress sequencing library preparation starting with DNA extraction is shown in Figure 2.4. gDNA was extracted from ~10^9 bacterial cells using RTP® Bacteria DNA Mini Kit (Stratec) following instructions for protocol 2 (Isolation of DNA from bacteria pellets (1 x 10^9 bacterial cells)). The extracted DNA was then quantified using Qubit. For TraDIS library preparation, 500 µL of 2 µg/mL of extracted DNA was processed. DNA was sheared with low intensity sonication (Frequency: 20kHz, Power: 160W) using following settings: 30 seconds on, 90 seconds off, 13 cycles, on a Diagenode bioruptor® plus. Fragmented DNA in a volume of 500 µL was concentrated to 55.5 µL using a concentrator (Eppendorf).

2.11.4 TraDIS library preparation for sequencing

The sequencing libraries were prepared using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® using the protocols provided with the kit. The fragmented DNA from section 2.11.3, has sticky ends which were repaired to achieve blunt ended double-stranded DNA fragments. Subsequently, an A tail was added to the 3' end, so that NEBNext Adaptor could be ligated on both sides of the fragments. Next, the uracil in the hairpin loop adaptors was removed using USER enzyme to create overhangs. Protocols 1 and 2 in the NEBNext Illumina kit manual were used for end repair and adaptor ligation respectively. Following adaptor ligation, ~250 bp DNA fragments were selected following instructions provided in protocol

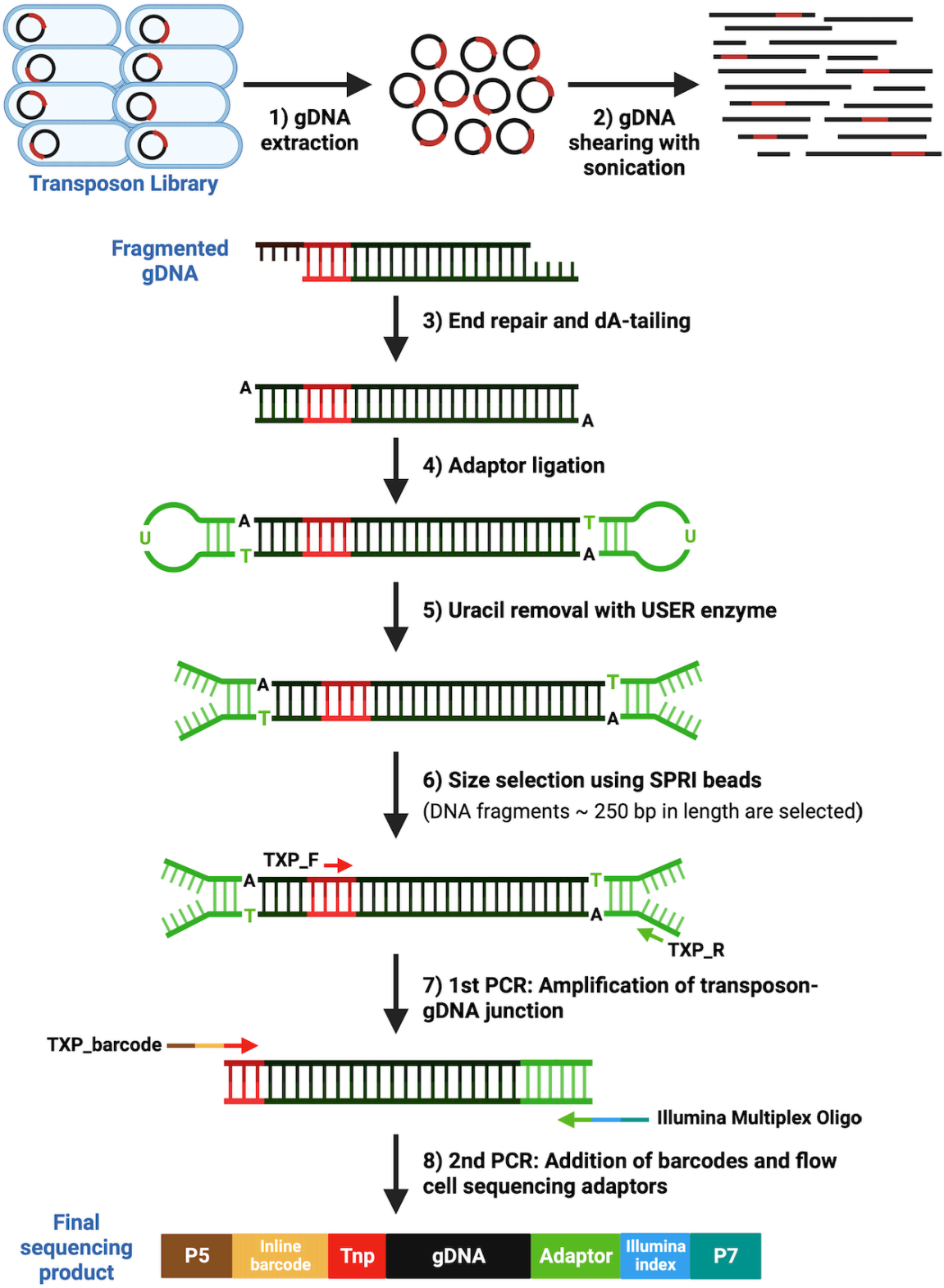


Figure 2.4: Steps involved in the preparation of a TraDIS-Xpress library for Illumina sequencing.
(Adapted from Milner (2021); Figure 2.7.1. Created with Biorender.com.)

3A. 44 µL and 20 µL of SPRI beads in the first and second bead selection steps respectively, were used. After size selection, the fragments containing transposon and neighbouring gDNA were amplified using PCR. The oligonucleotides TXP_F and TXP_R that bind respectively to the transposon and adaptor within the DNA fragments were used for this PCR step. PCR products were cleaned up of dNTPs, primers, salts and other contaminants using 36 µL of SPRI beads and following protocol 5 in the manual. The DNA in this clean up step was eluted in 17 µL of buffer EB instead of 33 µL as stated in the protocol. A final PCR was done using TXP_barcode and Illumina barcode oligos (NEBNext® Multiplex Oligos for Illumina®, Index Primers Set 1 or 2) to add inline and Illumina barcodes, as well as P5 and P7 sequence adaptors that anneal to their complementary sequences on the flow cell. Inline barcode was present on the transposon side of the DNA fragment and Illumina barcode on the gDNA side. Both barcodes allowed different samples to be multiplexed and included on the same sequencing run. The reagents and thermocycler conditions used in first and second PCR reactions are mentioned in Tables 2.6 and 2.7 respectively. The sequences of oligonucleotides TXP_F, TXP_R, and TXP_barcode are given in Table 2.3.

2.11.5 Preparation of SPRI beads

After adaptor ligation during TraDIS-Xpress library preparation, size selection was implemented to select the right size gDNA fragments. Furthermore, PCR products after PCR 1 and PCR 2 were cleaned to remove dNTPs, primers, salts, and other contaminants. The size selection and PCR clean-up were done using solid-phase reversible immobilization (SPRI) technique with in-house prepared SPRI beads, which are a cheaper substitute for AMPure XP beads by Beckman Counter. SPRI beads are magnetic beads with carboxyl groups present on their surface through which they can bind DNA reversibly in the presence of polyethylene

Table 2.6: Reagents used in the first and second PCR reactions during TraDIS-Xpress library preparation.

Reagent	Amount
NEBNext Q5 Hot Start HiFi PCR Master Mix	25 µL
10 µM of Primer 1 (TKK_F for PCR1; TKK_barcode for PCR2)	2.5 µL
10 µM of Primer 2 (TKK_R for PCR1; Illumina Multiplex Oligo for PCR2)	2.5 µL
DNA template	15 µL
Nuclease-free water	5 µL

Table 2.7: Thermocycler conditions used for the first and second PCR reactions during TraDIS-Xpress library preparation.

PCR Step	Temperature	Time	Number of Cycles
Initial Denaturation	98 °C	48 s	1
Denaturation	98 °C	15 s	10 for PCR1; 20 for PCR2
Annealing	65 °C	30 s	
Extension	72 °C	30 s	
Final Extension	72 °C	1 min	1
Hold	4 °C	∞	-

glycol (PEG) and salts (DeAngelis, Wang and Hawkins, 1995). The DNA bound to the beads is separated from the contaminants by pulling the beads down using a magnet, washing the beads multiple times with ethanol before eluting the DNA in buffer EB (Qiagen). DNA fragments approximately between 200 bp and 200 kb can be purified by changing the DNA to beads ratio.

The SPRI beads were prepared according to the protocol described by Jolivet and Foley (2015). The efficacy of the beads was confirmed by using 1 µg of fragmented DNA from Bioline DNA HyperLadder^{TM} 100 bp and processing it using both AMPure XP and in-house prepared SPRI beads through the first size-selection protocol of TraDIS-Xpress library preparation. The eluted gDNA fragments were run on 2% agarose gel. The presence of DNA fragments ranging between 300 and 500 bp confirmed the efficacy of SPRI beads.

2.11.6 TraDIS sequencing and data analysis

2.11.6.1 Sequencing and genome alignment of reads

The prepared TraDIS-Xpress libraries were sequenced using Illumina on NovaSeq 6000 system with S4 flow cell type generating 150 bp paired-end reads (Novogene). The sequencing reads from different samples were demultiplexed using Illumina barcodes and provided in separate fastq files by Novogene. The identification of reads containing the transposon sequence, followed by the trimming of the transposon sequence and alignment of the remaining gDNA sequence to the reference genome was done using a pipeline created by Ashley Robinson, Mathew Milner and Emily Goodall. The whole pipeline is available on a linux server provided by the MRC Cloud Infrastructure for Bioinformatics (CLIMB) project under the folder named 'TraDIS_analysis_v1.8m1' (Connor *et al.*, 2016). Briefly, the reads corresponding to a sample were identified using an inline barcode ranging between 6 and 9 bps.

Once the inline barcode was identified and trimmed, the reads were filtered in two additional steps based on the presence of sequences corresponding to the transposon. These steps ensured that only the sequencing reads which had the transposon sequence present in them were aligned to the reference genome. In the first step, the reads were screened for the presence of a 22 bp sequence within which 3 mismatches were allowed. The selected reads were then screened for the presence of a 10 bp sequence within which 1 mismatch was allowed. A final filtering step involved removing the reads which were either of poor quality or less than 20 bp. The remaining reads were then aligned to the reference genome of O157:H7 Sakai obtained from NCBI database (NC_002695.2; Hayashi *et al.*, 2001). The output files were provided for the number of unique transposon insertion sites per coding DNA sequence (CDS), and the number of reads per CDS and insertion site. Insertion index score which corresponds to number of unique insertion sites (UIS) normalised by gene length, and RPM which corresponds to number of reads per gene normalised by sequencing depth, were also calculated. Insertion sites were visualised using Artemis genome browser (Carver *et al.*, 2012). Circular genome diagrams were generated using DNA plotter.

2.11.6.2 Prediction of essential genes

The genes which are essential for the survival of O157:H7 Sakai were predicted using the method described in Goodall *et al.* (2018) and an R script gifted by Emily Goodall. The essential genes identified were used for pangenome analysis in Chapter 6, section 6.2.3. These genes were also excluded from the analysis of O157:H7 Sakai TraDIS data in Chapter 6. This is because essential genes either do not tolerate any or only a few insertions, therefore their conditional essentiality cannot be determined. The insertion index score described in section 2.11.6.1 follows a bimodal distribution when plotted on a histogram. Using an R (RStudio, version 1.1.456) package, R MASS library, an exponential distribution was fitted to

the left mode containing essential genes with low insertion index scores, and a gamma distribution was fitted to the right mode containing non-essential genes with high insertion index scores. Genes were assigned to the left or right mode depending on their log_2 likelihood ratio (LLR). Genes with LLR < -3.6 were considered essential and were 12 times more likely to be in the left mode than the right mode, and genes with LLR > 3.6 were considered non-essential and these were 12 times more likely to be in the right mode than the left mode. Genes with LLR between -3.6 and 3.6 were considered unclear.

2.11.6.3 Roary analysis

Pangenome analysis was done to identify the core and accessory genes of K-12 strains, MG1655 and BW25113, and O157:H7 Sakai using Roary (Page *et al.*, 2015). Roary takes the input genome, identifies the coding regions and converts them into protein sequences. BLAST is then done on those protein sequences to identify the homologous genes among different genomes and produce an output for genes present in all genomes (core genome), and the genes present only in individual genomes (accessory genome). Genome alignment was performed with reference genomes obtained from NCBI database: MG1655 (U00096.3; Riley *et al.*, 2006), BW25113 (CP009273.1; Grenier *et al.*, 2014), and O157:H7 Sakai (NC_002695.2; Hayashi *et al.*, 2001). For BLAST analysis, a cut-off of 95% sequence identity was used. TraDIS data obtained for essential gene analysis for MG1655 (Milner, 2021), BW25113 (Goodall *et al.*, 2018), and O157:H7 Sakai (this study) was manually aligned to the Roary output.

2.12 Optimisation for TraDIS experiments

2.12.1 Competition experiments

Competition experiments were done to compare the fitness of a lac$^+$ knockout strain (which was prepared by knocking the gene of interest out from WT MG1655, which is naturally lac$^+$) with lac$^-$ KH001 strain. KH001 strain was prepared by replacing the *lac* promoter with pKH3 cloning site to achieve a lac$^-$ strain which could be used in competition experiments. A separate O/N culture for each strain was set up in a universal container containing 5 mL LB. Next day, 5 mL of M9Ksupp pH 7 medium in a universal container was inoculated to a starting OD_{600} of 0.01 per strain. The strains were cultured to an OD_{600} of ~0.2 after which both cultures were normalised to an exact OD_{600} of 0.2 and mixed in equal proportions. 10-fold serial dilutions were prepared and 50 µL of relevant dilutions were plated on macConkey agar supplemented with 1% lactose. 1mL of mixed culture was pelleted down at 17,000 *g* for 2 mins, washed once in pre-warmed 1 mL M9Ksupp pH 7 medium, and resuspended in 990 µL of M9Ksupp pH 5.5 with either 10 µL of 100 mM indole dissolved in Ethanol or 10 µL of 100% Ethanol. The cultures were incubated at 37 °C for 0.5 h after which 10-fold serial dilutions were prepared and plated. The remaining 900 µL of culture was pelleted down and resuspended in pre-warmed M9Ksupp pH 2.5 medium. 10-fold serial dilutions were prepared, incubated at room temperature, and plated after 15, 30 and 60 mins. The plates were incubated at 30 °C overnight and number of red and white colonies were counted the next day. The survival was expressed as a percentage and calculated as follows:

$$\%\, Survival = \frac{Number\ of\ colonies\ at\ pH\ 5.5\ or\ pH\ 2.5}{Number\ of\ colonies\ at\ pH\ 7\ in\ 50{:}50\ mixed\ culture} \times 100$$

Fitness of the knockout strain (Strain A) relative to the WT strain (Strain B) was measured using selection rate, described in Travisano and Lenski (1996), which was calculated as follows:

$$Selection\ rate\ (s) = r_a - r_b = \ln\left(\frac{A_f}{A_i}\right) day^{-1} - \ln\left(\frac{B_f}{B_i}\right) day^{-1}$$

A_f/B_f and A_i/B_i refer to the final and intial colony counts at pH 5.5/pH 2.5 and pH 7 respectively.

For competition experiments with outgrowth, the culture acid stressed at pH 2.5 was incubated at room temperature for 30 mins. The culture was then pelleted down, resuspended in 5 mL of M9Ksupp pH 7 medium, and incubated in 37 °C shaking incubator for 3 hours of outgrowth. After 1, 2 and 3 h, the culture was removed from the incubator to prepare 10-fold serial dilutions which were plated straightaway. The experiments were done in at least triplicates.

2.12.2 Scaled up induction and indole inhibition assays with outgrowth

These assays were done as described in section 2.6.1.1. However, some differences were present. The volume of medium used to culture the strain, and the volume of cultures which were induced at pH 5.5 and acid stressed at pH 2.5 were all increased by 10 times. The strains were cultured in 250 mL shaker flasks and the induction and acid shock steps were done in 15 mL falcon tubes. The cells were pelleted down at 8,000 *g* for 7 mins. MG1655 was acid shocked at pH 2.5 for 30 mins whereas O157:H7 Sakai was acid shocked for 90 mins. After acid shock, the pH 2.5 cultures were pelleted down followed by separate resuspension in 50 mL of M9Ksupp pH 7 medium and incubated in 37 °C shaking incubator for 3 h. At 0

(right after resuspension in 50 mL of pH 7 medium), 0.5, 1, 2 and 3 h, the OD_{600} of the cultures was measured and 10-fold serial dilutions were prepared and plated.

2.13 TraDIS experiments

The TraDIS experiments with MG1655 and O157:H7 Sakai were done under four different conditions which are shown in Table 2.8. 50 mL of M9Ksupp pH 7 medium in a 250 mL shaker flask was inoculated to a starting OD_{600} of 0.01 (~10^7 transposon mutants) with the transposon library and cultured to OD_{600} of ~ 0.2 in a 37 °C shaking incubator. At this point, 10-fold serial dilutions were prepared and 8 µL from relevant dilutions was plated on LB agar plates using tilt plate method. Additionally, 10% glycerol stock of the culture was prepared for fossil records. From the remaining culture, 20 mL of culture was pelleted down at 8,000 *g* for 7 mins, washed once in pre-warmed 20 mL M9Ksupp pH 7 medium, and resuspended in 19.8 mL of M9Ksupp pH 5.5 with either 200 µL of 100 mM indole dissolved in Ethanol or 200 µL of 100% Ethanol. The culture was incubated at 37 °C for 30 mins after which it was equally split into 2 x 10 mL aliquots. Both aliquots were pelleted down and resuspended in 10 mL of M9Ksupp medium, one in pH 2.5 and the other in pH 7. The cultures were incubated at 25 °C for 30 mins for MG1655 and for 90 mins for O157:H7 Sakai, after which they were pelleted down as following:

- pH 5.5 + Indole -> 2.5 = Pellet entire 10 mL
- pH 5.5 + Indole -> 7 = Pellet 1 mL ('pH 5.5 + Indole -> 2.5' gives 10% survival, and 'pH 5.5 + Indole -> 7' should give ~100% survival, therefore one-tenth of the culture was used)

Table 2.8: The conditions under which TraDIS experiments were performed with MG1655 and O157:H7 Sakai.

Growth	Induction (30 mins)	Treatment (30/90 mins)	Outgrowth (3 h)	Day of the experiment
M9K pH 7	pH 5.5 + Indole (1mM)	pH 2.5	M9K pH 7	Day 1
		pH 7		
	pH 5.5 + EtOH (same volume as indole stock)	pH 2.5		Day 2
		pH 7		

- pH 5.5 + EtOH-> 2.5 = Pellet 1.4 mL (survival is 7 times higher than 'pH 5.5 + Indole -> 2.5')
- pH 5.5 + EtOH-> 7 = Pellet 1 mL

Each pellet was resuspended in 50 mL of M9Ksupp pH 7 medium and incubated in 37 °C shaking incubator for 3 h. At 0 and 3 h, the OD_{600} of the cultures was measured and 10-fold serial dilutions were prepared and plated. After 3 h of outgrowth, 10% glycerol stocks were prepared for fossil records and the remaining culture was pelleted down, washed once in M9Ksupp pH 7 medium and processed for genome extraction followed by library preparation as described in sections 2.11.4 and 2.11.5. All experiments were done in triplicates.

The forward primers used in PCR1 and PCR2 during MG1655 TraDIS and O157:H7 Sakai TraDIS-Xpress library preparation were different due to the difference in the sequences of the transposons present in MG1655 and O157:H7 Sakai transposon libraries. The primer sequences used in PCR1 and PCR2 for preparation of both sequencing libraries have been provided in Table 2.3.

2.14 Analysis of sequencing data from TraDIS experiments

2.14.1 Genome alignment of reads

The reads containing the transposon sequence were filtered as described in section 2.11.7.1. However genome alignment was performed using BioTraDIS pipeline described in Barquist *et al.* (2016), which requires the reference genomes to be in EMBL format (MG1655: GCA_000005845.2; O157:H7 Sakai: GCA_000008865.2). Output was provided for the number of reads and UIS per gene, as well as the insertion index score per gene. A custom R script

provided by Mathew Milner, which uses edgeR package to compare the read counts per gene among different samples, was used to perform principal component analysis (PCA). PCA was then used to visualise clustering among replicates of the same and different conditions.

2.14.2 Comparison between ITL and conditions using BioTraDIS

Comparison between Initial TraDIS Library (ITL) and TraDIS libraries from different conditions was done using BioTraDIS pipeline. BioTraDIS uses edgeR package to identify significant differences in read counts between two samples which is expressed as a log-fold change accompanied by an adjusted p value (q value) (Robinson, McCarthy and Smyth, 2010).

2.14.3 Gene set enrichment analysis (GSEA)

Gene set enrichment analysis (GSEA) was performed using GSEA software, version 4.2.3, downloadable from: http://www.gsea-msigdb.org/gsea/index.jsp (Mootha *et al.*, 2003; Subramanian *et al.*, 2005). The reference lists for the genes involved in different metabolic and regulatory pathways were prepared using information available in KEGG and Regulon DB databases respectively. The reference lists were kindly gifted by Dr Mathew Milner. The gene sets involved in metabolic and regulatory pathways were ranked depending on the log-fold change between ITL and each condition. Each gene set was then assigned an enrichment score (ES). ES was further normalised to account for the number of genes involved in a set and a normalised ES (NES) was calculated. The statistical significance of NES was calculated and corrected for multiple hypothesis testing (FDR q-value). A cut-off of FDR < 0.05 was applied and the genes contributing to the NES in these pathways were determined using leading-edge analysis.

2.15 TraDIS validation experiments

2.15.1 Competition experiments

The results of TraDIS experiments were only validated for MG1655. Two separate overnight cultures were set up, one for the mutant of interest and the other for KH001. Next day, 50 mL of M9Ksupp pH 7 medium in a 250 mL flask was inoculated to a starting OD_{600} of 0.01 using an equal proportion of the mutant and KH001. The culture was grown to OD_{600} of 0.2, 10-fold serial dilutions were created, and appropriate dilutions were plated on macConkey agar plates supplemented with 1% lactose. The plates were incubated at 30 °C overnight and the number of red and white colonies were counted the following day. The selection rate was calculated as described in section 2.12.1.

2.15.2 Survival assays

An overnight culture was set up for the mutant of interest. Next day, 50 mL of M9Ksupp pH 7 medium in a 250 mL flask was inoculated to a starting OD_{600} of 0.01. The culture was grown to OD_{600} of ~0.2. 1 mL of culture was acid stressed at pH 2.5 for 2 h at room temperature with and without pre-induction at pH 5.5 at 37 °C for 30 mins. 10-fold serial dilutions were created from cultures at pH 2.5 which were plated after 0.5 h and 2 h using tilt plate method. The plates were incubated overnight at 37 °C and colonies counted the following day. The survival was calculated as described in section 2.5.

2.16 Statistical analysis

Statistical analysis was performed using GraphPad Prism 9 for macOS, version 9.3.1 (350), downloadable from: www.graphpad.com.

The data for % survival and fold induction were log-transformed. Pairwise comparisons were done using Unpaired T-test and where more than two datasets were compared, One-way ANOVA was used instead. These tests were further followed by a multiple comparisons test to control the false discovery rate. The difference between two datasets was considered significant if the p value was below 0.05.

Chapter 3

Promoter probe assays confirm the role of EvgS and the impact of indole on induction of the AR2 response

3.1 Introduction

Bacteria use two-component signal transduction systems (TCSs) to sense and respond to their fluctuating environment. The EvgAS TCS, where EvgS is the sensor kinase and EvgA is the response regulator, activates one of the five and best characterised acid resistance systems, AR2, in exponentially growing *E. coli* cells (Masuda and Church, 2003; Itou, Eguchi and Utsumi, 2009). The activation of EvgS occurs in a pH range of 5.5 - 5.7 in minimal media supplemented with amino acid glutamate (Ma, Masuda and Foster, 2004; Burton *et al.*, 2010). The net effect of this activation is increased resistance against extreme acid stress of pH 2.5 and below. EvgS is homologous to BvgS of *Bordetella pertussis,* the causative agent of whooping cough (Herrou *et al.*, 2010; Sen *et al.*, 2017). Both kinases possess a large periplasmic domain made up of two Venus flytrap (VFT) domains. The periplasmic domain of sensor kinases is a potential target for antimicrobial agents. Indeed, nicotinate has been shown to bind the VFT domain of BvgS and inactivate it by shifting it to a phosphatase state (Dupré *et al.*, 2015). To date, no ligands have been identified that inhibit EvgS. In a search for possible EvgS ligands, a Masters student in the lab (Morissa Bradnick) screened a range of small molecules which could be the potential modulators of EvgS activity. Using promoter probe assays, it was then shown that bacterial signalling molecule indole inhibits the induction of AR2 promoter, *ydeP* (Boon *et al.*, 2020). This chapter, some of the results of which were published as part of Boon *et al.* (2020), shows that indole has an inhibitory effect on the induction of various AR2 promoters indicating that indole inhibits the complete AR2 acid resistance response. Further evidence is also provided for the likely target of indole inhibition to be EvgS. Additionally, an attempt at mapping the indole binding site within EvgS has been made.

3.2 Results

3.2.1 Induction of AR2 promoters and inhibition with indole using luciferase assay

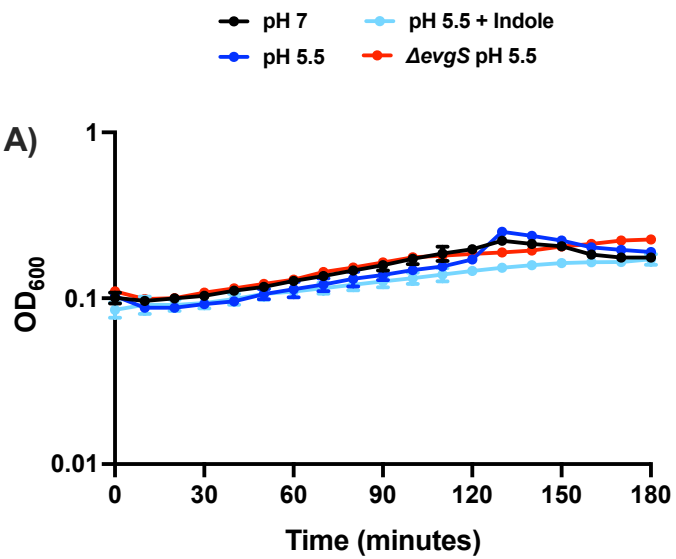
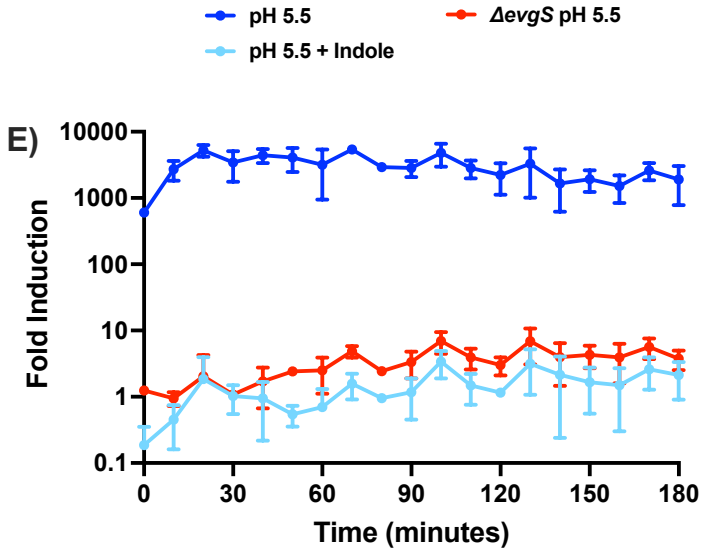
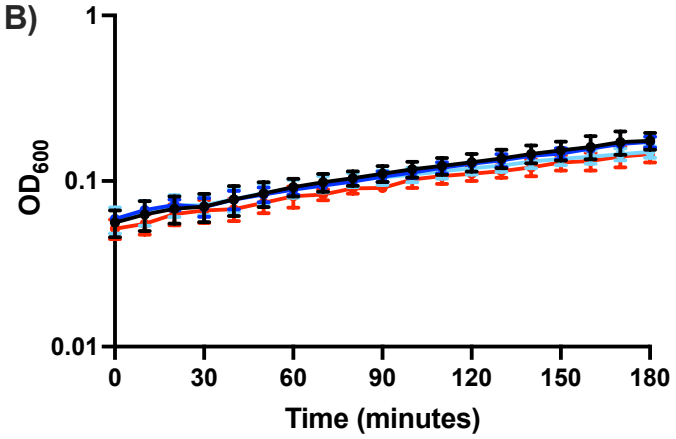
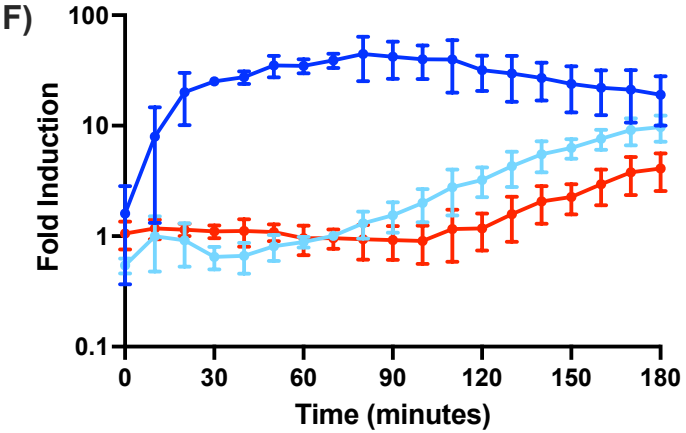
In Boon *et al.* (2020), we screened a range of different compounds such as cadaverine, γ-aminobutyric acid, glycine betaine, indole, ornithine, sarcosine, and spermine to measure the levels of activity from the *ydeP* promoter in the MG1655 *ΔevgS ydeP-lacZ* strain. These compounds were selected because they are produced within *E. coli* cells as metabolites and based on the predicted EvgS periplasmic domain structure, described in Sen *et al.* (2017), they could be the potential modulators of EvgS activity. *ydeP* is one of the promoters involved in the AR2 acid resistance response which is regulated directly by the EvgA response regulator (Burton *et al.*, 2010). The MG1655 *ΔevgS ydeP-lacZ* strain containing a chromosomal *ydeP-lacZ* fusion was used so that the activity of the *lacZ* reporter gene could be measured from the *ydeP* promoter. The strain lacked an *evgS* gene on the chromosome. However, a wild-type copy of the *evgS* gene was provided on pBAD24 plasmid, downstream of the L-arabinose-inducible *araBAD* promoter, allowing the gene to be expressed when needed (Guzman *et al.*, 1995). Of all the tested compounds, indole showed a complete inhibition of *ydeP* promoter activity at a concentration of 1 mM. Only a small inhibitory effect was observed after the addition of ethanol, in which the indole stock was prepared, providing evidence that the complete inhibition of induction response is due to indole only.

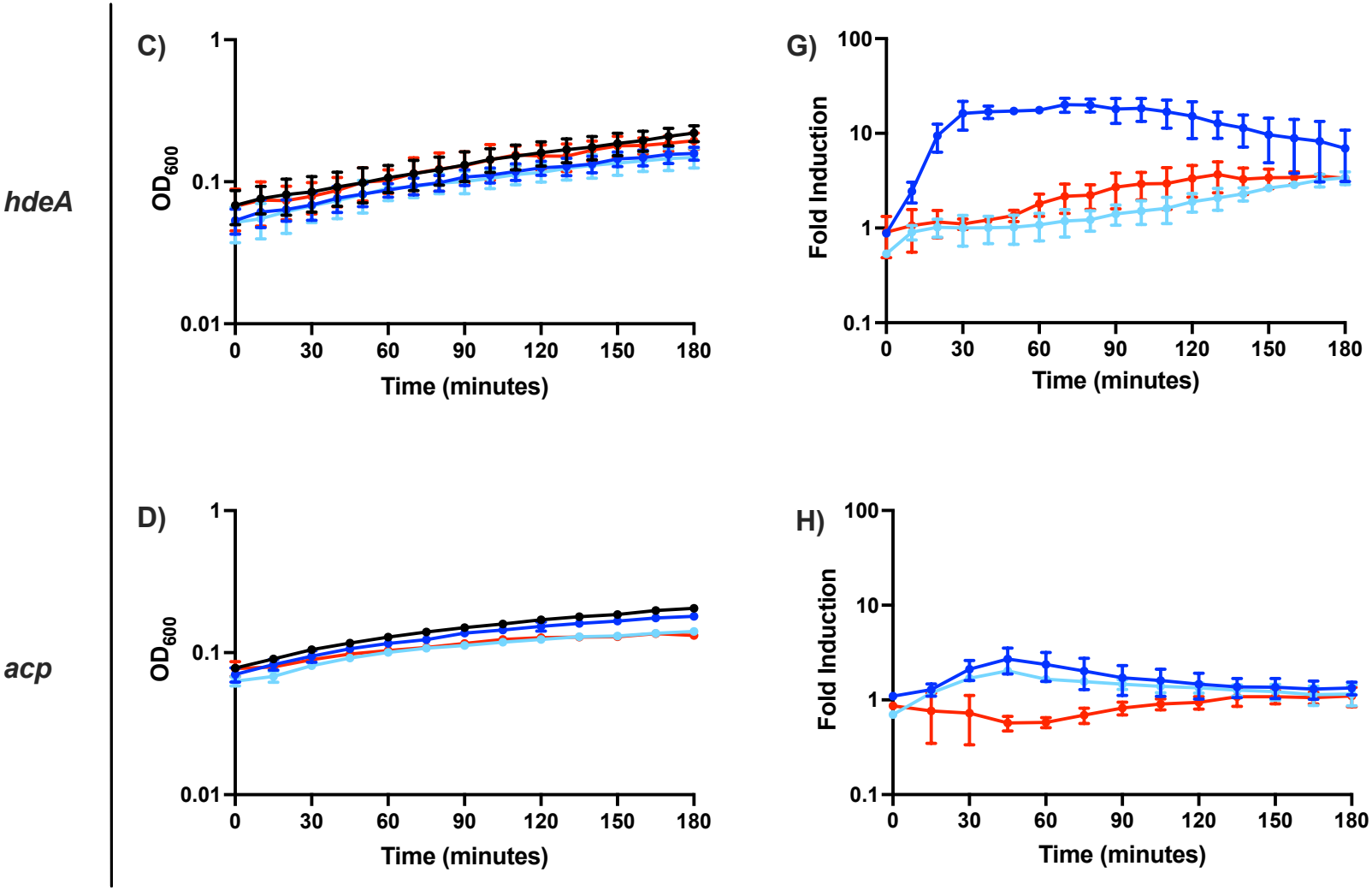
The inhibitory effect of indole could have been restricted to the *ydeP* promoter only, or it could be a more broad phenomenon with inhibition affecting all the genes of the AR2 system. To see which of these possibilities is more likely to be correct, luciferase assays were done with three AR2 promoters: *pydeP* (as a positive control), *pgadE*, and *phdeA*. All three promoters are regulated by EvgS sensor kinase; *pydeP* and *pgadE* are directly regulated by the EvgA response regulator whereas *phdeA* is regulated by GadE, the central regulator of

AR2 induction response. Additionally, a fourth promoter *pacp*, which is not induced significantly at pH 5.5 was used as a negative control. Each promoter sequence was fused with a luciferase reporter gene on a low-copy number plasmid as described in (Burton *et al.*, 2010).

MG1655 transformed with each plasmid: *pydeP-lux, pgadE-lux, phdeA-lux*, and *pacp-lux,* was grown to mid-exponential phase (OD_{600} of 0.2 - 0.4) before being exposed to pH 7 and pH 5.5 with and without 1 mM Indole. MG1655 *ΔevgS* strain transformed with each lux plasmid was also exposed to pH 5.5 as a negative control to ensure any effects seen required the activity of EvgS. Since the strain lacks EvgS, the sensor kinase that activates AR2 induction response, the AR2 promoters should fail to get induced resulting in no luminescence.

The growth of MG1655 and MG1655 *ΔevgS* containing each lux plasmid was monitored for 3 h at pH 7 and pH 5.5 with and without indole (Figure 3.1 A-D). Neither pH 5.5 nor the presence of indole affected the growth of cells during 3 h. The luminescence produced by cells under each condition was also measured for 3 h and fold-induction was calculated by dividing the luminescence normalised with optical density at pH 5.5 or pH 5.5 + indole by the normalised luminescence at pH 7. All three promoters *(pydeP, pgadE* and *phdeA)* were significantly induced at pH 5.5 and the induction response was inhibited on addition of indole (Figure 3.1 E-I). MG1655 *ΔevgS*, however, did not show any induction at pH 5.5 proving that the induction of AR2 promoters in exponentially growing *E. coli* cells requires EvgS. The control promoter *pacp* was also significantly induced at pH 5.5 but this response was not inhibited on addition of indole. Furthermore, the fold induction of *pacp* after 1 h was much smaller than the other three promoters. The fold induction for all promoters at 1 h is shown in Figure 3.1 I. Of all three AR2 promoters, *pydeP* showed the highest fold induction (100-fold higher than *pgadE* and *phdeA*) at 1 h.

Promoter	OD_{600}	Fold Induction
ydeP		
gadE		


hdeA
acp
C)
G)
D)
H)
OD600
Fold Induction
Time (minutes)
1
0.1
0.01
100
10
1
0.1
0
30
60
90
120
150
180

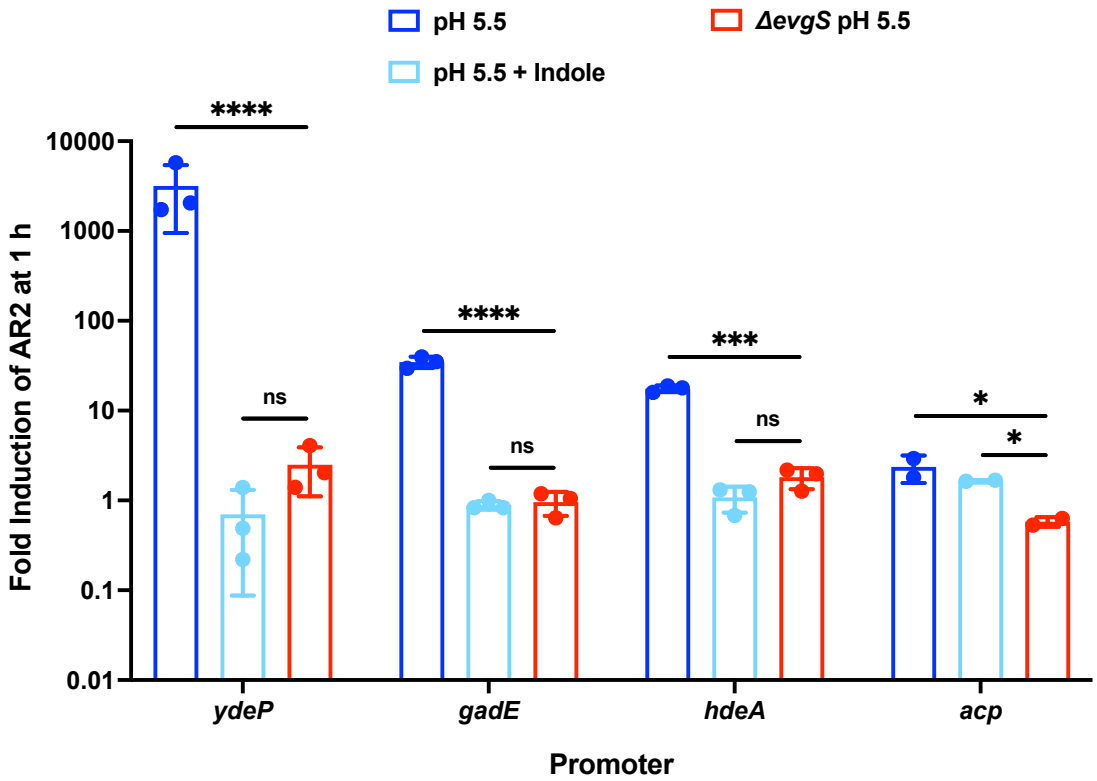


Figure 3.1: Growth profiles and fold induction of AR2 promoters measured using luciferase promoter-probe assay at pH 7 and 5.5 with and without indole.

(A-D) MG1655 transformed with pLux plasmid, containing fusions between AR2 promoter sequences and luciferase reporter gene, was cultured at pH 7 and pH 5.5 with and without 1 mM indole, and optical density was monitored for 3 h. pH 5.5 and indole did not affect the growth of cells relative to pH 7. (E-H) Fold induction at pH 5.5 with and without indole from AR2 promoters (*pydeP, pgadE, phdeA)* and control promoter (*pacp*) was monitored for 3 h. Fold induction was calculated by dividing the luminescence normalised with optical density at pH 5.5 or pH 5.5 + indole by the normalised luminescence at pH 7 for each promoter. (I) Comparison of fold induction at 1 h among AR2 and control promoters. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns= non-significant, *p < 0.05, ****p < 0.0001 (One-way ANOVA, followed by Tukey's multiple comparisons test).

3.2.2 Indole inhibits the AR2 system possibly through EvgS

It is clear from luciferase assays that indole has an inhibitory effect on AR2 promoters. However, it is not clear how indole inhibits the activity of these promoters. We hypothesised that indole inhibits the AR2 response through inhibiting EvgS activation. To test this hypothesis, we performed β-galactosidase assay with MG1655 Δ*evgS ydeP-lacZ* strain carrying a copy of constitutively active EvgS protein (EvgS-S600I), which is active at pH 7, on pBAD24 plasmid (Johnson *et al.*, 2014). MG1655 Δ*evgS ydeP-lacZ* pBADEvgS-S600I strain was grown to mid-exponential phase and the *lacZ* activity of *ydeP* promoter was measured at pH 7 in the absence and presence of indole. Results summarised in supplementary figure S4 of Boon *et al.* (2020) show that indole fails to inhibit the activity of *ydeP* promoter in the presence of constitutively active EvgS protein (Figure 3.2). If indole inhibited any of the AR2 components present downstream of EvgS, then indole would still be able to inhibit the activity of *ydeP* promoter in the presence of constitutively active EvgS protein. Therefore, these results are consistent with the hypothesis that indole inhibits the AR2 induction response through EvgS.

To further test the hypothesis that indole inhibits the AR2 response through EvgS, a competition assay was done between MG1655 and MG1655 *ΔevgS*. Since EvgS is necessary for the activation of AR2 response, as shown in previous section, then it would be expected that the MG1655 *ΔevgS* strain would be less fit than the corresponding wild-type strain when exposed to low pH following incubation at pH 5.5. However, if indole inhibits EvgS, then this fitness disadvantage of MG1655 *ΔevgS* strain would disappear when the incubation at pH 5.5 was done in the presence of indole. Both MG1655 and MG1655 *ΔevgS* produce red colonies on macConkey agar containing lactose. To distinguish MG1655 from MG1655 *ΔevgS,* MG1655 with a *lacZ* deletion (KH001), which produces white colonies on lactose-supplemented

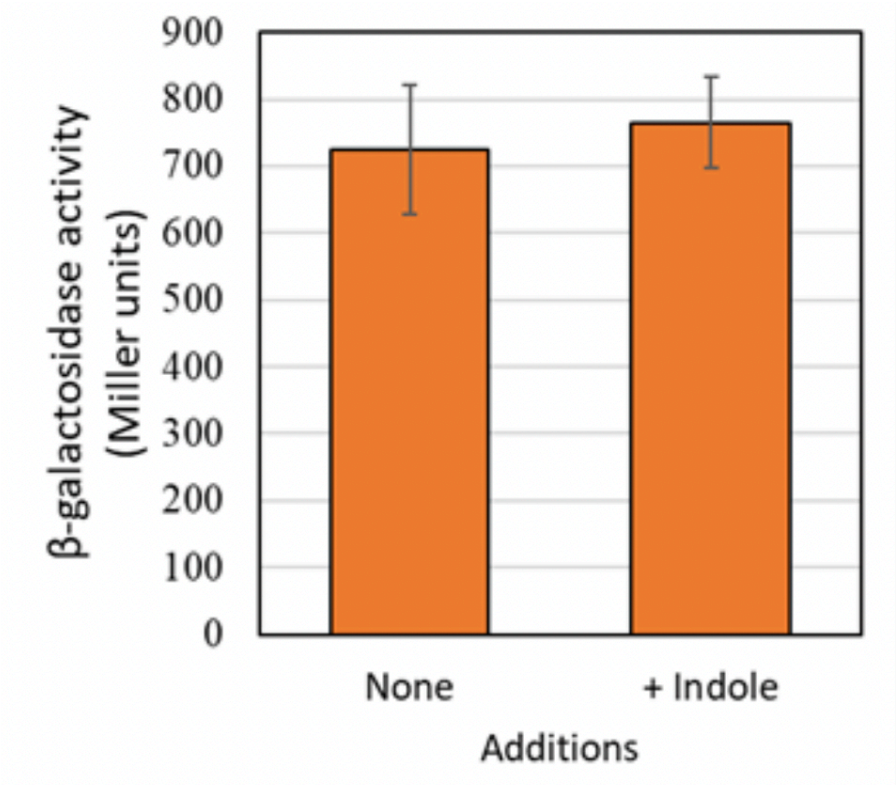


Figure 3.2: The effect of indole on the constitutive activity of EvgS-S600I.
The MG1655 Δ*evgS ydeP-lacZ* pBADEvgS-S600I strain was cultured to mid-exponential phase and the β-galactosidase activity was measured at pH 7 in the presence or absence of 1 mM indole. (Taken from Boon *et al.* (2020); Figure S4)

macConkey agar, was competed with MG1655 *ΔevgS* instead. To determine that KH001 has the same fitness as MG1655 and therefore can be used as its proxy, the competition assay was first performed between KH001 and MG1655.

KH001 and MG1655 were grown separately to mid-exponential phase (OD_{600} ~0.2), normalised to OD_{600} of 0.2, and mixed in equal proportions before being induced at pH 5.5 either in the presence of indole or ethanol. The induced cultures were further acid shocked at pH 2.5 for 15 minutes before being plated out on macConkey agar plates supplemented with lactose. MacConkey agar contains neutral red, a pH indicator which turns bright red or pink when pH falls below 6.8. Strains that can utilise lactose as a carbon source produce lactic acid which reduces the pH of agar resulting in red/pink colonies whereas white colonies are formed in the absence of lactose fermentation.

Selection rate (s) was used to determine the fitness of MG1655 and MG1655 Δ*evgS* against KH001 as described in Materials and methods, section 2.12.1. Selection rate of 0 shows no difference in the fitness of two strains, indicated by dashed red line in Figure 3.3. MG1655 showed no fitness advantage over KH001 under any tested conditions (pH 5.5 and pH 2.5) demonstrating that KH001 can be used as a proxy for MG1655 for comparison with MG1655 Δ*evgS*.

Competition between KH001 and MG1655 Δ*evgS* resulted in no difference in the fitness of two strains at pH 5.5 which shows that the induction step alone at pH 5.5 without the acid shock at pH 2.5 does not affect the fitness of the strains. At pH 2.5, however, MG1655 Δ*evgS* had a lower fitness than KH001 when the strains were induced in the presence of ethanol prior to acid shock at pH 2.5. This fitness disadvantage was not present when the strains were induced in the presence of indole before being acid shocked. These results

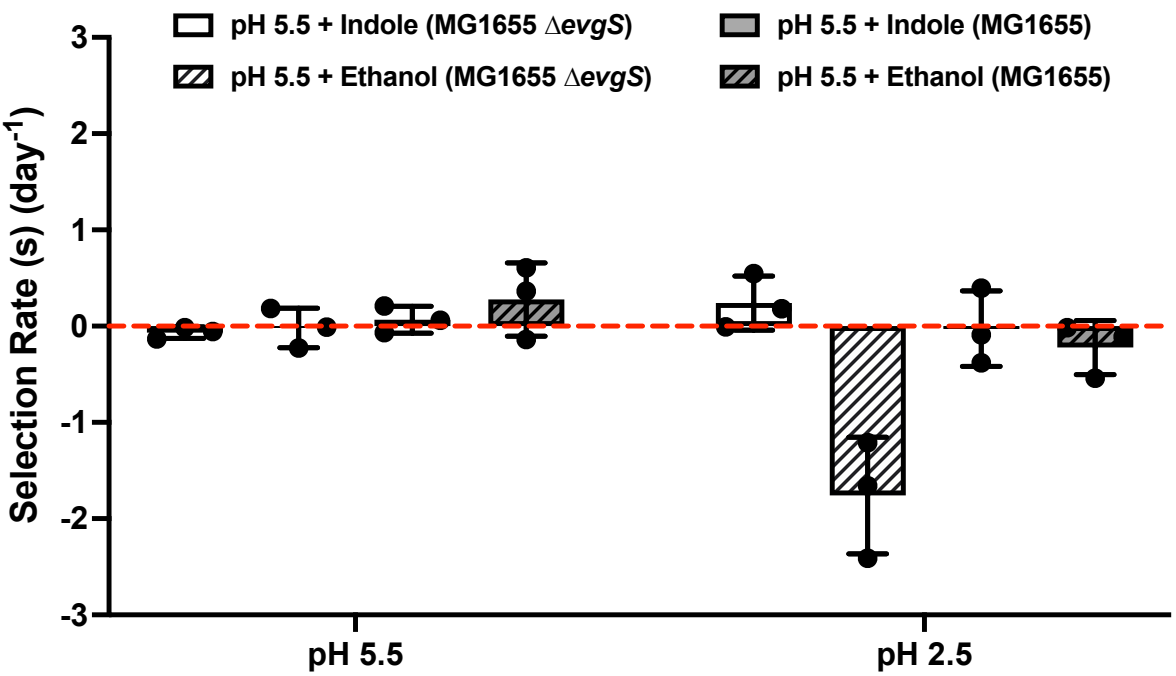


Figure 3.3: Competition of MG1655 and MG1655 Δ*evgS* against KH001 under AR2 induction conditions in the presence of 1 mM indole and the same volume of ethanol.
MG1655, MG1655 Δ*evgS* and KH001 were cultured separately to mid-exponential phase, normalised to OD_{600} of 0.2, and mixed in equal proportions before being induced at pH 5.5 with indole or ethanol, and acid shocked at pH 2.5. Selection rate of 0 shows no difference in the fitness of two strains (dashed red line). Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

suggest that KH001 has a fitness advantage over MG1655 Δ*evgS* when the two strains are induced in the presence of ethanol and acid shocked. This is because the AR2 system gets activated and provides higher acid resistance against pH 2.5 in KH001 than MG1655 Δ*evgS*. On the other hand, as MG1655 Δ*evgS* lacks EvgS, the AR2 system fails to get activated resulting in lower survival of this strain at pH 2.5 than KH001. In the presence of indole, however, both strains show almost equal fitness because indole blocks the induction of AR2 system in KH001.

This competition assay, therefore, shows that indole must inhibit a component that completely shuts down the AR2 response which results in equal fitness of wild-type and *evgS* knockout strains. Furthermore, β-galactosidase assay performed with constitutively active EvgS in Boon *et al.* (2020), which was described above, shows that there is a high probability that the component inhibited by indole is EvgS since indole is unable to block the induction of AR2 promoter, *ydeP,* in the presence of constitutively active EvgS protein. Other hypotheses about possible indole targets will be considered in the Discussion.

3.2.3 Modelling of a possible indole binding site within EvgS

During Covid-19, a Masters student, Andrew Fletcher worked on modelling the potential indole binding site within EvgS using I-TASSER protein structure prediction server. Several mutations within the cytoplasmic PAS domain of EvgS have been identified that switch the protein from inactive to active state even at neutral pH (Kato *et al.*, 2000; Johnson *et al.*, 2014). Therefore, PAS domain could be a potential binding site for indole. To date, the crystal structure of EvgS has not been solved. Therefore, I-TASSER was used to predict the crystal structure of EvgS PAS domain based on the structural templates which have been deposited in the Protein Data bank (PDB) (available at: https://www.rcsb.org). The PAS domain of the

HTR-like protein from the archaeal species *Haloarcula marismortui* was identified as a suitable template. This protein shares 20.8% sequence identity and 39.6% sequence similarity with the EvgS PAS domain. Furthermore, the crystal structure of this protein's PAS domain has been solved with ligand indole-3-carbaldehyde bound in it. The aspartate residue that coordinated the binding of indole-3-carbaldehyde in HTR-like protein's PAS domain is also conserved in EvgS PAS domain.

The models predicted by I-TASSER range from a confidence score of -5 to +2. Higher the confidence score, better the quality of the predicted model is likely to be. I-TASSER predicted five different models for EvgS PAS domain, with the confidence score ranging from -3.94 to +0.08. The model shown in Figure 3.4 had the highest confidence score of +0.08. As the crystal structure of the template protein (HTR-like protein) had indole-3-carbaldehyde bound to it, naturally this ligand was also predicted to bind EvgS with a confidence score of +0.45. Indole was modelled in the EvgS PAS domain by maintaining the predicted location of indole-3-carbaldehyde with respect to EvgS. According to this model, a hydrogen bond that forms between aspartate 625 (D625) residue of EvgS PAS domain and hydrogen present in -NH group of indole allows indole to bind EvgS (Figure 3.4).

To test the hypothesis that aspartate 625 (D625) is crucial for binding of indole to EvgS, site-directed mutagenesis was performed on pBADevgS plasmid containing a wild-type copy of EvgS protein to mutate aspartate 625 to alanine (D625A). Aspartate is a negatively charged amino acid and mutating it to alanine which is neutral should significantly alter the binding of indole if D625 is indeed crucial for binding of indole with EvgS. The D625A mutation in pBADevgS-D625A plasmid was constructed as described in Materials and Methods, section 2.9.2, and was confirmed through Sanger sequencing (data not shown). The MG1655 Δ*evgS ydeP-lacZ* strain was then transformed with the mutated plasmid. The transformed strain

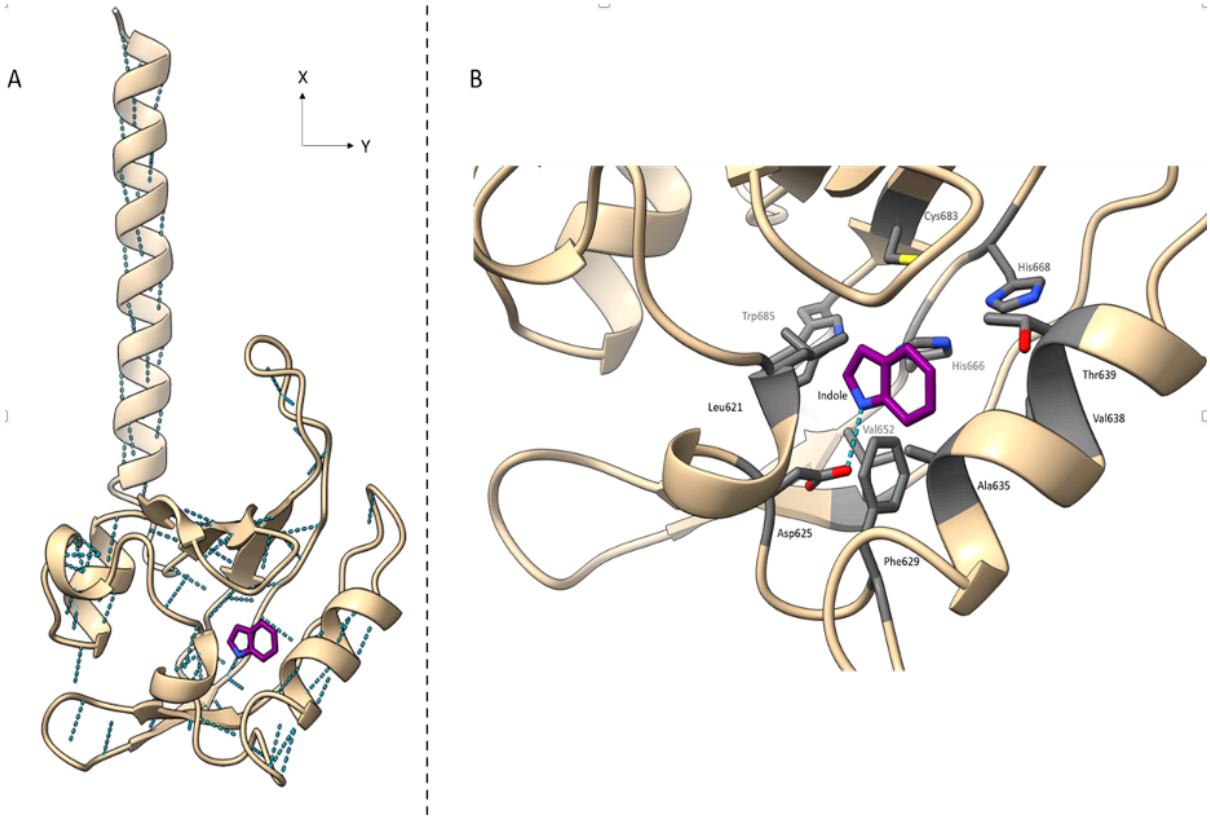


Figure 3.4: Model showing indole (purple) bound within the EvgS PAS domain.
(A) I-TASSER protein prediction server was used to predict the structure of EvgS PAS domain as well as binding of indole within this domain. The PDB entry 3BWL was used as a template (Yang and Zhang, 2015; Zhang, Freddolino and Zhang, 2017). Blue dashed lines indicate hydrogen bonds present within the protomer. (B) Residues which are 4 Å away from indole are shown in grey along with their sidechains. The blue dashed line shows the hydrogen bond that forms between aspartate 625 of EvgS PAS domain and hydrogen of -NH group in indole. (Taken from Fletcher (2021); Figure 9)

MG1655 Δ*evgS ydeP-lacZ* pBADevgS-D625A was cultured to mid-exponential phase and exposed to pH 7 and pH 5.5 in the presence of either indole or ethanol and the *lacZ* activity from *ydeP* promoter was measured. To ensure no activity was being produced from endogenous *lac* operon, *lacZ* activity in MG1655 was measured at pH 7 and pH 5.5, and was confirmed to be insignificant (Figure 3.5 A).

If the hypothesis that aspartate 625 (D625) is involved in the binding of indole by EvgS was true, then the MG1655 Δ*evgS ydeP-lacZ* pBADevgS-D625A strain would show a high *lacZ* activity at pH 5.5 which would not get repressed after the addition of indole to the induction medium. As expected, the recipient strain MG1655 Δ*evgS ydeP-lacZ* showed a low *lacZ* activity under all conditions as it lacks functional EvgS (Figure 3.5 B). The addition of pBAD24 vector plasmid to this strain did not alter the *lacZ* activity of the strain. However, the addition of pBADevgS plasmid led to a high *lacZ* activity at pH 5.5 which was repressed after addition of indole to the induction medium. The addition of pBADevgS-S600I plasmid resulted in a high *lacZ* activity at both pH 7 and pH 5.5, which did not get repressed with indole. The recipient strain transformed with pBADevgS-D625A plasmid showed a low *lacZ* activity at pH 7 but a high *lacZ* activity at pH 5.5 which was repressed with indole to almost the same level as pBADevgS-transformed strain. These results show that D625 is not a crucial residue for binding of indole with EvgS.

3.3 Discussion

EvgAS signal transduction system has been extensively studied in laboratory strain, *E. coli* K-12 where it has been shown to be crucial in the activation of AR2 acid resistance system among exponentially growing cells exposed to pH 5.5-5.7 (Masuda and Church, 2003; Itou, Eguchi and Utsumi, 2009; Burton *et al.*, 2010). The results from luciferase assays in this study

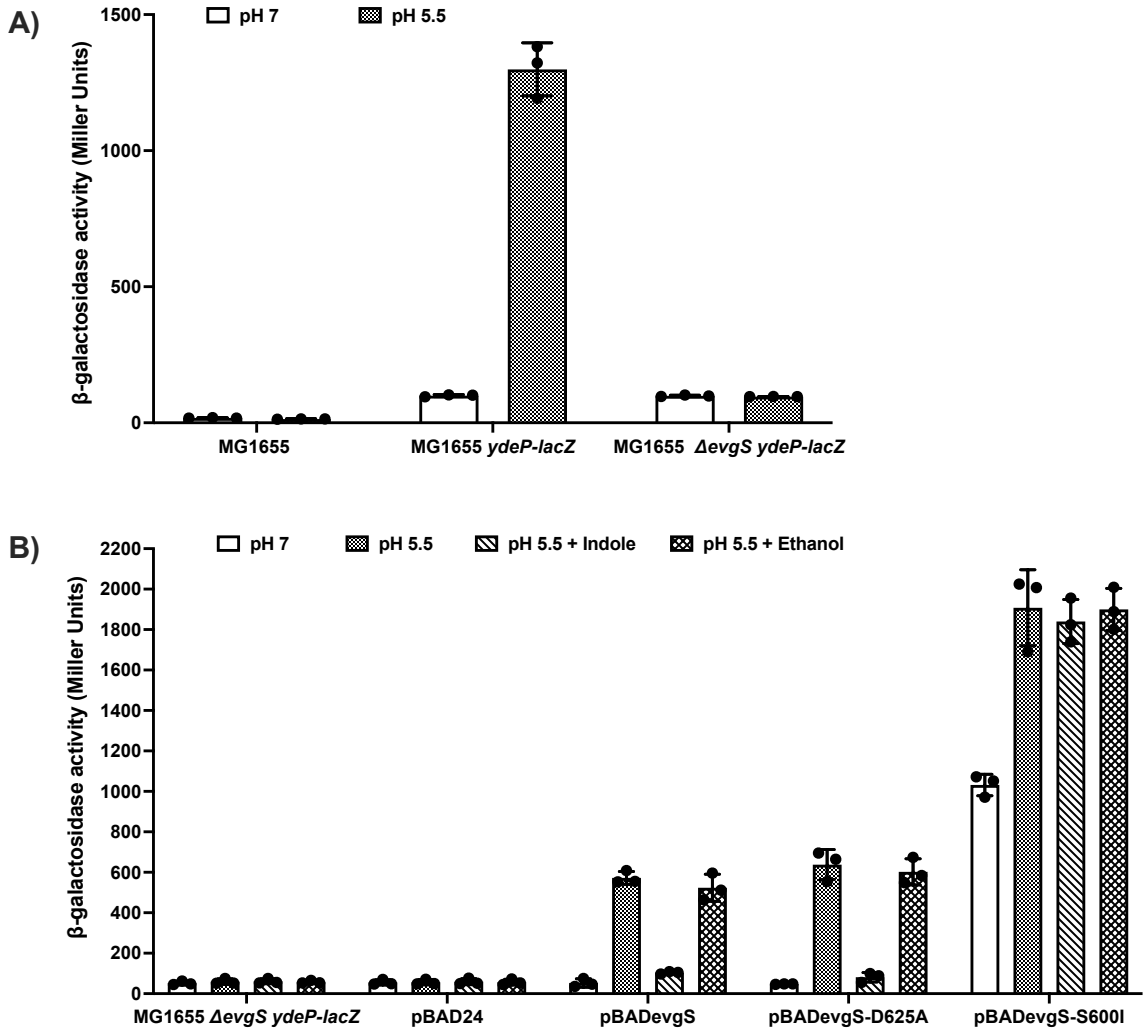


Figure 3.5: Measurement of β-galactosidase activity from *ydeP* promoter to test the potential binding site of indole within EvgS PAS domain.

Aspartate 625 (D625) was predicted to be a potential residue crucial for binding of indole within EvgS PAS domain by I-TASSER protein prediction server. Site-directed mutagenesis was performed on pBADevgS plasmid to mutate D625 to alanine (D625A). The mutated plasmid pBADevgS-D625 was transformed into MG1655 Δ*evgS ydeP-lacZ* strain and the *lacZ* activity was measured from *ydeP* promoter at pH 7 and pH 5.5 + indole/ethanol. (A) Endogenous *lac* operon in MG1655 produced low *lacZ* activity at both pH 7 and pH 5.5. (B) MG1655 Δ*evgS ydeP-lacZ* pBADevgS-D625A strain showed a high *lacZ* activity at pH 5.5 consistent with the induction of AR2 response and activation of *ydeP* promoter. The addition of indole repressed the high *lacZ* activity at pH 5.5 to almost the same level as pBADevgS-transformed strain showing that D625 is not involved in binding of indole with EvgS.

are consistent with this known role of EvgS because the activation of AR2 promoters could not be achieved in K-12 strain MG1655 Δ*evgS* (Figure 3.1). When induction could be achieved in wild-type strain, the addition of 1 mM indole to the induction medium inhibited the induction response from various AR2 promoters. Two of these promoters, *pydeP* and *pgadE* are under the direct regulation of EvgA response regulator, whereas the third promoter *phdeA* is regulated by GadE, the central regulator of AR2 response (Masuda and Church, 2003; Ma, Masuda and Foster, 2004; Itou, Eguchi and Utsumi, 2009; Seo *et al.*, 2015). Since indole had a complete inhibitory effect on all three promoters, it is likely that indole blocks the entire AR2 induction response rather than a part of it.

The competition assay between wild-type and *evgS* deletion mutant under AR2 induction conditions in the presence of indole and ethanol followed by extreme acid shock at pH 2.5 showed that *evgS* deletion mutant has a reduced fitness compared to its wild-type counterpart in the presence of ethanol but not indole, where both strains showed almost equal fitness (Figure 3.3). This result further indicates that indole is likely to inhibit the AR2 response completely. For indole to inhibit the AR2 response completely, it is likely to act on one or more of these targets: the sensor kinase EvgS, the response regulator EvgA, the central regulator of AR2 GadE, or the response regulator RcsB. RcsB forms heterodimer with GadE, and possibly with EvgA, to regulate the expression of AR2 genes and it has been shown that the induction of AR2 promoters is not possible in an *rcsB* deletion mutant (Castanié-Cornet *et al.*, 2010; Burton *et al.*, 2010; Johnson *et al.*, 2011).

β-galactosidase assay done with MG1655 *ΔevgS ydeP-lacZ* pBADevgS-S600I strain showed that indole does not inhibit the *ydeP* promoter activity in the presence of constitutively active EvgS protein (Figures 3.2 and 3.5 B) (Boon *et al.*, 2020). This result indicates that indole is likely to inhibit the AR2 response via EvgS, possibly through binding

and preventing EvgS from either sensing the signal or transducing it. Other possible explanations include that indole inhibits an unknown activator present upstream of EvgS which activates EvgS, or affects some property of the cell which is sensed by EvgS during induction at pH 5.5.

An attempt was made at mapping the indole binding site within EvgS using I-TASSER protein prediction server (Figure 3.4). The predicted residue crucial for indole binding, aspartate 625, present in the EvgS PAS domain, did not show promising results in β-galactosidase assay (Figures 3.4 and 3.5 B). Due to its role in low pH perception, identification of mutations that lead to constitutive activation of EvgS, and mutations (C671A and C683A) that can activate EvgS under anaerobic conditions, PAS domain is an important part of EvgS and therefore remains a potential candidate for indole binding (Kato *et al.*, 2000; Eguchi and Utsumi, 2014; Johnson *et al.*, 2014; Inada *et al.*, 2021). Other potential targets are the VFT domains which sense the alkali metal ions necessary for EvgS activation (Inada *et al.*, 2021).

Further biochemical experiments are required to check whether indole inhibits the AR2 response through binding with EvgS. One of these experiments include a Phos-tag assay which can detect the kinase activity of EvgS in the presence and absence of indole, through measuring levels of phosphorylated EvgA (Dupré *et al.*, 2015). Another experiment includes isothermal titration calorimetry (ITC) which can check for the direct interaction between indole and EvgS (Dupré *et al.*, 2015).

Chapter 4

Comparison of acid resistance and the impact of indole on induction of AR2 response among *E. coli* strains

4.1 Introduction

E. coli, being an enteric bacterium, is exposed to extremely low pH (1.5 - 3.5) while passing through the stomach to enter the intestines. The activation of AR2 system could be crucial for survival of *E. coli* in stomach since it is the most effective AR system against extremely low pH of 2.5 and below. The AR2 activation has been observed under two different conditions. First, when exponentially growing cells are exposed to mildly acidic conditions of pH between 5.5 and 5.7 (Ma, Masuda and Foster, 2004; Burton *et al.*, 2010; Eguchi and Utsumi, 2014). Second, when stationary phase cells are exposed to low pH (2 - 2.5) in the presence of glutamate (De Biase *et al.*, 1999).

Distinct regulatory networks control the transcription of AR2 genes in exponential and stationary phases. However, stationary phase sigma factor RpoS, plays a role in both of these growth phases in the activation of AR2 response. RpoS enhances the survival of cells under suboptimal growth conditions. For example, it regulates the expression of genes required to survive in nutrient deprived and stressful conditions such as oxidative stress, osmotic stress and acid stress (reviewed in Battesti, Majdalani and Gottesman, 2011; Schellhorn, 2020). In exponential phase, RpoS activates the expression of AR2 genes such as *gad* genes and genes encoding the periplasmic chaperones HdeA and HdeB, through regulating GadE levels (Burton *et al.*, 2010). In stationary phase, it activates the AR2 response by relieving the repression of *gad* genes (*gadA, gadB*, *gadC* and *gadX*) by global transcriptional regulator H-NS (Tramonti *et al.*, 2002; Giangrossi *et al.*, 2005).

The regulation of AR2 system has been extensively studied but mainly in laboratory K-12 strain MG1655. Although some work has been done on pathogenic and commensal strains, which was described in section 1.5 of Chapter 1, these have not been studied to the same extent as MG1655.

The presence of short chain fatty acids makes the ascending part of the colon mildly acidic (pH 5 - 6) which is appropriate for the activation of AR2 system. Indole is a bacterial signalling molecule which regulates gene expression under antibiotic stress, pH stress, heat stress, biofilm formation and persister cell formation (reviewed in Zarkan *et al.*, 2020). The evidence for presence of indole in human gut is that it is found in varying amounts (0.30 to 6.64 mM) in the faeces of healthy adults (Darkoh *et al.*, 2015). Furthermore, tryptophan which is an essential component of human diet can be metabolised into indole by gut microbiota using tryptophanase (*tnaA*) enzyme (Lee and Lee, 2010; Li and Young, 2013). Previously, we have shown in our lab that indole inhibits the induction of AR2 response in exponentially growing laboratory K-12 strains MG1655 and BW25113 (Boon *et al.*, 2020).

In this chapter, the induction of AR2 response under mildly acidic conditions of pH 5.5, was tested in three different growth stages: exponential, early stationary and stationary phase. This was done in laboratory K-12 strains MG1655 and BW25113, a pathogenic strain O157:H7 Sakai and a commensal strain Nissle 1917. Since bacterial cells that pass through the gastrointestinal system are in different growth stages, it is important to establish whether AR2 system can be activated in different growth phases, which will help us to understand how well the enteric pathogens and probiotic strains are able to survive the stomach acid.

Additionally, the survival of all four strains was compared with and without prior induction at pH 5.5, in the presence or absence of indole. This was to check whether the inhibitory effect of indole on AR2 response, previously shown in laboratory strains MG1655 and BW25113, is conserved in real world strains.

Finally, the effect of RpoS on innate acid resistance of *E. coli*, without AR2 induction was tested under different growth stages.

4.2 Results

4.2.1 Optimisation of conditions for AR2 induction and inhibition with indole

To study the impact of indole on the induction of AR2 system, it was important to ensure that the medium used to culture and induce the cells was indole free. *E. coli* can use tryptophanase, the product of *tnaA* gene, to metabolise tryptophan present in the growth medium to indole (Lee and Lee, 2010; Li and Young, 2013). Since the induction of AR2 system occurs in minimal medium and the medium should be free of tryptophan, M9 minimal medium supplemented with KCl, glucose and casamino acids (referred to as M9Ksupp medium in this study) was used to culture, induce and acid shock the cells. M9Ksupp medium lacks tryptophan because it is destroyed by acid treatment during the production of casamino acids. KCl is not included in typical M9 recipe but was included because the addition of alkali metal ions has been shown to better induce the AR2 system (Eguchi and Utsumi, 2014).

Before testing the AR2 induction and its effect on enhanced survival, the innate acid resistance of cells was determined across different growth stages to discern how susceptible the cells are to extreme acid stress and whether induction of AR2 is likely to play a significant role in improving their survival. To do this, MG1655 was cultured in M9Ksupp medium and the growth was monitored for 8 hours (Figure 4.1 A). Three growth stages, exponential (OD_{600} - 0.2), early stationary (OD_{600} - 0.9) and stationary (OD_{600} - 2) phase were selected for assessing the survival of cells after extreme acid stress at pH 2.5 (marked with red arrows in Figure 4.1 A). At each growth phase, the cells were acid shocked for 2 h at pH 2.5 and then plated out using two methods: spreading and the tilt plate method. The tilt plate is a method where a drop of culture is placed on top of a square plate and the plate is tilted vertically to allow the culture drop to run all the way to the bottom of the plate. This method was used as multiple dilutions can be plated on the same plate which saves time as less plates need to be prepared.

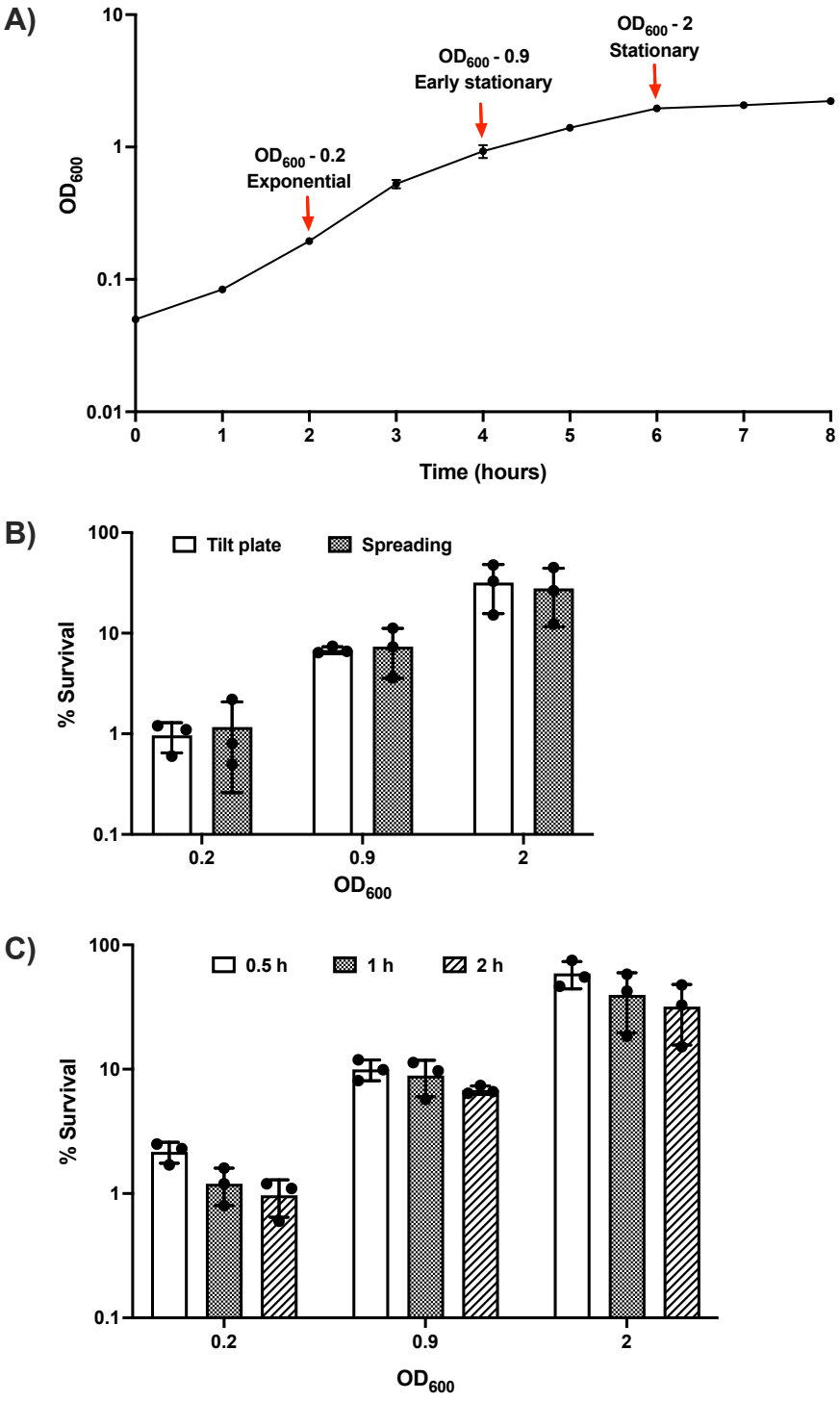


Figure 4.1: Survival of MG1655 at pH 2.5 in exponential, early stationary and stationary phases using different incubations times and plating methods.

(A) MG1655 was cultured in M9Ksupp pH 7 medium and optical density (OD_{600}) was measured every hour for 8 hours. Red arrows indicate the optical densities and growth stages at which the survival of cells was measured after being acid shocked at pH 2.5. The survival was calculated as described in Material and Methods, section 2.5. (B) Survival of MG1655 at OD_{600} of 0.2, 0.9 and 2 after 2 h of acid shock determined using tilt plate and spreading methods. (C) Survival of MG1655 at OD_{600} of 0.2, 0.9 and 2 after 0.5, 1 and 2 h of acid shock. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

However, while using tilt plate method, a small volume (8 µL) of culture is plated which can reduce the accuracy of results as compared to when larger volumes in a method like spreading (100 µL) are plated out. As shown in Figure 4.1 B, the two methods gave similar levels of survival. Therefore, tilt plate method was used in future experiments as it saves time and reagents.

To optimise the acid shock duration which would be used in the AR2 induction assays, the cells in each growth phase were acid stressed at pH 2.5 for 0.5, 1 and 2 h before being plated out using tilt plate method (Figure 4.1 C). The survival of cells was calculated, as described in Materials and methods section 2.5, by dividing the number of colonies at pH 2.5 after 0.5/1/2 h by the number of colonies at pH 7 after 0 h, and expressed as a percentage. The survival of MG1655 increased with increase in optical density suggesting that its innate acid resistance increases from exponential to stationary phase. Furthermore, at each growth stage, the survival of MG1655 decreased as the incubation time increased from 0.5 h to 2 h. At OD_{600} of 0.2, 1% of cells survived after being acid stressed for 2 h which is comparable to the results of a study done previously in our lab at the same OD_{600} and using an incubation time of 2 h (Johnson *et al.*, 2014), demonstrating the reproducibility of results. Therefore, the acid stress duration of 2 h at pH 2.5 was used for all the experiments in this chapter. After 2 h of exposure at pH 2.5, 6.8 % cells survived at OD_{600} of 0.9 and 31.9% at OD_{600} of 2.

4.2.2 Comparison of survival at pH 2.5 in MG1655 versus MG1655 *ΔrpoS*

4.2.2.1 Growth and acid shock in M9Ksupp

The innate acid resistance of MG1655 increases as cells transition from exponential to stationary phase as seen in Figures 4.1 B and C. The levels of stationary phase sigma factor, RpoS are low in exponential phase and increase as cells enter the stationary phase. One of

the reasons behind increased innate acid resistance of MG1655 from exponential to stationary phase, could be due to increased production of RpoS over time. To test this hypothesis, MG1655 and MG1655 Δ*rpoS* strains were cultured in M9Ksupp pH 7 medium to OD_{600} of 0.2, 0.6, and 1, acid shocked at pH 2.5 and their survival was measured (Figure 4.2). If our hypothesis was true and RpoS is the sole reason behind increased acid resistance of MG1655 from exponential to stationary phase, then the survival of MG1655 would be expected to increase with increase in OD_{600}. However, the survival of MG1655 Δ*rpoS* would remain the same at all growth stages. As shown in Figure 4.2, the survival of MG1655 increased from 1.5% at OD_{600} of 0.2 to 22.3% at OD_{600} of 1. However, the survival of MG1655 Δ*rpoS* also increased from 0.04% at OD_{600} of 0.2 to 0.7% at OD_{600} of 1, indicating that there is at least one other factor which is responsible behind increasing the innate acid resistance of MG1655 as growth proceeds towards stationary phase. As a general trend, the survival of MG1655 Δ*rpoS* was lower than MG1655 by more than a log-fold at all three growth stages suggesting that RpoS plays a role in protecting the cells against extreme acid stress in exponential and stationary phase.

4.2.2.2 Growth and acid shock in LB

The acid resistance of *E. coli* is known to be dependent on growth media (Lin *et al.*, 1996), so a comparison was made between innate acid resistance of MG1655 in M9Ksupp and LB. The effect of RpoS in the two different media was also compared. MG1655 and MG1655 Δ*rpoS* strains were cultured in LB broth to OD_{600} of 0.2, 0.6, and 1, acid shocked at pH 2.5 and their survival was measured (Figure 4.3). The unexpectedly high level of survival seen at OD_{600} of 0.2 as compared to OD_{600} of 0.6 and 1 could be due to residual levels of RpoS

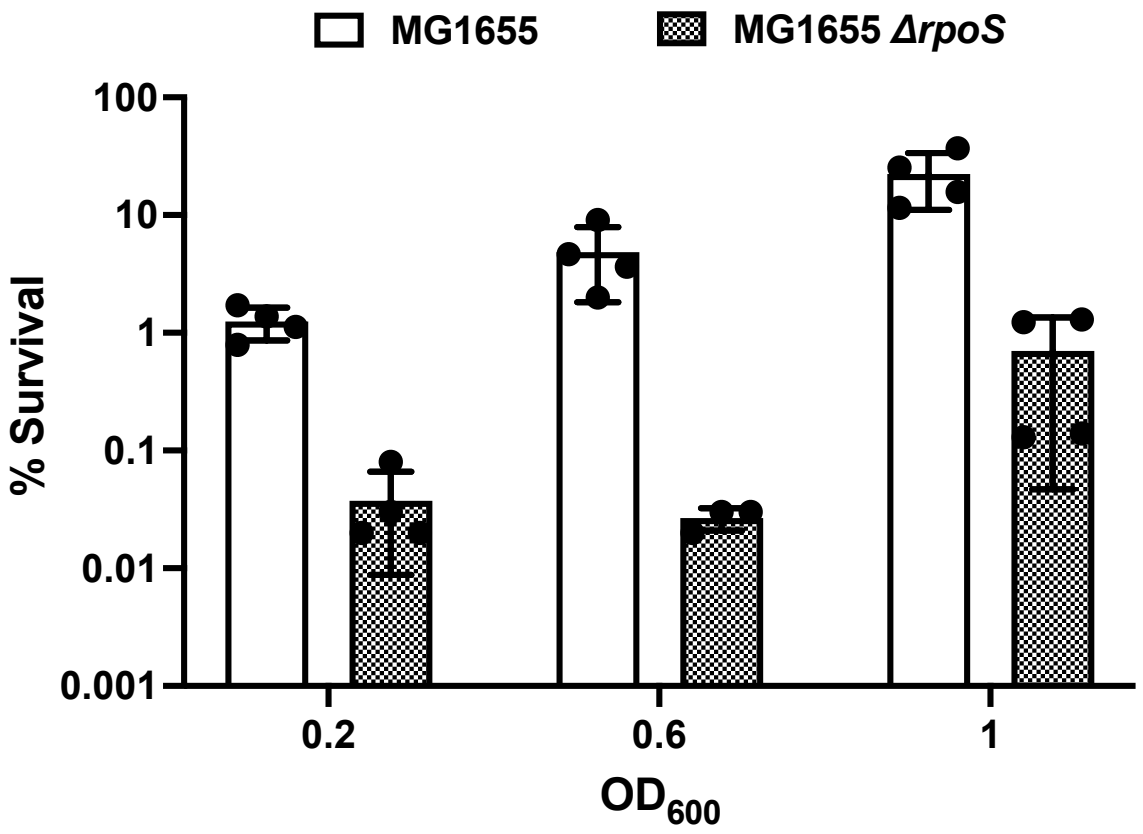


Figure 4.2: Comparison of innate acid resistance of MG1655 with MG1655 Δ*rpoS* cultured and acid stressed at pH 2.5 in M9Ksupp.

MG1655 and MG1655 Δ*rpoS* were cultured in M9Ksupp medium to OD_{600} of 0.2, 0.6 and 1, acid shocked at pH 2.5 for 2 h and their survival was measured. The survival was calculated as described in Material and Methods, section 2.5. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

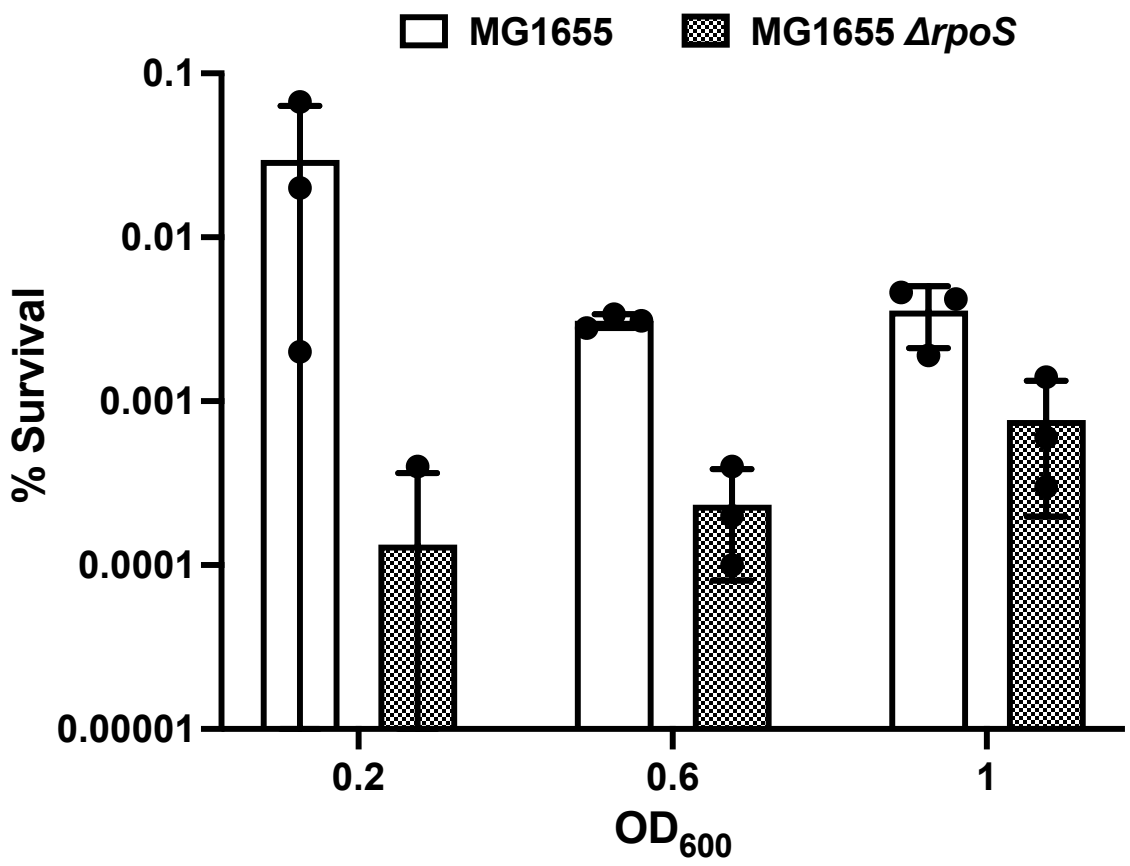


Figure 4.3: Comparison of innate acid resistance of MG1655 with MG1655 *ΔrpoS* cultured and acid stressed at pH 2.5 in LB.

MG1655 and MG1655 *ΔrpoS* were cultured in LB to OD_{600} of 0.2, 0.6 and 1, acid shocked at pH 2.5 for 2 h and their survival was measured. The survival was calculated as described in Material and Methods, section 2.5. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. Only one data point for MG1655 *ΔrpoS* has been shown at OD_{600} of 0.2 because no colonies were present on the plates for the other two biological replicates.

still present from stationary phase. Consistent with this, if cultures were grown from an initial OD_{600} of 0.05, rather than 0.01 as used here, an even higher level of survival was seen (15.2%). The survival of MG1655 Δ*rpoS* increased with increase in optical density, indicating that there is at least one other factor which is responsible behind increasing the innate acid resistance of MG1655 as growth proceeds towards stationary phase. However, the survival of MG1655 Δ*rpoS* was lower than MG1655 at all growth stages, which is consistent with the hypothesis that RpoS plays a role in extreme acid resistance in LB.

The survival of both strains, MG1655 and MG1655 *ΔrpoS,* in M9Ksupp and LB broth was compared (Figure 4.4). Both strains had a higher survival in M9Ksupp than LB broth. The decision was made to use M9Ksupp for all future experiments because it is a defined medium which increases the reproducibility of results. Additionally, the AR2 response in exponentially growing *E. coli* exposed to mild acid is not induced in LB for an unknown reason.

4.2.3 Induction of acid resistance and effect of indole in exponential phase cells

After determining the level of innate acid resistance of MG1655, the effect of AR2 induction with and without indole on the survival of MG1655 was tested at OD_{600} of 0.2, 0.9, and 2. MG1655 was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.2 and acid stressed at pH 2.5 for 2 h. Three additional cultures were pre-incubated at pH 5.5 for 30 minutes in the absence and presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone, before being acid stressed at pH 2.5. The uninduced cells exposed to pH 2.5 had a survival of 3% whereas the cells induced at pH 5.5 before acid shock had a survival of 126.7%, a 42-fold increase as compared to uninduced cells (Figure 4.5). The survival of over 100% can result from the growth that occurs during the induction step at pH

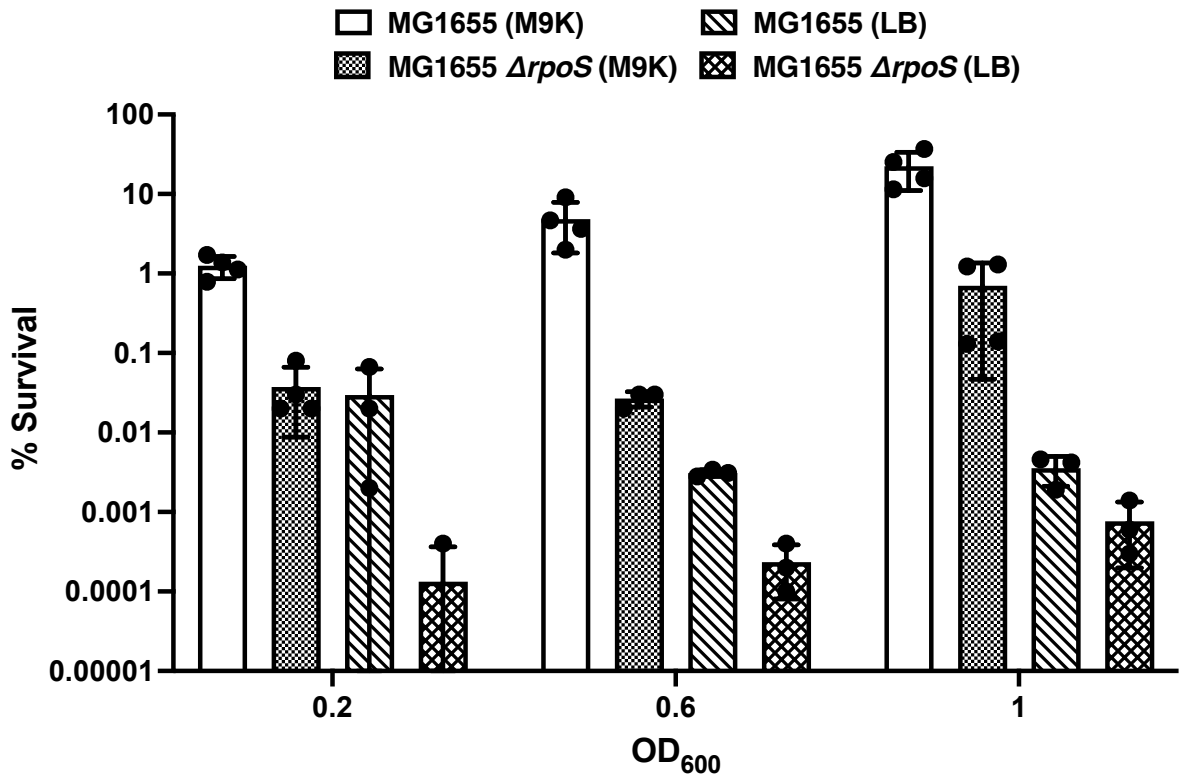


Figure 4.4: Comparison of innate acid resistance of MG1655 and MG1655 *ΔrpoS* cultured and acid stressed at pH 2.5 in M9Ksupp and LB.

MG1655 and MG1655 *ΔrpoS* were cultured either in M9Ksupp medium or LB broth to OD_{600} of 0.2, 0.6 and 1, acid shocked in the same medium at pH 2.5 for 2 h and their survival was measured. The survival was calculated as described in Material and Methods, section 2.5. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

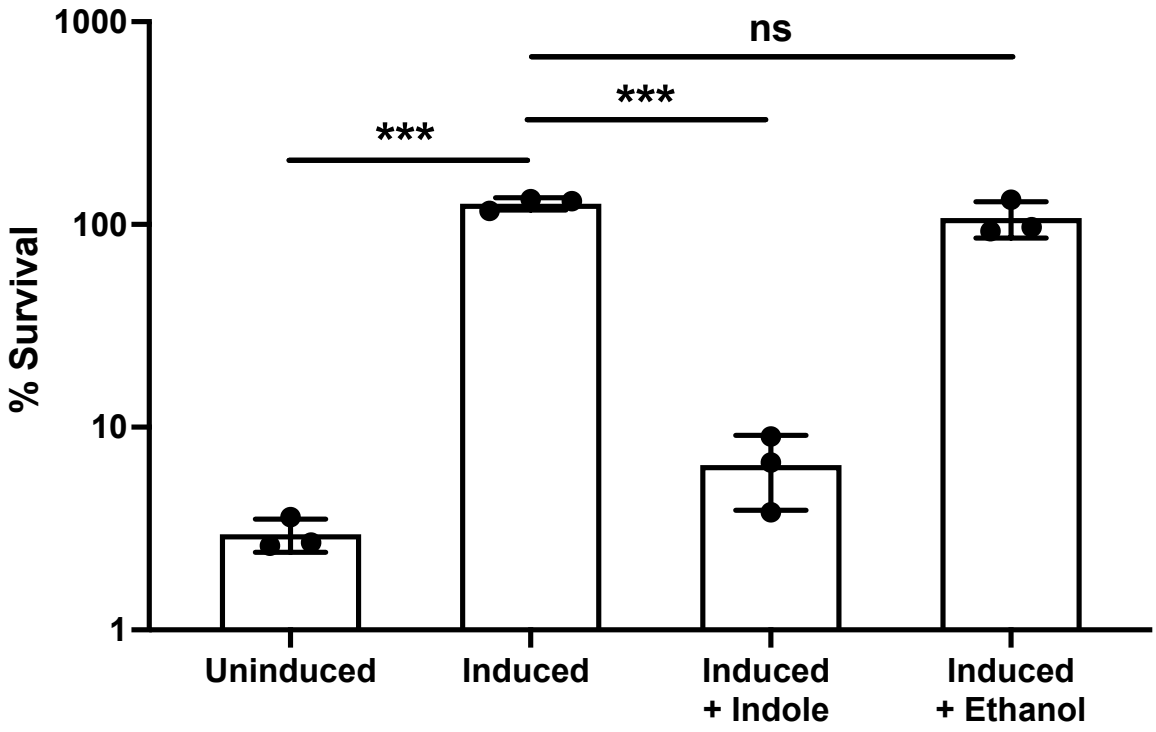


Figure 4.5: Induction of acid resistance and inhibition with indole in exponentially growing MG1655.

MG1655 was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.2 and acid stressed at pH 2.5 for 2 h with and without prior induction at pH 5.5 for 30 minutes in the absence or presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone. The survival was calculated as described in Material and Methods, section 2.5. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns= non-significant, ***p < 0.001 (One-way ANOVA, followed by Tukey's multiple comparisons test).

5.5 because the final survival is calculated by dividing the number of colonies at pH 2.5 after 2 h by the number of colonies at pH 7 obtained before the pH 5.5 step.

The addition of indole to the induction medium significantly reduced the survival of cells to 6.5%. As a control to check that the solvent was not causing this effect, MG1655 was induced in the presence of ethanol alone before being acid stressed. The addition of ethanol had a non-significant effect on the survival (107.5%) of MG1655, suggesting that the inhibitory effect observed previously is due to indole. This result, as shown previously by several studies, suggests that the acid resistance can be induced in exponentially growing *E. coli* enhancing the survival of cells exposed to extreme acid stress (Burton *et al.*, 2010; Johnson *et al.*, 2011; Eguchi and Utsumi, 2014; Roggiani, Yadavalli and Goulian, 2017; Boon *et al.*, 2020). Furthermore, the induced acid resistance can be blocked with indole such that the survival of cells decreases close to the same level as uninduced cells exposed to pH 2.5. The results of this plating assay are consistent with the results of promoter-probe assays in Chapter 3 (sections 3.2.1 and 3.2.3) where 1mM indole inhibited the induction of various AR2 promoters.

4.2.4 Induction of acid resistance and effect of indole in early stationary phase cells

4.2.4.1 Indole stock prepared in 100% ethanol

The results presented in section 4.2.1 show that MG1655 in a later phase of growth (OD_{600} 0.9 vs OD_{600} 0.2) had a higher survival following a 2 h exposure to pH 2.5. In this section, the effect of indole on increased innate acid resistance of MG1655 in early stationary phase was tested. Furthermore, whether the acid resistance can be induced and subsequently blocked with indole in this growth phase was determined. To test this, MG1655 was cultured to OD_{600} of 0.9 and acid shocked at pH 2.5 for 2 h. Additionally, two more cultures were pre-

incubated at pH 7 with either 1 mM indole or the same volume of ethanol for 30 minutes before being acid stressed. Three additional cultures were pre-incubated at pH 5.5 with and without 1 mM indole dissolved in ethanol, or ethanol alone, prior to acid stress. Indole inhibited the innate acid resistance of MG1655 by reducing the survival of cells from 29.9% to 15.9% (Figure 4.6 A). However, unexpectedly, the presence of ethanol alone reduced the survival of the cells (to 11.1%) even more than indole. Therefore, there is no evidence for any effect of indole on innate acid resistance of cells at OD_{600} of 0.9.

The pre-incubation at pH 5.5 followed by acid shock led to a non-significant increase in the survival of MG1655 from 29.9% (survival of uninduced cells) to 54.4%. Indole had a small but non-significant inhibitory effect on this induction and reduced the survival of cells to 42.7%. Ethanol, however, increased the survival of cells to 64.2%. Therefore, the inhibitory effect of indole cannot be determined with confidence due to the opposite effect of indole and ethanol on the induction response.

4.2.4.2 Indole stock prepared in dimethyl sulfoxide (DMSO)

Since ethanol had a negative effect on the survival of uninduced cells at pH 2.5 at OD_{600} of 0.9, an alternative solvent (DMSO) was tested to see if it had a similar effect. The final indole concentration was kept the same as above which is 1 mM. The volume of indole stock, prepared at a concentration of 100 mM, added to 1mL culture was 10 µL. The same volume of DMSO was added in the DMSO control. Indole showed an inhibitory effect on the innate acid resistance of MG1655 by reducing the survival of cells from 21.5% to 5.5 % (Figure 4.6 B). However, DMSO also reduced the survival of cells (6.7%) to almost the same level as indole. To reduce the volume of DMSO added to cells, a more concentrated stock of indole was prepared in DMSO (200mM), and 5 µL of this stock, or DMSO alone, were added to the cultures at pH 7

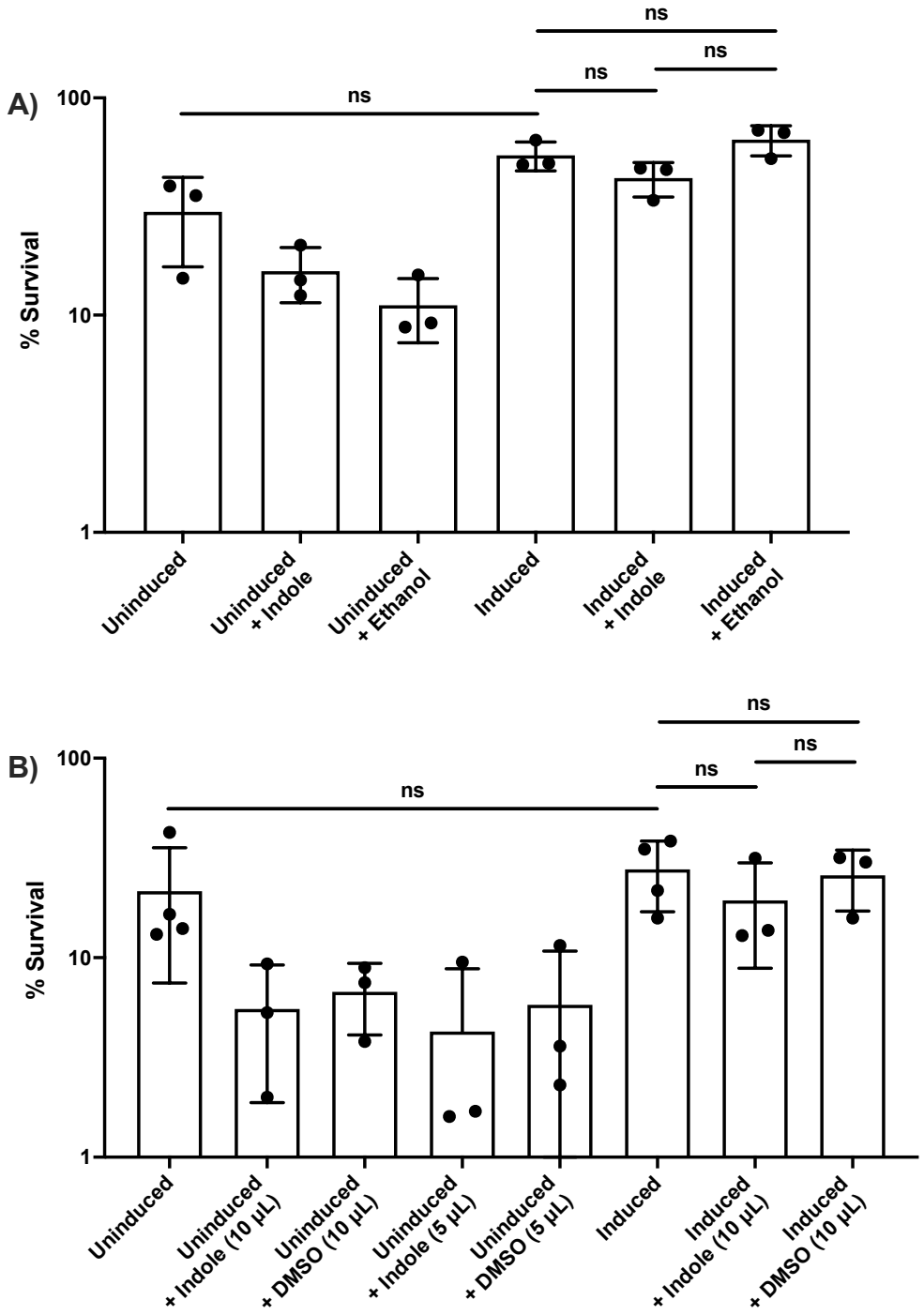


Figure 4.6: Induction of acid resistance and inhibition with indole in MG1655 grown to early stationary phase.

MG1655 was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.9 and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes. The effect of indole (dissolved in ethanol or DMSO) at a final concentration of 1 mM, or the same volume of ethanol/DMSO alone on both uninduced (pH 7) and induced cells (pH 5.5) was determined. The survival was calculated as described in Material and Methods, section 2.5. (A) The indole stock was prepared at 100 mM in Ethanol. (B) The indole stock was prepared either at 100 mM or 200 mM in DMSO, from which 10 µL and 5 µL was added to 1 mL cultures respectively. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns= non-significant (One-way ANOVA, followed by Tukey's multiple comparisons test).

before acid shock. DMSO, however, still had a negative effect on the survival (5.8%) of cells. Attempts were made to find an alternative solvent for preparing the indole stock, which included distilled water, phosphate buffer saline (PBS) and M9Ksupp pH 7 medium, none of which are likely to affect the survival of cells. But 100 mM of indole was insoluble in all of them.

The pre-incubation at pH 5.5 followed by acid shock led to a non-significant increase in the survival of MG1655 as observed before in Figure 4.6 A. Indole dissolved in DMSO had a slight but non-significant inhibitory effect on the induction response at pH 5.5. DMSO, on the other hand, also caused a small but non-significant reduction in the survival of cells.

Taken together, the results in Figure 4.6 suggest that indole does not inhibit the innate acid resistance of MG1655 in early stationary phase. Furthermore, the acid resistance cannot be induced significantly in MG1655 grown to early stationary phase and exposed to mildly acidic pH of 5.5 for 30 minutes.

4.2.5 Induction of acid resistance and effect of indole in stationary phase cells

The induction and inhibition of acid resistance at pH 5.5 along with the inhibition of innate acid resistance of MG1655 in stationary phase was tested. MG1655 was cultured to OD_{600} of 2 and tested under all the conditions described in previous section. The indole stock was prepared in DMSO rather than ethanol. This is because under oxygen-limiting conditions such as those encountered in stationary phase, *E. coli* uses mixed acid fermentation to generate ATP. In this process, glucose is converted into a variety of end products such as acetate, ethanol, lactate, formate and succinate which are secreted out of the cells. We were concerned that high levels of ethanol in the induction medium resulting from the ethanol in which the indole was dissolved, combined with ethanol produced during mixed acid fermentation, might have a negative effect on the survival of cells. Therefore, indole stock

was prepared in DMSO to perform the induction and indole inhibition assays in early stationary and stationary phases

Indole inhibited the innate acid resistance of MG1655 by reducing the survival of cells from 72.1% to 20.1% (Figure 4.7). However, DMSO also reduced the survival of the cells to 32.1%. Therefore, there is no evidence for any effect of indole on innate acid resistance of cells at OD_{600} of 2. MG1655 pre-incubated at pH 5.5 and acid shocked showed a decrease in survival (52%) as compared to uninduced cells (72.1%) which suggests that acid resistance cannot be induced by exposing MG1655 to mildly acidic conditions in stationary phase.

4.2.6 Induction of acid resistance and effect of indole in BW25113 grown to exponential, early stationary and stationary phases

We were interested to see whether the acid resistance through incubation at pH 5.5 can be induced in other *E. coli* strains. Additionally, we also wanted to check whether the inhibitory effect of indole on the induction of acid resistance at OD_{600} 0.2, previously seen with MG1655, was also seen in other *E. coli* strains. To test this, the survival of another laboratory K-12 strain BW25113, a pathogenic strain O157:H7 Sakai, and a commensal strain Nissle 1917, at pH 2.5 was tested with and without prior induction at pH 5.5 in the absence or presence of indole. This was repeated in exponential, early stationary and stationary phases in all three strains. In this section, the results of BW25113 will be described. The results of O157:H7 Sakai and Nissle 1917 will be described in sections 4.2.7 and 4.2.8 respectively. Lastly, the comparison of acid resistance at pH 2.5 with and without induction at pH 5.5 along with the impact of indole was drawn among all four strains, which will be discussed in section 4.2.9.

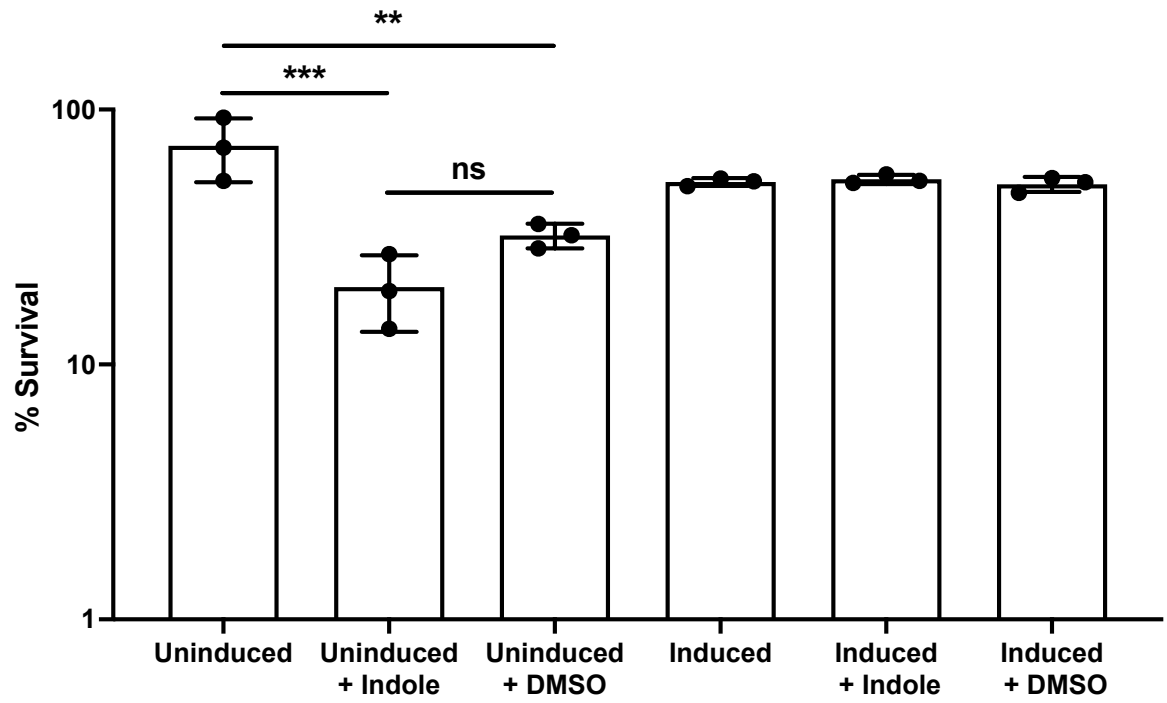


Figure 4.7: Induction of acid resistance and inhibition with indole in MG1655 grown to stationary phase.

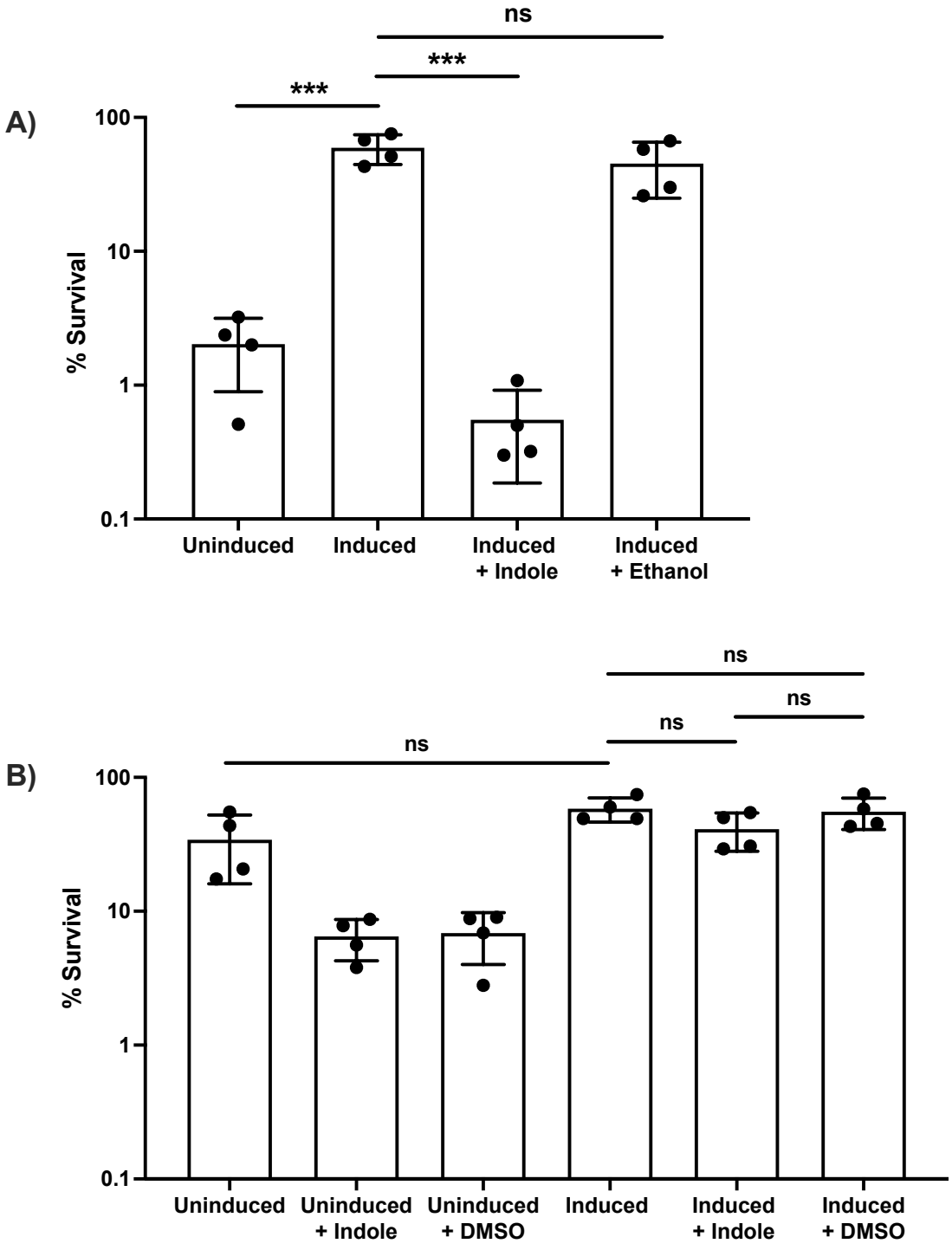
MG1655 was cultured in M9Ksupp pH 7 medium to OD_{600} of 2 and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes. The effect of indole (dissolved in DMSO) at a final concentration of 1 mM, or the same volume of DMSO alone on both uninduced (pH 7) and induced cells (pH 5.5) was determined. The survival was calculated as described in Material and Methods, section 2.5. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns= non-significant, **p < 0.01, ***p < 0.001 (One-way ANOVA, followed by Tukey's multiple comparisons test).

BW25113 was cultured to OD_{600} of 0.2 and acid stressed at pH 2.5 for 2 h, with and without induction at pH 5.5 and in the absence and presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone (Figure 4.8 A). Uninduced cells exposed to pH 2.5 had a survival of 2% whereas cells induced at pH 5.5 before acid shock had a survival of 59.3%, a 30-fold increase as compared to uninduced cells. This figure is close to the increase seen with MG1655 (42-fold, see section 4.2.3).

Addition of indole to the induction medium significantly reduced the survival of cells to 0.6%. Furthermore, addition of ethanol had a non-significant effect on survival (45.1%) suggesting that the inhibitory effect observed in survival of induced cells is due to indole.

To test the induction of acid resistance and effect of indole on uninduced and induced cells in early stationary phase, BW25113 was cultured to OD_{600} of 0.9 and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 (Figure 4.8 B). Furthermore, the effect of indole and DMSO on the survival of uninduced and induced cells was tested. Indole inhibited the innate acid resistance of BW25113 by reducing the survival of cells from 34.2% to 6.5%. However, DMSO also reduced the survival of cells to almost the same level (6.9%). Therefore, it can be concluded that indole does not reduce the survival of uninduced cells exposed to pH 2.5 in early stationary phase, as was also seen with MG1655. The pre-incubation at pH 5.5 followed by acid shock led to a non-significant increase in the survival of BW25113 from 34.2% (survival of uninduced cells) to 58.3%. Furthermore, indole had a slight but non-significant inhibitory effect on the survival of induced cells which decreased from 58.3% to 41.1%. These results suggest that the acid resistance cannot be induced by exposing BW25113 to mildly acidic pH in early stationary phase.

The induction of acid resistance and the subsequent effect of indole was also tested in stationary phase BW25113 cells under the same conditions as early stationary phase


A)
ns
***
***
100
10
1
0.1
% Survival
Uninduced
Induced
Induced + Indole
Induced + Ethanol
B)
ns
ns
ns
ns
100
10
1
0.1
% Survival
Uninduced
Uninduced + Indole
Uninduced + DMSO
Induced
Induced + Indole
Induced + DMSO

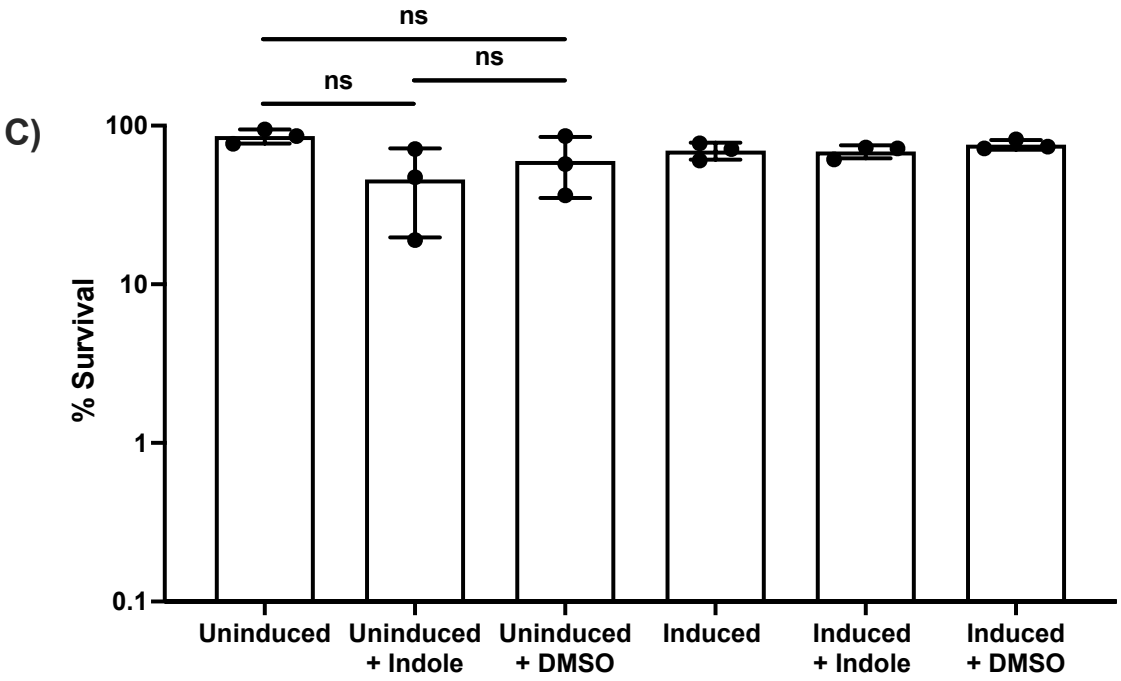


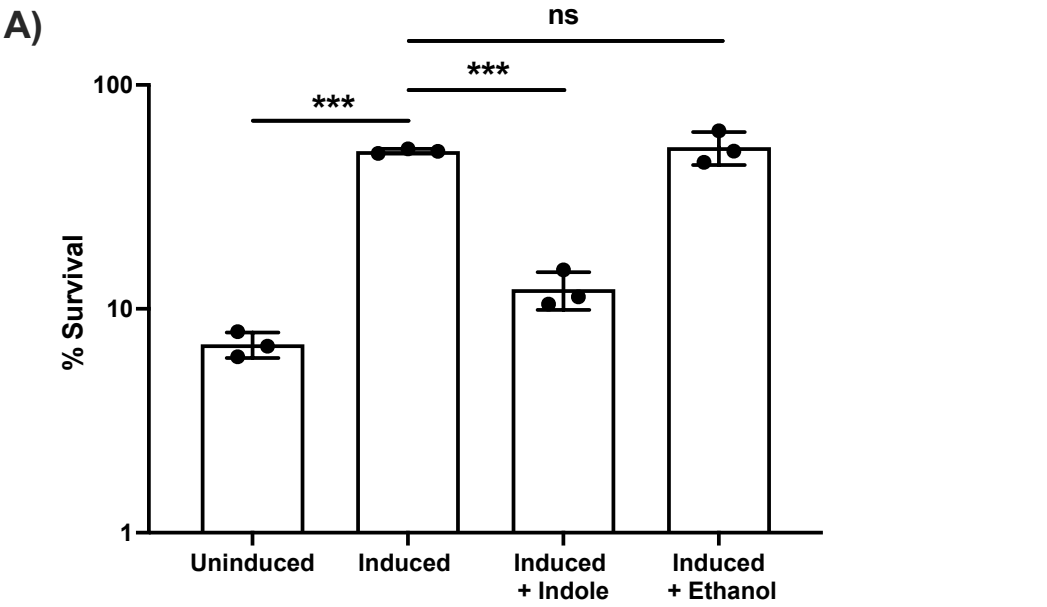
Figure 4.8: Induction of acid resistance and inhibition with indole in BW25113 in exponential, early stationary and stationary growth phases.
(A) BW25113 was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.2 and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes in the absence and presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone. The survival was calculated as described in Material and Methods, section 2.5. (B - C) BW25113 was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.9 (B) and 2 (C), and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes. The effect of indole (dissolved in DMSO) at a final concentration of 1 mM, or the same volume of DMSO alone, on both uninduced (pH 7) and induced cells (pH 5.5) was determined. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns= non-significant, ***$p < 0.001$ (Ordinary one-way ANOVA, followed by Tukey's multiple comparisons test).

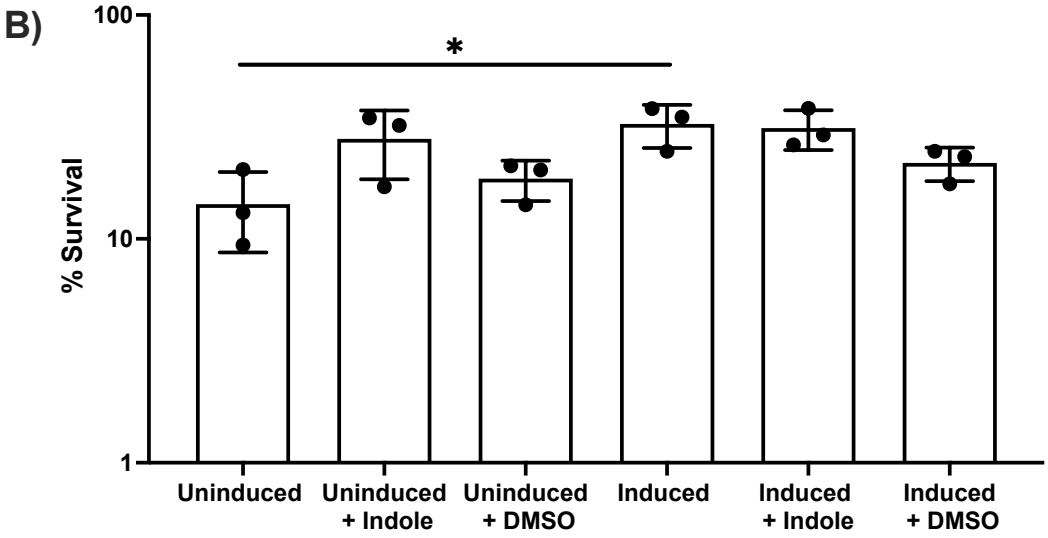
(Figure 4.8 C). Indole reduced the innate acid resistance by reducing the survival of uninduced cells from 85.8% to 45.8%. However, DMSO also reduced the survival of cells to 59.9% which means that indole does not have a significant inhibitory effect on the survival of uninduced cells exposed to pH 2.5 in stationary phase. After pre-incubation at pH 5.5, the survival of BW25113 reduced from 85.8% (survival of uninduced cells) to 69.6%. This suggests that the acid resistance cannot be induced by exposing BW25113 to mildly acidic pH in stationary phase.

Taken together, results in Figure 4.8 suggest that the innate acid resistance of BW25113 increases from exponential to stationary phase. However, indole has no significant inhibitory effect on innate resistance in early stationary and stationary phases. Furthermore, the acid resistance through exposure to mildly acidic pH can be induced and completely blocked with indole in BW25113 cells in exponential phase, but not in early stationary and stationary phases.

4.2.7 Induction of acid resistance and effect of indole in O157:H7 Sakai grown to exponential, early stationary and stationary phases

The induction of acid resistance and the inhibition of innate and induced acid resistance by indole was next tested in pathogenic strain, O157:H7 Sakai. O157:H7 Sakai was cultured to OD_{600} of 0.2 and tested under the conditions as described previously in section 4.2.3. Uninduced cells in exponential phase exposed to pH 2.5 had a survival of 6.9% which increased significantly to 50.6% after pre-incubation at pH 5.5, over 7-fold increase as compared to uninduced cells (Figure 4.9 A). The addition of indole to the induction medium significantly reduced the survival of cells to 12.2%. Furthermore, ethanol had a non-significant


A)
ns
***
***
100
10
1
% Survival
Uninduced
Induced
Induced + Indole
Induced + Ethanol


B)
*
100
10
1
% Survival
Uninduced
Uninduced + Indole
Uninduced + DMSO
Induced
Induced + Indole
Induced + DMSO

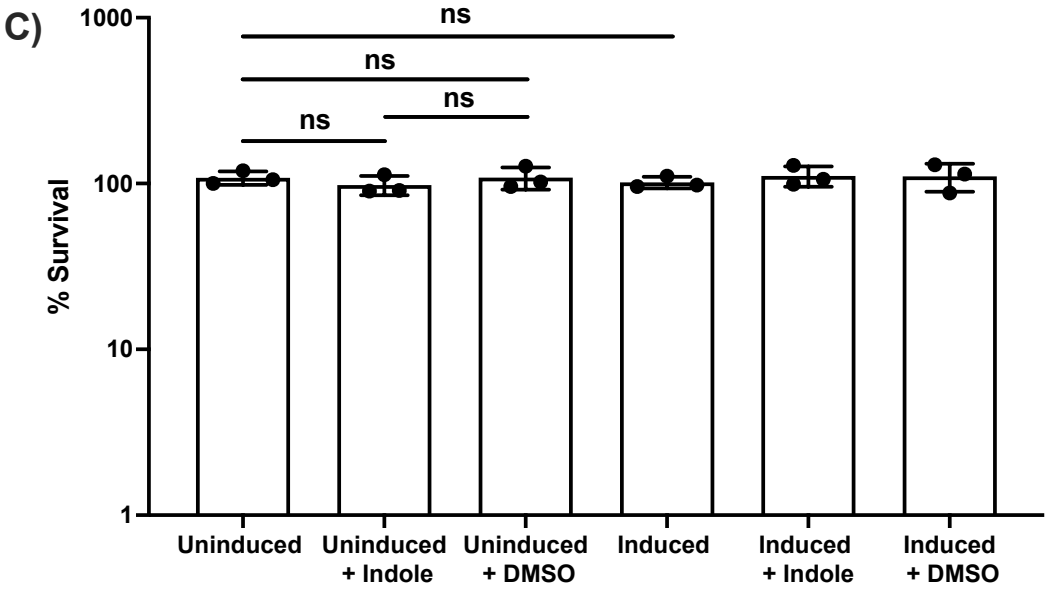

C)
ns
ns
ns
ns
1000
100
10
1
% Survival
Uninduced
Uninduced + Indole
Uninduced + DMSO
Induced
Induced + Indole
Induced + DMSO

Figure 4.9: Induction of acid resistance and inhibition with indole in O157:H7 Sakai in exponential, early stationary and stationary growth phases.
(A) O157:H7 Sakai was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.2 and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes in the absence and presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone. The survival was calculated as described in Material and Methods, section 2.5. (B - C) O157:H7 Sakai was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.9 (B) and 2 (C), and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes. The effect of indole (dissolved in DMSO) at a final concentration of 1 mM, or the same volume of DMSO alone, on both uninduced (pH 7) and induced cells (pH 5.5) was determined. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns = non-significant, *p < 0.05, ***p < 0.001 (Ordinary one-way ANOVA, followed by Tukey's multiple comparisons test).

effect on the survival of induced cells (52.7%) demonstrating that the inhibitory effect observed in survival of exponential phase cells is due to indole only.

The induction of acid resistance and its inhibition with indole was tested in early stationary phase at OD_{600} of 0.9 (Figure 4.9 B). Uninduced cells exposed to pH 2.5 had a survival of 14.3% which increased significantly to 32.6% after pre-induction at pH 5.5. These results suggest that the acid resistance through exposure to mildly acidic pH can be induced in O157:H7 Sakai grown to early stationary phase. These results are in contrast to the ones observed previously with MG1655 (Figure 4.6) and BW25113 (Figure 4.8 B) because neither of those two strains could be significantly induced in early stationary phase. The survival of induced O157:H7 Sakai cells reduced slightly in the presence of indole (31.2%). However, DMSO had a bigger negative effect on the survival of induced cells (21.8%) than indole which suggests that indole does not affect the induced acid resistance of O157:H7 Sakai in early stationary phase. Furthermore, indole had no inhibitory effect on the survival of uninduced cells (30%).

In stationary phase, there was a non-significant reduction in the survival of uninduced cells exposed to pH 2.5 from 108.3% to 101.5% after pre-induction at pH 5.5 (Figure 4.9 C). These results suggest that the acid resistance cannot be induced in stationary phase O157:H7 Sakai cells through exposure to mildly acidic pH. Furthermore, indole had a non-significant effect on the survival of uninduced and induced cells.

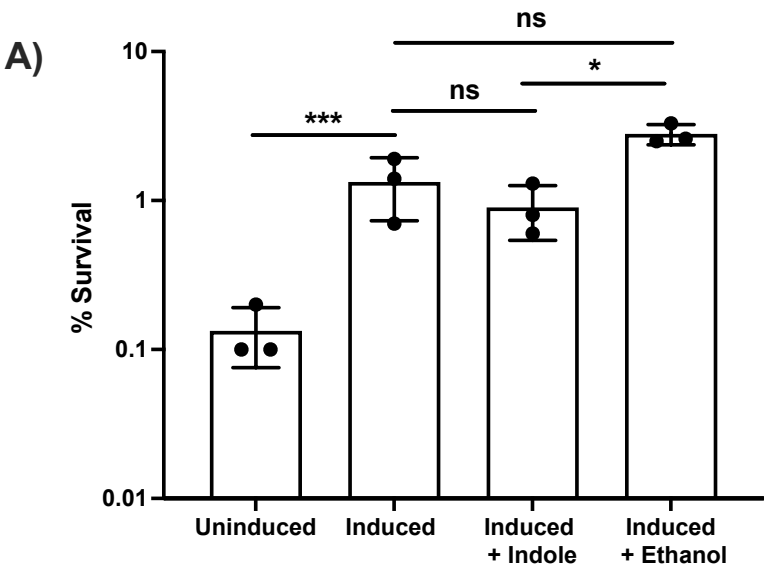
Taken together, the results in Figure 4.9 suggest that the innate acid resistance of O157:H7 Sakai increases from exponential to stationary phase. Indole has no significant inhibitory effect on the innate resistance in early stationary and stationary phases. Furthermore, the acid resistance through incubation at pH 5.5 can be significantly induced in

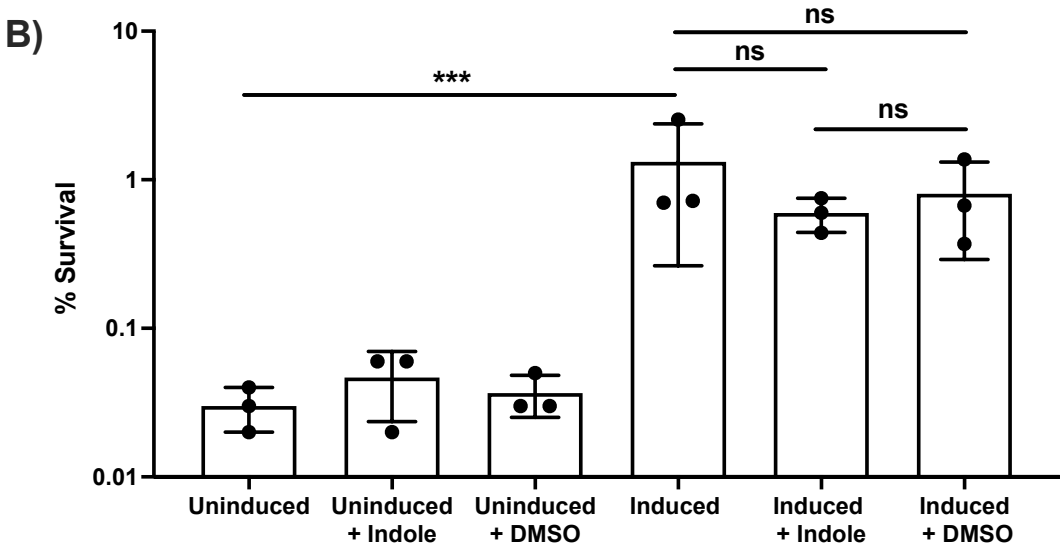
exponential and early stationary phases of growth but can be blocked with indole only in exponential phase.

4.2.8 Induction of acid resistance and effect of indole in Nissle 1917 grown to exponential, early stationary and stationary phases

Finally, the induction of acid resistance, and the effect of indole on innate and induced acid resistance was tested in the commensal strain Nissle 1917. Initial experiments showed that when cultured in M9Ksupp pH 7 medium, Nissle 1917 forms clumps due to which it was not possible to make accurate dilutions and perform survival assays with this strain. Our hypothesis was that deletion of *bcsA* gene which encodes the catalytic subunit of cellulose synthase enzyme might stop the cells from clumping together. Cellulose is one of the components present in extracellular matrix of biofilms. To test this hypothesis, the *bcsA* gene was deleted and the deletion was confirmed through PCR followed by agarose gel electrophoresis as well as whole genome sequencing (data not shown). Both wild-type and the Nissle 1917 *ΔbcsA* strains were then cultured in M9Ksupp medium at pH 7 and it was observed that the cells of Nissle 1917 *ΔbcsA* strain did not clump together while those of the wild-type strain did. The induction assays described in this section, therefore, were performed with Nissle 1917 *ΔbcsA* strain.

Nissle 1917 *ΔbcsA* strain cultured to OD_{600} of 0.2 (exponential phase) and exposed to pH 2.5 for 2 h had a survival of 0.13% which significantly increased to 1.3% after pre-induction at pH 5.5 (Figure 4.10 A). These results suggest that the acid resistance can be significantly induced in exponentially growing Nissle 1917 *ΔbcsA* exposed to mildly acidic pH. In the presence of indole, the survival of induced cells decreased from 1.3% to 0.9%. In contrast, in the presence of ethanol, the survival of induced cells increased to 2.8%. Since there was a


A)
ns
*
ns
***
10
1
0.1
0.01
% Survival
Uninduced
Induced
Induced + Indole
Induced + Ethanol


B)
ns
ns
***
ns
10
1
0.1
0.01
% Survival
Uninduced
Uninduced + Indole
Uninduced + DMSO
Induced
Induced + Indole
Induced + DMSO

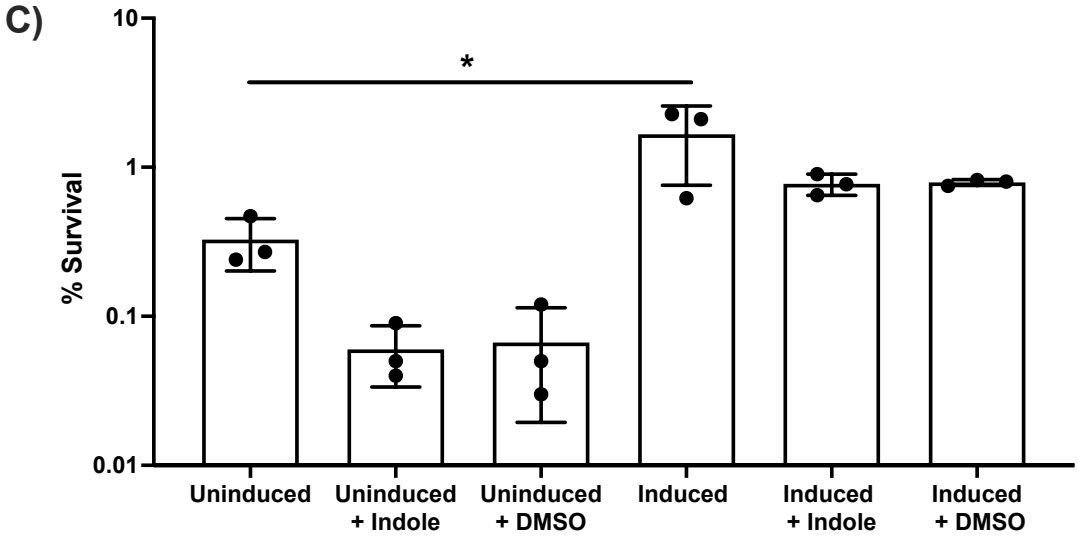

C)
*
10
1
0.1
0.01
% Survival
Uninduced
Uninduced + Indole
Uninduced + DMSO
Induced
Induced + Indole
Induced + DMSO

Figure 4.10: Induction of acid resistance and inhibition with indole in Nissle 1917 *ΔbcsA* in exponential, early stationary and stationary growth phases.
(A) Nissle 1917 *ΔbcsA* was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.2 and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes in the absence and presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone. The survival was calculated as described in Material and Methods, section 2.5. (B - C) Nissle 1917 *ΔbcsA* was cultured in M9Ksupp pH 7 medium to OD_{600} of 0.9 (B) and 2 (C), and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes. The effect of indole (dissolved in DMSO) at a final concentration of 1 mM, or the same volume of DMSO alone, on both uninduced (pH 7) and induced cells (pH 5.5) was determined. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns = non-significant, *p < 0.05, ***p < 0.001 (Ordinary one-way ANOVA, followed by Tukey's multiple comparisons test).

significant difference in the survival of induced cells in the presence of indole and ethanol, indole might have a significant inhibitory effect on the induction of induced cells.

In early stationary phase, at OD_{600} of 0.9, the uninduced cells exposed to pH 2.5 had a survival of 0.03% which significantly increased to 1.32% after pre-induction at pH 5.5 (Figure 4.10 B). These results suggest that the acid resistance can be significantly induced in Nissle 1917 *ΔbcsA* grown to early stationary phase followed by exposure to mildly acidic pH. Furthermore, indole increased the survival of uninduced cells and had no significant effect on the survival of induced cells.

In stationary phase, the uninduced cells exposed to pH 2.5 had a survival of 0.33% which significantly increased to 1.67% after pre-induction at pH 5.5 (Figure 4.10 C). These results suggest that the acid resistance can be significantly induced in Nissle 1917 *ΔbcsA* grown to stationary phase followed by exposure to mildly acidic pH. Indole had an inhibitory effect on innate acid resistance and reduced the survival of uninduced cells from 0.33% to 0.06%. However, DMSO also had an almost equal inhibitory effect on these cells (0.07%). These results suggest that indole does not inhibit the innate acid resistance of Nissle 1917 *ΔbcsA* in stationary phase. Furthermore, both indole and DMSO had an equal inhibitory effect on the survival of induced cells which reduced from 1.67% to 0.8% after addition of indole and DMSO. This suggests that indole does not inhibit the induced acid resistance of Nissle 1917 *ΔbcsA* in stationary phase.

The result in Figure 4.10 together suggest that the innate acid resistance of Nissle 1917 *ΔbcsA* first decreases from exponential phase to early stationary phase and then increases again in stationary phase. Furthermore, the acid resistance in Nissle 1917 *ΔbcsA* can be significantly induced in all three growth phases through incubation at pH 5.5. Finally, indole

has no significant effect on the survival of uninduced cells but may have a significant effect on the survival of induced cells in exponential phase.

4.2.9 Comparison of acid resistance profiles among MG1655, BW25113, O157:H7 Sakai, and Nissle 1917 *ΔbcsA*

The innate acid resistance of all four strains, MG1655, BW25113, O157:H7 Sakai and Nissle 1917 *ΔbcsA*, without the induced acid resistance was compared in exponential (OD_{600} of 0.2), early stationary (OD_{600} of 0.9) and stationary (OD_{600} of 2) phases by measuring the survival of cells after being exposed to pH of 2.5 for 2 h (Figure 4.11 A). MG1655, BW25113 and O157:H7 Sakai showed similar levels of survival, although O157:H7 Sakai had a slightly higher survival (6.9%) than MG1655 (3%) and BW25113 (2%) in exponential phase. Furthermore, the innate acid resistance of all three strains increased with increase in optical density. MG1655, BW25113 and O157:H7 Sakai had a survival of 72.1%, 85.8% and 108.3% in stationary phase. On the other hand, Nissle 1917 *ΔbcsA* showed a much lower survival than the remaining three strains across all three growth stages. Furthermore, the survival of Nissle 1917 *ΔbcsA* decreased from 0.13% in exponential phase to 0.03% in early stationary phase before increasing again to 0.33% in stationary phase.

All four strains showed an induced acid resistance in exponential phase, evident through higher survival of cells pre-induced at pH 5.5 than uninduced cells, acid stressed at pH 2.5 (Figure 4.11 B). Since uninduced MG1655, BW25113 and O157:H7 Sakai had similar level of survival, MG1655 showed the highest level of induction with 42-fold increase in the survival of induced cells as compared to uninduced cells. BW25113 and O157:H7 Sakai induced cells showed a 30-fold and a 7-fold increase in survival respectively, as compared to uninduced cells. Nissle 1917 *ΔbcsA* also showed a significant increase in survival after being

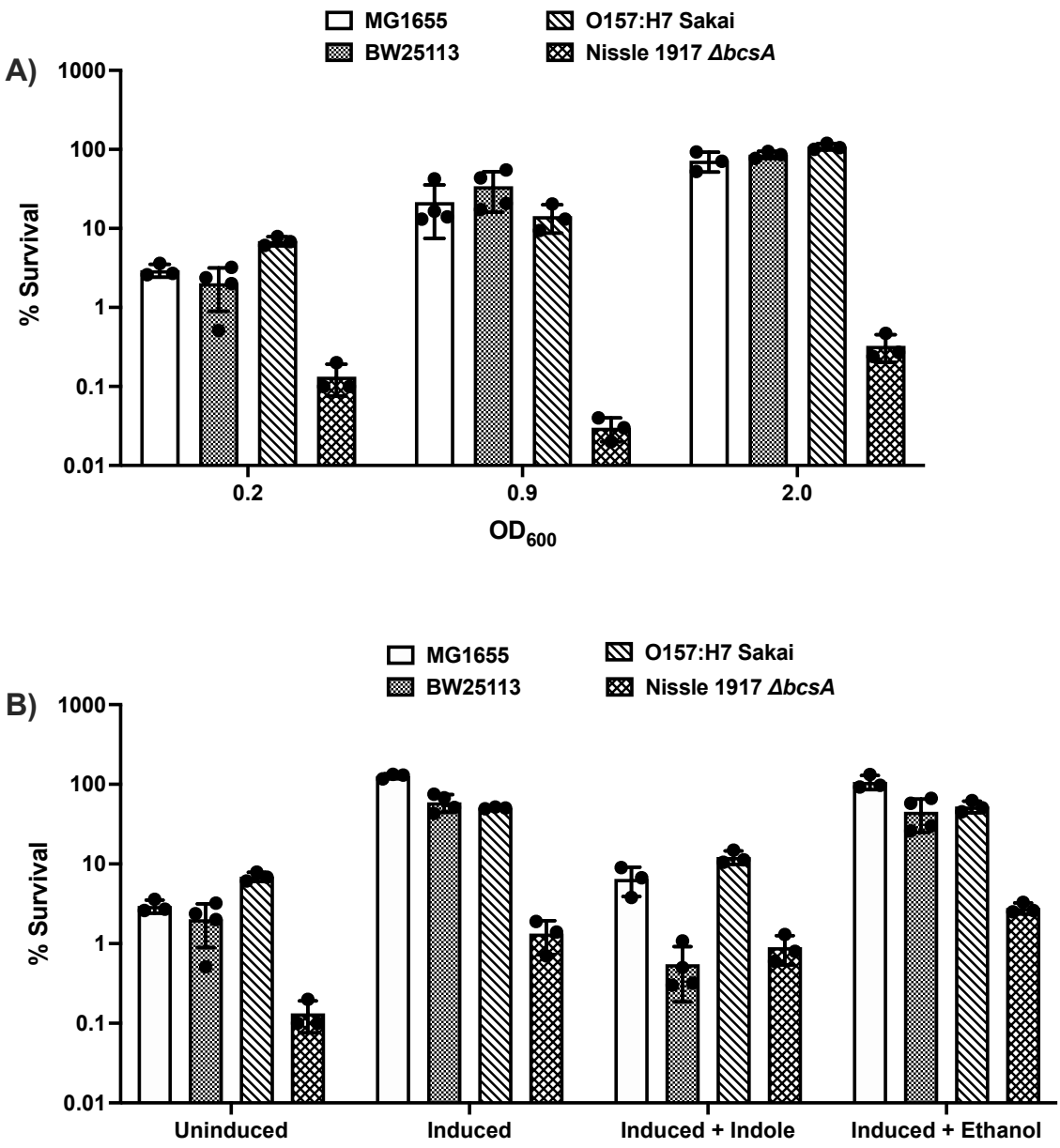


Figure 4.11: Comparison of survival among MG1655, BW25113, O157:H7 Sakai and Nissle 1917 *ΔbcsA* after being acid stressed at pH 2.5 with and without induction at pH 5.5.
(A) MG1655, BW25113, O157:H7 Sakai and Nissle 1917 *ΔbcsA* were cultured in M9Ksupp pH 7 medium to OD_{600} of 0.2, 0.9 and 2 before being acid stressed at pH 2.5 for 2 h. The survival was calculated as described in Material and Methods, section 2.5. (B) MG1655, BW25113, O157:H7 Sakai and Nissle 1917 *ΔbcsA* were cultured in M9Ksupp pH 7 medium to OD_{600} of 0.2 and acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes in the absence and presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

induced in exponential phase. However, the survival of Nissle 1917 *ΔbcsA* (1.3%) was much lower than MG1655 (126.7%), BW25113 (59.3%) and O157:H7 Sakai (50.6%). Indole showed a complete inhibition of the AR2 response in BW25113 (0.6%) and high levels of inhibition in MG1655 (6.5%) and O157:H7 Sakai (12.2%). There was a significant difference observed in the survival of induced Nissle 1917 *ΔbcsA* cells in the presence of indole and ethanol, therefore indole may have a significant inhibitory effect on induced Nissle 1917 *ΔbcsA* cells.

EvgS shows natural variation in its pH-sensing capacity, and some strains have been shown to not respond to low pH at all (Roggiani, Yadavalli and Goulian, 2017). To test whether the reason behind Nissle 1917 *ΔbcsA* not being able to get induced to the similar level as other strains at pH 5.5 was due to its inactive EvgS, Nissle 1917 *ΔbcsA* was complemented with a constitutively active copy of EvgS protein (EvgS-S600I) from MG1655 on pBAD24 plasmid (Johnson *et al.*, 2014). Since MG1655 showed high levels of induction in exponential phase, the addition of constitutively active copy of EvgS protein from MG1655 to Nissle 1917 *ΔbcsA* should give high levels of induction. Nissle 1917 *ΔbcsA* and Nissle 1917 *ΔbcsA* pBADevgS-S600I strains were cultured to OD_{600} of ~0.2 in M9Ksupp pH 7 medium, supplemented with 0.02% arabinose to induce the expression of EvgS-S600I protein. The strains were acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 in the absence and presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone. The addition of both pBAD24 vector plasmid and pBADevgS-S600I plasmid increased the survival of uninduced cells as compared to Nissle 1917 *ΔbcsA* (Figure 4.12). However, the presence of pBADevgS-S600I did not result in high levels of induction because the survival of induced cells only increased from 1.3% to 4.6% after the addition of plasmid. These results suggest that the inability of Nissle 1917 *ΔbcsA* to not get induced at pH 5.5 is not due to EvgS but another factor.

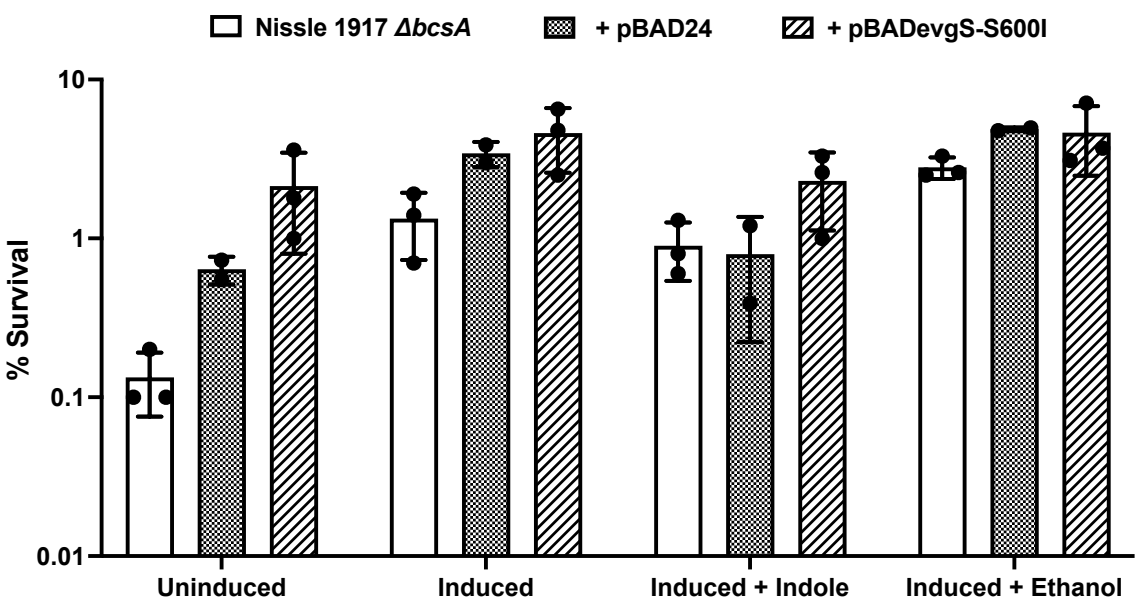


Figure 4.12: Comparison of survival of Nissle 1917 *ΔbcsA* in the presence and absence of constitutively active EvgS protein from MG1655, after being acid stressed at pH 2.5 with and without prior induction at pH 5.5.

Nissle 1917 *ΔbcsA* was transformed with pBADevgS-S600I plasmid containing constitutively active EvgS protein (EvgS-S600I) from MG1655. The parent and transformed strains were cultured in M9Ksupp pH 7 medium, supplemented with 0.02% arabinose to induce the expression of EvgS-S600I protein, to OD_{600} of 0.2. The strains were acid stressed at pH 2.5 for 2 h with and without induction at pH 5.5 for 30 minutes in the absence and presence of indole (dissolved in ethanol) at a final concentration of 1 mM, or the same volume of ethanol alone. The survival was calculated as described in Material and Methods, section 2.5. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

4.3 Discussion

In this study, the survival under extreme acid stress of pH 2.5, with and without the prior induction of acid resistance, in the absence and presence of indole, among laboratory (K-12 strains MG1655 and BW25113), pathogenic (O157:H7 Sakai) and commensal (Nissle 1917 *ΔbcsA*) *E. coli* strains was studied. Innate acid resistance of all strains except Nissle 1917 *ΔbcsA*, increased with increase in optical density (Figure 4.11 A). Furthermore, uninduced MG1655, BW25113 and O157:H7 Sakai showed similar susceptibility to acid with O157:H7 Sakai being slightly more acid resistant than the other two strains in exponential phase. Nissle 1917 *ΔbcsA*, however, showed much higher susceptibility to acid than the other three strains. This result was surprising because being a probiotic strain, Nissle 1917 has to pass through the stomach acid in order to exert its effects in the intestine. Therefore, Nissle 1917 *ΔbcsA* would be expected to be more acid resistant.

The effect of extreme acid stress (pH 2.5) across 2 laboratory (K-12 and MC4100), 3 commensal (isolated from different patients), and 11 pathogenic (O157:H7 EDL933, and 10 O157:H7 patient isolates) *E. coli* strains has been studied by Lin *et al.* (1996). The conditions used by Lin *et al.* (1996) were different to this study because the cultures were grown in E minimal media (73 mM $K_2 HPO_4$, 17 mM $NaNH_4 HPO_4$, 0.8 mM $MgSO_4$, and 10 mM citrate) (pH 7) supplemented with glucose, for 18-20 h, and acid stressed in the same media at pH 2.5 for 2 h at 37 °C. Most of the strains had a survival of less than 0.01%. Four pathogenic isolates had a slightly higher survival (0.02, 0.04, 0.03, and 0.06%) than 0.01%, and one commensal isolate had a survival of 4.1%. These survival values are much lower than the ones observed in our study in stationary phase, possibly due to the use of different media in both studies.

Comparison between acid resistance of MG1655 and MG1655 *ΔrpoS* showed that RpoS confers innate acid resistance to cells in exponential and early stationary phases. The

levels of RpoS are low during exponential phase but increase several fold in stationary phase (Schweder *et al.*, 1996; Becker, Klauck and Hengge-Aronis, 1999). During exponential phase, the adaptor protein RssB binds RpoS and marks it for proteolytic degradation by ClpXP protease (Becker, Klauck and Hengge-Aronis, 1999). During acid stress, the antiadaptor IraM binds RssB preventing its association with RpoS which increases the RpoS levels in the cell (Eguchi *et al.*, 2011). RpoS is an important sigma factor for stress management and thought to mainly control gene regulation in stationary phase. However, microarray analysis with MG1655 and MG1655 *ΔrpoS* has shown several genes which are regulated by RpoS in exponential phase including *dnaK*, which encodes a heat shock protein and gets induced at low pH (Heyde and Portalier, 1990; Dong, Kirchhof and Schellhorn, 2008). Various K-12 strains with *rpoS* deletion, including MG1655, cultured to stationary phase and acid stressed at pH 1 for 30 minutes showed lower survival than the parent strain (Thea *et al.*, 2004). O157:H7 EDL933 acid stressed at pH 2.5 and 1.5 in late-exponential and stationary phases (OD_{600} ranging from 0.5 - 2.5) showed a higher survival than the *rpoS* knockout which showed almost no survival (Coldewey *et al.*, 2007). The same study further showed that the acid susceptible phenotype of *rpoS* knockout could be restored completely through complementation with a wild-type copy of RpoS. In another study, MG1655 *ΔrpoS* strain cultured in M9 minimal medium, to exponential phase and acid stressed at pH 2.5 for 2 h had a log-fold reduced survival than the parent strain (Šeputienė *et al.*, 2006). These results suggest that RpoS improves the innate acid resistance of both laboratory and pathogenic strains in both exponential and stationary phase, which is consistent with the results of our study.

All four strains tested in our study: MG1655, BW25113, O157:H7 Sakai, and Nissle 1917 *ΔbcsA* showed increased acid resistance after getting induced at pH 5.5 in exponential phase (Figure 4.11 B). The induction of acid resistance at pH of 5.5 - 5.7 in MG1655 and

BW25113 has been reported previously using a combination of survival and promoter-probe assays (Burton *et al.*, 2010; Eguchi and Utsumi, 2014; Boon *et al.*, 2020; Inada *et al.*, 2021). Another study has reported that acid resistance can be induced in exponentially growing MG1655, commensal strain HS, enterotoxigenic strain H10407, and enteropathogenic strain E2348/69, but not in enterohemorrhagic strain O157:H7 EDL933, uropathogenic strain UTI89, and commensal strain Nissle 1917 (Roggiani, Yadavalli and Goulian, 2017). These findings are different from the results of our study where both O157:H7 and Nissle 1917 *ΔbcsA* showed significant induction of acid resistance.

The induction of acid resistance in MG1655, BW25113, O157:H7 Sakai, and Nissle 1917 *ΔbcsA* was also tested in early stationary and stationary phases. O157:H7 Sakai was the only strain that showed significant and high induction in early stationary phase (Figure 4.9 B), which to our knowledge has not been reported previously.

Natural variation in *evgS* sequence can alter its ability to sense and respond to low pH (Roggiani, Yadavalli and Goulian, 2017). In our study, Nissle 1917 *ΔbcsA* showed much lower levels of induced acid resistance in exponential phase as compared to the other three strains (Figure 4.11 B). In the study performed by Roggiani, Yadavalli and Goulian (2017), Nissle 1917 did not show any induced acid resistance when incubated at pH 5.7 in exponential phase. This phenotype, however, could be restored through complementation with a wild-type copy of EvgS from MG1655. This finding is contrary to the results of our study where complementation with EvgS from MG1655 could not increase the levels of induced acid resistance in exponentially growing Nissle 1917 *ΔbcsA* (Figure 4.12). The Nissle 1917 strains used in our study and the one by Roggiani, Yadavalli and Goulian (2017) have both been acquired from Ardeypharm, the company that sells this strain as a probiotic drug by the name of Mutaflor. However, when whole genome sequencing was performed on our strain, its

sequence was quite different from the reference strain submitted by Reister *et al.* (2014) (NZ_CP007799.1). It is possible that there are different Nissle 1917 isolates available with variations present in their genomic sequences. The variability in the findings of our study and the one by Roggiani, Yadavalli and Goulian (2017) could be due to the differences in the genomic sequences of Nissle 1917 isolates used in the two studies. The other possibility for the variation in the results of two studies could be due to *bcsA* deletion present in our strain.

Indole inhibited the induced acid resistance of MG1655, BW25113, and O157:H7 Sakai in exponential phase, but not the induced acid resistance of O157:H7 Sakai in early stationary phase (Figures 4.5, 4.8 A & B, 4.9 B). This suggests that the inhibitory effect of indole on induced acid resistance is unique to exponentially growing cells. In support of the findings of our study, another study has shown that the addition of 50 µM indole to the growth medium (LB broth) inhibits the GAD decarboxylase activity of MG1655 cells cultured to OD_{600} of 0.5 and induced at pH 5.5 for 30 minutes (Kanda *et al.*, 2020). A second study performed with BW25113 cultured to mid-exponential phase has shown that 2 mM indole reduces the acid resistance of cells at pH 2.5 (Lee *et al.*, 2007). Contrary to these results and the results of our study, 1 mM indole has been shown to enhance the acid resistance of MC4100 strain cultured to OD_{600} of 0.8, at pH 3.5 (Hirakawa *et al.*, 2010). The last two studies, however, were both performed in LB and without the pre-induction step at pH 5.5. Therefore, the results observed in these studies are not associated with the induced acid resistance.

In conclusion, the innate acid resistance of *E. coli* strains generally increases with increase in optical density, and part of this resistance is conferred by RpoS. Variability exists in the ability of different strains to induce their acid resistance in mildly acidic pH in exponential phase, which may or may not depend on EvgS. Indole successfully inhibits the induced acid resistance in exponential phase, among laboratory and pathogenic strains. The

acid resistance, through incubation at pH 5.5, can also be induced outside the exponential phase and this induction is unaffected by indole. Therefore, more studies should focus on the induced acid resistance, under mildly acidic conditions, in different growth phases and strains, to understand the full potential of this response and the variability among strains.

Chapter 5

Using Tradis to study the induction of the acid resistance response, and its inhibition with indole, in MG1655

5.1 Introduction

The main aims of this study were (a) to find genes that may be important in the ability of *E. coli* to survive at pH 2.5, (b) to find genes that are involved in the inhibitory effect of indole on the induction of acid stress response, and (c) to compare the acid stress response between laboratory, pathogenic, and commensal *E. coli* strains. The work comparing the acid stress responses between different strains was described in Chapter 4 where indole was shown to inhibit the induction of acid stress response in exponentially growing MG1655, BW25113 and O157:H7 Sakai strains. More work on comparison between a laboratory (MG1655) and a pathogenic (O157:H7 Sakai) strain will be described in Chapter 6.

In this Chapter, using TraDIS in MG1655, the genes that are needed for the optimal fitness of *E. coli* at pH 2.5 and the genes which are involved in inhibiting the acid stress response through indole will be determined. TraDIS involves growing a mutant library containing transposon insertions in every non-essential gene under a stress condition. The sequencing libraries, referred to as TraDIS libraries, are then prepared from control library and the library grown under stress. The control library can either be the initial transposon library (ITL) or the ITL grown under the experimental conditions but without the stress. By comparing a decrease or increase in the frequency of the mutants containing insertions in different genes, in the conditional TraDIS libraries relative to the control TraDIS library, the genes whose function or loss of function provides a fitness advantage under a stress condition can be determined.

The MG1655 transposon library grown to mid-exponential phase (OD_{600} of ~0.2) was incubated at pH 5.5 to induce the acid stress response, either in the presence of indole or ethanol, followed by acid stress at pH 2.5 and then recovery at pH 7 for 3 hours. The genes whose function provides a fitness advantage at pH 2.5 were identified as the ones where the

frequency of inserts in two conditional libraries declined, relative to the ITL. On the other hand, the genes that might be involved in inhibitory effect of indole were identified as the ones where the frequency of inserts increased in the indole-treated but not the ethanol-treated conditional library, relative to the ITL.

The advantage of using TraDIS is that a potential gene function can be identified without any prior evidence being required. Moreover, this process of connecting gene function with a phenotype is done at genome level rather than gene level, which can be time consuming. Furthermore, analysis of TraDIS sequencing data using an analytical tool such as Gene set enrichment analysis (GSEA) can help to identify the potential metabolic and transcriptional regulatory pathways which may have an important role in determining the fitness of cells under a particular stress. This helps to understand how an organism responds to a given stress at genome level.

To date, only one study has used TraDIS to study acid stress in bacteria. This study was done on *Salmonella* Derby 14T grown to stationary phase in LB at pH 4 (Gu *et al.*, 2021). It is not possible to draw direct comparisons between this study and ours because (a) *Salmonella* does not have an AR2 acid resistance response, and (b) the conditions used in our study are completely different- mid-exponential growth phase, M9K minimal medium, acid stress at pH 5.5 followed by pH 2.5. Therefore, ours is the first study which will provide an optimised protocol for using TraDIS to study AR2 acid resistance response in exponentially growing *E. coli* and examine the impact of indole on the induction of AR2 response.

5.2 Results

5.2.1 Optimisation of conditions for TraDIS experiments in *E. coli* K12 MG1655

TraDIS experiments are essentially competition experiments where a very large number of different mutants are competing against each other. The fitness of each mutant is then judged by seeing how much their relative proportions change over time, either compared to the initial traDIS library or compared to the same library grown under a suitable control condition. To improve the chances of a significant difference in fitness being observed, it is generally necessary to optimise the conditions of the experiment. Typical parameters that may be altered might include the length or intensity of the stress used, and the recovery time after the stress is removed. A condition which is too stressful may result in the loss of a large number of mutants and hence poor resolution of data, whereas a condition which is not stressful enough may not be sufficient to detect relatively small differences in fitness. Moreover, after stressing cells, a period of outgrowth under non-stress conditions may be needed to see the full effects of mutations that lead to different levels of fitness between different strains. Additionally, at the end of the outgrowth step, a sufficient number of live cells should be present in the culture so that a good DNA yield can be obtained to prepare the sequencing library. Therefore, before starting any actual TraDIS experiments, the conditions to be used need to be optimised. For our TraDIS experiments, the three parameters which were optimised were the acid stress duration at pH 2.5, the outgrowth time following acid stress, and the culture volume required for DNA extraction.

5.2.1.1 Optimisation of acid shock duration at pH 2.5

To optimise the duration for which the cells should be acid shocked at pH 2.5, after being induced at pH 5.5 in a TraDIS experiment, a competition assay was done between MG1655 and MG1655 *ΔevgS* strains. The MG1655 *ΔevgS* strain shows relatively poor fitness at pH 2.5, as EvgS is necessary for the induction of the AR2 response by a moderate pH drop (Ma, Masuda and Foster, 2004; Burton *et al.*, 2010). So MG1655 *ΔevgS* represents all the mutants which will show high sensitivity to low pH, such as the mutants with transposon inserted within different AR2 genes, in TraDIS experiments. The time point at which MG1655 and MG1655 *ΔevgS* strains show sufficient difference in the survival after being acid shocked post-induction, but at the same time a complete loss of survival of MG1655 *ΔevgS* is prevented, would be the ideal duration for acid shock at pH 2.5 in a TraDIS experiment. This time point would ensure that highly sensitive mutants to acid were represented, and at the same time the difference between the acid sensitive and acid resistant mutants can be determined.

Both MG1655 and MG1655 *ΔevgS* produce red colonies on macConkey agar containing lactose. To distinguish MG1655 from MG1655 *ΔevgS,* MG1655 with a *lacZ* deletion (KH001), which produces white colonies on lactose-supplemented macConkey agar, was competed with MG1655 *ΔevgS.* To determine that KH001 has the same fitness as MG1655 and can be used as its proxy, the competition assay was first performed between KH001 and MG1655.

KH001 and MG1655 were grown separately to mid-exponential phase (OD_{600} ~0.2), normalised to a precise OD_{600} of 0.2, and mixed in equal proportions before being incubated at pH 5.5 either in the presence of indole or ethanol. The induced cultures were further acid shocked at pH 2.5 for 15, 30 and 60 minutes before being plated out. The survival of strains

was calculated as described in Materials and Methods, section 2.5. As expected, there was no significant difference in the survival of KH001 and MG1655 acid stressed at pH 2.5 after 15, 30 and 60 minutes (Figure 5.1 A). This was true regardless of whether the strains were pre-induced at pH 5.5 in the presence of indole or ethanol.

When MG1655 Δ*evgS* was competed against KH001, the survival of both the strains decreased as the incubation time at pH 2.5 increased (Figure 5.1 B). As expected, when the strains were pre-induced at pH 5.5 in the presence of ethanol, KH001 had a higher survival than MG1655 Δ*evgS* after 15, 30 and 60 minutes of incubation at pH 2.5. This difference in survival is due to the absence of EvgS, which is required for the activation of AR2 response increasing the survival of cells exposed to pH 2.5. As indole is thought to inhibit the AR2 in exponential phase (Boon *et al.*, 2020), it was expected that the higher survival of KH001 would not be seen if the cells were pre-induced in the presence of indole. Confirming this, as can be seen in Figure 5.1 B, there was no significant difference in the survival of KH001 and MG1655 Δ*evgS* pre-induced in the presence of indole and acid stressed at pH 2.5.

As discussed previously, the acid shock duration at pH 2.5 which would be suitable for TraDIS experiments is the one which is stressful enough so that the small differences in fitness between different conditions can be detected while preventing complete loss of particularly acid-sensitive mutants. In the competition experiments, acid stress at pH 2.5 following pre-induction at pH 5.5 in the presence of indole led to similar levels of survival of KH001 and MG1655 Δ*evgS,* which was expected. However, pre-induction in the presence of ethanol, followed by acid shock, was expected to result in different levels of survival of these two strains. A maximum difference in the survival of KH001 and MG1655 Δ*evgS* pre-induced in the presence of ethanol was found when the strains were acid stressed for a period of 60 minutes. Although this duration would allow to determine the difference in the fitness of acid

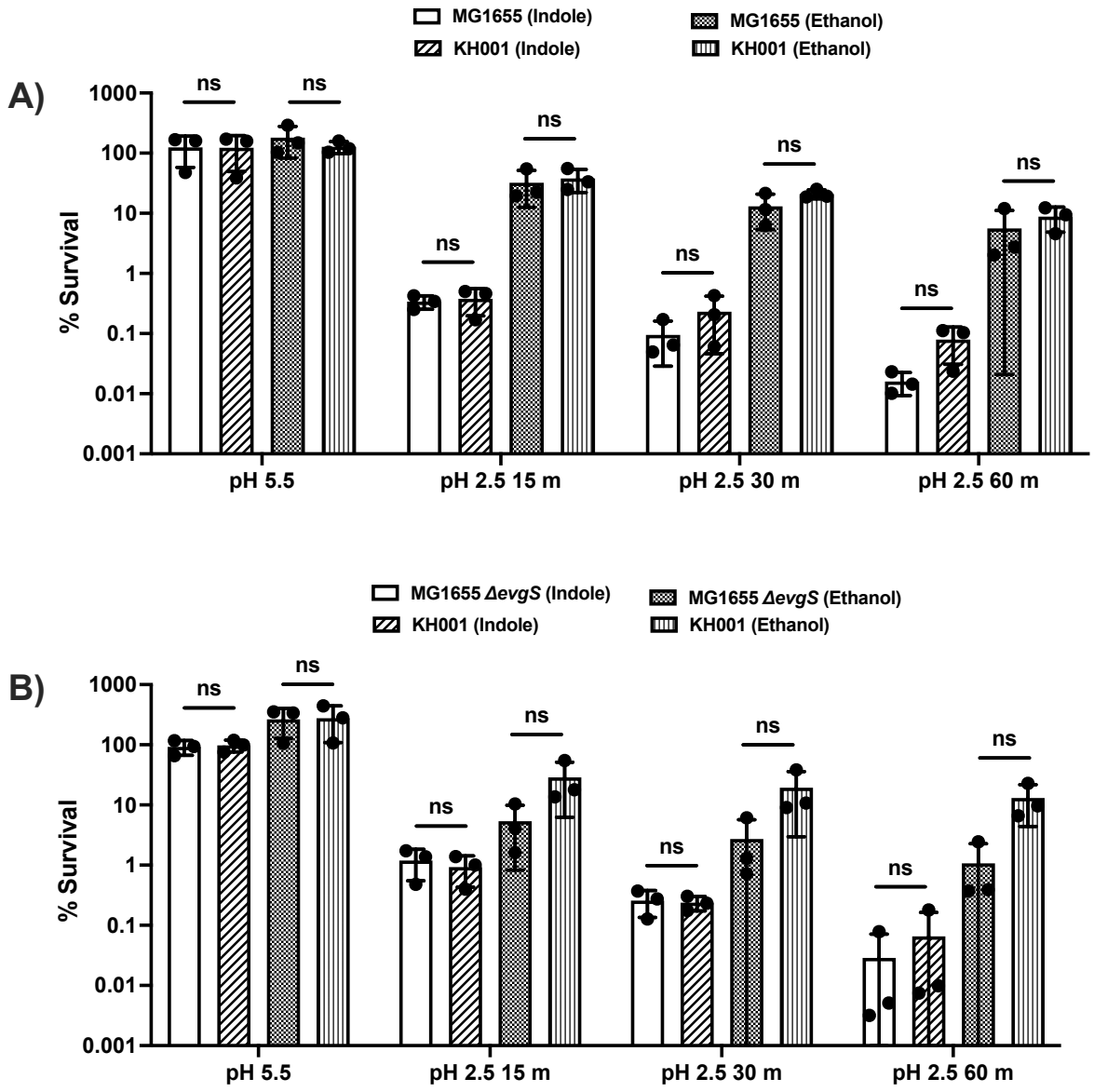


Figure 5.1: Competition of MG1655 or MG1655 *ΔevgS* against KH001 under AR2 induction conditions in the absence or presence of 1 mM indole, followed by different acid shock durations at pH 2.5.

MG1655, MG1655 *ΔevgS* and KH001 were cultured separately to mid-exponential phase, normalised to OD_{600} of 0.2, and mixed in equal proportions before being induced at pH 5.5 either in the presence of indole or ethanol, followed by acid shock at pH 2.5 for 15, 30 and 60 minutes. The survival was calculated as described in Material and Methods, section 2.5. (A) Comparison of survival between KH001 and MG1655. (B) Comparison of survival between KH001 and MG1655 *ΔevgS*. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns = not significant (Unpaired t-test, followed by Benjamini, Krieger, and Yekutieli multiple comparisons test).

sensitive and acid resistant mutants in a TraDIS experiment with most precision, the stress may be a bit too strong for indole sensitive mutants. Therefore, a duration of 30 minutes was selected for TraDIS experiments to ensure that indole sensitive mutants are represented in the sample.

5.2.1.2 Optimisation of culture outgrowth time after acid shock at pH 2.5

During TraDIS experiments, in order to see the effect of different mutations which impart either acid sensitive or acid resistant phenotype, the cultures need to be outgrown at neutral pH after acid shock at pH 2.5. The outgrowth step will allow the acid resistant mutants to increase in frequency whereas the frequency of acid sensitive mutants will either remain the same or change to a smaller extent relative to acid resistant mutants.

After the outgrowth step, the DNA is extracted, and sequencing libraries are prepared. During the analysis of sequencing reads, the frequency of mutants containing transposon insertions within different non-essential genes in the conditional TraDIS libraries and the initial TraDIS library (ITL) are compared. An increase or decrease in the frequency of mutants containing insertions in a particular gene tells whether that insertion mutation is beneficial or detrimental to the fitness of the cells.

To determine the optimum time for outgrowth, the competition assay between KH001 and MG1655 Δ*evgS,* as described in section 5.2.1.1, was repeated (Figure 5.2 A). The acid stressed cultures were outgrown for 3 hours and their CFU mL^{-1} was calculated every hour (Figure 5.2 B). The outgrowth duration of 3 hours, as opposed to a longer duration, was used to keep the cultures in exponential growth phase and prevent any additional stresses associated with stationary phase from acting on the cells.

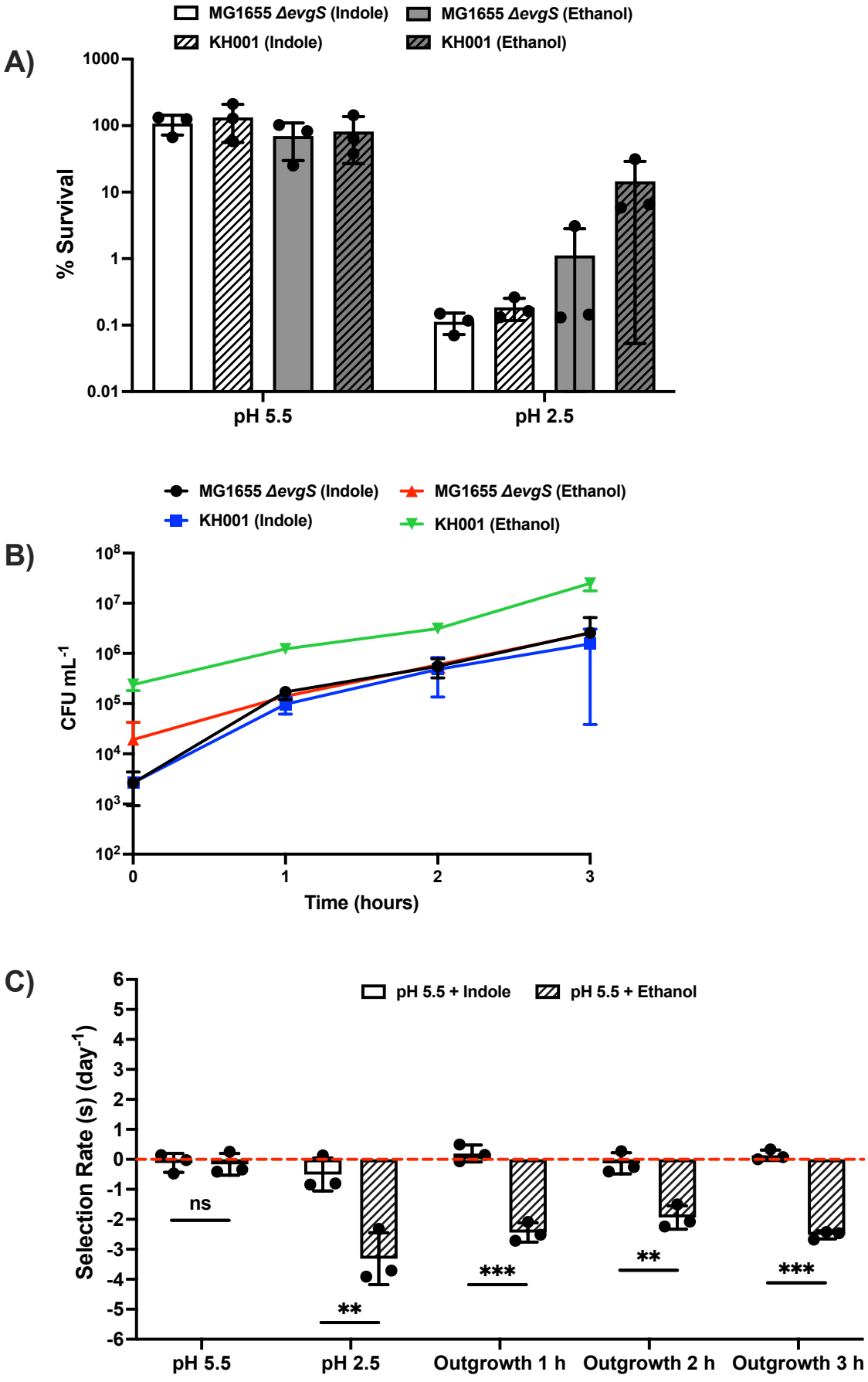

A)
MG1655 ΔevgS (Indole)
MG1655 ΔevgS (Ethanol)
KH001 (Indole)
KH001 (Ethanol)
% Survival
1000
100
10
1
0.1
0.01
pH 5.5
pH 2.5
B)
MG1655 ΔevgS (Indole)
MG1655 ΔevgS (Ethanol)
KH001 (Indole)
KH001 (Ethanol)
CFU mL-1
Time (hours)
C)
pH 5.5 + Indole
pH 5.5 + Ethanol
Selection Rate (s) (day-1)
ns
**
***
**
***
pH 5.5
pH 2.5
Outgrowth 1 h
Outgrowth 2 h
Outgrowth 3 h

Figure 5.2: Competition of MG1655 Δ*evgS* against KH001 under AR2 induction conditions in the absence and presence of 1 mM indole, followed by outgrowth.
MG1655 Δ*evgS* and KH001 were cultured separately to mid-exponential phase, mixed in equal proportions, induced at pH 5.5 either in the presence of indole or ethanol, followed by acid shock at pH 2.5 for 30 minutes. The acid stressed cultures were then outgrown at neutral pH for 3 hours. The selection rate was calculated as described in Material and Methods, section 2.12.1. (A) Comparison of survival between KH001 and MG1655 Δ*evgS*. (B) Increase in CFU mL^{-1} of KH001 and MG1655 Δ*evgS* during outgrowth. (C) The fitness of MG1655 Δ*evgS* relative to KH001, expressed as selection rate (s). There was no difference in the fitness of strains at selection rate of 0 (shown by red dotted line). Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns = not significant, **p < 0.01, ***p < 0.001 (Unpaired t-test, followed by Benjamini, Krieger, and Yekutieli multiple comparisons test).

As expected and as shown previously in section 5.2.1.1, when the strains were acid stressed following pre-induction at pH 5.5 in the presence of ethanol, KH001 had a higher survival than MG1655 Δ*evgS* (Figure 5.2 A). When the strains were pre-induced in the presence of indole followed by acid stress, both had similar survival. During outgrowth, both the strains showed a steady increase in CFU mL^{-1}, whether they were induced in the presence of indole or ethanol (Figure 5.2 B). This shows that the treatment applied (indole or ethanol) during the induction step does not affect the growth of the cells during outgrowth.

During TraDIS experiments, the outgrowth step should not affect the difference in the fitness of acid sensitive and acid resistant strains. The end of the outgrowth step should reflect the difference in survival which was present immediately after the acid shock step. To check whether this was true for KH001 and MG1655 Δ*evgS* in this competition assay, the fitness of MG1655 Δ*evgS* relative to KH001 was calculated using selection rate, as described in Materials and Methods, section 2.12.1, after the induction and acid shock steps, as well as during the outgrowth (Figure 5.2 C). After the acid shock at pH 2.5, MG1655 Δ*evgS* had a lower fitness than KH001 when the strains were induced in the presence of ethanol, but not in the presence of indole. The difference in the susceptibility of MG1655 Δ*evgS* to acid stress, relative to KH001, in the presence of indole and ethanol, was significant after the acid shock step as well as after 3 hours of outgrowth. This shows that the end of the outgrowth step reflects the differences in the fitness of strains that were present after the acid shock step.

The outgrowth duration should be long enough so that the mutants which have survived the acid stress can increase in frequency by sufficient amount so that they can be differentiated from the mutants which have not survived the acid stress. The CFU mL^{-1}, after 0 and 3 hours of outgrowth, of KH001 induced in the presence of ethanol followed by acid shock was 2.40×10^5 and 2.49×10^7, which is just over 6 doublings. With 6 doublings, the CFU

mL^{-1} of mutants containing insertions within any particular gene will increase by 64-fold, assuming that the cells are growing exponentially. An outgrowth duration of 3 hours is therefore sufficient for mutants which have survived the acid stress to grow and increase in frequency. Furthermore, during this time, the relative difference in the fitness of strains which is present after acid shock, does not change.

5.2.1.3 Optimisation of culture volume required for DNA extraction

The previous two optimisation steps were done with 5 mL culture volume grown in a 30 mL universal container. This is because all the survival assays with MG1655 in Chapter 4 were done using this volume and in order to compare the survival results of competition assays with the results of the survival assays in Chapter 4, the same volume was used. However, for TraDIS experiments, a bigger volume of culture needs to be used, so that the starting culture can be inoculated with a sufficient number of transposon mutants. Furthermore, the volume of culture in which the acid stressed cells are outgrown also needs to be increased so that a sufficient number of cells for DNA extraction can be obtained at the end of the outgrowth. The optimisation step was therefore repeated with a starting culture volume of 50 mL grown in a 250 mL shaker flask. The volume of culture acid stressed and outgrown was also increased by 10 times.

When the competition assay was done between KH001 and MG1655 Δ*evgS,* the indole- treated KH001 and MG1655 Δ*evgS* showed a rapid increase in CFU mL^{-1} after an hour of outgrowth (Figure 5.2 B). As mentioned in section 5.2.1.2, *E. coli* has a doubling time of ~30 mins during outgrowth. The CFU mL^{-1} of indole-treated MG1655 Δ*evgS* increased from 2,635 to 170,667 from 0 to 1 h of outgrowth. This amount of increase in CFU mL^{-1} is not theoretically feasible if *E. coli* has a doubling time of ~30 mins. A possible explanation of this result is that

when strains are acid shocked at pH 2.5 and plated out, additional killing happens on the plate because of the acidic media that the cells are transferred in onto the plate. Since the cells are cultured, and therefore plated, in pH 7 medium during outgrowth, this effect of additional killing on the plate does not happen during outgrowth. When the CFU mL^{-1} calculated after acid shock and during outgrowth are plotted on a graph, it looks like as if there is a rapid increase in CFU mL^{-1} after one hour of outgrowth. But this effect is only there because the CFU mL^{-1} calculated after acid shock is lower than the actual CFU mL^{-1}. Another hypothesis is that indole induces the cells into an acid-resistant state where if those cells are plated out, they are unable to grow but if they are added to a liquid medium, they are able to recover out of that non-platable state within an hour.

To test the first hypothesis, MG1655 was induced in the presence of indole, acid shocked at pH 2.5. Cells were either plated out immediately, or they were pelleted and resuspended in M9Ksupp pH 7 medium before being plated out. The cells plated out in pH 2.5 medium had a survival of 1.1% whereas those plated out in pH 7 medium had a survival of 11.3% (Figure 5.3 A). This result shows that plating the cells out in pH 2.5 medium leads to additional killing of cells on the plate. To test the second hypothesis, the culture volume was split equally after acid shock. One half was outgrown in the presence of 1 mM indole and the other without indole. Both cultures were plated out after 0.5, 1, 2 and 3 h and CFU mL^{-1} was calculated. If indole induces the cells into a non-platable state, then cells outgrown with indole would have a lower CFU mL^{-1} than the cells outgrown without indole. Figure 5.3 B shows that both the cultures, the one outgrown with indole and the other without indole, grew at a similar rate suggesting that indole does not induce the cells into a non-platable state.

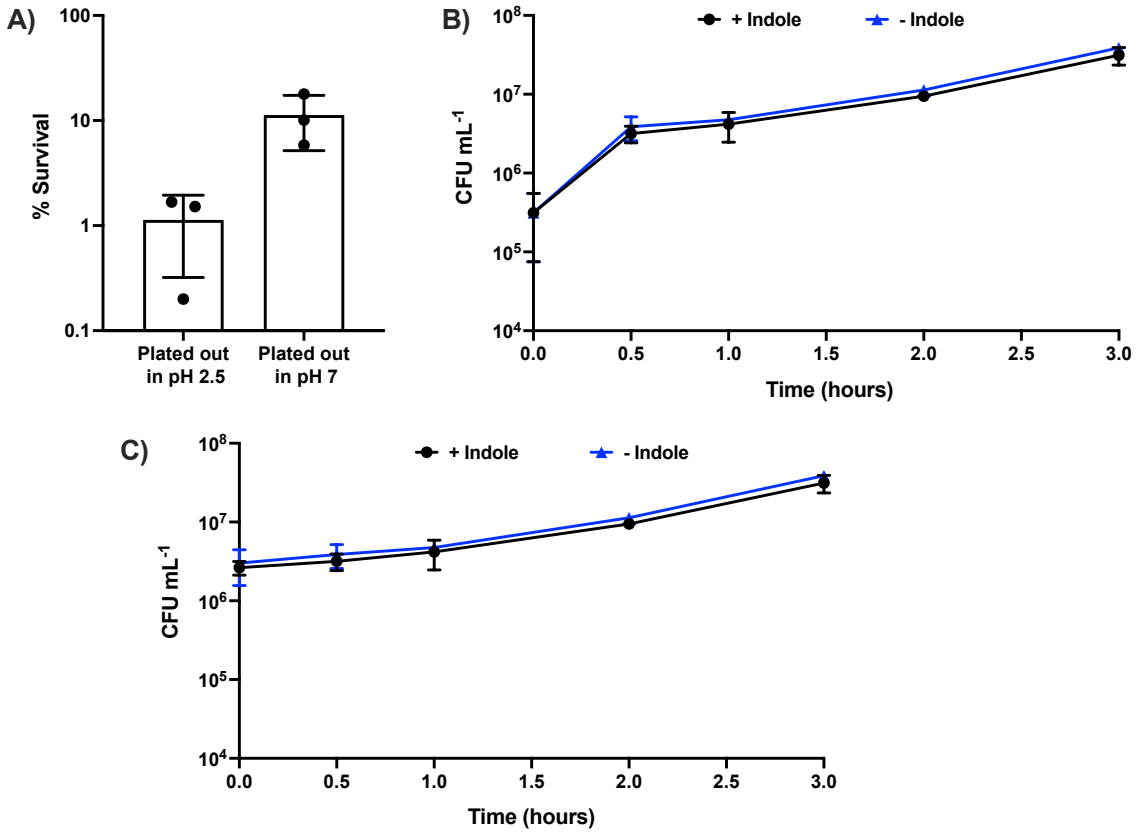


Figure 5.3: MG1655 induced in the presence of 1 mM indole, acid shocked at pH 2.5 and outgrown in the presence and absence of indole for 3 hours.

MG1655 was cultured to mid-exponential phase, induced at pH 5.5 in the presence of indole and acid shocked at pH 2.5 for 30 minutes. The acid stressed culture was plated out either immediately in pH 2.5 medium or after being resuspended in pH 7 medium. After acid shock, the cultures were outgrown for 3 hours in the presence and absence of 1 mM indole. (A) Survival of MG1655 plated out in either pH 2.5 or pH 7 medium, following induction and acid shock. (B-C) Increase in CFU mL^{-1} of MG1655 outgrown in the presence and absence of indole for 3 h after acid shock. CFU mL^{-1} at 0 h was either calculated from cells plated out in pH 2.5 medium (B) or pH 7 medium (C) after acid shock. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

When the CFU mL^{-1} of cells was calculated during outgrowth, the calculation at 0 h was done in two different ways. Calculation in Figure 5.3 B was done from the cells plated out in pH 2.5 medium whereas the calculation in Figure 5.3 C was done from the cells plated out in pH 7 medium. A rapid increase in CFU mL^{-1} during first half an hour did not occur when the calculation was done from the cells plated out in pH 7 medium. These observations suggest that the rapid increase in CFU mL^{-1} during first half an hour of outgrowth was only due to additional killing on plate at 0 h in pH 2.5 medium.

The effect of increasing the culture volume was checked by comparing the survival of MG1655 in this experiment with the survival of MG1655 induced in the presence of indole and acid stressed for 30 minutes in the competition assay in Figure 5.1 A. MG1655 had a survival of 1.1% as opposed to 0.1% in the competition assay. An increased survival using a higher culture volume is acceptable as long as there is a difference in the survival of MG1655 acid stressed following induction in the presence of indole, and ethanol. This was tested with MG1655 transposon library and will be discussed later in section 5.2.2.1.

MG1655 went through over 6 doublings during outgrowth, which is the same as shown previously in section 5.2.1.2. However, if the number of CFU mL^{-1} is recalculated from the cells plated in pH 7 after acid shock, then this reduces the number of doublings to over 3. This means that the CFU mL^{-1} of the mutants, from 0 h to 3 h of the outgrowth, will increase by 8-fold. Although this is a smaller increase than previously thought (64-fold), an 8-fold increase is still sufficient to differentiate between mutants which have survived and amplified versus which did not survive the acid stress.

Finally, whether the number of CFU mL^{-1} obtained at the end of the outgrowth is enough to provide a good DNA yield for library preparation was checked. The DNA extraction protocol recommends that at least 10^9 cells are processed for good DNA yield. After 3 hours

of outgrowth, 3.13 x 10^7 CFU mL^{-1} were present, which is equal to 1.57 x 10^9 CFU in a volume of 50 mL. If the entire volume of 50 mL after outgrowth was processed for DNA extraction, sufficient DNA yield would be obtained.

In conclusion, a higher culture volume of 50 mL can be used for TraDIS experiments. However, it is important to check and compare the survival of library acid stressed following induction in the presence of indole, and ethanol, to ensure that a difference in the survival is present under two conditions. The outgrowth time of 3 hours is sufficient for the acid sensitive mutants to increase in frequency and to obtain sufficient CFU mL^{-1} for getting good DNA yield.

5.2.2 TraDIS experiment results in MG1655

5.2.2.1 Workflow of TraDIS experiments

TraDIS experiments involve growing a strain under a particular condition or stress and then comparing the change in the frequency of the mutants containing insertions within different non-essential genes between the conditional TraDIS library and a control TraDIS library. The control TraDIS library is prepared from the initial transposon library either as it is or by growing it under the same conditions as the conditional TraDIS library but without the stress.

The analysis of the TraDIS sequencing data provides a list of genes which fit into one of the three categories: (1) **Enriched insertions:** Mutants containing insertions within the genes increase in frequency in the conditional TraDIS library relative to the control TraDIS library. This is because the loss of function of proteins encoded by these genes improves the fitness of *E. coli* under a given stress, (2) **Depleted insertions:** Mutants containing insertions within the genes decrease in frequency in the conditional TraDIS library relative to the control

TraDIS library. This is because the function of proteins encoded by these genes is necessary for full fitness of *E. coli* under a given stress, (3) Mutants containing insertions within the genes do not change in frequency in the conditional TraDIS library relative to the control TraDIS library. This is because the function of proteins encoded by these genes does not have a significant effect on the fitness of *E. coli* under a given stress.

The aim of the TraDIS experiments described in this chapter was to (1) identify genes that are important for survival of *E. coli* at pH 2.5 and (2) potentially detect genes that could be involved in the inhibitory effect of indole on the induction of acid stress response. To answer these questions, the MG1655 transposon library was cultured in M9Ksupp pH 7 medium to mid-exponential phase, induced at pH 5.5 either in the presence of indole or ethanol, followed by acid stress at pH 2.5 for 30 mins and finally outgrown for 3 hours in M9K supp pH 7 medium to allow the live cells to grow and amplify in numbers (Figure 5.4 A). The sequencing libraries were prepared from these two conditional libraries, namely 5.5 + I -> 2.5 and 5.5 + E -> 2.5 where I and E refer to indole and ethanol respectively, and the change in the frequency of the mutants containing insertions within different genes in the conditional TraDIS libraries relative to the control TraDIS library was determined (Figure 5.4 B).

In order to determine the genes through which indole inhibits the induced acid stress response, the TraDIS experiment needs to be performed by inducing the cells in the presence of indole, followed by acid shock at pH 2.5 and outgrowth. As a positive control, where the induced acid stress response is not inhibited, the experiment also needs to be performed by inducing the cells in the presence of ethanol, followed by acid shock and outgrowth.

The control TraDIS library, to compare with the conditional TraDIS libraries, was prepared from the initial MG1655 transposon library. Therefore, the control TraDIS library

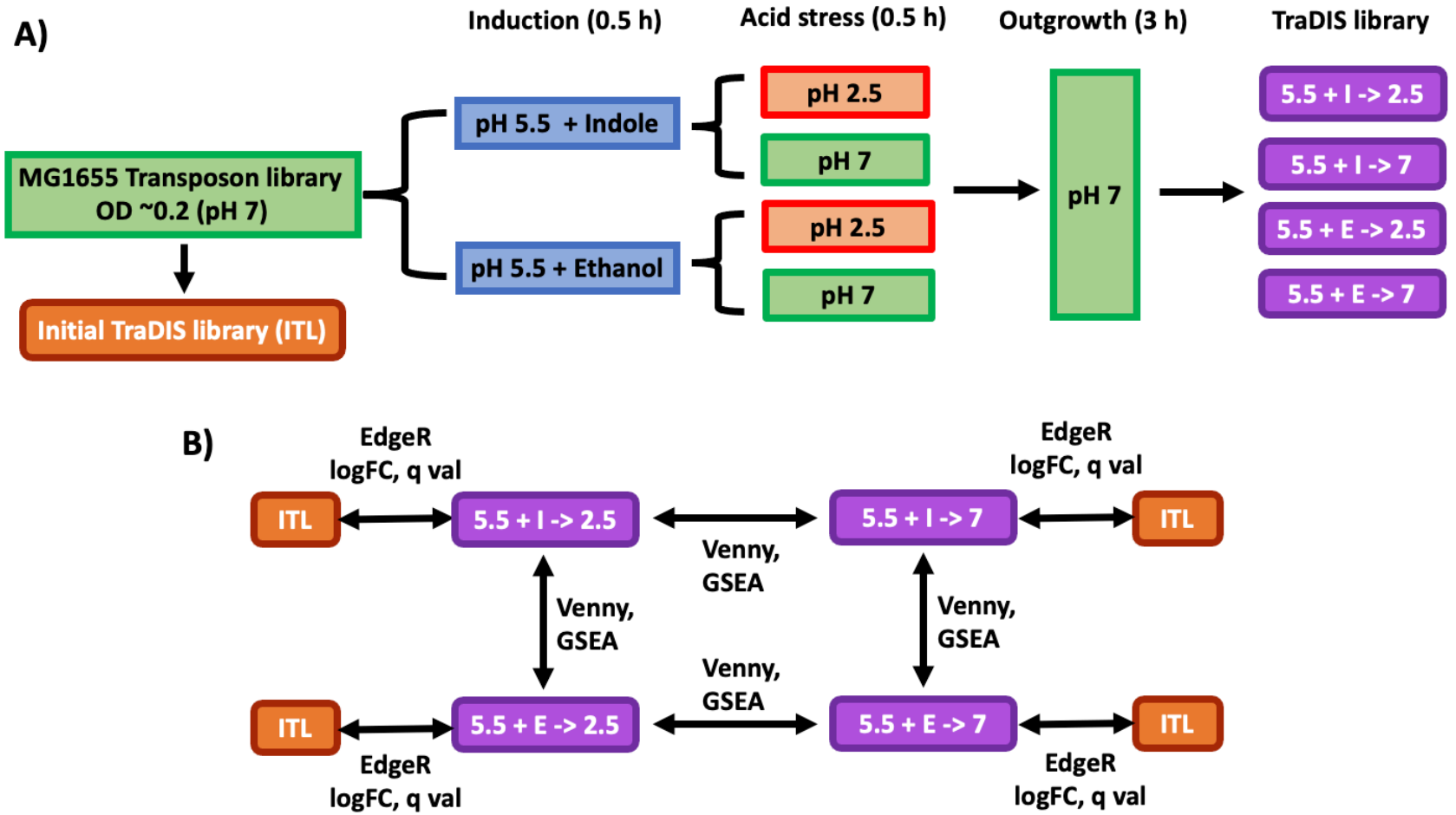


Figure 5.4: Workflow of TraDIS library preparation and comparison between control and contiditional TraDIS libraries.
(A) Conditions under which MG1655 transposon library was grown to study the AR2 acid resistance response and the impact of indole on the induction of this response. (B) The differences in the read counts between initial TraDIS library (ITL) and conditional TraDIS libraries was determined using EdgeR package of the BioTraDIS pipeline. The EdgeR outputs (log fold-change and q value) were used to compare different conditions in Venny and gene set enrichment analysis (GSEA).

will be referred to as the initial TraDIS library (ITL) from here on. Since the MG1655 transposon library was prepared by plating on LB agar at pH 7, any difference in the frequency of the mutants containing insertions in a particular gene, between our conditional TraDIS libraries and the ITL could be due to three reasons. The three possible reasons for the drop in the frequency of the mutants are (1) the function of the protein encoded by the gene in which insertions have dropped is needed for the optimal fitness of *E. coli* in M9Ksupp medium at pH 7, (2) the function of the protein is needed for the optimal fitness of *E. coli* at pH 5.5, (3) the function of the protein is needed for the optimal fitness of *E. coli* at pH 2.5. The TraDIS experiments were repeated by inducing the cells at pH 5.5 either in the presence of indole or ethanol, followed by incubation at pH 7 instead of pH 2.5, and then outgrowth. These libraries were named 5.5 + I -> 7 and 5.5 + E -> 7 respectively. When compared with ITL, the genes required for growth at pH 7, and genes required for survival at pH 5.5, would drop in all four libraries: 5.5 + I -> 2.5, 5.5 + E -> 2.5, 5.5 + I -> 7 and 5.5 + E -> 7. These genes can then be excluded from next steps of analysis.

So now to answer the first question, which is what genes are required for survival of *E. coli* at pH 2.5, the genes which are needed for the optimal fitness of cells at pH 7 and pH 5.5 are first removed as described above. Once the fitness of both 5.5 + I -> 2.5 and 5.5 + E -> 2.5 libraries has been determined relative to the ITL, the two libraries are then compared with each other. The genes in which inserts drop in either one or both of these libraries is determined. The function of these genes would be expected to be required for the optimal fitness of *E. coli* at pH 2.5. Furthermore, the genes in 5.5 + E -> 2.5 library would be the ones which are required for the induction of AR2 response whereas the genes in 5.5 + I -> 2.5 library would not be required for AR2 induction. Conversely, if the inserts within certain genes

increase in either one or both of these libraries, relative to ITL, then the mutations in these genes improve the fitness of *E. coli* at pH 2.5.

To answer the second question, which is what genes are involved in the inhibitory effect of indole on the induction of acid stress response, 5.5 + I -> 2.5 library can be compared with 5.5 + I -> 7 library. The frequency of the mutants containing insertions in genes that increase in either one or both of these libraries is determined. These genes would be expected to be involved in the inhibitory effect of indole on AR2 response.

5.2.2.2 Results of TraDIS experiments

When TraDIS experiments were performed, the survival of MG1655 transposon library exposed to each condition was calculated as a quality control (Figure 5.5). The library had a survival of 9.2% in 5.5 + I -> 2.5, 33.1% in 5.5 + E -> 2.5, 101.4% in 5.5 + I -> 7, and 123.6% in 5.5 + E -> 7 condition. As expected, and as seen previously with the MG1655 strain in Figure 5.1 A, the MG1655 transposon library acid stressed following induction in the presence of indole had a lower survival than when the library was acid stressed following induction in the presence of ethanol. The survival of indole- and ethanol-treated libraries exposed to pH 2.5 was higher than the survival of MG1655 in the competition assay done against KH001 (survival of MG1655 was 0.09% and 13.1% in indole- and ethanol-treated cultures respectively, Figure 5.1 A). The difference in the survival of MG1655 strain and MG1655 transposon library could be due to the MG1655 strain being plated out in pH 2.5 medium and MG1655 transposon library in pH 7 medium. As shown in Figure 5.3 A, the survival of MG1655 strain induced in the presence of indole and acid stressed at pH 2.5, increases from 1.1% to 11.3% if the culture is plated out after being resuspended in pH 7 medium as opposed to pH 2.5 medium. The

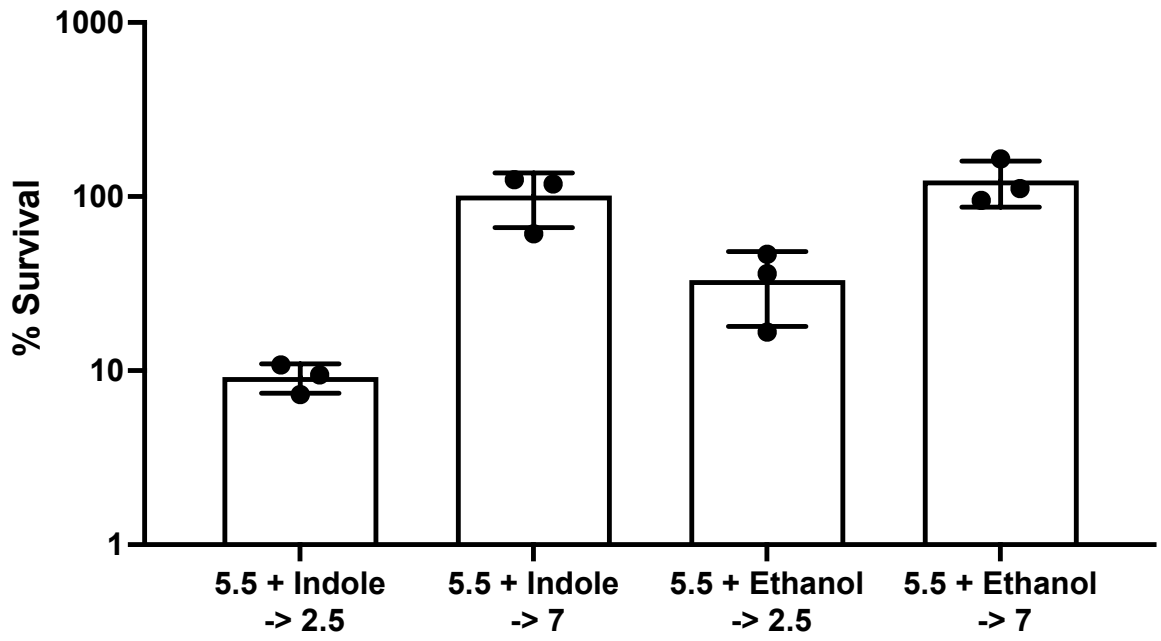


Figure 5.5: Survival of MG1655 transposon library induced either in the presence of 1 mM indole or the same amount of ethanol followed by exposure to pH 2.5 or pH 7.
MG1655 transposon library was cultured to mid-exponential phase, induced at pH 5.5 in the presence of either 1mM indole or the same volume of ethanol, followed by incubation at pH 2.5 or pH 7 for 30 minutes. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

survival of 11.3% is close to survival of 9.2% observed with MG1655 transposon library acid stressed following induction in the presence of indole. These observations suggest that the survival results obtained with MG1655 transposon library are consistent with what has been previously seen with MG1655 strain.

5.2.2.3 Confirmation of reproducibility among replicates

Two individual replicates (ITL1 and ITL2) of Initial MG1655 TraDIS library were sequenced. A total of 3,199,922 and 2,271,889 sequencing reads were aligned to the reference genome (EMBL accession: GCA_000005845.2) for ITL1 and ITL2 respectively. The number of unique insertion sites (UIS) present in ITL1 and ITL2 was 385,440 and 378,423, which averages to 381,931 UIS across the entire genome. With the genome size of 4,641,652 bp, the MG1655 transposon library therefore has a UIS present every 12.2 bps. The number of read counts and UIS present in the TraDIS libraries prepared under four conditions, as compared to the ITL, are shown in Figure 5.6. The number of UIS in the conditional TraDIS libraries is higher than the ITL because the conditional TraDIS libraries have been sequenced at a greater depth (Figure 5.6 B). This is shown by a higher number of read counts corresponding to the conditional TraDIS libraries than the ITL (Figure 5.6 A).

The reproducibility of replicate TraDIS sequencing libraries prepared under each condition was checked using principal component analysis (PCA) by comparing the number of read counts per gene normalised by sequencing depth (Figure 5.7). The replicates from indole- treated libraries, 5.5 + I -> 2.5 and 5.5 + I -> 7, clustered together whereas the first replicate from both ethanol-treated libraries, 5.5 + E -> 2.5 and 5.5 + E -> 7 showed variance as compared to the other two replicates, in the first component of the PCA analysis. It is worth noting that the experiments corresponding to the first replicates of both ethanol-treated

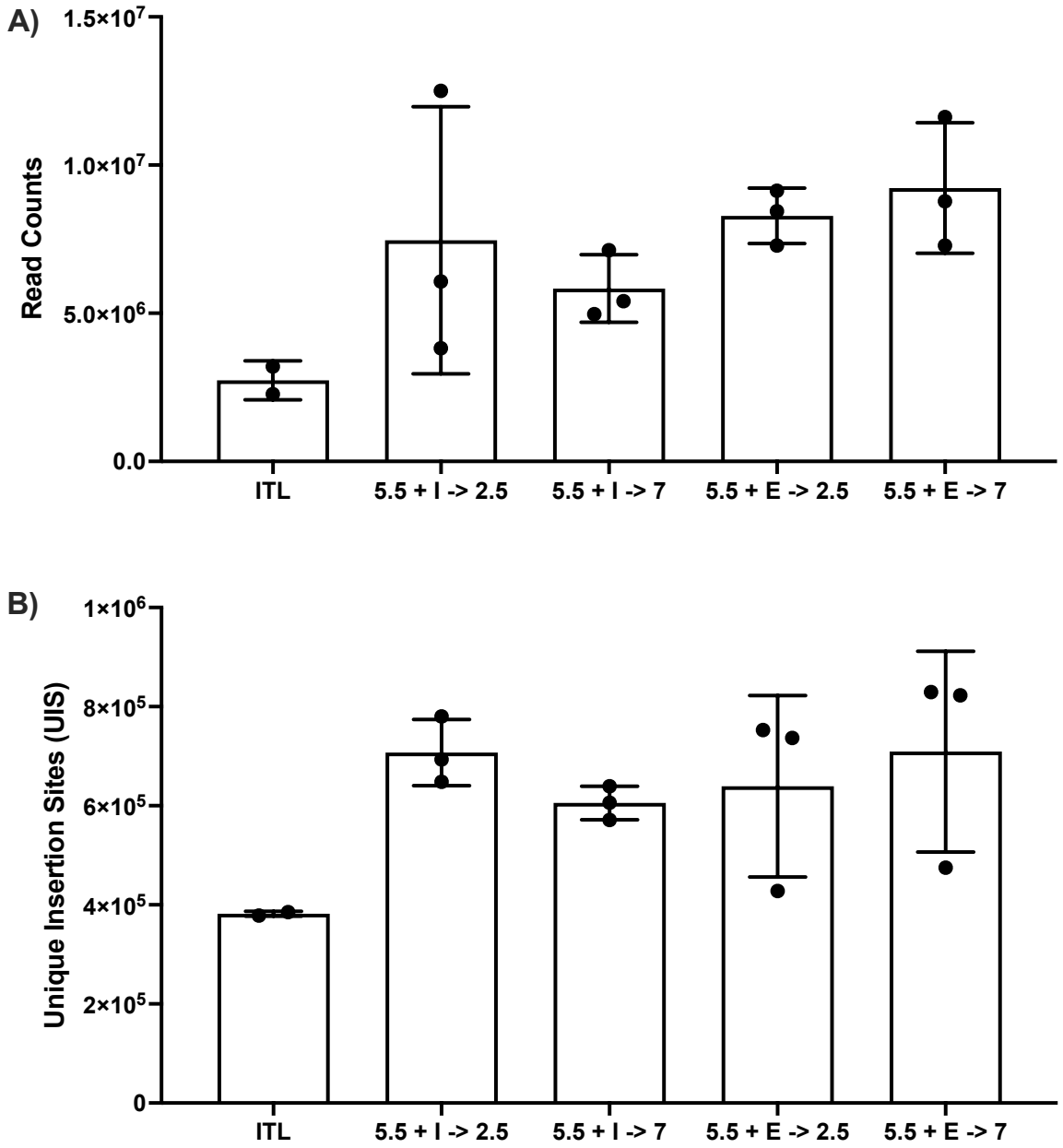


Figure 5.6: Comparison of total number of read counts (A) and unique insertion sites (UIS) (B) among initial TraDIS library (ITL) and the TraDIS libraries prepared under different conditions.

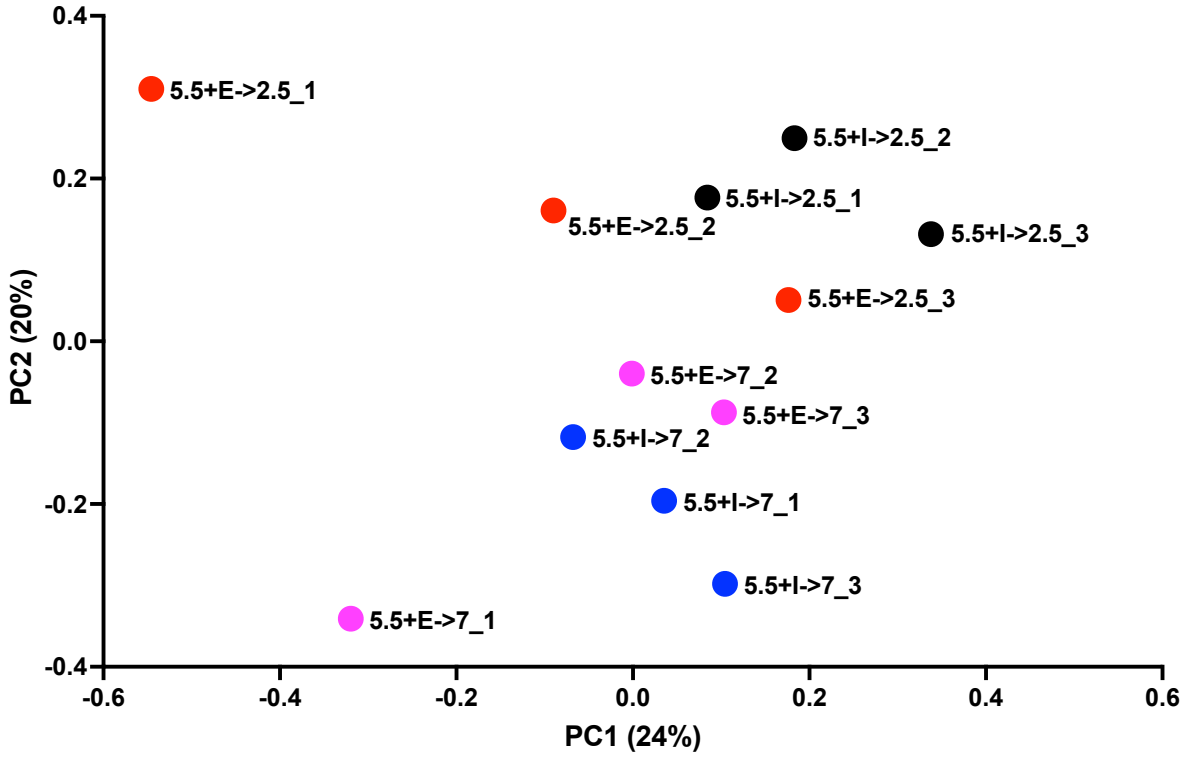


Figure 5.7: Clustering among replicates of same and different conditions visualised through principal component analysis done on read counts per gene normalised by sequencing depth.

The percentage variance explained by first (24%) and second (20%) principal components is shown on the X- and Y-axis respectively.

libraries were performed on the same day. Therefore, they could be true outliers or the variation could be a result of human error. Libraries from all four conditions clustered together but there was some separation between libraries acid stressed at pH 2.5 from those incubated at pH 7. Furthermore, ethanol-treated libraries clustered closer together than the indole-treated libraries.

The reproducibility among replicates was further checked using reads per million mapped reads (RPM) which is the number of reads per gene normalised by sequencing depth, and insertion index score which is the number of UIS per gene normalised by gene length. One plot for RPM and one for insertion index score from ITL and each conditional TraDIS library is shown in Figures 5.8 and 5.9 respectively. The plots for remaining replicates are shown in Appendix 8.2.2. The Spearman's correlation coefficient for both RPM and insertion index score was greater than 0.96 for all plots showing a high reproducibility among replicates of all conditions. Therefore, all three replicates from each condition were included in the EdgeR comparisons.

5.2.2.4 Visualisation of EdgeR comparisons between ITL and conditional TraDIS libraries using volcano plots

The outputs from EdgeR comparisons (log_2FC, logCPM and q value) between ITL and each conditional TraDIS library (5.5 + I -> 2.5, 5.5 + E -> 2.5, 5.5 + I -> I, 5.5 + E -> 7) were plotted on a separate volcano plot (Figure 5.10). An arbitrary cut-off of log_2FC (fold change) > 2 or < -2, with a q value < 0.05, was applied and these genes are shown in red. Another metric, logCPM (read counts per million), which is used as a proxy for the expression level of the genes in RNA-seq data, was used here as a proxy for the number of mutants containing transposon insertions within a gene. An arbitrary cut-off of logCPM > 8 was used and these

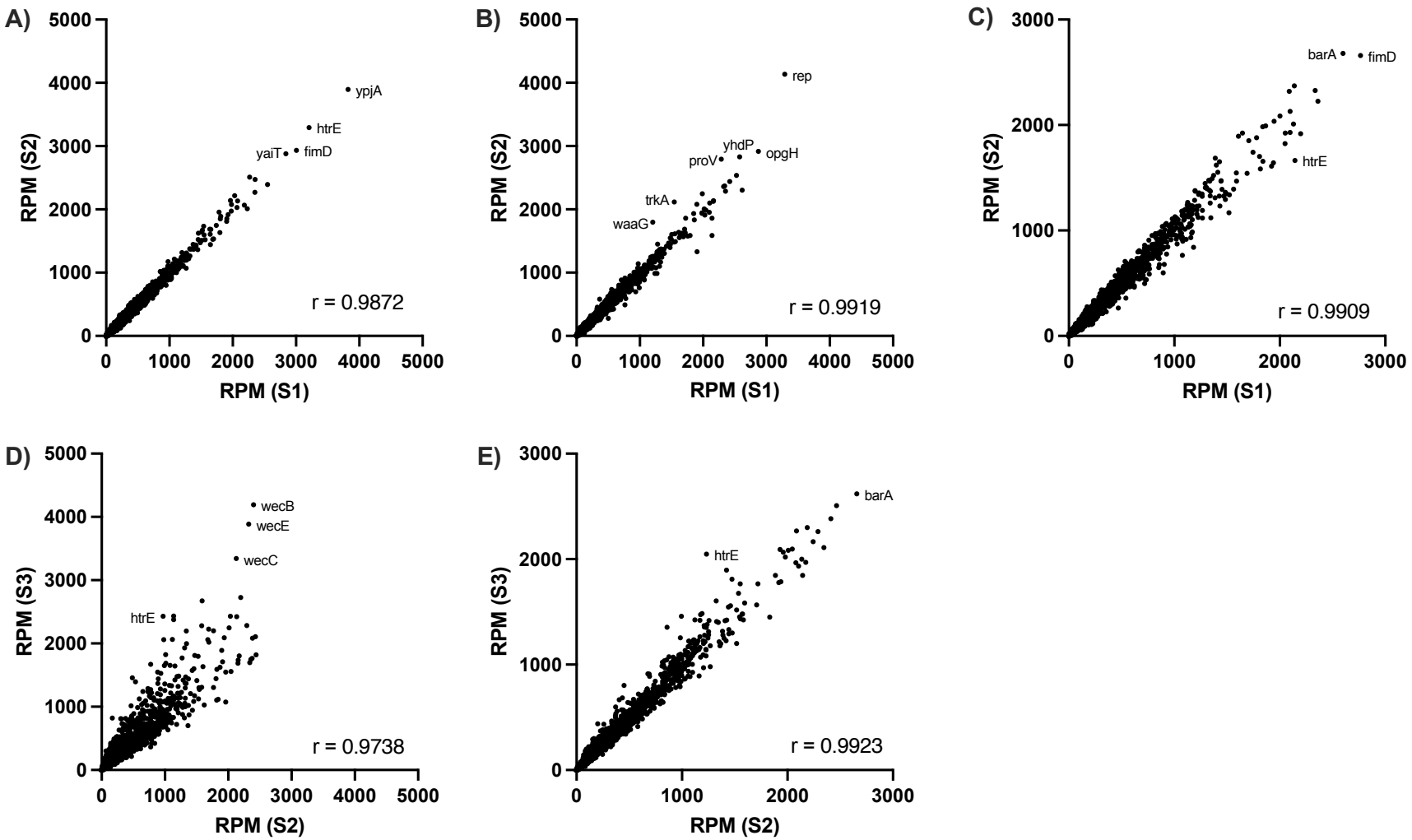


Figure 5.8: Comparison between RPM of two TraDIS libraries prepared from two separate biological replicates.
The conditions shown are as follows: A) ITL, B) 5.5 + I -> 2.5, C) 5.5 + I -> 7, D) 5.5 + E -> 2.5, E) 5.5 + E -> 7. The correlation was determined using Spearman's correlation coefficient shown at the bottom right corner.

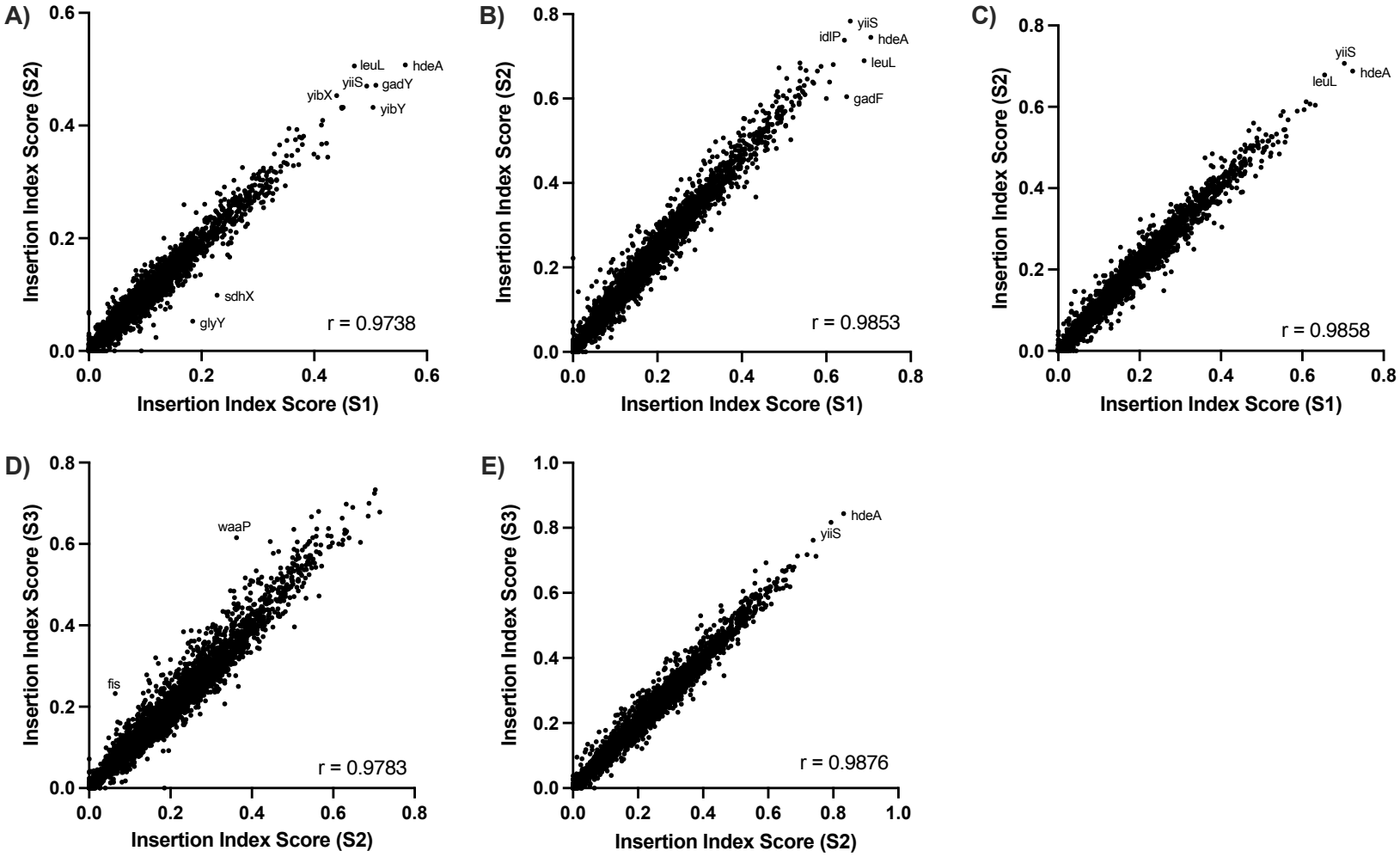
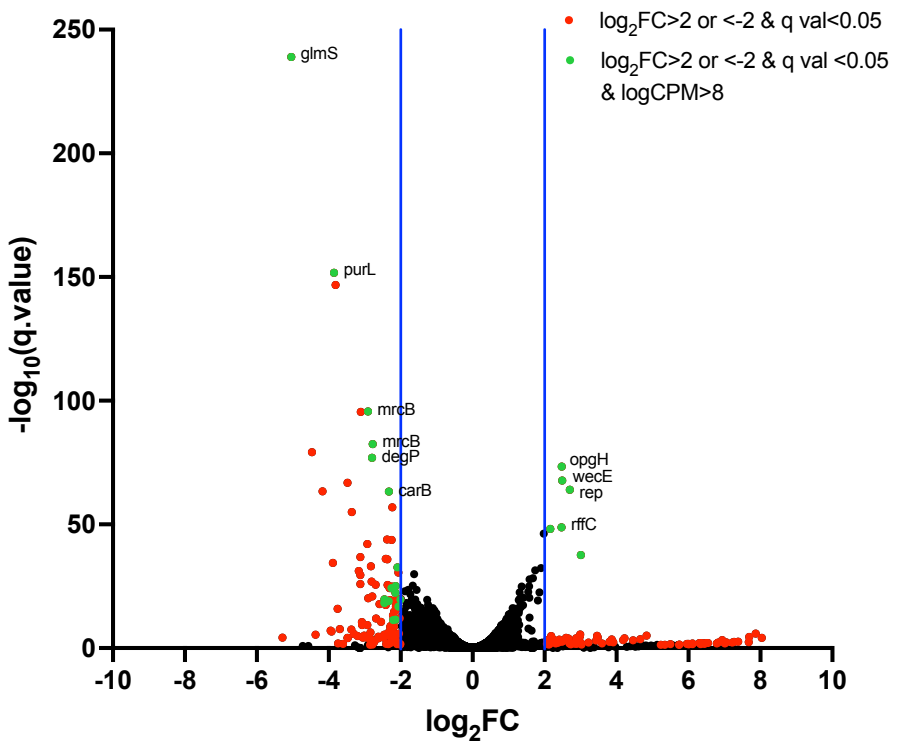
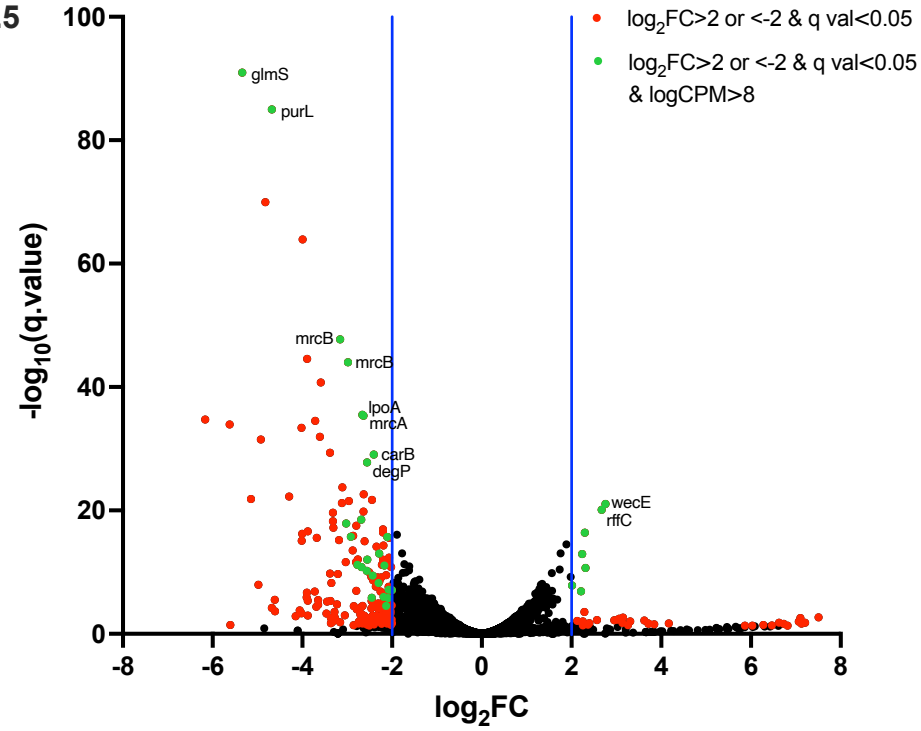


Figure 5.9: Comparison between insertion index scores of two TraDIS libraries prepared from two separate biological replicates.
The conditions shown are as follows: A) ITL, B) 5.5 + I -> 2.5, C) 5.5 + I -> 7, D) 5.5 + E -> 2.5, E) 5.5 + E -> 7. The correlation was determined using Spearman's correlation coefficient shown at the bottom right corner.

A) 5.5 + I -> 2.5



B) 5.5 + E -> 2.5



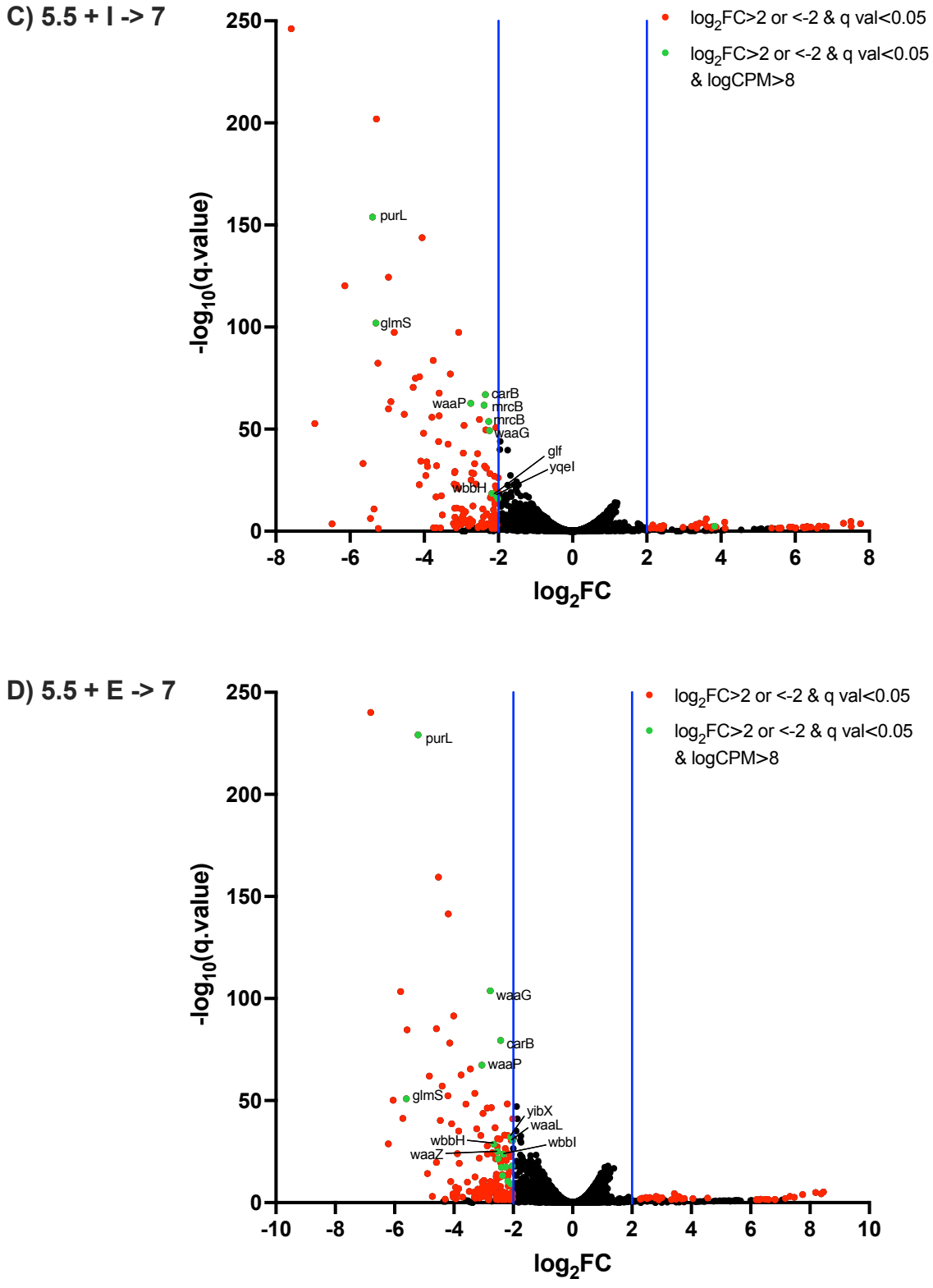


Figure 5.10: Volcano plots showing the log_2 fold-change and the statistical significance corrected for multiple comparisons (q value) of the mutants containing transposon insertions within different genes.

The comparisons between ITL and conditional TraDIS libraries: A) 5.5 + I -> 2.5, B) 5.5 + E -> 2.5, C) 5.5 + I -> 7, D) 5.5 + E -> 7, are shown. Top 10 genes with $log_2FC > 2$ or < -2, q value < 0.05, and logCPM > 8 in each plot are labelled.

genes are shown in green. Top 10 genes with $log_2FC > 2$ or < -2, q value < 0.05, and logCPM > 8 are labelled.

In all 4 volcano plots, the number of transposon mutants whose frequency decreased in the conditional TraDIS libraries relative to the ITL was higher than the mutants whose frequency increased in the conditional TraDIS libraries relative to the ITL. Additionally, a higher number of mutants with reduced frequency in the conditional TraDIS libraries relative to the ITL, had more significant q values than the mutants whose frequency increased in the conditional TraDIS libraries relative to the ITL.

The reason behind these two observations could be because a lot of genes containing transposon insertions are needed for the full fitness of cells in conditions under which TraDIS experiments were done. If majority of the genes, where the number of mutants containing insertions within them are decreasing in frequency, were required for the cells to survive at pH 2.5, then the frequency of such mutants would only decrease under 5.5 + I -> 2.5 and 5.5 + E -> 2.5 conditions, as opposed to all 4 conditions.

A higher number of mutants with increased frequency in the conditional TraDIS libraries relative to the ITL, had higher log_2FC values than the mutants whose frequency decreased in the conditional TraDIS libraries relative to the ITL. A lot of these genes are likely to be essential genes which generally would not have any transposon insertions in them. However, if a small number of mutants containing insertions within these genes were present and the frequency of these mutants changed slightly, then these changes are presented as high log_2FC by EdgeR as an artefact. An example where the majority of mutants with increased frequency relative to the ITL were found to contain insertions within essential genes is given in section 5.2.2.6.

A high number of genes with q value < 0.05 had logCPM < 8. LogCPM value of less than 8 for a gene means that the number of mutants containing insertions within this gene was lower than the number which is represented by the cut-off value of 8. If a cut-off of 8 for logCPM value was included, majority of the genes would be excluded from the analysis. Due to this reason, the cut-off for logCPM was not included in the data analysis in coming sections.

As can be seen in Figure 5.10, strains with insertions in some genes were reproducibly and significantly reduced under all conditions. This was true for *purL, glmS, mrcB* and *carB*. This result is as expected and acts as a control for the effectiveness of TraDIS in detecting fitness effects. This is because there is a high possibility that the function of these genes is needed for the optimal fitness of *E. coli* during growth in M9Ksupp medium at pH 7. M9K minimal media does not contain purines (and pyrimidines) and *pur* genes, including *purL*, are involved in the synthesis of purines (reviewed in Zhang, Morar and Ealick, 2008). Furthermore, the inserts in remaining *pur* genes including *purF, purM, purC,* and *purD* were also highly reduced ($log_2FC < -2$) in all conditions (data in Appendix 8.2.1).

mrcB encodes penicillin binding protein 1B (PBP1B) which is required for the synthesis of cell wall peptidoglycan (PG) (reviewed in Sauvage *et al.*, 2008). In Figure 5.10, two different *mrcB* genes are present which correspond to two different protein isoforms encoded by this gene. Due to its role in PG synthesis, one of the main cell components, it is highly likely that the function of *mrcB* is needed for the optimal fitness of cells at pH 7.

Further analysis on genes with $log_2FC > 2$ or < -2, and q value < 0.05 in each library will be discussed later in more detail in section 5.2.2.6, where log_2FC and q values of genes from all four libraries will be compared. In conclusion, the results discussed here suggest that the TraDIS experiments have been successful. This is because the mutants containing insertions

in *purL* and *mrcB* showed a drop in frequency in the conditional libraries where these genes are expected to have an important function.

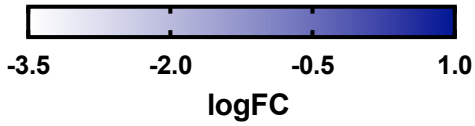
5.2.2.5 Comparison of fold change of mutants containing a transposon within AR2 genes across four conditions

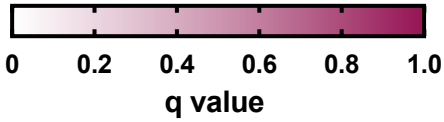
When the experimental conditions for TraDIS experiments were optimised as described in Section 5.2.1, the competition experiments described in this section confirmed that strains lacking EvgS were less fit than the wild type strains following incubation at pH 5.5 and then pH 2.5, as expected from the known role of EvgS in activating the AR2 system. This thus provided a useful way of checking the success of the TraDIS experiments, as it can be predicted that strains carrying insertion mutations in *evgS* (and *evgA*) will decline significantly relative to their frequencies in the ITL under these conditions.

Table 5.1 shows that the strains carrying transposon insertions within *evgS* and *evgA* showed the largest decline in fitness (the most negative $\log_2$FC) among the AR2 genes at pH 2.5 in the presence of ethanol. Furthermore, the $\log_2$FC for both genes were highly significant. Overall, this is consistent with expectations and suggests that the TraDIS experiments were successfully able to identify genes which are important under the conditions of the experiment. However, unexpectedly the *evgS* and *evgA* transposon mutants also showed a significant decline in fitness under all other conditions tested. This decline could be because these genes contribute to the fitness of *E. coli* either at pH 7, pH 5.5, or both. EvgS gets activated at pH between 5.5 - 5.7 but its function is required to provide resistance against low pH of 2.5 and below, which the strains in pH 7 conditions were not exposed to. Furthermore, the role of *evgA* and *evgS* at pH 7 has not been determined by any study yet. Whether these genes are required for full fitness of *E. coli* at pH 7 and/or pH 5.5, as is implied by these results,

Table 5.1: Comparison of log_2FC and q values of mutants containing transposon within the AR2 genes in four conditions under which TraDIS experiments were performed. The legend bars at the bottom show the colour corresponding to the log_2FC and q values. The data has been ranked from the smallest to the largest log_2FC in 5.5 + E -> 2.5 condition.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
evgA	-1.66	-3.31	-1.32	-1.62	8.46×10^{-14}	6.34×10^{-18}	9.96×10^{-8}	4.73×10^{-8}
evgS	-1.38	-2.92	-1.15	-1.31	2.46×10^{-13}	1.82×10^{-16}	4.27×10^{-9}	1.57×10^{-7}
gadE	-1.73	-2.20	-0.89	-1.20	1.79×10^{-22}	1.11×10^{-17}	1.82×10^{-4}	1.91×10^{-12}
hdeA	-1.56	-2.11	-0.88	-1.16	2.94×10^{-24}	1.93×10^{-16}	2.35×10^{-6}	9.61×10^{-15}
rcsB	-1.56	-1.83	-0.28	-0.25	7.83×10^{-11}	4.47×10^{-7}	0.30	0.28
hdeB	-1.46	-1.70	-0.91	-1.05	3.50×10^{-15}	1.13×10^{-10}	2.67×10^{-5}	5.86×10^{-8}
ydeO	-1.39	-1.66	-1.27	-1.94	1.03×10^{-3}	3.88×10^{-3}	0.02	1.46×10^{-4}
gadB	-1.00	-1.47	-0.98	-1.37	8.71×10^{-5}	6.28×10^{-4}	0.01	7.59×10^{-6}
ydeP	-1.09	-1.39	-0.81	-0.85	2.31×10^{-12}	1.58×10^{-9}	2.82×10^{-8}	1.11×10^{-7}
dctR	-1.24	-1.18	-0.78	-1.08	2.03×10^{-14}	5.21×10^{-5}	8.69×10^{-6}	1.85×10^{-13}
slp	-1.08	-1.15	-0.69	-0.97	2.02×10^{-11}	1.60×10^{-4}	7.15×10^{-5}	9.39×10^{-11}
hdeD	-0.92	-1.14	-0.75	-0.97	4.82×10^{-7}	9.16×10^{-5}	1.38×10^{-5}	1.18×10^{-12}
yhiD	-0.86	-1.06	-0.54	-0.73	6.96×10^{-8}	1.16×10^{-5}	0.01	5.69×10^{-7}
gadX	-0.97	-0.93	-0.46	-0.77	6.41×10^{-9}	1.34×10^{-3}	0.01	3.91×10^{-9}
lon	0.90	-0.75	-1.43	-1.21	1.05×10^{-5}	0.06	6.91×10^{-7}	3.21×10^{-6}
gadC	-0.61	-0.68	-0.59	-0.49	3.87×10^{-4}	0.02	1.34×10^{-4}	0.01
gadA	-0.48	-0.65	-0.32	-0.66	2.70×10^{-3}	0.03	0.06	1.73×10^{-6}
gadW	-0.50	-0.63	-0.24	-0.64	1.87×10^{-3}	0.03	0.19	3.71×10^{-7}
rssB	0.59	-0.44	-0.53	-0.58	1.11×10^{-3}	0.25	0.02	3.87×10^{-3}
rpoS	-1.60	-0.32	0.15	-0.14	4.90×10^{-18}	0.27	0.48	0.52
torR	0.01	-0.05	0.03	-0.03	0.98	0.88	0.92	0.93
phoP	-0.29	0.02	-0.45	-0.45	0.12	0.97	0.02	0.02
mgtA	0.14	0.21	0.16	0.10	0.38	0.45	0.40	0.54
phoQ	-0.12	0.35	-0.10	-0.02	0.51	0.20	0.61	0.93
torS	0.34	0.48	0.34	0.46	0.03	0.08	0.05	0.01





needs to be tested by competing the *evgS* and *evgA* mutants with the isogenic wild-type strain under these conditions.

Having examined the results for *evgA* and *evgS*, the results with other AR2 genes were also looked at. The frequency of mutants containing insertions in *gadE* as well as several other AR2 genes such as *hdeA, hdeB, ydeO, gadB, ydeP, dctR, slp, hdeD* also declined at pH 7. In contrast, the frequency of mutants containing insertions in *rcsB* did not decline much at pH 7. RcsB, along with GadE, are the central regulators of AR2 system. RcsB dimerises with GadE, and possibly with EvgA, to control the transcription of AR2 genes (Castanié-Cornet *et al.*, 2010; Burton *et al.*, 2010; Johnson *et al.*, 2011). A minute and insignificant decline in the frequency of *rcsB* mutants at pH 7 suggests that the decline observed in the frequency of *evgA, evgS*, and mutants containing insertions in other AR2 genes at pH 7 is not due to the function of these genes which is associated with survival at extremely low pH. Otherwise, the same decline in the frequency of *rcsB* mutants would also be observed. Whatever the function of EvgS is in pH 7 conditions, it seems that indole does not inhibit it because the frequency of mutants containing insertions in *evgS* reduced almost to the same extent in the presence of both indole and ethanol at pH 7. If indole inhibited the function of EvgS at pH 7, then mutants containing insertions in *evgS* would not have a fitness disadvantage since indole would disable the function of EvgS in all other strains which did not contain insertions in *evgS*. Consequently, the $\log_2$FC of *evgS* at pH 7 in the presence of indole and ethanol would not be similar.

The decline in the fitness of *evgS* and *evgA* transposon mutants was bigger at pH 2.5 in the presence of ethanol than in the presence of indole. Furthermore, the decline in the fitness of *evgS* and *evgA* transposon mutants at pH 2.5 in the presence of indole was almost to the same extent as the decline in the fitness at pH 7. This suggests that the function of EvgS, which caused the fitness of strains containing insertions in *evgA* and *evgS* to decline in

pH 7 conditions, is also important in pH 2.5 conditions. As stated before, this function of EvgS might contribute to the fitness of *E. coli* either at pH 7, pH 5.5, or both. Furthermore, this EvgS function is not inhibited by indole. However, a bigger decline in the fitness of *evgA* and *evgS* mutants at pH 2.5 in the presence of ethanol than indole suggests that indole interferes with the functioning of EvgS which is essential for the survival of cells at pH 2.5.

If the hypothesis that indole inhibits the function of EvgS required for the survival of cells at pH 2.5 was true, then indole would also inhibit the effect of all the AR2 genes since they are regulated by EvgS. This would mean that the fitness of strains carrying mutations in the AR2 genes at pH 2.5 in the presence of ethanol would decline more than the fitness of same strains at pH 2.5 in the presence of indole. This is because the impact on fitness of mutations in the AR2 genes goes away in the presence of indole, as all strains are equally affected, not just those with mutations in the AR2 genes.

There was a clear trend where several of the AR2 genes (*ydeP, slp*, *gadE, hdeA, hdeB, yhiD, hdeD, gadB,* and *gadC*) had a more negative $\log_2$FC in the presence of ethanol than indole, although the magnitude of difference in many cases was not big. Additionally, there was a bigger and significant decline in the fitness of strains containing insertions in some of the AR2 genes at pH 2.5 in the presence of indole, than at pH 7. These genes included *gadE, hdeA, hdeB, yhiD, gadX* and *gadC*. The strains with inserts in both *gadE* and *rcsB* also showed a significant decline in the fitness at pH 2.5 in the presence of indole. Lon protease is known to degrade GadE (Heuveling, Possling and Hengge, 2008). The strains carrying inserts within *lon* showed an increase in fitness at pH 2.5 in the presence of indole, but a decline in fitness at pH 2.5 in the presence of ethanol.

The TraDIS results for *gadE, rcsB* and *lon* suggest that the role of *gadE* and *rcsB* as AR2 regulators is crucial at pH 2.5 even in the presence of indole. So far its known that *gadE* gets

activated by EvgA, and the components present downstream of EvgA such as YdeO, GadX, and GadE itself (Ma, Masuda and Foster, 2004; Sayed, Odom and Foster, 2007). *gadE* can also get activated by RpoS sigma factor. Microarray analysis done on BW25113 cultured in glucose supplemented M9 minimal media at OD_{600} of 1 has shown that *rpoS* is required for *gadE* expression (Chattopadhyay *et al.*, 2015). Table 5.1 shows that the strains carrying inserts within *rpoS* show a big decline in fitness at pH 2.5 in the presence of indole. However, these strains do not show any decline in fitness at pH 2.5 in the presence of ethanol, or at pH 7. So it is possible that in the presence of indole, *gadE* is not activated by EvgA, but by RpoS, making the role of RpoS more important than usual.

Under neutral pH, RpoS levels remain low during exponential phase and increase as cells enter stationary phase. RssB, one of the factors that keep RpoS levels low during exponential phase, binds to RpoS, leading to its degradation by ClpXP protease (Becker, Klauck and Hengge-Aronis, 1999). The increased importance of RpoS at pH 2.5 in the presence of indole is further supported by the fact that the strains carrying inserts within *rssB* showed an increase in fitness at pH 2.5 in the presence of indole but a decline in fitness at pH 2.5 in the presence of ethanol. Although the change in the frequency of *rssB* mutants at pH 2.5 in the presence of both indole and ethanol was small, but the levels of RpoS in *rssB* mutants in the presence of indole would be higher than the wild-type cells which supports the hypothesis that the role of RpoS is more important for the functioning of AR2 in the presence of indole.

To check whether the genes whose role in acid stress has previously been studied showed any fitness advantage or disadvantage under the conditions used in our TraDIS experiments, the $\log_2$FC of strains containing inserts within 119 different genes across four TraDIS experimental conditions were compared. These are all the genes which were mentioned in the following three reviews on acid stress in *E. coli*: Kanjee and Houry (2013),

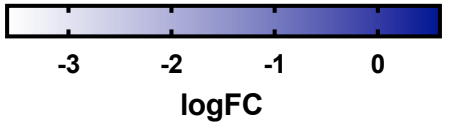
Lund, Tramonti and De Biase (2014), Li *et al.* (2020), and Schwarz *et al.* (2022). The list of these genes is provided in Appendix 8.2.5. The strains containing inserts within genes involved in AR3 (*adi* genes), AR4 (*cad* genes) and AR5 (*speF* and *potE*) systems did not show a significant decline in the fitness in any of the four conditions tested, suggesting that none of these systems were involved in the response being studied here (Table 5.2). AR1, in comparison to the other four AR systems, is poorly characterised. However, CRP and RpoS are known to be its main regulators (Lin *et al.*, 1995; Marie-Pierre *et al.*, 1999). Strains with inserts in *crp* did not show a significant decline in fitness under any of the four conditions. As discussed above, strains carrying inserts in RpoS showed a big decline in fitness at pH 2.5, but only in the presence of indole (Table 5.1). These observations along with the fact that AR1 is repressed by glucose, the main carbon source used in our medium, suggest that AR1 is not involved in the response being studied in our TraDIS experiments.

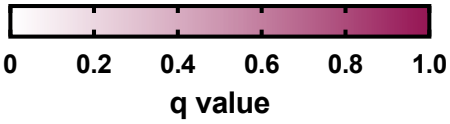
Strains containing inserts within *degP* showed a bigger decline in fitness at pH 2.5 than pH 7 (Table 5.2). DegP is a periplasmic serine endoprotease which degrades abnormal proteins in periplasm. It has been shown to increase the viability of *E. coli* at pH 2.3 (Fu *et al.*, 2018). So the results of our TraDIS experiments agree with the previous findings.

Another gene which has been shown to have an acid sensitive phenotype is *surA*. SurA is a periplasmic chaperone which maintains protein quality in the periplasm. SurA is protected by HdeA during extreme acid stress. Upon neutralisation, SurA assists HdeA in recovering other client proteins from damage endured during acid stress (Zhang *et al.*, 2011). In our TraDIS results, strains containing inserts within *surA* showed a smaller decline in fitness at pH 2.5 in the presence of indole than in the presence of ethanol, and at pH 7 (Table 5.2). Since SurA does not play an active role itself at low pH, it is understandable that the strains containing inserts in *surA* show a similar level of decline in fitness at pH 7 and at pH 2.5 in the

Table 5.2: Comparison of log_2FC and q values of mutants containing transposon within the genes which have been tested for their role in acid stress, in four conditions under which TraDIS experiments were performed. The legend bars at the bottom show the colour corresponding to the log_2FC and q values.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
adiA	0.24	0.42	0.20	0.23	0.13	0.12	0.26	0.11
adiC	0.21	0.34	0.13	0.18	0.19	0.20	0.48	0.23
adiY	-0.40	-0.59	-0.33	-0.49	0.02	0.05	0.05	1.35×10^{-4}
cadA	-0.12	-0.02	-0.02	-0.05	0.42	0.96	0.92	0.78
cadB	-0.34	-0.29	-0.18	-0.28	0.02	0.27	0.30	0.04
cadC	-0.46	-0.65	-0.38	-0.61	4.61×10^{-3}	0.03	0.03	5.04×10^{-6}
crp	0.45	0.30	-0.28	0.12	0.41	0.72	0.74	0.90
degP	-2.80	-2.56	-1.75	-1.86	1.25×10^{-77}	1.59×10^{-28}	1.85×10^{-40}	8.38×10^{-42}
potE	-0.31	-0.33	-0.35	-0.40	0.05	0.23	0.03	0.01
speF	-0.22	-0.22	-0.16	-0.24	0.13	0.42	0.36	0.11
surA	-0.68	-1.78	-1.75	-1.76	2.24×10^{-4}	9.24×10^{-14}	2.41×10^{-23}	3.82×10^{-24}
waaC	-1.65	-2.46	-3.60	-3.44	1.45×10^{-11}	5.83×10^{-10}	2.44×10^{-68}	3.53×10^{-66}





presence of ethanol. In the presence of indole at pH 2.5, however, due to inhibition of AR2 response by indole, cells must be producing lower levels of HdeA than in the presence of ethanol. Due to this, the cells are not able to protect SurA from acid stress. Since in the presence of indole, the strains containing inserts in *surA* are not at a bigger fitness disadvantage than the strains not containing inserts in *surA, surA* mutants do not decline in frequency to the same extent as the *surA* mutants in the presence of ethanol.

Strains containing inserts within *waaC,* which encodes LPS heptosyltransferase, an enzyme involved in LPS biosynthesis, showed a bigger decline in fitness at pH 7 than pH 2.5. *waaC* transposon mutants have been shown to be more resistant than the wild-type strain at extreme pH of 2.5 (Vivijs, Aertsen and Michiels, 2016). This might be the reason for a smaller decline in the fitness of *waaC* mutants at pH 2.5 than pH 7, in our TraDIS experiments, because *waaC* mutants are highly resistant at pH 2.5.

5.2.2.6 Comparison of fold change in genes with depleted and enriched transposon insertions across four conditions

In the previous section, the analysis of TraDIS data focussed on genes which are known from the literature to have an important role in different aspects of *E. coli*'s ability to survive at low pH. This analysis confirmed that, as expected, many of these genes also showed a strong signal in terms of the size and significance of their $\log_2$FC in the TraDIS data. Although some of these results were interpretable in the light of prior knowledge, others have suggested novel hypotheses about the functions of some of these genes. However, the main feature of TraDIS data is that it allows the identification of potential gene function without any prior evidence being required. In this section, therefore, the TraDIS data is analysed in terms of gene lists ranked according to the magnitude of the effect on fitness (positive or

negative) of insertions in each gene under each of the conditions (the log fold change), considered together with the significance of the change (the q value). This analysis will not include genes which were already considered in the previous section, although many of these were of course highly significant.

To identify potential genes of interest, the outputs of edgeR comparisons between the ITL and each of the four experimental conditions were inspected and lists of significant genes were generated, identifying genes as significant if they showed a $\log_2$FC > 2 or < -2, and a q value of < 0.05. Two Venn diagrams were constructed (one for $\log_2$FC < - 2 and one for $\log_2$FC > 2; Figures 5.11 A and B respectively) to show how many genes were in each of the four different categories, and the overlaps between them. It must be stressed at this point that this is a relatively crude analysis with a cut-off point which is somewhat arbitrary, and a potentially more powerful analytical method (Gene set enrichment analysis) will be discussed in section 5.2.2.7. Gene lists analysed in this way may include both false positives (genes which do not have a significant effect on fitness but which vary by random chance) and false negatives (genes where there is a genuine effect on fitness, but it is below the threshold used here. Note that many of the genes discussed in the previous section actually fall into this latter category).

Some general points can be made from these Venn diagrams, before looking at individual genes and pathways. First, the numbers of genes which overlap in all four categories are quite large: >25% of the genes in 5.11 A and >16% of the genes in 5.12 B are found in all four conditions. Because of time constraints, TraDIS experiments were all done with a pH 5.5 pre-incubation step, but the experimental approach used here would also detect genes with a strong positive or negative effect on fitness when mutated irrespective

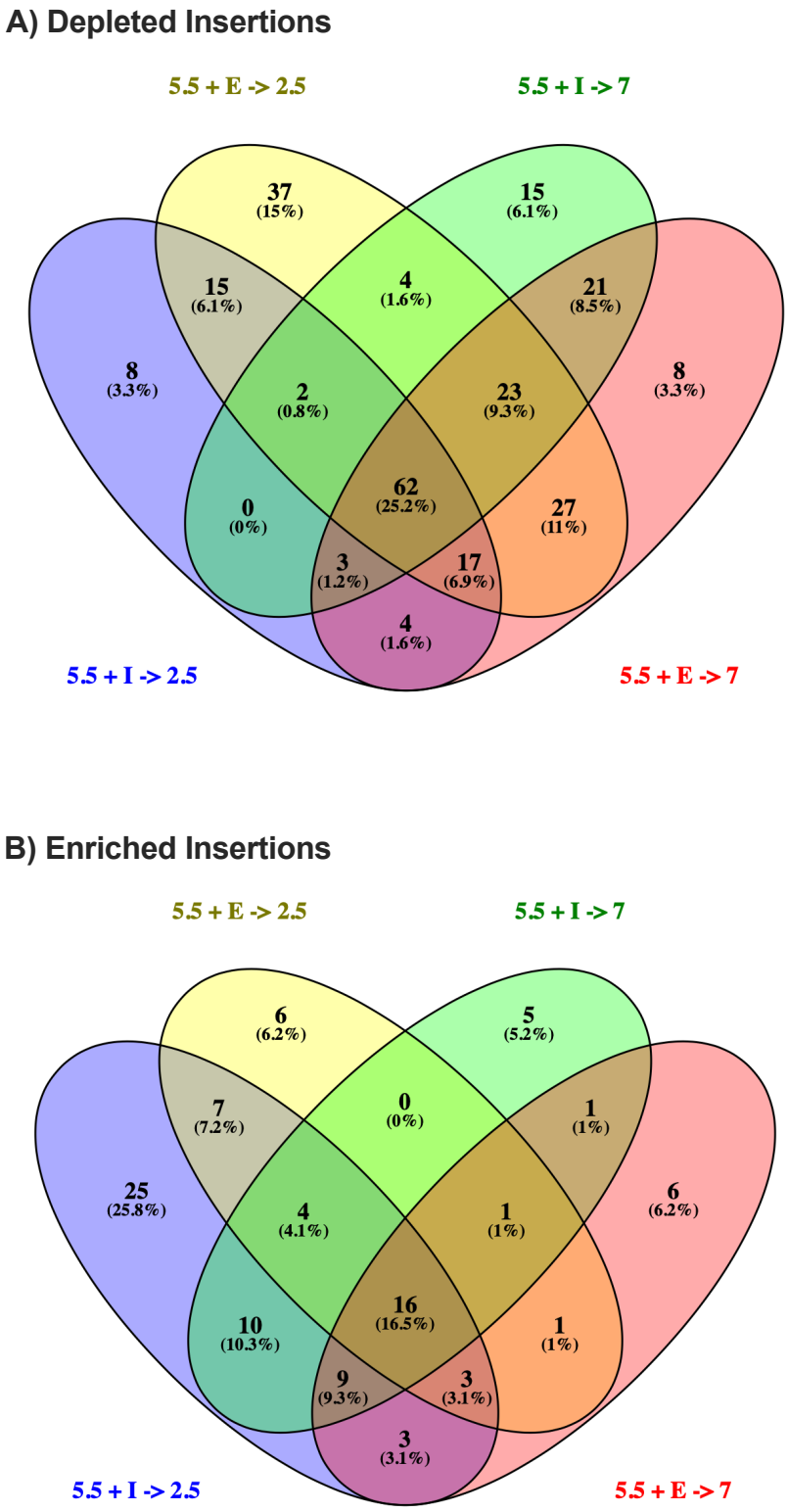


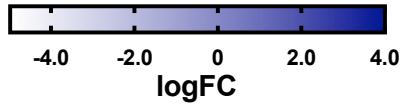
Figure 5.11: Venn diagrams showing genes where the number of insertions had either decreased (depleted insertions) (A), or increased (enriched insertions) (B), in the conditional TraDIS libraries as compared to the ITL.

of that step. The genes in this category may therefore represent genes whose impact on fitness would be seen even under normal growth conditions in the absence of any pH stress.

As discussed in section 5.2.2.1, the aim of the TraDIS experiments was to identify genes that are important for survival of *E. coli* at pH 2.5, and genes which are involved in the inhibitory effect of indole on AR2 induction. It was also mentioned in the same section that in order to identify genes which either confer a fitness advantage or disadvantage at pH 2.5, genes in which insertions are depleted or enriched in either or both of the 5.5 + I -> 2.5 and 5.5 + E -> 2.5 conditions need to be looked at. To identify genes involved in the inhibitory effect of indole, genes in which insertions are enriched in either or both of the 5.5 + I -> 2.5 and 5.5 + I -> 7 conditions need to be looked at. So within the Venn diagrams, genes falling in these categories were identified and their log_2FC and q values were compared across four conditions. These are given in Tables 5.3 (depleted insertions) and 5.4 (enriched insertions). A full list of genes in each category is given in Appendices 8.2.3 (depleted insertions) and 8.2.4 (enriched insertions). In Table 5.3, 4 AR2 genes (*evgA, evgS, gadE, hdeA*), 6 essential genes (*hipB, ymiB, iscU, iraM, topA, ydfO*) and 4 genes which were below 200 bp in length (*rrfF* - 120 bp, *yibY* - 111 bp, *ymgK* - 45 bp, *ynfT* -114 bp) were removed from the comparison. In Table 5.4, 40 essential genes and 3 genes below 200 bp in length (*ymdG* - 27 bp*, yncL* - 96 bp*, yqhJ* - 60 bp) are not shown. Since the comparison of log_2FC and q values in AR2 genes has already been discussed in section 5.2.2.5, these genes were excluded from in this section. The essential genes were excluded because they either do not tolerate any or only a few transposon insertions as they are absolutely necessary for the growth of cells in any condition. Therefore the change in the frequency of the mutants containing insertions in these genes cannot be looked at. When EdgeR performs the comparisons, it includes both essential and non-essential genes and essential genes need to be manually excluded from the gene lists

Table 5.3: Comparison of log_2FC and q values of transposon mutants whose frequency reduced (depleted insertions) under at least one TraDIS experimental condition compared to the ITL. The legend bars at the bottom show the colour corresponding to the log_2FC and q values.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
aroK	-1.14	-2.14	-1.74	-1.85	3.57×10^{-3}	2.88×10^{-6}	1.30×10^{-7}	1.17×10^{-6}
atpA	-1.44	-2.08	-1.07	-1.36	1.93×10^{-11}	2.35×10^{-8}	3.01×10^{-10}	5.62×10^{-15}
atpB	-2.12	-2.80	-1.67	-1.73	2.16×10^{-19}	3.00×10^{-18}	1.45×10^{-12}	2.49×10^{-12}
atpD	-1.69	-2.64	-1.26	-1.07	8.33×10^{-10}	1.55×10^{-20}	1.17×10^{-13}	2.19×10^{-7}
atpF	-1.55	-2.46	-1.33	-1.26	6.07×10^{-5}	7.48×10^{-5}	0.01	3.80×10^{-3}
atpG	-1.22	-2.13	-1.32	-1.50	2.18×10^{-7}	3.12×10^{-10}	3.85×10^{-5}	2.04×10^{-6}
clsA	-2.14	-1.43	-0.37	-0.21	7.62×10^{-16}	7.49×10^{-5}	0.13	0.41
cspF	-1.24	-2.10	-0.73	-1.35	0.01	4.78×10^{-3}	0.19	0.01
degP	-2.80	-2.56	-1.75	-1.86	1.25×10^{-77}	1.59×10^{-28}	1.85×10^{-40}	8.38×10^{-42}
emrR	-0.33	-2.15	0.23	0.17	0.30	2.86×10^{-5}	0.44	0.57
epmA	-1.77	-2.88	-1.38	-1.26	4.81×10^{-9}	2.96×10^{-14}	5.80×10^{-9}	2.35×10^{-5}
fepC	-1.25	-2.03	-1.64	-1.37	0.04	2.02×10^{-4}	2.43×10^{-3}	8.75×10^{-4}
fis	-2.84	-0.97	3.83	0.56	4.94×10^{-7}	0.35	3.96×10^{-3}	0.09
hfq	-2.79	-2.79	-0.23	-0.21	1.33×10^{-21}	3.73×10^{-12}	0.35	0.41
holC	-1.81	-2.06	-1.32	-1.82	0.01	0.01	0.04	0.04
hscB	-1.75	-3.72	-0.59	-1.51	2.23×10^{-4}	1.39×10^{-7}	0.33	0.02
ihfA	-2.33	-2.71	-0.74	-1.56	1.31×10^{-5}	7.00×10^{-4}	0.15	2.39×10^{-3}
ihfB	-2.68	-2.76	-0.69	-1.21	1.10×10^{-12}	8.95×10^{-13}	0.01	1.21×10^{-6}
insA2	-1.58	-2.64	-1.70	-1.96	0.01	2.35×10^{-3}	0.01	0.01
iscA	-1.58	-2.79	-0.84	0.25	0.01	1.06×10^{-4}	0.21	0.75
ldcA	-2.81	-2.19	-1.11	-1.16	1.31×10^{-27}	4.88×10^{-15}	3.83×10^{-6}	2.48×10^{-10}
lpoA	-2.14	-2.66	-0.92	-0.81	8.17×10^{-26}	3.17×10^{-36}	1.40×10^{-10}	1.56×10^{-11}
lpxM	-0.38	-2.26	-0.07	-0.29	0.16	1.41×10^{-7}	0.84	0.23
miaA	-1.84	-2.20	-0.57	-0.58	5.75×10^{-19}	3.70×10^{-17}	1.60×10^{-3}	0.01
minD	-2.25	-1.51	-1.02	-1.30	1.33×10^{-25}	8.79×10^{-9}	1.17×10^{-8}	6.26×10^{-12}
mrcA	-2.06	-2.64	-0.72	-0.70	1.69×10^{-17}	4.59×10^{-36}	8.15×10^{-7}	4.83×10^{-8}
mrcB	-2.78	-2.98	-2.26	-1.91	3.21×10^{-83}	9.55×10^{-45}	1.85×10^{-54}	8.62×10^{-36}
nlpI	-2.01	-2.07	-0.19	-0.11	1.20×10^{-20}	2.89×10^{-16}	0.34	0.58
rimP	-2.01	-1.28	-0.72	-1.07	0.02	0.22	0.31	0.29
rlmE	-1.51	-2.55	-1.72	-1.11	0.01	0.01	1.57×10^{-3}	0.06
rzoQ	-1.60	-2.48	-1.83	-1.48	3.58×10^{-3}	2.50×10^{-3}	6.77×10^{-4}	0.01
tusE	-1.80	-2.11	-1.41	-1.12	0.03	0.02	0.16	0.34
valT	-0.70	-2.02	-1.85	-1.48	0.41	0.04	0.05	0.17
valZ	-0.64	-2.21	-1.14	-1.61	0.35	0.01	0.13	0.03
waaB	-1.69	-2.29	-1.51	-1.93	2.74×10^{-10}	9.84×10^{-14}	5.40×10^{-11}	3.11×10^{-10}
waaU	-1.93	-2.30	-1.48	-1.92	1.87×10^{-16}	5.84×10^{-9}	1.25×10^{-13}	2.96×10^{-14}
wbbJ	-2.09	-2.17	-1.66	-2.00	2.66×10^{-33}	8.61×10^{-12}	7.93×10^{-20}	3.10×10^{-27}
ybfQ	-2.14	-2.16	-1.39	-1.90	2.43×10^{-3}	0.09	0.21	0.04
yciB	-3.08	-4.98	-1.97	-1.83	3.34×10^{-11}	1.13×10^{-8}	1.24×10^{-5}	6.68×10^{-4}
yciG	-2.28	-2.00	-1.27	-1.87	0.01	0.02	0.18	0.02
ydeQ	-1.68	-2.17	-1.58	-1.84	3.19×10^{-6}	3.87×10^{-6}	1.67×10^{-5}	9.65×10^{-7}
yehC	-1.39	-2.35	-1.06	-1.39	5.60×10^{-4}	1.15×10^{-3}	0.01	4.21×10^{-3}
yjbS	-1.81	-3.26	-1.78	-1.94	0.07	0.01	0.05	0.05
ymgF	-1.91	-2.06	-1.90	-1.83	6.24×10^{-15}	3.77×10^{-4}	1.70×10^{-9}	1.92×10^{-9}
yncI	-1.91	-2.13	-1.43	-1.86	1.41×10^{-10}	2.85×10^{-5}	6.44×10^{-5}	9.68×10^{-10}
yqeH	-1.72	-2.19	-1.63	-1.79	2.40×10^{-6}	6.19×10^{-5}	4.05×10^{-5}	2.43×10^{-6}



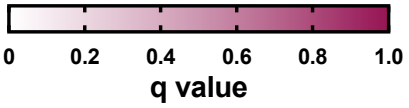
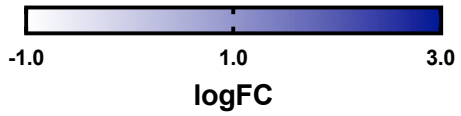
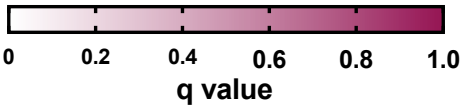


Table 5.4: Comparison of log_2FC and q values of transposon mutants whose frequency increased (enriched insertions) under at least one TraDIS experimental condition compared to the ITL. The legend bars at the bottom show the colour corresponding to the log_2FC and q values.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
nagA	3.01	2.21	1.08	1.00	2.65×10^{-38}	1.23×10^{-7}	3.95×10^{-10}	1.54×10^{-7}
nagC	1.34	2.01	0.81	0.74	2.41×10^{-18}	1.38×10^{-8}	1.43×10^{-7}	8.92×10^{-6}
opgH	2.48	1.89	0.28	0.43	3.97×10^{-74}	3.20×10^{-15}	0.16	1.90×10^{-3}
rep	2.70	1.48	-0.33	-0.40	1.08×10^{-64}	1.49×10^{-5}	0.04	0.01
rffC	2.47	2.67	1.18	1.03	1.62×10^{-49}	7.71×10^{-21}	3.43×10^{-13}	6.06×10^{-15}
wecB	1.68	2.23	0.98	0.89	5.88×10^{-29}	1.21×10^{-13}	8.86×10^{-11}	5.23×10^{-11}
wecC	1.74	2.30	1.04	0.87	4.04×10^{-32}	4.18×10^{-17}	1.44×10^{-11}	5.33×10^{-12}
wecE	2.49	2.76	1.08	1.06	1.80×10^{-68}	9.15×10^{-22}	8.66×10^{-9}	1.55×10^{-15}
wecF	2.16	2.31	0.95	0.78	6.46×10^{-49}	2.01×10^{-11}	8.56×10^{-11}	2.52×10^{-7}





during further analysis. It is also not possible to determine the conditional essentiality of short genes due to presence of no or a very small number of insertions.

It is worth noting that although Venn diagrams give a reasonable overview of the numbers of genes changing in each category, they are less useful when it comes to looking at individual genes and pathways, because of the arbitrary nature of the cut-off points discussed above. This is clearly illustrated by the data in Tables 5.3 and 5.4 which show the log_2FC and q values for all genes where at least one of the four conditions was judged to be significant by the criteria used to generate the gene lists for the Venn diagrams. It is evident from these data that many genes show values which suggest effects on fitness under multiple conditions, even if not all were above the cut-off value. For example, *atpA* shows log_2FC of -1.44, -2.08, -1.07 and -1.36 under the four conditions, all at very highly significant levels (all q values are $<10^{-7}$). It would be unreasonable to suggest that it was having an effect on fitness only when the log_2FC is -2.08 but not when it was -1.44. The discussion that follows therefore looks at some "low hanging fruit": cases (single genes, genes encoding a particular protein complex or involved in synthesis of a particular component) which are particularly striking in the datasets in Tables 5.3 and 5.4. A more detailed statistical analysis is then presented in section 5.2.2.7.

To know the biological processes, molecular functions, and the cellular components that the proteins encoded by the genes given in Tables 5.3 and 5.4 are involved in, the genes were analysed using the String database (https://string-db.org, version 11.5). Two sets of genes, each of which encode proteins involved in a common function were found. These were 5 *atp* genes (*atpA, atpB, atpD, atpF,* and *atpG*) which all encode components of F_1F_o ATP synthase, and 5 *wec* genes (*wecB*, *wecC, rffC/wecD, wecE,* and *wecF*) which are involved in the biosynthesis of enterobacterial common antigen, an outer membrane glycolipid present in all members of *Enterobacteriaceae* family.

Insertions in genes involved in synthesis of the F_1F_o ATP synthase generally have a negative effect on strain fitness under all four conditions (Table 5.3). The insertions in remaining three *atp* genes, (*atpC, atpH*, and *atpE)* which did not come up in Venny analysis, also mostly had a significant negative effect on strain fitness under all four conditions (data in Appendix 8.2.1). This is perhaps unsurprising as this enzyme is required for generation of most of the cell's ATP when cells are growing using respiration, which is likely to have been the case during this experiment, and cells with reduced ATP levels are likely to be generally less fit. It is however striking that the reduction in fitness caused by mutation of these genes is greater when cells were incubated at pH 2.5 than pH 7. This could simply be because the low levels of ATP in these cells makes them more susceptible to damage (or less able to repair such damage) caused by an extremely low pH. Another striking observation was that the fitness of the strains containing mutation in most of the *atp* genes was greater at pH 2.5 in the presence of ethanol than indole. It is also of interest that the F_1F_o ATP synthase has been proposed to be a component of the poorly-characterised AR1 system of *E. coli* (Richard and Foster, 2004), and although this system is mostly thought to be induced in stationary phase, and repressed by glucose, and provides relatively little protection at pH 2.5, its absence in the mutated strains may account for some of the reduced fitness at lower pH. If this is the case, it is intriguing that this effect is reduced in the presence of indole. As discussed in the previous section, incubation in the presence of indole may cause a general depression of fitness at pH 2.5, and so the relatively small effect of AR1 may be lost against this larger change.

Insertions in *wec* genes generally have a positive effect on strain fitness under all four conditions (Table 5.4). A striking result was that loss of function in all *wec* genes led to a fitter phenotype at pH 2.5 than pH 7, implying that loss of function of these genes can increase acid resistance. As discussed and shown previously in section 5.2.2.5, mutants with insertions in

waaC gene, which is involved in LPS synthesis, show a smaller decline in fitness at pH 2.5 than pH 7 (Table 5.2). Furthermore, *waaC* mutants have been shown to have a greater acid resistance than the wild-type strain at pH 2.3 (Vivijs, Aertsen and Michiels, 2016). The TraDIS results, for *waaC* and *wec* genes combined, suggest that loss of LPS at low pH is needed for the optimal fitness of cells. LPS is an integral component for maintaining cell integrity and protecting cells against environmental stresses. However, reducing LPS synthesis at low pH might help the cells to direct more energy and resources towards maintaining internal pH.

The data for inserts in *fis* gene is of particular interest. This is the only gene where inserts caused (a) a significant loss of fitness at pH 2.5 in the presence of indole, but no effect in the presence of ethanol, (b) a significant increase in fitness at pH 7 in the presence of indole but again, no effect in the presence of ethanol. These two results suggest that there is an important connection between the function of Fis and the effect of indole on cells. Fis is one of the 7 global regulators and controls the expression of a large number of genes. This makes it difficult to hypothesise the potential effect of Fis on indole-treated cells. The starting point would be to validate the results of TraDIS experiments by competing the *fis* mutant with the isogenic wild-type strain under the same conditions as the ones used in TraDIS experiments. The competition assays should then be followed by survival assays by checking whether *fis* mutant shows a compromised survival at pH 2.5, and an enhanced survival at pH 7, when induced in the presence of indole.

The data for several other genes given in Table 5.3 are of interest, some of which will be discussed here. Inserts in three genes (*clsA, mrcA*, and *nlpI)* caused a significant loss of fitness when cells were incubated at pH 2.5, but not pH 7. Inserts in *mrcB* and *yciB*, caused a significant loss of fitness under all four conditions. However, the reduction in fitness caused by mutation of these genes was greater when cells were incubated at pH 2.5 than pH

7. Inserts in *emrR* and *lpxM* reduced the fitness of strains at pH 2.5 but only in the presence of ethanol and not indole, which was a striking observation.

YciB is a protein of unknown function, which is predicted to be located in the inner membrane. Strains containing inserts in *yciB* showed the highest loss in fitness at pH 2.5, out of all 8 genes selected for validation (Table 5.3). It is known that *yciB* along with *dcrB* plays a synergistic role in cell envelope integrity, and *yciB dcrB* double mutant has a reduced proton motive force (PMF) (Mychack *et al.*, 2019). Maintaining high magnitude PMF is essential for the cells to maintain constant cytoplasmic pH (Terradot *et al.*, 2021). YciB appears to be necessary for optimal fitness of *E. coli* at pH 2.5. The reasons for this are not currently clear but they may be associated with maintenance of the PMF at extreme pH values.

mrcA and *mrcB* encode the penicillin binding protein 1A and 1B (PBP1A and PBP1B) respectively. PBPs are essential for the synthesis of cell wall peptidoglycan (PG) since they are involved in the polymerisation of glycan strands and the cross-linking of these glycan strands through short chain peptides (reviewed in Sauvage *et al.*, 2008). PBP1A-deficient cells have been shown to have a reduced growth and fitness in alkaline (pH 6.5-8.2) media, whereas PBP1B-deficient cells have a reduced growth and fitness in acidic (pH < 5.9) media (Mueller *et al.*, 2019). Our TraDIS results, however, show that the function of both *mrcA* and *mrcB* is required for the optimal fitness of *E. coli* at low pH. This is in contrast to previous findings that *mrcB*, but not *mrcA*, is needed for optimal fitness of *E. coli* at low pH.

nlpI encodes an outer-membrane anchored lipoprotein. NlpI functions as an adaptor for PG endopeptidases, a type of hydrolases involved in PG synthesis (Banzhaf *et al.*, 2020). It has been hypothesised that NlpI acts as a scaffold protein between PG hydrolases, specifically endopeptidases and PG synthases, specifically PBP1A (*mrcA*). Since our TraDIS results showed that inserts in *nlpI* and *mrcA* both cause a loss of fitness at pH 2.5 but not pH 7, it is possible

that NlpI is required for the function of PBP1A. While the role of *mrcA* in acid stress has been studied, the role of *nlpI* in low pH is still unexplored.

Inserts in *lpxM* which encodes lipid A biosynthesis myristoyltransferase, an enzyme involved in LPS synthesis, showed a decline in fitness at pH 2.5 but only in the presence of ethanol and not indole. This TraDIS result is interesting (a) because of the difference in the fitness in the presence of indole and ethanol, and (b) the previous TraDIS results with *waaC* (Table 5.2) and *wec* genes (Table 5.4) have suggested that the loss of LPS is needed for the optimal fitness of *E. coli* at low pH both in the presence of indole and ethanol. However, the TraDIS results for *lpxM* suggest that loss of LPS is needed for the optimal fitness of *E. coli* at low pH only in the presence of ethanol.

clsA encodes cardiolipin synthase A enzyme which produces cardiolipin, one of the three membrane phospholipids present in bacterial membranes. ClsA is the main cardiolipin synthase that operates during exponential phase, while the other two cardiolipin synthases, ClsB and ClsC, operate in stationary phase (Tan *et al.*, 2012). ClsA is involved in lipopolysaccharide (LPS) transport to the outer membrane, and shows a synthetic lethal phenotype with LpxM (Douglass, Cléon and Trent, 2021). The role of other membrane phospholipids in acid stress has been studied previously. Increased levels of membrane phospholipid phosphoethanolamine through overexpression of phosphatidylserine synthase (*pssA*) leads to improved tolerance to low pH (Tan *et al.*, 2017). According to our TraDIS results, inserts in *clsA* reduce the fitness of cells at pH 2.5 but not pH 7. Furthermore, the decline in fitness is bigger in the presence of indole than ethanol. One possible hypothesis for this observation is that ClsA reduces the diffusion of indole across the outer membrane into the cells through enhanced LPS transport to the outer membrane. Indole remains uncharged under the conditions used in the TraDIS experiments for this study, so it could easily diffuse

through lipid membranes. The reduced indole diffusion into the cells will have a smaller inhibitory effect on AR2 response, which in turn will increase the acid resistance and survival of cells at low pH.

EmrR is a DNA-binding transcriptional repressor which negatively regulates the expression of genes encoding the multidrug resistance efflux pumps such as *emrA* and *emrB* (Lomovskaya, Lewis and Matin, 1995). According to our TraDIS results, inserts in *emrR* reduce the fitness of cells at pH 2.5, but only in the presence of ethanol and not indole. *E. coli* has multiple efflux pumps where the drug export is known to be coupled with proton import (Paulsen, Brown and Skurray, 1996). Since protons are hydrophilic, it is possible that they can enter the cells at a faster rate through efflux pumps than by passive diffusion through cell membrane. Therefore, the transcriptional repression of efflux pumps by EmrR would increase the fitness of cells at low pH.

All these genes (*clsA, emrR, lpxM, fis, mrcA, nlpI,* and *yciB*), and *mrcB* to compare with *mrcA*, showed interesting TraDIS results and there is a potential that their function could change the fitness of *E. coli* under different conditions which were tested in TraDIS experiments. These genes were therefore selected for validation experiments (described in section 5.2.3) to verify the results of TraDIS experiments.

5.2.2.7 Comparison of metabolic and regulatory pathways important in four conditions

TraDIS experiments provide a huge dataset and looking at each gene individually can be a lengthy process. A more powerful analytical method is needed to analyse such data. Gene set enrichment analysis (GSEA) ranks gene sets, involved in different metabolic or regulatory pathways, depending on the difference in the log-fold change of all the genes present within a set between the ITL and the conditional TraDIS library. The gene sets present

in the ranked list are provided an enrichment score (NES) which shows that the gene set is associated with a phenotype being studied. These gene sets are further allocated a q value which gives the level of the significance of a particular NES score. Positive and negative NES scores present at the extreme ends of ranked lists correspond to the pathways which are either significantly enriched or depleted in the condition relative to the control. Compared to the analysis described in previous sections which involved taking arbitrary cut-offs, GSEA is an analytical and statistically powerful tool which reduces the risks of false-positives and false-negatives.

GSEA was used to determine the KEGG metabolic and transcriptional regulatory pathways that were either significantly depleted or enriched in at least one of the four TraDIS experimental conditions compared to the ITL (Tables 5.5 and 5.6). An enriched pathway (positive NES score) in a condition means that the frequency of the mutants containing insertions in genes involved in this pathway increased in conditional TraDIS library relative to the ITL. On the other hand, a depleted pathway (negative NES score) in a condition means that the frequency of the mutants containing insertions in genes involved in this pathway decreased in conditional TraDIS library relative to the ITL.

The KEGG pathway "purine metabolism" and the transcriptional pathway regulated by PurR, which includes genes involved in purine biosynthesis, were both significantly depleted in all four conditions (Tables 5.5 and 5.6). As discussed before in section 5.2.2.4, M9 minimal media lacks purines and pyrimidines. Therefore, mutants containing inserts in genes involved in purine metabolism will have reduced fitness in all conditions.

The transcriptional pathway regulated by EvgA, which includes genes involved in AR2 response, was depleted under all four conditions. However, the NES score was more negative and significant only at pH 2.5. As shown before in section 5.2.2.5 (Table 5.1), the mutants

Table 5.5: KEGG metabolic pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions relative to the ITL. The legend bars at the bottom show the colour corresponding to the normalised enrichment score (NES) and q value.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
Eco00061 (Fatty acid biosynthesis)	2.12	1.92	1.82	2.07	5.71×10^{-4}	0.02	0.04	2.10×10^{-3}
Eco00300 (Lysine biosynthesis)	1.84	1.92	1.76	1.82	0.02	0.02	0.06	0.03
Eco00440 (Phosphonate and phosphinate metabolism)	1.95	2.05	2.03	2.07	0.01	4.16×10^{-3}	3.63×10^{-3}	2.21×10^{-3}
Eco00630 (Glyoxylate and dicarboxylate metabolism)	1.56	1.82	1.16	1.35	0.15	0.03	0.52	0.31
Eco00740 (Riboflavin metabolism)	1.67	1.84	1.80	1.61	0.08	0.03	0.04	0.21
Eco00780 (Biotin metabolism)	1.81	1.14	-1.08	1.21	0.03	0.45	0.60	0.45
Eco00900 (Terpenoid backbone biosynthesis)	1.80	1.84	1.74	1.95	0.03	0.03	0.06	0.01
Eco00970 (Aminoacyl-tRNA biosynthesis)	1.86	1.37	1.58	1.72	0.03	0.30	0.16	0.08
Eco01210 (2-Oxocarboxylic acid metabolism)	1.85	1.50	1.15	1.03	0.02	0.20	0.51	0.62
Eco01212 (Fatty acid metabolism)	2.15	1.84	1.91	2.09	8.57×10^{-4}	0.03	0.02	1.79×10^{-3}
Eco02010 (ABC transporters)	1.75	1.70	1.40	1.46	0.05	0.07	0.40	0.26
Eco02040 (Flagellar assembly)	1.67	2.04	1.61	1.27	0.09	3.78×10^{-3}	0.14	0.39
Eco03010 (Ribosome)	2.71	2.62	2.45	2.75	0.00	0.00	0.00	0.00
Eco00230 (Purine metabolism)	-1.71	-1.83	-2.01	-1.95	0.04	0.01	0.00	6.14×10^{-4}
Eco00250 (Alanine, aspartate and glutamate metabolism)	-1.95	-1.78	-2.14	-2.07	2.97×10^{-3}	0.02	0.00	0.00
Eco00400 (Phenylalanine, tyrosine and tryptophan biosynthesis)	0.61	-1.32	-1.76	-1.72	0.95	0.42	0.02	0.02
Eco00540 (Lipopolysaccharide biosynthesis)	-1.35	-1.62	-1.73	-1.76	0.68	0.09	0.02	0.01
Eco00750 (Vitamin B6 metabolism)	-1.77	-1.99	-1.89	-1.94	0.03	1.05×10^{-3}	2.11×10^{-3}	8.28×10^{-4}
Eco00920 (Sulfur metabolism)	-1.17	-1.57	-1.73	-1.71	0.89	0.13	0.02	0.02
Eco01240 (Biosynthesis of cofactors)	1.36	-1.54	-1.73	-1.65	0.30	0.14	0.02	0.04

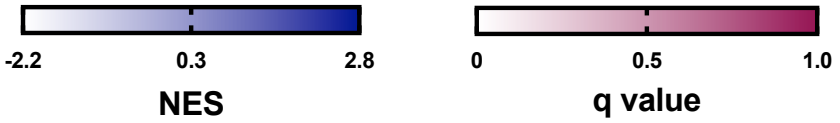


Table 5.6: Transcriptional regulatory pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions relative to the ITL. The legend bars at the bottom show the colour corresponding to the normalised enrichment score (NES) and q value.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
ArgP	1.88	2.03	1.81	1.92	0.02	2.64×10^{-3}	0.04	0.01
FadR	1.93	1.97	1.85	2.07	0.02	3.98×10^{-3}	0.05	0.01
FhlA	1.65	1.90	1.82	1.83	0.22	0.01	0.05	0.03
FlhDC	1.71	2.10	1.68	1.30	0.16	1.22×10^{-3}	0.18	0.48
ModE	1.53	1.83	1.82	1.92	0.26	0.03	0.06	0.01
MraZ	1.89	2.04	1.86	1.94	0.02	2.65×10^{-3}	0.10	0.01
PhoB	1.62	1.89	1.54	1.65	0.23	0.01	0.28	0.23
BirA	-1.42	-1.75	-1.78	-1.79	0.23	0.03	0.03	0.01
BluR	-1.74	-1.67	-1.67	-1.66	0.04	0.04	0.08	0.06
EvgA	-1.80	-1.71	-1.49	-1.61	0.03	0.03	0.23	0.08
OmpR	-1.76	-1.87	-1.65	-1.80	0.04	4.92×10^{-3}	0.08	0.02
PurR	-2.40	-2.23	-2.19	-2.29	0.00	0.00	0.00	0.00



containing insertions in genes involved in AR2 response including *evgA* and *evgS* showed a significant decline in fitness under all four conditions. However, the decline in fitness was greater at pH 2.5 in the presence of ethanol than indole, and pH 7. It was also proposed in that section that the decline in the fitness at pH 7 could be because the function of these genes provides fitness advantage at either pH 7 or pH 5.5. The function of these genes is most important at pH 2.5 in the presence of ethanol because they are involved in AR2 response, and the activation of AR2 response is required for the optimal fitness of cells at pH 2.5. In the presence of indole, however, the function of these genes becomes less important because indole blocks the AR2 response in all cells, not just those with mutations in *evgA* and *evgS*. Since the log_2FC of some of the AR2 genes was similar at pH 7 and at pH 2.5 in the presence of indole, it was hypothesised that the function of AR2 genes which is needed for the optimal fitness of cells at pH 7 is not inhibited by indole. Other genes such as *gadE* and *rpoS* had a more negative log_2FC at pH 2.5 in the presence of indole than at pH 7, suggesting that AR2 may still be able to function to some extent in the presence of indole, possibly by the RpoS-mediated activation of *gadE*. EvgA transcriptional pathway is depleted at pH 2.5 in the presence of both indole and ethanol, which also suggests that the function of AR2 genes may still be important for the optimal fitness of cells at pH 2.5 in the presence of indole.

The KEGG pathway "lipopolysaccharide biosynthesis" was depleted under all four conditions (Table 5.5). However, it was depleted more at pH 7 than pH 2.5. Furthermore, leading edge analysis showed that majority of the genes which contributed to the NES score were *waa* genes. It has been shown before in section 5.2.2.5 (Table 5.2), that inserts in *waaC* gene reduced the fitness more at pH 7 than pH 2.5. It was also shown in section 5.2.2.6 (Table 5.4) that mutants containing inserts in *wec* genes, which are involved in LPS synthesis, show a higher fitness at pH 2.5 than pH 7. All these observations suggest that loss of LPS is needed

for the optimal fitness of cells at pH 2.5. As hypothesised before in section 5.2.2.6, this could be because it allows the cells to direct more energy and resources towards maintaining internal pH.

The KEGG pathway "phenylalanine, tyrosine and tryptophan biosynthesis" was depleted at pH 7 and at pH 2.5 in the presence of ethanol (Table 5.5). Since amino acids are the building blocks of peptides and proteins, it is reasonable that inserts in genes involved in the biosynthesis of these amino acids reduce the fitness of cells. The depletion was smaller at pH 2.5 in the presence of ethanol than at pH 7. Additionally, at pH 2.5 in the presence of indole, this pathway was enriched. This is because at pH 2.5, the priority of cells is to survive the stress rather than to grow and divide. Since indole inhibits the induction of AR2 response, reducing the acid resistance of cells and making them more susceptible to damage, the priority of cells to cope with acid stress and improve their survival rather than to grow and divide increases even more in the presence of indole than ethanol.

The KEGG pathway "biosynthesis of cofactors" was depleted at pH 7 and pH 2.5 in the presence of ethanol (Table 5.5). However, at pH 2.5 in the presence of indole, this pathway was enriched. Since indole inhibits the induction of AR2 response, the cells might reduce the activity of biosynthesis pathways to save energy so that it can be utilised at maintaining the internal pH under extreme acid stress. In the presence of ethanol at pH 2.5, however, the activity of AR2 response helps the cells to maintain internal pH.

The KEGG pathway "glyoxylate and dicarboxylate metabolism" was enriched under all four conditions, but more at pH 2.5 than pH 7 (Table 5.5). This observation is consistent with the hypothesis by Maurer *et al.* (2005) that carbohydrate metabolism is preferred in alkaline pH to generate fermentation acids which can lower the internal pH whereas acid metabolism is preferred in low pH to raise the internal pH. If this is true in our TraDIS experiments, then

the enrichment of this pathway at pH 7 could be because of the pre-incubation step at pH 5.5.

The KEGG pathway "flagellar assembly" and the transcriptional pathway regulated by FlhDC, which includes the genes for flagellar assembly, were enriched in all four conditions (Tables 5.5 and 5.6). However, the enrichment of these pathways was highest at pH 2.5 in the presence of ethanol. Flagellar assembly requires ATP and cells under extreme acid stress may want to save energy so that it could be used to maintain the internal pH and repair the damage caused by acid stress. So it is understandable that inserts in flagellar assembly genes increase the fitness of cells at pH 2.5. It is striking however that mutants containing inserts in flagellar genes at pH 2.5 in the presence of indole are as fit as those at pH 7. Given that AR2 system is inhibited at pH 2.5 in the presence of indole, and these cells must require ATP to maintain internal pH, the loss of function of flagellar assembly genes should increase the fitness of these cells more than at pH 7.

5.2.3 Validation of TraDIS experiment results

As explained in section 5.2.2.5, 8 genes (*clsA, emrR, fis, lpxM, mrcA, mrcB, nlpI*, and *yciB*) were selected to validate the results of TraDIS experiments. An attempt was made to knock out these genes in MG1655 and create single gene-deletion mutants. However, only three mutants (*ΔmrcA, ΔmrcB* and *ΔnlpI*) could be made and due to time constraints, it was not possible to have more attempts at making the remaining mutants (Figure 5.12). The TraDIS results were therefore validated using MG1655 *ΔmrcA,* MG1655 *ΔmrcB* and MG1655 *ΔnlpI.* MG1655 *ΔgadE* was used as a control because *gadE* is known to be involved in the AR2 acid resistance response.

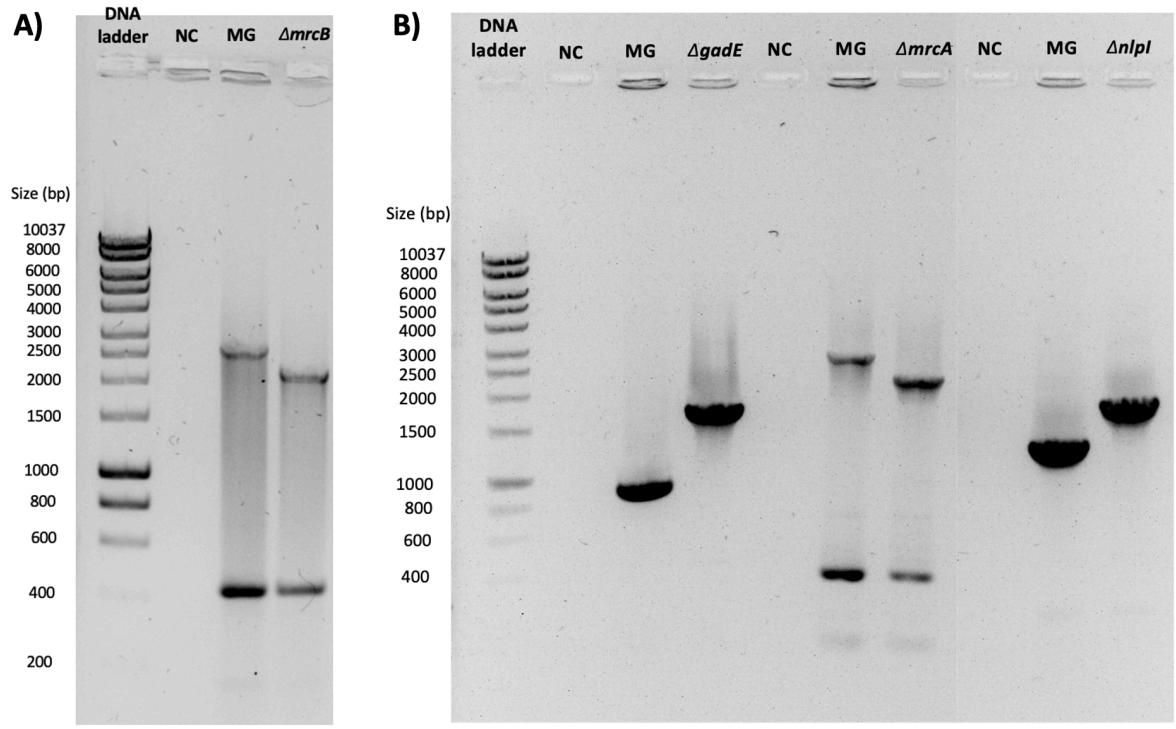


Figure 5.12: MG1655 *ΔmrcB* (A), MG1655 *ΔgadE,* MG1655 *ΔmrcA,* and MG1655 *Δnlpl* (B) deletion mutants prepared using P1 transductions.

gadE and *nlpI* genes were both replaced with a kanamycin resistance cassette, whereas *mrcA* and *mrcB* genes were replaced with a tetracycline resistance cassette. Gene-flanking primers (given in Material and Methods, Table 2.3) were used to perform the PCR reactions to verify the gene deletion. The PCR products were visualised on the agarose gel alongside the negative PCR controls (NC; PCR reactions done without template DNA) and the PCR reactions performed on wild-type MG1655 strain. The expected fragment lengths (in bp) for MG1655 and knockouts (KO) were: *ΔmrcB* (MG- 2705, KO- 2170), *ΔgadE* (MG- 866, KO- 1816)*, ΔmrcA* (MG- 2809, KO- 2256)*,* and *ΔnlpI* (MG- 1130, KO- 1723). The ~400 bp long bands observed in *ΔmrcA and ΔmrcB* gels were due to non-specific binding between the primers and the genomic DNA.

According to TraDIS results, inserts in *mrcA*, and *nlpI* caused a significant loss of fitness when cells were incubated at pH 2.5 (Table 5.3). Inserts in *gadE* and *mrcB* caused a significant loss of fitness under all four conditions (Tables 5.1 and 5.3). However, the reduction in fitness caused by mutation of *gadE* and *mrcB* was greater when cells were incubated at pH 2.5 than pH 7. The TraDIS results therefore suggest that the function of all four genes is needed for the optimal fitness of *E. coli* at pH 2.5.

During TraDIS experiments, the mutants were grown at pH 7 for a considerable proportion of time, before and after being induced and acid shocked. Therefore, the decline in the fitness of these mutants could be due to a combination of the fact that these mutants are able to grow less well at pH 7, and the function of these genes impart a fitness advantage to the cells at pH 2.5. Additionally, it is possible that these mutants have a reduced fitness under all conditions used in the TraDIS experiments. Indeed, both *mcrB* and *gadE* mutants showed a significant decline in fitness at pH 7 in TraDIS results. *mrcA* mutant showed a small decline in fitness, whereas *nlpI* mutant showed a minor decline in fitness at pH 7. Therefore, before testing whether the function of these genes is needed for the optimal fitness of *E. coli* at pH 2.5, the fitness of *gadE, mrcA, mrcB* and *nlpI* mutants was checked at pH 7 by growing them with KH001 to mid-exponential phase (the phase in which TraDIS experiments were performed) in competition assays. This is because if a mutant shows a compromised fitness when competed with other mutant(s) during growth at pH 7, that effect is likely to amplify when the cells are exposed to pH 2.5. So the reduced fitness of that mutant at pH 7 does not necessarily means that the function of the gene imparts a fitness advantage at pH 2.5.

The control mutant MG1655 *ΔgadE* showed no difference in fitness relative to MG1655 during growth at pH 7 (Figure 5.13 A). This implies that the function of *gadE* is not required during normal growth which is in contrast to the TraDIS results where *gadE*

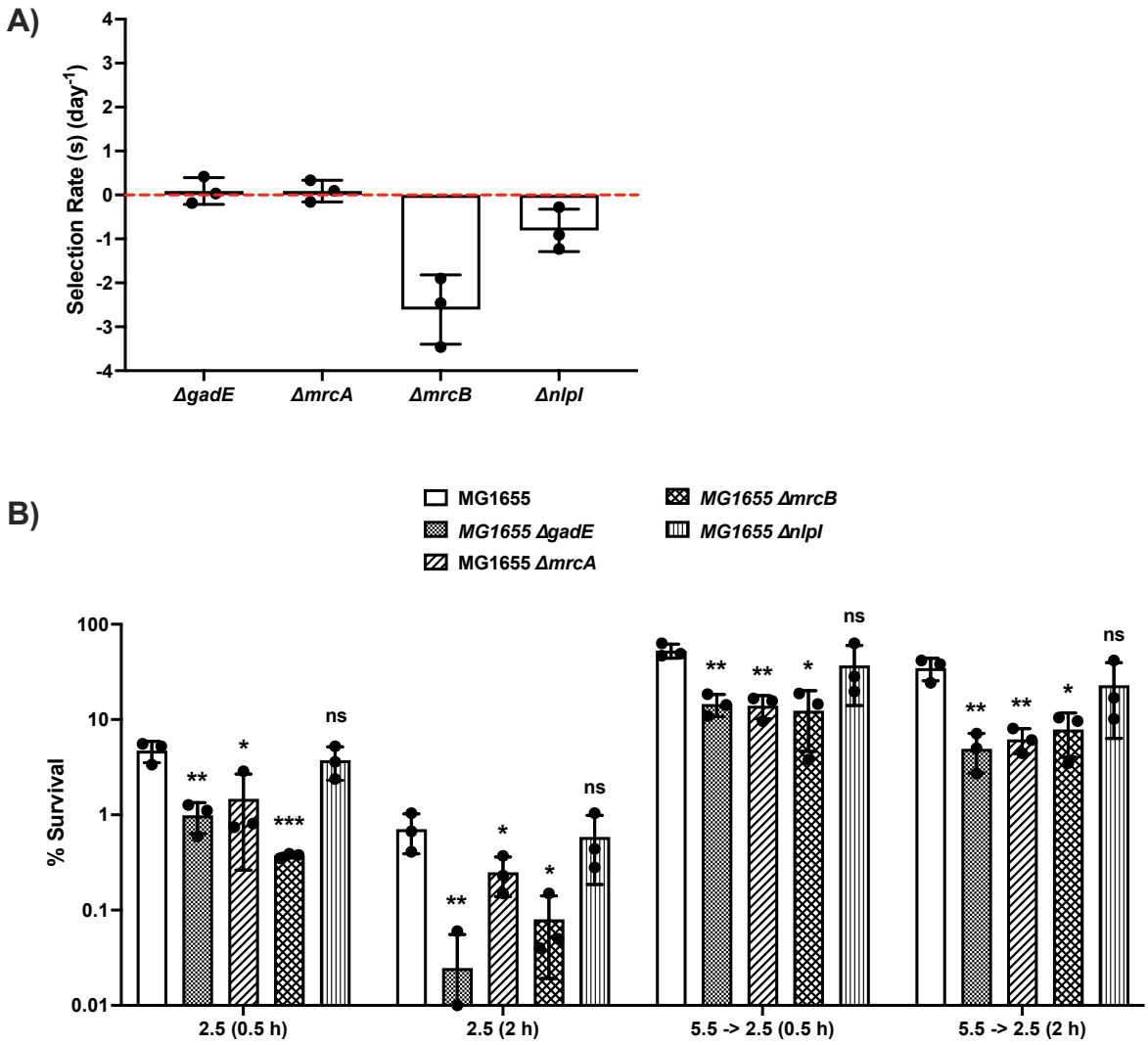


Figure 5.13: Validation of TraDIS experiment results using competition and survival assays. (A) MG1655 *ΔgadE/ΔmrcA/ΔmrcB/ΔnlpI* and KH001 were cultured together in M9Ksupp pH 7 medium to mid-exponential phase and plated on maConkey agar plates. The selection rate (s) was calculated relative to MG1655, as described in Material and Methods, section 2.12.1. (B) MG1655, MG1655 *ΔgadE*, MG1655 *ΔmrcA*, MG1655 *ΔmrcB*, and MG1655 *ΔnlpI* were cultured to mid-exponential phase and acid shocked at pH 2.5 for 30 minutes and 2 h, with and without prior induction at pH 5.5 for 30 minutes. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns = not significant, *p < 0.05, **p < 0.01, ***p < 0.001 (Unpaired t-test, followed by Benjamini, Krieger, and Yekutieli multiple comparisons test).

transposon mutant showed a significant decline in fitness at pH 7. MG1655 *ΔmrcA* also did not show any difference in fitness relative to MG1655, which to an extent fits with the TraDIS results where it only showed a small decline in fitness at pH 7. MG1655 *ΔmrcB* and MG1655 *ΔnlpI,* however, both had a lower fitness than MG1655. The results of *mrcB* are consistent with the TraDIS results where *mrcB* transposon mutant showed a significant reduced fitness at pH 7. The *nlpI* results, however, are in contrast with TraDIS results because *nlpI* transposon mutant only showed a minor decline in fitness at pH 7. It can therefore be concluded that TraDIS can give false positive and false negative results. Further experiments are always necessary to confirm the validity of TraDIS results.

To check whether the function of these genes is needed for the optimal fitness of *E. coli* at pH 2.5, as suggested by the TraDIS results, the survival of *gadE, mrcA, mrcB* and *nlpI* at pH 2.5 was measured using survival assays (Figure 5.13 B). The mutants were grown separately to mid-exponential phase, and acid shocked at pH 2.5 for 30 minutes (same duration as TraDIS experiments) and 2 h, with and without prior induction at pH 5.5 for 30 minutes, before plating them out. The survival of the mutants with induction in the presence of indole was not tested because none of the three mutants (*mrcA, mrcB,* and *nlpI*) showed a significant difference in fitness at pH 2.5 in the presence of indole, and ethanol. MG1655 and MG1655 *ΔgadE* were both used as controls in survival assays.

MG1655 *ΔgadE* had a significantly lower survival than MG1655 when cells were pre-incubated at pH 5.5 before being acid shocked. Since GadE is required for the induction of AR2 response, this result is as expected, and confirms the validity of the method used.

TraDIS results showed that there is no significant difference in the fitness of *mrcA, mrcB* and *nlpI* mutants at pH 2.5 in the presence of indole, and ethanol (Table 5.3). Since indole blocks the induction of AR2 response, this suggests that the function of these genes is

important for the ability of cells to survive exposure to pH 2.5. But their function is not required for the induction of AR2 response. Two of the mutants, *ΔmrcA and ΔmrcB*, had a significantly lower survival than MG1655 after being acid shocked at pH 2.5 (Figure 5.13 B), which is consistent with the TraDIS results. However, the survival of these mutants was also significantly lower than MG1655 when the cells were pre-incubated at pH 5.5 before being acid stressed. Both *mrcA* and *mrcB* encode penicillin binding proteins 1A and 1B respectively, which are required for the synthesis of cell wall peptidoglycan (PG). PG maintains the cell shape and integrity which is vital for the proper functioning of the cell. It is highly likely that the function of *mrcA* and *mrcB* is not directly involved in AR2 response but it is needed for the optimal fitness of cells at low pH due to their role in cell wall synthesis.

ΔnlpI did not show a significant difference in survival than MG1655 at pH 2.5 both with and without induction (Figure 5.13 B). These results are in contrast with TraDIS results where inserts in *nlpI* decreased the fitness of cells at pH 2.5 (Table 5.3).

In conclusion, the results for competition and survival assays suggest that the function of *mrcA* does not affect the growth of cells at pH 7 but it is needed for the optimal fitness of cells at pH 2.5. This is consistent with the TraDIS results because inserts in *mrcA* only declined the fitness of mutants significantly at pH 2.5. The competition and survival assays show that the function of *mrcB* is needed for the optimal fitness of cells at pH 7 and pH 2.5. This is also consistent with the TraDIS results where inserts in *mrcB* reduced the fitness of cells under all four conditions. Although it is possible that the decline in fitness of *mrcB* mutants at pH 2.5 is a result of their impaired ability to grow at pH 7 rather than a direct connection between the role of *mrcB* and pH 2.5. The competition and survival assays show that the function of *nlpI* is needed for the optimal fitness of cells at pH 7 but it does not affect the fitness of cells at pH 2.5. This is in contrast to TraDIS results where inserts in *nlpI* only showed a small decline in

fitness at pH 7 but a bigger decline in fitness at pH 2.5. The validation results of all four genes: *gadE, mrcA, mrcB* and *nlpI*, along with their TraDIS results, are shown in Table 5.7.

5.3 Discussion

The main aim of the study described in this Chapter was to identify genes whose function is needed for the optimal fitness of *E. coli* at pH 2.5, and genes that are involved in the inhibitory effect of indole on AR2 induction. Mutants containing insertions in several AR2 genes such as *evgS, evgA*, *ydeP, slp*, *gadE, hdeA, hdeB, yhiD, hdeD, gadB,* and *gadC* showed the biggest decline in fitness at pH 2.5 in the presence of ethanol. This is as expected, because the AR2 system provides resistance against extreme acid stress of pH 2.5 and below. The decline in the fitness of these genes at pH 2.5 suggests that the protocol used for TraDIS experiments can successfully identify genes whose function is needed for optimal fitness of cells at pH 2.5.

TraDIS results suggest that the function of several additional genes including *clsA, mrcA, yciB, emrR,* and *lpxM* is needed for optimal fitness of *E. coli* at pH 2.5. Furthermore, the function of *atp* genes which encode components of F_1F_o ATP synthase is also needed for optimal fitness of cells at pH 2.5. The loss of function of *wec* genes, on the other hand, which are involved in LPS synthesis, is suggested to improve the fitness of cells at pH 2.5. To our knowledge, the function of all these genes, except for *atp* genes, has not previously been reported to affect the fitness of cells at low pH. Competition and survival assays done with *mrcA* mutant demonstrated that the function of *mrcA* does not affect the growth of cells at pH 7 but is needed for the optimal fitness of cells at pH 2.5, confirming the TraDIS results.

Table 5.7: A summary of results from TraDIS experiments done with MG1655 transposon library and the validation experiments (competition and survival) done with MG1655 single gene mutants. Genes whose function is needed for the optimal fitness of MG1655 at pH 7 or pH 2.5 according to TraDIS and validation experiments is shown.

Gene	TraDIS results		Validation results	
	pH 7	pH 2.5	pH 7	pH 2.5
gadE	✓	✓	X	✓
mrcA	X	✓	X	✓
mrcB	✓	✓	✓	✓
nlpI	X	✓	✓	X

TraDIS data for *fis* gene showed interesting results. This is the only gene where inserts caused (a) a significant loss of fitness at pH 2.5 in the presence of indole, but no effect in the presence of ethanol, (b) a significant increase in fitness at pH 7 in the presence of indole but again, no effect in the presence of ethanol. These results suggest that is an interesting connection between indole and the function of Fis. Further work such as competition and survival assays are needed to validate these results.

Chapter 6

Using Tradis to study the induction of acid resistance response, and its inhibition with indole, in O157:H7 Sakai

Declaration

The optimisation experiments (not shown in the Chapter) for preparing the TraDIS-Xpress library in O157:H7 Sakai were done by the Masters student Nina Perry, with guidance from myself and Dr Mathew Milner. The final library was prepared by all three of us. The essential gene analysis in O157:H7 Sakai and pangenome analysis in MG1655, BW25113, and O157:H7 Sakai was done by Nina Perry with guidance from Dr Mathew Milner. The optimisation, final experiments, and data analysis to study acid stress in O157:H7 Sakai was done by myself.

6.1 Introduction

The main aims of this Chapter were (a) to find genes that may be important in the ability of *E. coli* O157:H7 Sakai to survive at pH 2.5, (b) to find genes that are involved in the inhibitory effect of indole on AR2 induction in O157:H7 Sakai, and (c) to compare the acid stress response between K-12 laboratory strain MG1655, and pathogenic strain O157:H7 Sakai. In previous Chapter, the results of TraDIS experiments done in MG1655 were described. In this Chapter, first the process of preparing a transposon library in O157:H7 Sakai is described. Following this, a brief pangenome analysis done on MG1655, BW25113, and O157:H7 Sakai is described. Finally, the results of TraDIS experiments done in O157:H7 Sakai, and the comparison of TraDIS results between MG1655 and O157:H7 Sakai are presented.

This is the first study to use TraDIS to study the effect of low pH in *E. coli*. It was shown in Chapter 5 that TraDIS was successfully able to identify genes which are known to be involved in low pH response, in the conditions used in our experiments. Furthermore, several genes whose function or loss of function has previously not been shown to be required for the optimal fitness of *E. coli* at pH 2.5 were also found. In this Chapter, we will extend this approach to find genes whose function or loss of function is needed for the optimal fitness of *E. coli* O157:H7 Sakai at pH 2.5 and genes which are involved in inhibiting the AR2 response through indole. Furthermore, using TraDIS results, we will compare the acid stress response in MG1655 with that of O157:H7 Sakai. The response of *E.coli* to low pH has been extensively studied but mostly using the laboratory strain MG1655. So the comparison between MG1655 and O157:H7 Sakai will help to deduce how much of what has been studied in laboratory strains can be applied to pathogenic strains. The outcomes of this comparison will also help to determine the extent to which MG1655 can be used as a model for the response to low pH in other *E. coli* strains. If this comparison shows that MG1655 is not the best model, then any

future studies focussing on studying acid stress in *E. coli* should aim to include real-world strains to understand the true impact of their findings. Of course, before any firm conclusions can be drawn, more pathogenic strains will need to be studied and compared with the laboratory strains. But this study will provide an optimised protocol for using a high-throughput technique such as TraDIS to study acid stress in *E. coli* at genome level which can be time-saving.

6.2 Results

6.2.1 Preparation of TraDIS-Xpress library in O157:H7 Sakai

The O157:H7 Sakai TraDIS-Xpress library was prepared using a mini-Tn5 transposon carrying an outward-facing *tac* promoter present downstream of the kanamycin resistance gene (Yasir *et al.*, 2020). The *tac* promoter can be induced through addition of IPTG (Isopropyl ß-D-1-thiogalactopyranoside) to the growth medium allowing the genes present downstream of the transposon to be either overexpressed or repressed. If the *tac* promoter is in the same orientation as the gene, then this leads to the overexpression of the gene. However, if the *tac* promoter is in the opposite orientation to the gene, then the anti-sense mRNA produced through *tac* promoter binds to the sense mRNA leading to gene repression. The advantage of using a TraDIS-Xpress library over a normal TraDIS library is that it allows the conditional essentiality of the essential genes, which cannot tolerate transposon insertions, to be determined.

To prepare the O157:H7 Sakai TraDIS-Xpress library, ultracompetent O157:H7 Sakai cells were prepared and transformed with the transposome as described in Materials and Methods, section 2.11.3. A transposome is a mixture of transposon DNA and the EZ-Tn5 transposase enzyme which recognises the 19-bp mosaic ends on either side of the transposon

DNA before binding it. Once the transposome is electroporated into the cells, the EZ-Tn5 transposase catalyses the insertion of transposon into the host DNA at random sites with a high efficiency. Once the insertions have taken place, they cannot move to new sites. This is because the transposon has no Tn5 transposase gene, so the Tn5 transposase enzyme cannot be produced by the cells to catalyse further transposition events.

Once the transformations are done, the cells are plated on a selective medium which allows the transformants to be selected. The Tn5 transposon carries a kanamycin resistance marker, so the transformants containing Tn5 transposon insertions can be selected on a medium supplemented with kanamycin. Therefore, prior to the transformations, the minimum inhibitory concentration (MIC) of kanamycin for O157:H7 Sakai was determined by plating the strain on successively higher levels of kanamycin (7.5, 15, 30 and 50 μg/mL). The concentration of 50 μg/mL resulted in no survivors. Therefore, the transformants were plated on LB agar plates supplemented with 50 μg/mL of kanamycin which would prevent the growth of satellite colonies. A trial run was done with 1 transformation which resulted in 2618 transformants. Followed by the trial run, 44 additional transformations were performed resulting in an average of 6,360 transformants per transformation. A total of approximately 282,000 transposon mutants were obtained. O157:H7 Sakai has a genome size of ~5.5 Mb and assuming that each transposon mutant has the transposon inserted at a unique site in the genome, this gives approximately 1 UIS every 19.5 bps.

The colonies from all transformations were pooled and two separate aliquots (ITL1 and ITL2) were sequenced. A total of 9,660,637 sequencing reads were aligned to the reference genome (NCBI database: NC_002695.2) for ITL1 and ITL2 combined. The number of unique insertion sites (UIS) present was 141,312. With the genome size of 5,498,578 bp, O157:H7 Sakai transposon library therefore has a UIS present every 39 bps. As stated

previously, it was expected that the final transposon library will have a UIS every 19.5 bps. This number was based on the fact that transposon in every cell will get inserted at a unique site. The reason behind the discrepancy between these two numbers could be because the final transposon library contains multiple insertions in the same sites.

The O157:H7 Sakai transposon library is less dense than the MG1655 transposon library which had a UIS every 12.2 bps. The transposon insertions in O157:H7 Sakai transposon library were distributed throughout the genome (Figure 6.1 A). The genomic regions around origin of replication (ori) are likely to have a higher frequency of insertions than the genomic regions away from ori. Furthermore, DNA sequences which either have an activator or a repressor protein bound to them are also likely to have a low frequency of insertions. The reproducibility among replicates was checked using RPM and insertion index score, both of which had a high Spearman's correlation coefficient of over 0.93 (Figures 6.1 B and C).

6.2.2 Assessment of essential genes in O157:H7 Sakai TraDIS-Xpress library

The insertion index score of all genes from O157:H7 Sakai TraDIS-Xpress library was plotted on a histogram where it can be seen to follow a bimodal distribution (Figure 6.2). An exponential distribution was fitted to the left mode and a gamma distribution was fitted to the right mode. Log_2 likelihood ratio (LLR) for all genes was calculated and genes were assigned either to the left or right mode depending on their LLR. Genes with LLR < -3.6 were considered essential, genes with LLR > 3.6 were considered non-essential and those with LLR between -3.6 and 3.6 were considered unclear. A total of 299 essential, 4,469 non-essential and 294 unclear genes were found in O157:H7 Sakai TraDIS-Xpress library.

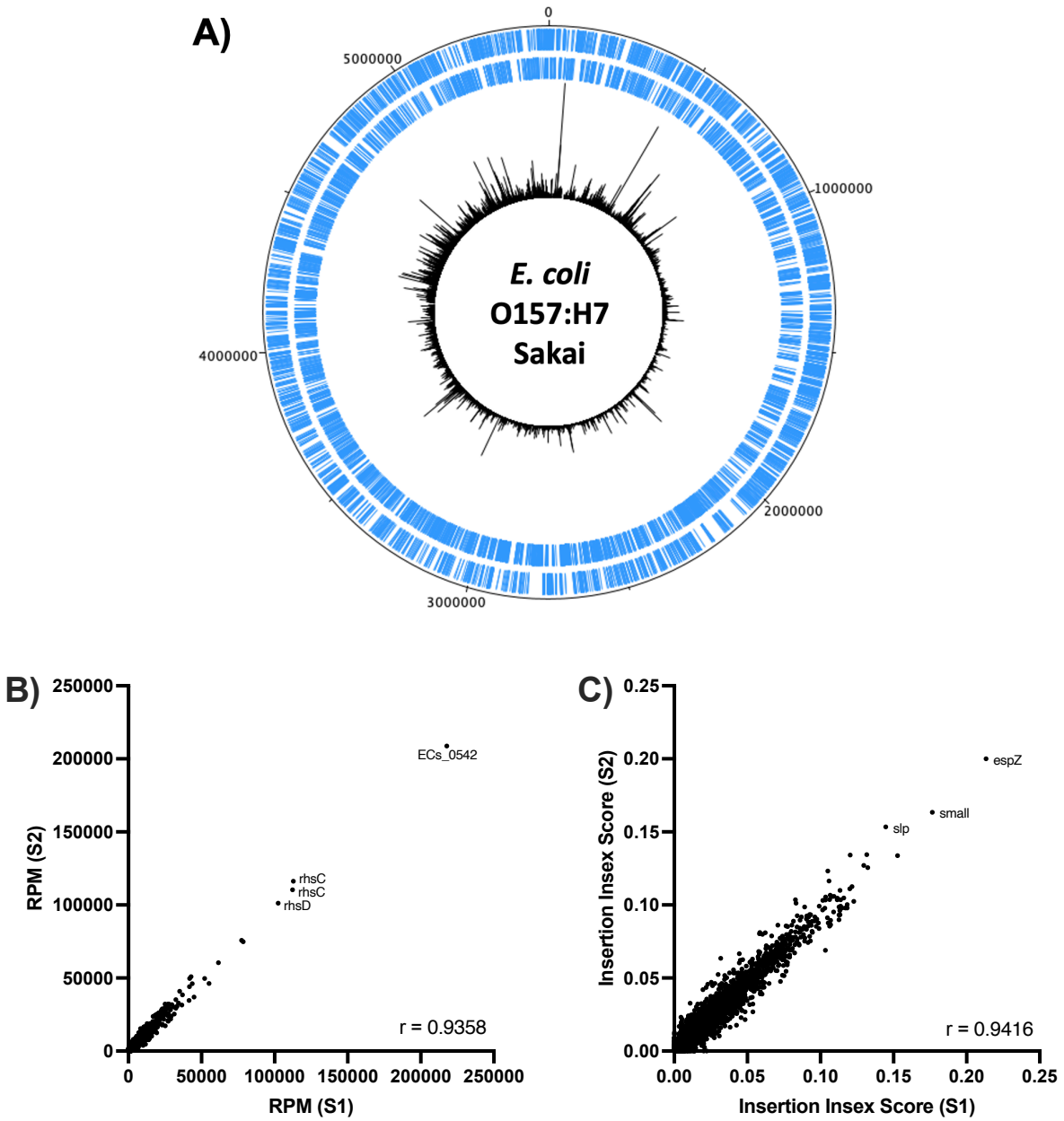
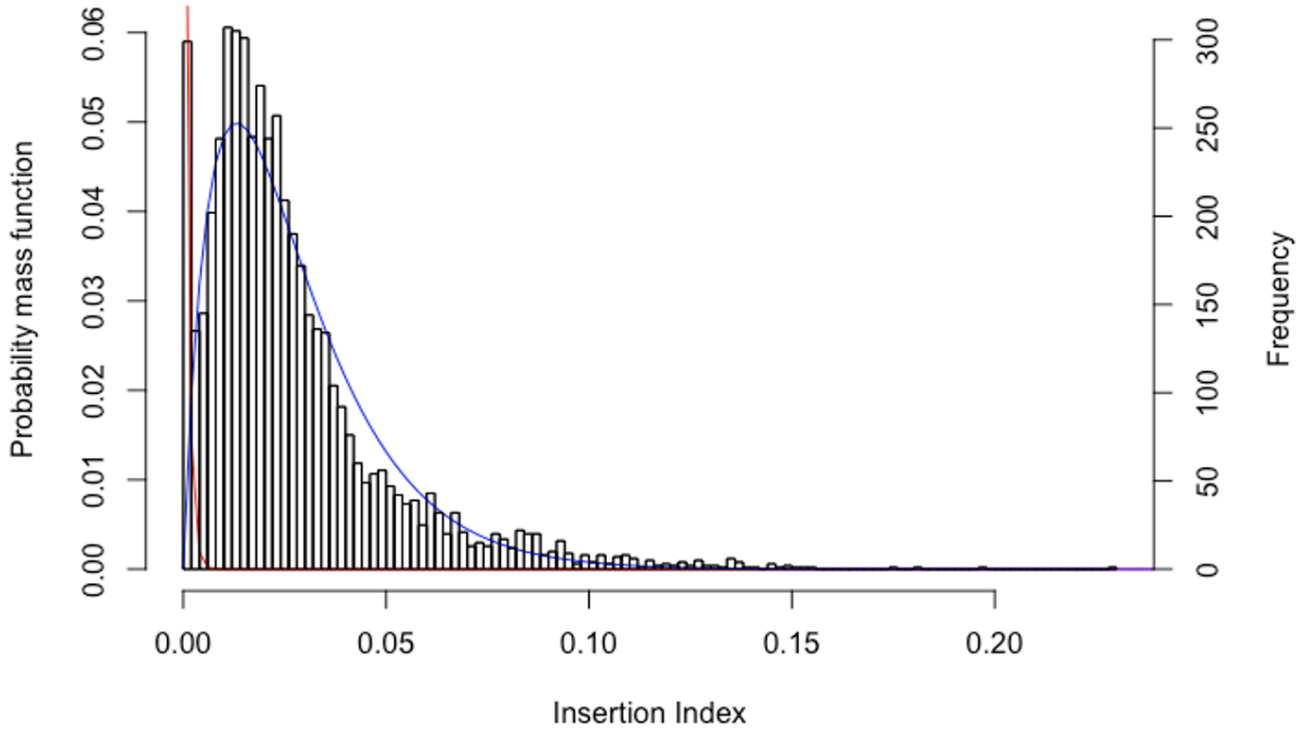


Figure 6.1: The position of unique transposon insertion sites and reproducibility between two biological replicates of O157:H7 Sakai TraDIS sequencing libraries.
(A) The position of unique transposon insertion sites across the O157:H7 Sakai genome. The outermost numbers correspond to the genome locations in base pairs. The next two blue tracks correspond to the sense and antisense coding strands respectively. The innermost black track corresponds to the position and frequency of transposon inserts. (B-C) Comparison between RPM (B) and Insertion index score (C) of two independently prepared initial TraDIS sequencing libraries. The correlation was determined using Spearman's correlation coefficient which is shown at the bottom right corner.



	Number of genes
Essential	299
Non-essential	4469
Unclear	294

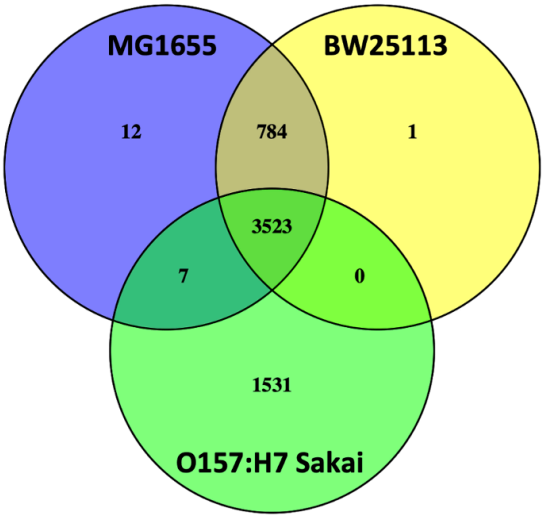
Figure 6.2: Distribution and frequency of insertion index score of genes from O157:H7 Sakai TraDIS-Xpress library.
The genes in left mode (red) were considered essential whereas those in right mode (blue) were considered non-essential. The number of essential, non-essential, and unclear genes in O157:H7 Sakai transposon library are given in the table on the right.

6.2.3 Pangenome analysis of K-12 strains and O157:H7 Sakai

Pangenome analysis using Roary as described in Materials and Methods, section 2.11.6.3, was done to determine the genes which are present and absent across three *E. coli* strains: MG1655, BW25113 and O157:H7 Sakai. 3,523 genes were present in all three strains and make up the core genome of these strains (Figure 6.3 A). 2,335 genes were present in either one or two of the three strains and make up the accessory genome of these strains. Pangenome analysis done on 1,324 *E. coli* genomes has revealed that ~25,000 genes are present in those *E. coli* strains. Out of those 25,000 genes, ~420 genes make up the core genome of those strains (Tantoso *et al.*, 2022). The difference in the number of genes that make up the core genome in two studies is because our study has only included three genomes. The core genome gets smaller as the number of genomes included in the pangenome analysis increase.

TraDIS can be used to determine which genes are essential and which genes are non-essential in a genome. In order to determine genes which are essential in one genome but not the others, the TraDIS data can be combined with Roary pangenome analysis as described in Figure 6.3 B. The TraDIS data on essential, non-essential, and unclear genes determined in MG1655 (Milner, 2021), BW25113 (Goodall *et al.*, 2018), and O157:H7 Sakai (this study) was overlapped with the Roary output of this study in order to determine genes which are essential in one strain but not the other two (Tables 6.1 and 6.2). For this analysis, only essential and non-essential genes were included, and unclear genes were excluded due to the uncertainty associated with them. Out of 446 genes which were essential in at least one strain, 395 genes were present in all three strains. 222 genes out of 395 were essential across all three strains and the remaining 173 were essential either in one or two of the strains. 51 genes which were essential in at least one strain, were also absent in at least one strain. 26

A)



B)

	BW25113	MG1655	O157:H7 Sakai	Roary Output
Gene 1				Core
Gene 2				Core
Gene 3			*Absent*	Accessory

TraDIS output → Essential Non-essential Unclear

Figure 6.3: Pangenome analysis performed on *E. coli* strains.
(A) The number of genes present or absent in comparison between MG1655, BW25113 and O157:H7 Sakai, determined using Roary pangenome analysis. (B) Illustration of how Roary pangenome output can be combined with TraDIS output to determine the core and accessory genome of *E. coli* strains (Adapted from Perry, 2022; Figure 8).

Table 6.1: The number of genes designated as essential, non-essential, or unclear in three *E. coli* strains, MG1655 (Milner, 2021), BW25113 (Goodall *et al.,* 2018) and O157:H7 Sakai (this study) determined using TraDIS.

Number of genes	BW25113	MG1655	O157:H7 Sakai
Essential	359	346	299
Non-essential	3756	3734	4469
Unclear	159	204	294
Total	**4274**	**4284**	**5062**

Table 6.2: The number of essential genes which are present in all three genomes (core genome) or absent in at least one genome (accessory genome).

Essential in at least one strain (446)

CORE GENOME Present in all three strains, Essential in at least one strain (395)		**ACCESSORY GENOME** Essential in at least one strain but absent in at least one strain (51)	
Essential in MG1655	10	Present in one or both K-12 strains	25
Essential in BW25113	27	Present only in O157:H7 Sakai	26
Essential in O157:H7 Sakai	38		
Essential in MG1655 and BW25113	85		
Essential in MG1655 and O157:H7 Sakai	7		
Essential in BW25113 and O157:H7 Sakai	6		
Essential in all three strains	222		

of these genes were absent in O157:H7 Sakai and 25 were absent from one or both of the K-12 strains.

6.2.4 Optimisation of conditions for TraDIS experiments in O157:H7 Sakai

6.2.4.1 Optimisation of acid shock duration at pH 2.5

One of the aims of this study is to compare how different strains respond to acid stress and how indole inhibits the induction of AR2 response in these strains. To answer this question, TraDIS was performed on a laboratory K-12 strain MG1655, the results of which were discussed in Chapter 5. In this Chapter, the results of TraDIS experiments performed with O157:H7 Sakai will be presented and discussed. The duration for which O157:H7 Sakai should be acid shocked at pH 2.5, following pre-incubation at pH 5.5 in the presence of indole and ethanol, was optimised to achieve levels of survival which are comparable to MG1655. This was crucial so that a direct comparison can be made between the TraDIS results of both strains later. O157:H7 Sakai was grown to mid-exponential phase, induced at pH 5.5 in the presence of indole or ethanol separately, and acid stressed at pH 2.5 for different durations (namely, 0.5, 1, 1.5 or 2 h). The survival of indole-treated cultures acid stressed at pH 2.5 was less than ethanol-treated cultures, consistent with our hypothesis that indole inhibits the induction of the AR2 response (Figure 6.4). After 1.5 h of acid stress at pH 2.5, the survival of indole- and ethanol-treated cultures was 9.4% and 21.5% respectively. These results are closer to the ones obtained with the MG1655 transposon library (9.2% and 33.1% for indole- and ethanol-treated cultures respectively, Figure 5.6). Therefore, the duration of 1.5 h for acid stress was selected for the TraDIS experiments in O157:H7 Sakai.

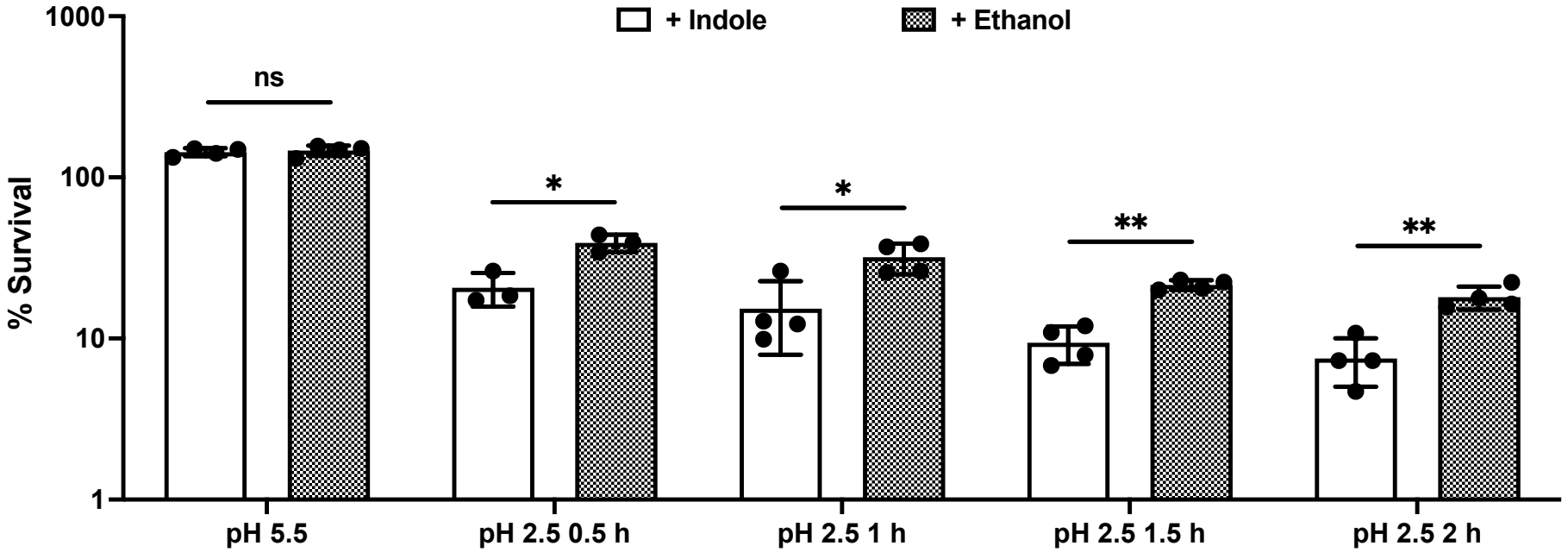


Figure 6.4: Survival of O157:H7 Sakai under AR2 induction conditions in the presence of 1 mM indole or the same volume of ethanol, followed by acid stress at pH 2.5 for different durations.

50 mL of O157:H7 Sakai culture in a 250 mL shaker flask was grown to mid-exponential phase, induced at pH 5.5 with either indole or ethanol, followed by acid shock at pH 2.5 for 0.5, 1, 1.5 and 2 h. The survival was calculated as described in Materials and Methods, section 2.5. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown. ns = not significant, *$p < 0.05$, **$p < 0.01$ (Unpaired t-test, followed by Benjamini, Krieger, and Yekutieli multiple comparisons test)

6.2.4.2 Optimisation of culture outgrowth time after acid shock at pH 2.5

As explained previously in Chapter 5, section 5.2.1.2, during TraDIS experiments the cultures need to be outgrown after acid shock step at pH 2.5. This is to allow the acid resistant mutants, that have survived the acid shock, to increase in frequency whereas the frequency of acid sensitive mutants either remains the same or changes to a smaller extent relative to acid resistant mutants. Furthermore, the outgrowth step allows the mutants to increase in numbers so that enough cells can be obtained for sufficient DNA yield required to prepare the sequencing libraries.

For TraDIS experiments in MG1655, the cultures were outgrown for 3 hours after the acid shock step at pH 2.5. To ensure that 3 hours of outgrowth duration was also appropriate in O157:H7 Sakai, the O157:H7 Sakai transposon library was induced with either indole or ethanol, acid stressed at pH 2.5 for 1.5 h and outgrown for 3 h in M9Ksupp pH 7 medium. The CFU mL^{-1} of the indole- and ethanol-treated libraries was monitored during outgrowth. The survival of indole- and ethanol-treated libraries after acid shock at pH 2.5 was 12% and 17.9% respectively (Figure 6.5 A). Both indole- and ethanol-treated libraries showed a steady increase in growth, which shows that the treatment applied during induction (indole or ethanol) does not affect the growth of the cells during outgrowth (Figure 6.5 B).

The outgrowth duration should be long enough so that the mutants which have survived the acid stress can increase in frequency by sufficient amount so that they can be differentiated from the mutants which have not survived the acid stress.. The CFU mL^{-1}, after 0 and 3 hours of outgrowth, of O157:H7 Sakai induced in the presence of ethanol followed by acid shock was 1.0×10^7 and 1.3×10^8, which is about 4 doublings. With 4 doublings, the CFU mL^{-1} of mutants containing insertions within any particular gene will increase by 16-fold, assuming that the cells are growing exponentially. An outgrowth duration of 3 hours is

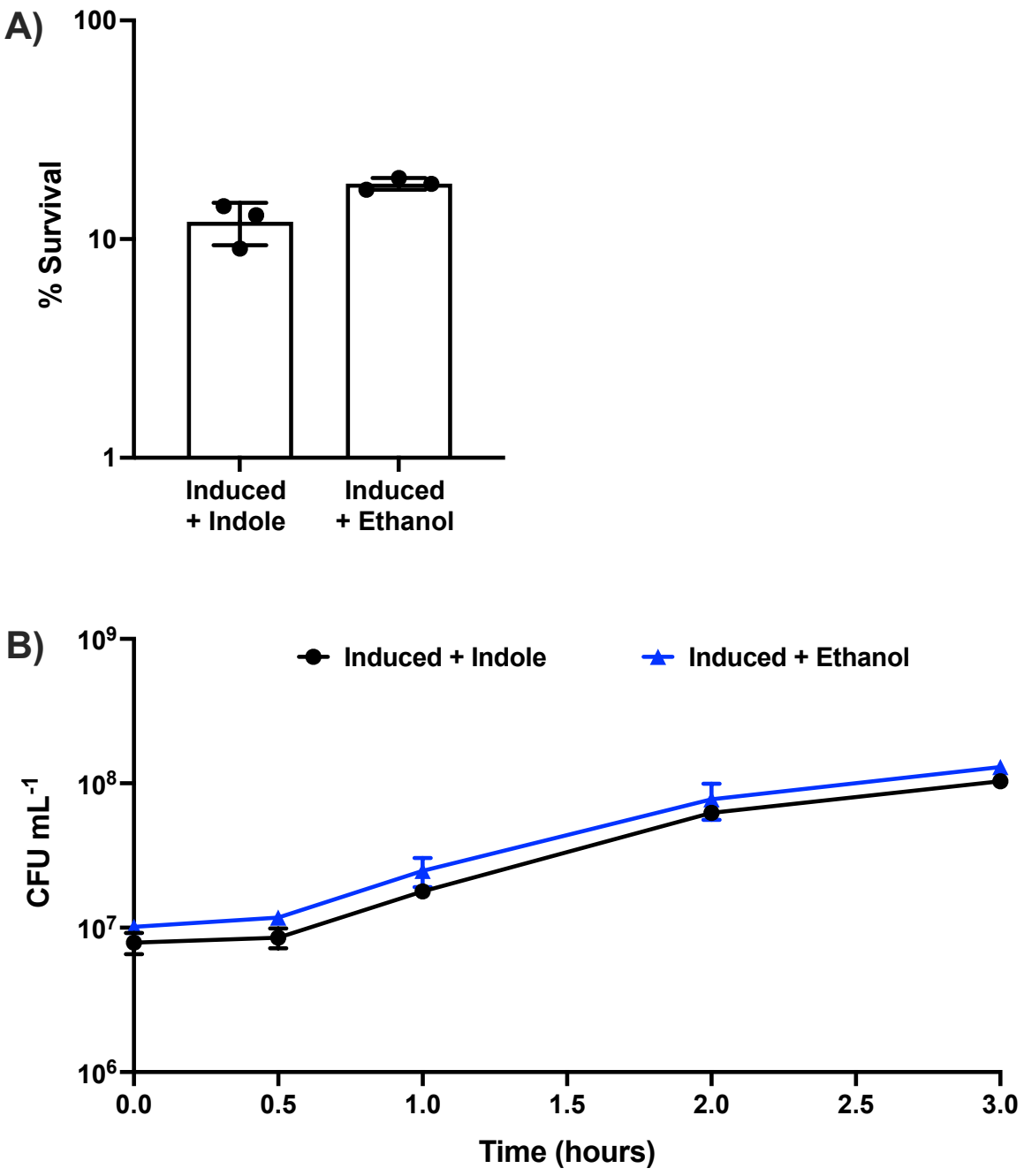


Figure 6.5: Survival of O157:H7 Sakai under AR2 induction conditions in the presence of 1 mM indole or the same volume of ethanol, followed by outgrowth.
50 mL of M9K supp pH 7 medium inoculated with O157:H7 Sakai transposon library, in a 250 mL shaker flask was grown to mid-exponential phase, induced at pH 5.5 with either indole or ethanol, followed by acid shock at pH 2.5 for 90 minutes. The acid stressed cultures were outgrown for 3 hours and plated out after 0.5, 1, 2, and 3 h. (A) Survival of O157:H7 Sakai transposon library after induction and acid shock. (B) Increase in CFU mL^{-1} of O157:H7 Sakai transposon library during outgrown. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

therefore sufficient for mutants which have survived the acid stress to grow and increase in frequency.

The DNA extraction protocol recommends that at least 10^9 cells are processed for good DNA yield. After 3 hours of outgrowth, ~ 1 x 10^8 CFU mL^{-1} were present in the indole-treated culture, which is equal to 5 x 10^9 CFU in a volume of 50 mL. This number of CFU is sufficient to get a good DNA yield for preparing the sequencing library.

6.2.5 Results of TraDIS experiments

6.2.5.1 Workflow of TraDIS experiments

The aim of the TraDIS experiments described in this chapter was to (1) identify genes that are important for survival of *E. coli* at pH 2.5, (2) potentially detect genes that could be involved in the inhibitory effect of indole on AR2 induction, and (3) compare the acid stress responses and the effect of indole on AR2 induction between MG1655 and O157:H7 Sakai.

To be able to draw direct comparisons between the TraDIS results of MG1655 and O157:H7 Sakai, the TraDIS experiments in O157:H7 Sakai were done under the same conditions as MG1655, as previously described in Chapter 5, section 5.2.2.1 (Figure 5.4). The O157:H7 Sakai transposon library was cultured to mid-exponential phase (OD_{600} of ~0.2), induced at pH 5.5 either in the presence of indole or ethanol, followed by acid stress at pH 2.5 or pH 7 for 90 mins (as opposed to 30 mins for MG1655 TraDIS library) and finally outgrown for 3 hours in M9K supp pH 7 medium. The sequencing libraries prepared under these conditions were named as 5.5 + I -> 2.5, 5.5 + E -> 2.5, 5.5 + I -> 7, and 5.5 + E -> 7, where I and E stand for induction in the presence of indole or ethanol respectively. 2.5 and 7 stand for the acid stress done either at pH 2.5 or pH 7 (i.e., the non-stressed control).

The rationale for performing the experiments under these conditions was explained in detail in Chapter 5, section 5.2.2.1. Briefly, all four libraries can be compared with the ITL and genes whose function is needed for the optimal fitness of *E. coli* at pH 7 and pH 5.5 can be determined. Then by comparing 5.5 + I -> 2.5 and 5.5 + E -> 2.5 libraries with each other, after comparison with ITL, the genes whose function or loss of function is required for the optimal fitness of *E. coli* at pH 2.5 can be determined.

To determine genes which are involved in the inhibitory effect of indole on AR2 induction, 5.5 + I -> 2.5 library is compared with 5.5 + I -> 7 library. The frequency of the mutants containing insertions in genes that increase in either one or both of these libraries is determined. These genes are candidates for ones that may be involved in the inhibitory effect of indole on AR2 response.

6.2.5.2 Survival results of TraDIS experiments

When TraDIS experiments were done, the survival of O157:H7 Sakai transposon library exposed to each condition was calculated as a quality control (Figure 6.6). The library had a survival of 6.5% in 5.5 + I -> 2.5, 6.9% in 5.5 + E -> 2.5, 155.6% in 5.5 + I -> 7, and 254.9% in 5.5 + E -> 7 condition. The survival of over 100% can result from the growth that occurs during the induction step at pH 5.5 because the final survival is calculated by dividing the number of colonies at pH 2.5 after 1.5 h by the number of colonies at pH 7 obtained before the pH 5.5 step.

Since indole inhibits the induction of AR2 response, the survival of library acid stressed at pH 2.5 following pre-induction in the presence of indole was expected to be lower than the one pre-induced in the presence of ethanol. As shown in Figure 6.4, when O157:H7 Sakai strain was acid stressed at pH 2.5, the cultures pre-induced in the presence of indole had a

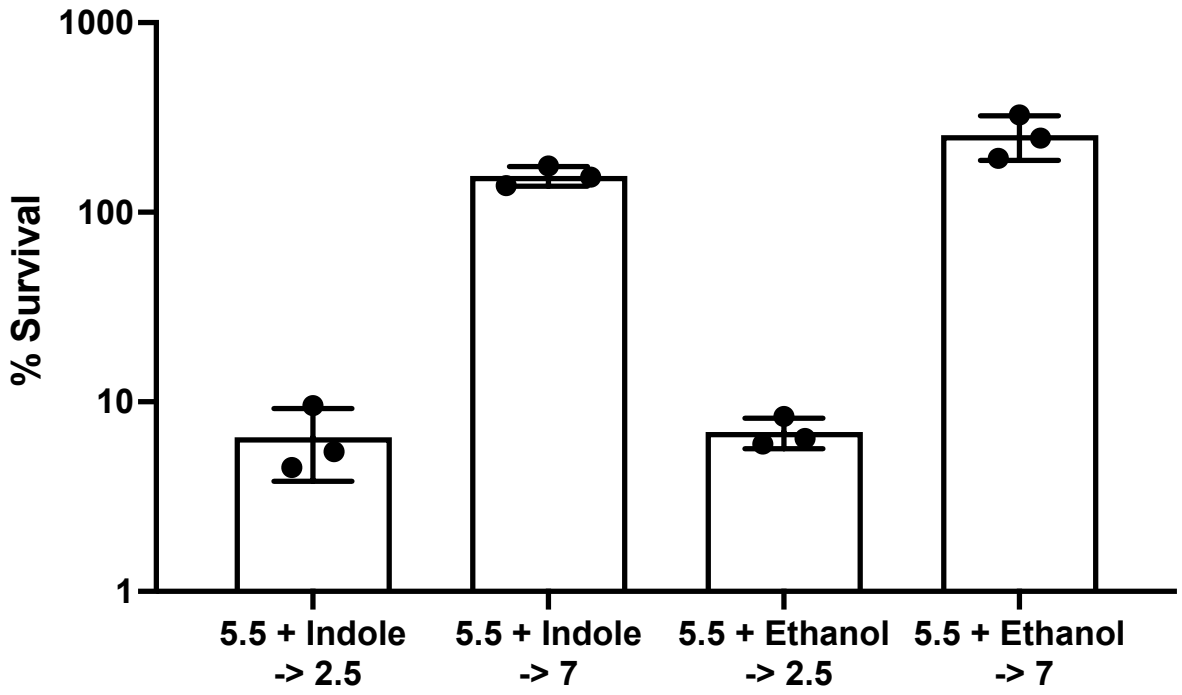


Figure 6.6: Survival of O157:H7 Sakai transposon library induced either in the presence of 1 mM indole or the same volume of ethanol followed by exposure to pH 2.5 or pH 7.
O157:H7 Sakai transposon library was cultured to mid-exponential phase, induced at pH 5.5 with either 1mM indole or the same volume of ethanol, followed by incubation at pH 2.5 or pH 7 for 90 minutes. Individual data points, and error bars corresponding the standard deviation of the replicate data points, are shown.

significantly lower survival than those induced in the presence of ethanol. When the experiment was repeated with O157:H7 Sakai transposon library, the indole-treated cultures had a lower survival than ethanol-treated cultures (Figure 6.5). However, the difference in the survival of indole- and ethanol-treated cultures was smaller than previously observed with O157:H7 Sakai strain after 1.5 h of acid stress at pH 2.5. For the final TraDIS experiments with O157:H7 Sakai transposon library, there was almost no difference in the survival of indole- and ethanol-treated cultures acid stressed at pH 2.5. The survival of indole-treated library (6.5%) was similar to the survival of indole-treated O157:H7 Sakai strain (9.4%). However, it was the ethanol-treated library (6.9%) which showed a much smaller survival than the ethanol-treated O157:H7 Sakai strain (21.5%). These observations suggest that the O157:H7 Sakai transposon library shows a smaller acid resistance than the O157:H7 Sakai strain, following pre-induction at pH 5.5.

6.2.5.3 Confirmation of reproducibility among replicates

After analysing the sequencing data with BioTraDIS pipeline for ITL and each conditional TraDIS library, the total number of read counts and the UIS across all genes were calculated (Figure 6.7). The total number of UIS among the conditional TraDIS libraries was higher than the ITL. This is because the conditional libraries have been sequenced at a greater depth than the ITL, which is evident from the higher read counts of these libraries than the ITL.

The reproducibility among replicate TraDIS sequencing libraries prepared under each condition was checked using principal component analysis (PCA) by comparing the number of read counts per gene normalised by sequencing depth (Figure 6.8). The first replicate of both indole-treated libraries, 5.5 + I -> 2.5 and 5.5 + I -> 7, showed the highest variance as

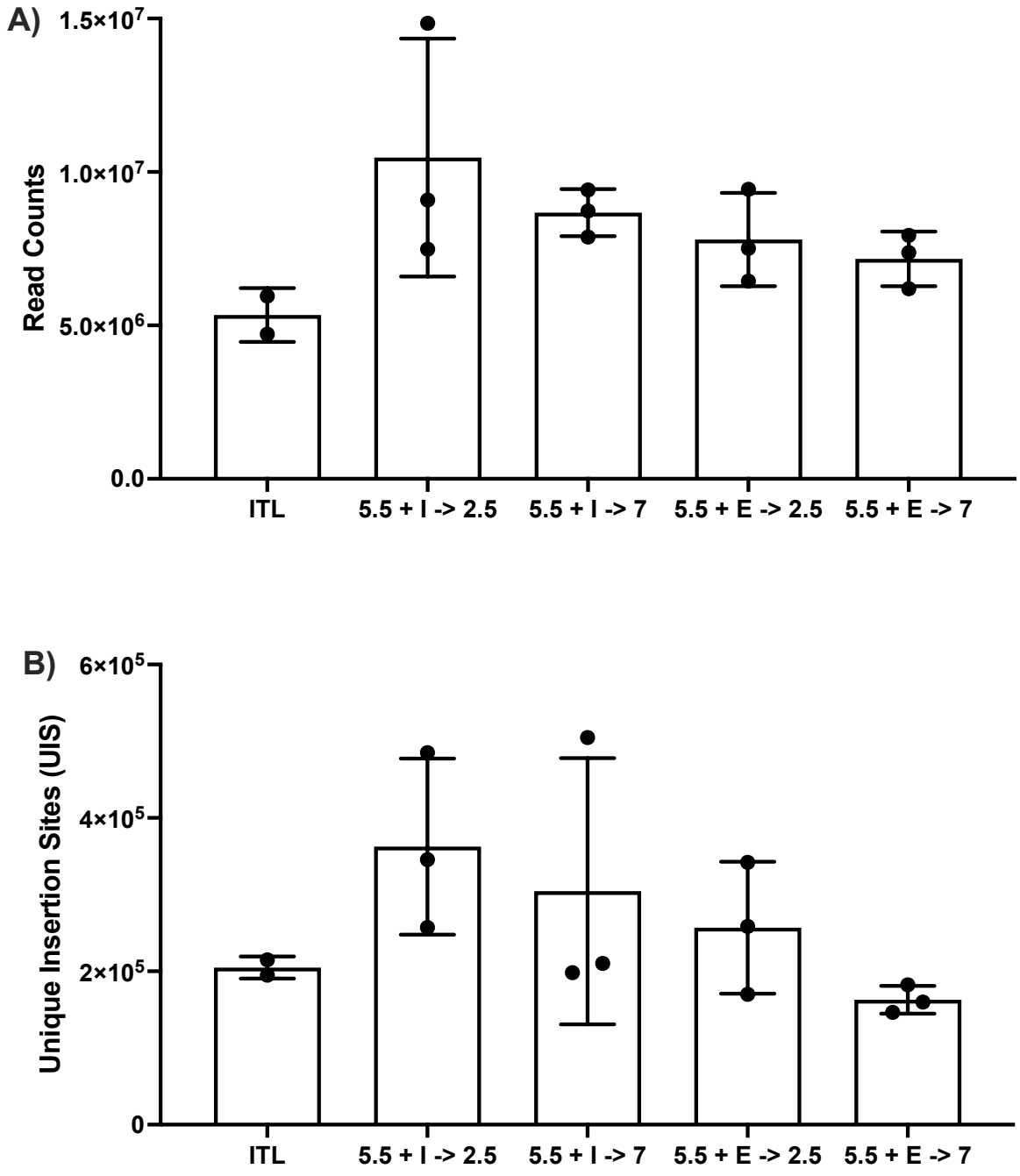


Figure 6.7: Comparison of total number of read counts (A) and unique insertion sites (UIS) (B) among initial TraDIS library (ITL) and the TraDIS libraries prepared under different conditions.

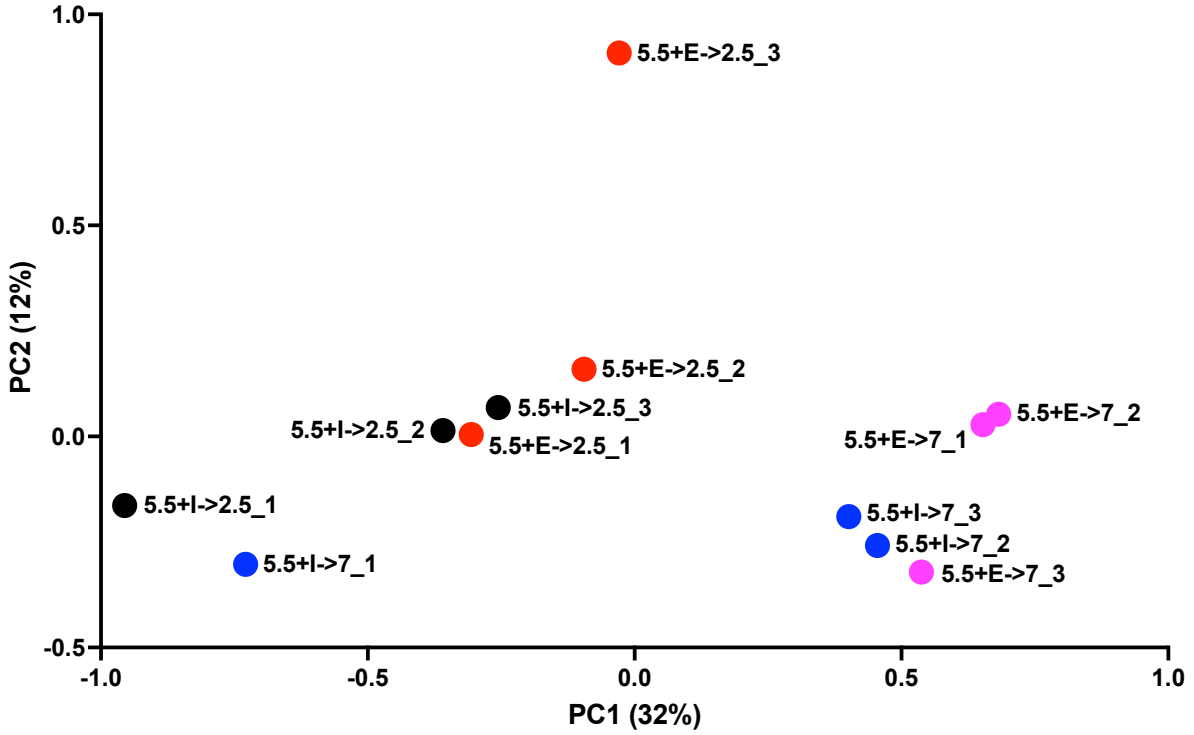


Figure 6.8: Clustering among replicates of same and different conditions visualised through principal component analysis done on read counts per gene normalised by sequencing depth.

The percentage variance explained by first (32%) and second (12%) principal components is shown on the X- and Y-axis respectively.

compared to the other two replicates, in the first component of the PCA analysis. For the ethanol-treated libraries, 5.5 + E -> 2.5 and 5.5 + E -> 7, the third replicate showed the highest variance as compared to the other two replicates, in the second component of the PCA analysis. There was a separation between libraries acid stressed at pH 2.5 from those incubated at pH 7. It is worth noting that the experiments corresponding to the first replicates of both indole-treated libraries were performed on the same day. Furthermore, the experiments corresponding to third replicates of both ethanol-treated libraries were also performed on the same day. Therefore, they could be true outliers or the variation could be a result of human error.

To check whether the variance present among the replicates was due to plasmid genes, PCA analysis was repeated after removing the genes encoded on two plasmids, pO157 and pOSAKI, present in the O157:H7 Sakai strain (data not shown). On repeating the PCA analysis with just the chromosomal genes, the variance among the replicates described above was still present. This suggests that plasmid genes are not responsible for a significant amount of the variance present between the replicates of TraDIS libraries.

The reproducibility among replicates was further checked using RPM and insertion index scores. One plot for RPM and one for insertion index score from ITL and each conditional TraDIS library is shown in Figures 6.9 and 6.10 respectively. The plots for remaining replicates are shown in Appendices 8.3.2.1 and 8.3.2.2. The Spearman's correlation coefficient for both RPM and insertion index score was greater than 0.92 for all plots showing a high reproducibility among replicates of all conditions.

The PCA analysis was repeated by removing the outliers present in the XY plots for RPM. However, this did not remove the variance observed previously between the replicates. This suggests that the variance observed among replicates is spread across the entire dataset,

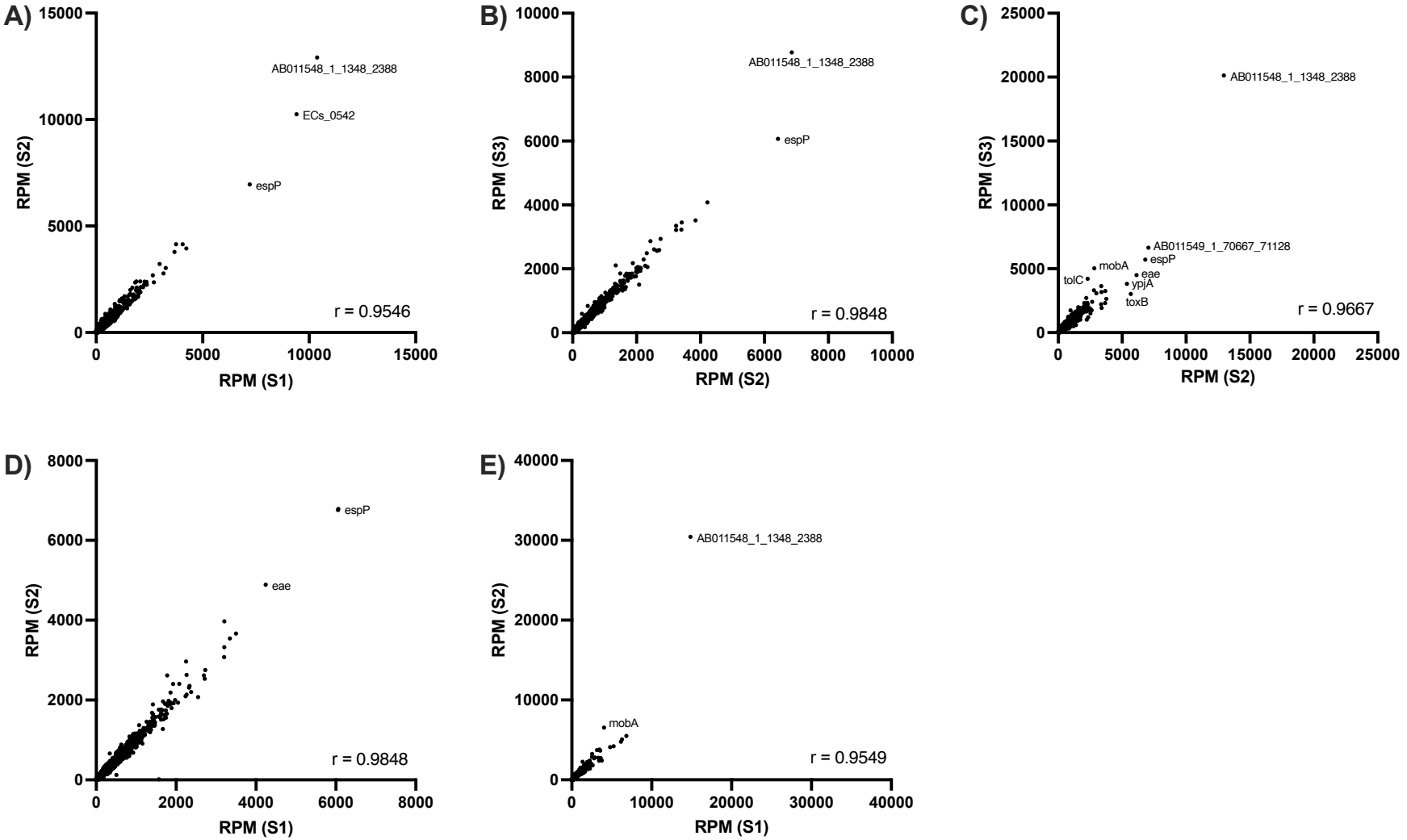


Figure 6.9: Comparison between RPM of two TraDIS libraries prepared from two separate biological replicates.
The conditions shown are as follows: A) ITL, B) 5.5 + I -> 2.5, C) 5.5 + I -> 7, D) 5.5 + E -> 2.5, E) 5.5 + E -> 7. The correlation was determined using Spearman's correlation coefficient shown at the bottom right corner.

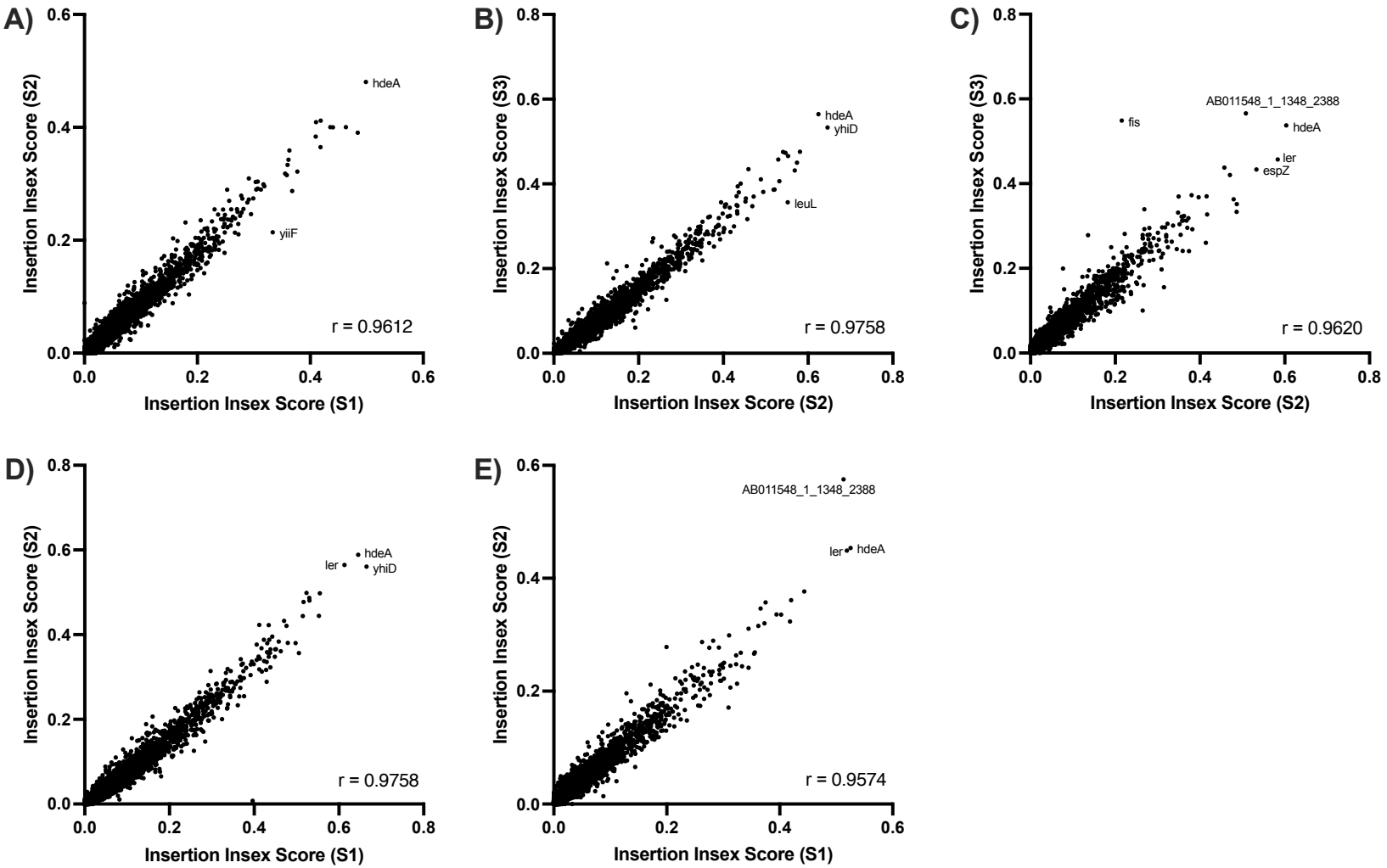


Figure 6.10: Comparison between insertion index scores of two TraDIS libraries prepared from two separate biological replicates.
The conditions shown are as follows: A) ITL, B) 5.5 + I -> 2.5, C) 5.5 + I -> 7, D) 5.5 + E -> 2.5, E) 5.5 + E -> 7. The correlation was determined using Spearman's correlation coefficient shown at the bottom right corner.

and it cannot be reduced by removing the outliers. Therefore, all three replicates, including all the genes present in each replicate, from all the conditions were included in the EdgeR comparisons.

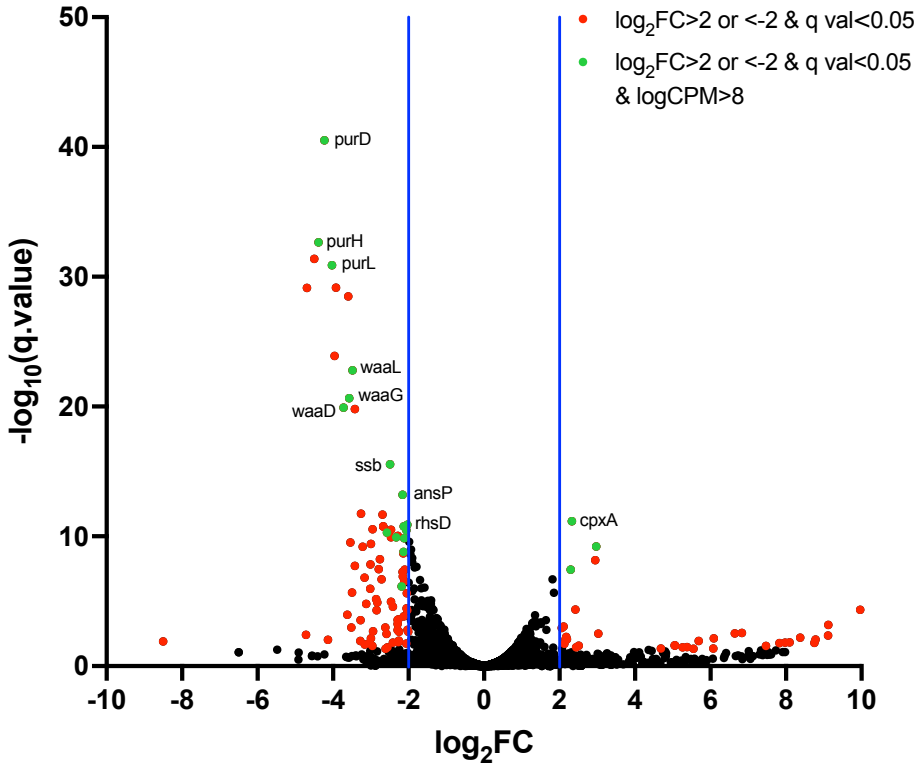
6.2.5.4 Visualisation of EdgeR comparisons between ITL and conditional TraDIS libraries using volcano plots

The outputs from EdgeR comparisons (log_2FC, q value and logCPM) between ITL and each conditional TraDIS library (5.5 + I -> 2.5, 5.5 + E -> 2.5, 5.5 + I -> I, 5.5 + E -> 7) were plotted on a separate volcano plot (Figure 6.11). An arbitrary cut-off of log_2FC > 2 or < -2, with q value < 0.05 was applied and these genes are shown in red. Another metric, logCPM (counts per million) which is used as a proxy for the expression level of genes in RNA-seq data, was used here as a proxy for the number of transposon mutants per gene. An arbitrary cut-off of logCPM > 8 was used and these genes are shown in green. Top 10 genes with log_2FC > 2 or < -2, q value < 0.05, and logCPM > 8 are labelled.

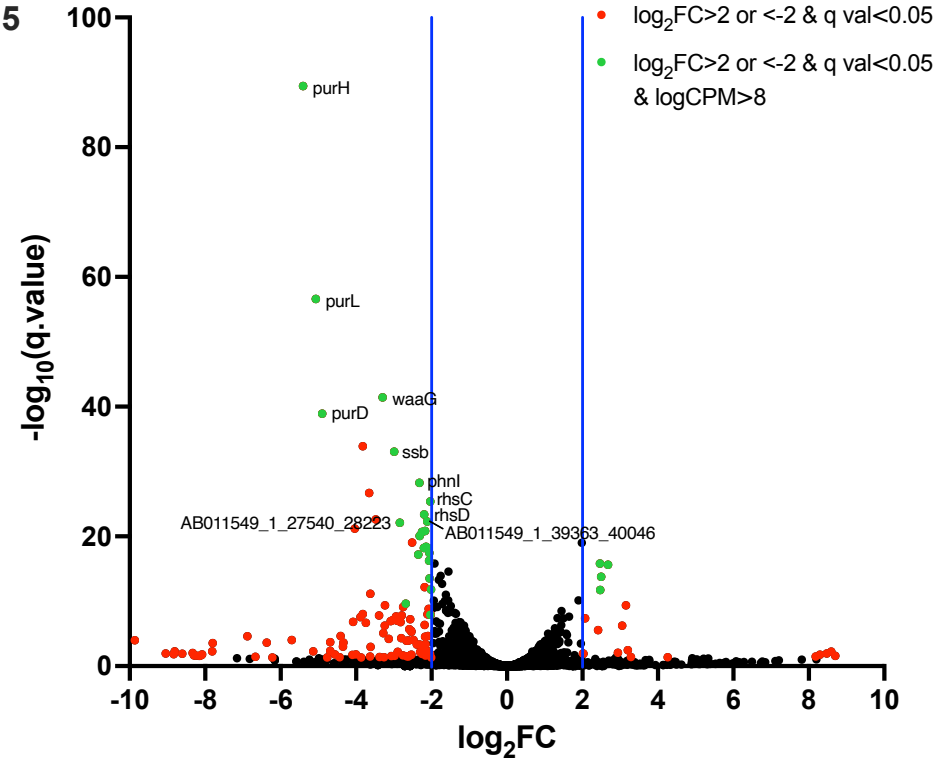
In all 4 volcano plots, the number of transposon mutants whose frequency decreased in the conditional TraDIS libraries relative to the ITL was higher than the mutants whose frequency increased in the conditional TraDIS libraries relative to the ITL. Additionally, a higher number of mutants with reduced frequency in the conditional TraDIS libraries relative to the ITL, had more significant q values than the mutants whose frequency increased in the conditional TraDIS libraries relative to the ITL.

The reason behind these two observations could be because a lot of genes containing transposon insertions are needed for full fitness of cells in conditions under which TraDIS experiments were performed. If majority of the genes, where the number of mutants containing insertions within them are decreasing in frequency, were required for the cells to

A) 5.5 + I -> 2.5



B) 5.5 + E -> 2.5



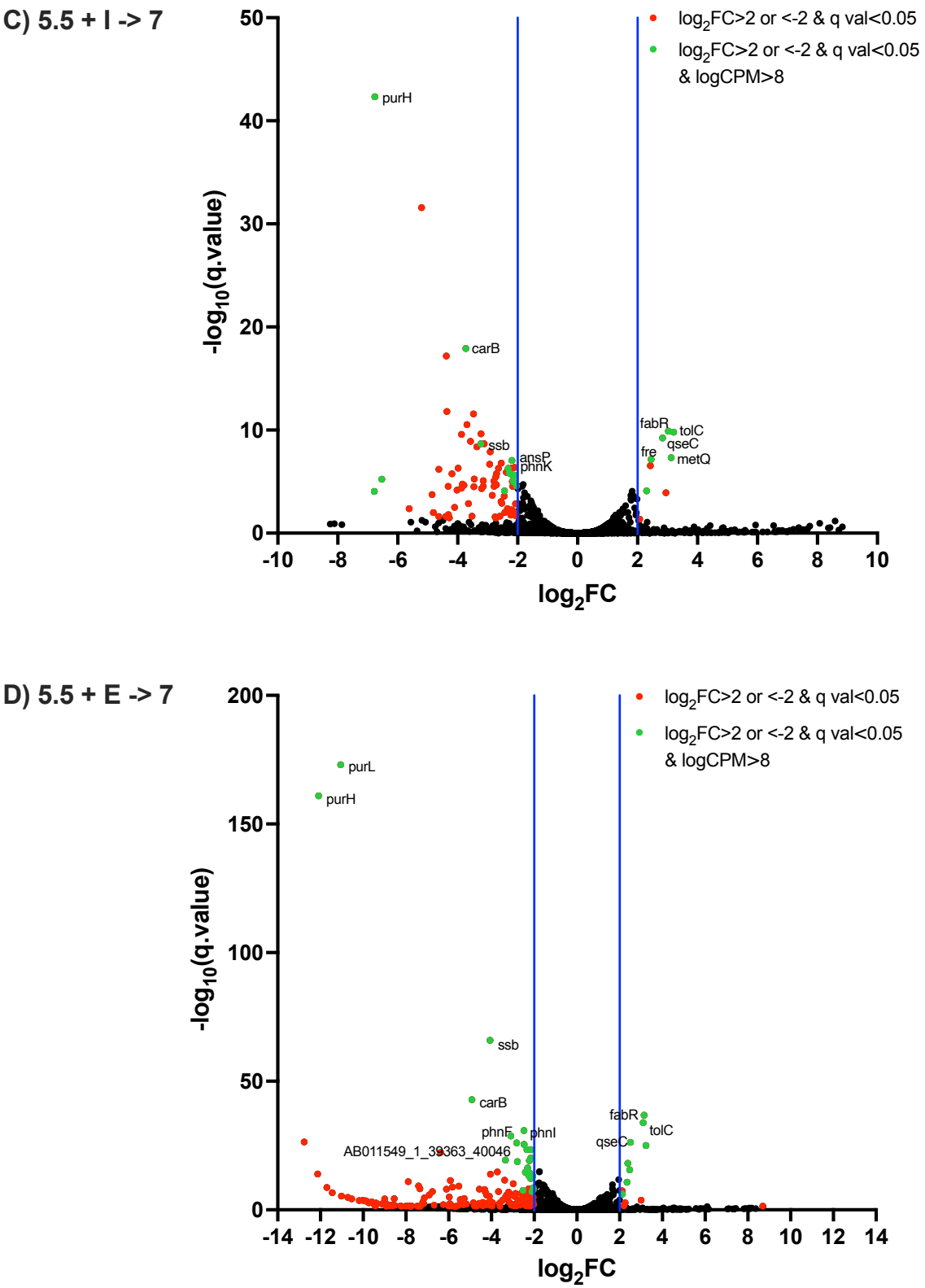


Figure 6.11: Volcano plots showing the log fold-change and the statistical significance corrected for multiple comparisons (q value) of the mutants containing transposon insertions within different genes.

The comparisons between ITL and conditional TraDIS libraries: A) 5.5 + I -> 2.5, B) 5.5 + E -> 2.5, C) 5.5 + I -> 7, D) 5.5 + E -> 7, have been shown. Top 10 genes with $log_2FC > 2$ or < -2, q value < 0.05, and logCPM > 8 in each plot are labelled.

survive at pH 2.5, then the frequency of such mutants would only decrease under 5.5 + I -> 2.5 and 5.5 + E -> 2.5 conditions, as opposed to all 4 conditions.

A large number of mutants with increased and decreased frequency in the conditional TraDIS libraries relative to the ITL, had very large log_2FC values. A large proportion of these mutants have q values which are below 0.05. A lot of these genes are likely to be essential genes which generally would not have any transposon insertions in them. However, if a small number of mutants containing insertions within these genes were present and the frequency of these mutants changed slightly, then these changes are presented as high log_2FC by EdgeR which is an artefact. An example where the majority of mutants with increased and decreased frequency relative to the ITL were found to contain insertions within essential genes is given in section 6.2.5.6.

A high number of genes with q value < 0.05 had logCPM < 8. LogCPM value of less than 8 for a gene means that the number of mutants containing insertions within this gene was lower than the number which is represented by the cut-off value of 8. If a cut-off of 8 for logCPM value was included, majority of the genes would be excluded from the analysis. Due to this reason, the cut-off for logCPM was not included in the data analysis in coming sections.

As can be seen in Figure 6.11, strains with insertions in some genes were reproducibly and significantly reduced under almost all conditions. This was true for *purH, purL* and *ssb*. As has been discussed previously in section 5.2.2.4 of Chapter 5, this result is as expected and acts as a control for the effectiveness of TraDIS in detecting fitness effects. There is a high possibility that the function of these genes is needed for the full fitness of *E. coli* during growth in M9Ksupp medium at pH 7. M9K minimal media does not contain purines (and pyrimidines) and *pur* genes are involved in the synthesis of purines (reviewed in Zhang, Morar and Ealick, 2008). So it is understandable that inserts in *purL* and *purH* were significantly reduced in all

conditions. Furthermore, the inserts in remaining *pur* genes including *purF, purM, purB, purC,* and *purD* were also highly reduced ($log_2FC < -2$) in all conditions (data in Appendix 8.3.1). Ssb is involved in DNA replication, repair and recombination. Upon checking the TraDIS results of ITL, *ssb* was found to be essential with no transposon insertions. So a high log_2FC produced by EdgeR for this gene under all conditions is an artefact.

Further analysis on genes with $log_2FC > 2$ or < -2, and q value < 0.05 in each library will be discussed later in more detail in section 6.2.5.6, where log_2FC and q values of genes from all four libraries will be compared.

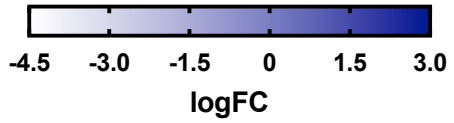
6.2.5.5 Comparison of fold change among mutants containing transposon within AR2 genes across four conditions

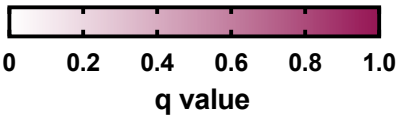
In Chapter 5, section 5.2.1, the success of the TraDIS experiments in MG1655 was checked by looking at the strains carrying insertions in *evgS* and *evgA.* It was predicted that among four TraDIS conditions, the mutants containing insertions in these two genes would show the largest decline in their frequencies, relative to the ITL, at pH 2.5 following pre-incubation at pH 5.5 in the presence of ethanol. This prediction was based on the fact that the strains lacking EvgS were less fit than the wild type strain in competition experiments performed under these conditions, which was expected from the known role of EvgS in activating the AR2 system (Chapter 5, section 5.2.1). Using the same principle, the success of TraDIS experiments in O157:H7 Sakai can be checked by looking at the *evgS* and *evgA* transposon mutants, which are expected to show the largest decline in their frequencies, relative to the ITL, at pH 2.5 in the presence of ethanol.

Table 6.3 shows that the strains carrying transposon insertions within *evgS* and *evgA* showed the largest and highly significant decline in fitness (the most negative log_2FC) at pH

Table 6.3: Comparison of log_2FC and q values of mutants containing transposon within the AR2 genes in four conditions under which TraDIS experiments were performed. The legend bars at the bottom show the colour corresponding to the log_2FC and q values. The data has been ranked from the smallest to the largest log_2FC in 5.5 + E -> 2.5 condition.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
rpoS	-2.79	-2.08	-1.39	-1.79	3.39×10^{-8}	1.50×10^{-9}	0.01	4.15×10^{-6}
gadE	-1.96	-1.46	-0.25	-2.98×10^{-3}	1.83×10^{-4}	0.01	0.94	1.00
lon	-0.44	-1.33	-2.72	-4.28	0.45	1.71×10^{-3}	3.34×10^{-6}	1.33×10^{-7}
ydeP	-1.54	-1.25	-0.72	-0.10	0.01	0.01	0.63	0.94
evgA	-0.96	-1.13	-0.25	-0.01	0.22	3.26×10^{-3}	0.96	1.00
evgS	-0.77	-1.13	-0.46	-0.08	0.30	7.75×10^{-4}	0.87	0.92
ydeO	-1.09	-0.90	-1.01	-1.02	0.18	0.16	0.52	0.33
rssB	-0.35	-0.79	-0.28	-2.17	0.79	0.57	0.97	0.31
gadW	-1.02	-0.56	-0.09	3.86×10^{-3}	0.03	0.19	0.99	1.00
dctR	-1.02	-0.44	-0.04	0.11	0.03	0.35	0.99	0.90
hdeD	-0.92	-0.42	-0.08	0.07	0.09	0.21	0.99	0.94
slp	-0.76	-0.37	-0.17	0.12	0.17	0.31	0.97	0.87
gadC	-0.27	-0.25	-0.06	-0.38	0.70	0.72	0.99	0.74
gadX	-0.07	-0.18	-0.09	0.18	0.92	0.70	0.99	0.79
hdeB	-0.41	-0.10	-0.13	0.01	0.54	0.89	0.98	1.00
hdeA	-0.09	0.02	0.15	0.34	0.91	0.98	0.98	0.42
gadB	0.16	0.08	0.52	1.13	0.85	0.95	0.85	0.05
mgtA	0.40	0.29	-0.01	0.24	0.43	0.46	1.00	0.71
yhiD	0.24	0.39	0.08	0.26	0.70	0.23	0.99	0.59
torR	1.11	0.43	0.85	0.62	0.18	0.80	0.83	0.86
gadA	0.67	0.51	0.46	0.46	0.11	0.15	0.80	0.28
torS	0.69	0.66	0.59	0.27	0.13	0.10	0.64	0.76
phoP	-2.13	0.68	1.32	1.39	1.46×10^{-4}	0.24	0.03	2.03×10^{-3}
phoQ	-1.60	0.74	0.95	1.15	0.01	0.20	0.19	0.01
rcsB	-1.08	2.78	-0.69	-0.47	0.32	0.30	0.82	0.82





2.5 in the presence of ethanol. Overall, this is consistent with expectations and suggests that the TraDIS experiments were successfully able to identify genes which are important under the conditions of the experiment. Furthermore, *evgS* and *evgA* transposon mutants only showed a small decline in fitness at pH 7 which suggests that the function of these genes does not confer a fitness advantage at pH 7. In contrast, during TraDIS experiments done with MG1655, *evgS* and *evgA* transposon mutants showed a large and significant decline in fitness at pH 7 which reflects the difference in the responses of two strains under our experimental conditions.

The frequency of mutants containing insertions in *gadE* as well as several other AR2 genes such as *ydeP, gadW, dctR, hdeD,* and *slp* showed a big decline in fitness at pH 2.5, but not at pH 7, consistent with the function of these genes needed for the optimal fitness of *E. coli* in extreme acid stress (Table 6.3). A rather unexpected result was that inserts in *rcsB* increased the fitness of mutants at pH 2.5 in the presence of ethanol. RcsB dimerises with GadE, and possibly with EvgA, to control the transcription of AR2 genes (Burton *et al.*, 2010; Castanié-Cornet *et al.*, 2010). It has also previously been shown that *rcsB* is crucial to increase the survival of exponentially-growing O157:H7 Sakai acid stressed at pH 2.4 following pre-incubation at pH 5.5 (Johnson *et al.*, 2011). Inserts in *rcsB*, therefore, were expected to reduce the fitness of mutants at pH 2.5 in the presence of ethanol, but the results of TraDIS experiments are completely opposite.

The decline in the fitness of *evgS* and *evgA* transposon mutants was bigger at pH 2.5 in the presence of ethanol than in the presence of indole. Furthermore, the decline in the fitness of *evgS* and *evgA* transposon mutants at pH 2.5 in the presence of indole was not significant. We have previously hypothesised that indole inhibits the AR2 response, possibly through interfering with the functioning of EvgS. A bigger decline in the fitness of *evgS* and

evgA transposon mutants in the presence of ethanol than indole suggests that this hypothesis could be true. As explained in section 5.2.2.5 of Chapter 5, this is because the impact on fitness of mutations in the AR2 genes goes away in the presence of indole, as all strains are equally affected, not just those with mutations in the AR2 genes.

The decline in the fitness of mutants containing inserts in several other AR2 genes such as *gadE, ydeP, gadW, dctR, hdeD, slp* and *hdeB,* was bigger at pH 2.5 in the presence of indole than ethanol. Furthermore, the $\log_2$FC of a number of these genes in the presence of indole was also significant. A bigger decline in the fitness of these mutants in the presence of indole than ethanol suggests that these genes could be getting activated by a transcription factor other than EvgA, since the function of EvgS and EvgA is likely to be inhibited by indole. GadE is the central regulator of AR2 response and controls the transcription of majority of the AR2 genes. It gets activated by EvgA, and the components present downstream of EvgA (Ma, Masuda and Foster, 2004; Sayed, Odom and Foster, 2007). It is possible that in the presence of indole, RpoS activates *gadE* expression and GadE would then activate the expression of other AR2 genes as usual. Mutants carrying inserts within *rpoS* showed a decline in fitness under all conditions (Table 6.3). But the decline in fitness was larger at pH 2.5 than pH 7, and largest in the presence of indole. These results suggest that RpoS is important for the ability of cells to survive exposure to pH 2.5. In MG1655, RpoS has been shown to activate the expression of *gad* genes and periplasmic chaperones *hdeA* and *hdeB,* through regulating GadE levels, in exponential phase (Burton *et al.,* 2010). This could be the reason for a bigger decline in the fitness of *rpoS* mutants at pH 2.5 than pH 7. We speculate that a bigger decline in the fitness of *rpoS* mutants in the presence of indole than ethanol could be because RpoS activates *gadE* in the presence of indole, to compensate for the reduced transcription of *gadE* by EvgA.

Mutants containing inserts in *lon* showed a smaller decline in fitness at pH 2.5 than pH 7, and the smallest decline in fitness in the presence of indole (Table 6.3). Lon protease is known to degrade GadE (Heuveling, Possling and Hengge, 2008). Mutants containing inserts in *lon* are likely to have higher GadE levels than wild-type cells which will increase their fitness at pH 2.5. The smallest decline in the fitness of *lon* mutants at pH 2.5 in the presence of indole, could be a result of the compensation for the inhibitory effect of indole on EvgS.

Mutants containing inserts in *phoP* and *phoQ* showed an increase in fitness under all conditions except at pH 2.5 in the presence of indole, these mutants showed a decline in fitness (Table 6.3). PhoPQ is a TCS where PhoQ is the sensor kinase and PhoP is the response regulator. It has both activating and repressing effects on AR2 response. PhoP directly represses *ydeO-safA* and *ydeP* promoters (Burton *et al.*, 2010). The decline in the fitness of mutants containing insertions in *phoP* and *phoQ* at pH 2.5 in the presence of indole is not likely due to its repressing effect on AR2 genes. This is because YdeO activates *gadE* which confers fitness advantage to cells at pH 2.5. PhoPQ also activates *iraM* (Yoko *et al.*, 2011). IraM associates with RssB, which prevents RssB from binding RpoS and directing it to ClpXP protease for its degradation. Therefore, mutants containing PhoP and PhoQ are likely to have higher levels of RpoS than mutants containing inserts in *phoP* and *phoQ*, which in turn can increase GadE levels if our hypothesis that RpoS activates *gadE* is true. EdgeR results for O157:H7 Sakai did not provide any output for *iraM*. When the nucleotide sequence of *iraM,* obtained from MG1655 genome, was searched in O157:H7 Sakai genome, it could not be found. The likelihood is that *iraM* is not present in O157:H7 Sakai genome. However, some other mechanism under PhoP and PhoQ regulation, which increases RpoS levels, could be present in O157:H7 Sakai.

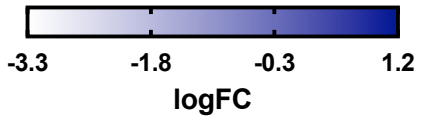
To check whether mutations in the genes whose role in acid stress has previously been studied caused any fitness advantage or disadvantage under the conditions used in our TraDIS experiments, the log_2FC of strains containing inserts within 106 different genes across four TraDIS experimental conditions were compared. This list was prepared from genes mentioned in the following three reviews on acid stress in *E. coli*: Kanjee and Houry (2013); Li *et al.* (2020); Schwarz *et al.* (2022). The list of these genes is provided in Appendix 8.3.5. The strains containing inserts within genes involved in AR3 (*adi* genes), AR4 (*cad* genes) and AR5 (*speF* and *potE*) systems did not show a significant decline in fitness in any of the four TraDIS conditions, suggesting that none of these systems were involved in the response being studied here (Table 6.4). CRP and RpoS are the main regulators of AR1 system (Lin *et al.*, 1995; Marie-Pierre *et al.*, 1999). Strains containing inserts in *crp* did not show a significant decline in fitness under any of the four conditions. Strains carrying inserts in RpoS, as previously discussed, showed a big decline in fitness under all four conditions (Table 6.4). However, the fact that AR1 is repressed by glucose, the main carbon source used in our medium, makes the likelihood of AR1 being active quite small.

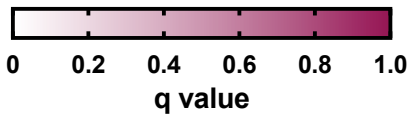
Strains containing inserts within *degP* showed a bigger decline in fitness at pH 2.5 than pH 7 (Table 6.4). DegP is a periplasmic serine endoprotease which degrades abnormal proteins in periplasm. It has previously been shown to increase the viability of *E. coli* at pH 2.3 (Fu *et al.*, 2018). So the results of our TraDIS experiments agree with the previous findings.

SurA is a periplasmic chaperone which maintains protein quality in the periplasm. SurA is protected by HdeA during extreme acid stress. Upon neutralisation, SurA assists HdeA in recovering other client proteins from damage endured during acid stress (Zhang *et al.*, 2011). In our TraDIS results, strains containing inserts within *surA* showed a decline in fitness under all four conditions (Table 6.4). This suggests that function of SurA improves the fitness of cells

Table 6.4: Comparison of log_2FC and q values of mutants containing transposon within the genes which have been tested for their role in acid stress, in four conditions under which TraDIS experiments were performed. The legend bars at the bottom show the colour corresponding to the log_2FC and q values.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
adiA	0.52	0.36	0.34	0.67	0.27	0.31	0.89	0.08
adiC	0.56	0.60	0.63	0.38	0.34	0.22	0.67	0.64
adiY	-0.19	-0.21	0.08	0.16	0.74	0.61	0.99	0.83
cadA	-0.01	0.09	-0.01	0.19	0.99	0.87	1.00	0.76
cadB	-0.16	-0.04	0.03	0.27	0.80	0.94	1.00	0.55
cadC	-0.37	-0.32	-0.32	-0.07	0.54	0.45	0.91	0.93
clcA	1.08	0.90	-0.29	-0.21	3.32×10^{-3}	4.16×10^{-3}	0.91	0.80
crp	0.37	-0.03	0.94	-0.56	0.92	0.98	0.95	0.94
degP	-0.79	-0.64	-0.15	-0.11	0.05	0.03	0.97	0.92
ompA	-1.94	-1.22	-0.65	-1.20	1.75×10^{-4}	0.05	0.75	0.26
ompC	-1.44	-1.77	-0.44	-0.28	0.01	6.51×10^{-4}	0.85	0.84
potE	0.24	0.29	0.31	0.13	0.71	0.62	0.92	0.92
speF	0.31	0.40	0.04	0.20	0.57	0.25	0.99	0.80
surA	-2.47	-3.23	-2.79	-2.92	1.20×10^{-10}	4.19×10^{-10}	2.89×10^{-5}	6.71×10^{-6}





at pH 7, or possibly at pH 5.5. The fitness of strains containing inserts in *surA*, showed a bigger decline in fitness at pH 2.5 in the presence of ethanol than in the presence of indole, and at pH 7. Although SurA does not play an active role itself at low pH, it can help HdeA later, once the stress is removed, in recovery of damaged proteins. So increased levels of HdeA in cells, resulting from the activation of AR2 response, may protect SurA from acid stress. This could be the reason why strains containing inserts in *surA* show a greater decline in fitness at pH 2.5 in the presence of ethanol than other three conditions. In the presence of indole at pH 2.5, however, due to inhibition of AR2 response by indole, cells must be producing lower levels of HdeA than in the presence of ethanol. Due to this, the cells would not be able to protect SurA from acid stress. Since in the presence of indole, the strains containing inserts in *surA* are not at a bigger fitness disadvantage than the strains not containing inserts in *surA, surA* mutants do not decline in frequency to the same extent as the *surA* mutants in the presence of ethanol. It is unlikely that SurA plays an active role in the recovery of damaged proteins at pH 2.5 because otherwise the decline in the fitness of *surA* mutants at pH 2.5 in the presence of ethanol and indole would have been similar.

Strains containing inserts in outer membrane porins (OMPs) *ompA* and *ompC* showed a bigger decline in fitness at pH 2.5 than pH 7 (Table 6.4). There are contradictory results regarding the role of OMPs in acid stress. *ompC* mutant has been shown to have a reduced survival at pH 2.5 with and without induction of AR2 response (Bekhit *et al.*, 2011). *ompA* mutant has also been shown to have a reduced survival at pH 3.6 (Wang, 2002). In contrast, He *et al.* (2019) have shown that periplasmic serine endoprotease DegP degrades OmpA and OmpC at pH 2, possibly to reduce proton influx into the cells. Our results however suggest that the function of both these OMPs is needed for optimal fitness of cells at pH 2.5, as shown by Bekhit *et al.* (2011) and Wang (2002). A possible explanation for this is that OMPs increase

the acid resistance of cells by transporting glutamate into the cells and by removing GABA out of the cells.

Strains containing inserts in H^+/Cl^- transporter *clcA* showed an increase in fitness at pH 2.5, but a small decline in fitness pH 7 (Table 6.4). It has been shown that during acid stress, *E. coli* reverse their transmembrane potential by making inside more positive which helps to prevent the influx of protons into the cells through charge repulsion (Iyer *et al.*, 2002; Kanjee and Houry, 2013). Once the stress is over, H^+/Cl^- transporters help to correct the transmembrane potential (inside negative) through efflux of protons in exchange for Cl^- ions from the media. The strains containing inserts in *clcA* have an increased fitness at pH 2.5 because they might be able to maintain a more positive transmembrane potential than the wild-type cells, reducing the flow of protons into the cells.

6.2.5.6 Comparison of fold change in genes with depleted and enriched transposon insertions across four conditions

In the previous section, the analysis of TraDIS data focussed on genes which are known to be required for optimum survival of *E. coli* at low pH. Many of these genes showed a high and significant log_2FC in conditions where their function, from prior knowledge, is known to contribute to optimum survival at low pH of *E. coli*. However, as discussed in Chapter 5, the main advantage of TraDIS data is that it allows the identification of potential gene function without any prior knowledge. In this section, therefore, the TraDIS data is analysed depending on the magnitude of the effect (log_2FC, whether positive or negative) that insertions within different genes can have on the fitness of cells, as well as the significance level of the log_2FC (q value). To identify potential genes of interest, each condition was compared to the ITL using EdgeR, and genes with $log_2FC > 2$ or < -2, and q value < 0.05 were identified. Venn diagrams

were then generated to identify genes which belong to only one or more than one condition. Separate Venn diagrams were generated for genes with $log_2FC < -2$, and those with $log_2FC > 2$ (Figures 6.12 A and B respectively).

As explained previously in Chapter 5, this is a rather crude analysis which involves using arbitrary cut-offs. Furthermore, a Venn diagram may suggest that a gene which is above the cut-off value in one condition is significant only in that category. In reality the log_2FC value for that gene could be similar in other categories, just below the cut-off value. Therefore, it would be unreasonable to suggest that the gene is only significant in one condition and not the others. But since TraDIS datasets are huge, Venn diagrams are appropriate for initial analysis to look at some "low hanging fruit". More analytical and statistically powerful analysis (Gene set enrichment analysis) will be discussed later in section 6.2.5.6. Some general points about Venn diagrams can be made before looking at individual genes or group of genes. Over 13% of the genes in Figure 6.12 A and over 6% of the genes in Figure 6.12 B are common in all four conditions. These are likely to be the genes whose function either provides a fitness advantage or fitness disadvantage under normal growth conditions and not essentially under acid stress.

As discussed in section 6.2.5.1, the aim of the TraDIS experiments was to identify genes that are important for survival of *E. coli* at pH 2.5, and genes which are involved in the inhibitory effect of indole on AR2 induction. It was also stated in the same section that in order to identify genes which either confer a fitness advantage or disadvantage at pH 2.5, genes in which insertions are depleted or enriched in either or both of the 5.5 + I -> 2.5 and 5.5 + E -> 2.5 conditions need to be looked at. To identify genes involved in the inhibitory effect of indole, genes in which insertions are enriched in either or both of the 5.5 + I -> 2.5 and 5.5 + I -> 7 conditions need to be looked at. So within the Venn diagrams, genes falling in

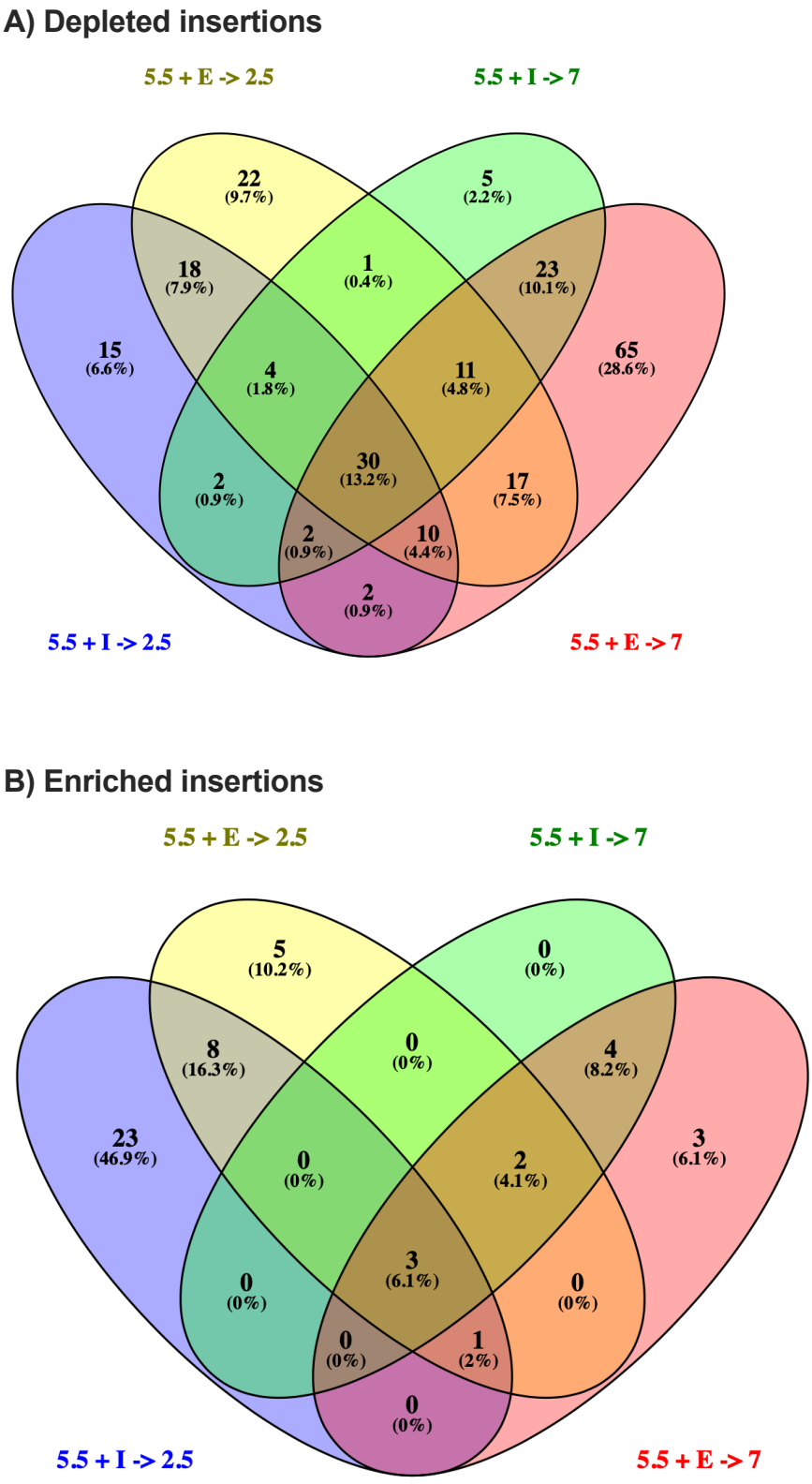


Figure 6.12: Venn diagrams showing genes where the number of insertions had either decreased (depleted insertions) (A), or increased (enriched insertions) (B), in the conditional TraDIS libraries as compared to the ITL.

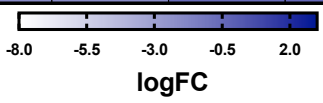
these categories were identified and their log_2FC and q values were compared across four conditions. These genes are given in Tables 6.5 (depleted insertions) and 6.6 (enriched insertions). A full list of genes in each category is given in Appendices 8.3.3 and 8.3.4.

In Table 6.5, 2 AR2 genes (*PhoP, rpoS*), 8 essential genes (*ECs_1837, facG_1, lolA, pdxH, ribE, rplV, rpmA, rpsB*) and 6 genes which were unclear in terms of whether they are essential or non-essential (*aceE, lepB, nuoA, rpsF, rpsR, yeeT*) were removed from the comparison. In Table 6.6, 19 essential genes, 1 unclear gene (*narY*), and 1 gene which was below 200 bp in length (*tisB*- 90 bp) are not shown. Since the comparison of log_2FC and q values in AR2 genes has already been discussed in section 6.2.5.5, these genes were excluded from this section. The essential genes were excluded because they tolerate a very small number of transposon insertions as they are absolutely necessary for the growth of cells in any condition. Therefore the change in the frequency of the mutants containing insertions in these genes cannot be looked at. The genes that fall into "unclear" category could either be essential or non-essential. Due to the uncertainty associated with their essentiality, these genes were also excluded from the comparison. Furthermore, it is also not possible to determine the conditional essentiality of short genes due to presence of no or very small number of insertions.

To know the biological processes, molecular functions, and the cellular components that the proteins encoded by the genes given in Tables 6.5 and 6.6 are involved in, the genes were analysed using the String database (https://string-db.org, version 11.5). 10 genes which are involved in the biosynthesis of extracellular polysaccharides were found. 7 of these genes (*arnB, waaO, waaL, waaR, waaD, wzy, wcaG*) are involved in LPS biosynthesis, and 3 genes (*gmd, wcaE, cpsB*) are involved in colanic acid biosynthesis. Insertions in all these genes have a negative effect on fitness at pH 2.5 (Table 6.5). Furthermore, the q values for most of these

Table 6.5: Comparison of log_2FC and q values of transposon mutants whose frequency reduced (depleted insertions) under at least one TraDIS experimental condition compared to the ITL. The legend bars at the bottom show the colour corresponding to the log_2FC and q values.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
AB011549_1_40162_40716	-1.79	-2.07	-1.58	-1.84	2.21×10^{-8}	5.64×10^{-17}	1.72×10^{-3}	8.10×10^{-11}
AB011549_1_43157_43546	-2.27	-2.55	-2.48	-1.50	2.50×10^{-4}	0.02	0.37	0.24
arnB	-3.26	-2.52	-0.22	-0.13	1.79×10^{-12}	8.77×10^{-20}	0.97	0.92
bamB	-2.69	-2.57	-0.78	-0.62	2.15×10^{-12}	5.80×10^{-8}	0.38	0.35
cpsB	-2.67	-1.95	0.31	0.65	1.76×10^{-11}	2.94×10^{-4}	0.94	0.27
ECs_0292	-2.01	-1.63	-1.02	-0.70	2.23×10^{-3}	0.02	0.60	0.61
ECs_1187	-3.00	-2.53	-2.29	-1.27	0.01	0.06	0.28	0.57
ECs_1985	-1.97	-3.07	-3.13	-3.00	0.23	0.03	0.65	0.39
ECs_2180	-2.06	-1.66	-1.28	-1.71	0.02	0.03	0.32	0.12
ECs_2840	-2.83	-2.60	0.45	0.40	1.24×10^{-5}	1.13×10^{-4}	0.93	0.73
ECs_2845	-3.28	-1.90	-0.01	-0.14	0.01	0.22	1.00	0.97
ECs_2958	-4.14	-2.06	-5.07	-2.01	0.01	0.44	0.08	0.72
espR2	-2.05	-1.75	-1.10	-1.16	3.61×10^{-5}	6.53×10^{-4}	0.39	0.10
gmd	-2.18	-2.03	0.20	0.46	7.28×10^{-7}	1.17×10^{-8}	0.98	0.37
gne	-2.67	-3.03	-0.84	-0.37	1.76×10^{-11}	8.61×10^{-8}	0.33	0.62
hldD	-2.86	-3.28	-3.18	-2.69	6.82×10^{-6}	8.36×10^{-6}	0.15	0.09
hscA	-1.65	-2.39	-1.34	-1.16	0.05	4.69×10^{-4}	0.28	0.45
hscB	-2.25	-2.76	-1.76	-2.12	2.21×10^{-3}	0.02	0.17	0.40
ihfB	-1.65	-3.26	-4.49	-4.72	0.33	0.04	0.06	0.13
mtn	-1.90	-3.63	-1.99	-4.57	0.28	0.03	0.77	0.15
neuD	-2.76	-2.89	-0.16	0.11	5.75×10^{-9}	2.80×10^{-8}	0.98	0.97
nhaA	-1.34	-2.01	-1.47	-2.02	0.07	0.05	0.33	0.34
nhoA	-1.66	-3.87	-1.71	-1.88	0.27	0.05	0.84	0.57
pal	-2.26	-2.04	-1.77	-2.68	0.01	0.01	0.13	0.06
pdxJ	-1.65	-4.00	-3.97	-3.79	0.08	0.02	0.14	0.10
priB	-2.43	-7.80	-4.75	-0.57	0.20	2.92×10^{-4}	0.24	0.92
proQ	-3.16	-2.29	-2.11	-3.33	0.02	0.13	0.48	0.23
purA	-2.27	-2.08	-2.65	-4.04	0.01	0.17	0.45	0.20
qseC	-2.47	-0.49	2.83	2.50	3.07×10^{-11}	0.25	5.90×10^{-10}	6.10×10^{-27}
rpmE	-2.61	-2.26	-2.73	-2.34	0.05	0.13	0.43	0.28
rseA	-2.41	-1.87	-1.43	-1.98	2.62×10^{-5}	0.04	0.17	0.03
waaD	-3.72	-2.68	0.27	0.42	1.19×10^{-20}	2.29×10^{-10}	0.96	0.35
waaL	-3.48	-2.05	0.47	0.54	1.60×10^{-23}	3.10×10^{-14}	0.87	0.20
waaO	-4.50	-3.82	-0.80	-0.82	4.28×10^{-32}	1.22×10^{-34}	0.65	0.11
waaR	-4.69	-3.65	-0.14	-0.20	7.17×10^{-30}	2.04×10^{-27}	0.98	0.87
wcaE	-3.52	-2.66	0.73	1.04	1.08×10^{-3}	0.03	0.86	0.33
wcaG	-1.42	-2.17	-1.06	-1.10	0.02	3.07×10^{-5}	0.54	0.20
wzy	-3.01	-1.08	2.03	2.13	1.10×10^{-6}	0.06	0.06	6.96×10^{-7}
yeeZ	-2.05	-2.02	-0.65	-0.57	2.48×10^{-6}	1.12×10^{-8}	0.54	0.38



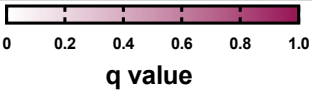
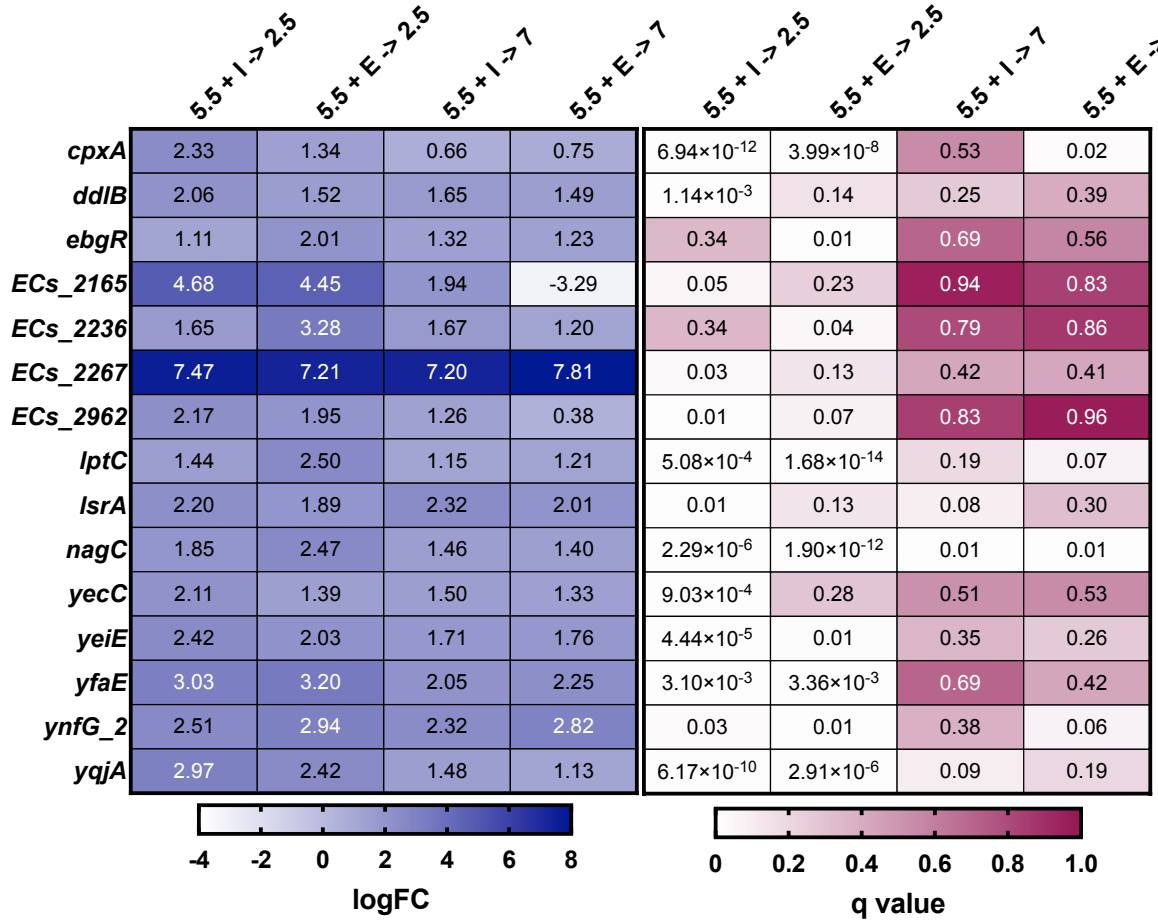


Table 6.6: Comparison of log_2FC and q values of transposon mutants whose frequency increased (enriched insertions) under at least one TraDIS experimental condition compared to the ITL. The legend bars at the bottom show the colour corresponding to the log_2FC and q values.



genes are highly significant. Insertions in these genes at pH 7, however, generally have a positive effect on fitness. If the effect on fitness at pH 7 is negative, then it is much smaller than the effect at pH 2.5. Insertions in another gene, *gne,* lead to a much bigger and more significant decline in fitness at pH 2.5 than pH 7 (Table 6.5). Gne is a UDP-galactose 4-epimerase which produces UDP-α-D-galactose, one of the components of colanic acid. These results, together, suggest that the function of extracellular polysaccharides is important for survival of cells at pH 2.5. It is possible that the production of increased amount of LPS and colanic acid might reduce the influx of protons into the cells at low pH.

Inserts in both ECs_2840 and ECs_2845 led to a decline in fitness at pH 2.5 but not pH 7, suggesting that the function of these proteins improves the fitness of cells at pH 2.5. ECs_2840 is a protein of unknown function. ECs_2845 is a glycosyl transferase but the specific process or the reaction in which it is involved is not known. Since the usual function of glycosyl transferases is to catalyse the transfer of sugar moieties to carbohydrates during oligosaccharide biosynthesis, it is possible that it is involved in LPS or colanic acid biosynthesis. Furthermore, both ECs_2840 and ECs_2845 are part of the same transcriptional unit as *cpsB, wcaG, gmd, arnB, wcaE,* and *wzy*. This raises a strong possibility that ECs_2840 and ECs_2845 may be involved in the synthesis of extracellular polysaccharide because the inserts in both these genes lead to a fitness disadvantage at pH 2.5, like the other genes described above.

Inserts in *qseC* led to an increase in fitness at pH 7 but a decline in fitness at pH 2.5 (Table 6.5). Furthermore, the decline in fitness at pH 2.5 was greater in the presence of indole than ethanol. QseC is a part of QseBC TCS, where QseC is the sensor kinase and QseB is the response regulator. QseB has been shown to regulate the expression of EptA, which adds phosphoethanolamine to lipid A core oligosaccharide of LPS (Hurst *et al.*, 2023). This LPS modification masks the negative charges on LPS reducing the influx of protons through charge

repulsion. Although inserts in *qseB* did not show a strong signal in our TraDIS data, inserts in *eptA* led to a decline in fitness at pH 2.5 but not pH 7 (data shown in Appendix 8.3.1). Furthermore, the decline in fitness of *eptA* mutants was greater in the presence of indole (log_2FC = -1.6) than ethanol (log_2FC = -1.3). There is a possibility that QseC regulates *eptA* expression through a different response regulator, and the inserts in *qseC* showed a decline in fitness because it provides acid resistance at pH 2.5 through LPS modification causing reduction in proton influx.

Inserts in *cpxA* led to an increase in fitness under all conditions (Table 6.6) However, the increase in fitness was greater and much more significant at pH 2.5 than at pH 7. CpxAR is a TCS which responds to envelope stress, whereby CpxA is the sensor kinase and CpxR is the response regulator. The activation of CpxAR TCS reduces the abundance of several proteins including those involved in AR2 response such as GadA, GadC, HdeA, HdeB and HdeD (Surmann *et al.*, 2016). Furthermore, the activation of this system leads to a complete loss of viability of cells at pH 2, which is recovered in the *cpxAR* mutant. Our TraDIS data, therefore, is consistent with the previous findings that the function of CpxA reduces the fitness of cells at pH 2.5.

6.2.5.7 Comparison of metabolic and regulatory pathways important in four conditions

In the previous section, the individual genes and group of genes whose function is required for full fitness of cells under different conditions were looked at. The cut-offs used in that section were arbitrary which can increase the risks of false-positives and false-negatives. In this section, a more powerful analytical method which is Gene set enrichment analysis (GSEA) was used to analyse the TraDIS dataset. GSEA was used to determine the KEGG metabolic and transcriptional regulatory pathways (Tables 6.7 and 6.8 respectively), in

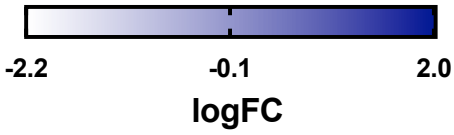
Table 6.7: KEGG metabolic pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions relative to the ITL. The legend bars at the bottom show the colour corresponding to the normalised enrichment score (NES) and q value.

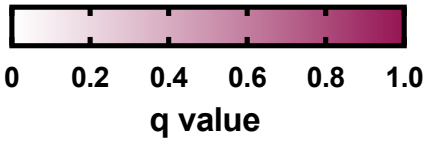
	5.5+I->2.5	5.5+E->2.5	5.5+I->7	5.5+E->7	5.5+I->2.5	5.5+E->2.5	5.5+I->7	5.5+E->7
Eco00020 (Citrate/TCA cycle)	-0.89	-1.03	-1.52	-1.58	0.97	0.85	0.16	0.04
Eco00230 (Purine metabolism)	-1.57	-1.87	-1.95	-1.65	0.20	0.01	6.83×10^{-4}	0.01
Eco00240 (Pyrimidine metabolism)	-1.24	-1.70	-1.97	-1.77	0.72	0.03	5.50×10^{-4}	9.29×10^{-4}
Eco00250 (Alanine, aspartate and glutamate metabolism)	-1.74	-1.68	-1.99	-1.64	0.05	0.04	1.10×10^{-3}	0.02
Eco00260 (Glycine, serine and threonine metabolism)	-0.92	-1.30	-1.70	-1.33	0.98	0.49	0.03	0.15
Eco00440 (Phosphonate and phosphinate metabolism)	-2.01	-1.86	-1.91	-1.59	3.13×10^{-3}	0.01	2.75×10^{-3}	0.04
Eco00670 (One carbon pool by folate)	-1.53	-1.80	-1.87	-1.61	0.24	0.01	2.48×10^{-3}	0.03
Eco00750 (Vitamin B6 metabolism)	-1.78	-1.83	-1.80	-1.66	0.05	0.01	0.01	0.01
Eco01240 (Biosynthesis of cofactors)	-1.20	-2.07	-1.90	-1.82	0.64	0.00	2.76×10^{-3}	0.00
Eco03010 (Ribosome)	1.95	-1.72	1.43	-1.72	0.01	0.03	0.28	3.84×10^{-3}
Eco00061 (Fatty acid biosynthesis)	1.81	1.34	1.61	-1.12	0.03	0.48	0.15	0.54
Eco00550 (Peptidoglycan biosynthesis)	1.73	1.66	1.74	1.26	0.07	0.31	0.05	1.00
Eco00900 (Terpenoid backbone biosynthesis)	1.73	1.24	1.88	1.12	0.08	0.56	0.02	0.93
Eco01212 (Fatty acid metabolism)	1.85	1.38	1.62	-1.09	0.03	0.49	0.17	0.56
Eco03060 (Protein export)	1.64	1.31	1.78	1.08	0.13	0.49	0.04	0.97
Eco03070 (Bacterial secretion system)	1.63	1.17	1.79	1.24	0.12	0.58	0.06	1.00



Table 6.8: Transcriptional regulatory pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions relative to the ITL. The legend bars at the bottom show the colour corresponding to the normalised enrichment score (NES) and q value.

	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7	5.5 + I -> 2.5	5.5 + E -> 2.5	5.5 + I -> 7	5.5 + E -> 7
BirA	-1.57	-1.59	-1.74	-1.55	0.26	0.44	0.03	0.16
PhoB	-1.84	-1.54	-1.71	-1.40	0.02	0.32	0.03	0.54
PurR	-2.18	-2.20	-2.16	-1.86	0.00	0.00	0.00	0.00
RbsR	-1.59	-1.49	-1.81	-1.57	0.26	0.44	0.01	0.14
FadR	1.87	1.35	1.74	-1.15	0.01	1.00	0.19	1.00
MraZ	1.99	2.01	1.72	1.30	9.47×10^{-4}	1.38×10^{-3}	0.14	1.00
PdhR	1.83	1.77	1.68	-0.87	0.02	0.12	0.13	1.00





which the inserts in different genes had either increased or decreased in frequency in a conditional TraDIS library relative to the ITL. The pathways where the inserts in genes increased in frequency in a conditional TraDIS library relative to the ITL were referred to as enriched pathways (positive NES score). The function of genes in these pathways confer a fitness disadvantage to cells under a given condition. The pathways where the inserts in genes decreased in frequency in a conditional TraDIS library relative to the ITL were referred to as depleted pathways (negative NES score). The function of genes in these pathways confer a fitness advantage to cells under a given condition.

The KEGG pathways "purine metabolism" and "pyrimidine metabolism", which include genes involved in purine and pyrimidine biosynthesis respectively, were both depleted in all four conditions (Table 6.7). Furthermore, the transcriptional pathway regulated by PurR, was also significantly depleted in all four conditions (Table 6.8). As discussed before in section 6.2.5.4, M9 minimal media lacks purines and pyrimidines. Therefore, mutants containing inserts in genes involved in purine and pyrimidine metabolism will not be able to grow under our experimental conditions. So this result is consistent with expectations, agrees with similar findings in MG1655 (see section 5.2.2.7), and acts as a positive control for these TraDIS experiments.

The KEGG pathway "Citrate/TCA cycle" was depleted under all four conditions (Table 6.7). Furthermore, it was depleted less at pH 2.5 than pH 7. TCA cycle is the main pathway that provides energy to cells, helping them to grow. Therefore, mutants containing inserts in genes involved in this pathway are expected to be less fit. These mutants, however, might be at a smaller disadvantage at pH 2.5 than pH 7, because the priority of cells at pH 2.5 is to survive the stress rather than to grow and divide.

6.2.6 Comparison between MG1655 and O157:H7 Sakai TraDIS results using GSEA

One of the aims of this study was to compare how MG1655 and O157:H7 Sakai respond to acid stress and how indole inhibits the induction of AR2 response in both the strains. TraDIS experiments were performed on both the strains and analysis done on AR2 genes described in previous (section 5.2.2.5) and this chapter (section 6.2.5.5) has shown that inserts in several AR2 genes including *evgS* reduce the fitness of both strains at pH 2.5 following pre-induction at pH 5.5. These results are consistent with the known function of EvgS which is to activate the AR2 response at pH 5.5. Furthermore, in both the strains, inserts in *evgS* led to a smaller reduction in fitness at pH 2.5 when the pre-incubation at pH 5.5 was done in the presence of indole. This observation is consistent with our hypothesis that indole might inhibit the function of EvgS to block the induction of AR2 response. Additionally, the promoter-probe assays done with MG1655 in Chapter 3 (sections 3.2.1 and 3.2.3) and the survival assays done with MG1655, BW25113 and O157:H7 Sakai in Chapter 4 (sections 4.2.3, 4.2.6, 4.2.7) also suggest that indole inhibits the induction of AR2 response in exponentially growing *E. coli*.

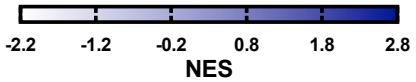
Several individual genes, and group of genes encoding different cell components, whose function might increase or decrease the fitness of cells at pH 2.5 in either MG1655 or O157:H7 Sakai have been discussed in the previous chapter (section 5.2.2.6) and this chapter (section 6.2.5.6). Furthermore, a more powerful statistical method (GSEA) than using arbitrary cut-offs, was also used to identify the KEGG metabolic and transcriptional regulatory pathways which get enriched or depleted in different TraDIS experimental conditions, relative to the ITL, in each strain (sections 5.2.2.7 and 6.2.5.7). The inserts in genes that are part of an enriched or a depleted pathway, either reduce or increase the fitness of a strain in a given condition.

In this section, GSEA will be used to identify the differences and similarities in enriched and depleted pathways in MG1655 and O157:H7 Sakai under the four TraDIS experimental conditions (Tables 6.9 and 6.10). The purpose of this analysis is to identify possible differences and similarities between the strategies employed by these two strains to increase their survival at low pH, and the potential pathways that may be affected by indole during inhibition of AR2 response. GSEA is an appropriate tool for such analysis because it helps to identify changes that happen at the pathway level rather than gene level. Gene level changes can be misleading or may not provide sufficient information about the processes that are affected in the cell.

The KEGG pathway "Citrate/TCA cycle" was depleted under almost all four conditions in both the strains (Table 6.9). However, it was depleted less at pH 2.5 than pH 7 in both strains. Since TCA cycle is the main energy producing pathway in the cells, it is not surprising that mutants containing inserts in genes involved in this pathway are less fit. But GSEA results suggest that these mutants are at a smaller disadvantage at pH 2.5 than pH 7 in both the strains. As suggested previously in section 6.2.5.7, this could be because the priority of cells at pH 2.5 is to survive the stress rather than to grow and divide. The KEGG pathway "Citrate/TCA cycle" was depleted less at pH 2.5 in the presence of indole than ethanol in O157:H7 Sakai. Furthermore, in MG1655, this pathway was enriched at pH 2.5 in the presence of indole but depleted in the presence of ethanol. These observations suggest that at pH 2.5, inserts in genes involved in TCA cycle, lead to a smaller decline in fitness in O157:H7 Sakai in the presence of indole than ethanol. In MG1655, the inserts in genes involved in TCA cycle, lead to an increased fitness at pH 2.5 in the presence of indole but not ethanol. This could be because since indole inhibits the induction of AR2 response, the acid resistance of cells decreases making them more susceptible to damage. So the priority of cells to cope with acid

Table 6.9: KEGG metabolic pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions and at least one of the strains, relative to the ITL. The legend bars at the bottom show the colour corresponding to the normalised enrichment score (NES) and q value.

	5.5 + I -> 2.5		5.5 + E -> 2.5		5.5 + I -> 7		5.5 + E -> 7		5.5 + I -> 2.5		5.5 + E -> 2.5		5.5 + I -> 7		5.5 + E -> 7	
	MG	O157	MG	O157	MG	O157	MG	O157	MG	O157	MG	O157	MG	O157	MG	O157
Eco00020 (Citrate/TCA cycle)	1.02	-0.89	-0.93	-1.03	-1.17	-1.52	-1.05	-1.58	0.60	0.97	0.86	0.85	0.47	0.16	0.62	0.04
Eco00061 (Fatty acid biosynthesis)	2.12	1.81	1.92	1.34	1.82	1.61	2.07	-1.12	5.70×10^{-4}	0.03	0.02	0.48	0.04	0.15	2.10×10^{-3}	0.54
Eco00230 (Purine metabolism)	-1.71	-1.57	-1.83	-1.87	-2.01	-1.95	-1.95	-1.65	0.04	0.20	0.01	0.01	0.00	6.80×10^{-4}	6.10×10^{-4}	0.01
Eco00240 (Pyrimidine metabolism)	-1.23	-1.24	-1.36	-1.70	-1.49	-1.97	-1.46	-1.77	0.77	0.72	0.39	0.03	0.19	5.50×10^{-4}	0.22	9.30×10^{-4}
Eco00250 (Alanine, aspartate and glutamate metabolism)	-1.95	-1.74	-1.78	-1.68	-2.14	-1.99	-2.07	-1.64	2.97×10^{-3}	0.05	0.02	0.04	0.00	1.10×10^{-3}	0.00	0.02
Eco00260 (Glycine, serine and threonine metabolism)	1.24	-0.92	1.22	-1.30	0.92	-1.70	0.94	-1.33	0.39	0.98	0.41	0.49	0.74	0.03	0.68	0.15
Eco00300 (Lysine biosynthesis)	1.84	1.45	1.92	1.14	1.76	1.14	1.82	-1.10	0.02	0.32	0.02	0.63	0.06	0.57	0.03	0.56
Eco00400 (Phenylalanine, tyrosine and tryptophan biosynthesis)	0.61	-0.48	-1.32	-0.77	-1.76	-0.84	-1.72	-0.82	0.95	1.00	0.42	1.00	0.02	0.86	0.02	0.82
Eco00440 (Phosphonate and phosphinate metabolism)	1.95	-2.01	2.05	-1.86	2.03	-1.91	2.07	-1.59	0.01	3.13×10^{-3}	4.16×10^{-3}	0.01	3.63×10^{-3}	2.75×10^{-3}	2.21×10^{-3}	0.04
Eco00540 (Lipopolysaccharide biosynthesis)	-1.35	1.34	-1.62	1.18	-1.73	1.45	-1.76	1.18	0.68	0.48	0.09	0.59	0.02	0.27	0.01	0.94
Eco00550 (Peptidoglycan biosynthesis)	1.26	1.73	1.22	1.66	1.31	1.74	1.60	1.26	0.38	0.07	0.42	0.31	0.41	0.05	0.20	1.00
Eco00630 (Glyoxylate and dicarboxylate metabolism)	1.56	1.09	1.82	0.85	1.16	-1.06	1.35	-0.91	0.15	0.67	0.03	0.89	0.52	0.65	0.31	0.75
Eco00670 (One carbon pool by folate)	-1.10	-1.53	-1.17	-1.80	-1.35	-1.87	-1.39	-1.61	0.92	0.24	0.68	0.01	0.29	2.48×10^{-3}	0.29	0.03
Eco00740 (Riboflavin metabolism)	1.67	1.23	1.84	-1.33	1.80	1.34	1.61	-1.26	0.08	0.53	0.03	0.47	0.04	0.34	0.21	0.41
Eco00750 (Vitamin B6 metabolism)	-1.77	-1.78	-1.99	-1.83	-1.89	-1.80	-1.94	-1.66	0.03	0.05	1.05×10^{-3}	0.01	2.11×10^{-3}	0.01	8.30×10^{-4}	0.01
Eco00780 (Biotin metabolism)	1.81	0.92	1.14	-1.52	-1.08	-1.32	1.21	-1.56	0.03	0.85	0.45	0.18	0.60	0.33	0.45	0.05
Eco00900 (Terpenoid backbone biosynthesis)	1.80	1.73	1.84	1.24	1.74	1.88	1.95	1.12	0.03	0.08	0.03	0.56	0.06	0.02	0.01	0.93
Eco00920 (Sulfur metabolism)	-1.17	-0.87	-1.57	-1.03	-1.73	-1.32	-1.71	-1.18	0.89	0.97	0.13	0.81	0.02	0.33	0.02	0.50
Eco00970 (Aminoacyl-tRNA biosynthesis)	1.86	1.69	1.37	-1.43	1.58	1.31	1.72	-1.82	0.03	0.09	0.30	0.29	0.16	0.37	0.08	0.00
Eco01210 (2-Oxocarboxylic acid metabolism)	1.85	-0.49	1.50	-0.54	1.15	-1.17	1.03	-1.49	0.02	1.00	0.20	1.00	0.51	0.50	0.62	0.14
Eco01212 (Fatty acid metabolism)	2.15	1.85	1.84	1.38	1.91	1.62	2.09	-1.09	8.60×10^{-4}	0.03	0.03	0.49	0.02	0.17	1.79×10^{-3}	0.56
Eco01240 (Biosynthesis of cofactors)	1.36	-1.20	-1.54	-2.07	-1.73	-1.90	-1.65	-1.82	0.30	0.64	0.14	0.00	0.02	2.76×10^{-3}	0.04	0.00
Eco02010 (ABC transporters)	1.75	0.67	1.70	-0.89	1.40	-0.87	1.46	-0.96	0.05	0.97	0.07	0.93	0.40	0.88	0.26	0.71
Eco02040 (Flagellar assembly)	1.67	0.97	2.04	1.42	1.61	0.93	1.27	0.87	0.09	0.76	3.78×10^{-3}	0.44	0.14	0.87	0.39	0.90
Eco03010 (Ribosome)	2.71	1.95	2.62	-1.72	2.45	1.43	2.75	-1.72	0.00	0.01	0.00	0.03	0.00	0.28	0.00	3.84×10^{-3}
Eco03060 (Protein export)	1.32	1.64	1.21	1.31	1.10	1.78	1.25	1.08	0.33	0.13	0.42	0.49	0.52	0.04	0.41	0.97
Eco03070 (Bacterial secretion system)	1.11	1.63	0.89	1.17	1.02	1.79	1.03	1.24	0.48	0.12	0.70	0.58	0.60	0.06	0.63	1.00



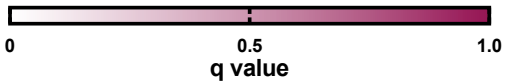
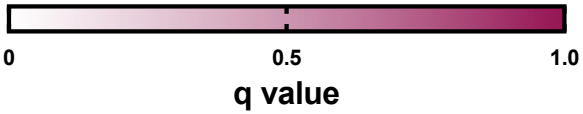


Table 6.10: Transcriptional regulatory pathways which were significantly depleted or enriched in at least one of the four TraDIS experimental conditions and at least one of the strains, relative to the ITL. Black space indicates where the pathway was neither depleted nor enriched in a condition. The legend bars at the bottom show the colour corresponding to the normalised enrichment score (NES) and q value.

	5.5 + I -> 2.5		5.5 + E -> 2.5		5.5 + I -> 7		5.5 + E -> 7		5.5 + I -> 2.5		5.5 + E -> 2.5		5.5 + I -> 7		5.5 + E -> 7	
	MG	O157	MG	O157	MG	O157	MG	O157	MG	O157	MG	O157	MG	O157	MG	O157
ArgP	1.88	1.64	2.03	1.05	1.81	1.54	1.92	-0.93	0.02	0.19	2.64×10^{-3}	0.90	0.04	0.69	0.01	1.00
BirA	-1.42	-1.57	-1.75	-1.59	-1.78	-1.74	-1.79	-1.55	0.23	0.26	0.03	0.44	0.03	0.03	0.01	0.16
BluR	-1.74		-1.67		-1.67		-1.66		0.04		0.04		0.08		0.06	
EvgA	-1.80	-1.63	-1.71	-1.45	-1.49	-1.38	-1.61	-0.79	0.03	0.23	0.03	0.58	0.23	0.48	0.08	1.00
FadR	1.93	1.87	1.97	1.35	1.85	1.74	2.07	-1.15	0.02	0.01	3.98×10^{-3}	1.00	0.05	0.19	0.01	1.00
FhlA	1.65	-0.63	1.90	-0.63	1.82	-0.60	1.83	-0.81	0.22	1.00	0.01	1.00	0.05	1.00	0.03	1.00
FlhDC	1.71	1.02	2.10	1.25	1.68	1.06	1.30	0.94	0.16	1.00	1.22×10^{-3}	1.00	0.18	1.00	0.48	0.95
ModE	1.53	0.81	1.83	0.74	1.82	0.93	1.92	0.94	0.26	1.00	0.03	1.00	0.06	1.00	0.01	0.93
MraZ	1.89	1.99	2.04	2.01	1.86	1.72	1.94	1.30	0.02	9.50×10^{-4}	2.65×10^{-3}	1.38×10^{-3}	0.10	0.14	0.01	1.00
OmpR	-1.76	-1.37	-1.87	-0.72	-1.65	-1.34	-1.80	-0.75	0.04	0.55	4.92×10^{-3}	1.00	0.08	0.54	0.02	1.00
PdhR	1.58	1.83	1.47	1.77	1.48	1.68	1.43	-0.87	0.20	0.02	0.29	0.12	0.39	0.13	0.40	1.00
PhoB	1.62	-1.84	1.89	-1.54	1.54	-1.71	1.65	-1.40	0.23	0.02	0.01	0.32	0.28	0.03	0.23	0.54
PurR	-2.40	-2.18	-2.23	-2.20	-2.19	-2.16	-2.29	-1.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RbsR	-1.50	-1.59	-1.68	-1.49	-1.71	-1.81	-1.65	-1.57	0.19	0.26	0.04	0.44	0.06	0.01	0.06	0.14





stress and improve their survival rather than to grow and divide, increases more in the presence of indole than ethanol. This strategy is used by both MG1655 and O157:H7 Sakai, according to the GSEA results.

The KEGG pathway "phenylalanine, tyrosine and tryptophan biosynthesis" was also depleted under almost all four conditions in both the strains (Table 6.9). Since amino acids are the building blocks of peptides and proteins, it is reasonable that inserts in genes involved in the biosynthesis of these amino acids reduce the fitness of cells. The loss of fitness, however, is smaller at pH 2.5 than pH 7 in both the strains. Furthermore, in O157:H7 Sakai, this pathway was depleted less at pH 2.5 in the presence of indole than ethanol. In MG1655, this pathway was enriched at pH 2.5 in the presence of indole but depleted in the presence of ethanol. As suggested above, this could be because the priority of cells at pH 2.5 is to survive the stress rather than to grow and divide. Since indole inhibits the induction of AR2 response, reducing the acid resistance of cells and making them more susceptible to damage, the priority of cells to cope with acid stress and improve their survival rather than to grow and divide increases even more in the presence of indole than ethanol.

The KEGG pathway "glyoxylate and dicarboxylate metabolism" was enriched under all four conditions, but more at pH 2.5 than pH 7 in MG1655 (Table 6.9). Furthermore, this pathway was depleted at pH 7 but enriched at pH 2.5 in O157:H7 Sakai. These results suggest that inserts in genes which are involved in the biosynthesis of carbohydrates improve the fitness of cells at pH 2.5 in both strains. This observation is consistent with the hypothesis by Maurer *et al.* (2005) that carbohydrate metabolism is preferred in alkaline pH to generate fermentation acids which can lower the internal pH. The enrichment of this pathway at pH 7 in MG1655 but depletion in O157:H7 Sakai could be because MG1655 is more sensitive to reduction in pH from 7 to 5.5 than O157:H7 Sakai. Due to this, MG1655 stops carbohydrate

metabolism as the pH drops to 5.5. Therefore, inserts in genes involved in "glyoxylate and dicarboxylate metabolism" increase the fitness of MG1655 at pH 7 but reduce the fitness of O157:H7 Sakai.

6.3 Discussion

The first aim of the study described in this Chapter was to identify genes whose function is needed for the optimal fitness of O157:H7 Sakai at pH 2.5, and genes that are involved in the inhibitory effect of indole on AR2 induction. Mutants containing insertions in several AR2 genes such as *evgS, evgA*, *gadE, ydeP, slp*, *gadW, dctR,* and *hdeD,* showed a bigger decline in fitness at pH 2.5 than pH 7 (section 6.2.5.5). This is as expected, because the AR2 system provides resistance against extreme acid stress of pH 2.5 and below. The decline in the fitness of several AR2 genes in TraDIS experiments performed with both MG1655 (Chapter 5, section 5.2.2.5) and O157:H7 Sakai suggests that the protocol used for TraDIS experiments in this study can successfully identify genes whose function is needed for optimal fitness of cells at pH 2.5.

TraDIS results in this Chapter suggest that the function of several additional genes including *arnB, waaO, waaL, waaR, waaD, wzy, wcaG, gmd, wcaE,* and *cpsB* are needed for optimal fitness of O157:H7 Sakai at pH 2.5 (section 6.2.5.6 and Table 6.5). All these genes are involved in the synthesis of extracellular polysaccharides such as LPS and colanic acid, suggesting that these cell components may impart an optimal fitness to cells at low pH. Furthermore, two genes ECs_2840 and ECs_2845, whose specific function is unknown, but they could potentially be involved in the synthesis of extracellular polysaccharide, are also suggested to provide optimal fitness to cells at pH 2.5. QseC, a sensor kinase, may also provide

optimal fitness to cells at pH 2.5. Additionally, its function may be more important in the presence of indole than ethanol, for the survival of cells at pH 2.5.

The second aim of the study in this Chapter was to compare the acid stress responses in MG1655 with those in O157:H7 Sakai. Analysis performed on individual genes revealed that in MG1655, loss of function of *wec* genes, which are involved in the biosynthesis of enterobacterial common antigen, led to a fitter phenotype at pH 2.5 than pH 7, implying that loss of function of these genes can increase acid resistance (Chapter 5, section 5.2.2.6 and Table 5.4). However, in O157:H7 Sakai, insertions in several genes involved in LPS and colanic acid biosynthesis had a negative effect on fitness at pH 2.5. The two strains therefore modify their surface capsule in different ways when exposed to extreme acid stress. Further experiments are required to establish the nature of these modifications which could involve an increase or decrease in the amount of surface capsule produced, changes in the net charge of the surface capsule etc.

GSEA was also used to compare the acid stress responses in MG1655 and O157:H7 Sakai because it is a statistically powerful tool and it allows to determine the processes which are changing in the cells at genome level. Comparison of GSEA results from MG1655 and O157:H7 Sakai suggests that same KEGG metabolic pathways such as "citrate/TCA cycle", "phenylalanine, tyrosine and tryptophan biosynthesis", and "glyoxylate and dicarboxylate metabolism" determine the fitness of cells at pH 2.5. However, subtle differences also exist in the response of two strains.

Chapter 7

Discussion

7.1 Summary of key results

The four main aims of this study were:

(a) to determine if the inhibitory effect of indole, previously reported with exponentially growing K-12 strains MG1655 and BW25113 on the induction of acid resistance, is conserved in pathogenic (O157:H7 Sakai) and commensal (Nissle 1917) strains;

(b) to determine whether the acid resistance can be induced and subsequently blocked with indole, through incubation at mildly acidic pH, in early stationary and stationary growth phases;

(c) to examine and compare the effect of low pH between MG1655 and O157:H7 Sakai using TraDIS;

(d) to understand the molecular mechanism behind the impact of indole on the acid stress response. The summary of key results in relation to these aims is in the following sections.

Using survival assays in Chapter 4, it was shown that the acid stress response, produced by incubation at pH 5.5, can be successfully induced in exponentially growing pathogenic strain O157:H7 Sakai and commensal strain Nissle 1917 (Figures 4.9 A and 4.10 A). This was evident through the higher survival of induced O157:H7 Sakai cells and Nissle 1917 *ΔbcsA* cells acid stressed at pH 2.5 than the uninduced cells. Furthermore, the induction of the acid stress response in O157:H7 Sakai can be blocked with indole, reducing the survival of induced cells (Figure 4.9 A). The induction of the acid stress response in laboratory, pathogenic and commensal strains was also tested in later growth phases (Figures 4.6 - 4.10). O157:H7 Sakai and Nissle 1917 *ΔbcsA* were the only two strains that showed significant induction of the acid stress response in early stationary phase but this induction response could not be blocked with indole. Only Nissle 1917 *ΔbcsA* of the four tested strains (MG1655,

BW25113, O157:H7 Sakai and Nissle 1917 *ΔbcsA*) showed significant induction of the acid stress response in stationary growth phase, but this response could not be blocked with indole. Additionally, the survival of Nissle 1917 *ΔbcsA,* even after the induction of acid stress response was really low in all three growth phases.

In Chapters 5 and 6, TraDIS was used to study the response of MG1655 and O157:H7 Sakai, respectively, to low pH. During TraDIS experiments conducted with both strains, mutants containing insertions in several AR2 genes such as *evgS, evgA*, and *gadE,* showed a bigger decline in fitness at pH 2.5 than pH 7 (Tables 5.1 and 6.3). This indicates that the function of these genes is needed for the optimal fitness of cells at pH 2.5, which is consistent with the known role of AR2 response in providing resistance against extreme acid stress of pH 2.5 and below. These results suggest that the protocol used for TraDIS experiments can successfully identify genes whose function is needed for optimal fitness of cells at low pH. Several genes which have previously not been shown to have a role in extreme acid stress also came up as significant at pH 2.5 for both MG1655 and O157:H7 Sakai (Tables 5.3 and 6.5). Additionally, validation experiments performed with an *mrcA* mutant demonstrated that the function of *mrcA* is needed for the optimal fitness of MG1655 at pH 2.5, confirming the TraDIS results (Figure 5.13).

Comparison between the acid stress responses of MG1655 and O157:H7 Sakai using GSEA suggests that same KEGG metabolic pathways such as "Citrate/TCA cycle", "phenylalanine, tyrosine and tryptophan biosynthesis", and "glyoxylate and dicarboxylate metabolism" are important in determining the fitness of cells at pH 2.5 (Table 6.9). Subtle differences, however, are present in the response of two strains.

TraDIS experiments performed with MG1655 showed that mutants containing insertions in several AR2 genes had a bigger decline in fitness at pH 2.5 in the presence of

ethanol than indole (Table 5.1). This observation is consistent with our hypothesis that indole blocks the induction of AR2 response, making the role of AR2 genes less important for the optimal fitness of cells at pH 2.5 in the presence of indole. Additionally, TraDIS data for mutants of the *fis* gene in MG1655 showed interesting results (Table 5.3). This is the only gene where inserts caused (a) a significant loss of fitness at pH 2.5 in the presence of indole, but no effect in the presence of ethanol, (b) a significant increase in fitness at pH 7 in the presence of indole but again, no effect in the presence of ethanol. These results suggest that there is an interesting connection between indole and the function of Fis, the mechanism of which remains to be elucidated.

7.2 Limitations of the study

The first limitation of this study is that the EdgeR outputs for comparisons between control and conditional TraDIS libraries contain both essential and non-essential genes. Essential genes normally do not tolerate any insertions, or tolerate only a few insertions. The conditional essentiality of such genes cannot be assayed unless a TraDIS-Xpress library is used and grown in the presence of IPTG. The expression of essential genes present downstream of the transposon can then be induced from the *tac* promoter in the transposon. In contrast, sometimes it is possible that only the function of a particular domain is essential and the remaining gene can tolerate a high number of insertions. The conditional essentiality of such genes can be determined through EdgeR comparisons.

For experiments done with a normal TraDIS library or a TraDIS-Xpress library grown without IPTG, EdgeR can provide a high $\log_2$FC values as well as low q values for essential genes with no or few insertions. One such example is *ssb* from Chapter 6. This gene was present among the top 10 genes with $\log_2$FC < -2 or > 2, q value < 0.05 and logCPM > 8 (Figure

6.11). Upon checking the TraDIS results of ITL, *ssb* was found to be essential with no transposon insertions. Such essential genes, therefore, need to be manually removed from EdgeR outputs.

The second limitation of this study is that TraDIS can give false positive results. An example of an apparent false positive result is *nlpI* in Chapter 5. According to TraDIS results, inserts in *nlpI* lead to a significant decline in fitness at pH 2.5 ($log_2FC < -2$) but not at pH 7 (Table 5.3). However, competition assays done between KH001 and MG1655 *ΔnlpI* at pH 7, and survival assays done with MG1655 *ΔnlpI* at pH 2.5, showed that the function of *nlpI* is needed for the optimal fitness of cells at pH 7 but not at pH 2.5, in contrast to TraDIS results (Figure 5.13). Validation experiments are therefore always necessary to check the reliability of TraDIS results for individual genes.

The third limitation of this study is that the indole stock used in TraDIS experiments was prepared in 100% ethanol. Ethanol has been shown to induce a variety of stress responses such as acid stress, osmotic stress and heat shock stress in *E. coli* strains exposed to 2.5% (v/v) of ethanol for 15 minutes (Cao *et al.*, 2017). Ethanol also reduces aerobic respiration and ATP production, and damages cell wall and membrane integrity. During TraDIS experiments, however, both MG1655 and O157:H7 Sakai were exposed to 1% (v/v) ethanol which is a smaller concentration than the one used by Cao *et al.* (2017). Furthermore, the strains were only exposed to ethanol for 30 minutes during the induction step at pH 5.5. When MG1655 and O157:H7 Sakai were induced in the presence and absence of ethanol followed by acid shock at pH 2.5, the difference in the survival under both conditions was non-significant for both strains (Figures 4.5 and 4.9 A). This result suggests that 1% ethanol does not affect the survival of these strains significantly. Nevertheless, TraDIS experiments

performed in the absence of ethanol during the induction step would have made it clear whether ethanol affects the fitness of these two strains or not.

7.3 Future work

Due to time constraints, the validation experiments to confirm TraDIS results in MG1655 in Chapter 5 could only be done with three mutants: *mrcA, mrcB* and *nlpI.* Several other genes were hypothesised to have a potential role in providing optimal fitness to cells at pH 2.5 in section 5.2.2.6 (Chapter 5), including *clsA, yciB, lpxM,* and *emrR.* More validation experiments with single knockout mutants lacking these genes are required to check the reliability of TraDIS results in MG1655.

Validation of TraDIS results in O157:H7 Sakai in Chapter 6 is also pending. The validation experiments in O157:H7 Sakai need to be performed with single mutants lacking genes which have been and will be tested to confirm the validity of MG1655 TraDIS experiments. Comparisons of results from validation experiments in MG1655 and O157:H7 Sakai will help to determine how similar the response of two strains is towards low pH. Other suitable candidates for validation experiments would be the genes hypothesised to have a potential role in providing optimal fitness to cells at pH 2.5 in section 6.2.5.6 (Chapter 6). Examples of such genes include *waaO, waaL, waaR, wzy, wcaG*, *gmd, wcaE, qseC* and *cpsB.*

TraDIS results in MG1655 suggest that there is an interesting connection between indole and the function of Fis. Validation experiments are first required to confirm TraDIS results. Further research is then needed on finding and testing the potential mechanisms that are involved in the effects exhibited by indole on cells through Fis.

The potential importance of essential genes in low pH was not assayed in this study due to time constraints. A TraDIS-Xpress transposon library, as opposed to a normal TraDIS

library, was prepared in O157:H7 Sakai so that future work could use this library to look at the role of essential genes in low pH. This could be done by repeating the TraDIS experiments performed in Chapter 6 but in the presence of IPTG to either overexpress or repress essential genes.

Appendices

8.1 Whole genome sequencing results of MG1655, BW25113, O157:H7 Sakai, and Nissle 1917

8.1.1 Predicted mutations in MG1655

The explanation for different types of evidence (MC, JC and RA) is provided below the table.

Evidence	Position	Mutation	Annotation	Gene	Description
MC JC	257,908	Δ776 bp		*[crl]*	[crl]
RA	803,662	C→A	L54I (CTT→ATT)	*ybhJ* →	aconitase family protein
MC JC	1,299,499	Δ1,199 bp	intergenic (+254/-485)	*ychE* → / → *oppA*	UPF0056 family inner membrane protein/oligopeptide ABC transporter periplasmic binding protein
RA	1,905,761	G→A	G25D (GGT→GAT)	*mntP* →	putative Mn(2+) efflux pump, mntR-regulated
MC JC	1,978,503	Δ776 bp		*insB1–insA*	insB1, insA
RA	2,173,363	Δ2 bp	intergenic (-1/+1)	*gatC* ← / ← *gatC*	pseudogene, galactitol-specific enzyme IIC component of PTS;transport; Transport of small molecules: Carbohydrates, organic acids, alcohols; PTS system galactitol-specific enzyme IIC/pseudogene, galactitol-specific

					enzyme IIC component of PTS;transport; Transport of small molecules: Carbohydrates, organic acids, alcohols; PTS system galactitol-specific enzyme IIC
RA	3,560,455	+G	intergenic (-2/+1)	*glpR* ← / ← *glpR*	pseudogene, DNA-binding transcriptional repressor;regulator; Energy metabolism, carbon: Anaerobic respiration; repressor of the glp operon/pseudogene, DNA-binding transcriptional repressor;regulator; Energy metabolism, carbon: Anaerobic respiration; repressor of the glp operon
RA	3,938,887	G→T	E221* (GAA→TAA)	*rbsR* →	transcriptional repressor of ribose metabolism
RA	4,296,381	+GC	intergenic (+587/+55)	*gltP* → / ← *yjcO*	glutamate/aspartate:proton symporter/Sel1 family TPR-like repeat protein

Missing coverage evidence (MC): Breseq examines read pileups to predict deletions when it encounters regions with missing and low read coverage in the reference genome.

New junction evidence (JC): Breseq looks for mosaic read alignments which indicate the presence of new junctions between disjoint regions of the reference genome in the sample.

Read alignment evidence (RA): Breseq examines read pileups mapped to each position in the reference genome to predict base substitution mutations and small indels.

Further information on different types of evidence to predict mutations in the sample genome can be found here: https://barricklab.org/twiki/pub/Lab/ToolsBacterialGenomeResequencing/documentation/methods.html#missing-coverage-evidence-mc.

8.1.2 Predicted mutations in BW25113

Evidence	Position	Mutation	Annotation	Gene	Description
RA	4,029,087	A→T	noncoding (199/1556 nt)	*BW25113_RS19980* →	16S ribosomal RNA

8.1.3 Predicted mutations in O157:H7 Sakai

8.1.3.1 Chromosome

Evidence	Position	Mutation	Annotation	Gene	Description
MC JC	1,267,023	1181 bp→84 bp		*stx2A–[stx2B]*	*stx2A, [stx2B]*
RA	2,427,339	A→C	G190G (GGT→GGG)	*astB* ←	succinylarginine dihydrolase
MC JC	2,925,414	Δ624 bp		*[stx1A]*	*[stx1A]*

RA	3,900,381	A→T	I415N (ATC→AAC)	*ygiQ* ←	radical SAM superfamily protein

8.1.3.2 plasmid po157

Evidence	Position	Mutation	Annotation	Gene	Description
RA	4,640	$(G)_{10\rightarrow 9}$	intergenic (+743/-170)	*repA* → / → *Z_RS28290*	incFII family plasmid replication initiator RepA/ribbon-helix-helix domain-containing protein
JC	22,675	$(T)_{21\rightarrow 17}$	intergenic (-111/-338)	*Z_RS28385* ← / → *stcE*	transposase/metalloprotease StcE
RA	25,085	C→T	T690T (ACC→ACT)	*stcE* →	metalloprotease StcE
RA	71,319	A→G	E269E (GAA→GAG)	*Z_RS28665* →	ParB/RepB/Spo0J family partition protein
JC	76,979	$(CCTCACA)_{8\rightarrow 10}$	intergenic (+217/-4)	*Z_RS28695* → / → *Z_RS33365*	hypothetical protein/APH(6) family putative aminoglycoside O-phosphotransferase
RA	79,753	T→C	N404N (AAT→AAC)	*toxB* →	toxin B
RA	83,563	G→A	G1674G (GGG→GGA)	*toxB* →	toxin B

8.1.3.3 plasmid pOSAK1

No mutations predicted.

8.1.4 Nissle 1917

8.1.5 Available at:

https://www.dropbox.com/sh/w7jdc4l2ai3yh9r/AAAZZz0GX5vLuF16hHN2t3Pda?dl=0Nissle 1917 *ΔbcsA*

Available at:
https://www.dropbox.com/sh/w7jdc4l2ai3yh9r/AAAZZz0GX5vLuF16hHN2t3Pda?dl=0

8.2 Results of TraDIS experiments in MG1655

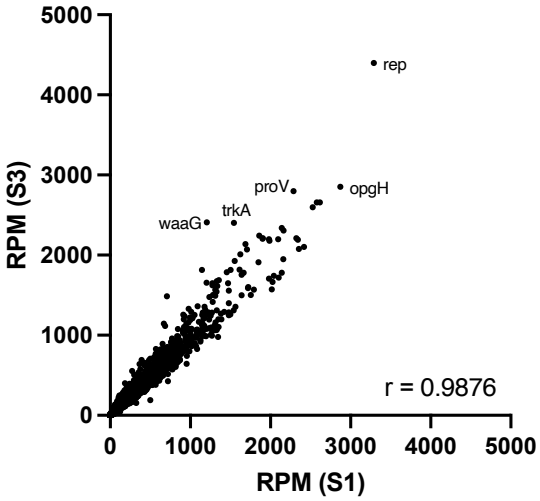
8.2.1 LogFC, q value and logCPM of all genes in each condition relative to ITL

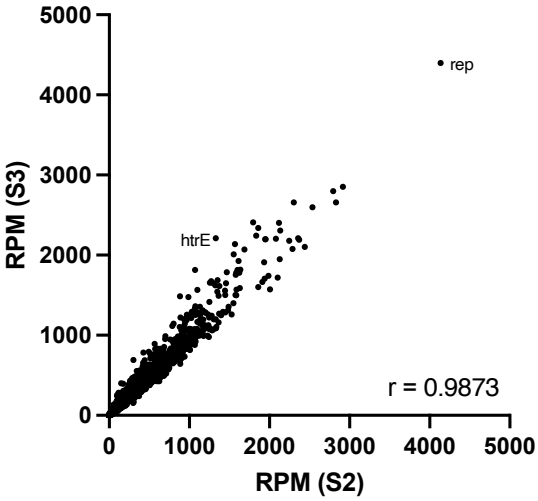
Available at:
https://www.dropbox.com/sh/w7jdc4l2ai3yh9r/AAAZZz0GX5vLuF16hHN2t3Pda?dl=0

8.2.2 Reproducibility among replicates in each experimental condition

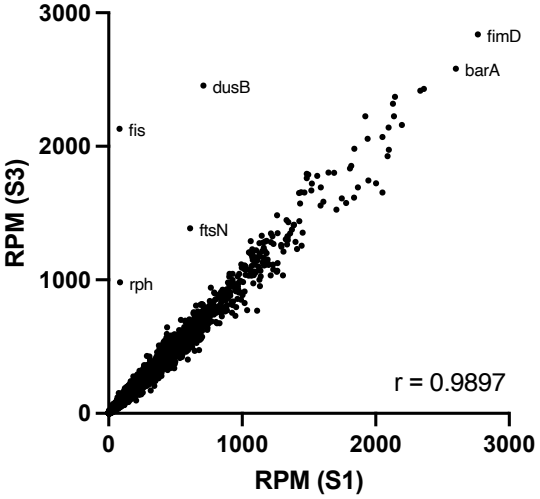
8.2.2.1 RPM

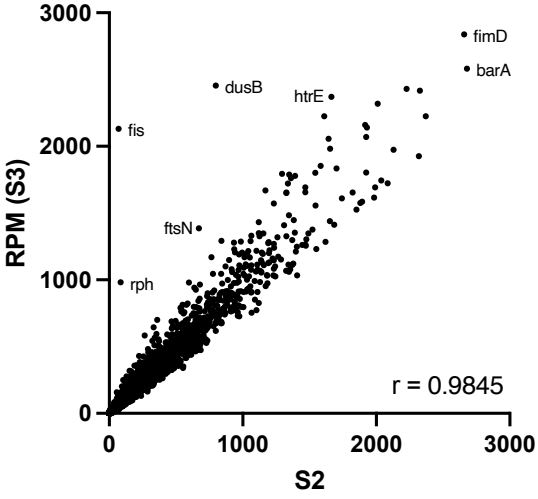
1) 5.5 + Indole -> 2.5



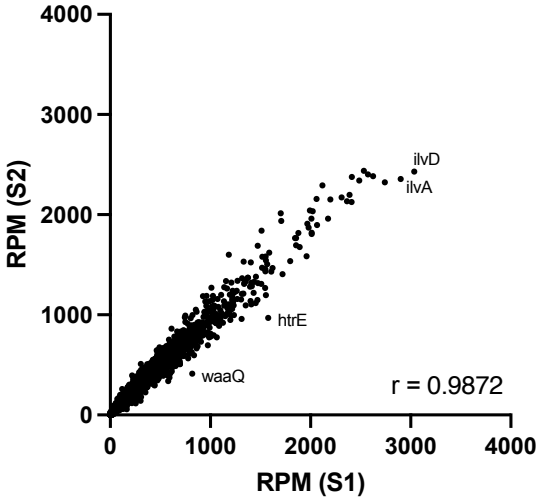


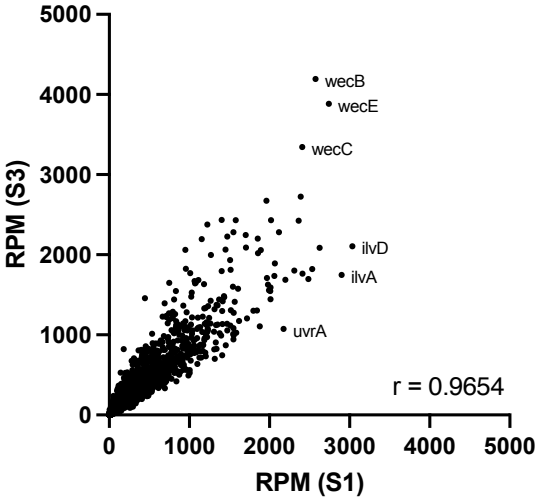
2) 5.5 + Indole -> 7



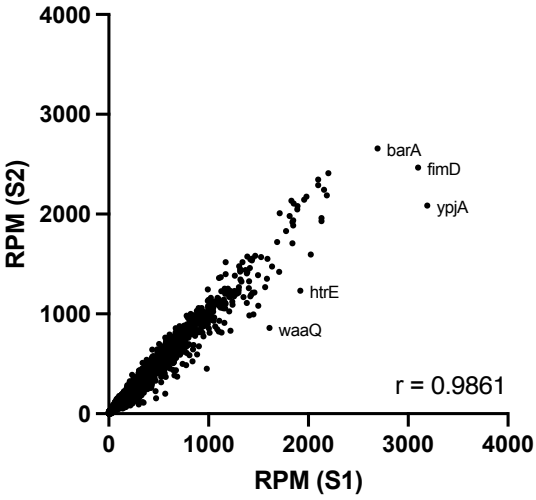


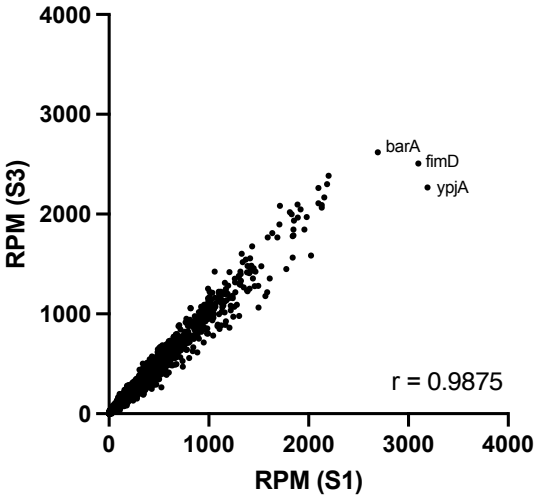
3) 5.5 + Ethanol -> 2.5





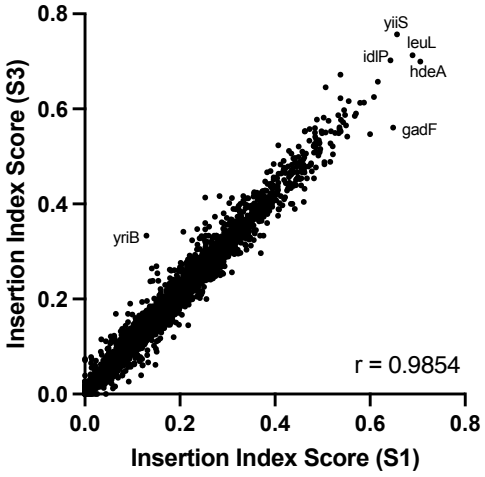
4) 5.5 + Ethanol -> 7

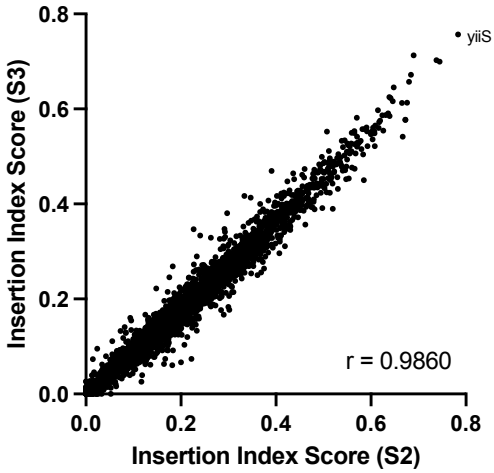




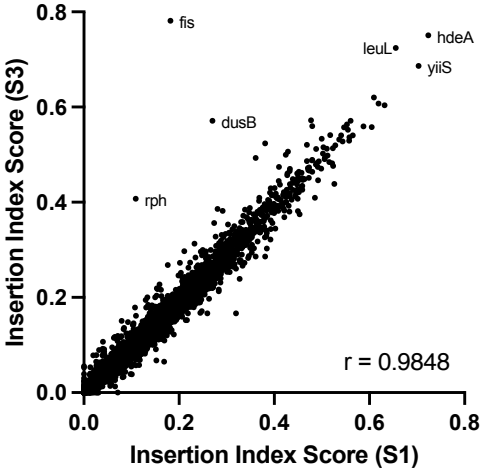
8.2.2.2 Insertion index score

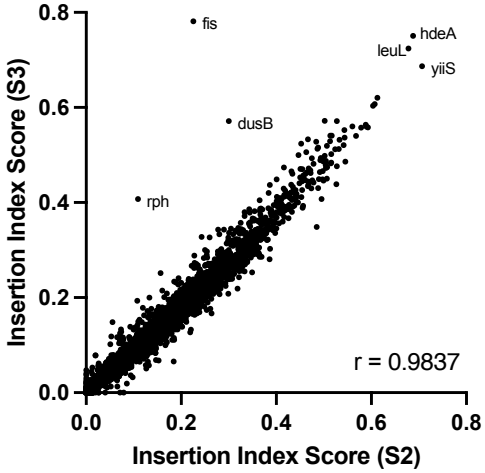
1) 5.5 + Indole -> 2.5



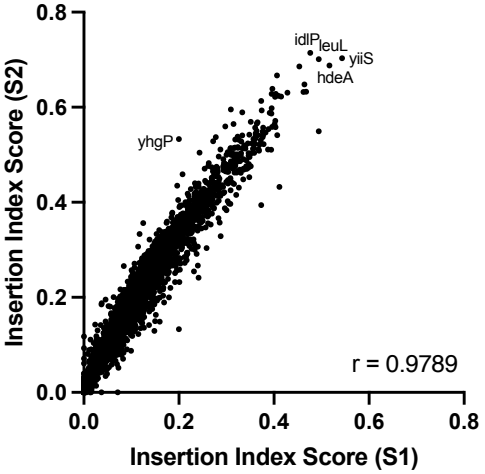


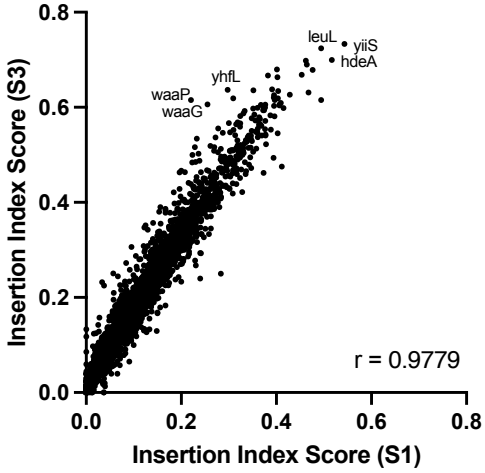
2) 5.5 + Indole -> 7



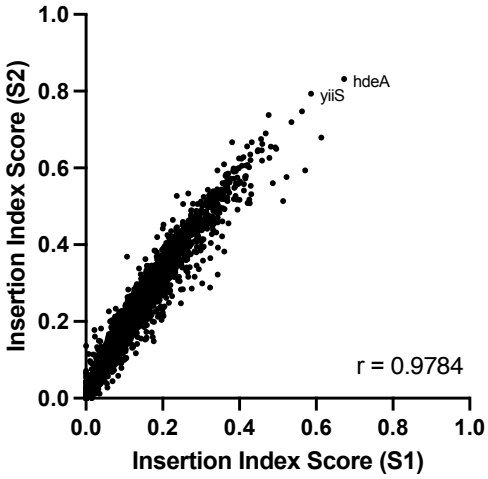


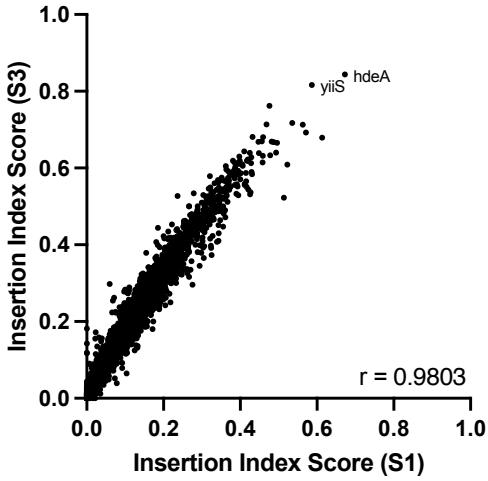
3) 5.5 + Ethanol -> 2.5





4) 5.5 + Ethanol -> 7





8.2.3 Genes with depleted insertions

8 genes exclusively in "5.5 + I -> 2.5"		37 genes exclusively in "5.5 + E -> 2.5"		15 common genes in "5.5 + I -> 2.5" and "5.5 + E -> 2.5"	
Gene	**Function**	**Gene**	**Function**	**Gene**	**Function**
hipB	antitoxin/DNA-binding transcriptional repressor HipB	*hscB*	[Fe-S] cluster biosynthesis co-chaperone HscB	*yciB*	inner membrane protein
ymgK	protein YmgK	*iscU*	scaffold protein for iron-sulfur cluster assembly	*ldcA*	murein L
ymiB	putative protein YmiB	*evgA*	DNA-binding transcriptional activator EvgA	*degP*	periplasmic serine endoprotease DegP
fis	DNA-binding transcriptional dual regulator Fis	*iraM*	anti-adaptor protein IraM	*hfq*	RNA-binding protein Hfq
minD	Z-ring positioning protein MinD	*yjbS*	uncharacterized protein YjbS	*ydfO*	Qin prophage; DUF1398 domain-containing protein YdfO
ybfQ	inactive transposase YbfQ	*evgS*	sensor histidine kinase EvgS	*ihfB*	integration host factor subunit beta
clsA	cardiolipin synthase A	*epmA*	EF-P-lysine lysyltransferase	*ihfA*	integration host factor subunit alpha
rimP	ribosome maturation factor RimP	*iscA*	iron-sulfur cluster insertion protein IscA	*yciG*	stress-induced bacterial acidophilic repeat motifs-containing protein YciG
		insA2	IS1 protein InsA	*ynfT*	Qin prophage; protein YnfT
		atpD	ATP synthase F1 complex subunit beta	*lpoA*	outer membrane lipoprotein - activator of MrcA activity

		rlmE	23S rRNA 2'-O-ribose U2552 methyltransferase	*atpB*	ATP synthase Fo complex subunit a
		rzoQ	Qin prophage; putative lipoprotein RzoQ	*wbbJ*	putative lipopolysaccharide biosynthesis O-acetyl transferase WbbJ
		atpF	ATP synthase Fo complex subunit b	*mrcA*	peptidoglycan glycosyltransferase/peptidoglycan DD-transpeptidase MrcA
		yehC	putative fimbrial chaperone YehC	*nlpI*	lipoprotein NlpI
		waaU	putative ADP-heptose:LPS heptosyltransferase 4	*bluF*	blue light- and temperature-regulated antirepressor BluF
		waaB	UDP-D-galactose:(glucosyl)lipopolysaccharide-1		
		lpxM	Lipid A biosynthesis myristoyltransferase		
		valZ	tRNA-Val(UAC)		
		gadE	DNA-binding transcriptional activator GadE		
		miaA	tRNA dimethylallyltransferase		
		yqeH	putative LuxR family transcriptional regulator YqeH		
		ydeQ	putative fimbrial adhesin protein YdeQ		
		emrR	DNA-binding transcriptional repressor MprA		
		aroK	shikimate kinase 1		

		atpG	ATP synthase F1 complex subunit gamma		
		yncI	pseudogene		
		tusE	sulfur transfer protein TusE		
		topA	DNA topoisomerase 1		
		hdeA	periplasmic acid stress chaperone HdeA		
		cspF	Qin prophage; cold shock protein CspF		
		yibY	protein YibY		
		atpA	ATP synthase F1 complex subunit alpha		
		holC	DNA polymerase III subunit chi		
		ymgF	inner membrane protein YmgF		
		fepC	ferric enterobactin ABC transporter ATP binding subunit		
		valT	tRNA-Val(UAC)		
		rrfF	5S ribosomal RNA		

8.2.4 Genes with enriched insertions

25 genes exclusively in "5.5 + I -> 2.5"		6 genes exclusively in "5.5 + E -> 2.5"		7 common genes in "5.5 + I -> 2.5" and "5.5 + E -> 2.5"		5 genes exclusively in "5.5 + I -> 7"		10 common genes in "5.5 + I -> 2.5" and "5.5 + I -> 7"	
Gene	**Function**	**Gene**	**Function**	**Gene**	**Function**	**Gene**	**Function**	**Gene**	**Function**
prfB	peptide chain release factor RF2	*nagC*	DNA-binding transcriptional dual regulator NagC	*wecF*	TDP-N-acetylfucosamine :lipid II N-acetylfucosaminyl transferase	*rpsC*	30S ribosomal subunit protein S3	*ftsW*	essential cell division protein FtsW
coaD	pantetheine-phosphate adenylyltransferase	*wecB*	UDP-N-acetylglucosamine 2-epimerase	*mreD*	cell shape determining protein MreD	*thyA*	thymidylate synthase	*tmk*	hypothetical protein
lptF	lipopolysaccharide transport system protein LptF	*can*	carbonic anhydrase 2	*rffC*	dTDP-4-amino-4,6-dideoxy-D-galactose acyltransferase	*frr*	ribosome-recycling factor	*pssA*	phosphatidylserine synthase
accB	biotin carboxyl carrier protein	*rpoH*	RNA polymerase, sigma 32 (sigma H) factor	*wecE*	dTDP-4-dehydro-6-deoxy-D-glucose transaminase	*fis*	DNA-binding transcriptional dual regulator Fis	*rpsL*	30S ribosomal subunit protein S12

yqhJ	protein YqhJ	*wecC*	UDP-N-acetyl-D-mannosamine dehydrogenase	*nagA*	N-acetylglucosamine-6-phosphate deacetylase	*rpmI*	50S ribosomal subunit protein L35	*rpsI*	30S ribosomal subunit protein S9
yncL	uncharacterized protein YncL	*adk*	adenylate kinase	*gapA*	glyceraldehyde-3-phosphate dehydrogenase A			*rplX*	50S ribosomal subunit protein L24
nrdB	ribonucleoside-diphosphate reductase 1 subunit beta			*kdsA*	3-deoxy-D-manno-octulosonate 8-phosphate synthase			*racR*	Rac prophage; DNA-binding transcriptional repressor RacR
pyrG	CTP synthetase							*rplW*	50S ribosomal subunit protein L23
nrdA	ribonucleoside-diphosphate reductase 1 subunit alpha							*pgsA*	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
opgH	osmoregulated periplasmic glucans (OPGs) biosynthesis protein H							*rpsS*	30S ribosomal subunit protein S19

prs	ribose-phosphate diphosphokinase								
alaS	alanine--tRNA ligase/DNA-binding transcriptional repressor								
rep	ATP-dependent DNA helicase Rep								
nadK	NAD kinase								
fabZ	*3-hydroxy-acyl-[acyl-carrier-protein] dehydratase*								
rpsB	30S ribosomal subunit protein S2								
rpoA	RNA polymerase subunit alpha								
rplN	50S ribosomal subunit protein L14								
ymdG	protein YmdG								

rpmJ	50S ribosomal subunit protein L36								
ribA	GTP cyclohydrolase 2								
rpsQ	30S ribosomal subunit protein S17								
rplJ	50S ribosomal subunit protein L10								
rpsM	30S ribosomal subunit protein S13								
rplL	50S ribosomal subunit protein L12								

8.2.5 Comparison of LogFC and q value of genes known to have a role in acid stress, among four conditions

Available at:
https://www.dropbox.com/sh/w7jdc4l2ai3yh9r/AAAZZz0GX5vLuF16hHN2t3Pda?dl=0

8.2.6 List of essential, non-essential and unclear genes in MG1655

Available at:
https://www.dropbox.com/sh/w7jdc4l2ai3yh9r/AAAZZz0GX5vLuF16hHN2t3Pda?dl=0

8.3 Results of TraDIS experiments in O157:H7 Sakai

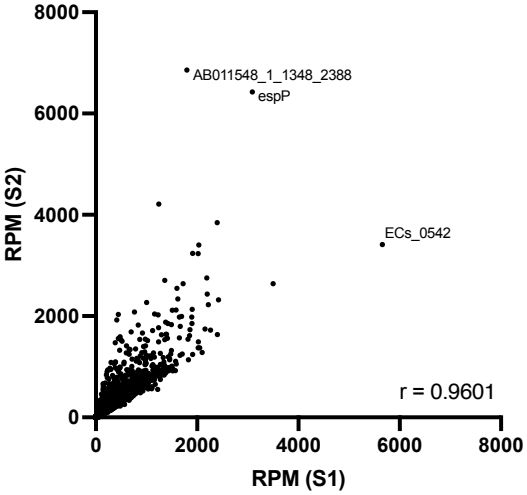
8.3.1 LogFC, q value and logCPM of all genes in each condition relative to ITL

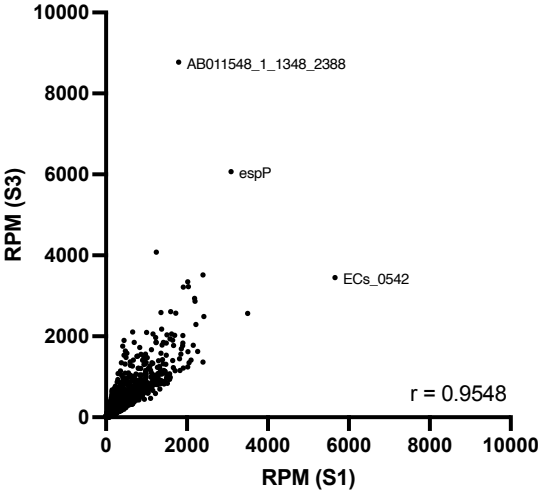
Available at:
https://www.dropbox.com/sh/w7jdc4l2ai3yh9r/AAAZZz0GX5vLuF16hHN2t3Pda?dl=0

8.3.2 Reproducibility among replicates in each experimental condition

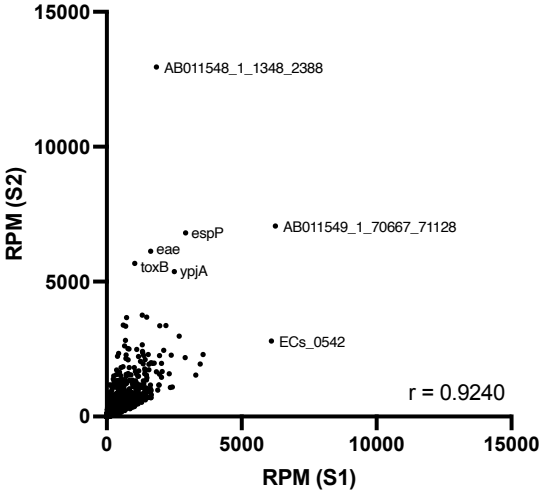
8.3.2.1 RPM

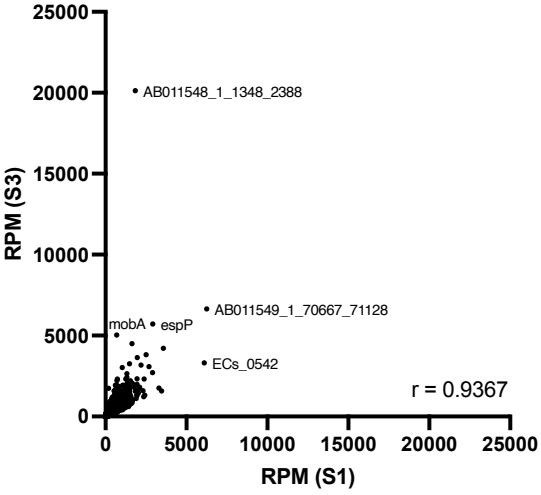
1) 5.5 + Indole -> 2.5



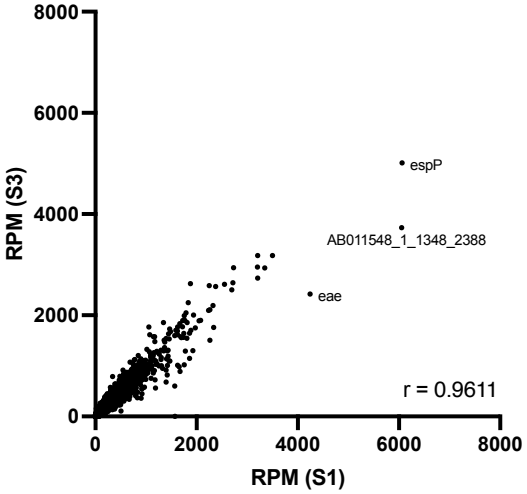


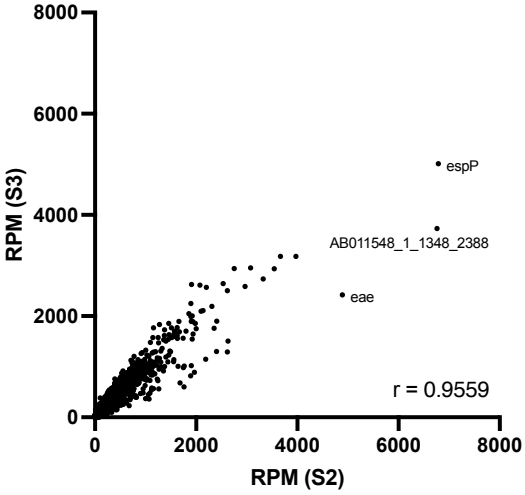
2) 5.5 + Indole -> 7



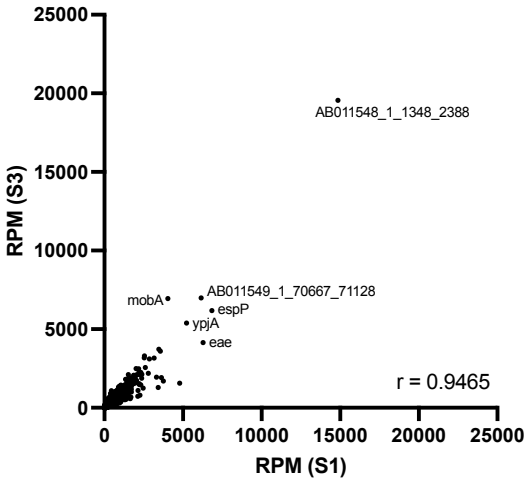


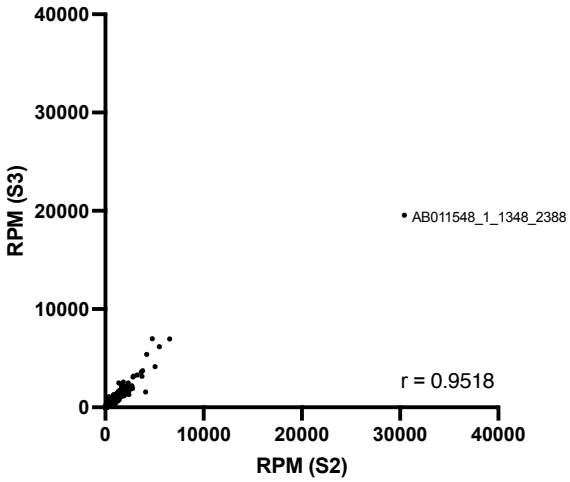
3) 5.5 + Ethanol -> 2.5





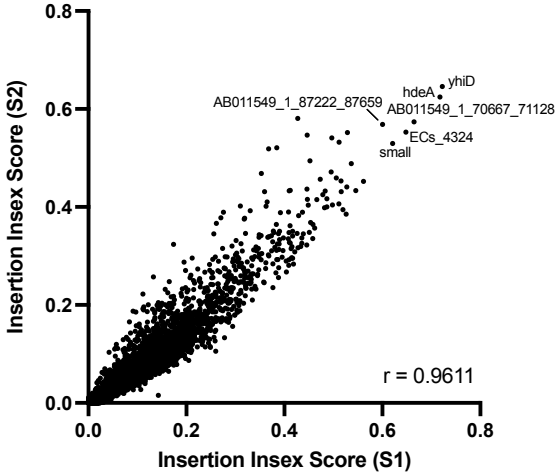
4) 5.5 + Ethanol -> 7

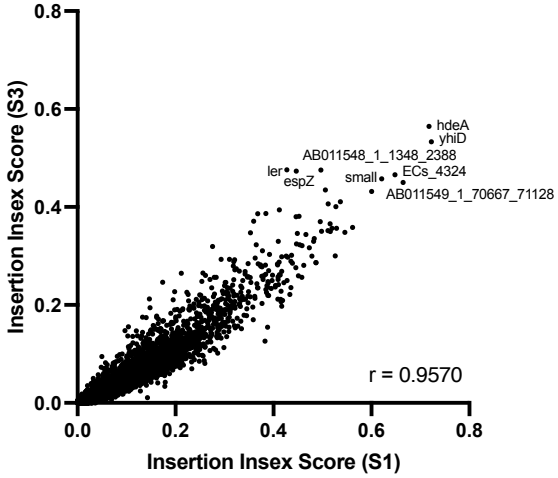




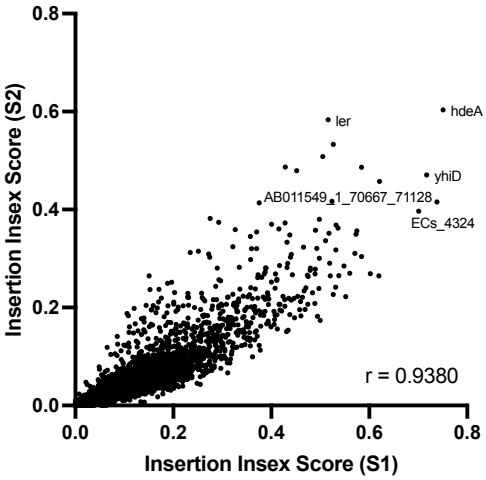
8.3.2.2 Insertion index score

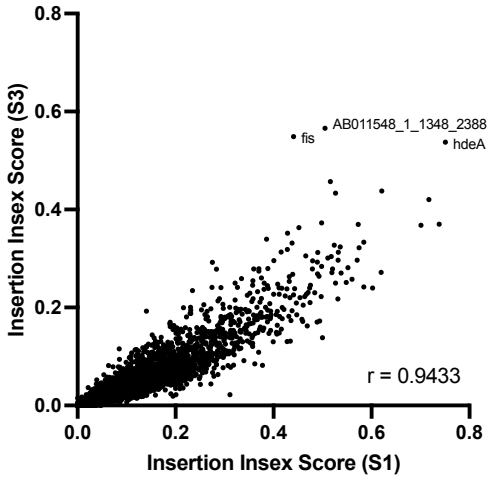
1) 5.5 + Indole -> 2.5



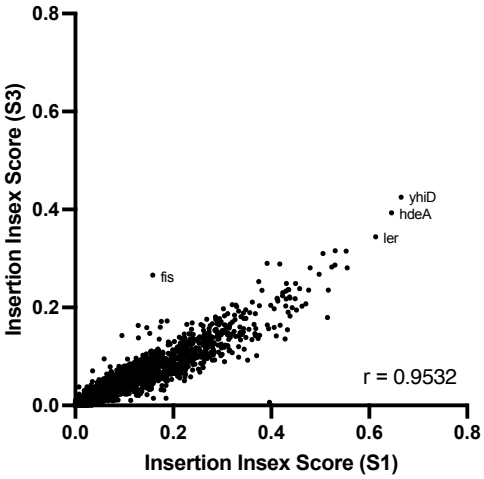


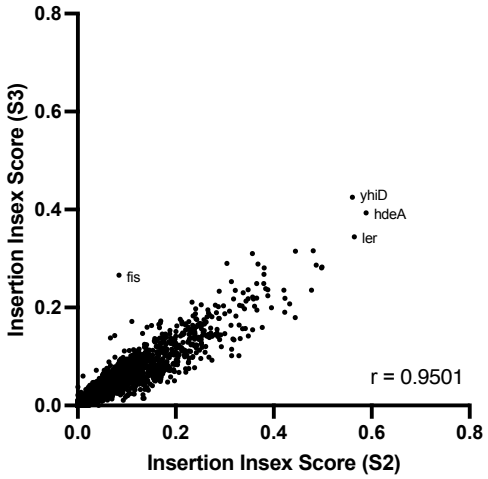
2) 5.5 + Indole -> 7



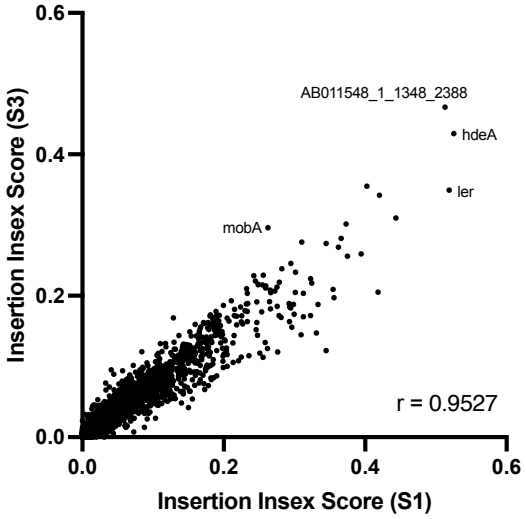


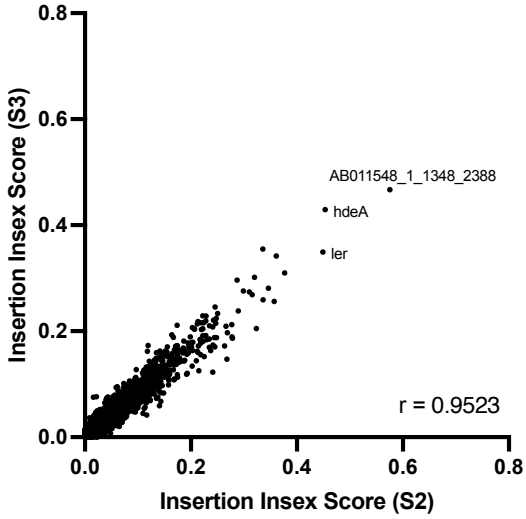
3) 5.5 + Ethanol -> 2.5





4) 5.5 + Ethanol -> 7





8.3.3 Genes with depleted insertions

15 genes exclusively in "5.5 + I -> 2.5"		22 genes exclusively in "5.5 + E -> 2.5"		18 common genes in "5.5 + I -> 2.5" and "5.5 + E -> 2.5"	
Gene	**Function**	**Gene**	**Function**	**Gene**	**Function**
ECs_2958	transposase	*lolA*	lipoprotein chaperone	*ECs_1837*	trp operon leader peptide
ECs_2845	glycosyl transferase	*rpmA*	50S ribosomal subunit protein L27	*waaR*	lipopolysaccharide 1
proQ	ProP effector	*rpsB*	30S ribosomal subunit protein S2	*waaO*	UDP-D-galactose:(glucosyl)lipopolysaccharide-alpha-1
wzy	O-antigen polymerase	*ribE*	riboflavin synthase beta chain	*waaD*	lipopolysaccharide 1
ECs_1187	phage regulatory protein CII	*fabG_1*	3-oxoacyl-[acyl-carrier-protein] reductase	*wcaE*	glycosyl transferase
cpsB	mannose-1-phosphate guanyltransferase	*rplV*	50S ribosomal subunit protein L22	*waaL*	O-antigen ligase
rpmE	50S ribosomal subunit protein L31	*rpsF*	30S ribosomal subunit protein S6	*arnB*	uridine 5'-(beta-1-threo-pentapyranosyl-4-ulose diphosphate) aminotransferase
qseC	quorum sensing sensory histidine kinase in two-component regulatory system with QseB	*priB*	primosomal protein N	*hldD*	ADP-L-glycero-D-mannoheptose-6-epimerase

rseA	anti-sigma factor	*pdxH*	pyridoxine 5'-phosphate oxidase	*ECs_2840*	hypothetical protein
rpsR	30S ribosomal subunit protein S18	*yeeT*	hypothetical protein	*rpoS*	RNA polymerase sigma S factor RpoS
purA	adenylosuccinate synthetase	*nuoA*	NADH-quinone oxidoreductase subunit A	*neuD*	shikimate 5-dehydrogenase
phoP	two-component regulatory system response regulator PhoP	*pdxJ*	pyridoxine 5'-phosphate synthase	*bamB*	BamABCDE complex OM biogenesis lipoprotein
ECs_2180	phage DNA packaging protein	*nhoA*	N-hydroxyarylamine O-acetyltransferase.	*gne*	UDP-galactose 4-epimerase
espR2	T3SS effector-like protein EspR.	*mtn*	5'-methylthioadenosine nucleosidase	*AB011549_1_43157_43546*	
ECs_0292	hypothetical protein	*aceE*	pyruvate dehydrogenase	*pal*	peptidoglycan-associated outer membrane lipoprotein
		ihfB	integration host factor DNA-binding protein beta subunit	*hscB*	co-chaperone HscB
		ECs_1985	phage minor tail protein	*gmd*	GDP-D-mannose dehydratase
		lepB	leader peptidase	*yeeZ*	hypothetical protein
		hscA	chaperone protein HscA		
		wcaG	GDP-L-fucose synthetase		
		AB011549_1_40162_40716			

		nhaA	pH-dependent sodium/proton antiporter		

8.3.4 Genes with enriched insertions

23 genes exclusively in "5.5 + I -> 2.5"		**5 genes exclusively in "5.5 + E -> 2.5"**		**8 common genes in "5.5 + I -> 2.5" and "5.5 + E -> 2.5"**		**0 genes exclusively in "5.5 + I -> 7"**	**0 common genes in "5.5 + I -> 2.5" and "5.5 + I -> 7"**
Gene	**Function**	**Gene**	**Function**	**Gene**	**Function**		
ddlB	D-alanine:D-alanine ligase	*ebgR*	transcriptional repressor	*yeiE*	transcriptional regulator		
yecC	ABC transporter ATPase	*nagC*	N-acetylglucosamine-inducible nag divergent operon transcriptional repressor	*ynfG_2*	oxidoreductase		
narY	nitrate reductase 1 beta (Fe-S) subunit	*lptC*	periplasmic membrane-anchored LPS-binding protein	*yqjA*	inner membrane protein		
ECs_2962	hypothetical protein	*ECs_2236*	phage tail assembly protein	*yfaE*	ferredoxin involved with ribonucleotide reductase		

lsrA	autoinducer 2 import ATP-binding protein	*dnaX*	DNA polymerase III/DNA elongation factor III tau and gamma subunits DnaX	*dnaG*	DNA primase		
cpxA	two-component system sensor histidine kinase CpxA			*mraY*	phospho-N-acetylmuramoyl-pentapeptide-transferase		
tisB	toxic membrane persister formation peptide			*yhaL*	hypothetical protein		
ECs_2165	phage minor tail protein			*waaA*	3-deoxy-D-manno-octulosonic-acid transferase		
coaD	pantetheine-phosphate adenylyltransferase						
fabB	3-oxoacyl-[acyl-carrier-protein] synthase I						
cydX	cytochrome d (bd-I) ubiquinol oxidase subunit X						
xseB	exonuclease VII small subunit						
fabD	malonyl-CoA-[acyl-carrier-protein] transacylase						

ribF	bifunctional riboflavin kinase/FAD synthetase						
tilS	tRNA(Ile)-lysidine synthetase						
murC	UDP-N-acetylmuramate:L-alanine ligase						
accD	acetyl-CoA carboxylase beta subunit						
dnaN	DNA polymerase III beta subunit DnaN						
ECs_2267	phage late gene regulator Q						
rplU	50S ribosomal subunit protein L21						
ECs_1064	hypothetical protein						
secE	preprotein translocase membrane subunit SecE						
ECs_1644	phage minor tail protein						

8.3.5 Comparison of LogFC and q value of genes known to have a role in acid stress, among four conditions

Available at:
https://www.dropbox.com/sh/w7jdc4l2ai3yh9r/AAAZZz0GX5vLuF16hHN2t3Pda?dl=0

8.3.6 List of essential, non-essential and unclear genes in O157:H7 Sakai

Available at:
https://www.dropbox.com/sh/w7jdc4l2ai3yh9r/AAAZZz0GX5vLuF16hHN2t3Pda?dl=0

References

Applebaum, D. M., Dunlap, J. C. and Morris, D. R. (1977) 'Comparison of the biosynthetic and biodegradative ornithine decarboxylases of Escherichia coli', *Biochemistry*. ACS Publications, 16(8), pp. 1580–1584. doi: 10.1021/bi00627a008.

Bak, G. *et al.* (2014) 'Roles of rpoS-activating small RNAs in pathways leading to acid resistance of Escherichia coli', *MicrobiologyOpen*. John Wiley & Sons, Ltd, 3(1), pp. 15–28. doi: https://doi.org/10.1002/mbo3.143.

Banzhaf, M. *et al.* (2020) 'Outer membrane lipoprotein NlpI scaffolds peptidoglycan hydrolases within multi-enzyme complexes in Escherichia coli', *The EMBO Journal*. John Wiley & Sons, Ltd, 39(5), p. e102246. doi: https://doi.org/10.15252/embj.2019102246.

Barquist, L. *et al.* (2016) 'The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries', *Bioinformatics*, 32(7), pp. 1109–1111. doi: 10.1093/bioinformatics/btw022.

Battesti, A., Majdalani, N. and Gottesman, S. (2011) 'The RpoS-Mediated General Stress Response in Escherichia coli', *Annual Review of Microbiology*. Annual Reviews, 65(1), pp. 189–213. doi: 10.1146/annurev-micro-090110-102946.

Bean, R. C., Shepherd, W. C. and Chan, H. (1968) 'Permeability of Lipid Bilayer Membranes to Organic Solutes ', *Journal of General Physiology*, 52(3), pp. 495–508. doi: 10.1085/jgp.52.3.495.

Becker, G., Klauck, E. and Hengge-Aronis, R. (1999) 'Regulation of RpoS proteolysis in Escherichia coli: The response regulator RssB is a recognition factor that interacts with the turnover element in RpoS', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 96(11), pp. 6439–6444. doi: 10.1073/pnas.96.11.6439.

Bekhit, A. *et al.* (2011) 'The Role of OmpC and OmpF in Acidic Resistance in *Escherichia coli*', *Biological and Pharmaceutical Bulletin*, 34(3), pp. 330–334. doi: 10.1248/bpb.34.330.

Bhagwat, A. A. *et al.* (2005) 'Characterization of Enterohemorrhagic Escherichia coli Strains Based on Acid Resistance Phenotypes', *Infection and Immunity*. American Society for Microbiology, 73(8), pp. 4993–5003. doi: 10.1128/IAI.73.8.4993-5003.2005.

Blattner, F. R. *et al.* (1997) 'The complete genome sequence of Escherichia coli K-12', *Science*, 277(5331), pp. 1453–1462. doi: 10.1126/science.277.5331.1453.

Blount, Z. D. (2015) 'The unexhausted potential of E. coli', *eLife*, 4, pp. 1–12. doi: 10.7554/eLife.05826.

Boon, N. *et al.* (2020) 'The Signaling Molecule Indole Inhibits Induction of the AR2 Acid Resistance System in Escherichia coli.', *Frontiers in microbiology*, 11, p. 474. doi: 10.3389/fmicb.2020.00474.

Browning, D. F., Hobman, J. L. and Busby, S. J. W. (2023) 'Laboratory strains of Escherichia coli K-12: things are seldom what they seem', *Microbial Genomics*. Microbiology Society, 9(2). doi: https://doi.org/10.1099/mgen.0.000922.

Burton, N. A *et al.* (2010) 'Novel aspects of the acid response network of E. coli K-12 are revealed by a study of transcriptional dynamics.', *Journal of molecular biology*. Netherlands, 401(5), pp. 726–742. doi: 10.1016/j.jmb.2010.06.054.

Cain, A. K. *et al.* (2020) 'A decade of advances in transposon-insertion sequencing', *Nature Reviews Genetics*, 21(9), pp. 526–540. doi: 10.1038/s41576-020-0244-x.

Cao, H. *et al.* (2017) 'Systems-level understanding of ethanol-induced stresses and adaptation in E. coli', *Scientific Reports*, 7(1), p. 44150. doi: 10.1038/srep44150.

Capitani, G. *et al.* (2003) 'Crystal structure and functional analysis of Escherichia coli glutamate decarboxylase', *The EMBO Journal*. John Wiley & Sons, Ltd, 22(16), pp. 4027–4037. doi: https://doi.org/10.1093/emboj/cdg403.

Carver, T. *et al.* (2012) 'Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data', *Bioinformatics*, 28(4), pp. 464–469. doi: 10.1093/bioinformatics/btr703.

Castanié-Cornet, M. P. *et al.* (2010) 'Acid stress response in Escherichia coli: Mechanism of regulation of gadA transcription by RcsB and GadE', *Nucleic Acids Research*, 38(11), pp. 3546–3554. doi: 10.1093/nar/gkq097.

Chang, Y.-Y. and Cronan, J. E. (1999) 'Membrane cyclopropane fatty acid content is a major factor in acid resistance of Escherichia coli', *Molecular Microbiology*. John Wiley & Sons, Ltd, 33(2), pp. 249–259. doi: https://doi.org/10.1046/j.1365-2958.1999.01456.x.

Chant, E. L. and Summers, D. K. (2007) 'Indole signalling contributes to the stable maintenance of Escherichia coli multicopy plasmids', *Molecular Microbiology*. John Wiley & Sons, Ltd, 63(1), pp. 35–43. doi: https://doi.org/10.1111/j.1365-2958.2006.05481.x.

Charles, D. *et al.* (2015) 'A Rapid and Specific Method for the Detection of Indole in Complex Biological Samples', *Applied and Environmental Microbiology*. American Society for Microbiology, 81(23), pp. 8093–8097. doi: 10.1128/AEM.02787-15.

Chattopadhyay, M. K. *et al.* (2015) 'Polyamines Stimulate the Level of the σ38 Subunit (RpoS) of Escherichia coli RNA Polymerase, Resulting in the Induction of the Glutamate Decarboxylase-dependent Acid Response System via the gadE Regulon', *Journal of Biological Chemistry*. Elsevier, 290(29), pp. 17809–17821. doi: 10.1074/jbc.M115.655688.

Choi, S. H., Baumler, D. J. and Kaspar, C. W. (2000) 'Contribution of dps to acid stress tolerance and oxidative stress tolerance in Escherichia coli O157: H7', *Applied and Environmental Microbiology*. Am Soc Microbiol, 66(9), pp. 3911–3916. doi: 10.1128/AEM.66.9.3911-3916.2000.

Coldewey, S. M. *et al.* (2007) 'Impact of the rpoS genotype for acid resistance patterns of pathogenic and probiotic Escherichia coli', *BMC Microbiology*, 7(1), p. 21. doi: 10.1186/1471-2180-7-21.

Connor, T. R. *et al.* (2016) 'CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community', *Microbial Genomics*. Microbiology Society, 2(9), p. e000086. doi: 10.1099/mgen.0.000086.

Dahl, J.-U. *et al.* (2015) 'HdeB Functions as an Acid-protective Chaperone in Bacteria', *Journal of Biological Chemistry*. Elsevier, 290(1), pp. 65–75. doi: 10.1074/jbc.M114.612986.

Datsenko, K. A. and Wanner, B. L. (2000) 'One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products', *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp. 6640–6645. doi: 10.1073/pnas.120163297.

Deamer, D. W. (1987) 'Proton permeation of lipid bilayers', *Journal of Bioenergetics and Biomembranes*, 19(5), pp. 457–479. doi: 10.1007/BF00770030.

DeAngelis, M. M., Wang, D. G. and Hawkins, T. L. (1995) 'Solid-phase reversible immobilization for the isolation of PCR products', *Nucleic Acids Research*, 23(22), pp. 4742–4743. doi: 10.1093/nar/23.22.4742.

Deatherage, D. E. and Barrick, J. E. (2014) 'Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq BT - Engineering and Analyzing Multicellular Systems: Methods and Protocols', in Sun, L. and Shou, W. (eds). New York, NY: Springer New York, pp. 165–188. doi: 10.1007/978-1-4939-0554-6_12.

De Biase, D. *et al.* (1999) 'The response to stationary-phase stress conditions in Escherichia coli: Role and regulation of the glutamic acid decarboxylase system', *Molecular Microbiology*, 32(6), pp. 1198–1211. doi: 10.1046/J.1365-2958.1999.01430.X.

De Biase, D. and Pennacchietti, E. (2012) 'Glutamate decarboxylase-dependent acid resistance in orally acquired bacteria: function, distribution and biomedical implications of the gadBC operon', *Molecular Microbiology*. John Wiley & Sons, Ltd, 86(4), pp. 770–786. doi: https://doi.org/10.1111/mmi.12020.

Deininger, K. N. W. *et al.* (2011) 'A Requirement of TolC and MDR Efflux Pumps for Acid Adaptation and GadAB Induction in Escherichia coli', *PLoS ONE*. Edited by A. Driessen. Public Library of Science, 6(4), p. e18960. doi: 10.1371/journal.pone.0018960.

del Alamo, D. *et al.* (2022) 'Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 119(34), p. e2206129119. doi: 10.1073/pnas.2206129119.

Dell, C. L., Neely, M. N. and Olson, E. R. (1994) 'Altered pH lysine signalling mutants of cadC, a gene encoding a membrane-bound transcriptional activator of the Escherichia coli cadBA operon', *Molecular microbiology*. Wiley Online Library, 14(1), pp. 7–16. doi: 10.1111/j.1365-2958.1994.tb01262.x.

Diez-Gonzalez, F. and Karaibrahimoglu, Y. (2004) 'Comparison of the glutamate-, arginine- and lysine-dependent acid resistance systems in Escherichia coli O157:H7', *Journal of Applied Microbiology*. John Wiley & Sons, Ltd, 96(6), pp. 1237–1244. doi: https://doi.org/10.1111/j.1365-2672.2004.02251.x.

Dong, T., Kirchhof, M. G. and Schellhorn, H. E. (2008) 'RpoS regulation of gene expression during exponential growth of Escherichia coli K12', *Molecular Genetics and Genomics*, 279(3), pp. 267–277. doi: 10.1007/s00438-007-0311-4.

Douglass, M. V, Cléon, F. and Trent, M. S. (2021) 'Cardiolipin aids in lipopolysaccharide transport to the gram-negative outer membrane', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 118(15), p. e2018329118. doi: 10.1073/pnas.2018329118.

Dupré, E. *et al.* (2015) 'Signal transduction by BvgS sensor kinase. BINDING OF MODULATOR NICOTINATE AFFECTS THE CONFORMATION AND DYNAMICS OF THE ENTIRE PERIPLASMIC MOIETY.', *Journal of Biological Chemistry*. Elsevier, 290(44), p. 26473. doi: 10.1074/jbc.A115.655720.

Eguchi, Y. *et al.* (2003) 'Transcriptional regulation of drug efflux genes by EvgAS, a two-component system in Escherichia coli', *Microbiology*. Microbiology Society, 149(10), pp. 2819–2828. doi: 10.1099/mic.0.26460-0.

Eguchi, Y. *et al.* (2007) 'B1500, a small membrane protein, connects the two-component systems EvgS/EvgA and PhoQ/PhoP in Escherichia coli', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 104(47), pp. 18712–18717. doi: 10.1073/pnas.0705768104.

Eguchi, Y. *et al.* (2011) 'Regulation of Acid Resistance by Connectors of Two-Component Signal Transduction Systems in Escherichia coli ', *Journal of Bacteriology*. American Society for Microbiology, 193(5), pp. 1222–1228. doi: 10.1128/JB.01124-10.

Eguchi, Y. *et al.* (2012) 'The connector SafA interacts with the multi-sensing domain of PhoQ in Escherichia coli', *Molecular microbiology*. Wiley Online Library, 85(2), pp. 299–313. doi: 10.1128/JB.01742-14.

Eguchi, Y. and Utsumi, R. (2014) 'Alkali metals in addition to acidic pH activate the EvgS histidine kinase sensor in Escherichia coli', *Journal of bacteriology*. 2014/06/23. American Society for Microbiology, 196(17), pp. 3140–3149. doi: 10.1128/JB.01742-14.

Escherich, T. (1988) 'The Intestinal Bacteria of the Neonate and Breast-Fed Infant', *Reviews of Infectious Diseases*, 10(6), pp. 1220–1225. doi: 10.1093/clinids/10.6.1220.

Fletcher, A. (2021) 'Exploration of the Induction Conditions of Escherichia coli EvgS in the Presence of Indole.' Unpublished MSci thesis. University of Birmingham.

Foster, J. W. (2004) 'Escherichia coli acid resistance: tales of an amateur acidophile', *Nature Reviews Microbiology*, 2(11), pp. 898–907. doi: 10.1038/nrmicro1021.

Frenzen, P. D., Drake, A. and Angulo, F. J. (2005) 'Economic Cost of Illness Due to Escherichia coli O157 Infections in the United States', *Journal of Food Protection*, 68(12), pp. 2623–2630. doi: https://doi.org/10.4315/0362-028X-68.12.2623.

Fu, X. *et al.* (2018) 'DegP functions as a critical protease for bacterial acid resistance', *The FEBS Journal*. John Wiley & Sons, Ltd, 285(18), pp. 3525–3538. doi: https://doi.org/10.1111/febs.14627.

Gaimster, H. *et al.* (2014) 'The indole pulse: A new perspective on indole signalling in Escherichia coli', *PLoS ONE*, 9(4). doi: 10.1371/journal.pone.0093168.

Gale, E. F. (1940) 'The production of amines by bacteria: The decarboxylation of amino-acids by strains of Bacterium coli', *Biochemical Journal*, 34(3), pp. 392–413. doi: 10.1042/bj0340392.

Gawronski, J. D. *et al.* (2009) 'Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 106(38), pp. 16422–16427. doi: 10.1073/pnas.0906627106.

Giangrossi, M. *et al.* (2005) 'Antagonistic role of H-NS and GadX in the regulation of the glutamate decarboxylase-dependent acid resistance system in Escherichia coli', *Journal of Biological Chemistry*. ASBMB, 280(22), pp. 21498–21505. doi: 10.1074/jbc.M413255200.

Gong, S., Richard, H. and Foster, J. W. (2003) 'YjdE (AdiC) is the arginine:Agmatine antiporter essential for arginine-dependent acid resistance in Escherichia coli', *Journal of Bacteriology*, 185(15), pp. 4402–4409. doi: 10.1128/JB.185.15.4402-4409.2003.

Goodall, E. C. A. *et al.* (2018) 'The Essential Genome of Escherichia coli K-12', *mBio*. American Society for Microbiology, 9(1), pp. e02096-17. doi: 10.1128/mBio.02096-17.

Goodman, A. L. *et al.* (2009) 'Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat', *Cell Host & Microbe*. Elsevier, 6(3), pp. 279–289. doi: 10.1016/j.chom.2009.08.003.

Grenier, F. *et al.* (2014) 'Complete Genome Sequence of Escherichia coli BW25113', *Genome announcements*. American Society for Microbiology, 2(5), pp. e01038-14. doi: 10.1128/genomeA.01038-14.

Groisman, E. A. *et al.* (2013) 'Bacterial Mg2+ Homeostasis, Transport, and Virulence', *Annual Review of Genetics*. Annual Reviews, 47(1), pp. 625–646. doi: 10.1146/annurev-genet-051313-051025.

Guan, N. and Liu, L. (2020) 'Microbial response to acid stress: mechanisms and applications', *Applied Microbiology and Biotechnology*, 104(1), pp. 51–65. doi: 10.1007/s00253-019-10226-1.

Gu, D. *et al.* (2021) 'Genome-wide identification of genes involved in acid stress resistance of salmonella derby', *Genes*, 12(4). doi: 10.3390/genes12040476.

Gut, H. *et al.* (2006) 'Escherichia coli acid resistance: pH-sensing, activation by chloride and autoinhibition in GadB', *The EMBO Journal*. John Wiley & Sons, Ltd, 25(11), pp. 2643–2651. doi: https://doi.org/10.1038/sj.emboj.7601107.

Guzman, L. M. *et al.* (1995) 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter', *Journal of Bacteriology*, 177(14), pp. 4121–4130. doi: 10.1128/jb.177.14.4121-4130.1995.

Hayashi, T. *et al.* (2001) 'Complete genome sequence of enterohemorrhagic Eschelichia coli O157:H7 and genomic comparison with a laboratory strain K-12', *DNA Research*, 8(1), pp. 11–22. doi: 10.1093/dnares/8.1.11.

Hayes, E. T. *et al.* (2006) 'Oxygen limitation modulates pH regulation of catabolism and hydrogenases, multidrug transporters, and envelope composition in Escherichia coli K-12', *BMC Microbiology*, 6(1), p. 89. doi: 10.1186/1471-2180-6-89.

He, D. *et al.* (2019) 'Protease-Mediated Protein Quality Control for Bacterial Acid Resistance', *Cell Chemical Biology*. Elsevier, 26(1), pp. 144-150.e3. doi: 10.1016/j.chembiol.2018.10.014.

Herrou, J. *et al.* (2010) 'Periplasmic domain of the sensor-kinase BvgS reveals a new paradigm for the Venus flytrap mechanism', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 107(40), pp. 17351–17355. doi: 10.1073/pnas.1006267107.

Heuveling, J., Possling, A. and Hengge, R. (2008) 'A role for Lon protease in the control of the acid resistance genes of Escherichia coli', *Molecular Microbiology*. John Wiley & Sons, Ltd, 69(2), pp. 534–547. doi: https://doi.org/10.1111/j.1365-2958.2008.06306.x.

Heyde, M. and Portalier, R. (1990) 'Acid shock proteins of Escherichia coli', *FEMS Microbiology Letters*, 69(1–2), pp. 19–26. doi: 10.1111/j.1574-6968.1990.tb04168.x.

Hirakawa, H. *et al.* (2010) 'Indole enhances acid resistance in Escherichia coli', *Microbial Pathogenesis*. Elsevier Ltd, 49(3), pp. 90–94. doi: 10.1016/j.micpath.2010.05.002.

Hobman, J. L., Penn, C. W. and Pallen, M. J. (2007) 'Laboratory strains of Escherichia coli: model citizens or deceitful delinquents growing old disgracefully?', *Molecular Microbiology*. John Wiley & Sons, Ltd, 64(4), pp. 881–885. doi: https://doi.org/10.1111/j.1365-2958.2007.05710.x.

Houry, W. A. (2001) 'Chaperone-assisted protein folding in the cell cytoplasm', *Current Protein and Peptide Science*. Bentham Science Publishers, 2(3), pp. 227–244. doi: 10.2174/1389203013381134.

Hrishiraj, S. *et al.* (2017) 'Structural and Functional Analysis of the Escherichia coli Acid-Sensing Histidine Kinase EvgS', *Journal of Bacteriology*. American Society for Microbiology, 199(18), pp. e00310-17. doi: 10.1128/JB.00310-17.

Hurst, M. N. *et al.* (2023) 'The QseB response regulator imparts tolerance to positively charged antibiotics by controlling metabolism and minor changes to LPS', *bioRxiv*, p. 2023.01.10.523522. doi: 10.1101/2023.01.10.523522.

Inada, S. *et al.* (2021) 'Acid-Sensing Histidine Kinase With a Redox Switch', *Frontiers in Microbiology*. Frontiers Media S.A., 12. doi: 10.3389/fmicb.2021.652546.

Itou, J., Eguchi, Y. and Utsumi, R. (2009) 'Molecular Mechanism of Transcriptional Cascade Initiated by the EvgS/EvgA System in Escherichia coli K-12', *Bioscience, Biotechnology, and Biochemistry*, 73(4), pp. 870–878. doi: 10.1271/bbb.80795.

Iyer, R. *et al.* (2000) 'Molecular basis for the polyamine-OmpF porin interactions: inhibitor and mutant studies11Edited by G. von Heijne', *Journal of Molecular Biology*, 297(4), pp. 933–945. doi: https://doi.org/10.1006/jmbi.2000.3599.

Iyer, R. *et al.* (2002) 'A biological role for prokaryotic ClC chloride channels', *Nature*, 419(6908), pp. 715–718. doi: 10.1038/nature01000.

Jin, Y. *et al.* (2009) 'Small noncoding RNA GcvB is a novel regulator of acid resistance in Escherichia coli', *BMC Genomics*, 10(1), p. 165. doi: 10.1186/1471-2164-10-165.

Johnson, M. D. *et al.* (2011) 'RcsB is required for inducible acid resistance in Escherichia coli and acts at gadE-dependent and -independent promoters', *Journal of bacteriology*. 2011/05/13. United States, 193(14), pp. 3653–3656. doi: 10.1128/JB.05040-11.

Johnson, M. D. *et al.* (2014) 'Characterization of mutations in the PAS domain of the EvgS sensor kinase selected by laboratory evolution for acid resistance in Escherichia coli', *Molecular Microbiology*, 93(5), pp. 911–927. doi: 10.1111/mmi.12704.

Jolivet, P. and Foley, J. W. (2015) 'Solutions for purifying nucleic acids by solid- phase reversible immobilization (SPRI)', *Ludmer Centre for Neuroinformatics and Mental Health*, 5(3), pp. 248–253.

Kanda, T. *et al.* (2020) 'RNase E-dependent degradation of tnaA mRNA encoding tryptophanase is prerequisite for the induction of acid resistance in Escherichia coli', *Scientific Reports*. Nature Research, 10(1). doi: 10.1038/s41598-020-63981-x.

Kanjee, U. and Houry, W. A. (2013) 'Mechanisms of acid resistance in escherichia coli', *Annual Review of Microbiology*, 67, pp. 65–81. doi: 10.1146/ANNUREV-MICRO-092412-155708.

Kashiwagi, K. *et al.* (1992) 'Excretion of putrescine by the putrescine-ornithine antiporter encoded by the potE gene of Escherichia coli.', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 89(10), pp. 4529–4533. doi: 10.1073/pnas.89.10.4529.

Kato, A. *et al.* (2000) 'Transcription of *emrKY* is Regulated by the EvgA-EvgS Two-Component System in *Escherichia coli* K-12', *Bioscience, Biotechnology, and Biochemistry*, 64(6), pp. 1203–1209. doi: 10.1271/bbb.64.1203.

Kawamura-Sato, K. *et al.* (1999) 'Role of multiple efflux pumps in Escherichia coli in indole expulsion', *FEMS Microbiology Letters*, 179(2), pp. 345–352. doi: 10.1111/j.1574-6968.1999.tb08748.x.

Khan, S. *et al.* (1995) 'Chemotactic signal integration in bacteria.', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 92(21), pp. 9757–9761. doi: 10.1073/pnas.92.21.9757.

Kobayashi, A. *et al.* (2006) 'Growth phase-dependent expression of drug exporters in Escherichia coli and its contribution to drug tolerance', *Journal of bacteriology*. Am Soc Microbiol, 188(16), pp. 5693–5703. doi: 10.1128/JB.00217-06.

Krikos, A. *et al.* (1985) 'Chimeric chemosensory transducers of Escherichia coli.', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 82(5), pp. 1326–1330. doi: 10.1073/pnas.82.5.1326.

Lai, Y.-C. *et al.* (2003) 'Identification and characterization of KvgAS, a two-component system in Klebsiella pneumoniae CG43', *FEMS microbiology letters*. Blackwell Publishing Ltd Oxford, UK, 218(1), pp. 121–126. doi: 10.1111/j.1574-6968.2003.tb11507.x.

Langridge, G. C. *et al.* (2009) 'Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants', *Genome Research*. Cold Spring Harbor Laboratory, 19(12), pp. 2308–2316. doi: 10.1101/gr.097097.109.

Large, T. M., Walk, S. T. and Whittam, T. S. (2005) 'Variation in Acid Resistance among Shiga Toxin-Producing Clones of Pathogenic Escherichia coli', *Applied and Environmental Microbiology*. American Society for Microbiology, 71(5), pp. 2493–2500. doi: 10.1128/AEM.71.5.2493-2500.2005.

Lee, J. *et al.* (2007) 'Structure and Function of the Escherichia coli Protein YmgB: A Protein Critical for Biofilm Formation and Acid-resistance', *Journal of Molecular Biology*, 373(1), pp. 11–26. doi: https://doi.org/10.1016/j.jmb.2007.07.037.

Lee, J.-H. and Lee, J. (2010) 'Indole as an intercellular signal in microbial communities', *FEMS Microbiology Reviews*, 34(4), pp. 426–444. doi: 10.1111/j.1574-6976.2009.00204.x.

Lee, J.-H., Wood, T. K. and Lee, J. (2015) 'Roles of Indole as an Interspecies and Interkingdom Signaling Molecule', *Trends in Microbiology*, 23(11), pp. 707–718. doi: https://doi.org/10.1016/j.tim.2015.08.001.

Li, G. and Young, K. D. (2013) 'Indole production by the tryptophanase TnaA in Escherichia coli is determined by the amount of exogenous tryptophan', *Microbiology*. Microbiology Society, 159(Pt_2), pp. 402–410. doi: 10.1099/mic.0.064139-0.

Lindahl, T. and Nyberg, B. (1972) 'Rate of depurination of native deoxyribonucleic acid', *Biochemistry*. American Chemical Society, 11(19), pp. 3610–3618. doi: 10.1021/bi00769a018.

Lin, J. *et al.* (1995) 'Comparative analysis of extreme acid survival in Salmonella typhimurium, Shigella flexneri, and Escherichia coli', *Journal of Bacteriology*. American Society for Microbiology, 177(14), pp. 4097–4104. doi: 10.1128/jb.177.14.4097-4104.1995.

Lin, J. *et al.* (1996) 'Mechanisms of acid resistance in enterohemorrhagic Escherichia coli', *Applied and environmental microbiology*, 62(9), p. 3094—3100. doi: 10.1128/aem.62.9.3094-3100.1996.

Li, Z. *et al.* (2020) 'The role of bacterial cell envelope structures in acid stress resistance in E. coli', *Applied Microbiology and Biotechnology*. Springer, pp. 2911–2921. doi: 10.1007/s00253-020-10453-x.

Lomovskaya, O., Lewis, K. and Matin, A. (1995) 'EmrR is a negative regulator of the Escherichia coli multidrug resistance pump EmrAB', *Journal of Bacteriology*. American Society for Microbiology, 177(9), pp. 2328–2334. doi: 10.1128/jb.177.9.2328-2334.1995.

Lund, P., Tramonti, A. and De Biase, D. (2014) 'Coping with low pH: Molecular strategies in neutralophilic bacteria', *FEMS Microbiology Reviews*, 38(6), pp. 1091–1125. doi: 10.1111/1574-6976.12076.

Lu, P. *et al.* (2013) 'L-glutamine provides acid resistance for Escherichia coli through enzymatic release of ammonia', *Cell Research*, 23(5), pp. 635–644. doi: 10.1038/CR.2013.13.

Ma, D. *et al.* (2012) 'Structure and mechanism of a glutamate-GABA antiporter.', *Nature*. England, 483(7391), pp. 632–636. doi: 10.1038/nature10917.

Ma, D., Lu, P. and Shi, Y. (2013) 'Substrate Selectivity of the Acid-activated Glutamate/ γ-Aminobutyric acid (GABA) Antiporter GadC from Escherichia coli', *Journal of Biological Chemistry*. Elsevier, 288(21), pp. 15148–15153. doi: 10.1074/jbc.M113.474502.

Majdalani, N., Hernandez, D. and Gottesman, S. (2002) 'Regulation and mode of action of the second small RNA activator of RpoS translation, RprA', *Molecular Microbiology*. John Wiley & Sons, Ltd, 46(3), pp. 813–826. doi: https://doi.org/10.1046/j.1365-2958.2002.03203.x.

Marie-Pierre, C.-C. *et al.* (1999) 'Control of Acid Resistance inEscherichia coli', *Journal of Bacteriology*. American Society for Microbiology, 181(11), pp. 3525–3535. doi: 10.1128/JB.181.11.3525-3535.1999.

Masuda, N. and Church, G. M. (2002) 'Escherichia coli Gene Expression Responsive to Levels of the Response Regulator EvgA', *Journal of Bacteriology*. American Society for Microbiology, 184(22), pp. 6225–6234. doi: 10.1128/JB.184.22.6225-6234.2002.

Masuda, N. and Church, G. M. (2003) 'Regulatory network of acid resistance genes in Escherichia coli', *Molecular Microbiology*. John Wiley & Sons, Ltd, 48(3), pp. 699–712. doi: https://doi.org/10.1046/j.1365-2958.2003.03477.x.

Maurer, L. M. *et al.* (2005) 'pH Regulates Genes for Flagellar Motility, Catabolism, and Oxidative Stress in Escherichia coli K-12', *Journal of Bacteriology*. American Society for Microbiology, 187(1), pp. 304–319. doi: 10.1128/JB.187.1.304-319.2005.

Ma, Z. *et al.* (2003) 'GadE (YhiE) activates glutamate decarboxylase-dependent acid resistance in Escherichia coli K-12', *Molecular Microbiology*. John Wiley & Sons, Ltd, 49(5), pp. 1309–1320. doi: https://doi.org/10.1046/j.1365-2958.2003.03633.x.

Ma, Z., Masuda, N. and Foster, J. W. (2004) 'Characterization of EvgAS-YdeO-GadE Branched Regulatory Circuit Governing Glutamate-Dependent Acid Resistance in Escherichia coli', *Journal of Bacteriology*. American Society for Microbiology, 186(21), pp. 7378–7389. doi: 10.1128/JB.186.21.7378-7389.2004.

Milner, M. T. (2021) 'Can Transposon Directed Insertion-Site Sequencing Be Used To Predict Possible Outcomes Of Evolution?' PhD thesis. University of Birmingham. Available at: https://etheses.bham.ac.uk.

Monteiro, C. *et al.* (2009) 'Characterization of cellulose production in Escherichia coli Nissle 1917 and its biological consequences', *Environmental Microbiology*, 11(5), pp. 1105–1116. doi: 10.1111/j.1462-2920.2008.01840.x.

Mootha, V. K. *et al.* (2003) 'PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes', *Nature Genetics*, 34(3), pp. 267–273. doi: 10.1038/ng1180.

Mueller, E. A. *et al.* (2019) 'Plasticity of Escherichia coli cell wall metabolism promotes fitness and antibiotic resistance across environmental conditions', *eLife*, 8. doi: 10.7554/eLife.40754.

Mujacic, M. and Baneyx, F. (2007) 'Chaperone Hsp31 contributes to acid resistance in stationary-phase Escherichia coli', *Applied and environmental microbiology*. Am Soc Microbiol, 73(3), pp. 1014–1018. doi: 10.1128/AEM.02429-06.

Mychack, A. *et al.* (2019) 'A synergistic role for two predicted inner membrane proteins of Escherichia coli in cell envelope integrity', *Molecular Microbiology*. John Wiley & Sons, Ltd, 111(2), pp. 317–337. doi: https://doi.org/10.1111/mmi.14157.

Nishino, K., Senda, Y. and Yamaguchi, A. (2008) 'The AraC-family regulator GadX enhances multidrug resistance in Escherichia coli by activating expression of mdtEF multidrug efflux genes', *Journal of Infection and Chemotherapy*. Elsevier, 14(1), pp. 23–29.

Noguchi, K. *et al.* (2010) 'Hydrogenase-3 Contributes to Anaerobic Acid Resistance of Escherichia coli', *PLOS ONE*. Public Library of Science, 5(4), p. e10132.

Oliveros, J. C. (2007-2015) *Venny. An interactive tool for comparing lists with Venn's diagrams.* Available at: https://bioinfogp.cnb.csic.es/tools/venny/index.html.

Opdyke, J. A. *et al.* (2011) 'RNase III Participates in GadY-Dependent Cleavage of the gadX-gadW mRNA', *Journal of Molecular Biology*, 406(1), pp. 29–43. doi: https://doi.org/10.1016/j.jmb.2010.12.009.

Page, A. J. *et al.* (2015) 'Roary: rapid large-scale prokaryote pan genome analysis', *Bioinformatics*, 31(22), pp. 3691–3693. doi: 10.1093/bioinformatics/btv421.

Paulsen, I. T., Brown, M. H. and Skurray, R. A. (1996) 'Proton-dependent multidrug efflux systems', *Microbiological Reviews*. American Society for Microbiology, 60(4), pp. 575–608. doi: 10.1128/mr.60.4.575-608.1996.

Perry, N. (2022) 'Construction of a TraDIS Xpress Library for Escherichia coli O157:H7 Sakai.' Unpublished MSci thesis. University of Birmingham.

Piñero-Fernandez, S. *et al.* (2011) 'Indole Transport across Escherichia coli Membranes', *Journal of Bacteriology*. American Society for Microbiology, 193(8), pp. 1793–1798. doi: 10.1128/JB.01477-10.

Rangel, J. M. *et al.* (2005) 'Epidemiology of Escherichia coli O157:H7 Outbreaks, United States, 1982–2002', *Emerging Infectious Disease journal*, 11(4), p. 603. doi: 10.3201/eid1104.040739.

Reister, M. *et al.* (2014) 'Complete genome sequence of the Gram-negative probiotic Escherichia coli strain Nissle 1917', *Journal of Biotechnology*, 187, pp. 106–107. doi: https://doi.org/10.1016/j.jbiotec.2014.07.442.

Repaske, D. R. and Adler, J. (1981) 'Change in intracellular pH of Escherichia coli mediates the chemotactic response to certain attractants and repellents', *Journal of bacteriology*, 145(3), pp. 1196–1208. doi: 10.1128/jb.145.3.1196-1208.1981.

Richard, H. and Foster, J. W. (2004) 'Escherichia coli glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential.', *Journal of bacteriology*, 186(18), pp. 6032–6041. doi: 10.1128/JB.186.18.6032-6041.2004.

Riley, M. *et al.* (2006) 'Escherichia coli K-12: A cooperatively developed annotation snapshot - 2005', *Nucleic Acids Research*, 34(1), pp. 1–9. doi: 10.1093/nar/gkj405.

Robins-Browne, R. M. *et al.* (2016) 'Are Escherichia coli Pathotypes Still Relevant in the Era of Whole-Genome Sequencing? ', *Frontiers in Cellular and Infection Microbiology* .

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, 26(1), pp. 139–140. doi: 10.1093/bioinformatics/btp616.

Roggiani, M., Yadavalli, S. S. and Goulian, M. (2017) 'Natural variation of a sensor kinase controlling a conserved stress response pathway in Escherichia coli', *PLOS Genetics*. Edited by P. H. Viollier. Public Library of Science, 13(11), p. e1007101. doi: 10.1371/journal.pgen.1007101.

Ruiz, N. and Silhavy, T. J. (2022) 'How Escherichia coli Became the Flagship Bacterium of Molecular Biology', *Journal of Bacteriology*. American Society for Microbiology, 204(9), pp. e00230-22. doi: 10.1128/jb.00230-22.

Salmond, C. V, Kroll, R. G. and Booth, I. R. (1984) 'The Effect of Food Preservatives on pH Homeostasis in Escherichia coli', *Microbiology*. Microbiology Society, 130(11), pp. 2845–2850. doi: https://doi.org/10.1099/00221287-130-11-2845.

Samartzidou, H. and Delcour, A. H. (1999) 'Distinct sensitivities of OmpF and PhoE porins to charged modulators', *FEBS Letters*. John Wiley & Sons, Ltd, 444(1), pp. 65–70. doi: https://doi.org/10.1016/S0014-5793(99)00030-7.

Sauvage, E. *et al.* (2008) 'The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis', *FEMS Microbiology Reviews*, 32(2), pp. 234–258. doi: 10.1111/j.1574-6976.2008.00105.x.

Sayed, A. K. and Foster, J. W. (2009) 'A 750 bp sensory integration region directs global control of the Escherichia coli GadE acid resistance regulator', *Molecular Microbiology*. John Wiley & Sons, Ltd, 71(6), pp. 1435–1450. doi: https://doi.org/10.1111/j.1365-2958.2009.06614.x.

Sayed, A. K., Odom, C. and Foster, J. W. (2007) 'The Escherichia coli AraC-family regulators GadX and GadW activate gadE, the central activator of glutamate-dependent acid resistance', *Microbiology*. Microbiology Society, 153(8), pp. 2584–2592. doi: https://doi.org/10.1099/mic.0.2007/007005-0.

Schellhorn, H. E. (2020) 'Function, Evolution, and Composition of the RpoS Regulon in Escherichia coli', *Frontiers in Microbiology*, *11*, p.560099. doi: 10.3389/fmicb.2020.560099.

Schwarz, J. *et al.* (2022) 'Bacterial battle against acidity', *FEMS Microbiology Reviews*, 46(6), p. fuac037. doi: 10.1093/femsre/fuac037.

Schweder, T. *et al.* (1996) 'Regulation of Escherichia coli starvation sigma factor (sigma s) by ClpXP protease', *Journal of Bacteriology*. American Society for Microbiology, 178(2), pp. 470–476. doi: 10.1128/jb.178.2.470-476.1996.

Segura, A. *et al.* (2017) 'Draft genome sequence and characterization of commensal Escherichia coli strain BG1 isolated from bovine gastro-intestinal tract', *Standards in Genomic Sciences*, 12(1), p. 61. doi: 10.1186/s40793-017-0272-0.

Seo, S. W. *et al.* (2015) 'Decoding genome-wide GadEWX-transcriptional regulatory networks reveals multifaceted cellular responses to acid stress in Escherichia coli', *Nature Communications*, 6(1), p. 7970. doi: 10.1038/ncomms8970.

Šeputienė, V. *et al.* (2006) 'Acid response of exponentially growing Escherichia coli K-12', *Microbiological Research*, 161(1), pp. 65–74. doi: https://doi.org/10.1016/j.micres.2005.06.002.

Slonczewski, J. L. *et al.* (2009) 'Cytoplasmic pH Measurement and Homeostasis in Bacteria and Archaea', in Poole, R. K. B. T.-A. in M. P. (ed.). Academic Press, pp. 1–317. doi: https://doi.org/10.1016/S0065-2911(09)05501-5.

Smith, D. K. *et al.* (1992) 'Escherichia coli has two homologous glutamate decarboxylase genes that map to distinct loci', *Journal of Bacteriology*. American Society for Microbiology, 174(18), pp. 5820–5826. doi: 10.1128/jb.174.18.5820-5826.1992.

Sonnenborn, U. (2016) 'Escherichia coli strain Nissle 1917-from bench to bedside and back: History of a special Escherichia coli strain with probiotic properties', *FEMS Microbiology Letters*, 363(19), pp. 1–6. doi: 10.1093/femsle/fnw212.

Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 102(43), pp. 15545–15550. doi: 10.1073/pnas.0506580102.

Sun, Y. *et al.* (2011) 'ATP Requirement for Acidic Resistance in Escherichia coli ', *Journal of Bacteriology*. American Society for Microbiology, 193(12), pp. 3072–3077. doi: 10.1128/jb.00091-11.

Sun, Y. *et al.* (2012a) 'Adenosine deamination increases the survival under acidic conditions in Escherichia coli', *Journal of Applied Microbiology*. John Wiley & Sons, Ltd, 112(4), pp. 775–781. doi: https://doi.org/10.1111/j.1365-2672.2012.05246.x.

Sun, Yirong *et al.* (2012b) 'Respiration and the F1Fo-ATPase Enhance Survival under Acidic Conditions in Escherichia coli', *PLOS ONE*. Public Library of Science, 7(12), p. e52577.

Surmann, K. *et al.* (2016) 'Molecular and proteome analyses highlight the importance of the Cpx envelope stress system for acid stress and cell wall stability in Escherichia coli', *MicrobiologyOpen*. John Wiley & Sons, Ltd, 5(4), pp. 582–596. doi: https://doi.org/10.1002/mbo3.353.

Tan, B. K. *et al.* (2012) 'Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 109(41), pp. 16504–16509. doi: 10.1073/pnas.1212797109.

Tantoso, E. *et al.* (2022) 'To kill or to be killed: pangenome analysis of Escherichia coli strains reveals a tailocin specific for pandemic ST131', *BMC Biology*, 20(1), p. 146. doi: 10.1186/s12915-022-01347-7.

Tan, Z. *et al.* (2017) 'Engineering Escherichia coli membrane phospholipid head distribution improves tolerance and production of biorenewables', *Metabolic Engineering*, 44, pp. 1–12. doi: https://doi.org/10.1016/j.ymben.2017.08.006.

Tenaillon, O. *et al.* (2010) 'The population genetics of commensal Escherichia coli', *Nature Reviews Microbiology*, 8(3), pp. 207–217. doi: 10.1038/nrmicro2298.

Terradot, G. *et al.* (2021) 'The proton motive force determines Escherichia coli's robustness to extracellular pH', *bioRxiv*, p. 2021.11.19.469321. doi: 10.1101/2021.11.19.469321.

Thea, K. *et al.* (2004) 'A Regulatory Trade-Off as a Source of Strain Variation in the Species Escherichia coli', *Journal of Bacteriology*. American Society for Microbiology, 186(17), pp. 5614–5620. doi: 10.1128/JB.186.17.5614-5620.2004.

Thomason, L. C., Costantino, N. and Court, D. L. (2007) ' E. coli Genome Manipulation by P1 Transduction ', *Current Protocols in Molecular Biology*, 79(1), pp. 1–8. doi: 10.1002/0471142727.mb0117s79.

Tomberlin, J. K. *et al.* (2017) 'Indole: An evolutionarily conserved influencer of behavior across kingdoms', *BioEssays*. John Wiley & Sons, Ltd, 39(2), p. 1600203. doi: https://doi.org/10.1002/bies.201600203.

Tramonti, A., De Canio, M. and De Biase, D. (2008) 'GadX/GadW-dependent regulation of the Escherichia coli acid fitness island: transcriptional control at the gadY–gadW divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites', *Molecular Microbiology*. John Wiley & Sons, Ltd, 70(4), pp. 965–982. doi: https://doi.org/10.1111/j.1365-2958.2008.06458.x.

Tramonti, A. *et al.* (2002) 'Functional Characterization and Regulation of gadX, a Gene Encoding an AraC/XylS-Like Transcriptional Activator of the Escherichia coli Glutamic Acid Decarboxylase System', *Journal of Bacteriology*. American Society for Microbiology, 184(10), pp. 2603–2613. doi: 10.1128/JB.184.10.2603-2613.2002.

Travisano, M. and Lenski, R. E. (1996) 'Long-Term Experimental Evolution in Escherichia coli. IV. Targets of Selection and the Specificity of Adaptation', *Genetics*, 143(1), pp. 15–26. doi: 10.1093/genetics/143.1.15.

Tsai, M.-F., McCarthy, P. and Miller, C. (2013) 'Substrate selectivity in glutamate-dependent acid resistance in enteric bacteria', *Proceedings of the National Academy of Sciences*, 110(15), pp. 5898 LP – 5902. doi: 10.1073/pnas.1301444110.

Tucker, D. L., Tucker, N. and Conway, T. (2002) 'Gene Expression Profiling of the pH Response in Escherichia coli', *Journal of Bacteriology*. American Society for Microbiology, 184(23), pp. 6551–6558. doi: 10.1128/JB.184.23.6551-6558.2002.

van Opijnen, T. and Camilli, A. (2013) 'Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms', *Nature Reviews Microbiology*, 11(7), pp. 435–442. doi: 10.1038/nrmicro3033

van Opijnen, T., Bodi, K. L. and Camilli, A. (2009) 'Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms', *Nature Methods*, 6(10), pp. 767–772. doi: 10.1038/nmeth.1377.

Vivijs, B., Aertsen, A. and Michiels, C. W. (2016) 'Identification of Genes Required for Growth of Escherichia coli MG1655 at Moderately Low pH', *Frontiers in Microbiology*, 7. doi: 10.3389/fmicb.2016.01672.

Wall, E., Majdalani, N. and Gottesman, S. (2018) 'The Complex Rcs Regulatory Cascade', *Annual Review of Microbiology*. Annual Reviews, 72(1), pp. 111–139. doi: 10.1146/annurev-micro-090817-062640.

Wang, Y. (2002) 'The Function of OmpA in Escherichia coli', *Biochemical and Biophysical Research Communications*, 292(2), pp. 396–401. doi: https://doi.org/10.1006/bbrc.2002.6657.

Watson, N. *et al.* (1992) 'Identification of elements involved in transcriptional regulation of the Escherichia coli cad operon by external pH', *Journal of bacteriology*. Am Soc Microbiol, 174(2), pp. 530–540. doi: 10.1128/jb.174.2.530-540.1992.

Wilks, J. C. and Slonczewski, J. L. (2007) 'pH of the cytoplasm and periplasm of Escherichia coli: Rapid measurement by green fluorescent protein fluorimetry', *Journal of Bacteriology*, 189(15), pp. 5601–5607. doi: 10.1128/JB.00615-07.

Yang, J. and Zhang, Y. (2015) 'I-TASSER server: new development for protein structure and function predictions', *Nucleic Acids Research*, 43(W1), pp. W174–W181. doi: 10.1093/nar/gkv342.

Yanofsky, C., Horn, V. and Gollnick, P. (1991) 'Physiological studies of tryptophan transport and tryptophanase operon induction in Escherichia coli', *Journal of Bacteriology*. American Society for Microbiology, 173(19), pp. 6009–6017. doi: 10.1128/jb.173.19.6009-6017.1991.

Yasir, M. *et al.* (2020) 'TRADIS-XPress: A high-resolution whole-genome assay identifies novel mechanisms of triclosan action and resistance', *Genome Research*, 30(2), pp. 239–249. doi: 10.1101/gr.254391.119.

Yuk, H.-G. and Marshall, D. L. (2004) 'Adaptation of Escherichia coli O157:H7 to pH Alters Membrane Lipid Composition, Verotoxin Secretion, and Resistance to Simulated Gastric Fluid Acid', *Applied and Environmental Microbiology*. American Society for Microbiology, 70(6), pp. 3500–3505. doi: 10.1128/AEM.70.6.3500-3505.2004.

Zarkan, A. *et al.* (2019) 'Indole Pulse Signalling Regulates the Cytoplasmic pH of E. coli in a Memory-Like Manner', *Scientific Reports*, 9(1), pp. 1–10. doi: 10.1038/s41598-019-40560-3.

Zarkan, A. *et al.* (2020) 'Local and Universal Action: The Paradoxes of Indole Signalling in Bacteria', *Trends in Microbiology*. Elsevier Ltd, 28(7), pp. 566–577. doi: 10.1016/j.tim.2020.02.007.

Zhang, C., Freddolino, P. L. and Zhang, Y. (2017) 'COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information', *Nucleic Acids Research*, 45(W1), pp. W291–W299. doi: 10.1093/nar/gkx366.

Zhang, M. *et al.* (2011) 'A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance', *Nature Chemical Biology*, 7(10), pp. 671–677. doi: 10.1038/nchembio.644.

Zhang, Y., Morar, M. and Ealick, S. E. (2008) 'Structural biology of the purine biosynthetic pathway', *Cellular and Molecular Life Sciences*, 65(23), pp. 3699–3724. doi: 10.1007/s00018-008-8295-8.