Examining the response of

Escherichia coli and *Pseudomonas aeruginosa*

to organic acid stress.

By

Francesca M. L. Bushell

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

DOCTOR OF PHILOSOPHY

Institute of Microbiology & Infection

School of Biosciences

University of Birmingham

September 2018

Abstract

Weak organic acids have been used for centuries to treat infections and preserve food. However, the detailed mechanisms by which they exert their effect on the cell are not fully understood. Two aspects of organic acid action are presented in this thesis. In the first study, a combination of transcriptomics, transposon-insertion sequencing (TraDIS) and evolution were used to probe the molecular mechanism by which acetic, propionic and butyric acid at both neutral and mildly acidic pH affect cellular process in the UPEC strain *E. coli* EO499. The effects were numerous and complex, with all three organic acids having a large impact on metabolic processes. Six populations of EO 499 evolved independently at pH 5.5 with 4 mM acetic acid showed increased fitness in that environment, but that fitness was not replicated in an environment of pH 5.5 alone. A cross-comparison of the data derived from these different approaches reveals a highly complex network of genes and responses which are not easily interpreted. Comparison of the methods revealed some overlap between the TraDIS and RNAseq data, with a large proportion of the genes which were considered significant in TraDIS also showing differential regulation. This indicates that with more sophisticated bioinformatics techniques, it should be possible to build up intricate networks of the response to organic acids. The second study examined the inhibition of growth and biofilm formation in *P. aeruginosa* PA01 and a clinical isolate PA1054 by a range of organic acids. This study was devised to help inform the use of organic acids as a topical treatment for burn wounds. A laboratory plate reader was used to generate growth data in eight different organic acids at three different concentrations and five different pH. This data was then analysed (in collaboration with colleagues) using three mathematical approaches, two parametric and one Bayesian. The plates used for growth were also stained to quantify biofilm formation. This study found that the most effective acid at inhibiting both growth and biofilm formation was propionic acid and the least effective was benzoic acid. The effects of all acids were significantly enhanced by reduction in pH. These data indicate that the use of organic acids as a topical treatment would likely inhibit wound infections. However, the currently used concentrations of ~800 mM could be reduced if combined with mildly acidic solution.

I dedicate this work to my parents, John and Roberta, without whom

it would not have been possible.

Thank you for everything.

Acknowledgements

My most sincere and heartfelt thanks go to Dr Pete Lund who has been an inspirational teacher and mentor. Five years ago I did not envisage I would ever be in this position and thanks to Pete I am able to realise a long-held dream and ambition. Something for which I will always be truly and profoundly grateful.

This project was a result of a number of collaborations with talented people to whom I extend my thanks. They are Dr Thippesh Sannasiddappa, Dr Sara Jabbari, Dr Peter Tonner, Dr Amy Schmid, Mathew Milner, Dr Francesco Falciani and Dr John Herbert who have all contributed significantly to the work in this thesis. I also thank Fatima Alatar for being a good friend, for introducing me to Arabic coffee and rahash, and for letting me use her competition data in this report.

I came to the lab from a law firm, and without the guidance of Thippesh S and Hrishiraj Sen, who taught me a huge amount, I would not have been able to do any of this work. They were generous with their time and incredibly patient. I am very grateful to them both for being such good lab-mates and friends. Working in T101 is a unique experience and one that I will treasure forever. I thank everyone in T101 who made my time there so immensely enjoyable.

I am grateful to Ewelina Kozlowska, who has been extraordinarily accommodating and understanding during the writing up of this thesis when I should have been at work!

I am grateful to my family for putting up with me during the last four years. Particularly to Harri for her supportive messages and packages of coffee. I am even more grateful to my parents for their constant and unwavering support in every way possible, but particularly for giving me the financial means to make all of this happen. I feel truly lucky to have you both in my life.

Last, I thank Ryan. My true love. Without his encouragement and support none of this would have happened and I cannot imagine my life any other way. It was Ryan's idea to email Pete and it was Ryan who has unfailing helped and encouraged me every day since, right to the last sentence. Together we can achieve anything.

Contents

Chapter One:

Introduction

1.1 Introduction

The ability both to survive and proliferate under acid stress is a vital adaptive mechanism for survival in the mammalian system for pathogenic and commensal bacterial species. Bacteria are required to tolerate extreme inorganic acid stress in the stomach when entering the host via ingestion (Castanie-Cornet et al., 1999). After transiting the stomach, which can reach a pH of as low as 2, bacteria then pass into the intestinal tract, with a pH ranging from 5.6±0.2 in the caecum to 6.6±0.1 in the colon (Cummings et al., 1987). This range of pH is achieved by a combination of both organic and inorganic acids that vary in both type and concentration across the intestinal tract. The organic acids in the intestines are a combination of those produced by the intestinal cells and also the fermentation processes of gut bacteria, which result in the production of short chain fatty acids (Macfarlane and Macfarlane, 2003).

The gastrointestinal tract has a very high concentration of short chain fatty acids (SCFA). This cocktail of organic acids is primarily made up of acetate, with butyrate and propionate as the next most abundant, with small amounts of others making up the remainder. These short-chain fatty acids are found in the human intestine at concentrations ranging from ~13mM in the ileum to ~130mM in the caecum and ~80mM in the descending colon. SCFAs are used as an energy source by the bacteria, as well as being produced as a by-product of fermentation. However, butyrate is also used as an energy source for colonocytes (the epithelial cells of the colon), adding to their integral role in the normal functioning of the gastrointestinal tract (Vinolo et al., 2011).

This project has investigated the effects of acids, both inorganic and organic, on the bacteria *E. coli* and *P. aeruginosa*, using a range of methods. This topic is important as it has ramifications for our understanding of how bacteria can survive passage through the human GI tract, and also because organic acids are widely used as antibacterial food additives and increasingly in medical contexts given

the steady and relentless increase in antibiotic resistance. Despite this wide usage, our understanding of much of the underlying biology of their effects on bacteria is surprisingly limited. This introduction is divided into several linked sections. First, there will be a description of the bacteria that have been used in this research and why. The focus will then be on the known biology of *E. coli* acid resistance mechanism and cell homeostasis. Then, specifically organic acid stress will be considered in detail.

1.2 *Escherichia coli* EO499

The role of the human microbiome is widely studied and is becoming understood more clearly, not only in terms of pathogens, but also their role within, and impact upon, normal gut conditions. The bacteria, fungi and other microorganisms that constitute the human microbiota make up a very large proportion of the cells within the human system. Most of these are located in the gastrointestinal tract and play a vital role in protection against disease, by outcompeting pathogens, and in digestion, by breaking down carbohydrates and proteins during fermentation. The concentration of the microorganisms within the gut, and therefore the relative local environment and activity of these organism (in both protective and metabolic terms), has been linked to a number of conditions recently which were previously thought to be unrelated to gut microbiota, such as obesity and mental health (Vinolo et al., 2011; Schmidt, 2015). This demonstrates the vital and varied role that the gut microbiota play in the health and wellbeing of humans (and of course other species), but also how pathogens might exploit a change in the populations and environment of 'resident' strains and species in order to cause opportunistic infection and disease. Species such as *E. coli* comprise many strains which play a role within the complex microbiota, some as commensals and others that only become pathogenic once the opportunity presents itself due to a change in the environment, possibly due to immunodeficiency caused by an unrelated factor. Other strains of *E. coli* are almost always pathogenic to humans.

The ability of pathogenic bacteria to colonise the host is dependent on a number of factors such as motility and adherence to host cells via *O* polysaccharide chains termed the K antigen. Also, in some cases the production of β-lactamase enzymes enables antibiotic resistance, namely to penicillin-derived β-lactam antibiotics such as cephalosporins, resulting from the production of CTX-M extended-spectrum β-lactimases (ESBL) (Lavigne et al., 2012; Phillips et al., 2001). β-lactam antibiotics contain a β-lactam ring which binds to and inhibits the transpeptidase active site, preventing the enzyme from performing the reaction vital to the synthesis of the bacterial cell wall i.e. peptidoglycan linkage. The bacterium then bursts open, due to osmotic instability, killing it (Lehninger et al., 2008). Some bacteria that have been widely exposed to these β-lactam antibiotics and derivatives such as cephalosporins have evolved to produce enzymes called β-lactamases which hydrolyse the β-lactam ring before it can bind to and inhibit transpeptidase. The primary strain of study in this thesis, *E. coli* EO499, belongs to serotype ST131, which produces CTX-M ESBL, (originally identified and named by Matsumoto et al., (1988) for its ability to hydrolyse and thereby inactivate cefotaxime, a cephalosporin antibiotic) (Bonnet, 2004). Since then, other types of CTX-M enzymes have been discovered in other strains of *E. coli*, as well as in different species, and now there are a large number of enzymes, which comprise the group CTX-M (123 as at December 2011). The genes responsible for these enzymes are chromosomal in *Kluyvera* species and in *E. coli* these enzymes are plasmid mediated (with uptake probably occurring by chance) (Cantón et al., 2012).

E. coli EO 499 is a uropathogen. Urinary tract infections are among the most common infections, infecting a reported 150 million individuals worldwide per year, with *E. coli* being the cause of approximately 80% of those infections. Common isolates include multi-locus sequence types (ST) 131, ST69, ST73 and ST95. These strains are also commonly isolated from the faecal flora, indicating their presence in the microbiota as commensal strains, and are, therefore, opportunistically pathogenic (Flores-Mireles et al., 2015; Schembri et al., 2015).

1.3 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is a ubiquitous gram-negative bacterium that is commonly isolated from a variety of environments ranging from kitchen sinks to soil. This is due to the bacterium's ability to grow in a wide range of environments, making it extraordinarily versatile (Youenou et al., 2014).

P. aeruginosa is an opportunistic human pathogen, mostly infecting those who are already immuno-compromised. The ability of *P. aeruginosa* to adapt to alternative environments means it can easily colonise the lungs of those with cystic fibrosis and cause pneumonia in intensive care patients, cause urinary tract infections in those with catheters and infect burns and other open wounds (Rybtke et al., 2015). The fact that *P. aeruginosa* is so versatile and such a successful opportunistic pathogen is due to a number of factors. Low membrane permeability provides an inherent resistance to many antimicrobials. *P. aeruginosa* has membrane permeability approximately 8% that of *E. coli* (Chevalier et al., 2017). Porins are defined in the literature as outer-membrane β-barrel proteins, which allow the passage of solutes and contribute to membrane stability. The term porin has historically been used to describe large non-specific outer-membrane channels, however there are two classes of outer-membrane channel in *Enterobacteriaceae;* these large non-specific channels, and also solute-specific channels. *P. aeruginosa* possess no large non-specific channels, only solute-specific (Chevalier et al., 2017; Henderson et al., 2016). This is one of the primary reasons for low membrane permeability. P. *aeruginosa* also possess several efflux pumps, which can extrude the majority of currently used antibiotics, as well as many other compounds, providing further resistance (Alav et al., 2018). Additionally, extended-spectrum β-lactamases have been identified in *P. aeruginosa* (Cholley et al., 2014). Together with efflux pumps, ESBLs provide *P. aeruginosa* with resistance to penicillin-derived antibiotics, aminoglycosides and fluoroquinolones (Potron et al., 2015). Finally, *P. aeruginosa* forms effective biofilms. Biofilms allow *P. aeruginosa* to colonise a wide variety of environments by encasing the cells in extracellular polymeric substances (EPS). The EPS holds the cells together and is comprised

mainly of extracellular DNA, exopolysaccharides and polypeptides, which form a hydrated mixture that acts as a means of providing the cells with structure. Biofilms provide a barrier between the cells and the external environment, preventing antimicrobials from reaching the bacteria. These bacterial communities encased in biofilm are then able to expand and are very difficult to treat. This makes *P. aeruginosa* infections of immune-compromised patients a challenging problem to overcome (Alav et al., 2018; Rasamiravaka et al., 2015). The particular issue of biofilms within infection will be discussed in more depth in Chapter 8.

Infections and sepsis are the leading cause of death in patients with a thermal injury who do not succumb to the injury itself. *Pseudomonas aeruginosa* has previously been isolated from one-third of burn wounds and 59% of those with extensive injuries. Due to the reduced efficacy of other treatments, particularly antibiotics, acetic acid is more commonly being utilised as a topical treatment for burn wounds to manage infection (Halstead et al., 2015).

This aspect of the research project will be discussed in greater detail in Chapter 8. The main focus of this research was *E. coli* EO499, and so the remainder of the introduction will be centred on the relevant acid resistance pathways within *E. coli* and organic acid stress in *E. coli.*

1.4 Acid resistance pathways of *E. coli*

Sabo et al., (1974a) described the optimal conditions by which to induce and purify lysine and arginine decarboxylase and noted an optimum pH of 5.6. The authors of this study referenced the work of Gale, (1946) who had hypothesised that the bacterial cell may be using these decarboxylases as a means to raise intracellular pH during mild acid shock. They observed that growth of mutants unable to induce arginine decarboxylase was very limited at pH 5.1, but showed no impact on growth at pH 6.1 (Sabo et al., 1974a).

Hengge-Aronis, (1993) described the role of stationary phase alternative sigma factor RpoS in the regulation of gene expression during transition to, and cell maintenance during, stationary phase. This sigma factor allowed the cells to adapt to stress conditions, such as reduced nutrient availability, by inducing glycogen synthesis. RpoS was also implicated in acid tolerance in stationary phase by regulation of genes which, at the time, were uncharacterised. The 'generalised' stress response phenotype that this study exhibited implicated RpoS as a global regulator, rather than a stress-specific response.

This study was followed by Lin et al., (1995) who demonstrated that there were at least three acid resistance pathways utilised by *E. coli* during extreme acid stress (defined as a pH below growth range). Their results showed that when cultivated in rich medium containing either the amino acids glutamate or arginine, *E. coli* cells were able to survive at pH 2-2.5 (the equivalent of the human stomach during fasting). However, in minimal media that was not supplemented with amino acids, survival rates were vastly reduced. The third mechanism by which *E. coli* survived acid stress was oxidative and glucose repressed. This is now known to be the RpoS-dependent acid stress response. Together with the lysine-dependent response, there are a total of four (three amino acid dependent and one amino acid independent) known acid resistance pathways in *E. coli*. These four acid resistance (AR) pathways have now been defined in much more detail (Lin et al., 1995).

The three amino acid-dependent acid resistance (AR) systems require amino acid decarboxylases GadAB, AdiA and CadA for the glutamate (AR2), arginine (AR3) and lysine (AR4) systems respectively. Together with their cognate antiporters GadC, AdiC and CadB these systems operate by exchanging the intracellular decarboxylase reaction products, namely glutamate/γ-aminobutyric acid, agmatine and cadaverine for their relevant extracellular amino acid. This process removes protons from the cell thereby decreasing intracellular pH (Tramonti et al., 2008a). The processes that govern these responses

will be discussed in more detail below starting with the acid fitness island and then the most widely studied acid resistance pathway, the GAD system.

1.5 The Acid Fitness Island

Many of the genes most associated with acid resistance as outlined above can all be found in the same region of the genome termed the Acid Fitness Island (AFI). The notable exception is the *gadBC* operon, which is located separately from the genes within the AFI. *GadA* is downstream of the *gadX* gene, which, along with *gadW,* encodes the transcriptional regulator of the GAD system. *GadA* can therefore be transcribed with *gadX*, or can be independently transcribed. GadX and GadW are both encoded within the AFI and this region is repressed by H-NS (histone-like nucleoid structuring protein) and controlled by RpoS (Mates et al., 2007; Tramonti et al., 2008b). Figure 1.1 below shows a schematic of the AFI.

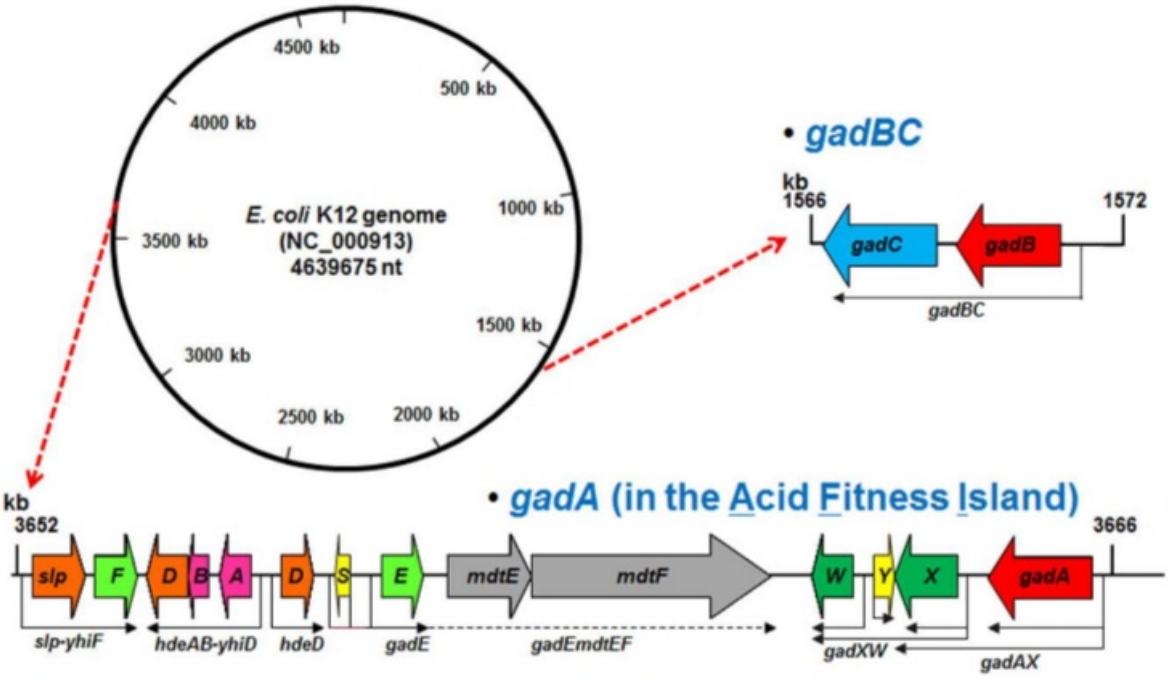


Figure 1.1: The Acid Fitness Island schematic and its location on the genome.
Genes coloured as follows: Glutamate decarboxylases: red; GABA antiporter GadC: blue;
Additional membrane proteins: orange; Acid stress periplasmic chaperones: pink;
LuxR-like transcriptional regulator: light green; AraC-like transcriptional regulators: dark green;
Multi-drug exporters: grey; Small regulator sRNAs: yellow.
Taken from De Biase and Pennacchietti, (2012)

GadE, a LuxR-type regulator, is a master regulator of the GAD system and is a key activator of the *gadA* and *gadBC* genes. GadE binds to the GAD box, which is a 20bp region 16bp upstream on the -35 element of the *gadA* promoter. Regulation of GadE is very complex, requiring RcsB, a response regulator, and part of the RcsCDB phosphorelay. RcsB can be induced together with, or independently to, this phosphorelay, and when induced separately is induced by RcsA, PhoP, TviA or RmpA which act as co-regulators (Castanie-Cornet et al., 2010).

1.6 AR2: The GAD system

AR2, the glutamate-dependent system, is the most effective, and therefore the most widely studied (Richard and Foster, 2007). AR2 is the glutamic acid-dependent (GAD) system which relies on glutamate decarboxylase isozymes GadA and GadB converting glutamate to GABA (glutamate/γ-aminobutyric acid) by taking up intracellular protons and replacing the α-carboxyl group in glutamate with a proton from the cytoplasm. This reduces the number of protons within the cell. The products of this reaction are CO_2 and GABA (See Figure 1.2 below). This is then removed from the cell via the antiporter GadC in exchange for glutamate (Foster, 2004). This results in the continuous consumption of protons.

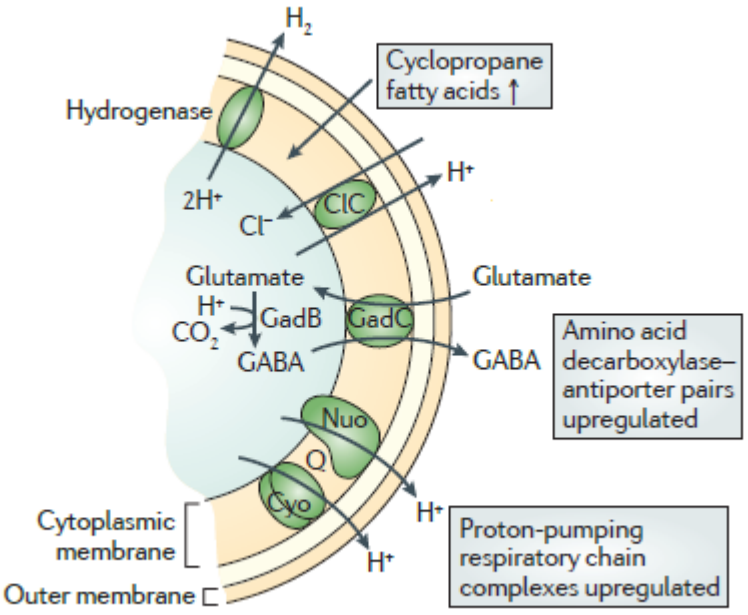


Figure 1.2: *E. coli* under acid challenge. Taken from Krulwich et al., (2011)

The GadC antiporter, which removes protonated GABA from the cell and brings in extracellular glutamate in a strict one-to-one exchange, has itself been the subject of study. Glutamate can exist in three differently charged carboxyl-group states, α-γ-double deprotonated Glu^-, α-γ-double protonated Glu^+ and α-deprotonated γ-protonated Glu^0. Similarly, the carboxyl-group of GABA can exist in two charged states: protonated $GABA^+$ and deprotonated $GABA^0$ (See Figure 1.3 below). This means there are six possible combinations for the exchange of GABA through the GadC antiporter: $Glu^-/GABA^+$, $Glu^+/GABA^+$, $Glu^0/GABA^+$ and $Glu^-/GABA^0$, $Glu^+/GABA^0$ and $Glu^0/GABA^0$. This degree of variation has implications on the cell and the efficacy of the AR2 system in that exchanging deprotonated GABA for protonated glutamate ($Glu^+/GABA^0$) would increase rather than decrease the proton concentration in the cell. Alternatively, exchanging neutrally charged GABA for neutrally charged glutamate ($Glu^0/GABA^0$) or positively charged GABA with positively charged glutamate ($Glu^+/GABA^+$) would have no net impact on the concentration of protons within the cell and would therefore provide no benefit during acid shock (Ma et al., 2013; Tsai et al., 2013). Glutamate carries a γ-carboxyl side-chain with a pK_a of ~4.2 and

an α-carboxyl side-chain with a pK_a of ~2.1. During the extreme acid stress of the stomach at around pH 2, therefore, the γ-carboxyl side-chain would be ~99% undissociated. At pH 2 50% of the α-carboxyl side-chain would be dissociated and 50% would be undissociated. This means that in the extreme acid stress of the stomach approximately 50% of glutamate would exist in the Glu^0 form and approximately 50% in the Glu^+ form (with around 1% in the Glu^- form) (Tsai et al., 2013). Upon entering the cytoplasm, with a pH nearer neutral, the undissociated α-carboxyl side-chain protons would dissociate, exacerbating the problem of free protons within the cytoplasm. It was therefore hypothesised that GadC demonstrated selectivity as to which form of glutamate was in fact imported into the cell, and conversely, which version of GABA was exported.

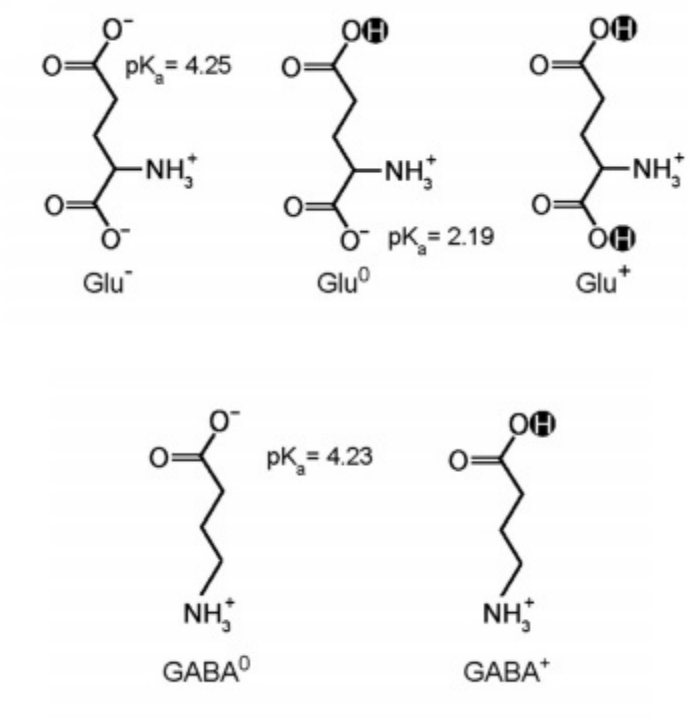


Figure 1.3: The differently charged states of Glutamate and GABA. Adapted from Ma et al., (2013).

At neutral pH GadC was shown by Ma et al., (2013) to be inactive and activated at pH 6.5. Both the study by Ma et al., (2013) and that of Tsai et al., (2013) published in the same month of that year show that

GadC is able to recognise the net charge of the glutamate that is imported into the cell (only importing Glu^0 and Glu^-) and is also able to recognise the charged state of GABA, only exporting $GABA^+$. In this way, the cell is able to effectively combat proton concentration within the cell without exacerbating the problem, by importing protonated glutamate in exchange for deprotonated GABA (Lund et al., 2014a; Ma et al., 2013; Tsai et al., 2013).

In a study by Castanie-Cornet et al., (2010) it was demonstrated that RcsB is a crucial co-activator, together with GadE, of *gadA*. RcsB, a response regulator, is regulated together with GadE to activate expression of *gadA/BC* in order to allow for acid resistance. GadE and RcsB form a heterodimer and it is this that binds to the GAD box initiating transcription, rather than just GadE which had previously been suggested. This was confirmed by Johnson et al., (2011) who demonstrated that in a Δ*rcsB* background, expression of *gadA* was significantly reduced. Johnson et al., (2011) also showed that there was a significant decrease in GadE-dependent promoters of the genes *hdeA* and *hdeB*. There was also a decrease in expression of GadE-independent promoters *gadX* and *gadW*. It has been previously shown that these promoters are regulated by the two component system EvgAS when the cell is exposed to acid stress. However, the expression of EvgAS was not affected when *rcsB* activity was suppressed. The authors hypothesised therefore that *rcsB* was involved in the function of the EvgAS two component system. Johnson et al., (2011) tested this hypothesis by introducing constitutively active EvgS into wild type and Δ*rcsB* strains. The Δ*rcsB* cells showed no survival at pH 2, despite constitutive EvgS activity, whereas wild type strains showing constitutive EvgS activity were able to survive at pH 2 both with and without induction at mild pH before exposure to extreme acid stress. The first major regulator downstream of EvgA, YdeO, is dependent on RcsB. This implies that RcsB has a direct role in the function of EvgA or EvgS, either via EvgS-dependent phosphorylation of EvgA or via regulatory activity of EvgA.

In summary, Johnson et al., (2011) showed that the effect of AR2 in mild acid shock was dependent on RcsB in both exponential and stationary phase and that this effect was not alleviated by introducing constitutively active EvgS. EvgS is part of a EvgAS, which is a two component system, a number of which are vital components in the acid resistance pathways of *E. coli.*

1.7 Two component systems

Acid resistance in *E. coli* requires a number of two and three component systems which are PhoPQ, EvgAS and RcsB (as described above). Two component signal transduction systems (TCS) are systems in bacteria which identify changes in the environment and send a signal via phosphorylation which induces a cellular response to allow for adaption to that environmental change (Eguchi and Utsumi, 2014).

EvgS is a sensor kinase membrane protein composed of a histidine kinase domain, receiver domain and Hpt transmitter domain. The EvgS sensor kinase activates a cascade of transcription factors: EvgA, YdeO and GadE. During activation of EvgS the phosphate at the initial histidine residue in the histidine kinase domain is transferred to an aspartate residue in the receiver domain. This is then relayed to the second histidine residue in the Hpt domain and is then transferred to the aspartate residue of EvgA. EvgS has a large periplasmic region which consists of two bacterial periplasmic solute-binding protein domains (PBPb). At the N-terminal of EvgS it is predicted that there is a signal peptide, then the two PBPb domains and a transmembrane region. Therefore, it is predicted that the N-terminal of the signal transmembrane domain of EvgS protrudes into the periplasmic space (Eguchi and Utsumi, 2014).

It has been shown that mild acid shock (pH 5.5) induces EvgSA activity in exponential phase cells, which in turn activates acid resistance. However, the mechanism via which EvgSA 'senses' acid stress is not currently understood. Eguchi and Utsumi, (2014) showed that activation of EvgS was dependent on factors such as Na^+ and K^+ concentration, as well as low pH. The authors cultivated cells in M9 medium

supplemented with KCl or NaCl and used *lacZ* as a reporter for EvgS activity. EvgS activation was also dependent on the concentration of these monovalent cations, with activation occurring at around 150mM of Na^+ plus K^+ in M9 minimal media.

There are known to be small membrane proteins that regulate the activity of two component systems. One such protein is the sensor associating factor A, or SafA, which is a 65 amino acid membrane protein that modulates the activity of PhoPQ. SafA is directly induced by EvgAS and so it was hypothesised on that basis that PhoPQ also plays a role in acid resistance (Eguchi et al., 2012, 2011). SafA, which is regulated by EvgA, binds to the PhoQ histidine kinase and activates the sensory domain of PhoQ in the periplasm. PhoQ has been shown to respond to a number of signals, including low pH and divalent cations such as Mg^{2+} and Ca^{2+}. Activation of the PhoPQ two component system improves acid resistance via regulation of RpoS levels, via the connector IraM, and therefore RpoS-regulated acid resistance genes (discussed in more detail below) (Eguchi et al., 2012; Ishii et al., 2013; Lund et al., 2014b).

PhoPQ negatively regulates transcription factor YdeO, which is involved in the EvgA→YdeO→GadE cascade. Specifics about the regulatory functions of YdeO are not completely understood. However, it has been determined that YdeO plays a critical role in the glutamate dependent acid resistance pathway in that activation of YdeO via EvgA in turn regulates activation of GAD components such as *gadA, gadBC* and *gadE*. Yamanaka et al., (2014) showed that YdeO was activated at pH 5.5 and pH 7 in anaerobic conditions implying a role in anaerobic survival as well as acid resistance. This result indicates that YdeO plays a vital role in the ability of *E. coli* to colonise the gut by providing the cells with the means to respond to both anaerobic and mildly acidic conditions simultaneously.

Yamanaka et al., (2014) also showed that YdeO stimulated the transcription of the *hdeAB* operon. HdeA and HdeB are periplasmic chaperones. The periplasm is more susceptible to damage caused by acid

stress because this region of the cell is subject to more extreme changes in pH. Exactly how these chaperones work is still under investigation. However, it is understood that these dimeric proteins dissociate into monomers under acid challenge. Activation of HdeA appears to be the protonation of negatively charged residues, which in turn contributes to dissociation of the dimer. HdeA and HdeB then block the aggregation of unfolded proteins, where denaturation has been caused by acid damage, by binding to the denatured proteins and inhibiting their acid-induced aggregation. HdeA and HdeB then promote the renaturation of these proteins once acid challenge has subsided and the cell is in recovery. This renaturation process coincides with a return to the dimeric inactive conformation of these proteins (Dahl et al., 2015; Lund et al., 2014b).

Dahl et al., (2015) showed that the optimal pH for the action of HdeA was pH 2, with partial dissociation of HdeA at pH 3.4 consistent with lower levels of activity. However, for HdeB the optimal pH was pH 4. This reflects the range of pH that *E. coli* will likely encounter in the human stomach and therefore provides the cell with a range across which it can respond effectively to acid challenge. The investigators also showed that at pH 4, when HdeB activity was optimal, HdeB was in its dimeric conformation. This implies a different mode of action to HdeA, although protonation of negatively charged residues is likely to be the point of activation in both cases.

1.8 AR3 & AR4: The arginine and lysine dependent systems

AR3 and AR4 are amino acid dependent response systems (arginine and lysine respectively) that operate in aerobic and anaerobic conditions. Arginine-dependent acid resistance relies on the decarboxylation of arginine to agmatine by the decarboxylase AdiA. This is then exported from the cell via the AdiC antiporter in a one-to-one exchange. Like glutamate, arginine can exist in different charged states. Therefore, like GadC, it is predicted that AdiC can distinguish which form of arginine to import to the cell by recognising the net charge of the amino acid so as not to exacerbate the problem of free protons

within the cytoplasm. The transcription factor AdiY initiates expression of AdiA and AdiC and is shown to be active at a pH of 6 or lower (Tsai and Miller, 2013).

The lysine-dependent AR4 system is regulated by the CadC protein which is a positive regulator of the *cadAB* operon. During acid challenge lysine in imported into the cell and decarboxylated to produce cadaverine by the decarboxylase CadA, which is then exported via the antiporter CadB (Casalino et al., 2010; Lund et al., 2014b).

It has been shown that these two acid resistance systems rely on the outer-membrane proteins OmpF and OmpC, which are acid-induced porins. In a study by Bekhit et al., (2011) they demonstrated that when these genes were deleted function of the AR3 and AR4 systems was reduced. However, in the case of the AR2, the GAD system, the addition of exogenous glutamate to the growth media enabled the normal function of this system (under aerobic conditions). However, when arginine and lysine were added to the media, the AR3 and AR4 systems still showed vastly reduced activity (under anaerobic conditions) and poor levels of survival of mutants. This implies therefore that OmpF and OmpC are required for full induction of AR3 and AR4, potentially via the import of the amino acids and export of their decarboxylation products. In support of this hypothesis, double mutants, with both proteins knocked out, were more susceptible to acid challenge than single mutants (Bekhit et al., 2011).

1.9 The RpoS-dependent system

The final acid resistance pathway that will be outlined here is the RpoS-dependent system, and the only known system that is not contingent on an amino acid decarboxylase. RpoS is the alternative sigma factor sigma^{38}. During the initiation of transcription the sigma subunit of RNA polymerase is responsible for recognising and binding a particular promoter (Hickey et al., 2007; Chattopadhyay et al., 2015). RpoS is under very tight regulation in order to ensure that it does not compete for core RNA polymerase when

it is not required. As a result it is found in very low levels during exponential phase and any RpoS detected during this growth phase is quickly degraded (Battesti et al., 2011). This degradation is mediated by the RssB protein, which directs RpoS to ClpXP, a protease. The activity of RssB is however repressed by PhoP, which is activated at low pH, allowing the accumulation of RpoS in the cell when under acid challenge (Lund et al., 2014b). Burton et al., (2010) showed that RpoS also appears to play a role in the regulation of the GAD AR2 system via the enhanced expression of *gadE*. GadE initiation is regulated by YdeO. However, in a Δ*ydeO* background there was evidence of GadE auto-regulation. This study showed that in wild type cells there was an initial surge in GadE activity, followed by a slower increase at later time points. There is a cascade of induction, with a gap of ~10 minutes between induction of *safA-ydeO* and *gadB*, with *gadE* falling between the two. However, although YdeO is required for the initial increase in AR2 activity, in a Δ*ydeO* background the system is still activated, albeit at later time points following a delay. After initial induction of the *gadB* promoter, RpoS is required for continued activation in later stages of growth, and in a strain with a *rpoS* deletion, AR2 activity is reproducibly lower at later time points. The results of this study indicate that RpoS activity does not rely on YdeO activity, and despite deletion of *ydeO*, promoters which are usually YdeO-dependent showed a delayed gradual increase in activity. This activity was however completely absent in a *rpoS-ydeO* double mutant (Burton et al., 2010). A representation of this system is shown below in figure 1.4.

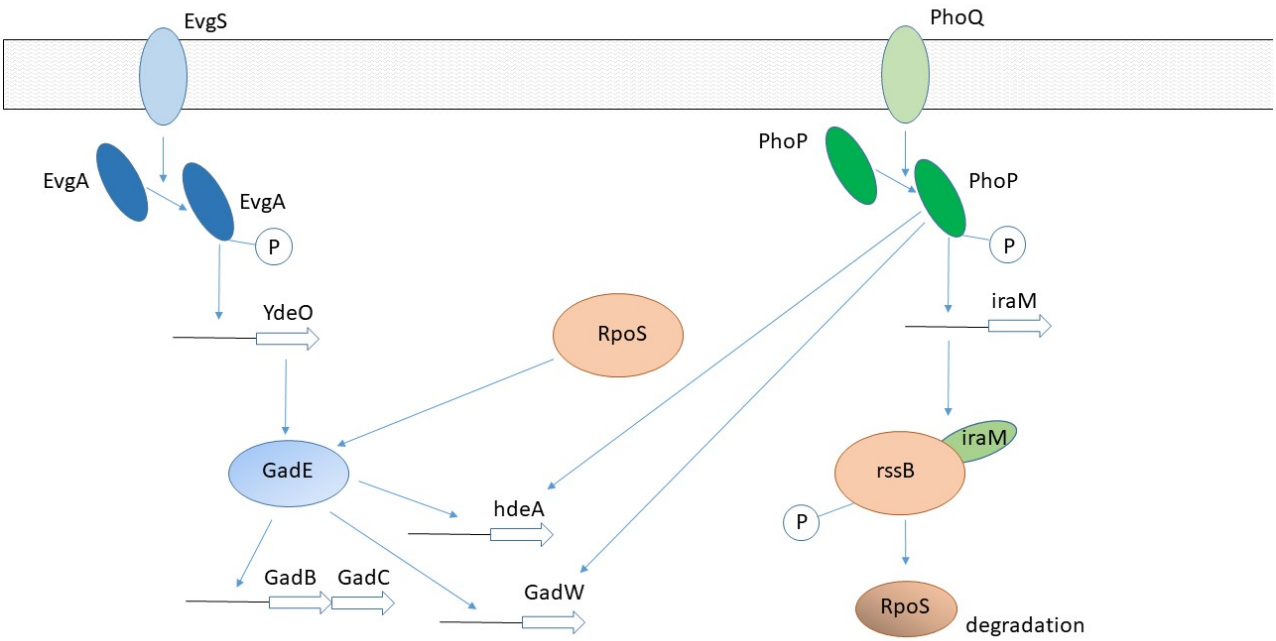


Figure 1.4 The regulatory cascade in the RpoS acid-resistance system. Adapted from Burton et al., (2010); Eguchi et al., (2011).

RpoS regulates the transcription of a significant number of genes required for cell maintenance during stationary phase in response to environmental challenges such as osmotic shock, temperature and carbon starvation. It is important to note that whilst the cell may have RpoS-dependent responses to these stresses, the cell also has a number of response pathways, such as those in the case of pH, which are RpoS-independent (Battesti et al., 2011).

The RpoS regulated response to pH is an oxidative response which is glucose repressed. Therefore, it becomes activated under carbon starvation, which implicates CRP. *E. coli* prefer glucose as a primary carbon source, though they are capable of utilising a wide range of other metabolites as a carbon source. When glucose becomes scarce or unavailable, the cells shift from metabolising glucose to other sources of available carbon. This transition induces activation of the adenylate cyclase CyaA, which produces high levels of cyclic AMP (cAMP). Binding of cAMP to the cAMP catabolite receptor protein

(CRP), which is a transcription factor, forms the cAMP-CRP complex, and directly regulates the expression of genes responsible for metabolism of alternative sources of carbon (Foster, 2004; Battesti et al., 2011; Donovan et al., 2013). In studies such as those described by Donovan et al., (2013) and Hengge-Aronis, (2002) Δ*crp* and Δ*ubiA* mutants showed increased levels of RpoS during exponential phase, indicating negative regulation of *rpoS* transcription. Hengge-Aronis, (2002) confirmed this with the addition of exogenous cAMP to growth media, which reduced levels of RpoS in exponential phase cells. The full mechanisms via which RpoS implements acid resistance are, to date, not fully understood. However, it is known that this system is based on the F0F1 ATPase, which pumps protons from the cell whilst consuming ATP. This system is regulated by both RpoS and CRP (Stincone et al., 2011). Maintenance of homeostasis in this way is incredibly important to the cell whilst under acid stress as the requirement to remove protons from the cell in order to maintain cytoplasmic pH can have a knock-on effect on other systems within the cell such as the electrochemical gradient.

1.10 Maintenance of cell homeostasis under acid challenge

A metabolic response to mild acid stress (~pH 5.5) utilised by *E. coli* is the upregulation of selected components of the electron transport chain including cytochrome *bo* oxidase, NADH dehydrogenase II, succinate dehydrogenase and NADH dehydrogenase I. These comprise part of the system by which proton motive force is generated. This occurs via the coupling of redox reactions with the direct or indirect export of protons from the cell. The cytochrome *bo* oxidase complex accepts electrons from ubiquinol-8 and catalyses the production of H_2O from intracellular oxygen and protons. This means that two protons are removed from the cell for each electron removed. By upregulating this process during acid stress *E. coli* can remove protons from the cell more efficiently (Kanjee and Houry, 2013).

Most bacteria possess membrane-bound proton pumps which remove protons from the cytoplasm in order to generate the proton motive force, which is an electrochemical gradient of protons across the

cell membrane coupled with a pH gradient (Mitchell, 1961; Booth, 1985). The electron transport carriers of *E. coli* are membrane associated. Electron transport systems have two functions: first, they mediate the transfer of electrons from donor to acceptor. Second, they capture some of the energy liberated in the breakdown of oxidised molecules like glucose. The cells use this partly to create a proton gradient which drives several key processes in the cell, most importantly ATP synthesis. These electron transport carriers are organised in the membrane so that when electrons are transported, electrons and protons are separated. NADH is the most important electron donor, although *E. coli* is capable of using multiple donors, such as formate and H_2. In the case of NADH, in aerobic respiration, the following reaction takes place: $2NADH + 2H^+ + O_2 \rightarrow 2NAD^+ + 2H_2O$. When glucose is metabolised to pyruvate, two molecules of NADH are made. Subsequently, pyruvate is decarboxylated to acetyl-CoA, which is then oxidised to CO_2. This reaction provides additional NADH. Succinate, an intermediary in the TCA cycle, can also be used by *E. coli* as an electron donor (Madigan and Brock, 2009; Gunsalus, 2011; Booth, 1985).

The electron transport chain is partly comprised of two key enzyme types, dehydrogenases, which oxidises an electron donor such as NADH, and cytochrome oxidases, which reduces the electron acceptor, for example O_2 to H_2O. *E. coli* has three cytochrome oxidases which are distinct in their ability to pump protons and their affinity for oxygen. All three cytochromes can accept electrons from a number of alternative dehydrogenases, depending on the availability of its substrate, for example organic or inorganic. Electrons are transferred from one to the other via quinones, such as ubiquinone and menaquinone (Unden and Bongaerts, 1997; Gunsalus, 2011). Carriers in the electron transport chain are ordered according to their reduction potential, with the final point being the terminal acceptor. The protons come from NADH, other electron donors and the dissociation of water to H^+ and OH^-. H^+ is then removed from the cell via membrane bound pumps and accumulates on the outside of the cell. An inevitable outcome of the removal of H^+ from the cell is the accumulation of OH^- inside the membrane. This results in a net negative charge inside the cell caused by OH^- and a net positive charge outside the

membrane caused by H^+. The result therefore of the electron transport chain is twofold. The first is the transmembrane pH gradient (ΔpH), which is alkaline inside the cell compared to outside (*E. coli* maintains a cytoplasmic pH of 7.8). The second part is the transmembrane electrical potential (ΔΨ) which is negative inside the cell compared to outside. *E. coli* has a small ΔpH at pH 7, but a significant ΔΨ (Krulwich et al., 2011). Together these two elements constitute the proton motive force. The proton motive force causes the membrane to be energised, and some of this energy is then conserved in the formation of ATP. The energised state of the membrane also plays a role in other cellular functions, such as ion transport and flagellar rotation (Booth, 1985; Madigan and Brock, 2009).

1.11 Proton motive force

The proton motive force drives ATP synthesis, which is required for a range of cellular functions such as reproduction and maintenance. When the proton motive force is disrupted, so too is ATP synthesis. This has a knock-on effect of disrupting ATP-dependent cellular functions. ATP synthesis is reliant on ATP synthase, or ATPase. This is a complex which converts the proton motive force into ATP. ATPase is made up of two parts, the F_1, which is a multi-protein extramembrane complex which faces into the cytoplasm, and the F_0 which is a proton-conducting intramembrane channel. ATPase functions by catalysing a reversible reaction between ADP + P_i to ATP. Protons cross the membrane through the F_0 channel, which dissipates the proton motive force and drives ATP synthesis (Madigan and Brock, 2009). ATP synthase has been shown to be downregulated during acid challenge, and upregulated during base challenge. This is consistent with its role as a proton import channel. However this downregulation during acid challenge results in the discontinuation of ATP synthesis via ATPase, which again would cause a knock-on effect to ATP-dependent cellular functions (Maurer et al., 2005).

Maintenance of energy is required for many metabolic processes, for example cell division, DNA repair and other cellular processes. Sun et al., (2011) showed that survival of *E.* coli under acid stress required

ATP both in the presence and absence of amino acids. Under mild pH 5.5 acid stress, ATP levels increased compared to levels at pH 7.5. At pH 2.5, levels decreased and were eventually lost. The authors were unable to draw solid conclusions as to why the levels of ATP increased under mild acid stress, but noted that membrane potential at pH 5.5 was low. They also suggested that it may be that cellular processes requiring ATP may be unable to function at pH 5.5, causing an accumulation of ATP in the cell. It was suggested that the decreased growth rate at pH 5.5 may support this. ATP is essential to DNA repair mechanisms and without these repair mechanisms functioning (in the absence of ATP) the cells were unable to survive acid shock. *E.* coli may have an unknown acid resistance mechanism and the data indicate an ATP-dependent repair mechanism that is essential to acid stress (Sun et al., 2011).

Under extreme acid stress conditions *E. coli* reverse their ΔΨ to a positive charge. This could be because of a build-up of positively charged decarboxylation by-products of acid resistance mechanisms, and could be to prevent the further uptake of protons, by charge repulsion. They could of course be inter-linked, with one leading to the other, as another form of acid resistance. The CIC protein, which exchanges H^+ ions for Cl^- ions, is thought to play a role in returning the ΔΨ back to its normal negative charge once the acid stress has subsided (Kanjee and Houry, 2013).

1.12 The role of ions in acid stress

West and Mitchell, (1974) recorded the maintenance of cell membrane potential via the addition of Na^+ after the removal of H^+ from the cell under acid stress. The role of Na^+ in pH homeostasis is now widely acknowledged (Padan, 2014; Krulwich, 1983; Pan and Macnab, 1990). The authors proposed that this came about by way of a Na^+/H^+ antiporter, which removes H^+ from the cell in exchange for Na^+, thereby maintaining pH homeostasis.

Environmental signals such as acid and osmotic stress are vital in the induction of acid resistance pathways. Na^+ levels affect *gadA/BC* expression during exponential phase and this appears to be through the modulation of GadX and GadW activity which control *gadE* expression. The uptake of Na^+ is mediated via the Na^+/H^+ antiporter system that helps maintain cytoplasmic pH when the bacteria are under acid stress. Although it is possible to conclude that Na^+ is important in acid resistance, it has not been possible to determine whether Na^+ plays a role in inducing acid resistance or whether Na^+ uptake is an actual mechanism of acid resistance. It has been concluded therefore that Na^+ can act as a counter-ion in the process that removes H^+ from the cell and this may be the only role played in acid resistance. However, two further possibilities are considered: the first that the increase in Na^+ as a result of this process induces acid resistance mechanisms and second: that Na^+ plays some physiological role in the acid resistance mechanisms themselves (Richard and Foster, 2007).

At pH 2.5 in media containing no Na^+ the cells were able to survive in the presence of glutamic acid. This indicates that Na^+ does not itself play a physiological role in acid resistance mechanisms. It is also noted that in cells grown in the absence of Na^+ the GadA/BC proteins are expressed at much lower levels than in those bacteria grown in the presence of Na^+. This therefore indicates that Na^+ plays a role in the induction of *gadA/BC* expression. Na^+ plays a role in GadE expression, where expression is three times higher when the bacteria are grown in high Na^+. It has not been possible yet to determine whether that is due to the presence of Na^+ being detected and causing induction or whether Na^+ is binding to the Gad proteins. The evidence strongly indicates that the increase in internal Na^+ concentrations brought about by the decrease in cytoplasmic pH is in part responsible for the induction of the GAD system (Richard and Foster, 2007).

The role of potassium in maintaining cell homeostasis is also important. As mentioned above, Eguchi and Utsumi, (2014) demonstrated that EvgS activity was dependent upon K^+ and Na^+ concentrations of

around 150 mM in the growth media. Potassium is imported into the cell via a number of K^+ transport systems. The main systems for K^+ transport are Kdp, which is a K^+-dependent ATPase, Trk, which is a lower-affinity system and acts as a H^+/K^+ symport, and Kup, which also appears to involve proton symport. Trk and Kup require proton motive force to maintain activity as they also import H^+ into the cytoplasm (Kitko et al., 2010). Under aerobic conditions Trk is driven by the PMF and the binding of ATP has a regulatory role in the influx of K^+. It is thought that under anaerobic conditions Trk forms a complex with the FoF1 ATP synthase, which is thought to act as a H^+/K^+ antiport. This subsequently promotes proton extrusion in order to maintain homeostasis, and the loss of antiport activity leads to an inability of the cell to increase cytoplasmic pH (Kitko et al., 2010; Kroll and Booth, 1981).

Kroll and Booth, (1981) describe the role of potassium across three key points:

i) There appears to be a relatively simple relationship between net K^+ uptake and ΔpH, which is independent of the presence of glucose (i.e. independent of respiratory rate).

ii) Extent of depolarisation of ΔΨ by net K^+ uptake is dependent on respiratory rate, and the ability of the cell to maintain ΔΨ during K^+ uptake is significantly enhanced in the presence of glucose.

iii) The PMF remains constant during K^+ uptake. However, distribution between ΔpH and ΔΨ is dependent on net K^+ uptake.

K^+ appears to play a significant role in maintenance of internal pH. It was hypothesised that the maintenance of ΔpH and ΔΨ is via the influx of a positive charge that is not H^+ during acid challenge, when the cell is attempting to increase internal pH through the efflux of protons, while also maintaining the transmembrane electrical potential (Kroll and Booth, 1981).

The external pH has a profound effect on the pH gradient generated during potassium uptake. At an external pH (pH_o) of 5.3, the internal pH (pH_i) increases rapidly from 6.3 to 7.6, decreases to 7.55 and then remains constant. At pH_o of 6.3, the pH_i rises to 8, acidifies to 7.6 and then, again, remains constant. The hypothesis was that K^+ uptake causes the cells to 'overshoot' the pH_i, and then undergo a controlled reduction to pH 7.6. However, in K^+-depleted cells, incubated at pH 5.3 but with an internal pH of 7.1 (and therefore with a pre-established ΔpH), this overshoot effect was not repeated on the addition of K^+ to the media. Instead, the pH_i rose from 7.1 to 7.6, where it remained constant. This result suggests that the mechanism of prevention of an excessively alkalised cytoplasm is directly related to ΔpH, as this reduction in alkalisation is more efficient as pH_o decreases. The authors of this work conclude, however, by noting that these results were replicated in Kdp and TrkA double mutants. This indicates that this result is mediated by overall pH_i-maintenance systems (including Na^+/H^+ exchange and Cl^-/HCO_3^- exchange) and not solely on the import of potassium ions. It may also be that the cell was utilising other means of importing potassium into the cell that the authors were not aware of at the time (Kroll and Booth, 1981).

A strain defective in all three main K^+ uptake systems, Kup, Kdp and Trk, shows a decreased growth rate at low pH when coupled with low K^+ levels. The cells also fail to maintain pH_i at pH_o of 5.9 with 100 mM K^+ in the growth media. However, addition of excess K^+ restores homeostasis in triple mutants. This could be via transport of potassium through multiple minor transport pathways (Kitko et al., 2010). The role of K^+ appears to extend further than driving efflux of protons and that another important function may be osmoregulation. The addition of K^+, Na^+ and osmolytes to wild type cells allows recovery of cytoplasmic pH back to pH 7.8, and the relative effect of each is equivalent. In the K^+-transport system triple mutants, exogenous KCl restores pH homeostasis during rapid acid shift. Growth rates were almost restored to those of wild type cells when excess KCl was added to growth media. The

mechanisms that link osmoregulation and cytoplasmic pH are unclear, but may involve maintaining cell volume and stabilising ion fluxes (Kitko et al., 2010).

The next section will focus specifically on weak organic acids and the stress these exert on the cell.

1.13 Weak organic acids

The role of cations and anions is vital when understanding the impact of organic acids upon cellular homeostasis under acid challenge. In the case of weak organic acid stress, weak acid freely diffuses across the membrane and then dissociates inside the cell causing a sudden increase in H^+ and the organic acid anion simultaneously. The impact upon the cell is multiple. The sudden influx of protons causes the acidification of the cytoplasm. The sudden accumulation of anions also impacts upon the cell, although the consequences of this are less clear and one of the key points of focus of this project is elucidating the genetic response of the cell to both the impact of the proton and the anion.

The Brønsted-Lowry definition of an acid is a substance that donates a proton (H^+). A base is a substance that accepts a proton. For example, when hydrochloric acid, HCl, gas is dissolved in water a reaction occurs whereby HCl donates a proton, and water acts as a base and accepts a proton. This results in H_3O^+ and Cl^-. The chloride ion is the conjugate base of the reaction, and the hydronium ion, H_3O^+, is the conjugate acid (McMurry, 1992). Put simply:

$$\underset{\text{Acid}}{H\text{-}A} + \underset{\text{Base}}{:B} \leftrightarrows \underset{\text{conjugate base}}{A:^-} + \underset{\text{conjugate acid}}{H\text{-}B^+}$$

Water is capable of acting as a base, as in the reaction stated above with HCl. However, water is also capable of acting as an acid, such as when it reacts with the amine ion, $^-NH_2$, by donating a proton to form the hydroxide ion OH^- and ammonia, NH_3.

Acids are described as being either weak or strong, depending on their ability to donate protons. Strong acids such as HCl react almost completely with water, whereas weak acids, such as acetic acid, react only partially with water. A standard equilibrium equation is used to determine the equilibrium constant K_{eq}, which determines the strength of an acid:

$$HA + H_2O \leftrightarrows A^- + H_3O^+$$

$$K_{eq} = \frac{[H_3O^+][A^-]}{[HA][H_2O]}$$

Where HA represents any acid

The acidity constant (K_a) of any acid, HA, is the equilibrium constant multiplied by the molar concentration of pure water, 55.5 M.

When considering the equilibrium equation '$HA + H_2O \leftrightarrows A^- + H_3O^+$', strong acids have their equilibrium towards the right of the equation, whereas weak acids have a smaller acidity constant, and therefore lie towards the left of the equation. The strength of an acid is usually expressed as pK_a, which is the negative logarithm of the acidity constant; $pK_a = -\log K_a$.

A strong acid has a large K_a and a small pK_a. A weak acid, such as acetic acid, has a low acidity constant, and a larger pK_a (McMurry, 1992).

It is important to note that the relationship of acid strength to that of its conjugate base is inverse. For example, though HCl is a strong acid, its conjugate base Cl^- is a weak base, and has a low affinity for a proton. Acetic acid is a weak acid and therefore the conjugate base acetate is stronger than Cl^-, with a stronger affinity for a proton. When acetic acid is mixed with water, water acts as a base as the

hydroxide ion has a stronger affinity for a proton than the acetate ion. This results in the following reaction:

$$CH_3COOH + OH^- \rightleftharpoons CH_3COO^- + H_2O$$

However, when acetic acid is mixed with hydrochloric acid it acts as a base, given that the acetate ion has a much higher affinity for a proton than the chloride ion. The following reaction occurs:

$$CH_3COO^- + HCl \rightleftharpoons CH_3COOH + Cl^-$$

If the pK_a of an acid is known, it is possible to determine the concentration of protonated, deprotonated and neutral forms at any given pH. This can be expressed using the Henderson-Hasselbalch equation, which is derived by rearranging the equation for pK_a:

$$pH = pKa + \log \frac{[A^-]}{[HA]}$$

$$\text{Or: } \log \frac{[A^-]}{[HA]} = pH - pKa$$

The pKa of acetic acid is 4.75. table 1.5 below shows a table of acetic acid and hydrochloric acid and the percentage dissociated versus undissociated at pH 7 and pH 5.5.

	pH	pKa	H^+ (10^pH)	K_a (10^pK_a)	HA ((H^+x2)/K_a)	% Diss (H^+/(A^-+HA) x 100)	% Undiss (100-%diss)
Acetic Acid	7	4.75	0.0000001	1.78E-05	5.62E-10	99.44	0.56
Acetic Acid	5.5	4.75	3.162E-06	1.78E-05	5.62E-07	84.90	15.10
HCl	7	-7	0.0000001	10000000	1E-21	100.00	0.00
HCl	5.5	-7	3.162E-06	10000000	1E-18	100.00	0.00

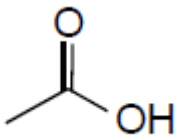
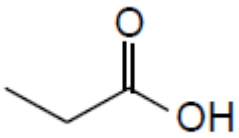
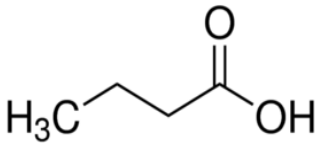
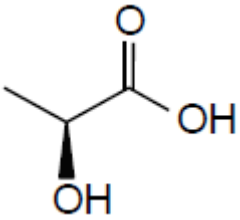
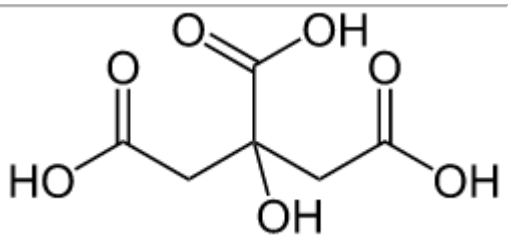
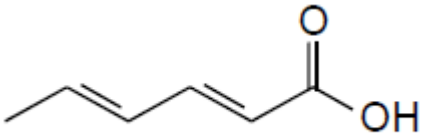
Table 1.1: Table demonstrating the percentage dissociated and undissociated values for acetic acid and hydrochloric acid at pH 7 and pH 5.5.

The mechanisms by which *E. coli* is able to survive the weak organic acids encountered in the gut are also important to understand better. Not only are the bacteria exposed to a combination of organic acids, primarily acetic acid, propionic acid and butyric acid, but *E. coli*, along with other commensals and pathogens, produce acetic acid as a by-product of fermentation. At a lower pH acetic acid is at higher concentrations in its undissociated (i.e unionised) form, as described above. In its undissociated form acetic acid is lipid soluble as it carries no charge. This allows the acid to cross the cell membrane upon which it enters the higher pH of the cytoplasm. When this occurs it dissociates, releasing protons and creating a subsequent build-up of anions and protons. This in turn lowers the intracellular pH of the bacterium, causing disruption to the cell (Roe et al., 1998).

1.14 Organic acids, their impact upon the cell and their role in cellular processes

Studies into the effectiveness of acetic acid, and other organic acids, as antimicrobials have focussed heavily on the food industry, with the use of acids to preserve food and prevent bacterial

contamination. Table 1.2 below shows weak organic acids which are commonly found in the gut and / or used in industry.

Acid	**HA**	**A^-**	**Chemical Structure**	**pKa**
Acetic Acid	CH_3COOH	CH_3COO^-		4.75
Propionic Acid	CH_3CH_2COOH	$CH_3CH_2COO^-$		4.87
Butyric Acid	$CH_3CH_2CH_2COOH$	$CH_3CH_2CH_2COO^-$		4.82
Lactic Acid	$CH_3CH(OH)CO_2H$	$CH_3CH(OH)COO^-$		3.86
Citric Acid	$H_3C_6H_5O_7$	$H_2C_6H_5O_7^-$		3.08
Sorbic Acid	$C_6H_8O_2$	$C_6H_7O_2^-$		4.75

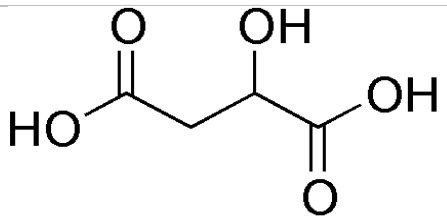
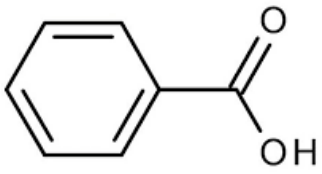
Malic Acid	$C_4H_6O_5$	$C_4H_5O_5^-$		3.40
Benzoic Acid	$C_7H_6O_2$	$C_7H_5O_2^-$		4.2

Table 1.2 Organic acids commonly found in the GIT and used in industry

The mechanisms by which acetic acid, as well as other organic acids, inhibit cellular processes have been widely studied but are still not fully understood. Roe et al., (1998) investigated the effects of organic acid on *E. coli* using acetic and benzoic acid. This study looked more closely at the build-up of acetate anions within the cell and noted that in order to compensate for this build-up, the bacteria reduced the size of other anion pools, such as glutamate anions. They also noted that once the bacteria were then transferred back to neutral pH, recovery of intracellular pH was dependent on glutamate synthesis. Therefore, changes to cellular function as a result of acetic acid cannot therefore be put down to the release of protons alone (Roe et al., 1998).

Diez-Gonzalez and Russell, (1997) examined the difference in acetate tolerance between *E. coli* O157:H7 and K-12. K-12 was able to grow well at pH 5.9 under anaerobic conditions. However, the addition of a small amount of sodium acetate, at concentrations of only 20 mM, was enough to have a marked impact on growth rate. O157:H7 was at least four times more resistant to acetate than K-12. In the presence of sodium acetate, both strains increased glucose consumption, which was converted to acetate, formate

and ethanol. However, as acetate concentrations increased, and subsequently glucose consumption increased, fermentation products shifted to D-lactate. O157:H7 produced more D-lactate than K-12. When exogenous acetate levels were at 140 mM, O157:H7 was producing D-lactate as its main fermentation product (K-12 growth having ceased at 60 mM acetate concentration).

At an external pH of 5.9, K-12 accumulated high amounts of intracellular acetate, though pH_i did not drop significantly. Intracellular acetate concentrations were as high as 500 mM in K-12. However, in *E. coli* O157:H7, the pH_i decreased when sodium acetate was added to the medium, and intracellular acetate concentrations never went higher than 300 mM. Both O157:H7 and K-12 had intracellular potassium levels of approximately 300 mM. However, K-12 increased its potassium levels as acetate accumulated to more than 300 mM. The ratio of intracellular to extracellular acetate anions was in close agreement with ΔpH at higher sodium acetate levels. Both K-12 and O157:H7 maintained a pH_i of approximately 6.8 when grown in media at pH 5.9. K-12 maintained a ΔpH of 0.9 units as sodium acetate concentrations increased. However, O157:H7 allowed its ΔpH to decrease by just 0.2 units as acetate concentrations increased. Differences in ΔpH caused large differences in intracellular acetate accumulation. Both strains increased intracellular potassium levels to match intracellular acetate concentrations. This ability of O157:H7 to tolerate a decrease in pH_i, and subsequently a decrease in ΔpH to only 0.2 units, marks out a significant advantage to withstand organic acid stress, which is not replicated in non-pathogenic K-12. The mechanisms by which O157:H7 exhibit this response are an interesting area of further research (Diez-Gonzalez and Russell, 1997).

The aim of a study conducted by Breidt et al., (2004) was to determine the effects of acetic acid and pH on *E. coli* O157:H7 independently and combined, mimicking conditions found in commercially pickled foods. The authors cultured *E. coli* O157:H7 in buffered growth medium, and acidified that medium with either HCl or acetic acid. They reported that for *E. coli* O157:H7 the minimum inhibitory pH of acetic acid

is one pH unit higher than that of HCl alone, with pH 5.5 providing the inhibitory value in the presence of acetic acid, compared to pH 4.5 in the presence of HCl. Temperature is a very important factor. A higher temperature resulted in increased effectiveness of organic acids as inhibitory to growth, noting that the effectiveness of lower concentrations of acetic acid was markedly increased as temperature increased (between 10^{o}C and 30^{o}C). The combined effect of pH and acetic acid was far more effective than the effect of pH alone. However, this research utilised acetic acid concentrations of between 100 mM and 400 mM (coupled with pH of between 3.1 and 4.1). These are conditions typical of acidified food products, the context in which this study was conducted (Breidt et al., 2004). For the purposes of this report these concentrations are too high, however the vast majority of research into the bacteriostatic and bactericidal effects of acetic acid have been done primarily in the field of food safety and preservation and it therefore may be difficult to draw relevant assertions from these studies.

One of the most important points when considering the effect of weak organic acids on the cell is that acetate is produced by the cell but can also be metabolised by the cell, as can propionic acid, citric acid and lactic acid. Figure 1.5 outlines the routes via which organic acids are used and produced in cellular metabolism.

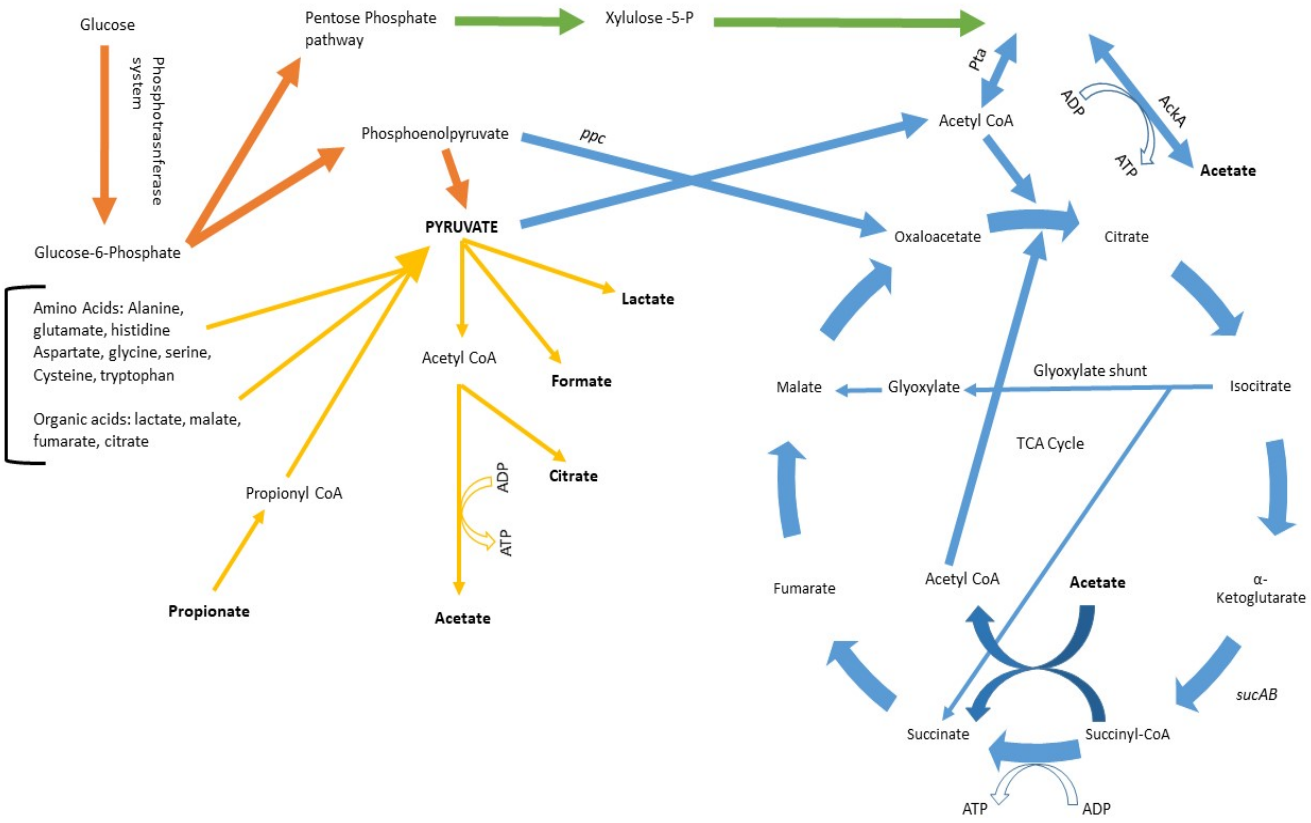


Figure 1.5 Pathways of organic acid use and production in cellular metabolism. Blue shows aerobic TCA cycle, orange shows glycolysis, green shows pentose phosphate pathway and yellow shows anaerobic fermentation. Adapted from Bernal et al., 2016; Kwong et al., 2017; Madigan et al., 2009.

These pathways are an additional consideration when trying to better understand the response of *E. coli* to organic acid stress, and will be explored in more detail throughout this thesis.

1.15 Aims & Objectives

The main aim of this project was to identify the mechanisms by which the commonly isolated UPEC *E. coli* strain EO 499 (NCTC 13441), of the serotype ST131, is able to respond to levels of organic acid in the human gut in order to survive, proliferate and cause opportunistic infection of the urinary tract. This was done by examining the genetic basis of the response of EO 499 to organic acid stress response by:

a) Using TraDIS genomic sequencing in order to examine the genes which appear to play a role in mild organic acid stress response. Survival in extreme organic acid stress was not examined in detail as this was not considered within the scope of how the bacteria would survive in the human system.

b) Generation of RNAseq data obtained under the same conditions as those used in TraDIS in order to examine whether the transcriptional response is comparable to the phenotypic response.

c) Evolution of twelve lines of EO 499, from a common ancestor, six at pH 5.5 and six at pH 5.5 with 4 mM acetic acid (again mirroring two of the conditions used to generate TraDIS and RNAseq data).

Once this data was analysed using bioinformatics, completed by our collaborators in Liverpool, genes and / or networks of genes which appear to be playing a significant role were looked at in more detail. This was done by using TraDIS mutants to examine whether this produces a change in phenotype under the conditions used in the preparation of samples for sequencing.

The last aspect of this project was the study of *Pseudomonas aeruginosa* PA01 (strain ATCC 15692) and the burns isolate PA1054 (isolated from a burn patient at the Queen Elizabeth Hospital). This was done by:

a) Growth of both strains in a range of pH both with and without a number of different organic acids at different concentrations.

b) A study of biofilm formation in the same conditions to see whether different organic acids affect biofilm formation and in what way.

The intention of this was grounded in the use of acetic acid as a topical treatment for burn wounds. We examined whether acetic acid is a suitable treatment for burn wound infections, whether other organic acids are more suited to this job, and whether organic acids are in fact a suitable treatment at all.

Chapter Two:

Materials & Methods

Chapter 2 Materials and Methods

2.1 Strains used in this thesis

The strains and source of each strain is listed below in Table 2.1.

Table 2.1 Strains used and relevant source or reference.

Species	Strain	Source/Reference
E. coli	ST131 EO 499 (NCTC 13441)	Discuva Ltd, Cambridge
E. coli	K12 MG1655	Blattner et al, (1997)
P. aeruginosa	PA01	ATCC 15692
P. aeruginosa	PA1054	QEHB clinical burn isolate

2.2 Culture media for *E. coli* and *P. aeruginosa* growth

Unless stated otherwise, all bacteria were grown in buffered M9 minimal media supplemented with glucose and cas-amino acids (referred to hereon as 'supplemented M9'). A breakdown of the media components and final working concentrations are contained in Table 2.2 below. Medium was brought to pH using either 4 M hydrochloric acid or 10 M sodium hydroxide, added drop by drop until the desired pH was reached. The starting pH of the media was pH 5.8 before addition of acid or base.

MacConkey lactose agar plates for competition experiments were made by combining 40 g/l MacConkey agar base with 10 g/l lactose.

Where stated, some overnight cultures were grown in LB broth or on LB agar plates. LB broth and agar contained 10g/l tryptone, 10g/l NaCl and 5g/l yeast extract and was made and provided by University of Birmingham School of Biosciences central services.

Table 2.2: Components of supplemented M9 media.

Compound	Formula	Final working concentration
di-Sodium hydrogen orthophosphate (anhydrous)	Na_2HPO_4	42.3 mM
Potassium dihydrogen orthophosphate	KH_2PO_4	22.1 mM
Sodium chloride	NaCl	8.56 mM
Ammonium chloride	NH_4Cl	18.7 mM
D-Glucose (anhydrous)	$C_6H_{12}O_6$	0.4% w/v; 22.2 mM
Casamino acids	Undefined	0.2% w/v
MOPS	$C_7H_{15}NO_4S$	100 mM
MES Hydrate	$C_6H_{13}NO_4S\ xH_2O$	100 mM (anhydrous basis)
Magnesium sulphate	$MgSO_4$-$7H_2O$	2 mM
Calcium chloride	$CaCl_2$	0.1 mM

M9 salts, which also contained 100 mM of MOPS (pKa 7.20) and 100 mM MES (pKa 5.96) was prepared and autoclaved. Magnesium sulphate and calcium chloride were prepared in 1 M batches, filter sterilised and added to the M9 salts on use. Glucose and casamino acids were made in 20% w/v batches, filter sterilised and added to the M9 salts on use.

2.3 Growth curves for determination of RNAseq and TraDIS growth conditions for sequencing

A number of conditions (specific pH with or without the addition of organic acid) were selected for growth in Erlenmeyer flasks. 25 ml of medium per 250 ml flask was prepared in advance in order to pre-warm overnight at 37^o C. An overnight culture grown from a single colony was used to inoculate each flask to a starting OD_{600} of 0.05. Cultures were grown at 37^o C with 180 rpm shaking. OD_{600} readings were then taken by pipetting culture into 1 ml cuvettes and measured using a spectrophotometer once an hour for eight hours. After three hours, the culture was diluted 1:10 before taking OD readings. Each growth experiment was conducted in biological triplicate.

2.4 Agarose gel electrophoresis

PCR products were separated and analysed using agarose gel electrophoresis. Depending on expected product size, a gel between 1% and 1.5% was made using agarose powder produced by Thermo Fisher Scientific combined with 1x TAE buffer (diluted from 50x TAE which contains 0.05 M EDTA, 1 M acetic acid and 2 M Tris, pH 8.3). Midori green stain was added to the agarose-TAE solution at a concentration of 4 µl per 100 ml solution. Agarose gels were left to set and DNA was loaded alongside the appropriate ladder (either 100bp or 1kb ladders from Bioline). Gels were run using Biorad gel tanks filled with 1x TAE buffer set to 100v for 45 – 60 minutes. Gels were viewed and photographed using the Geneflow gel doc produced by Syngene. In the case of gel extractions, bands were viewed using the non-UV blue light illuminator (Nippon).

2.5 Collection of samples for RNAseq

For RNA sample collection, three cultures were grown overnight, each inoculated from a different single colony grown in supplemented M9. Three 250 ml Erlenmeyer flasks were prepared containing 50 ml of supplemented M9 at pH 7. This was placed overnight in an incubator at 37^{o} C to pre-warm. This process also ensured there were no contaminants in the medium. Each flask was inoculated with a different overnight culture to a starting OD_{600} of 0.05. Flasks were set on shakers at 180 rpm and incubated at 37^{o} C. Growth was periodically monitored by taking the OD_{600} measurement as described above. Once an OD of 1.8 was reached, the medium in each flask was replenished in order to promote bacterial growth. This was done by removing 5 ml of culture and adding 5 ml of fresh pre-warmed medium. This was repeated twice with a 30 minute gap in-between each replenishment. After the second replenishment, bacteria were grown for a further 30 minutes. After this period, a further 5 ml of culture was removed from the flask, and 5 ml of the relevant solution as shown in figure 2.3 was added to induce the desired stress response. Growth was then continued for 15 minutes, when 5 ml of culture was removed from each

flask and transferred to an equal amount of Qiagen RNA Protect and mixed thoroughly. Growth then continued for a further 10 minutes, when a final 5 ml of culture was again transferred to 5 ml Qiagen RNA Protect. These samples were spun at 4000rpm for 5 minutes in a chilled centrifuge, the supernatant was poured away and the pellet then frozen at -80^{o} C for downstream RNA purification. All growth conditions used to grow EO 499 for RNA sample collection were cultured both aerobically and anaerobically. Anaerobic cultures were grown at 37^{o} C, with 180 rpm shaking, in a Don Whitley Scientific anaerobic cabinet. A gas mix of 5% CO_2 and 95% N_2 was used for anaerobic culturing.

2.6 RNA Purification

RNA was purified from the pellets collected as above using the Qiagen RNeasy kits according to the instructions included. The optional on-column DNA digestion step was completed for all samples to ensure there was no DNA contamination of samples. PCR amplification of the *evgS* gene was used to test for the presence of genomic DNA. All samples were then checked for quality using a Prokaryote Total RNA Nanochip, analysed using the Agilent Bioanalyser in order to provide a RNA integrity number (RIN). Only samples with a RIN of 7 or higher were used for sequencing. For Further quality control, Nanodrop and Qubit were used to ensure there was no DNA contamination and in order to quantify RNA yield. Samples were also tested for DNA contamination by PCR. Ribo-depletion was done using the Illumina Ribo-zero rRNA removal kit according to the instructions. Samples were again checked for quality using Qubit RNA quantification to measure yield.

All samples of a satisfactory standard and final yield were then transferred to the Centre for Genomic Research at the University of Liverpool for sequencing, using the Illumina 2500 HiSeq platform.

2.7 Collection of samples for TraDIS

Samples for TraDIS were collected by pre-warming three 250 ml Erlenmeyer flasks containing 50 ml supplemented M9 media pre-adjusted to the required stress condition (as indicated in Table 2.1). Flasks were incubated at 37° C and rotated at 180 rpm. Each flask was inoculated with 10 µl of the UPEC TraDIS library, giving a starting OD_{600} of 0.01. Cultures were grown for 12 hours when OD_{600} was checked. If OD had reached at least 1, then 10 ml of culture was removed from the sample. This was centrifuged for 5 minutes at 4000 rpm and the pellet frozen at -20°C for downstream processing. Growth was continued for a further 12 hours and another sample of 10 ml was centrifuged and frozen as described. All conditions were repeated anaerobically as well as aerobically. Anaerobic cultures were grown at 37° C, with 180 rpm shaking, in a Don Whitley Scientific anaerobic cabinet.

Genomic DNA was purified using the Qiagen DNeasy Blood and Tissue kit according to the instructions provided.

Table 2.3 Stress conditions used for growth of bacteria

Stress Condition	Details of addition to media to induce acid challenge
pH 7 with 40 mM acetic acid	5 ml supplemented M9 at pH 7 with 400 mM acetic acid.
pH 5.5 with 4 mM acetic acid	5 ml 1 M HCl mixed with 40 mM acetic acid. This simultaneously drops pH and dilutes to 4 mM acetic acid concentration.
pH 7 with 30 mM propionic acid	5 ml Supplemented M9 at pH 7 with 300 mM propionic acid.
pH 5.5 with 2 mM propionic acid	5 ml 1 M HCl mixed with 20 mM propionic acid.
pH 7 with 30 mM butyric acid	5 ml Supplemented M9 at pH 7 with 300 mM butyric acid.
pH 5.5 with 2 mM butyric acid	5 ml 1 M HCl mixed with 20 mM butyric acid.

Each stress condition and the component added to the culture in order to induce that stress condition identified for RNA sequencing.

2.8 RNAseq data processing

Data were analysed by Dr John Herbert from the University of Liverpool by the following method. To save on sequencing costs, a reference set of biological replicates were sequenced from EO 499 grown in either pH 7 or pH 5.5 medium to produce 3 biological replicates per condition. For each of the organic acids RNA was pooled prior to sequencing (the pools contained equal concentrations of each replicate). The resulting sequence reads were mapped to the EO 499 genome using STAR and mapped reads assigned to EO 499 features using featureCounts. An error model was then used to find significantly differentially expressed genes between pools of organic acids and matched pool controls, using the Voom function from the Limma Bioconductor package. Briefly, each organic acid pool contrast analysis contained 8 samples and 4 treatment groups (the combined 6 biological replicates of pH 7 and pH 5.5 samples, plus the two pooled samples to be compared). A Limma contrast was performed between the pools, using the information from the pH 7 and pH 5.5 replicated treatment groups to calculate statistical significance between pools.

Limma reference; https://www.ncbi.nlm.nih.gov/pubmed/25605792

STAR reference; https://www.ncbi.nlm.nih.gov/pubmed/23104886

FeatureCounts reference; https://www.ncbi.nlm.nih.gov/pubmed/24227677

2.9 TraDIS data processing

Data were analysed by Dr John Herbert from the University of Liverpool by the following method. The ESSENTIALS: Software for Rapid Analysis of High Throughput Transposon Insertion Sequencing Data was used to calculate statistical significance of conditionally essential genes between organic acids and control TraDIS library reads. The ESSENTIAL pipeline uses the edgeR package internally. More specifically it uses the "exactTest" function from edgeR within ESSENTIALS, which includes in the results a false discovery rate (FDR) adjusted P-value.

ESSENTIALS reference; https://www.ncbi.nlm.nih.gov/pubmed/22900082

2.10 Long-term lab-based evolution

E. coli EO 499 was grown in 10 ml buffered M9 minimal medium at pH 7, supplemented with glucose and casamino acids as detailed above, overnight at 37^{o} C, shaking at 180 rpm in a glass 20 ml universal bottle. An aliquot of 600 µl culture mixed with 400 µl 50% glycerol was frozen for the fossil record. This culture was then streaked out onto an agar plate and incubated overnight at 37^{o}C. Twelve colonies were selected from the plate and each used to set up a single evolution experiment. Six lines were evolved at pH 5.5 (acidified with HCl), termed 5.5/0-1 to 5.5/0-6, and six lines were evolved at pH 5.5 with 4 mM added acetic acid, termed 5.5/4-1 to 5.5/4-6. Each of the twelve lines was grown each day for 22 – 24 hours in 5 ml supplemented M9, in a 20 ml glass universal, at 37^{o} C with 180 rpm shaking. Following this period of growth 100µl of culture was then transferred directly into 5 ml of fresh medium for a total of 11 months. Every two weeks a 600 µl aliquot of each population was mixed with 400 µl 50% glycerol and stored at -80^{o}C in order to compile a comprehensive fossil record.

2.11 Primers for Contamination Checks

Primer	Sequence 5' to 3'	Reference
rfb.1bis	ATACCGACGACGCCGATCTG	Clermont et al., (2007)
rfbO25b.r	TGCTATTCATTATGCGCAGC	Clermont et al., (2007)

2.12 PCR conditions and reagents

The following conditions were used for the PCR reaction with rfb/O25b primers:

- 12.5 µl MyTaqRed (Bioline)
- 1 µl forward primer (stock conc. 10 pmol/ml)
- 1 µl reverse primer (stock conc. 10 pmol/ml)
- 0.2 µl $MgCl_2$
- 1 µl boiled colony lysate or 1 colony transferred via sterile pipette tip directly into reagent mixture.
- 9.3 µl water

In order to ensure there were no contaminants in the evolved populations, they were regularly tested for contamination by plating a sample of the population and using a minimum of 10 colonies selected at

random per population for individual colony PCR using EO 499 specific primers shown above. For each PCR cycle one control reaction was also prepared containing 1 µl water instead of lysate. PCR cycle was as follows: 5 minutes at 95°C, followed by 30 cycles of 10s 95°C, 10s at 60°C, 10s at 72°C with a final extension step of 5 minutes at 72°C. PCR product was then loaded on a 1.5% agarose gel and run for 45 minutes at 100v. The expected product from this reaction was 427bp.

2.13 Isolating mutants from the TraDIS library

Individual strains containing single transposon insertions within specific genes were isolated from the TraDIS library using a pooling and dilution technique as follows. The library was diluted to 1000 cells per ml (200 cells per 200 µl), 500 cells per ml, 100 cells per ml, plating 200 µl per well in 77 wells of a 96 well plate (leaving the last column and bottom row empty) and then incubating overnight at 37° C, 180 rpm covered with an evaporation resistant lid and sealed with parafilm. 20 µl of each sample was then pooled into the empty wells around the edge and PCR used to test for the mutant of interest using gene specific primers paired with a transposon specific primer (TnR1 or TnR2) shown in the table below. Pools which gave a band on 1.2% agarose gel were then plated on LB agar and grown overnight. Individual colonies were then again tested using target specific primers in order to isolate the mutant of interest. This method will be explained in detail in Chapter Seven.

Primer	Sequence 5' to 3'	Reference
lacZF1	AATACGCAAACCGCCTCTCC	This study
lacZR1	CGGCGTATCGCCAAAATCACC	This study
lacZF2	TGATCCTTTGCGAATATGCCCACG	This study
lacZR2	AGCCAGAAAACAAACTGATTATTGATGGT	This study
lacZF3	TCGGACCACTGTTACAATACAACAT	This study
lacZR3	AAATTGAAATTTGCATAAAAATTGCGGC	This study

TnR1	TCTTACGTGCCGATCAACGTCTC	This study
TnR2	ACATCTACCGGGTCGAATTTGC	This study
nuoM_F	CCTGGGCATATAAGAACGTAGGTCG	This study
nuoM_R	AAGCCCAATAACCGAGAGCGTAG	This study
nuoM-int1	TTGACGGTCTGTCGTTGCTGATGGT	This study
nouM-int2	TGGCAATCTGGCTTTTCGCT	This study
ytfP_F	GTCTGCAAGTGAAATACGGCGT	This study
ytfP_R	GCGGCTCGGAGAGAAAAGCAA	This study
ytfP-int1	AATATTTGTCTACGGCAGTTTACGC	This study
ytfP-int2	CTGTCTAACCAGTCGCCGCTT	This study
rssB_F	GTAAAGCCAGTCAGGGGAGAGAAC	This study
rssB_R	TTTAGCCCGTGTTATCGTTTGC	This study
apaH_F	GCCGTTCCCACACTCATTCAT	This study
apaH_R	TGCTCGACGGGATTTAAGACG	This study

Per reaction:
12.5 µl My Taq Red (Bioline)
1 µl forward primer
1 µl reverse primer
0.2 µl $MgCl_2$
4 µl template (pooled culture)
6.3 µl DNA/RNA free water

Cycle: 95°C for 3 minutes, then 30 cycles of 94°C for 30 second; 55°C for 30 second; 72°C for 5 minutes,

final extension step 5 minutes at 72°C.

2.14 Gel extraction

DNA was extracted and cleaned of PCR reagents using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. In summary, the gel was dissolved into pH stabilised buffer at 50^{o}C, which was then added to the DNA binding column. DNA bound to the silica membrane within the column and was eluted into elution buffer for downstream sequencing.

2.15 Competition experiments

An overnight culture of each evolved population and the Lac^{-} phenotype mutant isolated from the TraDIS library (acting as proxy for the ancestor strain) were grown in 5 ml supplemented M9 medium at pH 7 in a 20 ml glass universal. Following overnight growth, the evolved population was combined with the Lac^{-} mutant in 5 ml supplemented M9 medium set to pH 5.5 either with or without 4 mM acetic acid in a 20 ml glass universal to a starting combined OD600 of 0.05 (0.025 per starting culture). A 100 µl sample of this initial combination was plated on MacConkey Lactose agar and incubated overnight at 37^{o} C. A colony count was taken of these plates and recorded as time point zero (t0). The combined cultures were then grown at 37^{o} C, 180 rpm for 24 hours, serially diluted and again plated on MacConkey Lactose agar and incubated overnight at 37^{o} C. These colonies were counted and recorded at time point 24 (t24). This assay is adapted from Lenski et al., (1991).

Fitness was calculated using the equation as described in Lenski et al., (1991);

$$W=[\ln(R_{t24}/R_{t0})/\ln(V_{t24}/V_{t0})]$$

W is defined as the relative competitive index and R and V represent the two competing populations. T0 is the colony count at time point zero and T24 is the colony count at the 24 hour time point.

2.16 Whole genome sequencing of the evolved populations

Each evolved population was analysed using whole-population sequencing. Paired end sequencing on all 12 evolved populations (250 cycles) was performed by MicrobesNG (http://www.microbesng.uk) using the Illumina Miseq 2500 platform. A mean read depth of greater than 100 times was obtained for each sample. MicrobesNG prepared Fastq files, which were provided after adapter sequences were trimmed. Data were then analysed by Mathew Milner, PhD student using the Breseq tool (Deatherage and Barrick, 2014). Genome alignment and subsequent genome resequencing were analysed using the Breseq computation pipeline against an EO 499 genome (NCTC 13441) using both paired and unpaired reads. The Breseq variant calling output table was converted to .csv file and sample outputs were then combined into one table using an in-house script (devised by Mathew Milner). Mutations were only considered if they were greater than 5% frequency within a population.

2.17 High-throughput growth experiments with *Pseudomonas aeruginosa* (Chapter Eight)

Pseudomonas aeruginosa strains PA01 and clinical isolate PA1054 were grown in buffered M9 minimal medium supplemented with glucose and casamino acids as described in section 2.2. Growth kinetics of PA01 and PA1054 were recorded using the CLARIOstar automated microplate reader, produced by BMG Labtech. The bacteria were grown in a clear, flat-bottomed 96 well Costar plate. The edges of the plate (rows A and H and columns 1 and 12) contained only media and were therefore recorded as the blank. Media were brought to a pH of 7, 6.5, 6, 5.5 or 5 using either 10 M sodium hydroxide or 4 M hydrochloric acid added drop by drop (starting pH of medium is pH 5.8). Media were subsequently autoclaved and supplements were added on use. The position of pH on the plate was shifted to the right by two columns each day to negate any possible bias across the plate. For example on day one pH 7 was in columns 2 and 3, pH 6.5 in columns 4 and 5, pH 6 in columns 6 and 7, pH 5.5 in columns 8 and 9 and pH 5 in columns 10 and 11. On day two, medium with a pH of 5 was in columns 2 and 3, pH 7 in columns 4 and 5, pH 6.5 in columns 6 and 7 and so on. Organic acid was added to a final concentration of either 0

mM (control), 5 mM, 10 mM or 20 mM each in six replicates. One plate containing the conditions in triplicate which was then repeated the following day (albeit with pH position having shift two columns to the rights as described above). Figure 2.2 below shows visual representation of the plate layout.

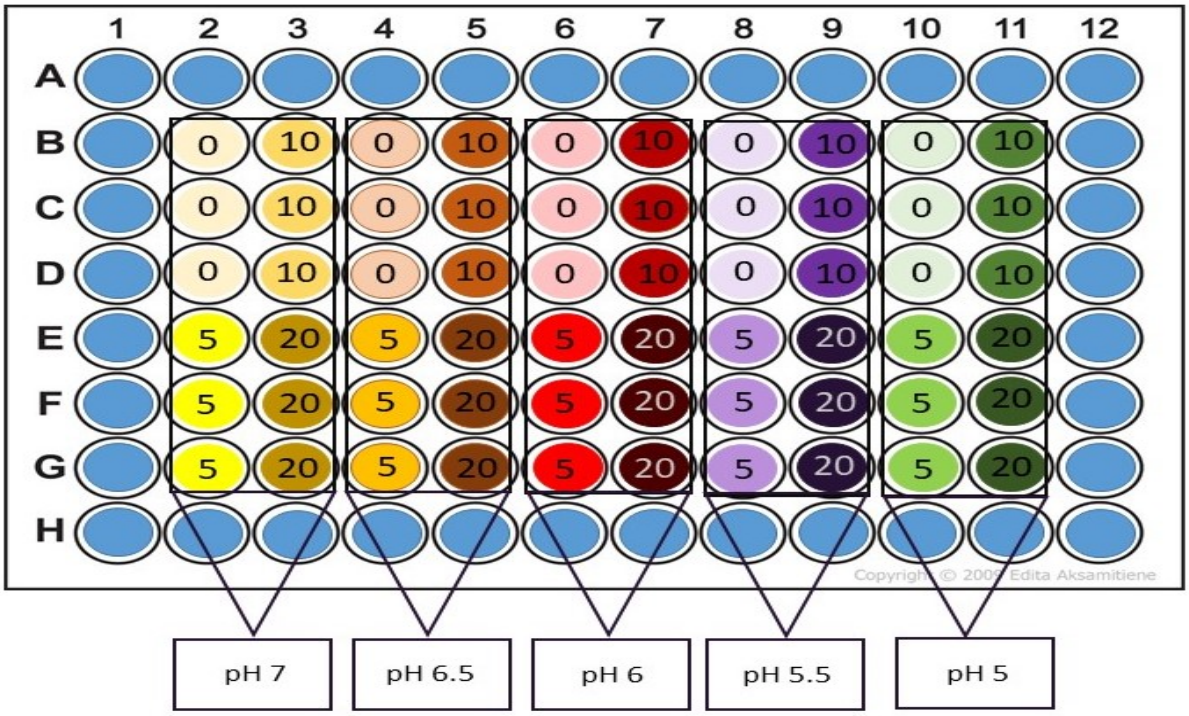


Figure 2.2 Plate layout for a triplicate growth experiment. Numbers 0, 5, 10 and 20 represent organic acid concentration in mM. pH values are that of the media within the corresponding wells. The blue wells around the edge represents position of the blanks.

Three cultures were grown overnight, each from a single colony. The optical density of the overnight culture was determined by diluting the culture 10 fold into M9 media, with M9 used as the blank. This value was used to calculate a final dilution of 0.05 starting OD, with acidified media making up the volume to 200µl in each well. The plates were incubated in the reader at 37^{o}C, with 300rpm shaking continuously and rigorous 500rpm shaking for 10 seconds prior to each read to reduce clumping. The OD_{600} was recorded every 15 minutes for a total of 24 hours. Plates were covered with an evaporation

resistant lid in order to maintain aerobic conditions whilst minimising evaporation. Each condition was repeated a minimum of six times.

2.18 Biofilm Assay

P. aeruginosa strains PA01 and PA1054 were grown in supplemented M9 media in the presence or absence of acetic acid, propionic acid, butyric acid, citric acid, lactic acid, sorbic acid, malic acid and benzoic acid across a range of concentrations and pH as described above. Biofilm formation was then quantified by emptying out the culture and gently washing the plate in water to remove loose cells. The plate was then stained using 220 µl of 0.1% crystal violet and left to stand for 20 minutes. The crystal violet was tipped out, the plate washed twice with water and blotted vigorously to remove excess dye, covered with paper towelling and left to dry overnight. Stained biofilm was re-solubilised in 30% acetic acid and the absorbance at OD_{550} nm read using the CLARIOstar microtitre plate reader with 30% acetic acid as the blank, as per O'Toole, (2011) microtitre biofilm assay.

2.19 Computational modelling

Growth data were analysed using two independent methods. Code was written to fit the data to two parametric models, Baranyi and Logistic, and executed using Matlab, a numerical computing tool which allows complex analyses of multiple variables. The code for Matlab analysis was written by Dr Sara Jabbari from the School of Mathematics at University of Birmingham. The second method used was Gaussian computational growth analysis, written and executed in its entirety by Dr Peter Tonner from Computational Biology and Bioinformatics at Duke University.

The parametric approach estimated growth parameters by generating the best fit from the data to the logistic equation:

$$\frac{Ky_0 \exp(rt)}{K + y_0(\exp(rt) - 1)},$$

Where K is the carrying capacity (i.e. the value the data settles down to, or the steady state),

IC is the initial estimated OD after curve fitting of the data,

r is the exponential growth rate.

The Matlab code makes an initial estimate for these parameters and calculates the error to be the difference between the data and the resulting curve (using the square root of the square of the distance between the data and the curve, which eliminates negative values). The code repeats this process many times until it finds a set of parameters that minimise this error. The parameter values for growth rate and carrying capacity as calculated by the code are then visualised as a heat map. The code also generates curve fits for the data, which are displayed with error bars showing standard error of the mean.

The Baranyi method was based on the code as described above, except an additional fourth parameter λ for lag phase before the onset of exponential phase. This is particularly useful where the growth is not logistic, or where the bacteria have taken a long time to enter log phase.

$$y(t) = \frac{Ky_0 \exp\big(r(t-\lambda)\big)\big(1 - \exp(-rt) + \exp\big(-r(t-\lambda)\big)\big)}{K + y_0 \exp\big(r(t-\lambda)\big) - y_0 \exp(-\lambda r)}$$

Phenom Gaussian regression analysis was done in its entirety by Peter Tonner. Briefly, raw growth data were pre-processed prior to input into the model. The first five time points were removed in order to minimise error as a result of poor accuracy in early time point readings by the CLARIOstar plate reader. Data were then log2 transformed and normalised. The data were then standardised to mean = 1 and standard deviation = 1 for input into the model. Standardised data were then converted back to the original mean and standard deviation but still log2 transformed for plotting.

Chapter Three:

RNA sequencing analysis

Chapter 3 – RNA Sequencing Analysis

3.1 Analysis of the genetics of *E. coli*

The development of high-throughput sequencing technologies has allowed the development of novel methods for studying the genetic processes of organisms in a multitude of environments and conditions. In the past two decades, microarrays have been a useful tool in studying the changes in gene expression of an organism under specific conditions (Fodor, 1997). There have been a number of studies to date focussing specifically on gene expression of *E. coli* under acid stress, both organic and inorganic (Huang et al., 2007; King et al., 2010; Lambert et al., 1997; Nishino et al., 2003; Sandoval et al., 2011; Stincone et al., 2011; Tucker et al., 2002).

The methods being utilised in this thesis are TraDIS and RNAseq. This chapter will focus on the data obtained from RNAseq and the analysis of that data. RNAseq is a digital process that analyses the transcriptome by recording frequencies and alterations of transcripts. There are a number of advantages to using RNAseq over microarrays. It is a more sensitive tool and therefore more likely to identify low-abundance reads. Also, unlike microarrays, it does not require pre-existing knowledge of the target sequence, which means unidentified genes will not be missed. The digital detection of RNAseq is more sensitive, so there is less room for error. However, RNAseq is very expensive. Therefore, there is ongoing discussion regarding the number of replicates required in order to determine statistical significance. There is currently no consensus on this point (Lei et al., 2015).

In this study, RNAseq data was generated for 12 different conditions. These conditions were bacterial growth at pH 7 or pH 5.5, aerobic or anaerobic growth conditions in the presence of either acetic, propionic or butyric acid. For each of these conditions two lists of genes were generated, one of significantly upregulated genes (counted as those with an increased log2 fold change of 1 and a p-value of 0.05) and significantly downregulated genes (a decreased log fold change of 1 and p-value of 0.05 or

less). These lists were compared against one another and those genes which were common between conditions and unique to each condition were examined in more detail. This was done mainly by clustering of genes into pathways.

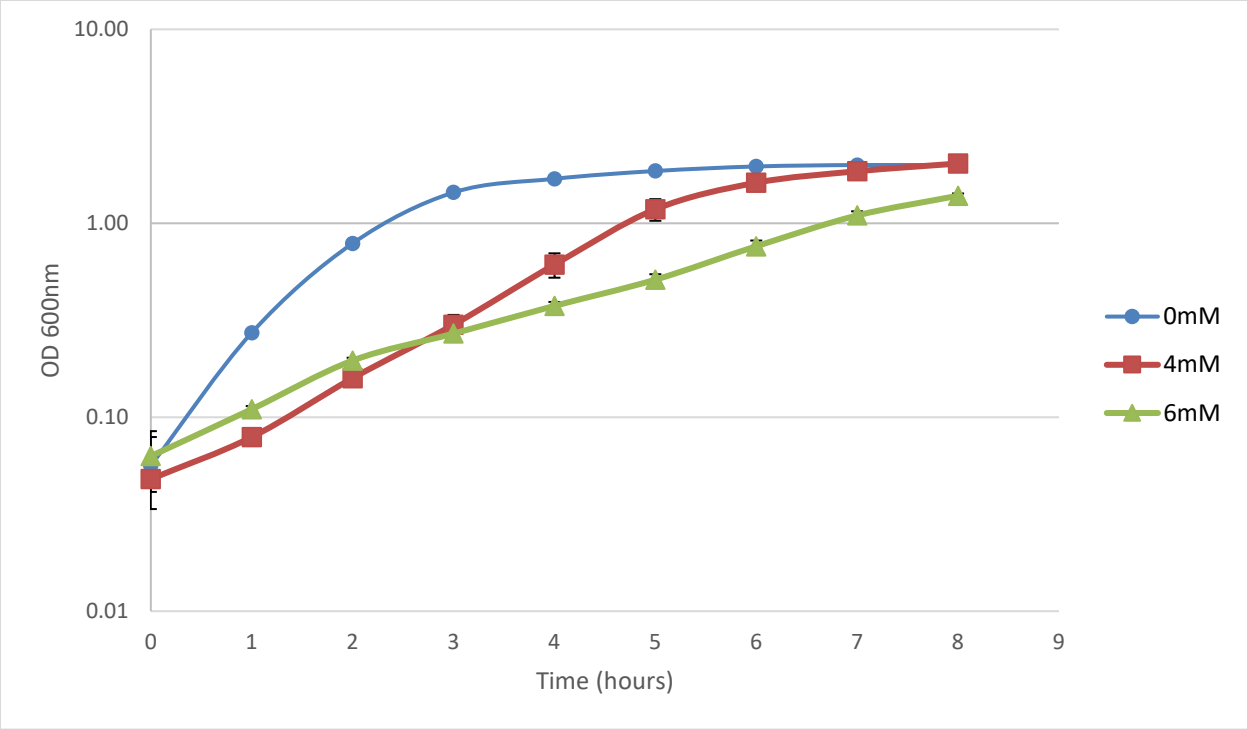
This chapter will first examine the conditions which were chosen for the subsequent sequencing (which were used for both RNAseq and TraDIS which will be discussed in the next chapter). It will then go on to look at each condition and the respective up- and down-regulated gene expression and pathway clustering. The chapter will conclude with some exploration into the ST131-specific genes of unknown function which appear in the respective lists.

3.2 Choosing conditions for downstream sequencing for gene expression and TraDIS experiments

The first step was to determine the best conditions for this analysis on the basis of a certain set of criteria. The analysis was part of a larger project being led by Dr Thippesh Sannasiddappa, the main aim of which was to build a stress response network of *E. coli* to a large number of stresses. That part of the project is ongoing, however the criteria for the stress response was that the bacteria were able to grow well (for the practical reasons of having enough cells to collect for DNA and RNA collection and analysis), but more importantly to mimic as closely as possible the conditions which the bacteria would encounter within the gut. The pH levels within the gut are neutral to mildly acidic, depending on the location within the intestines (not counting the very acidic stomach). On that basis, the pH of 5.5 was chosen for these experiments in order to replicate an environment that would be encountered within a host organism. Second, although the combination and concentration of acid within the gut fluctuates, again depending on location, bacteria live and reproduce there. Therefore, concentrations of organic acid were chosen that would allow the bacteria to grow at a reduced rate than a non-stressed control, but not such stringent conditions that there was no growth. Third, we also wanted to examine the genetic impact of organic acids alone, without the additional combinatorial effect of low pH, and therefore it was decided

that we should study the effect of a higher concentration of organic acid at pH 7. We also selected conditions whereby there would be the same impact on growth i.e. the growth impact at pH 5.5 with organic acids would be as close as possible to that seen at pH 7 with organic acids because we did not want big differences in overall growth rate between pH conditions and between organic acids so that the results were as comparable as possible.

Below at figures 3.1 to 3.3 is a series of growth curves showing potential conditions tested for the RNAseq and TraDIS experiments. In each case the organic acid was added to media at various concentrations with a pH of either 7 or 5.5 and the OD600 measured over the course of eight hours. Final conditions were selected as per the criteria laid out above.



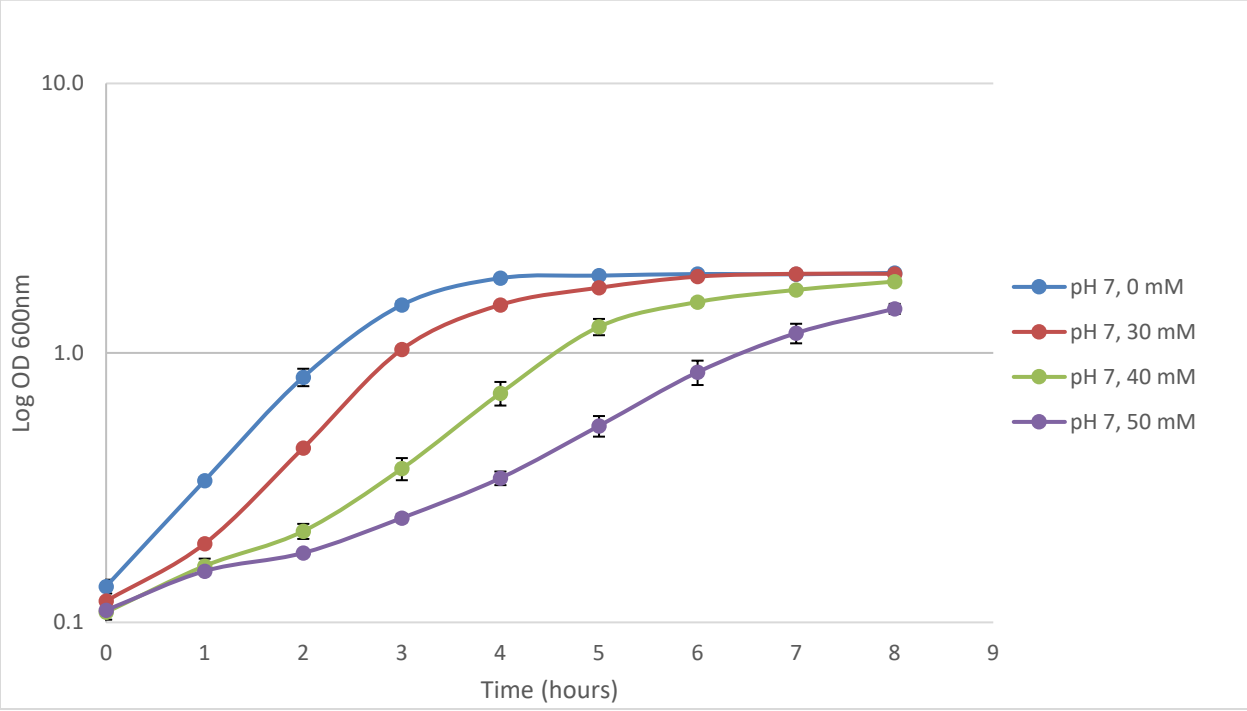
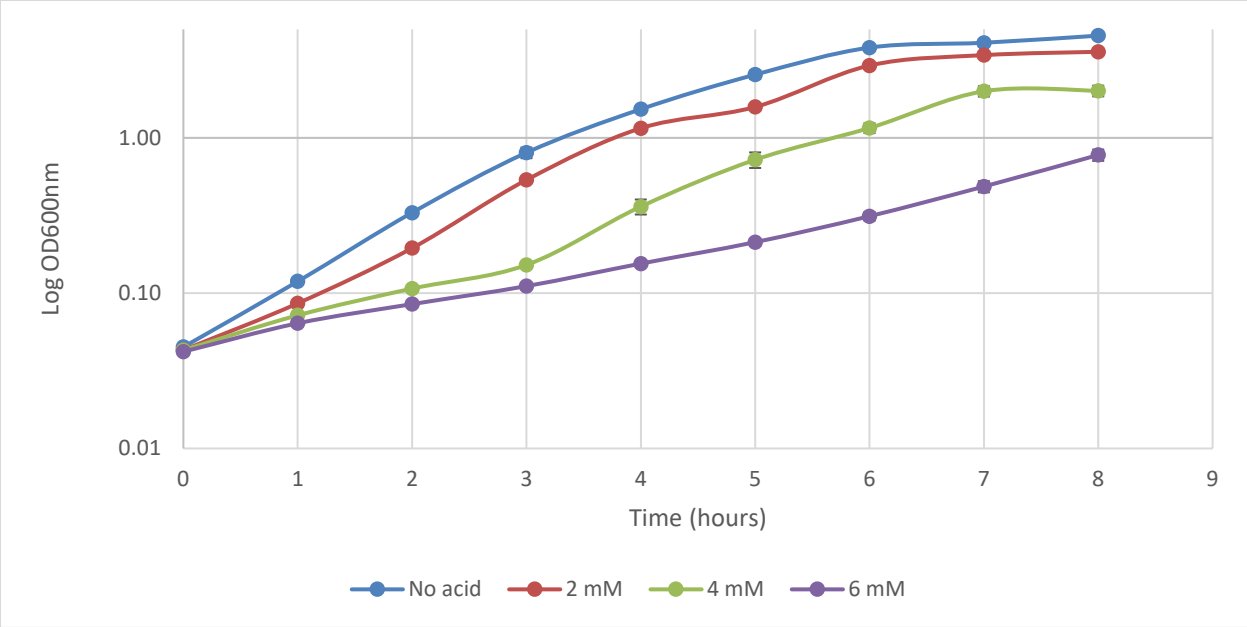


Figure 3.1 Growth curves with added acetic acid. Top: pH 5.5 with either no added acetic acid (control), 4 mM and 6 mM. Bottom: pH 7 with no added acetic acid (control), 30 mM, 40 mM and 50 mM. Error bars shows standard error of the mean (SEM) of three biological replicates.



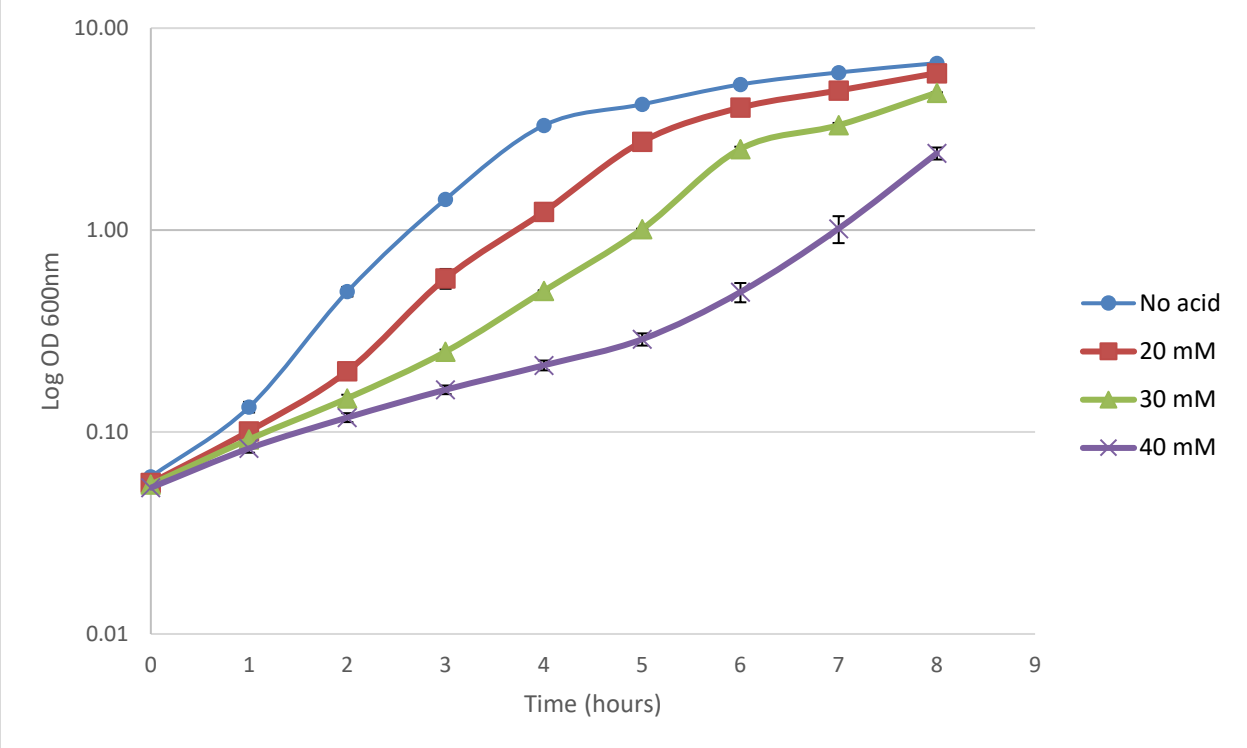


Figure 3.2 Growth curves with added propionic acid. Top: pH 5.5 with either no added propionic acid (control), 2 mM, 4 mM and 6 mM. Bottom: pH 7 with no added propionic acid (control), 30 mM, 40 mM and 50 mM. Error bars shows standard error of the mean (SEM) of three biological replicates.

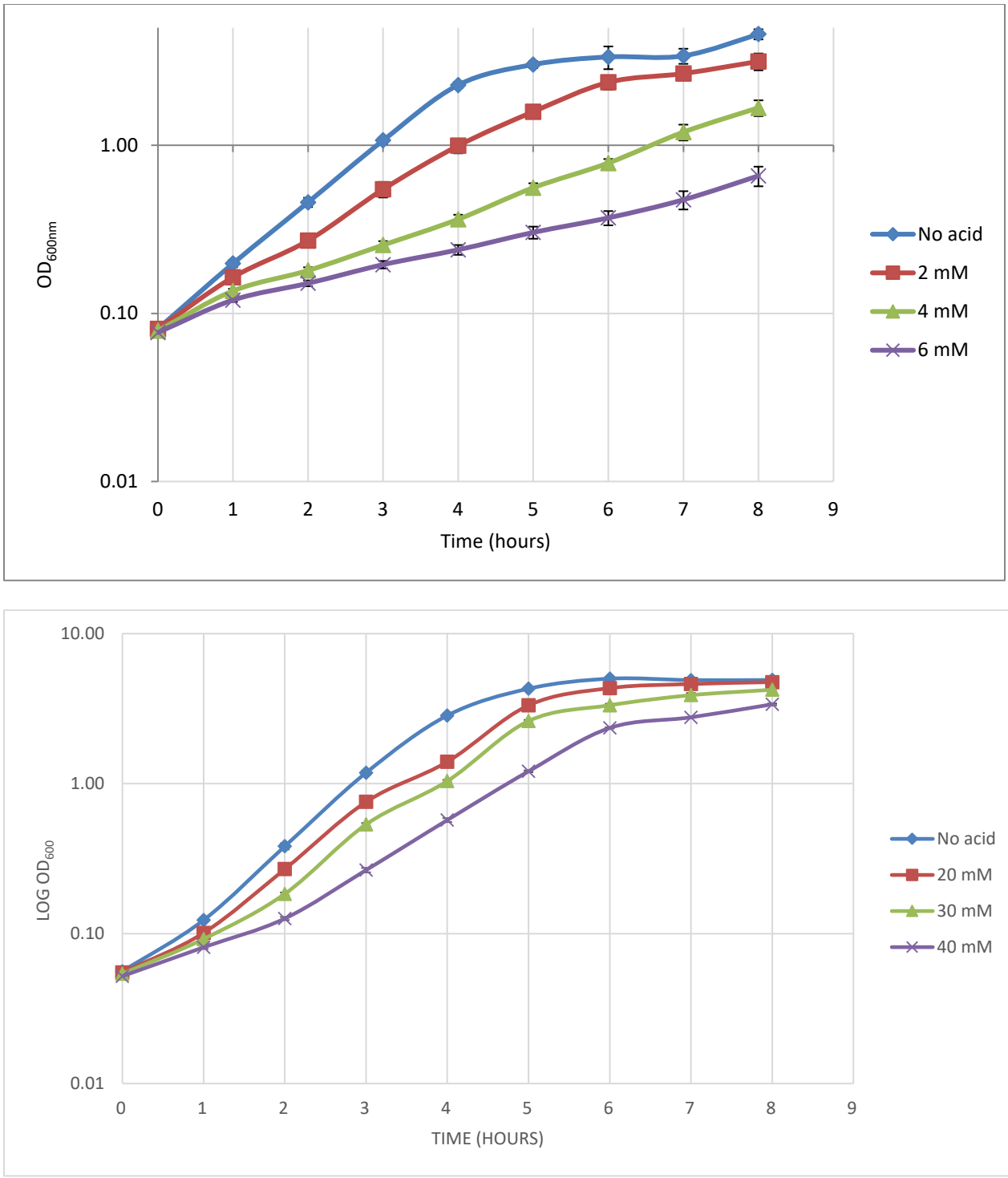


Figure 3.3 Growth curves with added butyric acid. Top: pH 5.5 with either no added butyric acid (control), 2 mM, 4 mM and 6 mM. Bottom: pH 7 with no added butyric acid (control), 30 mM, 40 mM and 50 mM. Error bars shows standard error of the mean (SEM) of three biological replicates.

In order to choose conditions, there were a number of factors that needed to be considered. In consultation with Dr Sannasiddappa, we chose conditions whereby the cells were mildly inhibited, but where final optical density ultimately reached that of the control condition (without any added organic acid). The focus of this study is mild stress, not harsh or extreme stress conditions, and so conditions which ultimately provided the cells with the ability to reach the same final optical density as the non-stressed control was a primary motivator. Second, the RNAseq and TraDIS experiments were repeated in anaerobic conditions. This added a further stress on growth of the bacteria, and so the addition of an anaerobic growth environment made it difficult for the populations to grow. For RNA pellet collection, the culture needed to reach OD600 of 1.8 to 2 before the media replenishment step (Chapter Two). In anaerobic conditions, the cultures grown at pH 7 with 40 mM acetic acid took 6.5 hours to reach OD600 of 1.75. The graphs in figures 3.2 and 3.3 show that butyric acid in particular, and propionic acid to a lesser extent, had a greater impact on growth than the addition of acetic acid. On that basis, and given the need as outlined above for a good level of overall growth, lower concentration of butyric and propionic acid than acetic acid were used in the final experiments. The final conditions chosen were: pH 7: 40 mM acetic acid; 30 mM propionic acid and 30 mM butyric acid. At pH 5.5: 4 mM acetic acid; 2 mM propionic acid and 2 mM butyric acid. Cells were grown both aerobically and anaerobically as described in Chapter Two. DNA and RNA was collected, purified and sequenced as described in Chapter Two.

3.3 RNAseq Data Analysis

This section will focus on the key pathways and, in some cases, the levels of gene expression, which were up-regulated and down-regulated in the RNA sequencing data. This section will start with an overview of the overlap between the three acids across the conditions, and will then focus on acetic acid in particular and the pathways which are specific to acetic acid, though for the sake of brevity, this analysis will be less comprehensive. In order to complete the analysis, gene lists have been analysed using the online tool DAVID 6.8: Functional Annotation Tool (david.ncifcrf.gov, Huang et al., 2009a,

2009b). This tool utilises the genome of several strains of *E. coli,* however none of them are the same strain as used in this study. For the purposes of this evaluation, the ST131 genes of unknown function have been removed from the analysis and will be dealt with separately later on in the chapter. The remaining genes have been analysed using the K-12 MG1655 genome for reference. A full list of differentially expressed genes in every condition can be found in the supplementary data provided.

3.3.1 Genes upregulated when cells were grown aerobically at pH 7 plus organic acid

The RNA sequencing data analysis was partly completed by Dr John Herbert from University of Liverpool (outlined in Chapter Two under 'data processing').

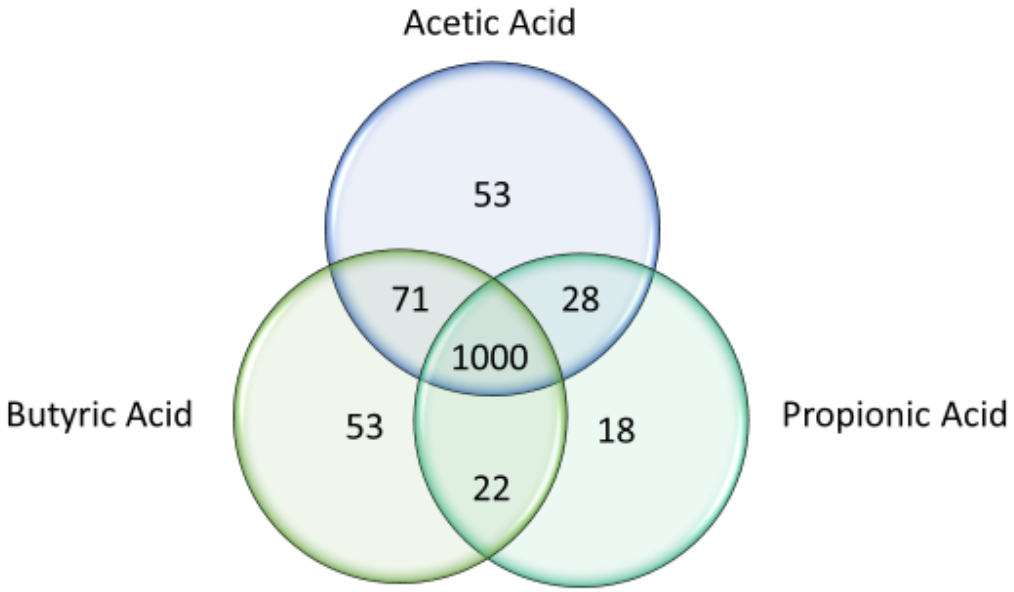


Figure 3.4 Number of genes significantly upregulated at pH 7 with each organic acid, grown aerobically.

Of the combined 1245 significantly upregulated genes across all acids at pH 7 in aerobic growth conditions, 1000 of those were shared amongst all three acids as shown above in Figure 3.4. Below at Figure 3.5 is a table showing Kegg Pathways and enrichment scores for the pathways to which these genes belong, determined via clustering. Assessing which biological pathways are being utilised or

disused by a population from RNA sequencing data is typically the first step in analysing long gene lists. Simply determining the function of each individual gene is overwhelming in the scale of the task and ultimately unhelpful in understanding fully the nature of any biological changes happening within an organism. Therefore, this thesis will focus on analysis using significantly enriched Kegg pathways as a means for analysing the RNAseq data, rather than looking at long gene lists. The term 'enrichment' applies to the amount to which a set of genes is over-represented compared to the background genome. For example, if genes regulating porins make up 1% of the genome, but are represented in 20% of the list, then by using simple statistical tests it is possible to determine how significant, or not, that biological process is in comparison to the control condition (Tipney and Hunter, 2010).

The table below gives a better understanding of the biological processes which are being utilised by the bacteria under organic acid stress. These pathways are enriched compared to a control condition of pH 7 with no added organic acid and a p-value has been calculated to determine the probability of a set of genes which belong to a single pathway having been represented purely by chance. The cut off chosen for the p-value for the clustering is 0.05, as this is the most widely used accepted cut-off.

As the pH of the media remained constant at pH 7 in all cases, it is unlikely that this response is a simple acid stress response, as it is not in response to a change in external pH. It is of course possible that the pathways being upregulated are in response to changes in internal pH after organic acids have diffused across the membrane and dissociated (as explained in detail in Chapter One), as opposed to a specific organic acid response (i.e. a response to the presence of protons and/or anions). However, at pH 7 the percentage of undissociated propionic acid is only 0.74%, for butyric acid it is 0.66% and for acetic acid it is 0.57%. At pH 5.5 the percentage of undissociated propionic acid increases to 18.99%, for butyric acid it is 17.28% and for acetic acid it is 15.4%. This shows that there is more to consider here than cytoplasmic pH stress.

Later on in this chapter there will be closer examination of the pathways which have altered regulation in the presence of organic acids at pH 5.5, which is compared against a control of pH 5.5 alone. This may give a clearer indication of those pathways which are organic acid specific, rather than a generic acid response.

Table 3.1 Kegg Pathways upregulated at pH 7, shared across all three organic acids.

KEGG TERM	P-VALUE	ENRICHMENT SCORE
FLAGELLAR ASSEMBLY	2.8E-16	4.2
PHOSPHOTRANSFERASE SYSTEM (PTS)	8.1E-10	3.2
FRUCTOSE AND MANNOSE METABOLISM	1.7E-07	3
PENTOSE AND GLUCURONATE INTERCONVERSIONS	1.3E-06	3.3
GALACTOSE METABOLISM	6.9E-06	2.8
PROPANOATE METABOLISM	0.000033	2.7
BACTERIAL CHEMOTAXIS	0.0027	2.7
GLYOXYLATE AND DICARBOXYLATE METABOLISM	0.004	2
BENZOATE DEGRADATION	0.0041	3.3
LYSINE DEGRADATION	0.0041	3.3
BACTERIAL SECRETION SYSTEM	0.0098	2.1
PHOSPHONATE AND PHOSPHINATE METABOLISM	0.011	3.7
VALINE, LEUCINE AND ISOLEUCINE DEGRADATION	0.012	3.1
FATTY ACID DEGRADATION	0.019	2.6
CITRATE CYCLE (TCA CYCLE)	0.024	2.1
MICROBIAL METABOLISM IN DIVERSE ENVIRONMENTS	0.041	1.2
TWO-COMPONENT SYSTEM	0.046	1.3

Table 3.1 above shows that the most up-regulated pathway in all three organic acids was flagellar assembly followed by the phosphotransferase system (PTS). Bacterial chemotaxis is also a significantly upregulated pathway. Flagellar assembly and bacterial motility is well established as being differentially regulated during acid response and flagellar assembly is a known virulence factor allowing the bacteria to invade the host and seek out favourable environments (Lackraj et al., 2016; Mellies and Lorenzen,

2014). The significant upregulation of both flagellar assembly and bacterial chemotaxis, as well as secretion systems, is a strong indication that virulence factors are being induced by the bacteria in the presence of organic acid. The upregulation of virulence factors could possibly be as a result of being exposed to conditions which would be associated with the human gut.

The second-most upregulated pathway on this list is the phosphotransferase system (PTS). The PTS is a major mechanism in *E. coli* for the uptake of sugars. The PTS is made up of a family of proteins, of which some are sugar specific and some are non-specific. During the uptake of sugars such as glucose, fructose and mannose, the sugars are modified by phosphorylation. Each protein in the pathway is also phosphorylated in a cascading fashion, which is energetically expensive. However, when for example the glucose molecule exits the PTS it has been phosphorylated to glucose-6-phospate, the first step in the glucose metabolic pathway (Madigan et al., 2009). The next most upregulated pathway in the list is fructose and mannose metabolism, which is closely tied with the PTS pathway in that many of the genes clustered into this pathway, such as *fruA, srlA, srlB, manY* and *manZ* are components of the PTS, but are mannose or fructose specific (Charbit et al., 1996; Saier Jr., 2015). Again, the next pathway on the list, galactose metabolism, is tied into the previous two pathways, with *gatA* and *gatB* being PTS-specific components. Also in this set of enriched genes, the entire Aga operon is upregulated, containing genes such as *agaBCDV,* which are galactose-specific components of the PTS (Brinkkötter et al., 2000; Saier Jr., 2015). The upregulation of these pathways implies high energy usage by the cell under these conditions. However, given the upregulation of energy-expensive pathways such as flagellar assembly, this is perhaps not surprising. Sim et al., (2017) noted an increase in flagellar assembly during high growth rate, however they also note the previous studies have reported repression of flagellar assembly during glucose metabolism, indicating that motility is upregulated in the search for nutrients. Sim et al., (2017) however, report the opposite of this and note that regulation of flagellar assembly is likely to be layered.

The presence of organic acids has also caused upregulation of several pathways in the metabolism of carboxylates: propionate metabolism, benzoate metabolism, glyoxylate and dicarboxylate metabolism and fatty acid degradation, as well as the TCA cycle. The *ato* operon is included in the cluster for all these different pathways, as well as the two component system cluster, as the regulator for the AtoDAEB operon is the two component system AtoSC, a response regulator, which is also upregulated across all three organic acids. The *ato* operon is responsible for the catalysis of acetoacetate to acetate and acetyl CoA, the first step of the TCA cycle (discussed in detail in Chapter One) (Jenkins and Nunn, 1987; Kyriakidis and Tiligada, 2009; Pauli and Overath, 1972). The gene *lpd* is upregulated across all conditions and has been clustered into the carboxylate and dicarboxylate metabolism pathway, the valine, leucine and isoleucine degradation pathway and the TCA cycle cluster. *lpd* stands for lipoamide dehydrogenase and is part of the pyruvate dehydrogenase complex. During aerobic growth, *E. coli* converts pyruvate to acetyl CoA by oxidative decarboxylation, which then enters the TCA cycle. Also within these clusters are the *fadABDEHIK* genes responsible for fatty acid degradation and *sucADC* and *sdhCD* which are involved in the conversion of succinate to succinyl CoA for use in the TCA cycle. The TCA cycle is directly involved in acetate metabolism, which could reduce acetate stress if upregulated. However, in the case of the other two organic acids, it may be that the upregulation of the TCA cycle is for increased energy usage of those pathways being utilised to remove H^+ and anions from the cell.

Within the cluster designated as Pentose and Glucuronate Interconversions is the *araBAD* operon. This operon is responsible for degradation of L-arabinose to form D-xylulose-5-phosphate which then enters the pentose phosphate pathway. The pentose-phosphate pathway again involves metabolism of acetate and therefore upregulation of this pathway may aid in the stress response of the cell. In the list of upregulated genes, but not included in the clustering are the *araFGH* genes, responsible for transport of L-arabinose. Both operons share the same regulator *araC*, which is significantly downregulated in this condition.

The Kegg Pathway called 'Two Component Systems' refers to all two component systems within the cell which respond to changes in the environment. The genes from the list specifically included in this cluster are *cheAB* involved in chemotaxis. *kdpAB* are involved in import of potassium ions which is outlined in Chapter One aids in maintaining cell homeostasis during acid challenge. Also included is c*itAB* which autophosphorylates in response to the presence of citrate and activates the other *cit* genes inducing citrate fermentation, again aiding in central metabolism, all of which are upregulated under all three conditions (Yamamoto et al., 2009).

Of the 748 genes that were not clustered, there are several genes and operons which are found on the list that might be anticipated under these conditions, such as *gadW, cadB, cadC* and *safA* which are all known to be involved in the acid stress response (see Chapter One for details). Also on the list of genes which were not considered to be significantly clustered were *csgA-F* involved in expression of curli, a fundamental part of biofilm formation, and *cyoB, cyoC* and *cyoD* which are subunits of the cytochrome *bo* terminal oxidase complex, again discussed in Chapter One, subsections 1.10 and 1.11 which outline the role of organic acids in disrupting the proton motive force. Several genes associated with multidrug efflux pumps were also significantly upregulated.

Included not just in this list of non-clustered genes but in the non-clustered upregulated list for pH 5.5 are the genes *eutAHJMNPQST*. In the list of genes shared between propionic and butyric acid is *eutK,* between acetic and propionic is *eutB* and in the list for acetate alone is *eutL* and *eutD*. The Eut operon is involved in the utilisation of ethanolamine, which is converted to acetaldehyde and then to acetyl CoA. EutD catalyses the same reaction as phosphate transferase *pta*, a reversible reaction in the metabolism of acetate between acetylphosphate and acetyl CoA which results in either utilisation or production of acetate. Pta is downregulated in all three acids at pH 5.5 and pH 7 in aerobic conditions. Under

anaerobic conditions Pta is significantly upregulated at pH 7 in acetic acid and propionic acid and at pH 5.5 in butyric acid (Bologna et al., 2010). This shows a complicated response from the cell between reducing acetate production whilst simultaneously producing acetyl-CoA for use in metabolic pathways such as the TCA cycle.

Two pathways were considered significantly upregulated in acetic and propionic acid only: alanine, aspartate and glutamate metabolism ($p = 0.022$) and butanoate metabolism ($p = 0.028$). Of the 28 genes shared across the two conditions, 23 were not included in the clustering. Of the two pathways considered significantly clustered, included in the 'alanine, aspartate and glutamate metabolism pathway' were the genes *gabT* and *gabD*, which are involved in the degradation of GABA, which has been discussed in Chapter One in the context of its importance during acid challenge. The third gene clustered into this pathway is *putA* which is part of the proline degradation pathway which results in the production of glutamate, NADH and H^+.

Within the butanoate metabolism cluster are the genes *gabT* and *gabB* again, and also *sdhA* which is part of the same pathway as *sdhCD*, referred to above.

Of the 23 genes not clustered, heat shock protein *ibpA* was upregulated, which is involved in prevention of permanent denaturation of proteins during heat shock and oxidative stress. This is possibly a generic stress response, rather than a condition-specific response. Within the list of genes not clustered across all three acids was *ibpB*, part of the same process. This gene though, along with *ipbA,* was also unregulated at pH 5.5 with no added organic acid, and is therefore upregulated here as a generic stress response, rather than organic acid specific response.

There were 22 genes in common between butyric and propionic acid in this condition, not shared with acetic. Of the 22 genes, the pathway nitrotoluene degradation (p-value = 0.026) was considered to be significantly upregulated due to the presence of *hyaA* and *hyaB*, the small and large subunits of hydrogenase 1, responsible for oxygen tolerance and expression of the operon is induced in anaerobic conditions (Sawers and Boxer, 1986). On that basis it is difficult to hypothesise why the presence of propionate or butyrate might initiate or increase transcription of this operon.

There were 71 genes in common between acetic and butyric acid, however none of these clustered into significantly upregulated pathways. For acetic acid only (53 genes in total) again there were no uniquely upregulated pathways. However, included in this gene list is *gadA,* part of the glutamate-dependent acid response in *E. coli* (discussed in Chapter One).

The next section will look at whether the genes and pathways upregulated at pH 5.5 share any similarities to those outlined above.

3.3.2 Genes upregulated when cells were grown aerobically at pH 5.5 plus organic acid

This section will examine the pathways and genes upregulated at pH 5.5 with the addition of either acetic, butyric or propionic acid and combinations between them. The control condition against which expression of these genes has been compared is pH 5.5 with no added organic acid. Below at figure 3.5 is a Venn diagram showing numbers of significantly upregulated genes for each condition and for each combination of conditions. Of the 1250 genes upregulated at pH 5.5 with organic acid, 729 are shared between all three conditions. There is a much larger proportion of the genes upregulated in acetic acid which are not shared between the other acids, and the majority of the genes upregulated in butyric acid are shared with acetic. Conversely, there are few in common between propionic and butyric or

propionic and acetic, with the majority of genes upregulated in propionic acid also being upregulated in acetic and butyric and, therefore, shared between all three acids.

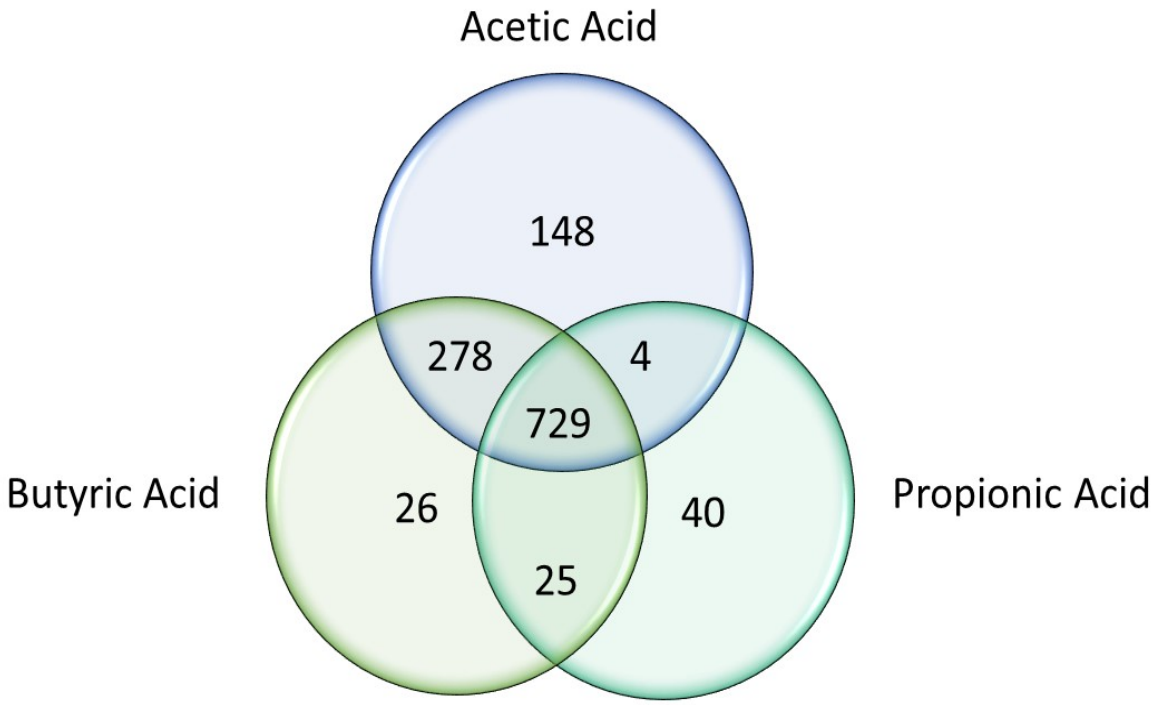


Figure 3.5 Venn diagram displaying number of significantly upregulated genes.

The highest number of overlapping genes occurs between all three acids (729 genes). This list (and the genes involved) is almost identical to that in the section above, clustered from genes upregulated at pH 7 during aerobic growth. Flagellar assembly is the most significantly upregulated pathway, with 3.1% of the gene list involved in this pathway (whilst comprising approximately 1% of the genome) (Chilcott and Hughes, 2000). Flagellar assembly is seen repeatedly in the lists produced in this study. Table 3.2 below shows the Kegg pathways, with associated p-value and enrichment scores, of the genes shared between acetic, propionic and butyric acid.

Table 3.2 Significantly upregulated Kegg pathways clustered from genes shared between acetic, butyric and propionic acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
FLAGELLAR ASSEMBLY	2.5E-10	4.2
PHOSPHOTRANSFERASE SYSTEM (PTS)	2.5E-08	3.5
FRUCTOSE AND MANNOSE METABOLISM	2.9E-08	3.6
GALACTOSE METABOLISM	0.000015	3.2
PENTOSE AND GLUCURONATE INTERCONVERSIONS	0.00026	3.1
BACTERIAL CHEMOTAXIS	0.00026	3.6
BACTERIAL SECRETION SYSTEM	0.0031	2.6
PROPANOATE METABOLISM	0.0039	2.4
PHOSPHONATE AND PHOSPHINATE METABOLISM	0.023	4.1
DEGRADATION OF AROMATIC COMPOUNDS	0.025	2.8
BENZOATE DEGRADATION	0.025	3.3
LYSINE DEGRADATION	0.025	3.3
GLYOXYLATE AND DICARBOXYLATE METABOLISM	0.038	1.9

In acetic acid only, there were 148 unique upregulated genes. Only one pathway was significantly clustered from that list, the TCA cycle (p-value 0.026). This was due to the presence of *sdhABC*, which are all upregulated at pH 7 and have been described above.

Of the 40 genes that were significantly upregulated in propionic acid alone, there was one significantly clustered pathway: microbial metabolism in diverse environments (p-value 0.024). Of the 26 genes uniquely upregulated in butyric acid, there were no significantly clustered pathways, however within genes for the condition was *rpoS*. As detailed in Chapter One section 1.9, *rpoS* is the key regulator in AR1, the RpoS-dependent acid response of *E. coli*. It is interesting that this should be upregulated in butyric acid alone.

There are 25 genes shared between propionic and butyric acid, but no significant pathways. However, within the list of shared genes is a large number of genes within the *arg* pathway encoding arginine biosynthesis (*argABCDEGHIJ)*. It may be an artefact of the sequencing that these genes have not also been measured as upregulated in acetic acid, or it may be that the particular stress exerted by propionic

and butyric acid is such that this pathway is more useful here than under acetic acid stress. This would require more detailed investigation.

There are considerable overlaps between the genes being upregulated at pH 7 and those being upregulated at pH 5.5, and subsequently the pathways involved. Later on in this chapter there will be an overview of the genes and pathways shared between the acetic acid conditions.

Of the 148 genes upregulated in acetic acid alone, these clustered into one significant pathway: the TCA cycle (0.02).

3.3.3 Genes upregulated when cells were grown aerobically at pH 5.5 only (versus pH 7)

In the analysis of genes upregulated at pH 5.5 only versus pH 7, there were only 38 significantly upregulated genes. Only one of those, *clpB,* was shared between all three acids and pH 5.5 only. The majority, 32 genes, were in the list for pH 5.5 only and not shared with any of the other acids. These included acid resistance genes such as *cadA* and *cadB*. One of the reasons these genes may not be in common between the conditions and the control is because the measure of significant upregulation in these conditions was in comparison to this pH 5.5 control condition, and therefore any upregulated genes in the analysis of genes upregulated at pH 5.5 with organic acids were those considered differentially regulated compared to the control. Therefore, these acid response genes were not additionally upregulated compared to the control.

When compared against the genes upregulated at pH 7 plus organic acid, which was compared against a control condition of pH 7 alone, there were seven genes in common, this time including the *cad* genes, implying a generalised acid response in these pH 7 conditions as well as an organic acid response.

However, twenty of the genes from the pH 5.5 control were not shared with the pH 7 plus organic acid conditions. This probably reflects the differing nature of these stress conditions.

3.3.4 Genes downregulated when cells were grown aerobically at pH 7 plus organic acid

This section and the one after it at 3.3.5 will focus on the genes downregulated across the two conditions examined above; pH 7 with organic acid and pH 5.5 with organic acid in aerobic conditions. Below at figure 3.6 is the Venn diagram for pH 7 with added organic acid demonstrating the number of genes downregulated under this condition and how many of those genes were shared across conditions, or otherwise. In total 1265 genes were significantly downregulated (compared to 1245 genes upregulated under the same conditions). What is noticeable is the more even distribution of genes across the diagram, where less than half are shared among all three, compared to 1000 of the 1245 upregulated genes shared across the three conditions.

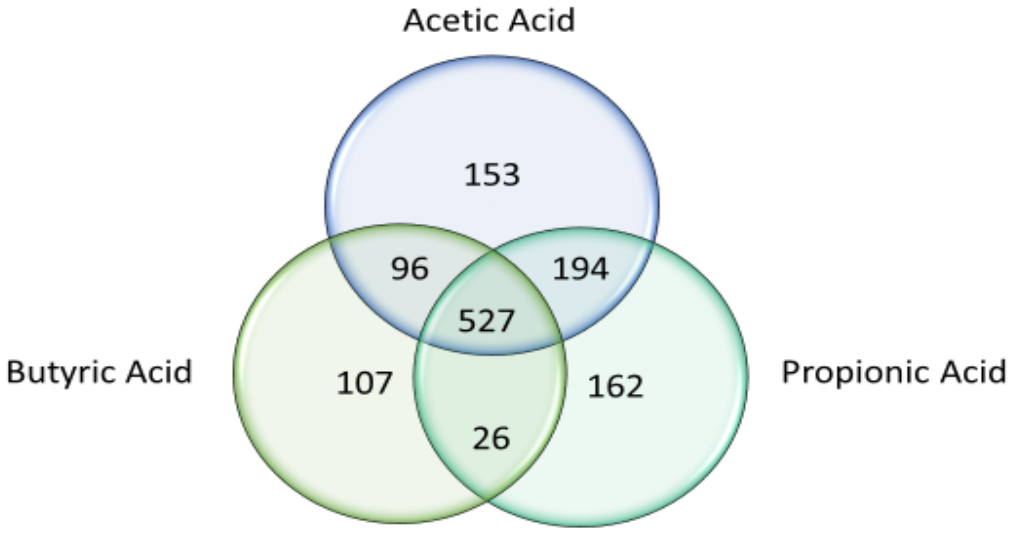


Figure 3.6 Venn diagram displaying number of significantly downregulated genes.

Table 3.3 below shows the Kegg pathways most significantly downregulated across the three acids at pH 7. The table was generated from a list of 527 genes, and 317 of those genes were not included in the clustering. Of the remaining 210 genes, the following pathways were clustered:

Table 3.3 Kegg pathways clustered from significantly downregulated genes in all three acids

KEGG TERM	P-VALUE	ENRICHMENT SCORE
RIBOSOME	1.7E-19	3.5
PURINE METABOLISM	0.00032	1.9
FATTY ACID BIOSYNTHESIS	0.0038	3.4
SULFUR METABOLISM	0.0093	2.2
PYRIMIDINE METABOLISM	0.012	1.7
RNA POLYMERASE	0.02	5.5
BIOSYNTHESIS OF ANTIBIOTICS	0.036	1.3
PROTEIN EXPORT	0.041	2.3

The most downregulated pathway, with a p-value of $1.7x10^{-19}$ is the Kegg pathway called 'Ribosome'. The Ribosome Kegg pathway incorporates 78 ribosomal subunit proteins, of which 50 were included in this list as downregulated across the organic acid stress conditions. It is possible that under stress, the cells are downregulating synthesis of all non-essential proteins synthesis and this has resulted in wholesale downregulation of the ribosomal pathway.

The fatty acid biosynthesis pathway is listed due to the cluster of genes *fabADGHI* and *accACD*. In this list of genes not clustered is *fabR*, the regulator of the *fab* genes.

In this list of non-clustered genes was *atpABCDEFGHI* which encode the F0 and F1 subunits of ATP synthase involved in transport of H^+ into the cell. This corroborates the literature reviewed in chapter one noting the downregulation of this system during acid challenge (Maurer et al., 2005).

The genes *cysABCDHIJMNPSUM* are all significantly downregulated across the three acids. These genes encode sulphur uptake and metabolism. Significant downregulation may be in order to conserve energy. In all three lists for each organic acid alone there were no significantly clustered pathways. However, in the gene list for acetic acid only as significantly downregulated was *nuoAEFHIJM.* The *nuo* genes encode NADH dehydrogenase 1, described in Chapter One, section 1.11 as contributing to the maintenance of the proton motive force. That these genes should be downregulated is interesting as the literature outlined in chapter one indicates upregulation of this pathway to remove protons from the cell. The reasons for its downregulation in the presence of organic acid would need further investigation.

In the genes shared between the significantly downregulated genes in acetic and propionic acid, there were five pathways, shown below at table 3.4. Beneath that at table 3.5 are the pathways shared between acetic and butyric acid. There were no significant pathways in the list of genes shared between propionic and butyric acid.

Table 3.4 Table of Kegg pathways from significantly downregulated genes shared between acetic and propionic acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
METABOLIC PATHWAYS	0.000011	1.5
BETA-LACTAM RESISTANCE	0.0026	5.7
BIOSYNTHESIS OF SECONDARY METABOLITES	0.013	1.5
DNA REPLICATION	0.017	4.8
BIOSYNTHESIS OF ANTIBIOTICS	0.042	1.6

Table 3.5 Table of kegg pathways from genes significantly downregulated shared between acetic and butyric acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
ARGININE BIOSYNTHESIS	0.00000015	16.3
BIOSYNTHESIS OF AMINO ACIDS	0.0000049	4.3
BIOSYNTHESIS OF SECONDARY METABOLITES	0.00021	2.4
2-OXOCARBOXYLIC ACID METABOLISM	0.0043	7
BIOSYNTHESIS OF ANTIBIOTICS	0.01	2.3
METABOLIC PATHWAYS	0.015	1.4
THIAMINE METABOLISM	0.039	9.2
CYSTEINE AND METHIONINE METABOLISM	0.04	5.1

It is interesting that of the pathways shown in figure 3.5 (acetic and butyric acid) arginine biosynthesis is the most significantly clustered. Within the list of genes is *argABCDEGHI* which are significantly upregulated at pH 5.5. This suggests that the cells are relying on different acid resistance pathways at pH 5.5 with organic acid and pH 7 with organic acid. More investigation would need to be done to determine why that might be and whether the external pH at 5.5 is initiating the arginine response which is not being replicated when the external pH is neutral.

3.3.5 Genes downregulated when cells were grown aerobically at pH 5.5 plus organic acid

This section will focus on those genes downregulated at pH 5.5 in the presence of organic acids when grown aerobically. At figure 3.7 below, the Venn diagram shows that, like at pH 7, there are fewer genes shared between acids than in aerobic conditions, with a more even distribution.

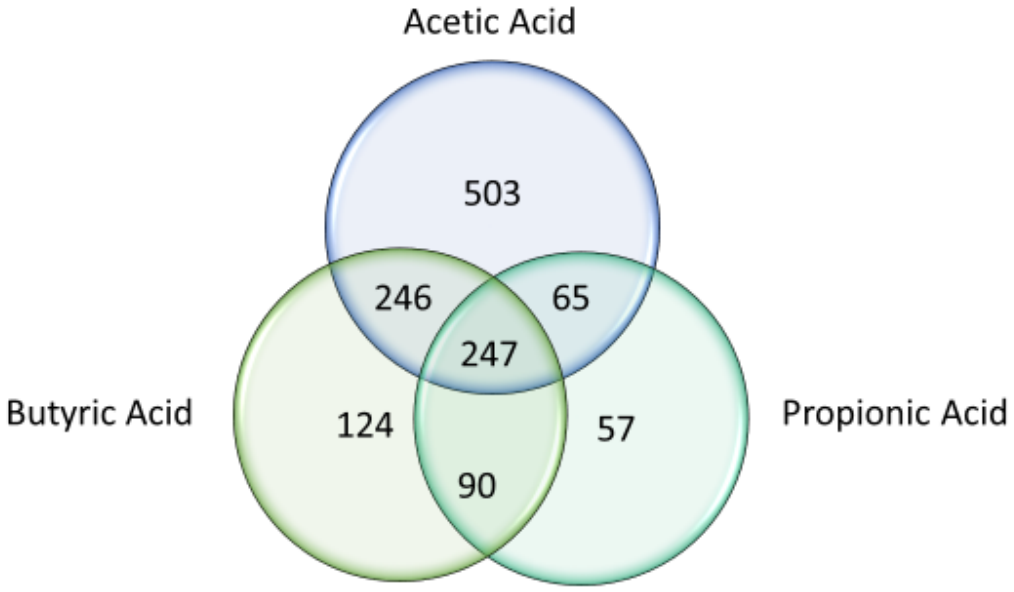


Figure 3.7 Venn diagram showing numbers of genes significantly downregulated in each acid and shared across acids.

Table 3.6 Table showing Kegg pathways from genes significantly downregulated in all three acids.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
RIBOSOME	1.4E-08	3.7
PROTEIN EXPORT	0.0031	4.4
PURINE METABOLISM	0.0038	2.2
OXIDATIVE PHOSPHORYLATION	0.016	2.6
BACTERIAL SECRETION SYSTEM	0.032	2.8
BIOSYNTHESIS OF ANTIBIOTICS	0.042	1.5

As with pH 7 and pH 5.5 upregulated, many of the pathways here are shared between the pH conditions. Included here but not in the list for pH 7 is oxidative phosphorylation due to the presence of genes in the *nuo* operon. As described above, it is interesting that these should be significantly downregulated. At pH 7 the genes were downregulated in acetic acid alone, however at pH 5.5 they are

downregulated across all three acids tested. This pathway also clusters *atpABEH*, also described previously as being downregulated at pH 7.

For the genes shared across acetic and propionic acid there were no significant clusters. For acetic and butyric acid there were again pathways previously seen in other groups such as the Ribosome pathway, as well as homologous recombination and mismatch repair. Again, this may be due to the cells downregulating pathways which are energetically expensive when under acid challenge.

Table 3.7 Kegg pathways from genes significantly downregulated in acetic and butyric acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
PURINE METABOLISM	0.0022	2.4
RIBOSOME	0.0023	2.5
METABOLIC PATHWAYS	0.0041	1.3
BIOSYNTHESIS OF SECONDARY METABOLITES	0.0087	1.5
HOMOLOGOUS RECOMBINATION	0.0093	3.6
MISMATCH REPAIR	0.016	3.8
GLYCOLYSIS / GLUCONEOGENESIS	0.019	2.8
DNA REPLICATION	0.027	4.1
CARBON METABOLISM	0.035	1.8
METHANE METABOLISM	0.037	3.1

Table 3.8 Kegg pathways from genes downregulated in acetic acid only.

KEGG PATHWAYS	P-VALUES	ENRICHMENT SCORES
METABOLIC PATHWAYS	0.000016	1.3
BIOSYNTHESIS OF ANTIBIOTICS	0.00049	1.6
BIOSYNTHESIS OF SECONDARY METABOLITES	0.0031	1.4
THIAMINE METABOLISM	0.0079	3.5
PENTOSE PHOSPHATE PATHWAY	0.019	2.2
MISMATCH REPAIR	0.021	2.4
GLYCOLYSIS / GLUCONEOGENESIS	0.026	1.9
CARBON METABOLISM	0.048	1.4
PYRIMIDINE METABOLISM	0.05	1.6

In acetic acid only there were again several pathways already noted, shown above in table 3.8. For the genes shared between propionic and butyric acid, there were no significantly clustered pathways.

3.3.6 Genes downregulated when cells were grown aerobically at pH 5.5 versus pH 7

There were only 13 significantly downregulated genes at pH 5.5 when compared against the control of pH 7. Six of those genes were *argABCDGH* which, as explained above, are involved in arginine biosynthesis. Three of the other seven were *hisJMQ* which encodes an ATP-dependent histidine transporter. The small number of genes here has been seen before in RNAseq experiments which measure differential expression at pH 5.5 versus pH 7. Zorraquino et al., (2016) measured only 21 differentially expressed genes at pH 5.5 compared to the control.

3.3.7 Genes upregulated when cells were grown anaerobically at pH 7 plus organic acid

The next sections in this chapter will look at the same conditions in anaerobic growth to determine whether the pathways being utilised by the cell are the same as under aerobic organic acid challenge and, if not, how they are different. Below is the Venn diagram showing the number of genes upregulated in each acid at pH 7 in the presence of organic acids. There are fewer genes overall being upregulated under anaerobic conditions than aerobic conditions (889 here compared to 1245 aerobically). More genes are upregulated in butyric acid alone and a large proportion are shared across butyric with the other two acids than in those acids alone or with each other.

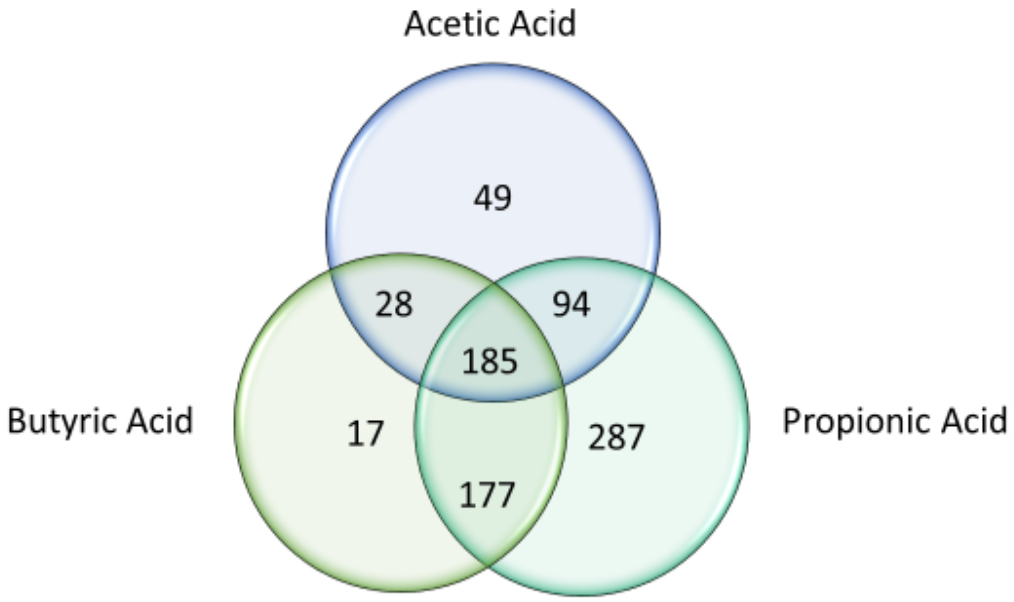


Figure 3.8 Venn diagram displaying number of significantly upregulated genes

Across the three acids there were many of the same genes and pathways seen previously, though not entirely the same. The elements of aerobic respiration that were upregulated at pH 7 in aerobic growth conditions are downregulated here due to the change to anaerobic respiration. Under the pathway of biosynthesis of antibiotics, there are genes involved in lots of the different metabolic pathways, for example *aroHL* which are steps of chorismate pathway leading to aromatic amino acid synthesis, *ispADEFH* which are genes in the MEP pathway, and the *pur* genes, involved in purine biosynthesis. These genes are in fact clustered under the Kegg pathways of antibiotic biosynthesis, secondary metabolite biosynthesis, metabolic pathways and purine metabolism. This is an example of how the software will generate multiple pathways from the same set of genes and consider them significant because the list itself is in fact quite small. The same genes are repeatedly clustered into almost all the pathways here, with *aroHL* (shared between all three acids and *aroABCG* appearing in other lists as downregulated here) involved in the chorismate pathway, clustered into almost every Kegg pathway.

Table 3.9 Kegg pathways clustered from significantly upregulated between all three acids.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
BIOSYNTHESIS OF ANTIBIOTICS	1.6E-13	3.3
BIOSYNTHESIS OF SECONDARY METABOLITES	7.3E-13	2.6
METABOLIC PATHWAYS	9.1E-06	1.5
PURINE METABOLISM	0.00031	2.8
BIOSYNTHESIS OF AMINO ACIDS	0.00038	2.4
GLYCOLYSIS / GLUCONEOGENESIS	0.00077	3.8
CARBON METABOLISM	0.0012	2.4
PENTOSE PHOSPHATE PATHWAY	0.0024	4
TERPENOID BACKBONE BIOSYNTHESIS	0.0078	5.8
METHANE METABOLISM	0.028	3.4

Between the conditions of propionic and butyric acid there were two significantly upregulated Kegg pathways, biotin metabolism (p-value = 0.014) and cationic antimicrobial peptide (CAMP) resistance (p-value = 0.043). The biotin metabolism pathway includes the genes *bioABF*. At pH 7 in aerobic growth, these genes (including *bioCD*) were all downregulated as well as biotin protein ligase *birA* which binds to and inhibits the *bio* operon in the presence of biotin. Research indicates that biotin sensing is a virulence factor in *E. coli* O156:H7 and implicated biotin availability in reduced incidence of colonisation and infection of the large intestines in mice (Yang et al., 2015). It is apparent that the combination of mildly acidic pH, organic acids and anaerobic conditions has stimulated the pathway in this study.

Below are the pathways for acetic and propionic acid (table 3.10) and beneath that acetic and butyric acid (table 3.11). These are pathways that have been discussed previously in this chapter and so will not be looked at in detail again.

Table 3.10 Kegg pathways clustered from genes shared across acetic and propionic acid

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
MICROBIAL METABOLISM IN DIVERSE ENVIRONMENTS	0.019	1.8
ALANINE, ASPARTATE AND GLUTAMATE METABOLISM	0.023	4.4
NITROTOLUENE DEGRADATION	0.024	11.6

Table 3.11 Kegg pathways clustered from genes shared across acetic and butyric acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
CITRATE CYCLE (TCA CYCLE)	0.0088	18.2
GLYCOLYSIS / GLUCONEOGENESIS	0.02	11.8
PYRUVATE METABOLISM	0.032	9.3

Table 3.11 above shows the pathways shared between acetic and butyric acid. The most upregulated pathway is TCA cycle, which is repressed under anaerobic conditions. However, the presence of *aceEF* involved in pyruvate decarboxylation to acetyl CoA and *lpd* (already described) in this list has resulted this pathway being considered significantly upregulated despite a short gene list of only 28 genes.

There were 287 genes in this condition which were differentially expressed in propionic acid only. This is more than shared between conditions and more than any other individual condition. This points to propionic acid having its own, separate effect on the cell under anaerobic conditions. Below at table 3.12 is the list of pathways

Table 3.12 Kegg pathways clustered from genes significantly upregulated in propionic acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
METABOLIC PATHWAYS	0.00066	1.3
FATTY ACID BIOSYNTHESIS	0.0024	5.7
HOMOLOGOUS RECOMBINATION	0.0043	3.6
MISMATCH REPAIR	0.0062	3.9
PROTEIN EXPORT	0.015	3.9
FATTY ACID METABOLISM	0.022	3.5

These pathways have been discussed already, however metabolic pathways here includes genes such as *menCEH* and *nadC* which will be discussed in more detail in the next chapter.

The next section will look at the upregulated genes at pH 5.5 in anaerobic conditions and compare these to those outlined above.

3.3.8 Genes upregulated when cells were grown anaerobically at pH 5.5 plus an organic acid

This section will examine the genes upregulated in organic acid stress at pH 5.5 in anaerobic conditions. Below is a Venn diagram showing the number of genes shared between the three conditions and unique to each. What is surprising is the very low number significantly upregulated in propionic acid compared to the other two acids, and subsequently the low number shared between propionic and the other acids.

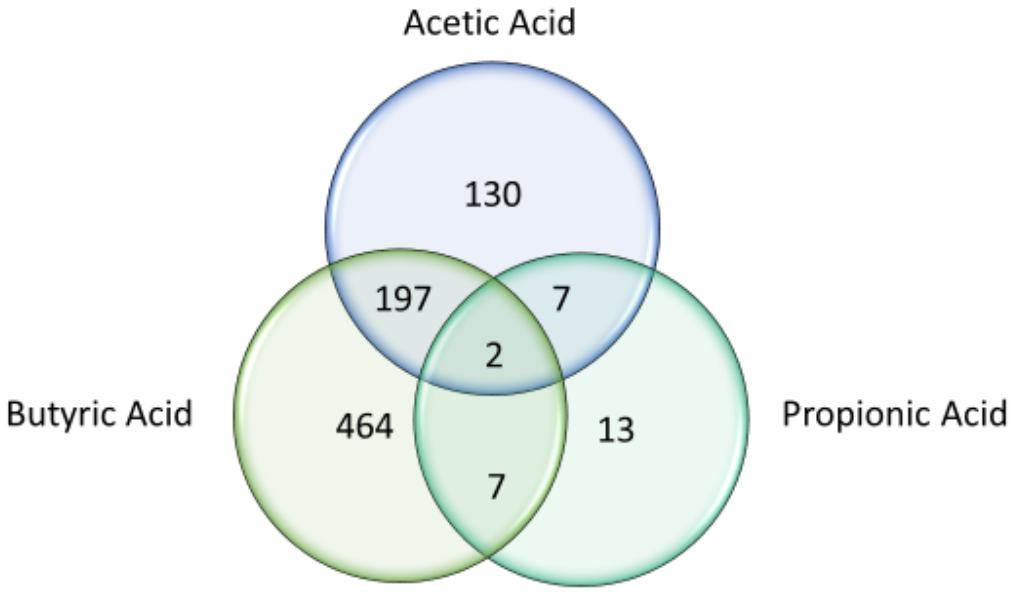


Figure 3.9 Venn diagram showing numbers of genes upregulated for each condition.

There are only two genes in common between all three acids, *metK* and *oppA*. The gene *metK* is a methionine adenosyltransferase that catalyses the formation of S-adenosylmethionine (SAM), which in turn is responsible for a large range of metabolic processes including the methylation of DNA and RNA. It has been shown to be essential (El-Hajj et al., 2013). In the list of genes shared between acetic and propionic acid in this condition is *metBFJR* which are also involved in methionine biosynthesis. This implies that the entire pathway is unregulated in this condition however it is possible that these genes were not considered upregulated in butyric acid due to an error or artefact of the sequencing.

The other gene shared between all three acids is *oppA*, which is a periplasmic binding protein of an oligopeptide ABC transporter. Studies have indicated that *oppA* knockouts are more prone to heat shock and suggest that OppA is involved in repairing and/or preventing protein denaturation during heat challenge (Krisko et al., 2014). OppA expression is regulated by the *bgl* operon and research indicates that mutants with transcriptionally active *bgl* operon show growth advantage in stationary phase (GASP). The same study showed that *bgl* regulates *oppA* via *gcvAB* (Harwani et al., 2012). The *bgl* operon forms part of the PTS.

In the RNAseq data presented here, *oppA* was significantly upregulated in this condition, but also *oppABCD* were all upregulated at pH 7 in anaerobic growth. At both pH 7 and pH 5.5 in aerobic conditions, *bglBFGHJ* was upregulated across all three acids. At pH 5.5 in aerobic growth, *gcvA* is significantly downregulated in the presence of all three acids and, in the same condition, *oppACDF* are all significantly downregulated in acetic and propionic acid. This implies that though in aerobic conditions the cells are relying heavily on the PTS (demonstrated through upregulation here of the *bgl* operon), the cells are not utilising the OppA transporter system in aerobic conditions, but only in anaerobic conditions.

In the presence of acetic acid only one pathway was significantly clustered: DNA recombination (p-value 0.044). This is due to the presence of *dnaT, dnaE* and *recA* which is involved in DNA recombination and repair included in the list. This may indicate that the cell is prioritising DNA repair in this condition.

The pathways upregulated across acetic and butyric acid are shown in table 3.13. The pathway of carbon metabolism includes the gene *ackA* which is involved in acetate metabolism. This gene is significantly downregulated in aerobic conditions. Also, in this list but not clustered were *gadA* and *gadX.*

Table 3.13 Kegg Pathways clustered from gene significantly upregulated in acetic and butyric acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
BIOSYNTHESIS OF ANTIBIOTICS	0.000001	2.6
BIOSYNTHESIS OF AMINO ACIDS	0.00021	2.6
CARBON METABOLISM	0.00024	2.7
BIOSYNTHESIS OF SECONDARY METABOLITES	0.00045	1.8
METHANE METABOLISM	0.00055	5.1
GLYCOLYSIS / GLUCONEOGENESIS	0.0014	3.9
METABOLIC PATHWAYS	0.018	1.3
PYRUVATE METABOLISM	0.023	2.7

Between propionic and butyric acid there were only seven shared genes, and no clusters. The longest list of genes was for the condition of butyric acid alone. Below is the table showing the clusters from this gene list. Many of the clusters and genes have already been described and so will not be again.

Table 3.14 Kegg pathways clustered from gene significantly upregulated in butyric acid alone.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
PURINE METABOLISM	0.000032	2.2
METABOLIC PATHWAYS	0.000037	1.3
BIOSYNTHESIS OF SECONDARY METABOLITES	0.00055	1.5
BIOSYNTHESIS OF ANTIBIOTICS	0.00063	1.6
TERPENOID BACKBONE BIOSYNTHESIS	0.0058	3.7

FATTY ACID BIOSYNTHESIS	0.0058	3.7
UBIQUINONE AND OTHER TERPENOID-QUINONE BIOSYNTHESIS	0.017	2.8
ALANINE, ASPARTATE AND GLUTAMATE METABOLISM	0.025	2.2
NICOTINATE AND NICOTINAMIDE METABOLISM	0.034	2.7
BIOTIN METABOLISM	0.039	3

3.3.9 Genes downregulated when cells were grown anaerobically at pH 7 plus organic acid

The next two sections will examine the genes and pathways downregulated in anaerobic organic acid challenge. First, pH 7 with organic acid. Below is the Venn diagram showing the numbers of significantly downregulated genes and how they are distributed amongst the conditions. There are far fewer genes significantly downregulated in acetic acid compared to the other two, with the majority of genes in the list for acetic acid being shared across all three acids. Beneath figure 3.10 is table 3.15 which is the table of pathways resulting from the clustering of the 136 genes shared across all three acids.

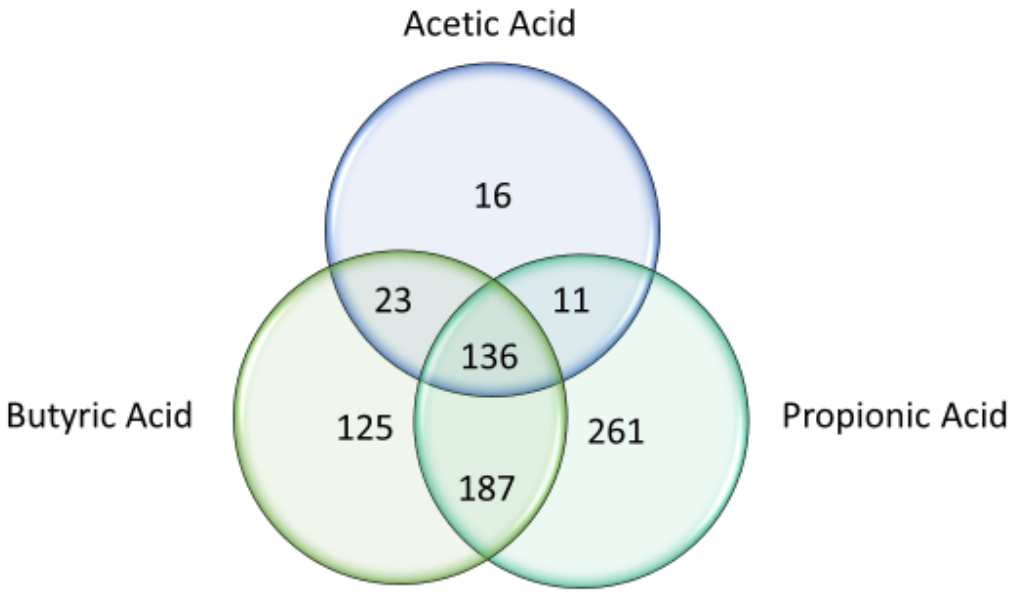


Figure 3.10 Venn diagram showing numbers of genes downregulated in each condition and across conditions.

Table 3.15 The downregulated Kegg pathways shared between all three acids under this condition.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
SULFUR METABOLISM	3.1E-07	8
ARGININE BIOSYNTHESIS	3.3E-06	10.6
2-OXOCARBOXYLIC ACID METABOLISM	4.7E-06	8.3
BIOSYNTHESIS OF AMINO ACIDS	0.00001	3.4
BIOSYNTHESIS OF SECONDARY METABOLITES	0.00001	2.3
BIOSYNTHESIS OF ANTIBIOTICS	0.000049	2.6
METABOLIC PATHWAYS	0.0014	1.4
VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS	0.0033	7.5
C5-BRANCHED DIBASIC ACID METABOLISM	0.0065	9.5
GLYCEROPHOSPHOLIPID METABOLISM	0.029	4.1
ALANINE, ASPARTATE AND GLUTAMATE METABOLISM	0.036	3.8

The pathways above are very similar to those listed as downregulated at both pH 7 and pH 5.5 in aerobic conditions. This is due to the presence of many of the same genes such as *cysADHIJMNPUW* and *argABCDEGHIR* which have been previously described.

Of the 16 genes down-regulated in acetic acid only, none of these cluster into a single pathway. The table below at table 3.16 shows the pathway clustered from the 11 genes shared between propionic and acetic acid: ABC transporters. This is due to the presence of *artI* and *artJ* in this list, which encode elements of the arginine transport system (discussed above). Beneath that (table 3.17) is the table of pathways shared between acetic and propionic acid. These are largely focused on the biosynthesis of amino acids, as well as the ribosome pathway, which indicates a down-regulation of protein synthesis.

Table 3.16 Kegg pathways downregulated shared between acetic and propionic acids.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
ABC TRANSPORTERS	0.06	5.6

Table 3.17 Kegg pathways downregulated between acetic and butyric acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
RIBOSOME	0.0014	8.4
VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS	0.005	24.6
BIOSYNTHESIS OF AMINO ACIDS	0.0065	5.6
BIOSYNTHESIS OF ANTIBIOTICS	0.037	3.4
BIOSYNTHESIS OF SECONDARY METABOLITES	0.038	2.7

Despite a large number of genes being up-regulated in common between propionic and butyric acid (187 genes), there are no significantly downregulated pathways.

As with genes upregulated at pH 7 in anaerobic conditions, a large number of genes were differentially downregulated in propionic acid only. The pathways are shown below at table 3.18. This again points to the particular effect of propionic acid which may be due to the bacteria using propionate as a carbon source.

Table 3.18 Kegg pathways clustered from significantly downregulated genes in propionic acids

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
PHOSPHOTRANSFERASE SYSTEM (PTS)	0.0066	3.9
ABC TRANSPORTERS	0.036	1.9
MICROBIAL METABOLISM IN DIVERSE ENVIRONMENTS	0.05	1.6
GALACTOSE METABOLISM	0.05	3.5

The next section, showing genes downregulated anaerobically at pH 5.5, will compare the above pathways, and those downregulated aerobically with one another.

3.3.10 Genes downregulated when cells were grown anaerobically at pH 5.5 plus organic acid

This section will look at what pathways were downregulated at pH 5.5 in anaerobic conditions. The Venn diagram below shows the numbers of downregulated genes in each condition and shared between conditions.

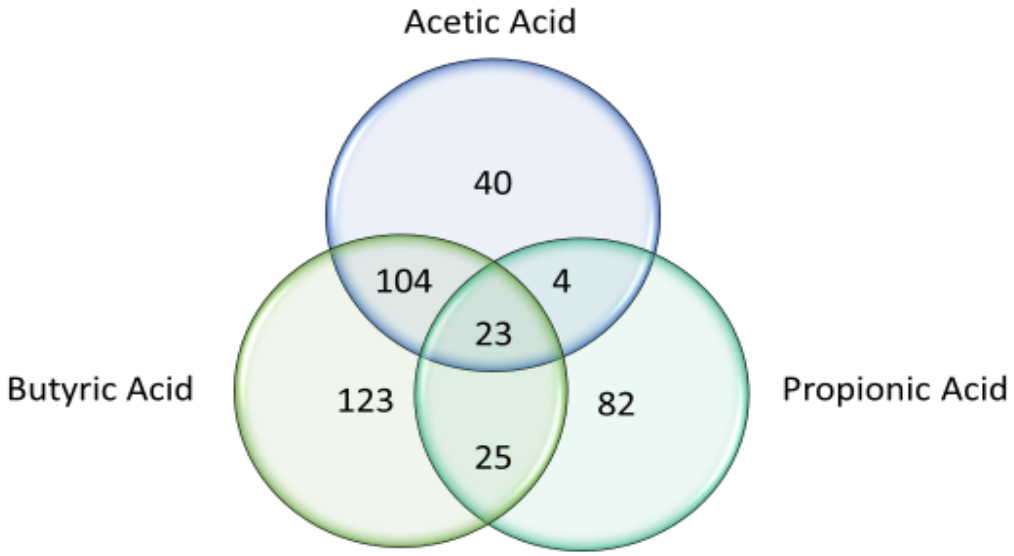


Figure 3.11 Venn diagram showing number of downregulated genes in each condition.

Only 23 genes were shared between all three acids, and two pathways were considered significantly clustered, which are shown below in table 3.19.

Table 3.19 Kegg Pathways downregulated across all three acids.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
GLYCEROPHOSPHOLIPID METABOLISM	6.2E-06	30.2
BIOSYNTHESIS OF SECONDARY METABOLITES	0.0075	3.6

The most significantly clustered is the glycerophospholipid metabolism pathway. This is due to the significant downregulation of *glpABCDFKQT.* These genes are all involved in the utilisation of glycerol in the cell. Notably *glpABC* are involved in anaerobic respiration so it is interesting that they should be downregulated in this condition (compared to the control of pH 5.5 with no added organic acid in

anaerobic conditions). The relationship between glycerol utilisation and organic acids requires more detailed investigation.

Below are four tables (3.20, 3.21, 3.22 and 3.23) showing pathways clustered from those genes down-regulated in acetic acid only, in butyric acid only, shared between butyric and acetic acid and shared between butyric and propionic acid. Again, these pathways have been discussed previously and so will not be repeated here, but the continued appearance of the same genes and pathways is a good indication that the data obtained from the RNAseq was reliable in that the results do not appear to be random.

Table 3.20 Kegg pathways downregulated in acetic acid only.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.0064	9.3
GALACTOSE METABOLISM	0.037	9

Table 3.21 Kegg pathways downregulated in butyric acid only.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
BIOSYNTHESIS OF AMINO ACIDS	1.5E-08	4.6
2-OXOCARBOXYLIC ACID METABOLISM	1.1E-06	9.9
BIOSYNTHESIS OF ANTIBIOTICS	8.8E-06	2.9
ARGININE BIOSYNTHESIS	0.000017	11.1
BIOSYNTHESIS OF SECONDARY METABOLITES	0.00003	2.3
VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS	0.00013	10.7
SULFUR METABOLISM	0.00066	6.1
C5-BRANCHED DIBASIC ACID METABOLISM	0.0039	11.5
METABOLIC PATHWAYS	0.0084	1.4
ALANINE, ASPARTATE AND GLUTAMATE METABOLISM	0.02	4.6

Table 3.22 Kegg pathways downregulated in acetic and butyric acid.

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
PHOSPHOTRANSFERASE SYSTEM (PTS)	0.0014	6.7
MICROBIAL METABOLISM IN DIVERSE ENVIRONMENTS	0.0052	2.4
METABOLIC PATHWAYS	0.011	1.5
CITRATE CYCLE (TCA CYCLE)	0.014	7.3
PYRUVATE METABOLISM	0.018	4.7
CARBON METABOLISM	0.019	3.1
GALACTOSE METABOLISM	0.032	5.5
OXIDATIVE PHOSPHORYLATION	0.048	4.7

Table 3.23 Kegg pathways downregulated in butyric and propionic acid.

KEGG PATHWAYS	P-VALUE	ENRICHMENT SCORES
SULFUR METABOLISM	0.0084	17.9
BIOSYNTHESIS OF ANTIBIOTICS	0.045	4

For the genes significantly downregulated in propionic acid alone, there were no significant clusters. However, within this gene list are several of the *gad* genes and *rpoS*, previously outlined for their role in acid stress response. That they should be downregulated in this condition, but not any of the others and when compared against the anaerobic control of pH 5.5 with no added organic acid (a background of acid stress), is interesting and implies that the cell is utilising different approaches to dealing with this particular stress than other two organic acids tested.

3.3.11 Genes upregulated in acetic acid only.

This section will briefly look at overlaps in the upregulated genes in the four conditions associated with acetic acid, pH 7 and pH 5.5 aerobic and anaerobic. Below at figure 3.12 is a Venn diagram showing the overlaps in gene number between the four conditions. There are only five genes shared between all four conditions. The vast majority of genes upregulated at pH 5.5 and pH 7 aerobic were shared between the two conditions (1053 of 1236). Between the anaerobic conditions, 132 genes are shared, however there are a larger number of genes unique to these two conditions, with 195 upregulated at pH 7 anaerobic and 177 upregulated at pH 5.5 anaerobic that are not shared between any other condition.

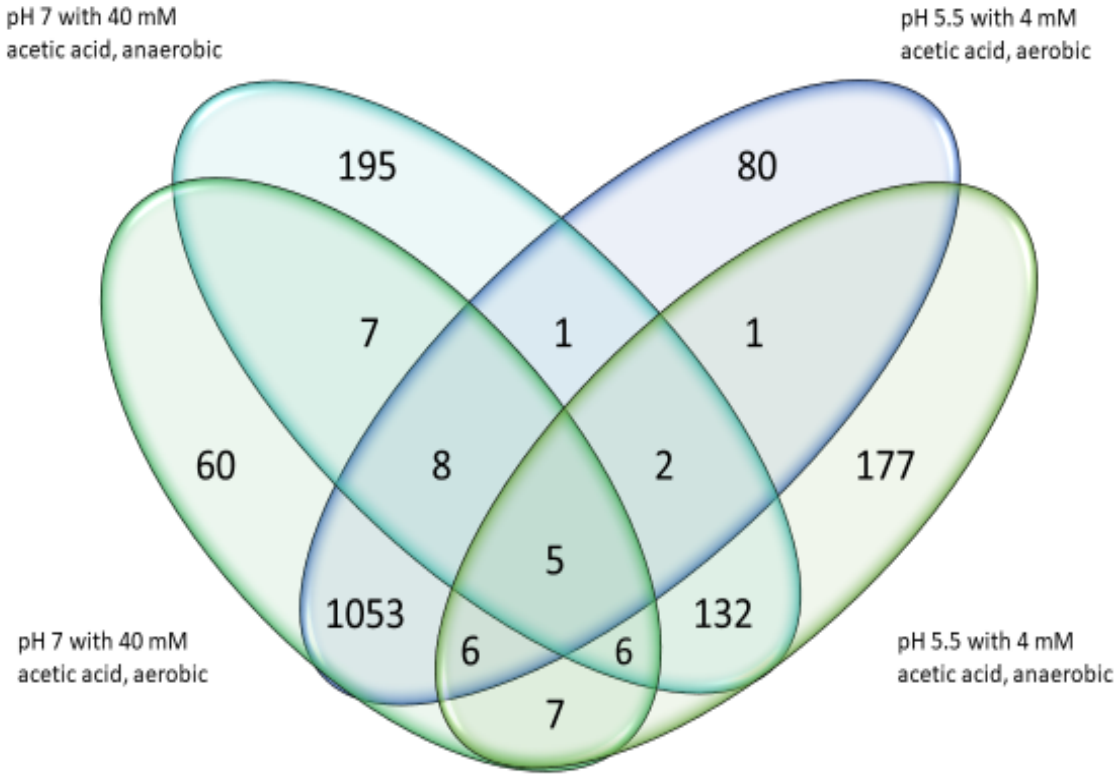


Figure 3.12 Venn diagram showing numbers of upregulated genes in acetic acid only.

Below at table 3.24 is the table of pathways shared across pH 7 and pH 5.5 with acetic acid, aerobic (1053 genes). This table is highly similar to the table outlined above at section 3.3.1 and so will not be discussed again in detail.

Table 3.24 Kegg pathways clustered from significantly upregulated genes in acetic acid across pH and respiratory conditions

KEGG TERM	P-VALUE	ENRICHMENT SCORE
FLAGELLAR ASSEMBLY	1.1E-15	4
PENTOSE AND GLUCURONATE INTERCONVERSIONS	0.000000042	3.5
PHOSPHOTRANSFERASE SYSTEM (PTS)	0.000000092	2.8
FRUCTOSE AND MANNOSE METABOLISM	0.00000041	2.8
PROPANOATE METABOLISM	0.000014	2.7
GALACTOSE METABOLISM	0.000014	2.7
BENZOATE DEGRADATION	0.00084	3.5
BACTERIAL CHEMOTAXIS	0.00088	2.8
TWO-COMPONENT SYSTEM	0.0049	1.4
LYSINE DEGRADATION	0.0053	3.1
FATTY ACID DEGRADATION	0.006	2.8
PHOSPHONATE AND PHOSPHINATE METABOLISM	0.013	3.5
BACTERIAL SECRETION SYSTEM	0.014	2
VALINE, LEUCINE AND ISOLEUCINE DEGRADATION	0.015	3
BUTANOATE METABOLISM	0.021	1.9
TRYPTOPHAN METABOLISM	0.025	3.1
CITRATE CYCLE (TCA CYCLE)	0.033	2
GLYOXYLATE AND DICARBOXYLATE METABOLISM	0.036	1.7

The five genes shared across all four conditions are *acrA, lpd, macB, ptsH* and *tktA.* The gene *lpd* has been described already.

AcrA is part of the AcrAB-TolC system which belongs to the resistance-nodulation-cell division (RND) family of transporters and is a major multidrug efflux pump in *E. coli*. Also on the list is MacB, which is part of the MacAB-TolC system, a member of the ABC transporter family. MacAB-TolC has been shown to infer antibiotic resistance on cells which are lacking the AcrAB-TolC transporter. This implies common functionality under certain conditions which, in this case, is acetic acid resistance. These transporters

may be involved in expulsion of protons or anions from the cell in order to restore homeostasis (Hayashi et al., 2016; Lu and Zgurskaya, 2012). At pH 7 and 5.5, *acrABDEFRS* are all significantly upregulated.

PtsH is a non-sugar-specific component of the PTS system, which has come up repeatedly as one of the most highly upregulated systems under organic acid stress, so its presence in this list is unsurprising. The final gene in the list *tktA* is a transketolase, an enzyme which catalyses a reversible reaction between glycolysis and the pentose-phosphate pathway which again have come up multiple times in this analysis. This gene has been discussed in more detail in Chapter Six.

Below is a list of pathways clustered from the 132 genes shared between the two anaerobic conditions. These are also similar to those listed previously.

Table 3.25 Kegg pathways upregulated in pH 7 anaerobic & pH 5.5 anaerobic:

KEGG PATHWAYS	P-VALUES	ENRICHMENT SCORES
BIOSYNTHESIS OF ANTIBIOTICS	0.0011	2.2
GLYCOLYSIS / GLUCONEOGENESIS	0.0013	4.5
CARBON METABOLISM	0.0019	2.7
PYRUVATE METABOLISM	0.021	3.1
BIOSYNTHESIS OF SECONDARY METABOLITES	0.022	1.6
METHANE METABOLISM	0.028	4.2
BIOSYNTHESIS OF AMINO ACIDS	0.029	2.1

There are a high number of genes not shared between each separate condition here, indicating the different responses by the cell to the varying environmental conditions other than the addition of acetic acid. This in itself is not surprising. Of the 177 genes upregulated in pH 5.5 in anaerobic growth only, there were no significantly clustered pathways. In the case of the 195 genes upregulated in pH 7 anaerobic only, the list of pathways reflected those already presented above in section 3.3.7 and will not be covered again.

For the purpose of further investigation, those genes shared between conditions are perhaps a better indication of the cells response specifically to acetic acid (or organic acids in general) than those only upregulated under, say, anaerobic conditions. However, those genes cannot be disregarded, and have all been measured as significantly upregulated compared against a control condition which was also grown anaerobically, so they cannot be put down to anaerobic growth alone. This gene list requires more comprehensive analysis than possible with this approach and that is the intention for this data in terms of future publication.

3.3.12 Genes downregulated in acetic acid only

The final part of this section will examine the genes downregulated in acetic acid and below at Figure 3.13 is a Venn diagram showing the numbers of genes overlapping in each condition or not. What is immediately obvious are the very low numbers of genes overall, and overlapping. The majority of genes are unique to each condition, with the most at pH 5.5 during aerobic growth. There are 59 genes shared between the two aerobic condition, but only one shared between the two anaerobic conditions. There are no genes shared between all four conditions.

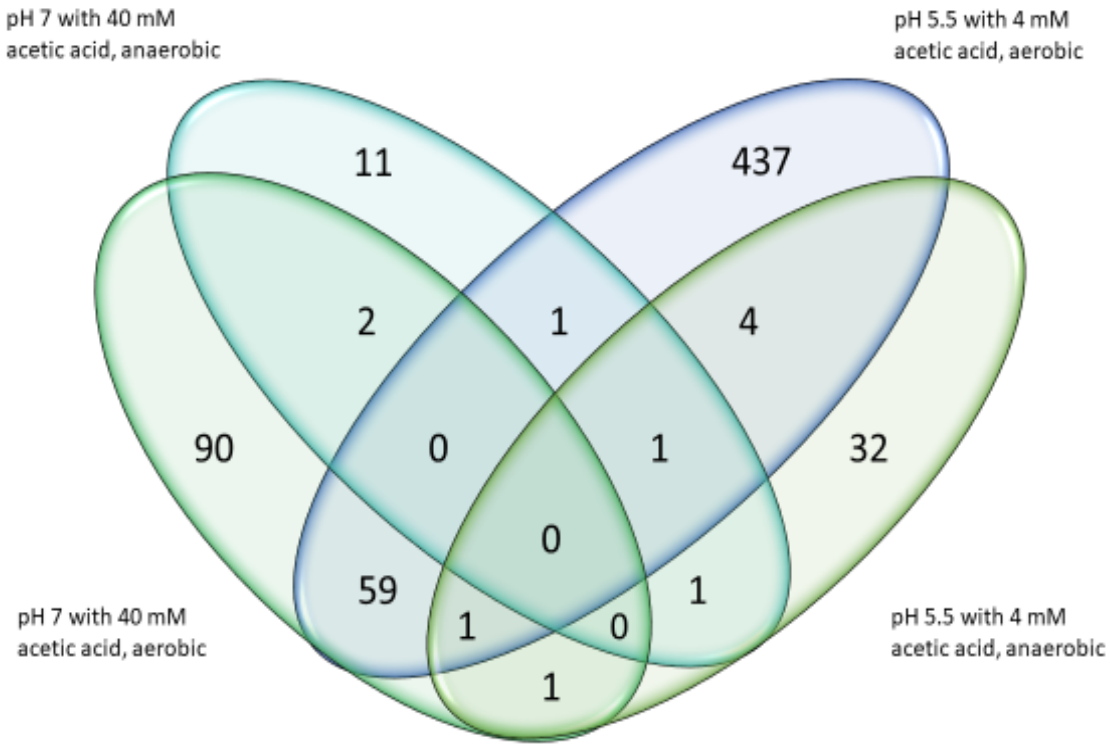


Figure 3.13 Venn diagram showing numbers of downregulated genes in acetic acid only.

Of the 59 genes shared between the two aerobic conditions, the following pathways shown in table 3.26 were clustered:

Table 3.26 Kegg pathways shared between pH 7 and pH 5.5, aerobic, in the presence of acetic acid.

KEGG PATHWAYS	P-VALUE	ENRICHMENT SCORES
METABOLIC PATHWAYS	0.0047	1.5
PYRIMIDINE METABOLISM	0.044	3.6
VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS	0.045	8.4

The most genes on any individual list were those unique to pH 5.5 during aerobic growth. Table 3.27 shows the pathways clustered from those genes.

Table 3.27 Pathways clustered from genes downregulated at pH 5.5 during aerobic growth only:

KEGG PATHWAY	P-VALUE	ENRICHMENT SCORE
BIOSYNTHESIS OF ANTIBIOTICS	0.00057	1.6
METABOLIC PATHWAYS	0.00076	1.2
THIAMINE METABOLISM	0.0035	4.1
BIOSYNTHESIS OF SECONDARY METABOLITES	0.0075	1.4
PENTOSE PHOSPHATE PATHWAY	0.02	2.3
GLUTATHIONE METABOLISM	0.033	2.7
CARBON METABOLISM	0.046	1.5
GLYCOLYSIS / GLUCONEOGENESIS	0.049	1.9

The first pathway, biosynthesis of antibiotics, contains the genes *pfkA* and *pfkB*. These two genes encode 6-phosphofructokinase I and 6-phosphofructokinase II. These two enzyme are key in regulating the glycolysis pathway (which is also a significantly downregulated pathway in this list), with 90% of the activity attributed to PFK-I and 5% attributed to PFK-II. That these two key enzymes catalysing the same reaction should be jointly downregulated at pH 5.5 aerobic only is interesting. These enzymes are inhibited by high levels of ATP, so it is possible that ATP levels in this condition are significantly different to those in the other conditions (Cabrera et al., 2011). This would require further investigation.

The pathways and gene combinations studied in this chapter are complex and intertwining with many of the pathways and genes all influencing and inducing one another. For example, the thiamine metabolism pathway above is present in the clustering because the genes *thiCDEFGH* are all present in the list. These genes are involved in the synthesis of thiazole, which is primarily utilised under anaerobic conditions. On that basis it is perhaps not strange that it is downregulated under aerobic conditions. However, it is significantly downregulated compared to the control condition, which is also aerobic, which implies organic acid is inducing further downregulation. The gene *thiC* is a member of the SAM family, referred to above in section 3.3.9 – genes upregulated at pH 5.5 anaerobic – when discussing the gene *metK* which catalyses formation of SAM (Kriek et al., 2007). Again, this highlights the complicated

interactions between these genes and pathways which requires a more sophisticated method of analysis.

The next and final part of this chapter will examine the genes designated 'ST131v2_' which have unknown function.

3.4 ST131 Genes of unknown function

The last set of data to be presented here is that relating to the genes of unknown function from the ST131 genome that have been significantly up- or down-regulated in the conditions tested. Of the 855 ST131 specific genes on the genome and mega-plasmid, between 250 and 375 were significantly upregulated in aerobic conditions between the three acids and pH, whereas between 18 and 38 genes were significantly downregulated. In anaerobic conditions the ST131 genes were much less likely to be either up or down regulated, with between 3 and 27 genes appearing in lists across conditions for upregulation and 4 – 120 being downregulated. Interestingly, propionic acid was the condition with the largest and smallest list of both up and downregulated genes, with 3 genes upregulated at pH 5.5 in anaerobic growth conditions, and 4 genes downregulated and 27 genes upregulated at pH 7 in anaerobic growth and 120 genes downregulated.

The potential function of these genes was investigated by doing BLASTx (blast.ncbi.nlm.nih.gov) searches against the nucleotide sequences to find proteins of similar sequence. Given the high number of genes, searches were limited to those genes shared across all three acids and where a number of genes were placed consecutively on the genome.

Of the genes shared at pH 7 aerobic across all three acids, the BLAST search revealed high similarity of genes named ST131v2_00958 – ST131v2_00986 (over 99% match) to Type VI secretion system proteins

including TssE, TssF, TssJ, VasK, VasL, VgrG and TssK, which encode various subunits of the system including the cell puncturing tip, baseplate and tail sheath (Journet and Cascales, 2016). Whilst it is not clear why this entire pathway is being upregulated under organic acid stress, a study by Wang et al., (2015) showed that when under oxidative stress *Yersinia pseudotuberculosis* utilised its T6SS for uptake of zinc ions. They hypothesise that the T6SS carries out a diverse range of unknown functions, particularly under environmental stress in order to survive in detrimental environments. This presents the interesting possibility that this strain of uropathogenic *E. coli* is utilising T6SS under organic acid stress as a means of importing or exporting ions for maintenance of the transmembrane electrical potential.

Also among the genes shared by all three acids at pH 7 aerobic were the genes ST131v2_1801 – ST131v2_1813, which are equally distributed between unknown hypothetical proteins and members of the IS21 transposase family. The genes ST131v2_01414 – 01417 were matched to an oxidoreductase, proteins of unknown function and an AlpA family transcriptional regulator. AlpA regulates Alp, which stands for alternative Lon protease. Lon is an ATP-dependent protease which degrades misfolded proteins and rapidly degraded regulatory proteins. Lon is significantly down-regulated in this condition.

At pH 5.5 during aerobic growth, there were 375 upregulated ST131 genes in acetic acid, 330 upregulated genes in propionic acid and 346 upregulated genes in butyric acid. Of those genes, 304 were common between all three. Of the list in common between all three acids, 299 of those were also shared with pH 7 upregulated across all three acids. There were 5 genes which were shared between the three acids but unique to pH 5.5, and 16 shared between the three acids but unique to pH 7.

Of the 16 genes upregulated at pH 7 but not pH 5.5, these were mostly of unknown function (either listed as hypothetical or domain of unknown function). Two of the genes associated with type VI

secretion system appeared on the list and one transposase. Of the 5 genes upregulated at pH 5.5 but not pH 7, three were on the plasmid and share high similarity with TrbA and TrbE, conjugal transfer proteins. The other two were a transposase and an unknown protein.

More than 15% of the genes in the EO 499 genome are ST131 specific and have not been annotated. Almost half of those are collectively up- or down-regulated in these stress conditions. On that basis they make for a fascinating prospect of further study in regards to their function and how they contribute to the ability of this pathogen to survive and proliferate.

3.5 Discussion

This chapter has presented data from 12 independent RNA-sequencing experiments using three different organic acids, resulting in 24 lists of up to 1200 genes. Given the very large amount of data, the examination of individual gene profiles was not realistic in the scope of this thesis, and so the focus was mainly on pathways that clustered from lists of significantly up- and down-regulated genes.

The results showed that the response of EO 499 to organic acids was quite different to the response to low pH only, demonstrated by the alternative up and downregulated pathways at pH 5.5 compared to both pH 7 with organic acids and pH 5.5 with organic acids. Zorraquino et al., (2016) showed differential expression of 21 genes at pH 5.5. In this study, 53 genes were differentially expressed at pH 5.5 with no organic acids compared to pH 7. The genes shown in that research as differentially regulated included *relA, lon, gadA,* which were also see in this study. At pH 5.5 with 4 mM acetic acid 1558 genes were significantly upregulated and 1098 significantly downregulated.

The key pathways which were commonly upregulated in all three organic acids, at both pH tested were those involved in metabolism, namely glycolysis, the phosphotransferase system, the pentose

phosphate pathway and the TCA cycle. Given the key metabolic pathways seen in this analysis both utilise and produce acetate, thereby increasing both production and use of anions within the cytoplasm, it makes biological sense to upregulate these pathways to use up the additional acetate within the cytoplasm. It is possible that by upregulating key metabolic pathways such as the TCA cycle, more acetate is used than produced, which would have net positive impact on reducing stress on the cell. In regards to the other organic acids, it is possible these pathways are being upregulated in order to provide enough energy to the cell for the pathways it is utilising to remove butyrate and propionate from the cell. However, *E. coli* can also utilise propionate (but not butyrate) via its conversion to pyruvate which can then enter the TCA cycle (Brock et al., 2002). In anaerobic conditions at pH 7 a large number of genes were unique to propionic acid alone, and this may be indicative of how propionate is utilised by the cell, particularly under anaerobic conditions.

A number of virulence factors were also upregulated, including flagella assembly, chemotaxis and secretion systems. Analysis of unannotated genes showed that a large number of these genes likely belong to a type 6 secretion system, which has to date not been described in this strain.

The data presented here will be visited again in Chapters Four and Six.

Chapter Four:

Analysis of genes of phenotypic importance in organic acid stress using TraDIS

Chapter 4: Analysis of genes of phenotypic importance in organic acid stress using TraDIS

The previous chapter described the results of the RNA sequencing, focusing on the pathways involved in the response to organic acid stress at both neutral and mildly acidic pH. This chapter will focus on the TraDIS results from the same conditions.

Transposon Directed Insertion Sequencing (TraDIS) is a method by which every gene in the genome can be analysed in terms of its contribution to fitness in a particular environment. Fitness is a measure of how well a cell competes with its neighbour to reproduce. A cell with low fitness will compete poorly, and reproduce more slowly than a cell with a higher fitness. Ultimately the cell with lower fitness may be completely outcompeted and disappear entirely from a population. There are also measures of fitness in between this extreme, where a cell may be slightly outperformed, or reproduce slightly better. This can then be quantified. TraDIS allows large scale competition of millions of single-transposon mutants. Each gene within the genome contains potentially thousands of individual transposon insertions in different insertion sites. The library is sequenced to determine where specifically each transposon insert is located and the frequency of insertions into each gene. This mutant library is then cultured under a specific condition and the culture is re-sequenced in order to determine transposon positions and frequencies of the final population (see Figure 4.1 below). The individual mutants which have a transposon inserted in a gene which is required for growth in the stress condition will not grow, or growth will be relatively poor, and therefore that gene will be considered conditionally essential. Some mutants may have relatively improved growth, and therefore by inhibiting gene function there has been a fitness advantage provided. In this way, it is possible to determine those genes which are required under certain conditions, those which are not and those where loss of function improve fitness in the condition tested (Langridge et al., 2009).

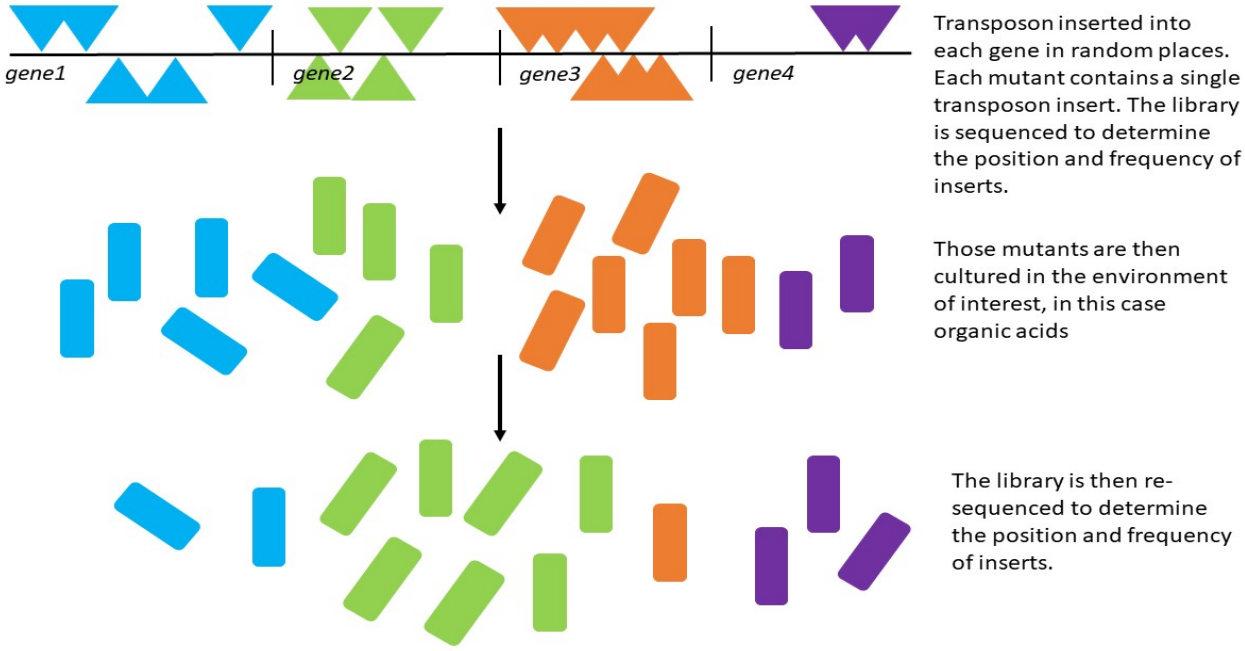


Figure 4.1 An overview of TraDIS

Goodall et al., (2018) have used TraDIS to map the essential genome of BW25113 (the background strain of the Keio library) and have noted inconsistencies in terms of the numbers of genes in this strain which have previously been reported as essential which are not also considered essential in a TraDIS experiment. They conclude this is at least in part attributable to different definitions of what is considered essential, however they also note insertion bias, whereby there is a high transposon insertion density in the 3' end of a gene which cannot be deleted in its entirety. They also concluded that insertions could be made in otherwise essential genes, but that these insertions were at low frequency when compared to non-essential regions of the genome. They used particular statistical tests in order to determine the difference between truly essential genes and those which are conditionally or non-essential.

In this study, the transposon used to construct the library encoded the *cat* gene, which conferred chloramphenicol resistance for subsequent selection of successful transposon mutants. The transposon

itself has flanking regions of palindromic repeats, which are used to construct primers for PCR amplification and subsequent sequencing to determine transposon frequency and insertion site. The palindromic flanking regions allow PCR amplification regardless of the orientation of the transposon. The full transposon sequence is included in the supplementary data.

In this chapter, data will be analysed from TraDIS populations grown in the conditions as described in Chapter Three. The TraDIS library was made and provided by Discuva Ltd (now defunct). It was cultured, genomic DNA purified and sequenced as described in Chapter Two. Lists were generated from the data processing completed by Dr John Herbert, also described in Chapter Two. Gene insertions within a particular gene were classed as significantly over- or under-represented in the final population on the basis of a p-value of 0.2, once adjusted for a false discovery rate of 10%. This high p-value was chosen as under some conditions, particularly in anaerobic growth, a more stringent p-value resulted in lists with either very low numbers of genes, or none at all. The p-values of genes being analysed in this chapter have been listed in the data below in order to show that the majority of genes used in the analysis had a much lower p-value than 0.2 and are therefore less likely to be a false positive.

The TraDIS analysis has resulted in lists of far fewer genes that were considered significant than in the RNAseq analysis. On that basis, this chapter will focus more on the roles of individual genes, and how they may contribute to the strain's organic acid resistance pathways, rather than on pathways themselves. This is primarily because the lists are too short for pathway analysis (there were no significant clusters from the lists).

What is also notable throughout this chapter is that there are far fewer genes shared amongst pH/respiratory conditions as well, with only one occasion where a single gene is shared between all four conditions (butyric acid genes required for fitness 4.1.4). There is however similarity between acids and,

therefore the first sub-section of each main part of the chapter will focus on those genes shared between acids. There will then be a breakdown of the genes which came up as significant (either improving or limiting fitness) in each acid individually with focus shifting to differences between the pH and respiratory conditions. Unlike in the previous chapter, the ST131 specific genes will be incorporated into this main analysis (and are denoted simply by the number at the end of the gene, as all have the same prefix of ST131v2_xxx). These genes have been included within the analysis of genes of known function because this analysis is focused on roles of individual genes rather than pathways, and so it is possible to incorporate these genes into the main analysis (which was not possible with pathway analysis).

This chapter will start with an analysis of genes for which the frequency of transposon inserts in the final population were lower when compared to the starting population. This implies that the mutants carrying those transposon inserts have a limitation on fitness under organic acid stress. The following section will consider those genes whereby insertion frequency increased. The conditions tested were pH 7 with 40 mM acetic acid, pH 7 with 30 mM propionic acid and pH 7 with 30 mM butyric acid, which have all been compared to a control condition of pH 7 with no added organic acid in order to determine significant change in insertion frequency compared to the control (in order to remove changes due to other factors such as media composition or aerobic/anaerobic growth). The library grown pH 5.5 with 4 mM added acetic acid, 2 mM propionic acid or 2 mM butyric acid was compared against a control condition of pH 5.5. The TraDIS data for the library grown at pH 5.5 only has also been compared against a control condition of pH 7 in order to provide comparison for those genes which appear to contribute to fitness in acid stress in general (as opposed to organic acid stress in particular). The library was cultured as above in both aerobic and anaerobic conditions and where the data shows a change in transposon frequency under aerobic conditions, this is against a control condition also cultured aerobically. Where the TraDIS data for anaerobic conditions were compared to a control cultured grown

anaerobically. This chapter will conclude with a discussion about the possible cellular processes involved in the organic acid stress response.

4.1 Genes implicated as being required for fitness under organic acid stress

As described above, this section will investigate genes for which the frequency of transposon inserts has significantly decreased by Log2 fold (significance considered to be an FDR adjusted p-value of 0.2 or less). The data was analysed using the ESSENTIALS tool, as described in Chapter Two, and was processed in its entirety by Dr John Herbert of the University of Liverpool. Dr Herbert subsequently calculated fold change as control/stress, which has resulted in a positive number for decreas in insertion frequency, and a negative number for increase in insertion frequency. The cut off is 1 or -1 of log2 transformed data. The graph below at Figure 4.2 shows the data for all mutants in the library grown at pH 5.5 aerobically in the presence of 4 mM acetic acid, compared against a control of pH 5.5 during aerobic growth. The arrows show the point above which or below which genes have been analysed, provided they have an FDR adjusted p-value of 0.2 or less. The majority of genes here fall across the middle of the graph at or around zero. This shows that the majority of single-transposon inserts had no impact on the cell and these mutants were neither more or less fit than most of the other cells in the population. This graph shows that the gene mutation with the largest limitation on fitness under this condition, when frequency of transposon inserts was compared at the start and finish of the experiment against the control, was *apaH.* The gene whereby fitness was most improved was *rpmF.* This highlights an important issue with TraDIS experiments mentioned above. The gene *rpmF* is essential, therefore this result may simply be an artefact of very few initial insertions increasing in frequency by a small amount, which is then interpreted as being highly significant (Goodall et al., 2018).

A full list of TraDIS results for each conditions can be found in the supplementary data provided.

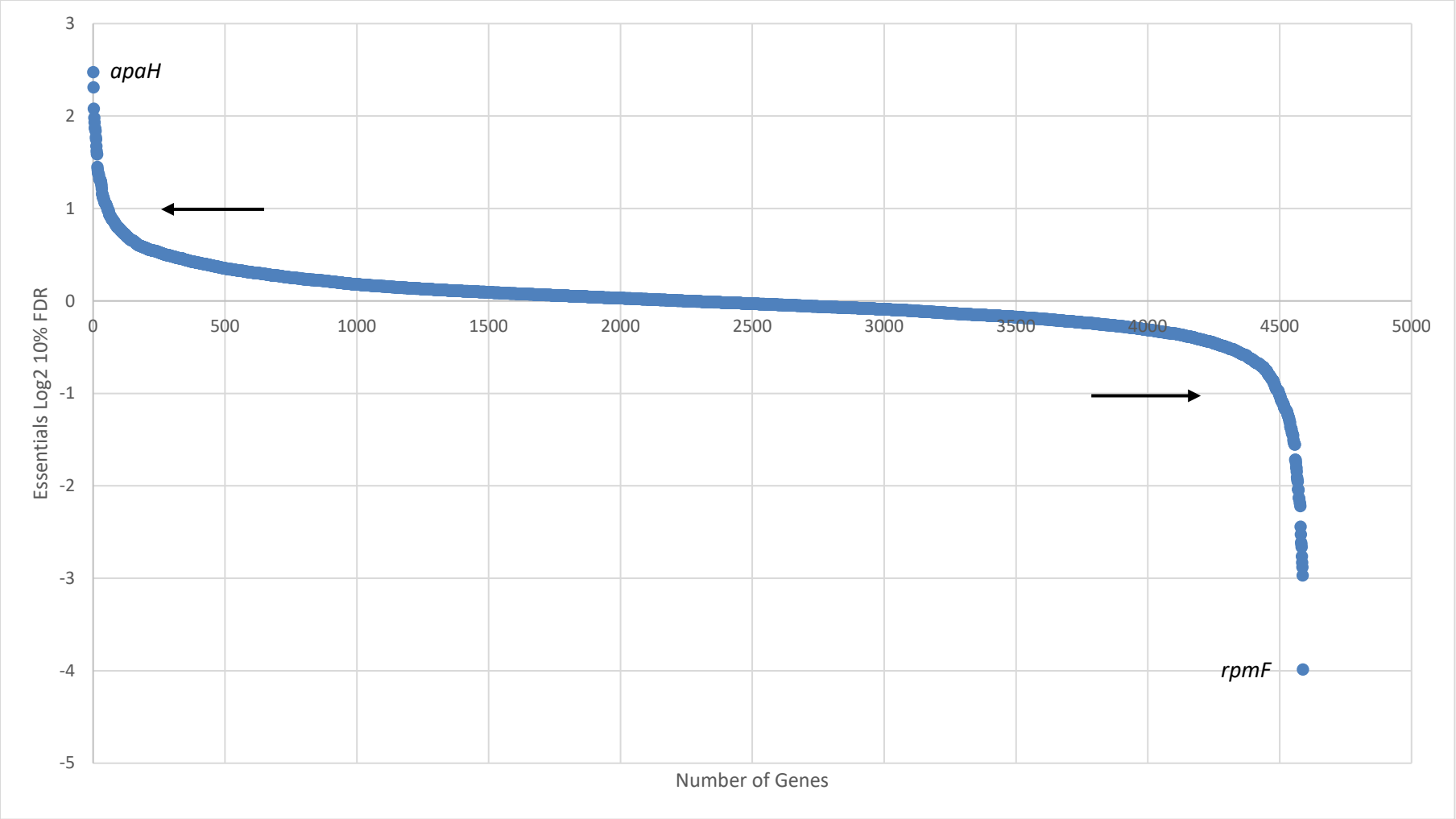


Figure 4.2 All genes in the genome once frequency of transposon inserts has been calculated following culture at pH 5.5 with 4 mM acetic acid, compared against the control condition of pH 5.5 only. Average from three replicates. Arrows show point at which genes above or below were used for analysis provided p-value was considered as significant.

4.1.1 Genes with a decrease in insertion frequency during aerobic growth in organic acids

The first of the two tables below is a table of genes which have been implicated as required by the cell to some extent in organic acid stress at pH 5.5, in aerobic growth. None of these genes appears in the list of those required at pH 5.5 alone (data shown in the second table above at Table 4.2). Genes are ranked according to Essentials Log2 score. A higher score denotes a greater difference in frequency of transposon insertions compared to the control condition.

Acetic	Essentials Log2 Score	FDR adjusted P-value	Propionic	Essentials Log2 Score	FDR adjusted P-value	Butyric	Essentials Log2 Score	FDR adjusted P-value
apaH	2.474	6.92E-09	*yejM*	2.142	3.72E-18	*yejM*	1.778	1.06E-13
lon	1.980	0.000193	*nuoL*	1.493	3.18E-06	*nuoN*	1.711	1.69E-07
yejM	1.423	0.00241	*nuoE*	1.480	0.01045	*lon*	1.584	4.29E-08
nuoF	1.312	0.000696	*nuoN*	1.468	4.00E-06	*apaH*	1.578	1.12E-05
nuoN	1.158	0.099614	*apaH*	1.314	0.003127	*nuoE*	1.214	0.011634
nuoE	1.124	0.104333	*nuoF*	1.282	1.09E-03	*nuoL*	1.203	1.12E-05
sucD	1.117	7.10E-05	*lon*	1.206	0.000199	*nuoF*	1.176	0.000737
nuoL	1.098	0.012023	*sucD*	1.125	0.000154	*sucD*	1.139	2.06E-07

Table 4.1 Genes with a Log2 essentials score of 1 or more which were shared across all organic acids at pH 5.5 with 4 mM acetic acid, aerobic (versus pH 5.5 aerobic control), ranked according to score with associated p-value.

Gene	Essentials Score	FDR adjusted P-value	Gene	Essentials Score	FDR adjusted P-Value
1. *dgkA*	3.751	8.79E-46	10. *hscA*	1.200	0.017806
2. *sapC*	1.684	1.77E-07	11. _03155	1.158	0.060197
3. *sapD*	1.599	4.84E-05	12. _00545	1.138	0.063402
4. *pal*	1.510	0.000607	13. _00544	1.138	0.063402
5. *trkA*	1.471	8.74E-05	14. *ribF*	1.077	0.108461
6. *dsbB*	1.398	0.047687	15. *hokE*	1.031	0.01108
7. *atpB*	1.327	0.001588	16. _03178	1.011	0.048678
8. _00938	1.279	0.06706	17. *insB*	1.011	0.036454
9. _00977	1.224	0.109025			

Table 4.2 Genes with a Log2 essentials score of 1 or more which were implicated as impacting on fitness in pH 5.5 aerobic alone (pH 5.5 v pH 7 aerobic control condition), ranked according to score with associated p-value.

Acetic & Butyric	**Essentials Score**	**FDR adjusted P-value**	**Acetic, Butyric & pH 5.5**	**Essentials Score**	**FDR adjusted P-value**
clp	1. 1.641, 2. 1.852	1. 0.018, 2. 1.06E-13	*sapD*	1. 1.232, 2. 1.113, 3. 1.599	1. 0.121, 2. 0.0256, 3. 4.84E-05
lon	1. 1.970, 2. 1.038	1. 1.61E-06, 2. 0.059			
Butyric & pH 5.5	**Essentials Score**	**FDR adjusted P-value**	**Propionic, butyric & pH 5.5**	**Essentials Score**	**FDR adjusted P-value**
_03178	1. 1.139, 2. 1.011	1. 0.063, 2. 0.049	_00977	1. 1.381, 2. 1.338, 3. 1.224	1. 0.035, 2. 0.071, 3. 0.109
Acetic, Propionic & pH 5.5	**Essentials Score**	**FDR adjusted P-value**			
_03155	1. 1.0138, 2. 1.905, 3. 1.158	1. 0.039, 2. 0.0005, 3. 0.060			

Table 4.3 Genes in common across more than one acid with a Log2 essentials score of 1 or more at pH 7 with 40 mM acetic acid (compared to a control of pH 7 aerobic) with associated p-value.

Table 4.3 above shows those genes shared across organic acids at pH 7 aerobic, and shared with pH 5.5 alone. There were fewer shared genes, and no gene shared across all three organic acids at pH 7 in aerobic conditions. The three ST131 specific genes shown above, _03178, _00977 and _03155 were analysed using BLASTx online tool and were found to be of unknown function.

In table 4.1 showing genes shared at pH 5.5, the first on the list for acetic acid and also listed for propionic and butyric acid is *apaH,* which encodes diadenosine tetraphosphate, which hydrolyses

diadenosine-5', 5'''-P^1, P^4-tetraphosphate (Ap4A) into two molecules of ADP. This gene will be discussed in more details in Chapter Six when comparing this data against the results of the RNAseq and experimental evolution.

The gene at the top of the list in Table 4.1 for propionic acid and butyric acid, and third on the list for acetic acid is *yejM*. AlthoughtThis gene is considered essential, the 3' end has been shown to be dispensable and therefore can support transposon insertions (Goodall et al., 2018). Its inclusion in this list, therefore, is likely a result of small changes in transposon insertion frequencies being considered more significant than they are. However, this data was compared against the control condition of pH 5.5 over three replicates and has a very low p-value. This implies that although there may be bias of transposon insertions in the 3' end of this gene, the addition of organic acids may have impacted further upon the ability of mutants to proliferate in the conditions tested. This is supported by other research into this gene whereby *yejM* knockout strains are not viable but cells with mutations to rather than deletion of *yejM* are shown to be more sensitive to a range of stresses including temperature and antibiotics (De Lay and Cronan, 2008).

Within the list in Table 4.1 of those genes considered as significant across the three organic acids at pH 5.5 are *nuoEFLN.* Not all of the data is shown in the tables above. However, in the subsections below there are figure showing the genes whereby loss of fitness (calculated as reduction in transposon frequency) have been plotted. These figures show that across all the data sets, the loss of any one of the genes *nuoEFHLMN* resulted in loss of fitness in mutants cultured in the three acids at pH 5.5 and/or pH 7 in aerobic conditions. In acetic acid only, NuoJ is also in the list. Shared between acetic and butyric is NuoG and shared between propionic and butyric is NuoI. NuoB was also in the list for butyric acid. Together these are ten of the 13 subunits of the NADH:ubiquinone oxidoreductase respiratory Complex

I. As discussed in Chapters One and Three, this system is involved in pumping protons across the membrane and, as such, plays a key role in maintaining proton motive force (Erhardt et al., 2012; Spehr et al., 1999). Erhardt et al., (2012) showed that disruption of any one of these subunits resulted in a only partial stability and functionality of Complex I in *E. coli*. That ties in with the data seen here. This system appears key in the regulation of PMF whilst under organic acid stress.

Two genes on the list in Table 4.3 are *clpP* and *lon* which are both ATP-dependent proteases, and are shared between butyric and acetic acid at pH 7. *Lon* in the list of significant genes for propionic acid growth at pH 5.5 and *clp* for propionic at pH 7 in anaerobic conditions (data shown below). Neither gene however appears in the lists for pH 5.5 anaerobic for any acid (or pH 5.5 alone).

ClpP and Lon are both responsible for removal of misfolded and redundant proteins from the cell during normal cell regulation. This implies that protein folding, misfolding and subsequent removal from the cell are important under organic acid stress. This may be because organic acids or lower intracellular pH as a result of H^+ influx over and above that caused by pH 5.5 alone interfere with precise and effective protein folding which subsequently need to be efficiently removed from the cell and therefore interruption of these proteases negatively impacts on the cells ability to reproduce when either of these proteases is inactivated (Miller et al., 2013; Vera et al., 2005).

In table 4.3 *sapD* is shown as shared between acetic acid, butyric acid and at pH 5.5 alone. For propionic acid, *sapD* was listed under pH 7 in anaerobic growth. This is an ATP binding protein involved in putrescine efflux. However, it has also been shown to be involved in K^+ uptake which, as noted in Chapter One, is important for maintenance of cell homeostasis particularly under low pH. Inactivation of *sapD* almost completely inhibited action of the SapAF ABC binding cassette and, interestingly, *sapF*

appears on the list as improving fitness when function is lost in propionic acid at pH 7 in aerobic conditions (shown below in section 4.2.4). Inactivation of *sapF* alone did not show inhibition of the ABC binding cassette however (Schmid et al., 2001). It may be that inactivation of *sapF* stimulates expression of an alternative system.

Also on the list of genes shared between all three acids at pH 5.5 is *sucD*. In the list for propionic only was *sucB* and in the list for butyric and acetic was *sucAC*. All these genes are involved in the TCA cycle (as discussed briefly in Chapter Three). It is interesting that they should appear both in this list and that for the RNAseq and this will be discussed in more detail in Chapter Six.

The next subsection will consider those genes whereby loss of function has been implicated in loss of fitness in organic acids under anaerobic growth conditions.

4.1.2 Genes with a decrease in insertion frequency during anaerobic growth in organic acids

Tables 4.4, 4.5 and 4.6 below show the genes which were shared in all three acids as being required for fitness in anaerobic growth. There are a much larger number of genes shared at pH 5.5 than at pH 7 with organic acids and also more than in pH 5.5 alone when compared to pH 7.

Acetic	Essentials Log2 Score	FDR adjusted P-value	Propionic	Essentials Log2 Score	FDR adjusted P-value	Butyric	Essentials Log2 Score	FDR adjusted P-value
mobA	9.170	0.103	*mobA*	9.199	0.098	*mobA*	9.242	0.099
moaD	8.868	0.103	*moaD*	9.025	0.098	*moaD*	9.047	0.099
moaA	8.724	0.103	*moaE*	8.461	0.098	*moaC*	8.672	0.099
moaC	8.702	0.103	*moeB*	8.442	0.098	*moaA*	8.640	0.099
moeA	8.614	0.103	*moaA*	8.437	0.098	*moeA*	8.521	0.099
moeB	8.407	0.103	*moaC*	8.429	0.098	*moeB*	8.391	0.099
moaE	8.336	0.103	*moeA*	8.417	0.098	*moaE*	8.344	0.099
mog	7.846	0.103	*narQ*	7.608	0.098	*mog*	7.626	0.113
narQ	7.592	0.103	*mog*	7.534	0.137	*narQ*	7.597	0.099
narG	6.468	0.131	*narG*	6.393	0.183	*narG*	6.428	0.170
narH	5.992	0.131	*narH*	5.886	0.183	*pykF*	6.173	0.099
pykF	5.812	0.126	*pykF*	5.534	0.193	*narH*	5.894	0.170
moaB	5.530	0.126	*moaB*	5.300	0.183	*moaB*	5.591	0.115
typA	3.005	0.013	*typA*	2.368	0.098	*typA*	3.004	0.006

Table 4.4 Genes with a Log2 essentials score of 1 or more which were shared across all organic acids at pH 5.5 anaerobic (versus pH 5.5 anaerobic control), ranked according to score with associated p-value.

pH 5.5 v pH 7 anaerobic	Essentials Log2 Score	FDR adjusted P-value
lepB	2.327	0.096
dgkA	1.795	0.039
dicB	1.418	0.119
mdh	1.369	0.096
kup	1.144	0.173

Table 4.5 Genes with a Log2 essentials score of 1 or more which were implicated as impacting on fitness in pH 5.5 anaerobic alone (pH 5.5 v pH 7 anaerobic control condition), ranked according to score with associated p-value.

Acetic & Propionic	Essentials Log2 Score	FDR adjusted P-value
_04682	a. 1.921, b. 1.566	a. 0.007, b. 0.024
clpP	a. 1.287, b. 2.354	a. 0.033, b. 1.07E-05

Table 4.6 Genes in common across more than one acid with a Log2 essentials score of 1 or more at pH 7 with organic acid anaerobic (compared to a control of pH 7 anaerobic) with associated p-value.

The first group of genes in table 4.4 *moaABCDE, mobA, moeAB, mog* and *narGH* were all shared between all three acids at pH 5.5 in anaerobic growth. These genes are all involved in anaerobic respiration via molybdoenzymes and nitrate reductase, which subsequently contribute to the maintenance of the proton motive force under anaerobic conditions (Anderson et al., 2000; Cotter and Gunsalus, 1989; Leimkühler, 2014; McNicholas et al., 1997). NarQ is a nitrate dependent sensor kinase which, together with response regulator NarP contributes to the bacterium's anaerobic respiration gene expression and is responsible for direct induction of genes *narGH* above (Chiang et al., 1992).

The gene *pykF* is a pyruvate kinase. Pyruvate is a key intermediary in the glycolytic pathway. PykF catalyses the conversion of phosphenol pyruvate and ADP to pyruvate and ATP, thus playing a central role in the metabolism and the transfer of energy (Al Zaid Siddiquee et al., 2004).

The next three sub-sections in this Chapter will look more closely at the pH and respiratory conditions for each individual organic acid, starting with acetic acid, then propionic acid and finally butyric acid.

4.1.3 All genes implicated by TraDIS as being required for fitness in acetic acid stress

This subsection will consider those genes which are implicated as being required for fitness in acetic acid stress only. Below at Figures 4.3 and 4.4 are the genes which were considered significant (in terms of frequency of transposon inserts before and after growth when compared against the control condition of culture grown at either pH 7 grown aerobically or anaerobically or pH 5.5 grown aerobically or anaerobically depending on the experimental growth condition).

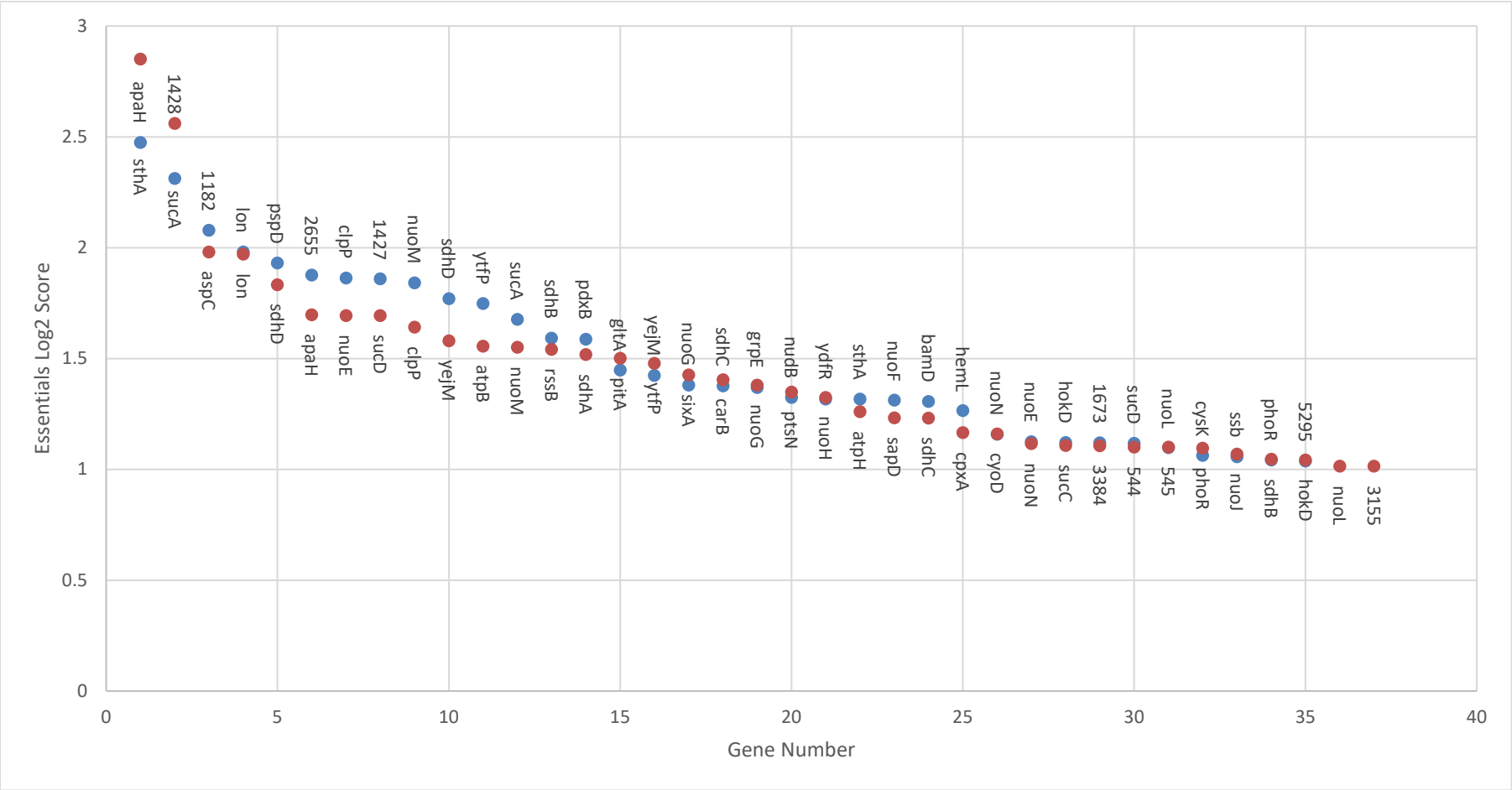


Figure 4.3 Genes where transposon frequency was significantly lower at the end of the aerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 4 mM acetic acid, red shows pH 7 with 40 mM acetic acid. Data is average of three replicates.

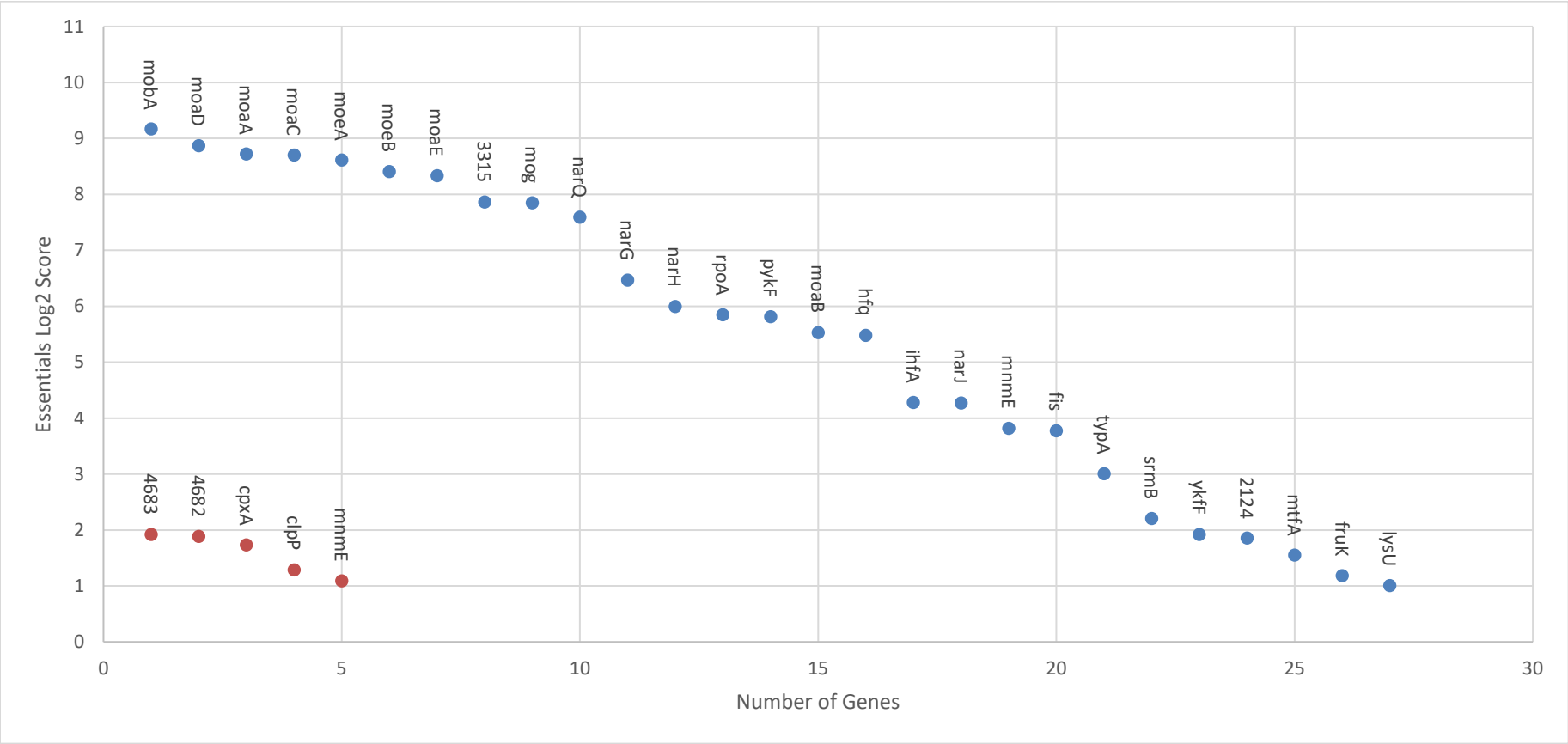


Figure 4.4 Genes where transposon frequency was significantly lower at the end of the anaerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 4 mM acetic acid, red shows pH 7 with 40 mM acetic acid. Data is average of three replicates

Figure 4.3 shows the genes for which loss of function was considered to significantly impact fitness in aerobic growth in the presence of acetic acid. The gene *atpH* is shown here (p-value 0.01), which does not appear in any other acid under any other condition. However, shared with propionic acid at pH 7 aerobic and in the list for pH 5.5 with no organic acid is *atpB,* and in propionic only are *atpA* and *atpD* (p-values all less than 0.0001) (see Figure 4.5). These genes encode subunits of the F_oF_1 ATP synthase which transports H^+ into the cell. It was shown in Chapter Three that this entire system is downregulated under organic acid stress (which is in agreement with published research). It is interesting, therefore, that interruption of various subunits of this system causes a growth defect in the cell under organic acid stress. Disruption of this complex, even if downregulated, may have a detrimental knock-on effect to other systems under organic acid stress.

Also in the list for acetic acid only, at pH 5.5 in aerobic growth, are *sdhABCD* (p-values all less than 0.0001) which are all the genes in the *sdh* operon. It is a repeated pattern to see all, or almost all, of certain operons in this data (as seen above with *nuo* and *moa*). This will be considered in more detail in the discussion at the end of this chapter. These genes encode subunits of succinate dehydrogenase and are directly involved in the TCA cycle.

RssB is the regulator of *rpoS* which, as previously discussed, is responsible for one of the known responses of *E. coli* to acid stress. Its inclusion in this list therefore is not surprising as loss of this regulator may impact on the cells ability to respond to acid challenge, however it is interesting that it appears only in the list for acetic acid and not the other two organic acids tested.

For acetic acid only, the interruption of *rpoA* caused a significant growth impairment, but only at pH 5.5 in anaerobic conditions. RpoA is an RNA polymerase subunit which, amongst other things, has been

shown to regulate the arginine decarboxylase system, involved in acid resistance (Shi and Bennett, 1994). This is another essential gene which may have a bias in terms of position of transposon insertions, whereby change in frequency is exaggerated by changes to already low numbers. The p-value for this result is high, at 0.12, and the use of a more conservative threshold would have removed this gene from the list of those being analysed as it would have been considered not significant.

The next sub-section will look more closely at those genes which are implicated in fitness in propionic acid only, which do not appear in the lists for the other organic acids.

4.1.4 All genes implicated by TraDIS as being required for fitness in propionic acid stress

This subsection will consider those genes which are implicated as being required for fitness in propionic acid stress only. Below at Figures 4.5 and 4.6 are the genes which were considered significant (in terms of frequency of transposon inserts before and after growth when compared against the control condition of either pH 7 aerobic or anaerobic or pH 5.5 aerobic or anaerobic depending on the experimental growth condition).

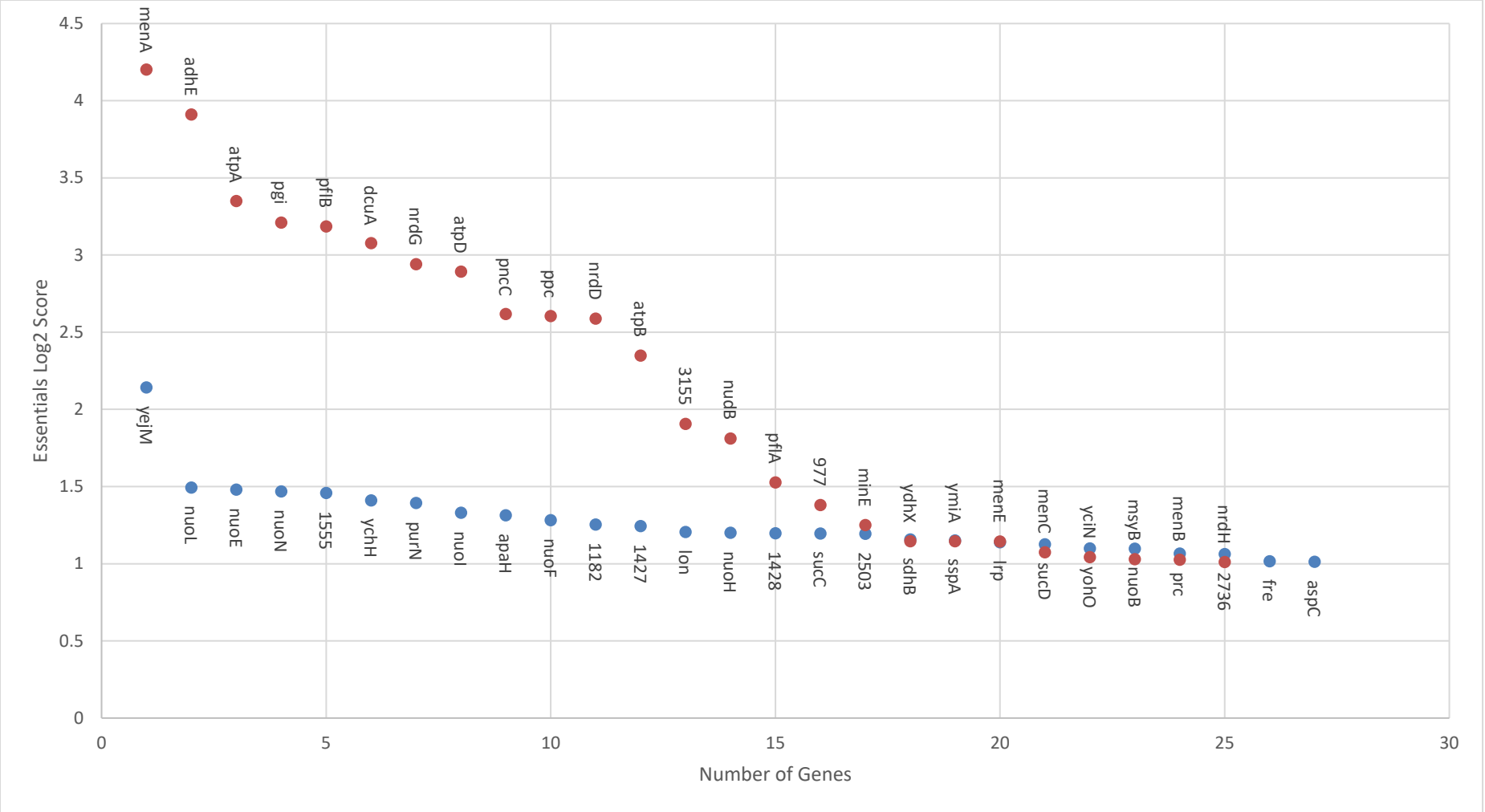


Figure 4.5 Genes where transposon frequency was significantly reduced at the end of the aerobic growth period compared to the start, when considered against the relevant control condition. Blue shows data for growth at pH 5.5 with 2 mM propionic acid, red shows pH 7 with 30 mM propionic acid. Data is average of three replicates.

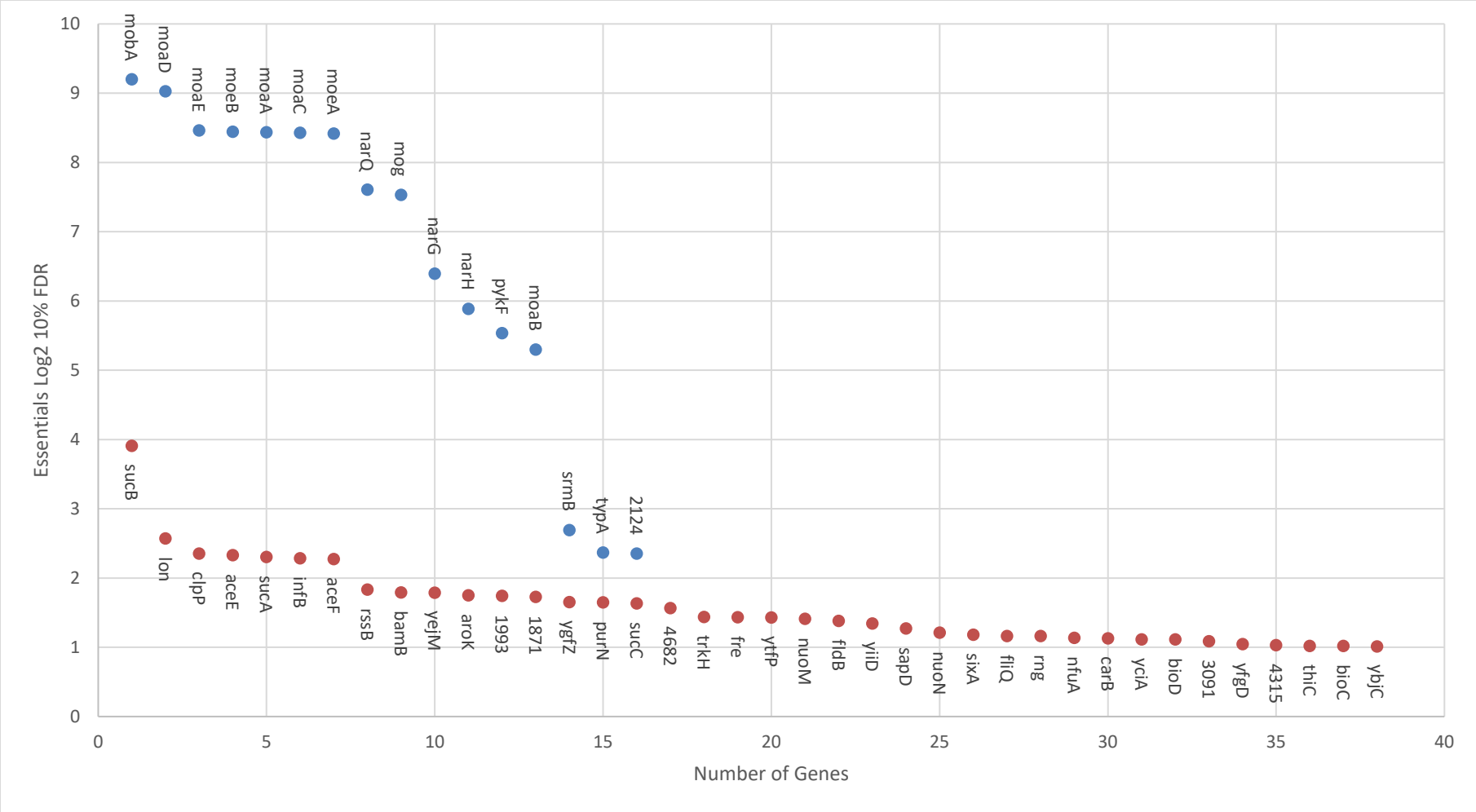


Figure 4.6 Genes where transposon frequency was significantly reduced at the end of the anaerobic growth period compared to the start, when considered against the relevant control condition. Blue shows data for growth at pH 5.5 with 2 mM propionic acid, red shows pH 7 with 30 mM propionic acid. Data is average of three replicates.

The data for propionic acid has delivered some unusual results. Included in Figure 4.5 for pH 7 aerobic are *menABCE* (p-values of 0.00003 or less) which are involved in anaerobic respiration, genes *nrdDGH* (p-value 0.09 or less) which are involved in DNA synthesis in anaerobic conditions, *dcuA* (p-value $0.16x10^{-84}$) which is involved in C4-dicarboxylate transport under aerobic, microaerobic and anaerobic conditions and *pflAB* (p-values less than 0.001) which is a pyruvate formate-lyase activating enzyme, active under both aerobic and anaerobic conditions. In data shown below at Section 4.2.4 all of these mutants are included in the list for propionic acid as showing improved fitness but under anaerobic conditions at pH 7 (as opposed to aerobic conditions here). This result is bizarre and there is no clear indication from the literature what may be happening in the presence of propionic acid and propionate at neutral pH during either aerobic or anaerobic growth which might be inducing this response. This result requires considerably more investigation.

The next section with consider those genes appearing only in the list for butyric acid.

4.1.5 All genes implicated by TraDIS as being required for fitness in butyric acid stress

This subsection will consider those genes which are implicated as being required for fitness in butyric acid stress only. Below at Figures 4.7 and 4.8 are the genes which were considered significant (in terms of frequency of transposon inserts before and after growth when compared against the control condition of either pH 7 aerobic or anaerobic or pH 5.5 aerobic or anaerobic depending on the experimental growth condition).

The gene, *mnmG,* shared between all four conditions with butyric acid, did not appear in the lists for either propionic and acetic acid. In butyric acid, shared between all conditions except pH 5.5 anaerobic was *mnmE*. Together, these two genes form a heterotetramer which is involved in tRNA modification. Modifications of tRNAs contribute to the fidelity and efficiency of mRNA translation and maintain stability of the codon-anticodon pairing. These two genes specifically are involved in decreasing +2 frameshift errors, ensuring correct translation (Brégeon et al., 2001; Shi et al., 2009). The inclusion of these genes in butyric acid suggests that in this condition there are high levels of errors in mRNA translation being caused by the presence of the organic acid. This ties in with the proteases discussed above at section 4.1.1 which implicate a high level of misfolded proteins, possibly as a result of poor translation.

The next section of this chapter will focus on the lists of genes whereby a transposon insert provided a growth advantage under the conditions tested.

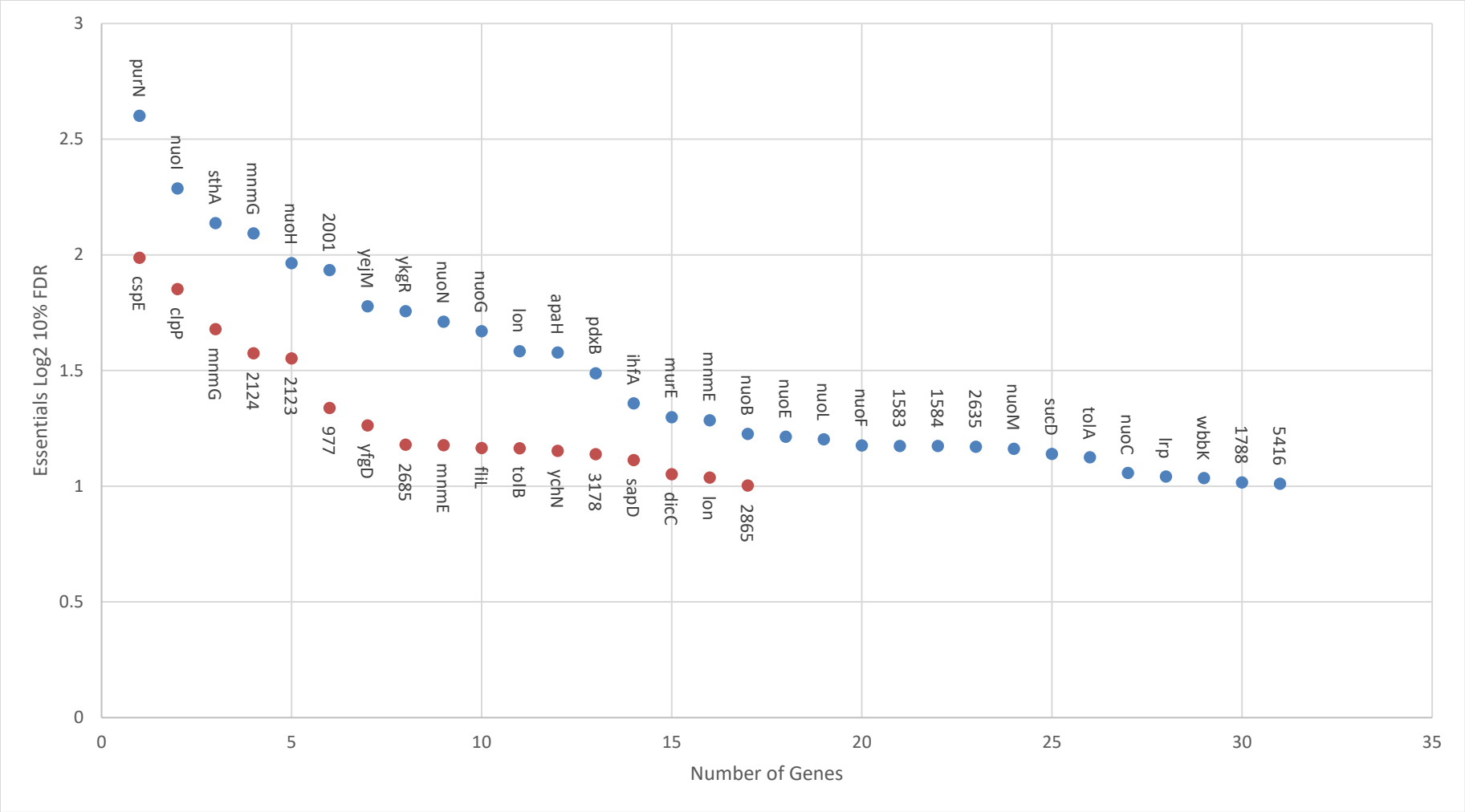


Figure 4.7 Genes where transposon frequency was significantly reduced at the end of the aerobic growth period compared to the start, when considered against the relevant control condition. Blue shows data for growth at pH 5.5 with 2 mM butyric acid, red shows pH 7 with 30 mM butyric acid. Data is average of three replicates.

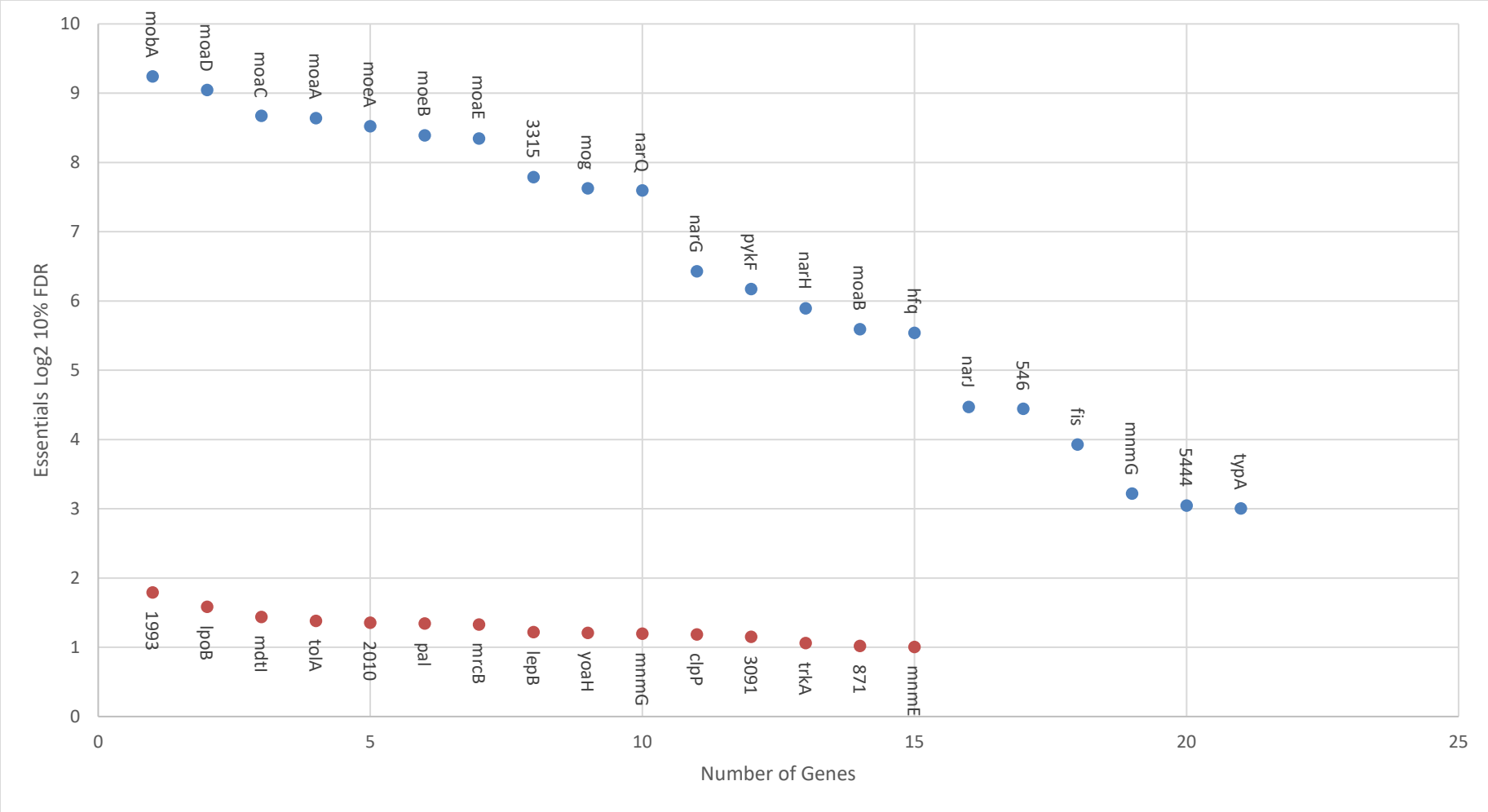


Figure 4.8 Genes where transposon frequency was significantly reduced at the end of the anaerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 2 mM butyric acid, red shows pH 7 with 30 mM butyric acid. Data is average of three replicates.

4.2 Genes where loss of function improves fitness under organic acid stress

This section and sub-sections will examine the results which showed that the presence of the transposon insert into a particular gene provided a growth advantage to the cell compared to the control condition (where those populations grown at pH 7 with organic acid are compared against pH 7 without added acid, pH 5.5 with organic acid compared against pH 5.5 without organic acid). First, this section will consider those genes appearing across the lists for all three acids, and will then move onto a brief look at one or two genes that showed significantly improved growth in each acid individually.

4.2.1 Genes with a decrease in insertion frequency during aerobic growth in organic acids

Table 4.7 below shows genes which, under the conditions tested, provided a growth advantage to mutants when function was prevented or stopped via the transposon insert and were shared at pH 5.5 during aerobic growth.

Acetic	Essentials Log2 Score	FDR adjusted P-value	Propionic	Essentials Log2 Score	FDR adjusted P-value	Butyric	Essentials Log2 Score	FDR adjusted P-value
aroE	-2.662	1.77E-18	*aroE*	-2.449	1.15E-11	*aroE*	-2.077	8.10E-13
bioH	-2.147	6.64E-20	*aroC*	-2.098	3.25E-07	*fruK*	-1.906	1.33E-23
bioF	-2.132	1.48E-08	*aroH*	-1.875	2.69E-06	*aroH*	-1.874	1.19E-08
aroC	-2.128	1.52E-10	*fruK*	-1.819	4.02E-15	*aroC*	-1.440	0.000537
aroH	-1.943	3.05E-06	*aroB*	-1.444	1.40E-06	*bioF*	-1.415	0.00011
cvpA	-1.797	0.00189	*bioF*	-1.312	0.034458	*cvpA*	-1.408	1.76E-06
bioA	-1.722	3.96E-14	*cvpA*	-1.309	0.000381	*bioH*	-1.248	3.42E-08
fruK	-1.548	7.38E-10	*bioH*	-1.276	0.000216	*bioA*	-1.240	4.98E-10
aroB	-1.422	1.7E-06	*bioA*	-1.134	0.006089	*aroB*	-1.050	0.00011

Table 4.7 Genes with a Log2 essentials score of -1 or more less which were shared across all organic acids at pH 5.5 aerobic (versus pH 5.5 aerobic control), ranked according to score with associated p-value.

The next four tables show the genes whereby loss of function improved fitness in pH 5.5 versus pH 7 (table 4.8), between all three organic acids and at pH 5.5 (table 4.9), those in common between the three organic acids (table 4.10) and in different combinations of organic acid and pH 5.5 (table 4.11).

pH 5.5 v pH 7 aerobic	**Essentials Log2 Score**	**FDR adjusted P-value**
panB	-1.939	7.04E-05
pdxB	-1.665	0.00171
mobA	-1.489	0.195167
2001	-1.304	0.072665
2503	-1.161	0.118339
aroA	-1.132	0.002116
794	-1.063	0.063781
ychF	-1.054	0.003158
surA	-1.026	0.138671
ppc	-1.013	1.71E-05
pabB	-1.012	1.31E-06

Table 4.8 Genes with a Log2 essentials score of -1 or less which where loss of function improved fitness in pH 5.5 aerobic alone (pH 5.5 v pH 7 aerobic control condition), ranked according to score with associated p-value.

a. Acetic, b. Propionic, c. Butyric d. pH 5.5	**Essentials Log2 Score**	**FDR adjusted P-value**
panB	a. -1.808, b. -1.215, c. -1.363, d. -1.939	a. 0.0001, b. 0.0002, c. 2.06E-07, d. 7.04E-05
aroA	a. -1.455, b. -1.079, c. -1.255, d. -1.132	a. 0.00000253, b. 0.0026, c. 7.61E-09, d. 0.0021

Table 4.9 Genes with a Log2 essentials score of -1 or more which were shared across all organic acids and pH 5.5 only in pH 5.5 aerobic (organic acids versus pH 5.5 aerobic control and pH 5.5 versus pH 7 control), ranked according to score with associated p-value.

Acetic	Essentials Log2 Score	FDR adjusted P-value	Propionic	Essentials Log2 Score	FDR adjusted P-value	Butyric	Essentials Log2 Score	FDR adjusted P-value
bioF	-2.462	4.97E-15	*bioF*	-2.986	1.01E-24	*nadA*	-2.139	5.70E-07
nadA	-2.254	9.36E-11	*bioH*	-2.730	1.40E-33	*bioF*	-2.077	2.87E-11
bioH	-2.011	4.97E-15	*bioD*	-2.695	1.37E-19	*bioH*	-1.776	1.89E-08
bioD	-1.902	1.28E-09	*bioA*	-2.578	9.41E-32	*aroE*	-1.737	1.97E-12
aroE	-1.820	9.79E-17	*aroE*	-2.533	1.12E-27	*bioC*	-1.705	1.10E-11
nadC	-1.815	0.000276	*aroB*	-2.298	2.25E-19	*bioD*	-1.587	8.79E-05
aroB	-1.788	7.53E-12	*bioC*	-2.160	1.43E-16	*aroH*	-1.510	5.66E-08
bioC	-1.778	6.07E-17	*aroH*	-2.002	8.45E-10	*aroB*	-1.480	3.61E-06
bioA	-1.767	6.98E-09	*nadA*	-1.724	0.000142	*bioA*	-1.465	0.000475
aroH	-1.506	1.92E-11	*nadC*	-1.213	0.060835	*nadC*	-1.411	0.013493
fruK	-1.355	5.94E-17	*fruK*	-1.195	4.45E-09	*fruK*	-1.365	8.36E-17

Table 4.10 Genes with a Log2 essentials score of -1 or less which were shared across all organic acids at pH 7 aerobic (versus pH 7 aerobic control), ranked according to score with associated p-value.

a. Acetic b. Butyric	Essentials Log2 Score	FDR adjusted P-value	a. Acetic, b. Propionic c. pH 5.5	Essentials Log2 Score	FDR adjusted P-value
ackA	a. -3.364, b. -2.173	a. 4.35E-12, b. 8.18E-06	*surA*	a. -1.804, b. -1.1421, c. -1.026	a. 7.63E-15, b. 0.0012, c. 0.139
pta	a. -2.669, b. -1.430	a. 4.34E-43, b. 1.25E-10	**a. Acetic b. pH 5.5**	**Essentials Log2 Score**	**FDR adjusted P-value**
panC	a. -2.405, b. -2.128	a. 6.19E-08, b. 0.0007	*ppc*	a. -1.207, b. -1.013	a. 6.03E-23, b. 1.71E-05
panD	a. -2.128, b. -1.398	a. 2.92E-07, b. 0.078			
nadB	a. -1.85, b. -1.75	a. 8.19E-16, b. 0.0069			
pgi	a. -1.72, b. -1.86	a. 9.60E-17, b. 5.98E-25			
ptsG	a. -1.260, b. -1.129	a. 2.47E-18, b. 1.02E-11			

Table 4.11 Genes in common across more than one acid with a Log2 essentials score of -1 or less at pH 7, aerobic (compared to a control of pH 7 aerobic) with associated p-value.

In the list of genes shared across acetic and butyric acid (table 4.11) are AckA-Pta. Both of these genes are downregulated in these growth conditions in the RNAseq data for all three acids, and the TraDIS results show an improvement in fitness in acetic and butyric acid when either one of the *ackA* or *pta* genes is interrupted. AckA-Pta catalyse a reversible reaction whereby acetyl CoA is converted into acetyl phosphate and then into acetate. Given this pathway is reversible, it can also metabolise acetate into acetyl CoA. Cells will metabolise glucose initially before switching to acetate metabolism and will switch to acetate metabolism once glucose is depleted. However, it has been shown that *E. coli* will simultaneously metabolise both. In these experiments then it would seem beneficial to the cell for these genes to be functional i.e. to catalyse the metabolism of acetate into acetyl CoA. However, as with other pathways that will be discussed in this Chapter, there appears to be some benefit to the cell of producing less acetyl CoA under these organic acid stress conditions. In the case of AckA-Pta it may be that loss of both reactions, both the production of additional acetate and the production of acetyl CoA is in some way beneficial to the cell (Enjalbert et al., 2017).

Within table 4.10 are the genes *aroABCDEH* which are all involved in the chorismate pathway which leads to synthesis of aromatic amino acids. What is interesting about this result is that other studies have shown that *aroABCE* mutants are unable to grow in M9 supplemented with glucose. However, that data was derived using the Keio library, which uses the background *E. coli* strain of BW25113, not EO 499 (as used here) (Baba et al. 2006). On that basis it is not a directly comparable result, and the uropathogenic strain of *E. coli* used for this TraDIS experiment may be utilising organic acids as an additional carbon source for this purpose. As has been discussed earlier in this Chapter, there is some uncertainty surrounding results in TraDIS experiments in genes which are otherwise considered essential. It is possible that interruption of any one of the subunits of this operon induces higher activity

of an alternative pathway to compensate. This may result in improvement in growth overall, however it would be expected that this would be replicated in the control conditions against which this data was compared. That being said, it is interesting that interrupting this pathway, and genes which have been shown to be conditionally essential in other strains of *E. coli,* should improve growth under organic acid stress. This is an interesting route of further investigation (Joyce et al., 2006; Patrick et al., 2007).

Also in table 4.10 are the genes *bioACDFH,* which are all involved in biotin metabolism which was briefly discussed in Chapter Three due to several of the genes of this pathway being both up- and down-regulated. These genes were shown to be upregulated in propionic and butyric acid under anaerobic conditions in Chapter Three. The genes *bioC* and *bioD* are also included in the lists in Section 4.1.4 above as showing reduced fitness in propionic acid at pH 7 in anaerobic conditions. The list of *bio* genes here, however, has been generated under aerobic conditions, which is in contradiction to the anaerobic conditions where fitness was impaired. Biotin is an essential cofactor in carboxylase enzymes including acetyl-CoA carboxylase and is utilised by the cell in production of long-chain fatty acids via the carboxylation of acetyl Co-A. It appears that disruption of multiple parts of this pathway is beneficial to the cell under organic acid stress (Blanchard et al., 1999).

In table 4.11, one of the genes shared between acetic and butyric acid is *pgi.* Though not in the tables above, shared between propionic and acetic acids at pH 7 aerobic and butyric acid at pH 5.5 aerobic is *zwf* (all p-values less than 0.005). The gene *zwf*, is a glucose-6-phosphate dehydrogenase and the first enzyme in the pentose phosphate pathway. This is coupled with *pgi* which is a phosphoglucoisomerase and part of the pathway which produces 30 – 40% of NADH when glucose is the sole carbon source. Strains lacking in *pgi* show reduced growth in media with glucose as the sole carbon source (as per the control conditions used here). In organic acids therefore there is an additional carbon source, which may

counter this problem. On that basis, rather than the interruption of the gene being beneficial in organic acid stress, it may be that compared to the control condition, the addition of the alternative carbon source has improved growth (Olavarria et al., 2014; Zhao, 2004).

The next section will consider the genes whereby loss of function improved growth in anaerobic conditions.

4.2.2 Genes shared across the three organic acids in anaerobic conditions

The tables below show the genes where loss of function improved fitness in organic acids relative to the control condition. Due to very low numbers of genes, 9 in total across all three acids, there were no genes in common at pH 5.5 in the presence of organic acids in anaerobic growth conditions.

Acetic	**Essentials Log2 Score**	**FDR adjusted P-value**	**Propionic**	**Essentials Log2 Score**	**FDR adjusted P-value**	**Butyric**	**Essentials Log2 Score**	**FDR adjusted P-value**
ptsG	-1.27227	0.145517	*ptsG*	-2.99347	3.03E-24	*arnE*	-1.16268	0.069416
arnE	-1.07644	0.045395	*arnE*	-1.02752	0.028562	*ptsG*	-1.01828	0.046927

Table 4.12 Genes with a Log2 essentials score of -1 or more less which were shared across all organic acids at pH 7 anaerobic (versus pH 7 anaerobic control), ranked according to score with associated p-value.

a. Acetic b. Butyric	**Essentials Log2 Score**	**FDR adjusted P-value**	**a. Acetic b. pH 5.5**	**Essentials Log2 Score**	**FDR adjusted P-value**
ydiZ	a. -1.0211, b. -1.2496	a. 0.0288, b. 2.16E-03	*moaD*	a. -1.061, b. -9.974	a. 0.0368, b. 0.096
yciN	a. -1.115, b. -1.359	a. 0.158, b. 0.0166	*prc*	a. -1.660, b. -1.333	a. 5.01E-10, b. 0.131
aroA	a. -1.1625, b. -1.1022	a. 0.002, b. 0.001			
surA	a. -1.391, b. -1.287	a. 0.004, b. 0.099	**a. Acetic b. butyric c. pH 5.5**	**Essentials Log2 Score**	**FDR adjusted P-value**
nrdD	a. -2.045, b. -1.345	a. 8.69E-13, b. 1.70E-04	*nadA*	a. -2.486, b. -2.351, c. -2.440	a. 6.37E-07, b. 4.18E-06, c. 0.096
rpmF	a. -2.310, b. -2.045	a. 0.00068, b. 0.037			
a. Acetic b. Propionic	**Essentials Log2 Score**	**FDR adjusted P-value**	**a. Propionic, b. butyric c. pH 5.5**	**Essentials Log2 Score**	**FDR adjusted P-value**
arcB	a. -1.137, b. -1.398	a. 1.70E-06, b. 3.59E-08	*panE*	a. -1.515, b. -1.368, c. -1.971	a. 2.02E-06, b. 0.003, c. 0.157
metJ	a. -1.257, b. -1.716	a. 0.134, b. 0.0003			
nadC	a. -1.464, b. -1.385	a. 0.0099, b. 0.024			
ydfZ	a. -1.656, b. -1.239	a. 0.0009, b. 0.182			
nadB	a. -1.716, b. -1.705	a. 2.93E-05, b. 0.0005			

Table 4.13 Genes in common across more than one acid with a Log2 essentials score of -1 or less at pH 7 with organic acid anaerobic (compared to a control of pH 7 anaerobic) with associated p-value.

pH 5.5 v pH 7 anaerobic	Essentials Log2 Score	FDR adjusted P-value
moaD	-9.974	0.095934
mobA	-9.244	0.119479
moaC	-8.563	1.19E-01
moaA	-8.150	0.130486
moeA	-8.102	0.130486
moeB	-8.017	0.130486
moaE	-7.761	0.163721
narQ	-7.415	0.130486
mog	-7.303	0.157075
narG	-6.523	0.157075
narH	-5.794	0.173054
pykF	-5.510	0.173054
moaB	-5.232	0.173054
fis	-4.575	0.119479
nadA	-2.440	0.095934
panE	-1.971	0.157075
prc	-1.333	0.130486

Table 4.14 Genes with a Log2 essentials score of -1 or less which where loss of function improved fitness in pH 5.5 anaerobic alone (pH 5.5 v pH 7 anaerobic control condition), ranked according to score with associated p-value.

NadABC are all involved in NAD biosynthesis. In acetic acid, *nadC* mutants showed improved fitness in all pH and respiratory conditions, with *nadA* and *nadB* in the lists for all except pH 5.5 in anaerobic conditions. In propionic acid, *nadC* is showed improved fitness in all conditions except pH 5.5 in anaerobic conditions, *nadA* at pH 7 and pH 5.5 in aerobic conditions and *nadB* in pH 7 in anaerobic conditions. For butyric acid, *nadB* and *nadC* are in the list for pH 7 in aerobic conditions and *nadA* for pH 7 in aerobic and anaerobic conditions. So, for all but one gene in one acid, none of these genes appear in the lists generated at pH 5.5 in anaerobic conditions. This points to the disruption of the use of NAD in redox reactions as being beneficial under organic acid stress, perhaps by preventing additional H^+ being released into the cytoplasm during the reduction of NAD to NADH. However, this may not be the case in growth at pH 5.5 in anaerobic conditions. It may be that at pH 5.5 in the presence of organic acids in

anaerobic conditions there are higher energy requirements for the cell and so disruption of this pathway is proportionally advantageous despite this additional H^+ in the cytoplasm.

On the list as improving fitness through loss of function in acetic and butyric acid at pH 7 and propionic acid at pH 5.5 is *surA* (p-values 0.09 or less). SurA is an isomerase which is required for the efficient folding of outer membrane proteins such as OmpF and OmpC, acid induced porins, which are involved in acid response (see Chapter One). On that basis, therefore, loss of this gene via transposon insert would seem counter to successful growth in organic acids due to a decrease in the efficient folding of these porins. However, here loss of *surA* is showing a fitness advantage. A study by Lazar et al., (1998) showed that *rpoS* compensates for loss of *surA*, possibly through induction of alternative chaperones. It could be that loss of *surA* via transposon insertion induces higher than normal *rpoS* activity, which may aid growth in organic acid conditions, leading to the result seen here.

Almost all of the genes listed above in table 4.14 showing an improvement in fitness in pH 5.5 alone with no added organic acid are the same as those listed in table 4.4 which shows genes where loss of function caused a reduction in fitness across all three organic acids at pH 5.5 under anaerobic conditions. When there are no organic acids present, the loss of function of these genes improves fitness, whereas when organic acids are present this has a significant impact on fitness of the cells. This result requires more investigation.

The next three subsections will focus on genes which appear in the list for each acid individually. This Chapter will then conclude with a brief discussion.

4.2.3 All genes implicated by TraDIS as conferring a fitness advantage in acetic acid stress when insertion present i.e. loss of function leads to improved fitness

The two figures below 4.9 and 4.10 show the genes where mutants showed a significant improvement in fitness in acetic acid only. The first figure, 4.9 shows aerobic conditions and the second figure, 4.10, shows anaerobic.

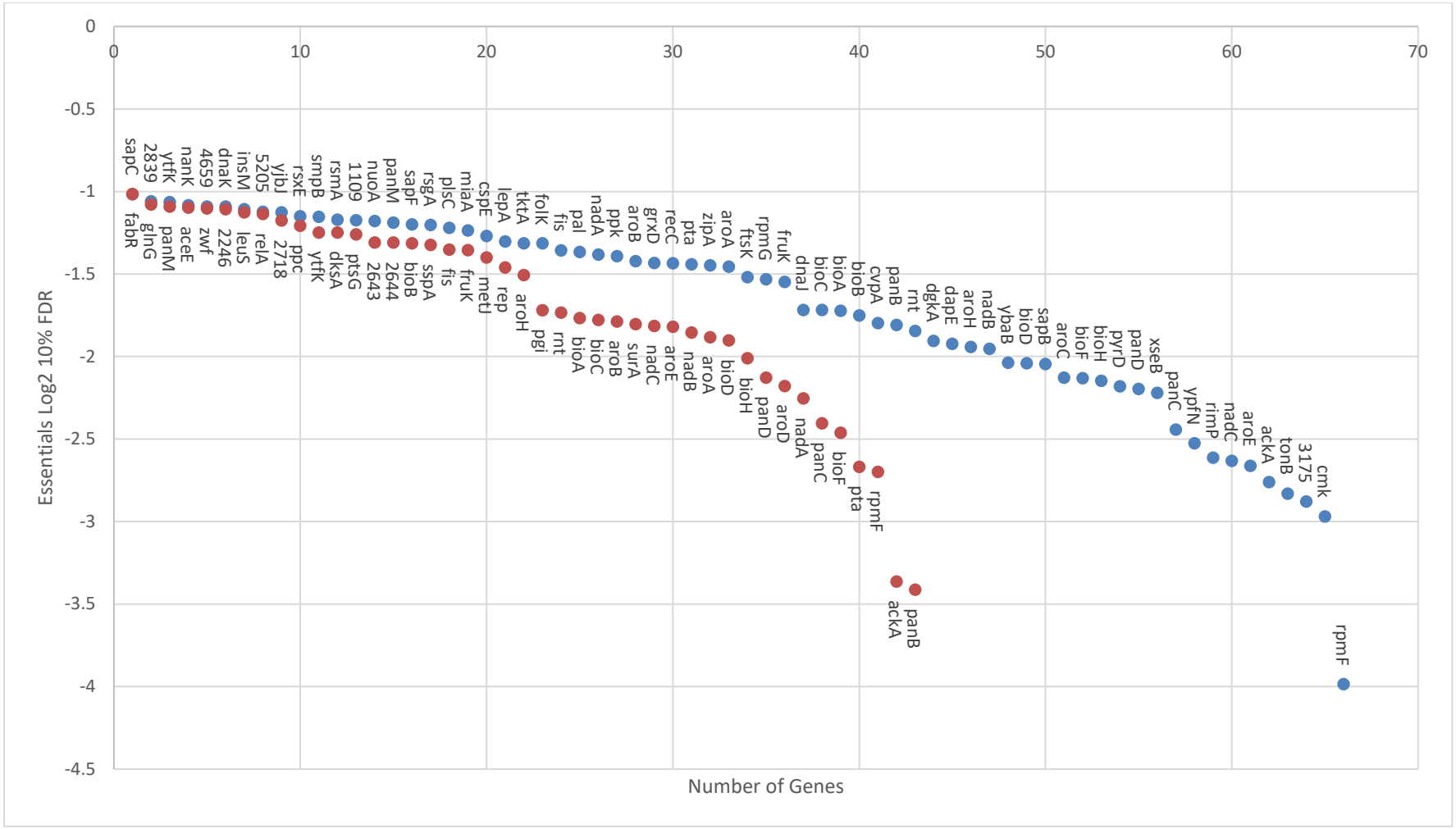


Figure 4.9 Genes where transposon frequency was significantly higher at the end of the aerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 4 mM acetic acid, red shows pH 7 with 40 mM acetic acid. Data is average of three replicates.

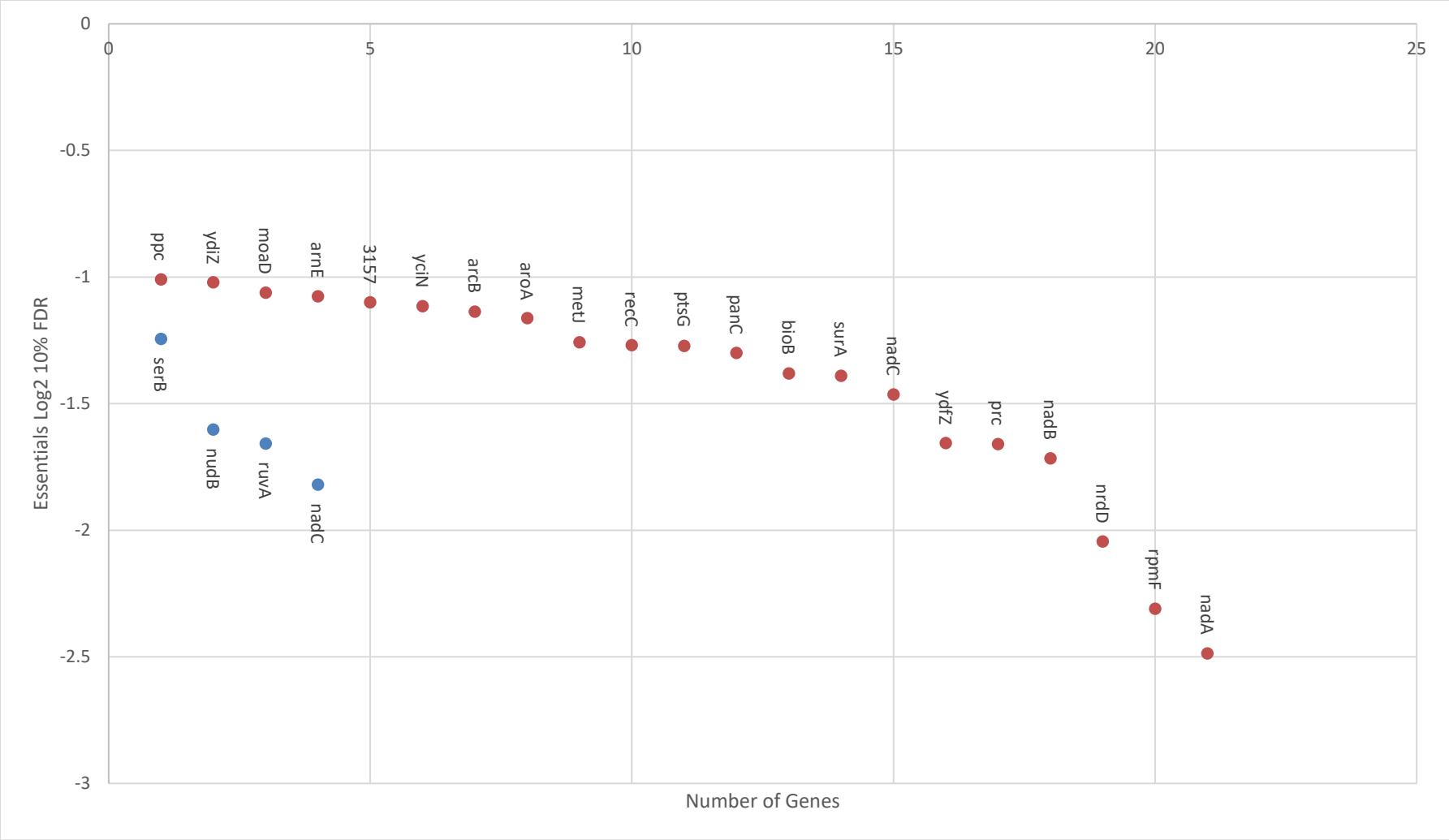


Figure 4.10 Genes where transposon frequency was significantly higher at the end of the anaerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 4 mM acetic acid, red shows pH 7 with 40 mM acetic acid. Data is average of three replicates.

In the list for acetic acid only is the gene *panM, panCD* are included in both acetic and butyric and *panE* in propionic, butyric acid and pH 5.5 only. In the list as being shared across all three acids is *panB*. All of these genes are involved in the biosynthesis of pantothenate which is involved in the production of coenzyme A. It is interesting that these sorts of pathways, like biotin metabolism, are coming up repeatedly in this analysis. The downregulation or interruption of pathways which result in the production of vital growth factors may be beneficial to the bacteria under organic acid stress, possibly via the use of alternative pathways of achieving the same end which utilise organic acid anions thereby more quickly remedying the internal stress on the cell. It may be that changes in the rates of acetate metabolism is, overall, somehow beneficial to the cell in organic acid stress. This result demonstrates the complex nature of cellular metabolism and it is not possible to fully elucidate the full picture of what is happening on a cellular level from looking at these individual genes in the lists generated here. This requires a wider network approach, which may be undertaken with this data at a later date.

As mentioned above, *dnaK* and *dnaJ* are in the list for acetic acid at pH 5.5 in aerobic growth conditions. As discussed previously, this may indicate a higher level of protein misfolding which, when interrupted, results in less damage-control having to be utilised by the cell under these conditions and minimising energy expenditure on these processes.

The next subsection will focus on the genes associated with propionic acid only.

4.2.4 All genes implicated by TraDIS as conferring a fitness advantage in propionic acid stress when insertion present i.e. loss of function leads to improved fitness

The two figures below 4.11 and 4.12 show the genes where mutants showed a significant improvement in fitness in acetic acid only. The first figure, 4.11 shows aerobic conditions and the second figure, 4.12, shows anaerobic.

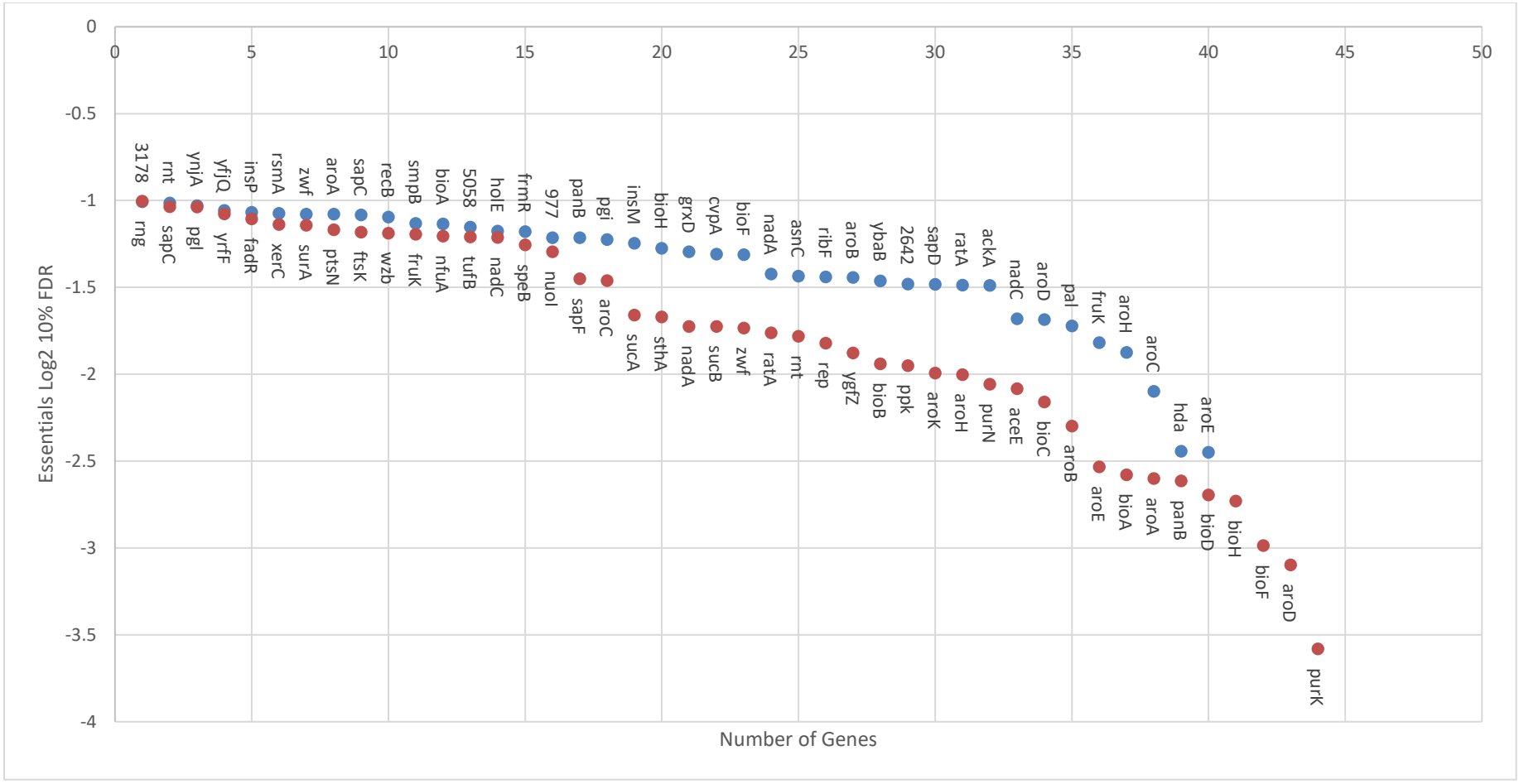


Figure 4.11 Genes where transposon frequency was significantly higher at the end of the aerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 2 mM propionic acid, red shows pH 7 with 30 mM propionic acid. Data is average of three replicates.

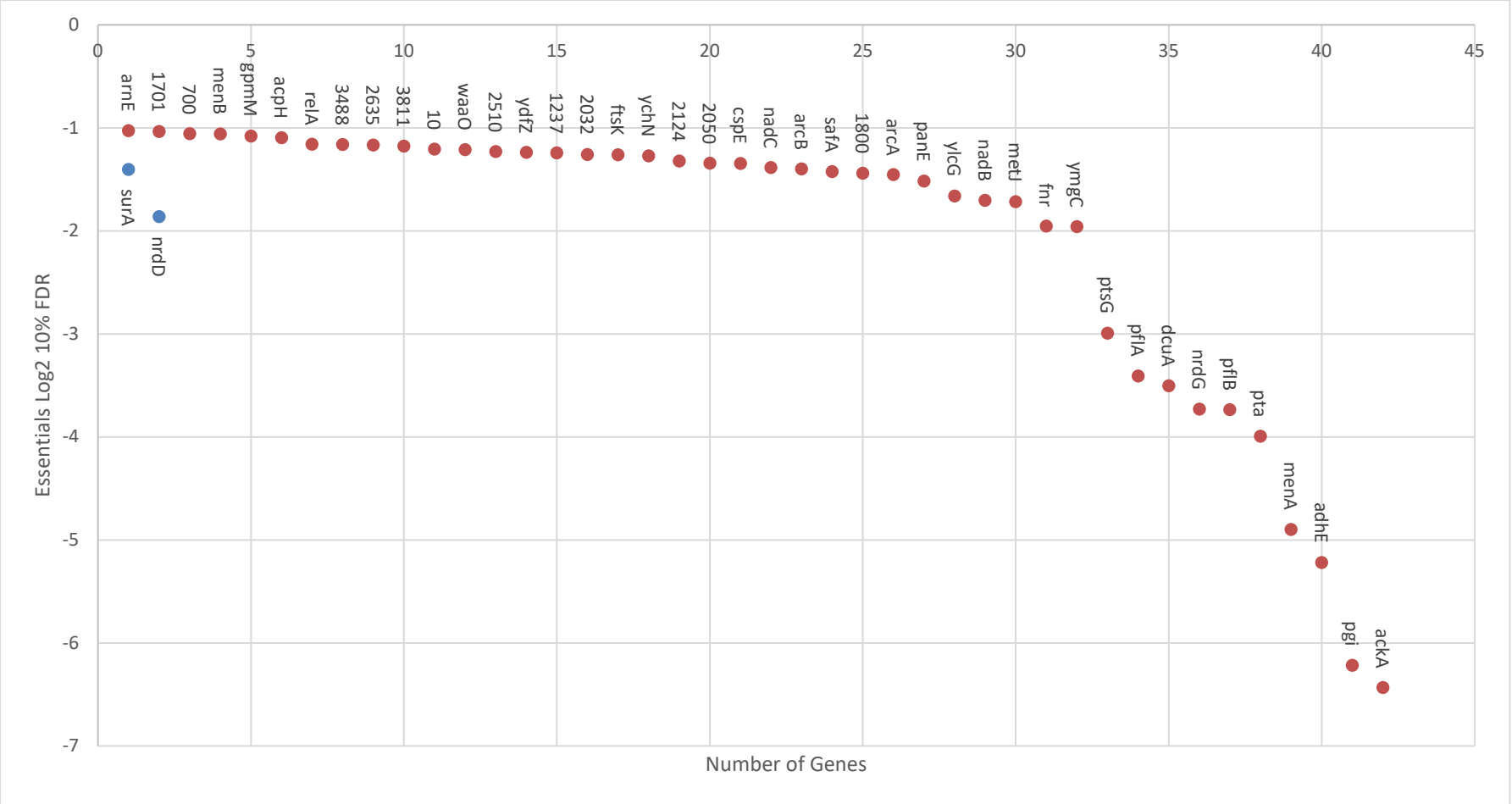


Figure 4.12 Genes where transposon frequency was significantly higher at the end of the anaerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 2 mM propionic acid, red shows pH 7 with 30 mM propionic acid. Data is average of three replicates.

The two genes *pflAB* appear here in the list for propionic acid only at pH 7 under anaerobic growth conditions. As noted above at section 4.1.4, these genes are in the list at pH 7 under aerobic conditions in the presence of propionic acid as causing a significant reduction in fitness. The genes *pflAB* are involved in production of acetyl CoA from pyruvate under aerobic, microaerobic and anaerobic conditions. Again, interruption of this system could result in less acetate metabolism by the cell, relieving stress. In this case, that may then result in higher rates of metabolism of propionate instead, thereby relieving propionate stress on the cell. However, if that were the case, it is interesting that a) these genes to not appear in the list for butyric acid and b) that they should be in the list for propionic acid under aerobic conditions as causing a growth deficit when interrupted by a transposon, but here as conferring a fitness advantage when function is lost under anaerobic conditions. This result requires more investigation because, as mentioned previously, these individual gene snapshots do not reveal the full picture.

The two component system ArcAB is in the list for propionic acid, *arcA* in propionic only and *arcB* in both propionic and acetic. ArcAB are in the list at pH 7 in anaerobic conditions, however not pH 5.5 anaerobic. Also in the list for genes at pH 7 in anaerobic conditions is *fnr,* also in propionic only. The genes *arcAB* are a two-component system and major transcriptional regulator under anaerobic and microaerobic conditions. Likewise, Fnr is also a major regulator under anaerobic conditions. It is interesting that these genes have come up as conferring a growth advantage when interrupted in mutants grown in propionic acid, but only at pH 7, and not in the presence of the other organic acids tested. Given that the overall lists for propionic have higher numbers of genes than the other two acids, it is clear that the addition of propionate has resulted in an impact on the cell not replicated by the other two organic acids and which requires further, more detailed, investigation.

4.2.5 All genes implicated by TraDIS as conferring a fitness advantage in butyric acid stress when insertion present i.e. loss of function leads to improved fitness

Last, the list of genes generated under butyric acid will be considered in isolation. The two figures below 4.13 and 4.14 show the genes where mutants showed a significant improvement in fitness in acetic acid only. The first figure, 4.9 shows aerobic conditions and the second figure, 4.10, shows anaerobic

In the list for butyric acid only, under pH 5.5 aerobic are *atpABD*. These genes have been discussed earlier in this chapter at section 4.1.3 as being required under acetic and propionic acid. That the reverse is true in butyric acid requires further investigation. These genes were significantly downregulated in organic acid stress and so, under those circumstances, it is perhaps not surprising that by interrupting gene function that growth is improved. Clearly, the relationship between up or down regulation of pathways and loss of those pathways is a complicated one with different outcomes for the cell. As with propionic acid, there are features of the results for all three organic acids which point to some quite unique effects on cellular processes imparted by all three.

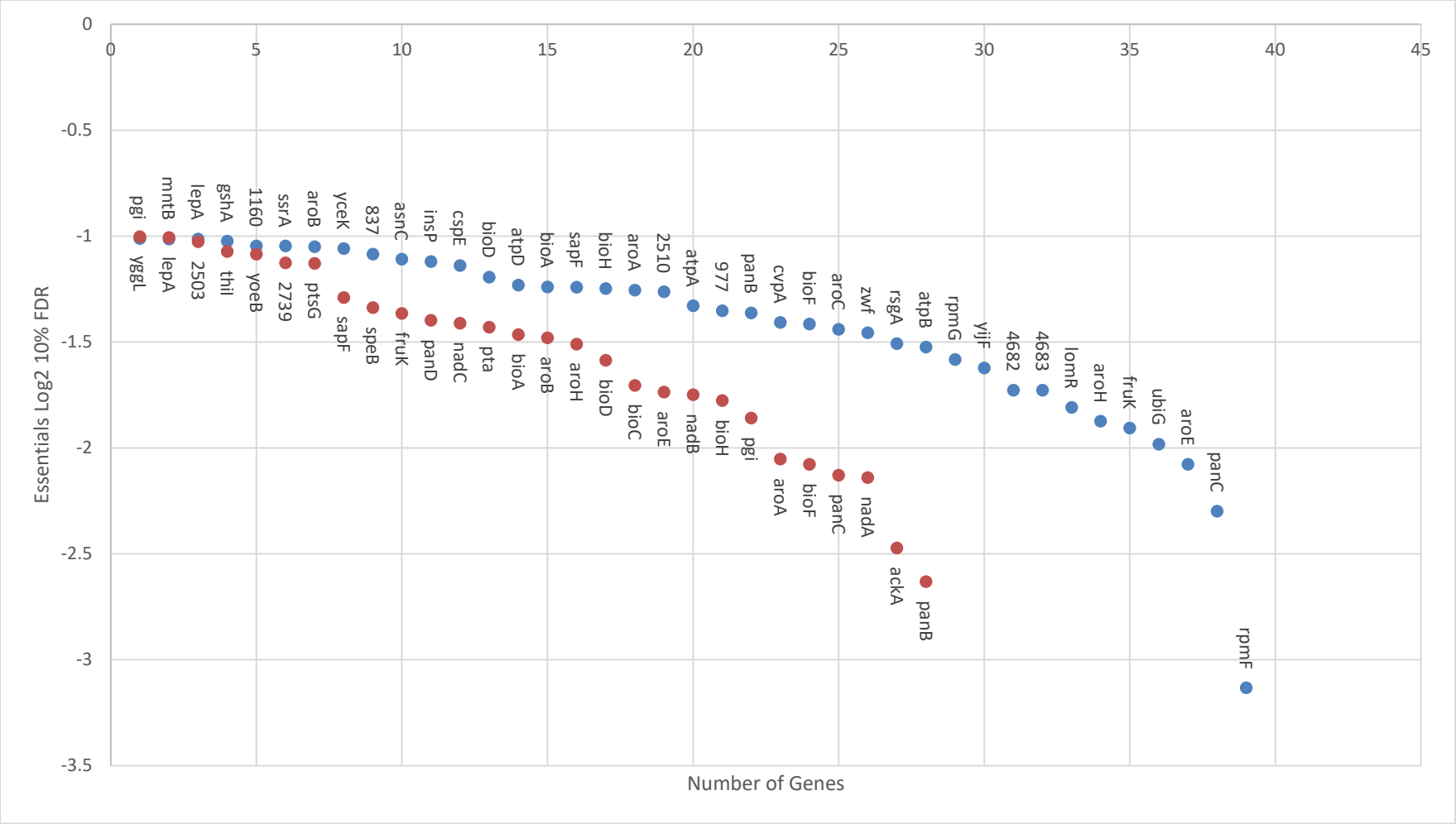


Figure 4.13 Genes where transposon frequency was significantly higher at the end of the aerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 2 mM butyric acid, red shows pH 7 with 30 mM butyric acid. Data is average of three replicates.

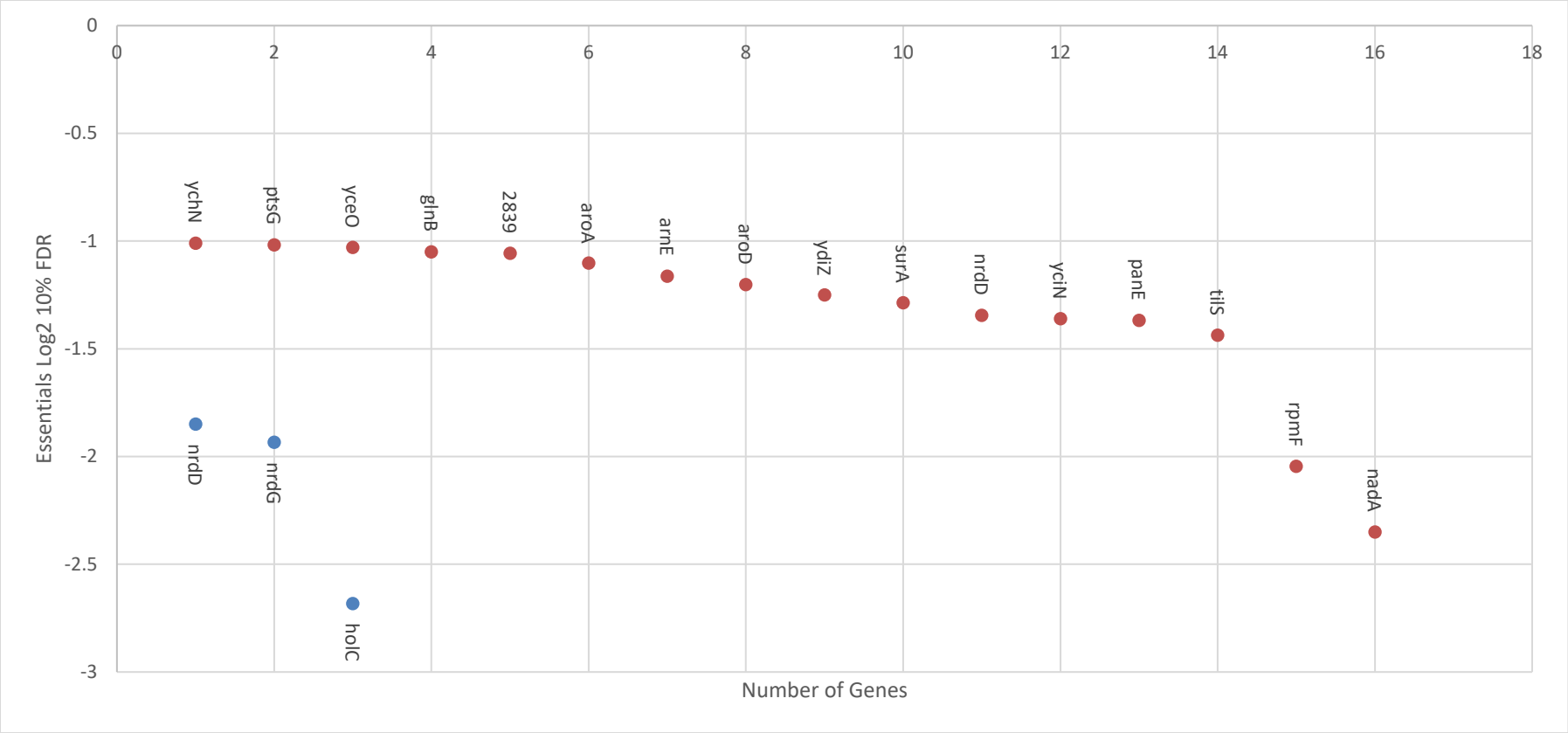


Figure 4.14 Genes where transposon frequency was significantly higher at the end of the anaerobic growth period compared to the start, when considered against the relevant control condition. Blue shows pH 5.5 with 2 mM butyric acid, red shows pH 7 with 30 mM butyric acid. Data is average of three replicates.

4.3 Discussion

A large proportion of genes which are important for fitness seem to be involved in respiration and maintenance of the proton motive force. Given the influx of H^+ and anions brought about by the presence of organic acids, is it possible that the cell is utilising these systems in order to maintain cellular homeostasis. By limiting certain pathways, such as the *nuo* pathway by way of transposon insertions into vital subunits, the ability to remove protons from the cytoplasm is reduced, thereby limiting its ability to both reduce internal pH and maintain proton motive force. By inhibiting respiration and the release of energy there are implications on other cellular processes which could limit the ability to metabolise excess organic acid anions, and/or remove them from the cell through pumps which require high energy usage. This could all play a role in limiting the fitness of these mutants to respond to their environment, but not preventing them entirely from growing.

There were a number of operons which appeared in the lists, both as causing a fitness advantage and disadvantage when a single gene within the operon lost its functionality. Operons in *E. coli* are transcribed into a single long polycistronic mRNA, which means the peptides are translated one after the other by a single ribosome. Given the transposon would interrupt transcription of polycistronic mRNA, this would result in incomplete translation of the entire operon and disrupt function of all genes within it. Therefore, a transposon insert into any part of the *sdh* operon, for example, has resulted in all subunits being included in the lists here in TraDIS. It is not possible, therefore, to state that loss of one of the genes has a greater impact on cellular function than any of the others using this method (Madigan et al., 2009).

Mutations which impacted on the processes of protein folding, repair and removal were also seen to decrease fitness. This implies high levels of misfolded proteins under organic acid stress. There is a possibility that the presence of organic acids either via the organic acid anions or because of low cytoplasmic pH is causing incorrect and inaccurate translation which results in protein misfolding.

Genes whereby loss of function provides a fitness advantage are diverse, but there are similarities wherein pathways which metabolise acetate are not only not required under these conditions, but the loss of these pathways allows higher levels of growth than when they are present. Given the bacteria need to reduce acetate pools within the cytoplasm rather than increase or maintain them this seems strange. In the case of propionic and butyric acid, it is possible that the cells are instead utilising these anions as a source of carbon over acetate, and that partial loss of acetate metabolising pathways (via the loss of function of a single gene) forces the cell to switch carbon source, relieving stress. In the case of acetate, the reasons are more obscure and it is difficult to propose a reasonable hypothesis.

There were higher numbers of mutants, both where fitness was impaired and where it was improved, in propionic acid, and some of those results require more investigation in order to fully understand their presence on the list. Particularly interesting was the presence of global regulators under anaerobic conditions whereby loss of function appeared to improve growth, but also the presence of genes which are involved in anaerobic respiration appearing in the list for propionic acid under aerobic growth conditions. It may be that the presence of propionate in high concentrations requires higher oxygen usage by the cells, forcing them to switch to anaerobic respiration sooner than in acetic and butyric acid. If this was the case, however, one would expect to see the same genes appear in the list for the anaerobic conditions as well showing impaired growth. The genes do appear in the list for anaerobic growth conditions as reducing fitness, but are in fact shown as improving fitness when function is lost. It appears that propionate is having some impact on the cell which is unknown at this time and is grounds for further research.

The results discussed in this section will be considered again in Chapter Six when bringing together the results from Chapter Three and the next chapter, which will investigate the mutations seen in populations evolved at pH 5.5 and at pH 5.5 in the presence of acetic acid.

Chapter Five:

Long-term laboratory-based evolution of *E. coli* ST131 EO499

Chapter 5: Long-term Lab Based Evolution of *E. coli* ST131 EO 499

5.1 Experimental Bacterial Evolution

To examine further whether a core set of genes are responsible for the organic acid stress response in *E. coli* ST131 EO 499, the method of long-term lab-based evolution was utilised. The intention of this study was to determine whether the same genes that are shown to be up-regulated and down-regulated in RNAseq, and those which are shown to confer or decrease fitness in *E. coli* during the course of the TraDIS experiment, are the same genes which have been mutated or deleted in lines of *E. coli* that have been evolved from the same strain under the same conditions as those used for part of the RNAseq and TraDIS experiments.

Bacteria and fungi have been used for several decades as model organisms for laboratory based experimental evolution (Helling et al., 1987; Paquin and Adams, 1983). This is because bacteria have a short generation time and a large population, which makes it possible to study evolution over hundreds and thousands of generations. Also, the advent of cheap sequencing technology means that bacterial genomes can be sequenced quickly and cheaply in order to study the phenotype of the evolved culture and its genotype simultaneously, which makes the identification of mutations in the evolving population relatively uncomplicated. Consequently bacterial evolution experiments are being used in an attempt to answer questions about evolution which include whether the same selective pressure produces the same evolutionary response. Also, whether the same fitness landscape can be achieved via multiple genetic pathways and whether evolutionary process can be predicted. Ultimately, this kind of knowledge could be used to predict an evolutionary time course, allowing researchers to predict bacterial evolution to other situations, such as antimicrobial resistance or for use in industry to metabolise or produce industrially useful compounds (Lenski et al., 2015).

Long-term evolution experiments provide a means to look at how relatively mild selective pressures, such as low concentrations of weak organic acid, may result in improved fitness in response to that pressure. This can happen over tens, hundreds or thousands of generations. The maintenance of a fossil record allows the study of gradual changes over the timeframe of the experiment, which provides for the pinpointing of changes in the fitness landscape.

The most famous long-term evolution experiment is that of Richard Lenski, who has been evolving *E. coli* for 25 years, over more than 60,000 generations. The populations have been exposed to limited glucose, forcing the use of alternative carbon sources. At most recent publication, Lenski et al., (2015) report that even after 40,000 generations, there is still increasing fitness in the evolved populations, with an increase between 40,000 and 50,000 generations and the same increase again between 50,000 and 60,000 generations. This raises an important question whether bacteria can improve their fitness to an unchanging environment indefinitely. If there is a 'limit' to fitness, when is this reached and how is it determined? There are limits to how results from an experiment of this design can be applied to 'real world' situations in that there would be almost no scenarios in nature where a bacterial strain would be in this kind of rigidly defined environment for this length of time. However, in terms of evolving populations for use in industry, the implications could be more directly applied.

5.1.1 Short-term evolution of *E. coli*

In terms of *E. coli* that has been experimentally evolved to withstand acid stress, the literature is limited. Johnson et al., (2014) used a different approach in order to elucidate an acid resistant phenotype. Rather than gradually evolve *E. coli* MG1655 over hundreds or thousands of generations to increasingly acidic conditions, the bacteria were exposed to bactericidal levels of hydrochloric acid in LB medium brought to pH 2.5. This was repeated for three weeks, and was ended once all cultures showed some level of acid resistance. This method produced five evolved populations with a much

greater acid resistant phenotype compared to their wild type ancestor. The evolved strains showed 59% to 37% survival at pH 2.5, where the wild type ancestor showed just 1% survival. The evolved strains were also able to survive for up to 2 hours at pH 1.5, whereas the ancestor did not survive at all. A single clone taken from each population was sequenced, which revealed that all had different mutations in *evgS*, which provided constitutive activity. As discussed in Chapter One, EvgS is part of the EvgAS two component system which activates the GAD AR2 system at pH 5.5, which allows for survival in the highly acidic environment of the stomach. When the mutations which conferred *evgS* with constitutive activity were introduced into the ancestor strain, the highly acid resistant phenotype seen in the evolved population was reproduced. This method was very useful for pinpointing a particular regulator of significance under extremely acidic conditions, which provided for further, more focused study of that one particular gene and subsequently the AR2 network as a whole.

5.1.2 Long-term evolution of *E. coli*

In terms of long-term evolution over hundreds of generations, Harden et al., (2015) evolved 24 populations of *E. coli* K-12 from a single common ancestor at pH 4.8 for 730 generations and then pH 4.6 for 2000 generations. By generation 730, all evolved lines showed increased fitness at pH 4.6 compared to the wild type ancestor. By the 2000th generation, all populations showed growth advantage compared with the 730th generation. There was no growth disadvantage at pH 7 or at pH 9. This implies that the evolved fitness to pH 4.6 was acid-specific.

Isolates from the 2000th generation were screened for degradative lysine decarboxylase (CadA), which was negative. This is very surprising given that AR4 is lysine-dependent. In the ancestor strain the lysine dependent acid response pathway was upregulated when grown at pH 4.6. The loss of lysine decarboxylase activity was gradual, with lower levels of activity being recorded in more isolates as generation time increased. However, when these clones were incubated for several days

lysine-decarboxylase levels were detected. This implies that *cadA* had not been deleted or lost function through mutation, but was in fact being down-regulated. This surprising result remains to be explained. However, the fact that all 24 populations demonstrated a down-regulation of this system implies a fitness advantage, rather than disadvantage, to long-term evolution at pH 4.6 (Harden et al., 2015).

The researchers identified clones with mutations in *adiY*, part of the arginine-dependent acid resistance pathway. Other strains had mutations in the *nuo* operon, which codes for NADH: ubiquinone oxidoreductase, part of the electron transport chain. A number of clones had mutations in a subunit of the RNA polymerase (RNAP) holoenzyme, specifically *rpoB* and *rpoC*. Mutations in RNAP may alter rates of transcription of acid-inducible genes. This is being investigated further by the researchers (Harden et al., 2015).

Of particular importance to this thesis is the observation that clones from four of the evolved populations were tested for improved fitness at pH 6.5 in the presence of benzoate, which would depress cytoplasmic pH. None of the clones showed improved fitness in this condition. This implies that any observed increase of fitness to acidic pH was not the result of an increased tolerance to acidic cytoplasmic pH (Harden et al., 2015). It should be noted that the populations were not evolved under organic acid stress, so although this result is interesting, its direct relevance is limited.

A study by Sen et al., (unpublished) also from the Lund group, demonstrated acid resistance in six populations of *E. coli* MG1655 to acid stress. These six populations were evolved for 740 generations at pH 4.5. All six populations demonstrated significantly enhanced fitness at pH 4.5 compared to the wild type ancestor. Isolated evolved clones of each population also demonstrated cross-stress enhanced fitness, including in osmotic and organic acid stress. These clones were shown to have loss

of function of *arcA*, a global regulator. Using RNAseq, they showed changes in metabolic regulation via changes in expression of genes involved in the TCA cycle and amino acid metabolism.

Studies detailing the evolution of pathogenic *E. coli* under organic acid stress have not yet been conducted. We therefore evolved twelve lines of EO 499 from a common ancestor. Six populations were evolved at pH 5.5, acidified with hydrochloric acid. Six populations were evolved at pH 5.5, acidified with hydrochloric acid, with 4 mM acetic acid added (to mirror the conditions used for RNAseq and TraDIS experiments). The intention of this study was to see whether increased fitness could evolve under organic acid stress. Examination of sequenced populations was then used identify nonsynonymous changes in gene sequence (those which caused amino acid change) in order to attempt to determine the pathways by which pathogenic *E. coli* manages continuous reproduction in organic acid and, as such, may shed some light on survival and proliferation within the human GIT.

The Chapter will first describe whether a phenotype was observed in the evolved populations when grown in competition with a non-evolved *lacZ* mutant of the same strain. The Chapter will then go on to consider the non-synonymous mutations within the populations, those which were shared across all or nearly all twelve populations, those which were confined to populations evolved at pH 5.5 and those which were confined to populations evolved at pH 5.5 in the presence of acetic acid. In doing so it may be possible to elucidate from the results the way in which fitness is conferred upon each population depending on the mutations which are shared and which are unique to each group.

5.2 Determining the phenotype of the evolved populations

E. coli EO 499 was evolved for 11 months and approximately 1900 generations. Six lines of EO 499 were evolved at pH 5.5 and six lines were evolved at pH 5.5 with 4 mM acetic acid. These conditions were chosen to reflect those used for the traDIS and RNAseq experiments, the aim being to

determine whether a more acid-tolerant and organic acid-tolerant phenotype could be evolved, and whether any mutations within the evolved populations reflected the genes determined to be important to organic acid resistance as outlined in Chapters Three and Four.

Every day 100 µl of culture grown in the chosen conditions was transferred into 5 ml of fresh medium and every two weeks an aliquot of culture was frozen for the fossil record, as described in Chapter 2.

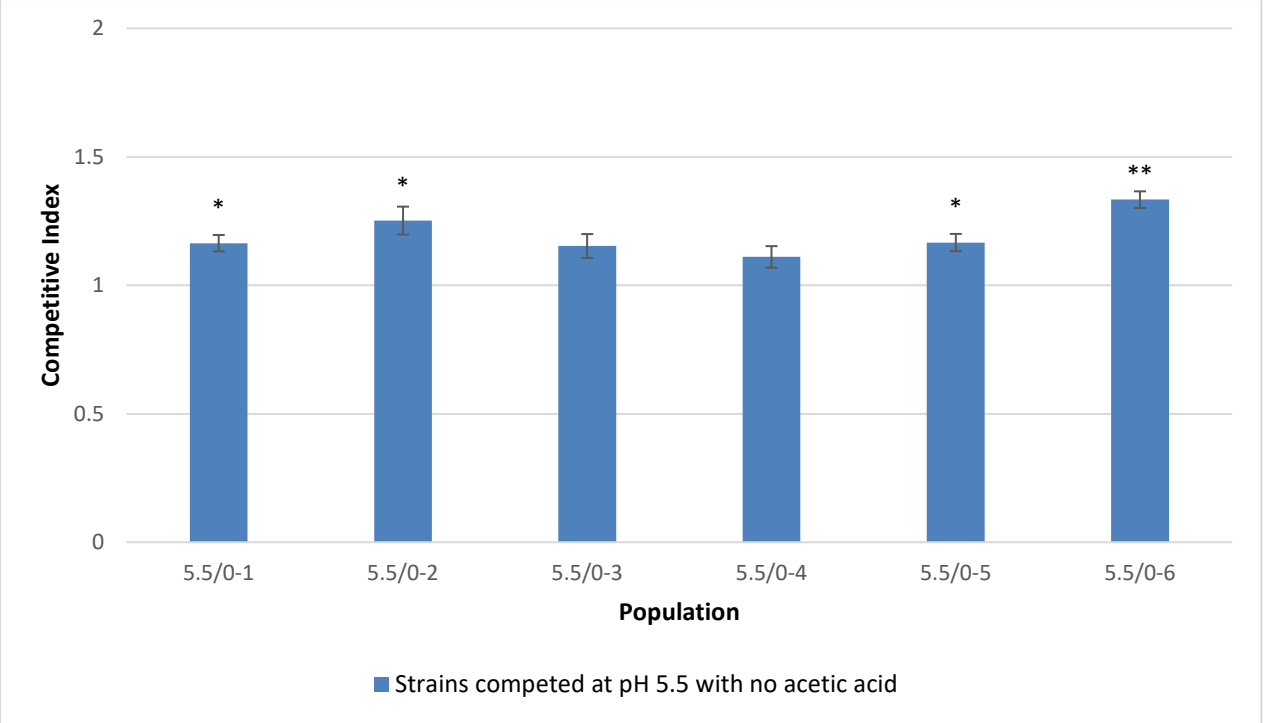
Following 11 months of evolution, the evolved populations were tested for an evolved phenotype using competition experiments. Each evolved population was competed in the condition under which it was evolved against the TraDIS Lac$^-$ transposon mutant (isolated from the transposon library as described in Chapter Seven). Prior to the competition experiments using the evolved population, the *lacZ* transposon mutant was first competed against the EO 499 ancestor strain in order to determine whether the transposon insert in the *lacZ* gene conferred any sort of competitive advantage or disadvantage at pH 7, pH 7 with 50 mM acetic acid, pH 5.5 and pH 5.5 with 4 mM acetic acid, which it did not (data shown in Chapter Seven). Fitness was calculated using the following equation (Lenski et al., 1991):

$$W=[\ln(R_{t24}/R_{t0})/\ln(V_{t24}/V_{t0})]$$

W is defined as the relative competitive index and R and V represent the two competing populations. T0 is the colony count at time point zero and T24 is the colony count at the 24 hour time point.

All competitions were repeated between three and nine times. In order to ensure there was no competitive difference between the TraDIS Lac$^-$ mutant and the ancestor strain, these two strains were first competed in all test conditions and at pH 7. Figure 5.1 below shows the results of the

competition experiments with the evolved populations. The first graph shows the populations evolved at pH 5.5 with no added acid (denoted as 5.5/0- followed by 1 – 6 for each of the six populations) and the second graph shows the competition for populations evolved at pH 5.5 with 4 mM acetic acid (denoted as 5.5/4- followed by 1 – 6 for those six populations). The six lines evolved at pH 5.5 with 4 mM acetic acid showed a more pronounced evolved phenotype versus the ancestor strain than those evolved at pH 5.5 only. Two of the strains evolved at pH 5.5 without added organic acid, populations 5.5/0-3 and 5.5/0-4 showed no significant difference to the *lacZ* transposon mutant.



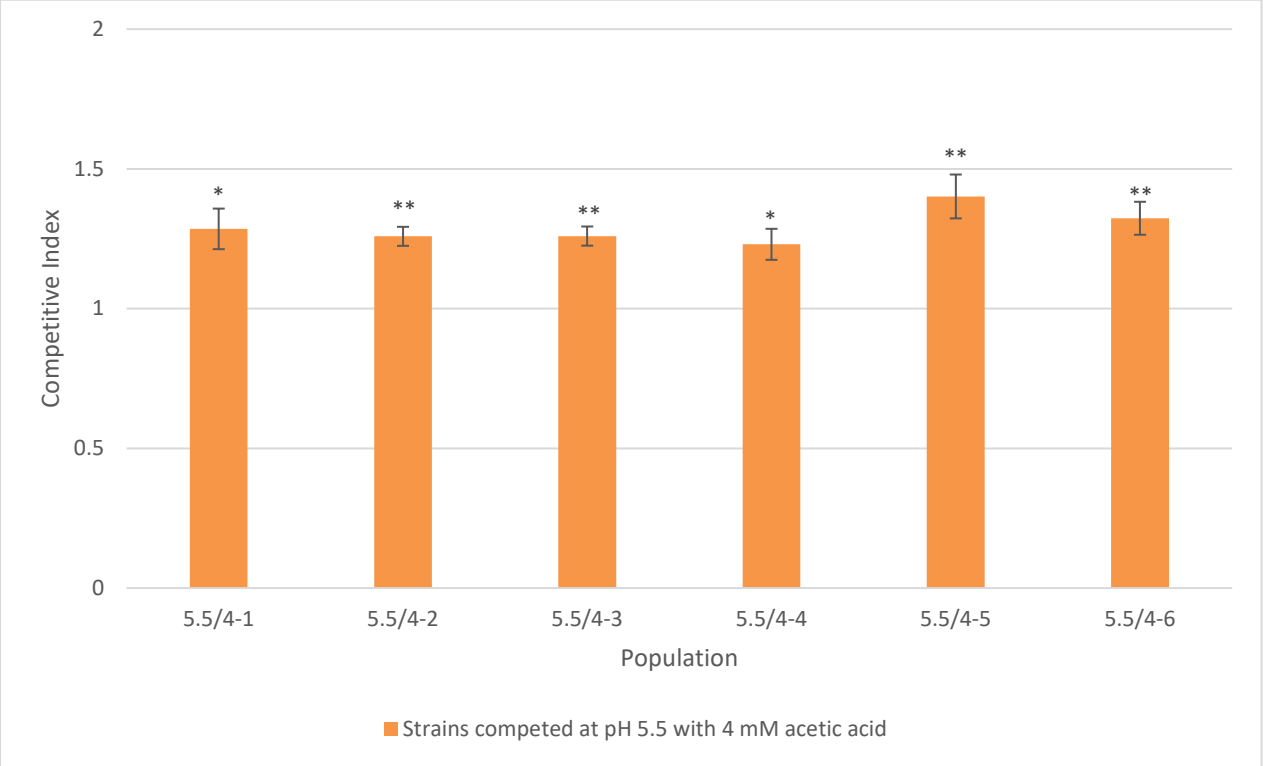
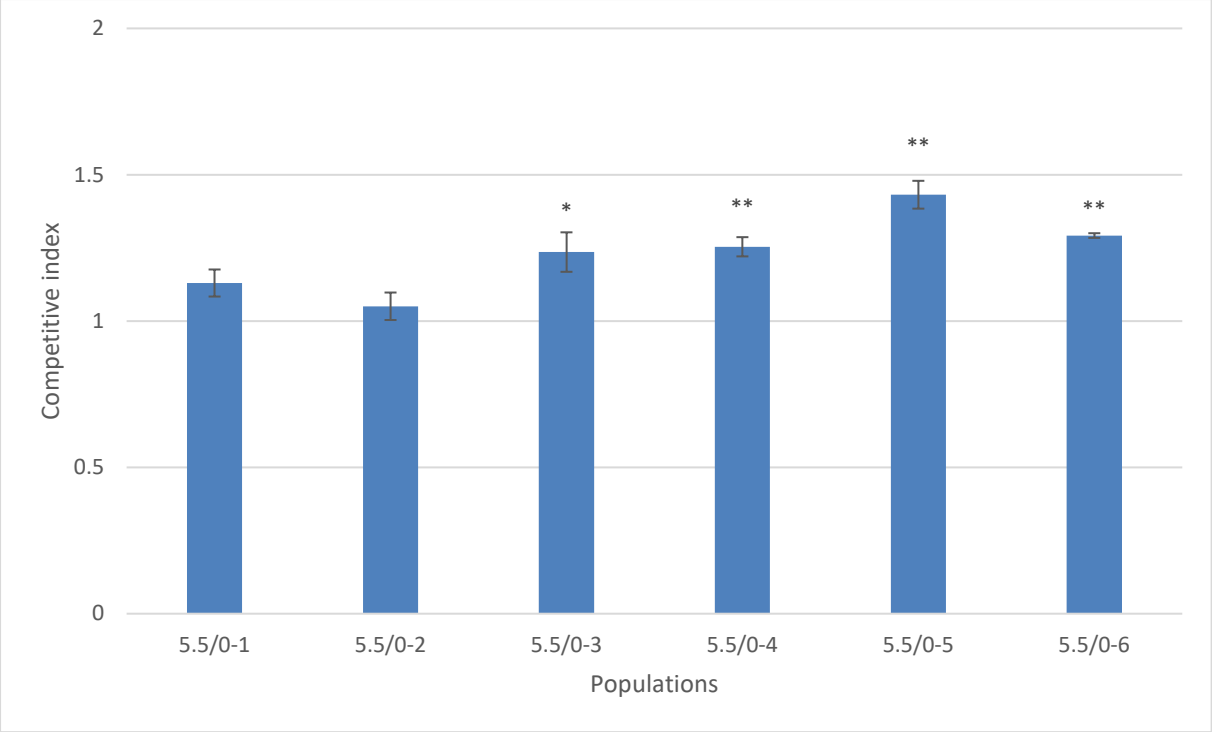


Figure 5.1 Top: Competition experiment between evolved populations 5.5/0 1 to 6 and the *lacZ* mutant. N=9. Error bars show SEM. Populations were competed at pH 5.5 Bottom: competition experiment between evolved populations 5.5/4 1 to 6 and the *lacZ* mutant. N=9. Error bars show SEM. Competed at pH 5.5 with 4 mM acetic acid. A single asterisk denotes a P-value of less than 0.05. A double asterisk denotes a P-value of less than 0.01.

Following the competition experiments in the condition in which the populations were evolved, the conditions were swapped to see if the populations showed an evolved phenotype in the alternative condition. Figure 5.2 shows the competition experiments of populations 5.5/0 1 to 6 in pH 5.5 media with 4 mM added acetic acid and beneath that is the competition experiment for populations 5.5/4 1 to 6 in media at pH 5.5 but with no added organic acid. What is interesting about this result is that four of the populations showed significant growth advantage over the *lacZ* mutant at pH 5.5 in the presence of acetic acid (5.5/0-1 and 5.5/0-2 did not). However, none of the populations evolved in acetic acid showed a fitness advantage over the *lacZ* mutant when competed in the absence of organic acids. It also appears that populations 5.5/4-2, 5.5/4-4 and 5.5/4-5 have lost fitness compared to the transposon mutant. These results though did not show any significant difference to the transposon mutant (p values of 0.09 – 0.3).



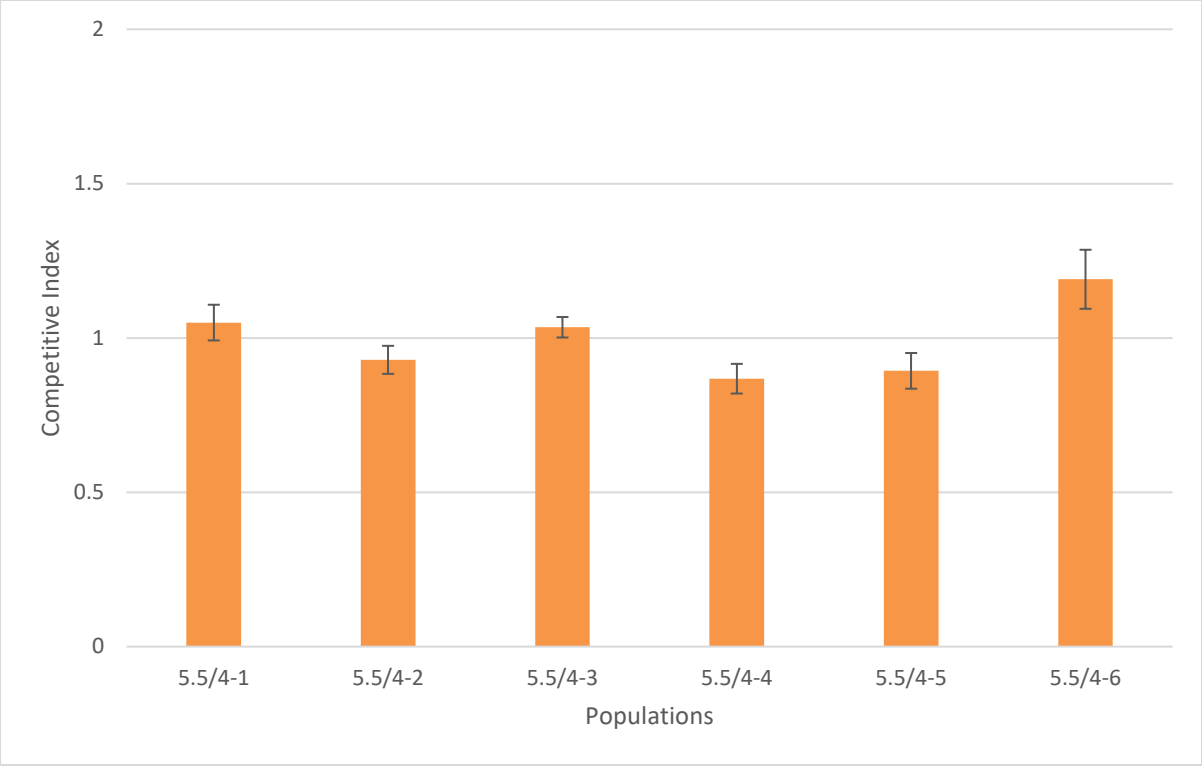


Figure 5.2 Top: Competition experiment between evolved populations 5.5/0 1 to 6 and the *lacZ* mutant. N=3. Error bars show SEM. Populations were competed at pH 5.5 with 4 mM acetic acid. Bottom: competition experiment between evolved populations 5.5/4 1 to 6 and the *lacZ* mutant. N=3. Error bars show SEM. Populations were competed at pH 5.5. A single asterisk denotes a P-value of less than 0.05. A double asterisk denotes a P-value of less than 0.01.

This result indicates that the fitness advantage seen in the 5.5/4 populations in the presence of acetic acid is potentially an acetic acid specific advantage. In order to test this, all twelve populations were competed against the TraDIS Lac$^-$ mutant at pH 7 in the presence of 50 mM acetic acid.

Figure 5.3 shows the results of the competition experiments between the Lac$^-$ transposon mutant and the evolved populations at pH 7 in the presence of 50 mM added acetic acid. The top graph shows the populations evolved at pH 5.5 and the bottom graph the populations evolved in the presence of acetic acid. The phenotype of improved fitness has been restored in the 5.5/4 populations. This supports the hypothesis that the fitness advantage is indeed acetic acid specific, rather than pH specific. Four of the six populations evolved at pH 5.5 also showed significant fitness advantage over the Lac mutant in this condition.

The next step in investigation the fitness advantage of the evolved strains was the sequence the populations. The next section of this chapter examines the sequencing data obtained from populations sequencing.

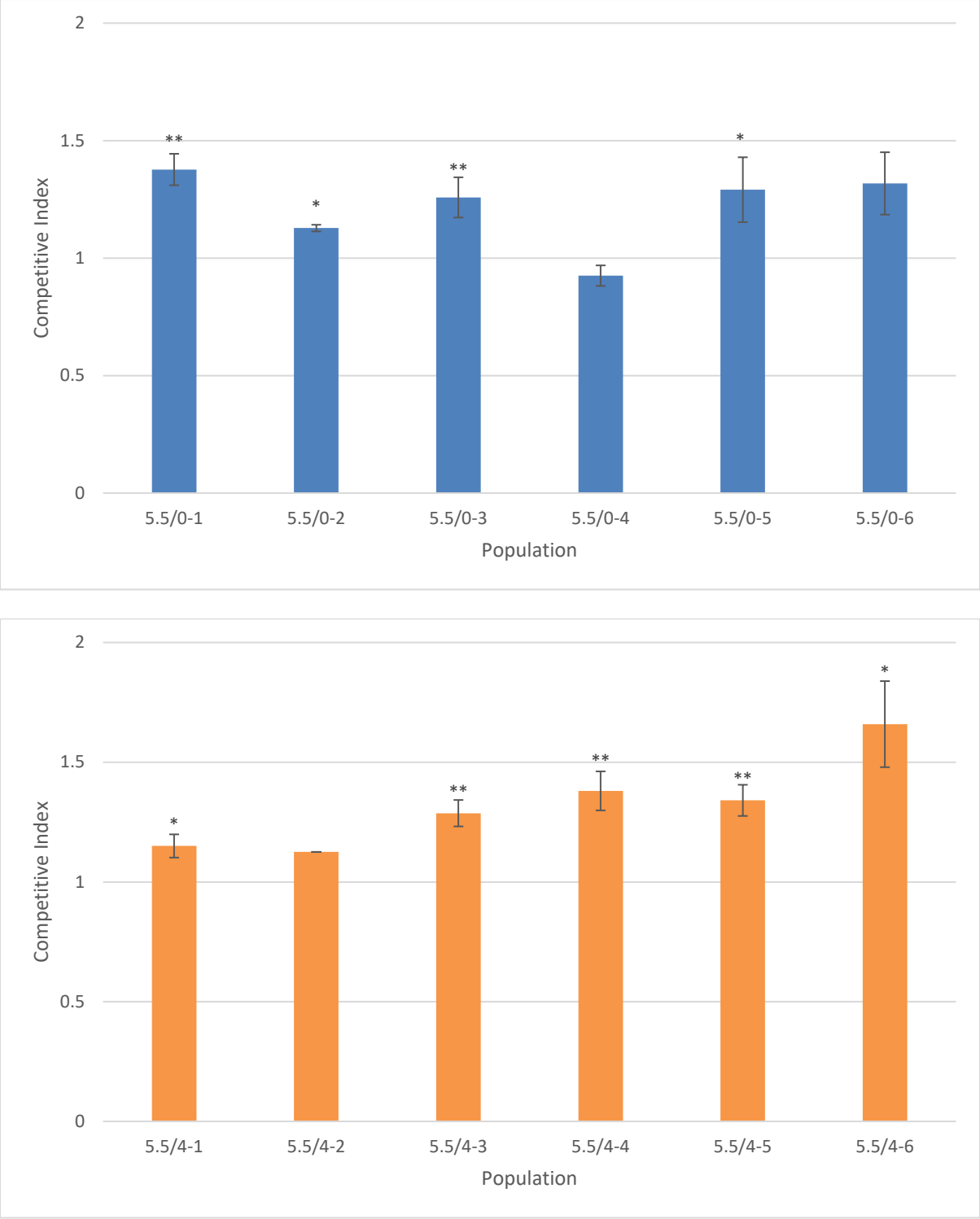


Figure 5.3 Top: Competition experiment between evolved populations 5.5/0 1 to 6 and the *lacZ* mutant. N=3. Error bars show SEM. Population were competed at pH 7 with 50 mM acetic acid. Bottom: competition experiment between evolved populations 5.5/4 1 to 6 and the *lacZ* mutant. N=3. Error bars show SEM. A single asterisk denotes a P-value of less than 0.05. A double asterisk denotes a P-value of less than 0.01.

5.3 Investigation the genotypic landscape of the evolved populations

This section will review the non-synonymous mutations within the evolved populations. Non-synonymous mutations indicate a change in the amino acid sequence and this may impact upon function. Only non-synonymous mutations were considered in this analysis due to the larger analysis into these evolved populations being in its infancy at the time of writing. A more in depth analysis, which includes the specifics of each mutation and also synonymous mutations is underway but will not be presented here. A full list of mutations is included in the supplementary data.

It was not possible to isolate single clones with evolved phenotype and so whole populations were sequenced using deep sequencing on the Illumina MiSeq 2500 platform as explained in Chapter Two. Data was then analysed by Mathew Milner, a PhD student, using Breseq which is a computational pipeline for finding mutations when compared against a reference genome. In this instance, the reference genome was EO 499, the parent strain (see Chapter Two).

This section will begin with the genes that were mutated in all twelve evolved populations and those which were shared between the two evolved groups, even if not in every population. It will then consider those mutations confined either to group one, populations evolved at pH 5.5 and then group two, populations evolved at pH 5.5 in the presence of acetic acid. The work has not yet been done to determine whether these are loss of function, gain of function or neutral mutations and that aspect of this project is on-going.

5.3.1 Genes with nonsynonymous mutations shared across evolved populations from both groups

Table 5.1 shows a list of genes which have been mutated in more than one population and across the two groups. A number of the genes shown are either putative/hypothetical proteins and are therefore unknown in function. The gene ST131v2_02626 which is commonly mutated in three

populations evolved at pH 5.5 and two populations evolved at pH 5.5 with acetic acid is a phage portal protein. There are also a number of other phage related proteins in the list. The high level of mutation within phage-related genes may play no role in the adaptation to acid and organic acid stress.

There are a number of genes in the list, which have been discussed in earlier chapters, such as *lpd, rpoA* and *rpoS.* The next chapter will consider these in more detail. It should be noted here, however that *rpoS* was not mutated in one of the populations evolved at pH 5.5 but in fact a large region of approximately 12kb starting *rpoS* and finishing at *mutS* (a DNA repair protein) was deleted entirely. The intervening genes are largely hypothetical proteins. Why this region was deleted would require more detailed investigation.

All twelve populations have a mutation in *nagE*. NagE is an N-acetylglucosamine specific transporter and part of the PTS. This system has come up repeatedly in this thesis as being of high importance to the cell during acid and organic acid challenge. The fact that it is mutated in all twelve populations could simply be down to it being mutated in the ancestor strain. However, mutated in all the populations in group two, and half the populations from group one are the genes *nanR, nanM* and *nanK.* NanR is the regulator of the N-acetlyneuraminc acid metabolism which is subsequently catabolised by *nanM* and *nanK* (with others) into N-acetlyglucosamine. Degradation of N-acetlyglucosamine is regulated by *nagC* which represses expression of *nagE* (Condemine et al., 2005; Plumbridge, 2001). These mutations are clearly all interconnected and the more detailed investigation into the effects of these mutations is on-going.

Gene name	5.5/0-1	5.5/0-2	5.5/0-3	5.5/0-4	5.5/0-5	5.5/0-6	5.5/4-1	5.5/4-2	5.5/4-3	5.5/4-4	5.5/4-5	5.5/4-6
tfaE/pinE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
yoeG	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
nagE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
lpd	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ydbA	✓	✓	✓	✓	0	✓	✓	✓	✓	✓	✓	✓
_02626	✓	✓	0	0	✓	0	✓	✓	✓	✓	✓	✓
nanR	✓	✓	✓	0	0	0	✓	✓	✓	✓	✓	✓
nanM	✓	✓	✓	0	0	0	✓	✓	✓	✓	✓	✓
nanK	✓	✓	✓	0	0	0	✓	✓	✓	✓	✓	✓
_03596	✓	✓	0	✓	0	0	✓	✓	✓	0	✓	✓
rpoS	✓	✓	✓	✓	✓	0	✓	✓	0	0	✓	0
imm	✓	✓	0	✓	0	0	✓	0	✓	✓	✓	0
_03597	✓	✓	0	✓	0	0	0	✓	✓	✓	✓	0
ldrD	✓	✓	✓	✓	0	0	✓	✓	✓	0	✓	0
stfE	0	0	✓	0	0	0	✓	✓	✓	✓	0	✓
nadR	✓	✓	0	0	0	0	✓	✓	0	✓	0	✓
papB/yraI	✓	0	✓	✓	✓	0	0	0	0	✓	✓	0
mdh	✓	0	0	0	0	✓	✓	0	0	✓	✓	0
yhjV/ldrD	✓	0	0	0	0	0	0	✓	✓	✓	0	0
ompR	✓	0	✓	0	✓	0	0	0	✓	0	0	0
rsxC	0	0	0	✓	0	0	0	✓	✓	✓	0	0
topA	0	✓	0	0	0	0	✓	0	0	✓	0	✓
mreB	0	0	✓	0	0	✓	0	0	✓	0	0	0
_02613	✓	0	0	0	0	0	0	0	0	✓	✓	0
kdsD	0	✓	0	0	✓	✓	0	0	0	0	0	0
_03123	0	0	0	✓	0	0	✓	✓	0	0	0	0
traD	0	0	✓	0	0	✓	0	0	0	✓	0	0
rpoA	0	0	0	0	0	✓	0	0	0	0	✓	0
_01828	0	0	✓	0	0	0	0	0	✓	0	0	0
_02047	0	0	0	0	0	✓	0	0	0	0	✓	0
yebT	0	0	0	0	0	✓	0	0	0	0	✓	0
nlpC/ydiV	0	0	0	0	0	0	✓	0	0	0	✓	0
ppsA	0	0	0	0	0	✓	0	0	0	0	✓	0
lrp	0	0	✓	0	0	0	0	0	✓	0	0	0
lpxC	0	0	0	0	0	✓	0	0	0	0	✓	0
rluF/lsrR	0	0	0	0	0	✓	0	0	0	0	✓	0
trkH	0	0	0	0	0	✓	0	0	0	0	✓	0
Total =	20	17	17	14	9	16	19	19	20	20	25	14

Table 5.1 List of genes with non-synonymous mutations together with populations in which they are mutated. A tick indicates the presence of a mutation and a zero indicates no mutation when compared to the ancestor strain.

Two of the genes in the list *mdh* and *ndh* also belong to the same pathway. The two dehydrogenases facilitate redox reactions as part of the TCA cycle, and *mdh* has been implicated in the direct transfer of NADH into NADH:ubiquinone oxidoreductase complex I (encoded by the *nuo* genes) (Amarneh and Vik, 2005).

OmpR accumulated mutations in four of the twelve populations. OmpR regulates the porins OmpF and OmpC which have been discussed in Chapter One section 1.8 due to their known role in *E. coli* acid resistance. OmpR has also been connected to genes involved in aspects of metabolism, such as glycolysis and pyruvate metabolism, and also components of the GAD system (Stincone et al., 2011). It is also directly implicated in the regulation of other genes which may be involved in acid resistance. Also in the list above is *trkH* which is involved in potassium transport. As discussed in Chapter One, potassium transport plays a key role in cell homeostasis during acid stress.

Next, the genes shared across populations evolved at pH 5.5 will be considered in more detail.

5.3.2 Genes with nonsynonymous mutations in populations evolved at pH 5.5

Other than those mutations above, *kdsD* was the only gene to have accumulated mutations in more than one population evolved at pH 5.5. Levels of nonsynonymous mutations were low, with an average of 27 nonsynonymous mutations per population, the lowest being 14 nonsynonymous mutations in population 5.5/0-5 and 57 in population 5.5/0-6.

There are two remaining genes shared across more than one population. The first, *kdsD,* is involved in biosynthesis of the cell wall. The gene *kdsD* encodes D-arabinose-5-phophate, which is a precursor for KDO, a vital component of the cell wall (Martorana et al., 2011). In the list of genes shared across both groups (two from group one and one from group two) is *mreB*, which is involved in maintenance of the rod shape of the cell and has also been implicated in inner-membrane organisation (Nurse and Marians, 2013). Within the list of genes mutated in only one population from this group was *mrdB* which is also involved in cell wall maintenance.

The next section will focus in on the genes mutated in the populations evolved at pH 5.5 in the presence of acetic acid.

5.3.3 Genes with nonsynonymous mutations in populations evolved at pH 5.5 with 4 mM acetic acid

Due to the high number of mutations that occurred in these population, this analysis will concentrate on those mutations which occurred in more than one population, unless it is of particular note. There were just under three times as many mutations across the strains evolved at pH 5.5 with acetic acid than at pH 5.5 alone. This is likely due to the greater selective pressure on the cell.

Table 5.2 below lists the genes that were mutated in more than one population. The majority of these mutations were shared across two populations, with only four being shared across three. None were shared across four or more without being also shared with populations from group one.

Those shared across three populations were *ptsP, uacT* and *gltP*. They will be discussed in more detail below.

Gene name	5.5/4-1	5.5/4-2	5.5/4-3	5.5/4-4	5.5/4-5	5.5/4-6
acpT	✓	0	0	0	✓	0
adiA	0	✓	0	0	0	✓
apaH	0	0	0	✓	✓	0
clpA	✓	0	0	0	0	✓
csiD/smpB	✓	0	0	0	✓	0
dhaR	✓	0	0	0	✓	0
eptA	0	0	0	✓	✓	0
fabB	0	0	0	✓	✓	0
fadH	✓	0	0	0	✓	0
fecC	0	0	0	✓	✓	0
fic	✓	0	0	0	✓	0
glcF	✓	0	0	0	✓	0
gltP/nrfG	✓	✓	0	0	✓	0
hokE/ybdK	0	✓	✓	0	0	0
holA	✓	0	0	0	✓	0
kup/ravA	✓	0	0	0	✓	0
ldtE	✓	0	0	0	✓	0
leuO/leuA	0	✓	0	0	0	✓
lpxK	✓	0	0	0	✓	0
malM	✓	0	0	0	✓	0
mobA	✓	0	0	0	✓	0
mobB	✓	0	0	0	✓	0
mprA	✓	0	0	✓	0	0
mrcA	✓	0	0	0	✓	0
nlpC/ydiV	✓	0	0	0	✓	0
nusG	✓	0	0	✓	0	0
papB/yraI	0	0	0	✓	✓	0
pbpG	✓	0	0	0	✓	0
pepA	✓	0	0	✓	0	0
ptsP	0	0	✓	✓	✓	0
purL/glrK	✓	0	0	0	✓	0
rsmH	✓	0	0	0	✓	0
sapC	✓	0	0	0	✓	0
ST131v2_03127	0	✓	0	0	0	✓
ST131v2_05411	✓	0	0	0	✓	0
ST131v2_05434	✓	0	0	0	✓	0
sufS	✓	0	0	0	✓	0
tktA	✓	0	0	0	✓	0
treB	0	✓	0	0	0	✓
uacT	0	✓	0	✓	0	✓
wecA	0	✓	0	✓	0	0
xdhD	✓	0	0	0	✓	0
yaiT/hemB	✓	0	0	0	✓	0
ybaT	✓	0	0	0	✓	0
yddM	✓	0	0	0	✓	0
ydhF	✓	0	0	0	✓	0
yeeJ	0	0	✓	0	0	✓
yghA	✓	0	0	0	✓	0
yhdJ	✓	0	0	0	✓	0
yjbE/pgi	✓	0	0	0	✓	0
yjiY	✓	0	0	0	✓	0
Total =	39	9	3	12	41	6

Table 5.2 Genes with mutations in two or more populations evolved at pH 5.5 with acetic acid. A tick indicates the presence of a mutation and a zero indicates no mutation when compared to the ancestor strain.

Of particular note are the genes *adiA, apaH, clpP, fadH, mobA* and *mobB.* These have all been discussed in earlier chapters and this will be explored in more detail in Chapter Six. Two of the populations had large deletions. These deletions, though not the same, were very similar and overlapped, with both populations dumping the *fli* genes which encode flagellar synthesis. This has been seen before in lab evolved populations due flagellar assembly both having high energy requirements and being generally unnecessary in the manufactured lab environment of constant shaking at 180 rpm (Plague et al., 2017, Sen et al., unpublished).

In the list of genes which are mutated in more than one population is *gltP*. This is a H^+ / glutamate symporter, which would be particularly useful under acid stress in removing H^+ from the cell under acid stress. In the list of genes mutated in one population is *gltS,* which is also a glutamate transporter, but this population is not one of those carrying the *gltP* mutation. This gene is a glutamate/Na^+ transporter. This may play a role under organic acid stress in maintaining cell homeostasis as described in Chapter One.

Also mutated in multiple populations is *ptsP,* another gene in the phosphotransferase system, and *dhaR* which also regulates genes involved in the PTS.

UacT, mutated in three of the six populations in group two, is a urate/H^+ symporter. Included in the list of genes mutated in one population only are *cadC,* involved in the lysine-dependent acid resistance system, *nuoF* and *nuoG* which have been described in earlier chapters, and *trkA* which is involved in potassium transport. The genes listed here are a snapshot of the sequencing data for the evolved populations and a more detailed analysis is underway.

5.4 Discussion

In this part of the study EO 499 was evolved for almost 2000 generations at pH 5.5. and at pH 5.5 with 4 mM acetic acid. After this period of evolution, the populations did show an increase in fitness compared to the wildtype, however in the populations evolved in acetic acid, that fitness advantage did not extend to mildly acidic pH only. The populations evolved just at pH 5.5 however did retain their fitness advantage when competed in organic acids. This indicates that the fitness advantage gained from evolving at mildly acidic pH can be utilised in organic acid stress response. However, the mutations which contribute to a fitness advantage in organic acid do not also confer an advantage to pH 5.5 only. This is an interesting result. Multiple laboratory-based evolution experiments have shown that an evolved fitness advantage in one stress often confers a fitness advantage in multiple stresses (Sen et al, unpublished, Dragosits et al., 2013; Rudolph et al., 2010). This was indeed validated to some extent by the pH 5.5 evolved population showing a fitness advantage at pH 7 in the presence of 50 mM acetic acid. The fact that cross-stress fitness, to a condition very like the one in which it was evolved, did not appear to be shown in the population evolved in acetic acid needs much more rigorous analysis than was possible in the timescales of this thesis. There is still a wealth of work to be done on this aspect of the project, and that is one of the recommendations for future research to come out of this report.

Deep sequencing of the twelve populations revealed mutations in the genes *rpoS, rpoA, cadC, cadA* and *adiA.* These are known acid resistance genes and this result is similar to that seen in other acid-evolved populations including by Harden et al., (2015) who evolved MG1655 in mildly acidic conditions, and also Rajaraman et al., (2016) who evolved *E. coli* at pH 7 on 85 mM acetate as the sole carbon source. They noted mutations in *rpoA.* Other genes which were mutated in these populations are members of the aerobic respiratory chain and the PTS. This will be discussed in more detail in the next chapter.

Chapter Six:

A comparison of RNAseq, TraDIS and Evolution data

Chapter 6: A comparison of RNAseq, TraDIS and evolution data

The previous three chapters have described in detail the transcriptional, genomic and evolutionary response of EO 499 to organic acid stress. Many of the processes which appear to be involved in the response of the cell to organic acid stress are centred around fermentation, the PTS, PMF and the TCA cycle. These systems have been discussed in Chapter One, describing the role of organic acids in cellular metabolism. Other pathways which appear from the evidence to be playing a central role in the response to organic acids are glycolysis and the pentose-phosphate pathway, which run parallel to one another and tie in with the other pathways in terms of contributing to metabolism.

This chapter will look at those genes which have been identified as likely to be significant in more than one of the previous three chapters. As has been shown in the earlier analyses in this thesis, there will inevitably be uncertainty in the data provided by all three of the methods used here. There is a change in regulation of systems that do not appear to play a direct role in organic acid response, or at least there is no obvious explanation, for example upregulation of flagella biosynthesis in Chapter Three, the inclusion of essential genes in the TraDIS analysis, or the mutation of phage genes in evolution. By focusing therefore on the genes and pathways which are in common between the three methods, it may be possible to paint a more focused picture of what is happening on a cellular and molecular level under these stress conditions. TraDIS and RNAseq do not measure the same thing, and therefore it is reasonable to assume there would be few genes which are considered significant in only one of the two methods as a gene which impacts on fitness may not be differentially expressed, and vice versa. However, it is also distinctly possible that the same networks and pathways will be differentially expressed and have an impact on fitness.

6.1 How comparable are these different approaches?

There are few studies which have used this particular approach in order to study cellular response, and only one study found in the literature has used all three. The first described here used an approach of RNAseq and evolution. Zorraquino et al., (2016) used lab-based evolution over 500 generations combined with RNAseq after 12 hours exposure to the stresses of interest, n-butanol, osmotic, acidic (pH 5.5) and oxidative, individually and in combination in order to build up a stress-response network in *E. coli*. Fitness of the evolved strains was tested using competition assays. In terms of over-lap between the stresses, they found that 18 of the 21 differentially expressed genes under acid stress were shared with the other stresses, and 7 of those 18 were shared between all four stresses. Overall, they identified no genes in common between the two methods as being both differentially expressed and mutated in more than one population (4 populations were evolved in each stress). However, they found that many of the genes were part of common networks and shared similar regulators and though there were no genes shared between the two methods, they were able to build up network responses to the stresses by looking at relationships between genes in each data set.

Jensen et al., (2016) used a combination of RNAseq and TraDIS (TnSeq) to elucidate the response of *S. pneumoniae* to nutrient depletion. They found that there was little correlation between phenotypically important genes (which contribute to fitness) and those which were differentially expressed. However, as with the study outlined above, they determined that whilst they saw few of the same genes across the two methods, they were able to construct networks based on the data due to many of the same pathways being both differentially regulated and contributing to fitness. Despite a date of 2016, this paper was only available as a pre-print that had not been peer reviewed.

The same laboratory has also released a new preprint dated August 2018 in which they incorporate short-term lab-based evolution into their study of the *S. pneumoniae* response to nutrient depletion. They cultured *S. pneumoniae* in nutrient limited media and reduced the nutrient concentration by 15% every three days until they had a stable population in severely nutrient depleted media (after around 40 generations). They assessed the transcriptional response to the nutrient limitation of these strains and transcriptional response in the same environment to the parent strain and, again, used the data from all three methods to build a network of the response to nutrient limitation. They found that adaptation led to transcriptional 'rewiring' and that the adapted strains had a more focussed transcriptional response to the stress environment, with fewer differentially expressed genes. They extended this approach to look at antibiotic resistance and also to *Pseudomonas aeruginosa* in a bid to predict the evolutionary trajectory using TnSeq and RNAseq from PA01. They argue that by using machine learning, RNAseq and TnSeq could be used to predict adaptation and disease progression, with a particular interest in burn wounds (Zhu et al., 2018).

From these studies then it appears that whilst the data from these various approaches is often not directly comparable in the sense that the same genes appear as differentially expressed, as contributing to fitness or as mutated through evolution, they are often part of the same networks which imply that the bacteria are using these particular networks and pathways in response to the stress and/or environment.

This Chapter will bring together the results from the previous three in order to make comparisons between them and identify the genes and networks that EO 499 is utilising under organic acid stress at both neutral and mildly acidic pH. The approach used is not as sophisticated as that employed by Zhu et al., (2018) referenced above. It has not been possible to use advanced bioinformatics in order to analyse

the pathways which are shared. As stated above, there may be few genes in common between the methods used, but those genes may belong to the same networks. It is the intention for more sophisticated methods to be used for analysis prior to publication of this data.

6.2 Genes which overlap in RNAseq, TraDIS and lab-based evolution

The table below shows the genes which overlap in all three methods. When considering the data, it might be reasonable to assume that a gene which is required in TraDIS might also be upregulated, and that it would not be mutated as this may affect function. However, as mentioned above, these methods are looking at different things. It may also be the case that a mutated gene somehow improves function, and therefore despite it being required and upregulated, it is also mutated. There are several possibilities and they will be explored in more detail below.

The genes included from the RNAseq analysis are those which were significantly differentially regulated (log2-fold change of 1 and p-value of 0.05 or less). Those from the TraDIS analysis are either labelled as TraDIS-improved, meaning that fitness was improved under the test conditions, or TraDIS-required meaning that that fitness was reduced under the test conditions. Those labelled as mutated are those which were mutated in at least one population evolved at pH 5.5 with 4 mM acetic acid. The genes in the table below have been compiled from the data generated in acetic acid, in order that it is comparable to the evolution data which was only generated in acetic acid. It has not been separated into different pH conditions, though symbols have been used to identify which were shared across the different pH and which were not. Only aerobic data was used for this three-way comparison as the evolution experiment was done under aerobic conditions. The data for pH 7 plus 40 mM acetic acid was used for the comparison as the evolved populations also showed fitness advantage in this condition so it

was theorised that genes upregulated or downregulated or improving or impacting fitness in the condition may also have been genes mutated in the evolution experiment.

TraDIS-Improved, Mutated & Upregulated	TraDIS-Improved, Mutated & Downregulated	TraDIS-Required, Mutated & Upregulated	TraDIS-Required, Mutated & Downregulated
nanK	*cmk*	0	*apaH*†
	*pgi**		*cysK*
	*ppc**		*hokD*†
	pyrD		*nuoF*
	*relA**		*nuoG*†
	*smpB**		*ptsN**
	tktA		

Table 6.1 Genes which overlap across the analyses for all three methods used in chapter three, four and five. Symbols denote the following: † = in the lists for both pH 7 and pH 5.5. * = in the list for pH 7 but not pH 5.5 (TraDIS). No symbol = in the list for pH 5.5 but not pH 7 (TraDIS).

The first gene here is *nanK* which is mutated in all six populations evolved at pH 5.5 with 4 mM acetic acid and in three of the populations evolved at pH 5.5 only. In the RNAseq experiment *nanK* was measured of having an upregulated Log2-fold differential expression of 6.48 and a p-value of 0.0006 in pH 5.5 aerobic, and in the TraDIS data a Log2-fold change in transposon frequency of 1.08, though the p-value was high at 0.13. However, despite the high p-value for the TraDIS data, it appears likely that this gene has an important function under organic acid stress, though that may or may not be through loss of function. It is at this stage unfortunately not possible to know the impact of the mutations on the evolved populations. NanK is the N-acetylmannosamine kinase which adds a phosphate group to N-acetylmannosamine, produced by the removal of a pyruvate group from N-acetylneuraminic acid (NANA). This allows bacteria to use NANA as a carbon source by eventually converting it into fructose-6P for use in glycolysis. With regards to the TraDIS result, where fitness was improved, the loss of this additional route of carbon utilisation may cause increased usage of organic acids as a carbon source instead, easing organic acid stress in the cytoplasm. NANA is as a precursor to UDP-N-acetylglucosamine, used in cell wall synthesis. Acetate is utilised in the formation of UDP-N-acetylglucosamine, and

therefore upregulation of this pathway may result in higher utilisation of excess acetate in the cytoplasm, relieving stress on the cell (Anba-Mondoloni et al., 2013; Plumbridge and Vimr, 1999; Rao and Mendicino, 1978). It is difficult to elucidate from the literature, however, why either removal or upregulation of this pathway might play a particularly important role under organic acid stress. Where RNAseq and TraDIS give apparently conflicting data like this, it can be difficult to understand from looking simply at gene function what is happening on a wider scale within the cell.

Cytidine monophosphate kinase, *cmk*, was downregulated in all three acids at pH 5.5 (data for acetic acid only: -1.76, p = 0.0007) and TraDIS mutants showed improved growth in acetic acid at pH 7 and pH 5.5 (-2.97, p = 0.029), though not in propionic and butyric acid. It is mutated in one of the populations evolved at pH 5.5 with 4 mM acetic acid. This kinase is involved in DNA replication and plays an important role in synthesis of nucleoside precursors, though is not essential. In the RNAseq analysis, there were several pathways associated with DNA replication which were downregulated, and it may be that this process is being downregulated to conserve energy in the stress conditions (Fricke et al., 1995; Walker et al., 2012).

In the same group of genes that were downregulated in RNAseq (-2.52, p = 0.0002) and where loss of function improved fitness in TraDIS is *tktA* (-1.315, p-value 0.036), there is a transketolase which provides a reversible link between the pentose phosphate pathway and glycolysis. *E. coli* has two transketolases *tktA* and *tktB* which complement one another. In data presented in Chapters Three and Four, *tktB* was significantly upregulated, and the TraDIS mutants showed little change in transposon frequency. It is possible that downregulation of *tktA* causes upregulation of *tktB,* which could be a more efficient pathway. On that basis, it is also possible that loss of function via transposon insertion in *tktA* increased *tktB* function to compensate (though we cannot know this from the TraDIS data), which was

not replicated with *tktB* was interrupted with the transposon. Expression of *tktB* is *rpoS* dependent, so it is possible that this pathway would be upregulated in organic acid stress due to the *rpoS*-dependent acid response system (Rahman et al., 2008).

Also in this list is *ppc* which stands for phosphoenolpyruvate carboxylase, which replenishes oxaloacetate in the TCA cycle. Peng and Shimizu, (2004) showed that *ppc* knockout mutants produced less acetate and enzymes of the TCA cycle were significantly upregulated whereas genes involved in glycolysis and the pentose phosphate pathway were significantly downregulated. In the data presented in this thesis, *ppc* was significantly downregulated at both pH 5.5 plus acetic acid (-1.936, p = 0.0005) and pH 7 with acetic acid (-1.78, p = 0.0008) and the TraDIS mutant showed improved fitness at pH 7 with acetic acid (-1.21, p = $0.6x10^{-24}$) though interestingly, not at pH 5.5 with acetic acid. This gene is also mutated in a single population evolved at pH 5.5 with 4 mM acetic acid.

Considered as significant in the RNAseq for acetic acid only (-1.397, p = 0.0007) and TraDIS analyses for acetic, butyric and propionic acid (for acetic acid: 2.474, p = $0.7x10^{-10}$), and also as mutated in two populations evolved at pH 5.5 with 4 mM acetic acid is *apaH,* which encodes diadenosine tetraphosphate, which hydrolyses diadenosine-5', 5'''-P^1, P^4-tetraphosphate (Ap4A) into two molecules of ADP. Marriott et al., (2015) showed that Ap4A may be responsible for, or play a significant role in, prevention of DNA damage in mammalian cells under stress. However, in *E. coli* it has been shown that Ap4A binds to a number of proteins associated with heat and oxidative shock, particularly DnaK and GroEL. ApaH has not been shown to induce the response to heat and oxidative shock, but possibly plays a role in modulation, in which case its role may go beyond that of these two stress conditions, to include others such as organic acids (Johnstone and Farr, 1991). Vivijs et al., (2016) used a transposon insertion library in *E. coli* K12 MG1655 to study the response to a moderately low pH of 4.5. They also showed

increased sensitivity of *apaH* mutants to this pH. Furthermore, they were subsequently able to demonstrate pH sensitivity in MG1655 *apaH* knockout mutants (compared to wildtype). They postulate that the inability of the *apaH* mutants to hydrolyse Ap4A to ADP causes a build-up of Ap4A, which inhibits proteins involved in stress response such as DnaK and GroEL leaving the cell unable to effectively combat cellular stresses. In the data presented here *apaH* mutants were considered significantly inhibited compared to a control condition of pH 5.5, where *apaH* mutants did not demonstrate growth inhibition. It may be that in the cited study, the additional reduction to pH 4.5 induced the same response as seen here at pH 5.5 with acetic acid (as opposed to at pH 5.5 alone which did not induce this response and furthermore, this gene was only mutated in the population evolved in the presence of acetic acid, and not in the population evolved at pH 5.5 alone). Also, in data presented in Chapter Four, a transposon insertion in *dnaK* improved growth in acetic acid at pH 5.5 aerobically. The RNAseq data also showed that *dnaK* was significantly downregulated in propionic and butyric acid. Additionally, *apaH* was significantly downregulated in acetic acid at pH 5.5 aerobically, but significantly upregulated in all three acids under anaerobic conditions at pH 7, and in butyric acid at pH 5.5 anaerobic. It does not, however, appear in the TraDIS lists for these conditions. This suggests either an additional function of *apaH* or that its role in stress response and, possibly, prevention of DNA damage is not fully understood (Farr et al., 1989; Johnstone and Farr, 1991; Marriott et al., 2015; McLennan et al., 2001; Vivijs et al., 2016).

Cysteine synthase A, *cysK,* was significantly downregulated in RNAseq in acetic and butyric acid, but not propionic (data for acetic acid: -1.303, p =0.003), in TraDIS in acetic acid only showed a decrease in fitness when function was lost (1.062, p = 0.041), and was also mutated in one population evolved at pH 5.5 with 4 mM acetic acid. CysK is involved in cysteine biosynthesis in a reversible reaction whereby acetate is displaced from O-acetly-L-serine resulting in L-cysteine, acetate and H^+. That is the preferred

direction of the reaction, and may explain why the reaction is downregulated under acetic and butyric acid stress. However, the TraDIS experiment showed that loss of function of *cysK* caused a limitation of fitness in acetic acid stress. It may be that, although the gene is downregulated, it still functions in the reverse reaction under acetic acid stress to remove both acetate and H^+ from the cytoplasm and therefore loss of function is detrimental to the cell. It would be very interesting to know how the mutation in the evolved population is affecting phenotype in this instance.

Two of the genes in this list, *nuoF* and *nuoG* have come up repeatedly in this study. Given there are two genes from the *nuo* operon mutated in the evolved populations, though both genes were only mutated in a single population and they were not mutated in the same population, it seems as though this system, which has been shown in previous studies to be upregulated in acid stress, is almost certainly involved in the organic acid stress response (Kanjee and Houry, 2013). The RNAseq data here shows the reverse of that, where the system is in fact downregulated, however the interruption of any one of the *nuo* operon subunits showed a decrease in fitness in the TraDIS experiment. It would, again, be incredibly useful to know in what way, if any, the mutations in the evolved populations are impacting on phenotype.

The next section will look at the genes which were considered as significant in the TraDIS analysis, and which have been mutated in one or more of the evolved populations.

6.2 Genes which overlap in TraDIS and lab-based evolution

This section will look more closely at the genes which were shared only between TraDIS and were found to be mutated in the evolved population, but did not show differential expression in the RNAseq experiment. It is possible that genes whereby fitness was improved in the TraDIS experiment would also

be mutated in an evolved population which demonstrated improved fitness. However, long term evolution does not accurately mirror a short-term TraDIS experiment and therefore there could also be little or no genes in common between the two experiments. A short-term growth advantage over a period of 12 hours, as seen in TraDIS, may have detrimental impacts on the cell over hundreds or thousands of generations. Table 6.2 below shows the genes which overlap between the two methods. There are only four genes in common between the two. The will be discussed in more detail below.

TraDIS-Improved & Mutated	**TraDIS-Required & Mutated**
aroE	*mobA**
glnG	
sapC	

Table 6.2 Genes which overlap across the analyses for TraDIS and evolution, but were not differentiall expressed. Symbols denote the following: * = in the list for pH 7 but not pH 5.5 (TraDIS). No symbol = in the list for pH 5.5 but not pH 7 (TraDIS).

The first listed here *aroE* is an essential gene when *E. coli* is grown in supplemented M9 (as it was here) (Joyce et al., 2006; Patrick et al., 2007), which brings into question the validity of the TraDIS result, an issue which has been discussed in Chapter Four. However here the transposon frequency was compared to the control which was cultured at pH 5.5 and there was a significant increase in fitness compared to the control. There is clearly a difference between the two conditions here on transposon frequency at the end of the experiment, even if those transposons are in a dispensable part of the gene, though it is not possible to say how that might affect expression. The gene *aroE* is mutated in a single population. In a separate population, *aroF* is also mutated. In the RNAseq experiment *aroABCGHK* were all significantly downregulated, *aroBHKL* were significantly downregulated in butyric acid and *aroBH* were

downregulated in propionic acid. There is more work to do here on fully understanding the role of this operon in organic acid stress.

The other genes in the table above have been discussed in earlier chapters and will not be revisited here.

6.3 Genes which overlap in TraDIS and RNAseq

As discussed above in the introduction to this section, the amount to which TraDIS and RNAseq data sets can be compared has been the topic of some discussion and will likely become a more pertinent issue as these two methods are more frequently used simultaneously.

Figures 6.1 and 6.2 below show the genes that were considered significant in both the TraDIS and the RNAseq experiments in acetic acid aerobic (Figure 6.1) and anaerobic (Figure 6.2). There were many fewer genes, overall, which were considered significant in the TraDIS analysis than in the RNAseq analysis, with a total of 2567 genes either up (1558) or downregulated (1009), compared to 102 genes in total for TraDIS, with 35 genes improving fitness and 66 genes reducing fitness in that analysis. In total, there were 50 genes in common between the TraDIS and RNAseq experiments for pH 5.5 with 4 mM acetic acid under aerobic conditions. This accounts for 49.02% of the TraDIS genes considered significant, but only 1.95% of the genes which were differentially expressed in RNAseq. At pH 7, a total of 79 genes were considered significant in the TraDIS experiment and 2567 differentially expressed in RNAseq. There were 43 genes shared between the two, which was 54.43% of the TraDIS genes and 1.7% of the RNAseq genes. Given TraDIS and RNAseq are looking at two very different responses, one transcriptional and the other phenotypic, it is not unexpected that the results would be different, in the sense that whilst transcription of a gene may change in the environment of interest, that does not mean

that loss of function of that gene would necessarily have a phenotypic impact on the cell, resulting in improved or reduced fitness. The reverse is also true, in that a gene may be crucial to fitness, but not change in expression under the tested condition. This may be because the biological process for which the gene is required is not being utilised under organic acid stress, so it is downregulated as it is not required, and therefore removal of function has no effect. It may also have a role which is replicated in another gene or pathway, and therefore loss via transposon insert has no effect on the cell. In respect of genes that are significant in TraDIS but show no difference in regulation, these genes may be being utilised by the cell in a process which involves multiple genes, which are not differentially expressed and are not essential, but removal of one gene, in an operon for example, inhibits function of the entire biological process, thereby causing a growth defect in the cell in certain conditions. There is, however, some overlap between the two methods as mentioned above. Even if every gene which was considered significant in TraDIS was differentially expressed, at pH 5.5 with 4 mM acetic acid this would still only be 3.97% of the total number of differentially expressed genes. Given the total in common of 1.95% of differentially expressed genes and 49.02% of genes impacting fitness, there does seem to be quite a good level of agreement between the two methods, and in the case of pH 7 plus 40 mM acetic acid, this is increased to more than half of the TraDIS genes. This slightly higher figure could potentially be because of inclusion of some genes which are differentially expressed because of acid stress, rather than organic acid stress, as this condition was compared against a control of pH 7 with no added acid, whereas the data generated at pH 5.5 with added organic acid was compared against a control of pH 5.5.

With regards to the data generated under anaerobic conditions, there were far fewer shared between RNAseq and TraDIS in the acetic acid data sets, with only four for each pH condition. This is largely because there were fewer genes found to be significant in the anaerobic data sets. At pH 5.5 with acetic

acid under anaerobic conditions there were 345 significantly upregulated genes and 192 significantly downregulated genes and a total of 30 genes in the TraDIS analysis, 26 of which were required for fitness and 4 which improved fitness. There were four in common between the two sets, which is 0.74% of differentially expressed genes and 13% of the TraDIS genes. At pH 7 there were again 4 in common, which was 0.7% of the differentially expressed genes and 15.4% of the TraDIS genes.

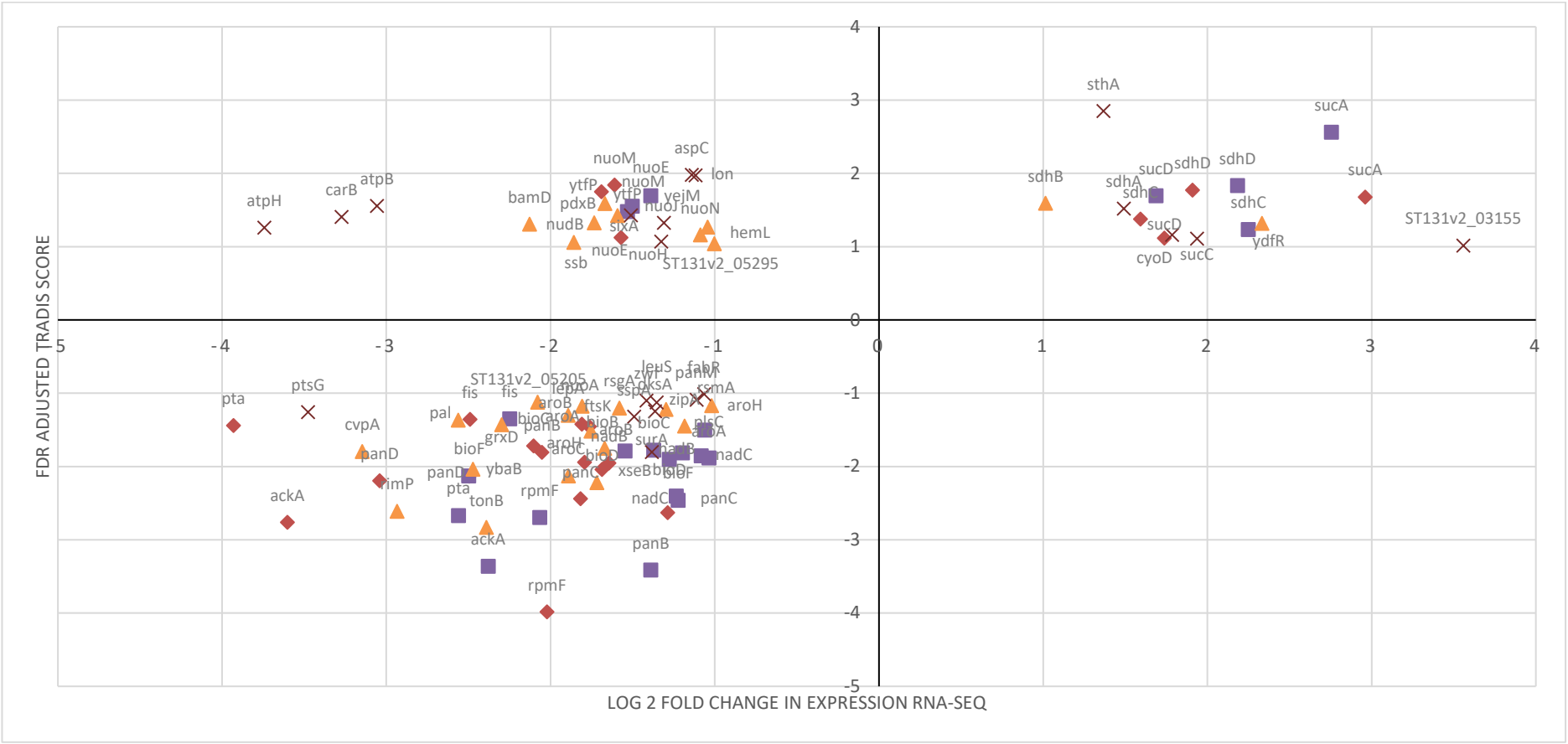


Figure 6.1 Genes in common between the TraDIS and RNAseq aerobic acetic acid data sets comparing the RNAseq log2-fold change in expression plotted against the FDR adjusted Essentials score indicated loss or gain of fitness in TraDIS. Each pH condition has been compared to its partner condition in the other data set i.e. pH 5.5 plus acetic acid RNAseq has been compared against pH 5.5 plus acetic acid TraDIS. X = In pH 7 data aerobic only, Δ = In pH 5.5 aerobic data only, ◊ = pH 5.5 aerobic data, shared across both pH conditions □ = pH 7 aerobic data, shared across both pH conditions (i.e. ◊ and □ show the same genes, but values are different depending on data set).

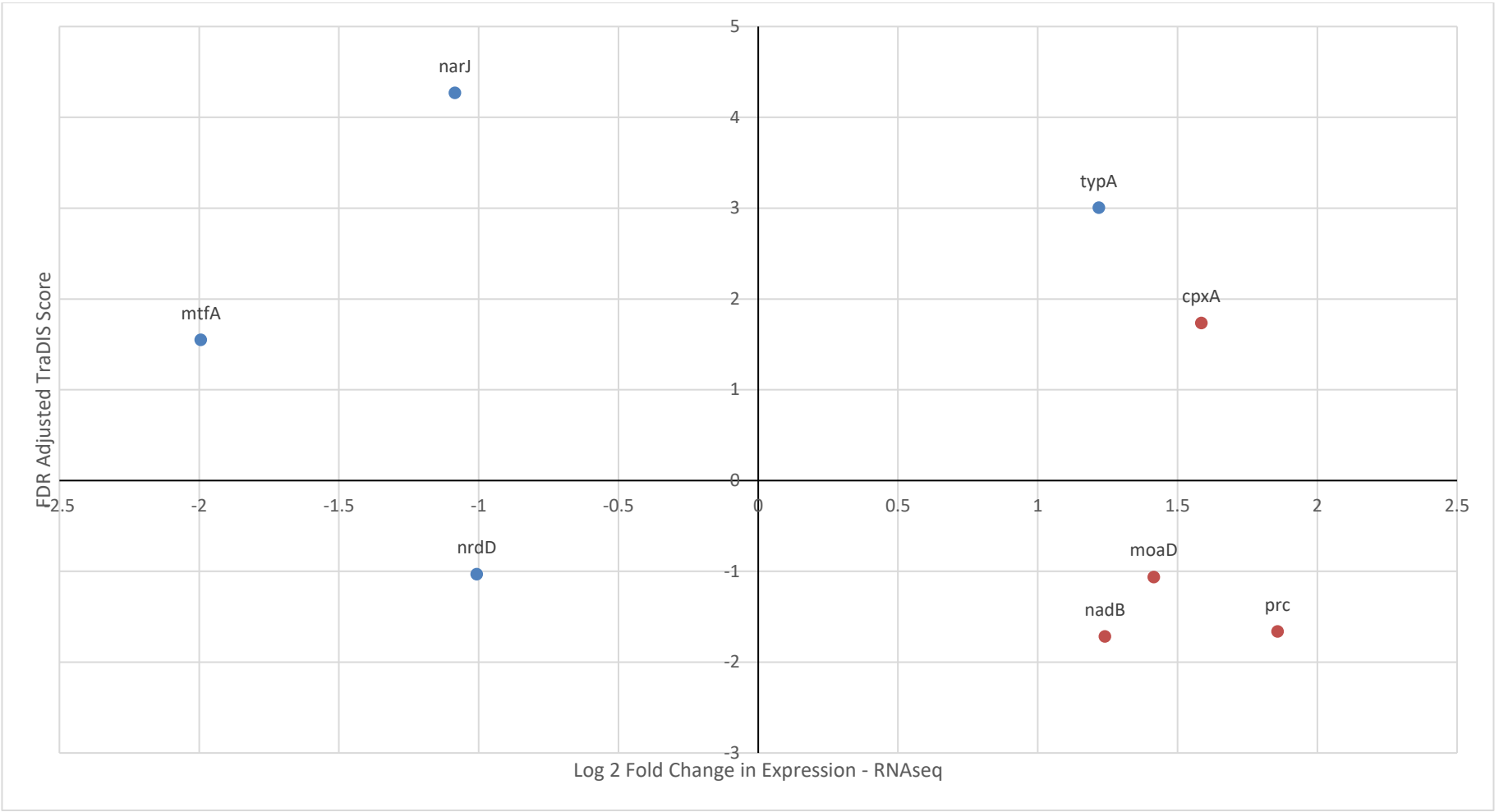


Figure 6.2 Genes in common between the TraDIS and RNAseq anaerobic acetic acid data sets comparing the RNAseq log2-fold change in expression plotted against the FDR adjusted Essentials score indicated loss or gain of fitness in TraDIS.. Each pH condition has been compared to its partner condition in the other data set i.e. pH 5.5 plus acetic acid RNAseq has been compared against pH 5.5 plus acetic acid TraDIS. Blue = pH 5.5, red = pH 7

6.4 Genes which overlap between RNAseq and lab-based evolution

This section will consider briefly the genes shared between those which were mutated and those which were differentially expressed in RNAseq. Genes that were mutated in at least one strain in the lab based evolution experiment, and also either significantly up-regulated or down-regulated, are shown in Figures 6.3 and 6.4 respectively.

There were a total of 368 genes with non-synonymous mutations across the six populations evolved at pH 5.5 with 4 mM acetic acid. Of those, 123 were shared with the genes significantly upregulated at pH 5.5 and pH 7 with acetic acid (figure 6.3), with lower numbers shared between just the mutated genes and those upregulated at either pH 5.5 or pH 7 with acetic acid. There were 51 genes shared between those that were mutated and those that were significantly downregulated at pH 5.5 and pH 7 with acetic acid (figure 6.4).

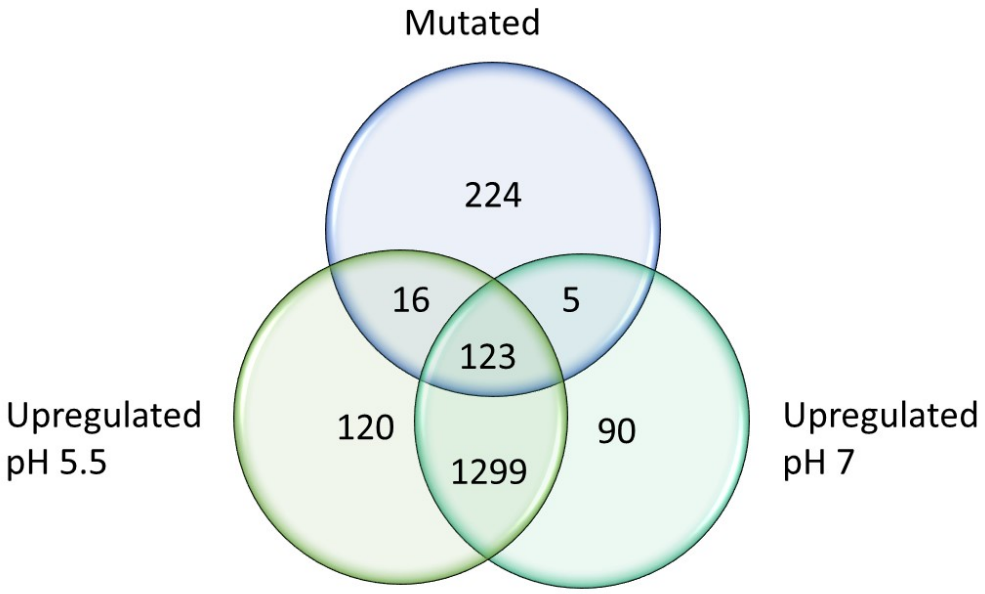


Figure 6.3 Venn diagram showing numbers of genes which overlap between those which are mutated and those which are significantly upregulated in the presence of acetic acid.

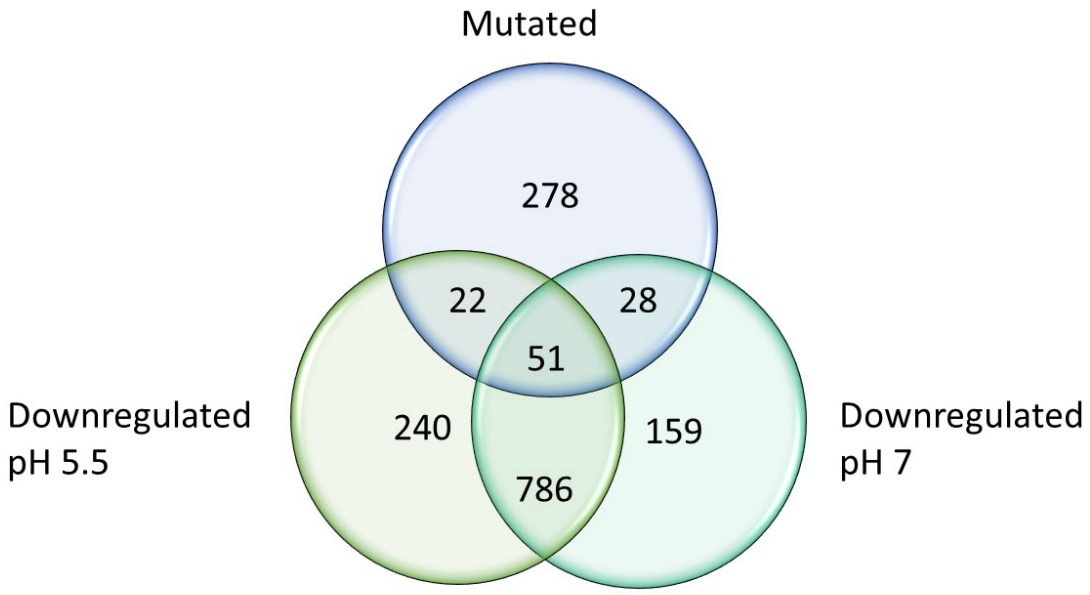


Figure 6.4 Venn diagram showing numbers of genes which overlap between those which are mutated and those which are significantly downregulated in the presence of acetic acid.

The genes that overlapped did not group into any significantly clustered pathways, however the list included genes such as *bglB, bglJ, cadC, fadE* and *fadH* which have all been discussed in Chapters Three and Four. This implies that there is some overlap in the genes identified by each of the two methods as to the pathways and genes being utilised under organic acid stress.

6.5 Discussion

The aim of this chapter was to look in more detail at the overlap between the three methods. As with other studies as outlined in the introduction to this chapter, there were few genes which came up across all three studies, and a very small number of differentially expressed genes which were also considered to have an impact on fitness, either negative or positive. That is largely due to the difference in how many genes were considered significant in either study, with fitness being a very different measure of cellular response to expression. However, in contrast to Zorraquino et al., (2016) who saw no genes in common

between their evolved populations and differentially expressed genes, the same was not true here, with quite a large proportion of the mutated genes also being differentially expressed. However, in terms of their actual results, there were genes in common with the data presented here. They saw mutations in their populations evolved in acid (pH 5.5) in *lon, pgi* and *relA*. Lon, a protease, has been discussed earlier in this thesis, however both *pgi* and *relA* are included in the list of genes presented in section 6.2 as having come up in all three methods used for acetic acid.

There were not enough genes in common here to do pathway analysis, however with more sophisticated analysis in combination with the evolution data once that has been more fully analysed, it would be possible to build up networks that are being utilised by the bacteria under organic acid stress. There are, however, clearly a number of pathways which have come up repeatedly in the last Chapters which would be a sensible route of further investigation. In terms of maintenance of cell homeostasis some of these are already relatively well understood, for example ion transport, such as those involving potassium ions and proton motive force. With regards to metabolism and energy transfer, these are glycolysis and fermentation in anaerobic cultures, the pentose phosphate pathway, the PTS and the TCA cycle and, again these are well understood. In Chapter One, at figure 1.5 there is a diagram showing how these pathways converge on the production of acetate. In terms of processes whereby organic acids may be impacting on other cellular functions, there were genes relating to protein misfolding and removal and downregulation of processes involved in DNA replication and turnover.

Here we have shown that all of these pathways are interlinked in the response to organic acid stress. That is likely because all of these various pathways either result in production or utilisation of acetate, and/or the utilisation of propionate. As described in Chapter One, the role of acetate in metabolism is well known and has been quite extensively investigated, both in order to understand cellular biology but for industrial purposes. The effects of butyrate and propionate however are less extensively covered in the literature,

and despite the industrial importance and of both compounds and the particular medical relevance of butyric acid given its role as an energy source for colonocytes, they are less well understood. The data here indicates that the responses to both butyric and propionic acid were in some ways quite unique. This is understandable given they are different compounds which could ultimately have very different effects on the cell.

Overall, however, this study indicates that the response of the cell to all three organic acids is generally the same, with a large proportion of the same genes showing both up and down regulation, many of the same genes showing improved or impacted fitness indicating the same major pathways as being of most use to the cell under organic acid stress, as outlined above. There were of course some differences which cannot be explained by looking solely at gene function. There were some particularly interesting results from the data derived from the propionic acid experiments, particularly in regards to genes which normally function under anaerobic conditions being apparently important under aerobic growth. It is possible that propionate is more readily utilised by the cell under anaerobic growth and so strategies employed by the cell to combat propionate stress under anaerobiosis are utilised under aerobic conditions as well.

The response to butyric acid again showed some interesting and individual pathways, such as tRNA modification genes which were not considered significant in any of the other conditions. There is an implication therefore that butyric acid interferes with mRNA translation. This is an interesting route of further investigation.

Overall, the pathways as outlined above are those of key interest. In terms of key genes of interest, the next step is to test these in *in vivo* to determine whether they afford EO 499 additional fitness in organic acid stress. This will be considered in Chapter Seven, where the genes for further investigation were

selected from those which were considered significant in both the RNAseq and TraDIS experiments, for the reason that those genes which are considered significant in more than one of the methods used in this thesis were more likely to play an important role in organic acid stress.

Chapter Seven:

Towards validation of RNAseq & TraDIS data: isolation & testing of a mutation in *ytfP*

Chapter 7 – Towards Validation of RNAseq and TraDIS data: isolation and testing of a mutation in *ytfP.*

The results of the last four chapters have identified genes which appear to be important in organic acid stress. The next step was to test whether the results of the TraDIS and RNAseq analysis could be confirmed in the laboratory using knockouts of genes which appeared to play a role in the response of EO 499 to organic acid stress. We initially intended to make knockouts and analyse the genes of interest in EO 499, and this was explored at great length by Dr Thippesh Sannasiddappa and then Dr Hadi Mohammed, post-doctoral researchers in the laboratory. EO 499 carries a mega-plasmid which encodes several nucleases, this potentially caused the destruction of the introduced plasmids or PCR products intended for targeting and knocking out the gene. It was, therefore, not possible to make knockouts in EO 499. This would have been the preferred method of exploring the results of the TraDIS and RNAseq experiments, as more genes could have been tested. We therefore decided to attempt to find mutants from the TraDIS library instead, given that the library should, by definition, contain all the mutants of interest. This had drawbacks for several reason. First, isolating the mutants proved difficult, which will be explained in more detail below. Second, the TraDIS mutants are not true knockouts, as the transposon insertions mean that the gene is still present in the genome, it has just been interrupted. Depending on where the transposon is inserted within the gene determines how much of the gene is translated into protein. As was discussed in Chapter Four, there are essential genes which contain transposons in dispensable parts of the gene, usually at the 3' end, indicating that the presence of the transposon does not always lead to loss of function. Therefore, even if it was shown to be possible to isolate the mutants, there was no guarantee that these mutants would give the phenotype of true knockouts. However, after considering the options available, including using an alternative background strain such as the Keio library and looking at alternative routes of making knockouts in EO 499, it was decided that finding the transposon mutants from the TraDIS library was worthy of exploration as an alternative way of analysing phenotypes.

7.1 Using a pooling strategy to isolate mutants

In order to find mutants in the TraDIS library the following plan of pooling, followed by PCR, was devised. The idea behind the approach was that it would be possible to calculate an estimated number of a particular mutant within the library. For example, if there were 5000 genes on the genome, and the transposons are equally distributed (this is an incorrect assumption) then on a plate of 15,400 individual mutants (200 mutants per well across 77 wells), then there would be at least 3 mutants on the plate carrying the mutated gene. Despite knowing that the assumption of equal distribution of insertions was incorrect, by ensuring that only genes with high frequency of transposon inserts were targeted, this would improve chances of finding a match.

First, 10 µl of the TraDIS library was grown overnight in 10 ml of buffered, supplemented M9 medium at pH 7 in a 20 ml glass universal bottle. This culture was then diluted to 1000 cells per 1 ml fresh M9 at pH 7 and 200 µl of this dilution was pipetted into each well of a 96 well plate (equating to 200 cells per 200 µl). The cultures were grown as described in Chapter Two. The following day, the plate was removed from incubation and 20 µl from each well was pooled either vertically or horizontally into the corresponding empty row or column. A fresh tip was used every time to prevent carry-over into different wells. Figure 7.1 shows a diagram of the pooling strategy, where the mutant is in well D4. Column 12 and Row H were left empty in each case so that a sample from each well could be pooled into the corresponding well. For D4, the mutant would be pooled in D12 and H4 and the next step would involve amplifying DNA from these wells using PCR.

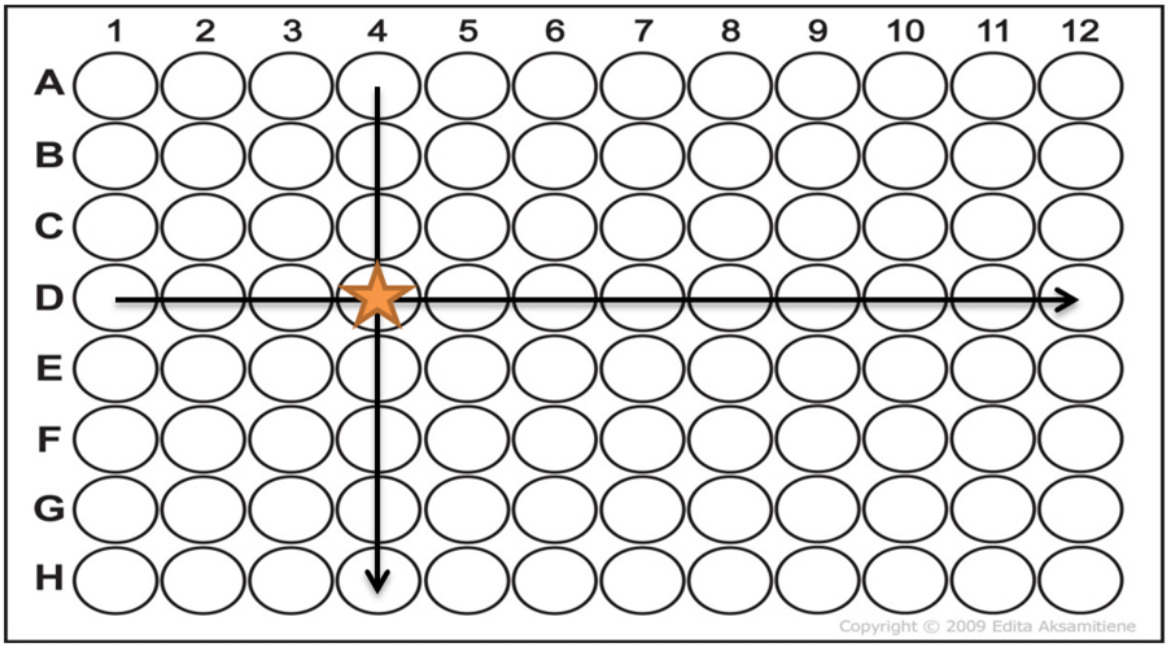


Figure 7.1 The layout of the 96 well plate, where the mutant of interest is in well D4 (represented by the star).

The pools were then tested using PCR to isolate the gene carrying the transposon insert. This was done using a combination of forward primer for gene of interest (F), paired with the transposon primer (Tn), or reverse primer for gene of interest (R) paired with transposon primer (Tn). The primer for the transposon was outward facing in order that it would pair with the forward or reverse primer of gene of interest, regardless of the orientation of the transposon and matched the 5' end of the transposon, amplifying the first 54 bases of the transposon only. Figure 7.2 below shows a diagram of the potential bands using the lac operon as an example. PCR reagents and cycling conditions can be found in Chapter Two. In total, for each plate there were eleven wells across row H (1 – 11) and seven wells down column 12 (A – G) making 18 pools and therefore 36 PCR reactions (18 reactions of F+Tn and 18 of R+Tn) plus controls per plate.

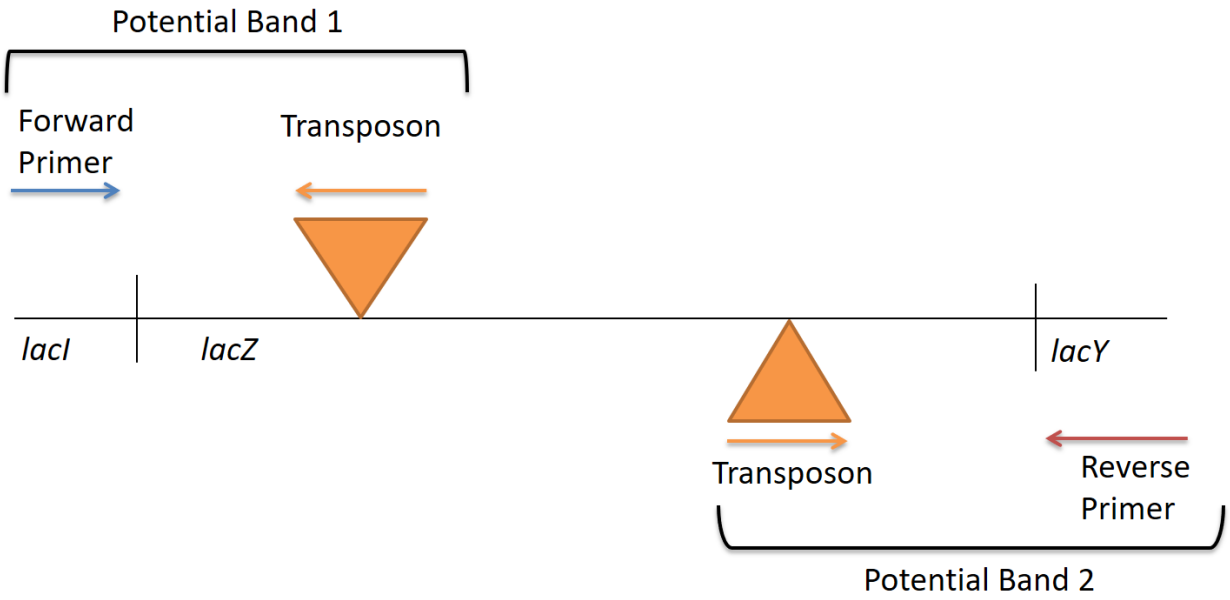


Figure 7.2 The *lac* operon and how the forward or reverse primer for the gene of interest (in this case *lacZ*) would pair with the transposon primer depending on orientation of the transposon.

Before using the method to attempt to find a specific mutant, the first step was to check the sensitivity of the method by testing its ability to detect a known mutation in a wildtype background. To do this, a Lac$^-$ mutant was first isolated from the library simply by plating the library on MacConkey Lactose agar and looking for white colonies. Plating around 10,000 colonies revealed three white colonies. These colonies were grown overnight and a PCR reaction was used to determine where within the *lac* operon the transposon was located using three primer pairs, spanning *lacZ* and *lacY,* the genes within the lac operon which could give rise to the appearance of white colonies if they were interrupted with a transposon. A diagram of these is shown in Figure 7.3. The primer pair which gave a band approximately 1KB larger than the known gene size would indicate the presence of the transposon.

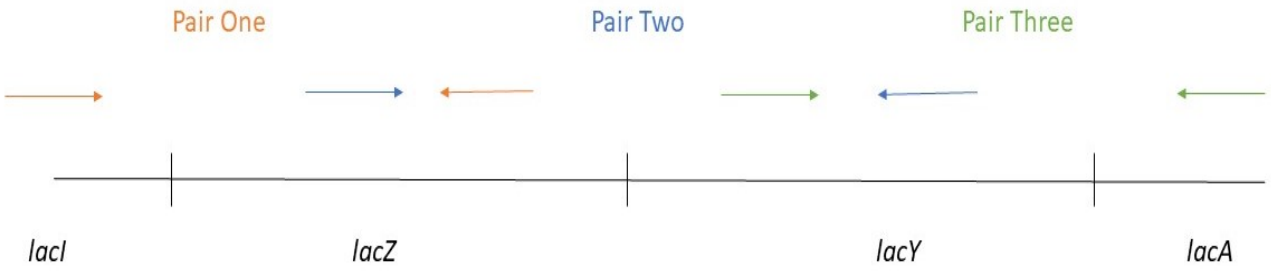


Figure 7.3 Three overlapping primer pairs were used to map the operon using PCR.

The location of the transposon within the three white colonies was then determined by combining each forward and reverse primer with the transposon primer, as described above, with the subsequent band size indicating how close to the start or end of the gene the transposon lay. Figure 7.4 shows the position of the transposon within each of the three isolated mutants.

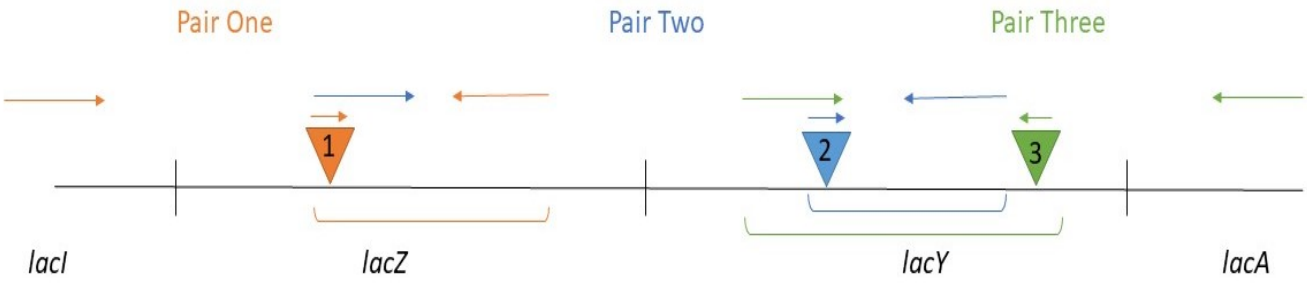


Figure 7.4 Position of the tranposons within the *lac* operon in the three isolated mutants and the primer pairs used to confirm their position. In mutant 1 (orange) the reverse primer mapping region one was paired with the transposon (tn) primer, in mutant 2 (blue) the reverse primer of region two was paired with the tn primer, and for mutant three (green) the forward primer of region three was paired with the tn primer.

These mutants were then used to artificially spike a culture of the transposon library at different dilutions of 50/50, 1 in 10, 1 in 100, 1 in 1000, 1 in 10,000 and 1 in 100,000. These cultures were used for a PCR reaction using the *lac* primer and transposon primer, which would establish the sensitivity of the primer pair against the background of the other transposon mutants. Figure 7.5 shows a gel of this reaction and reveals a band was visible up to a 1 in 100,000 dilution. However, it was becoming faint and in the context of searching for unknown bands, where the number of target mutants in the starting library was unknown, this would be unreliable. Therefore, the maximum dilution that would be considered reliable was 1/10,000.

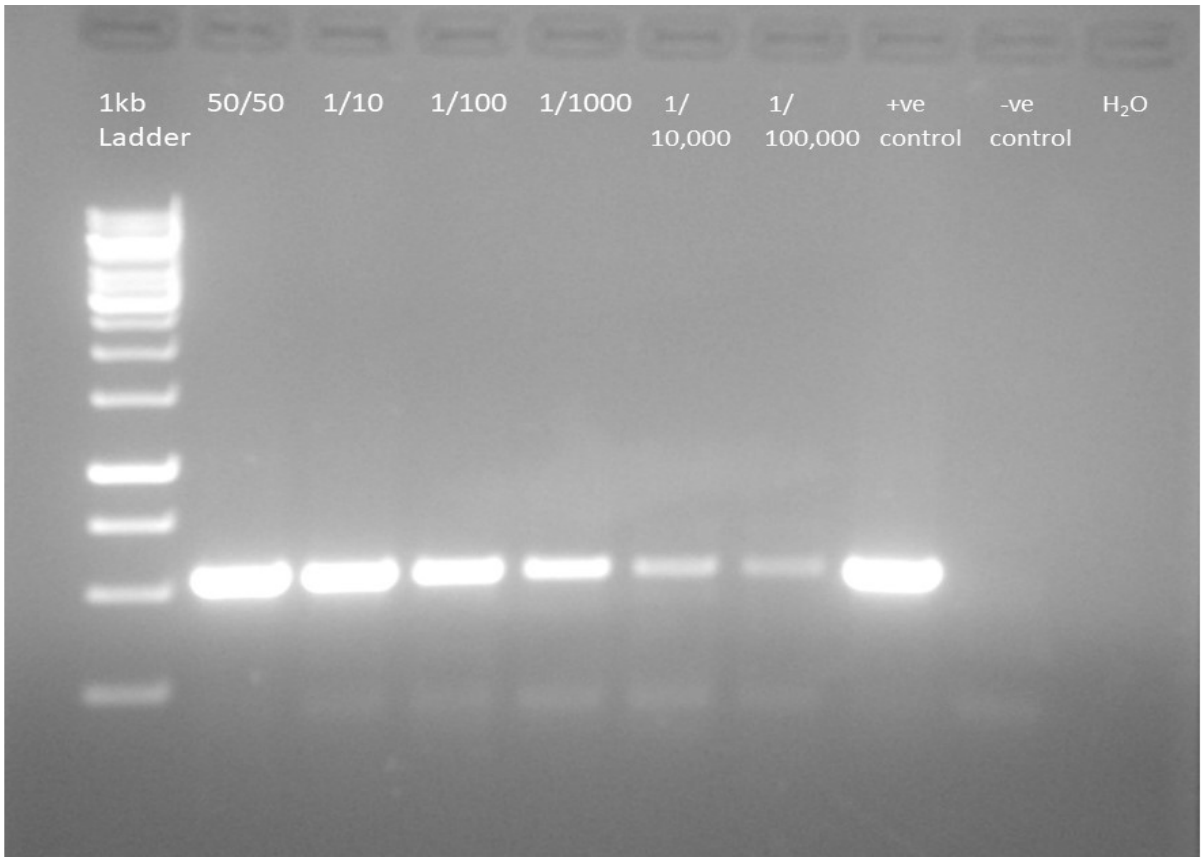


Figure 7.5 PCR results of dilutions of the known *lac* mutant into the traDIS library to ascertain sensitivity of the method.

The next step was to use the *lac* mutant in the pooling experiments with a library background to establish how sensitive the method of screening was for the plate pooling method. After the mutants proved relatively easy to find using a dilution of up to 1 in 100,000 cells against the transposon library background culture, an overnight of the transposon library was again seeded with the known mutant at a dilution of 1 in 10,000, diluted to 200 cells per well and plated. At this dilution, the plate would in theory give at least one well, and hopefully more, with the *lac* mutation and should mirror a real experiment of finding unknown mutants. Figure 7.6 below shows the results of the PCR reactions of each well, with matching bands circled. There are four pairs of band which have been circled. Two of these match the spiked mutant and two are different. This culture was only spiked with one of the *lac* mutants, not three (with an expected band size of approximately 500 bases), therefore the additional two bands were new mutants which had not been artificially added. The gel in figure 7.6 also shows lots of additional bands. This is likely due to the primers in the reaction matching all of the mutants in the reaction (just not as a pair) which has resulted in amplification of random parts of the genome where the transposon sits, or in some cases may pair with the *lac* primer but the transposon primer further upstream or downstream of the *lac* operon (and not within the *lac* operon). This was not investigated further.

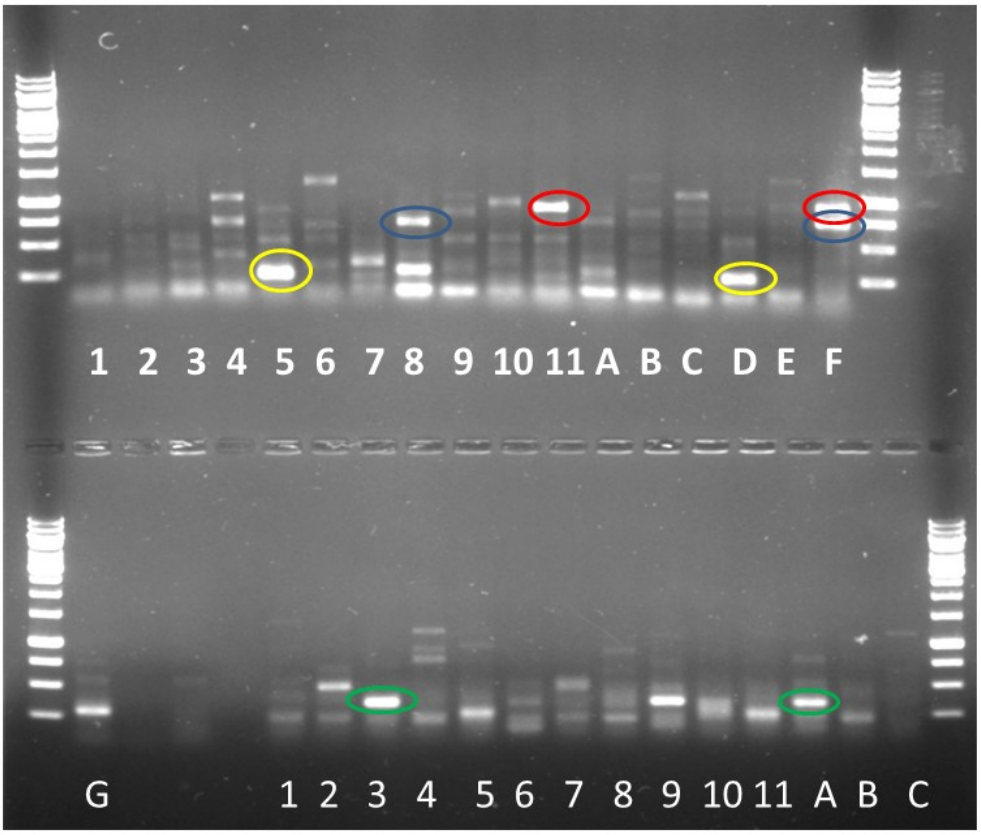


Figure 7.6 PCR result of the 96 well plate experiment using the known *lac* mutant. Numbers represent plate column and letters represent plate rows. The gel shows a large number of different signals, however there are some matching pairs which have been outlined in yellow (D5), blue (F8), red (F11) and green (A3).

All of these four wells were plated on MacConkey Lactose agar to see whether they actual did have *lac* mutants within them. All plates showed white colonies. Figure 7.7 below shows the plate of well A3 after overnight incubation at 37^{o} C. The plate had six white colonies. This is good indication that this was one of the spiked wells as the band matched that of the previously isolated *lac* mutant. The additional two mutants found shows that the estimation of approximately 1 in 5000 mutants being the mutant of interest would yield approximately 3 mutants per plate is roughly correct, but cannot be held as a rule given the uneven distribution of inserts, depending on the gene, and the frequency of transposons within

the gene. However, this was a positive result and gave cautious encouragement of this method working to find other mutants which may be found in the library at similar frequency to the lac mutants.

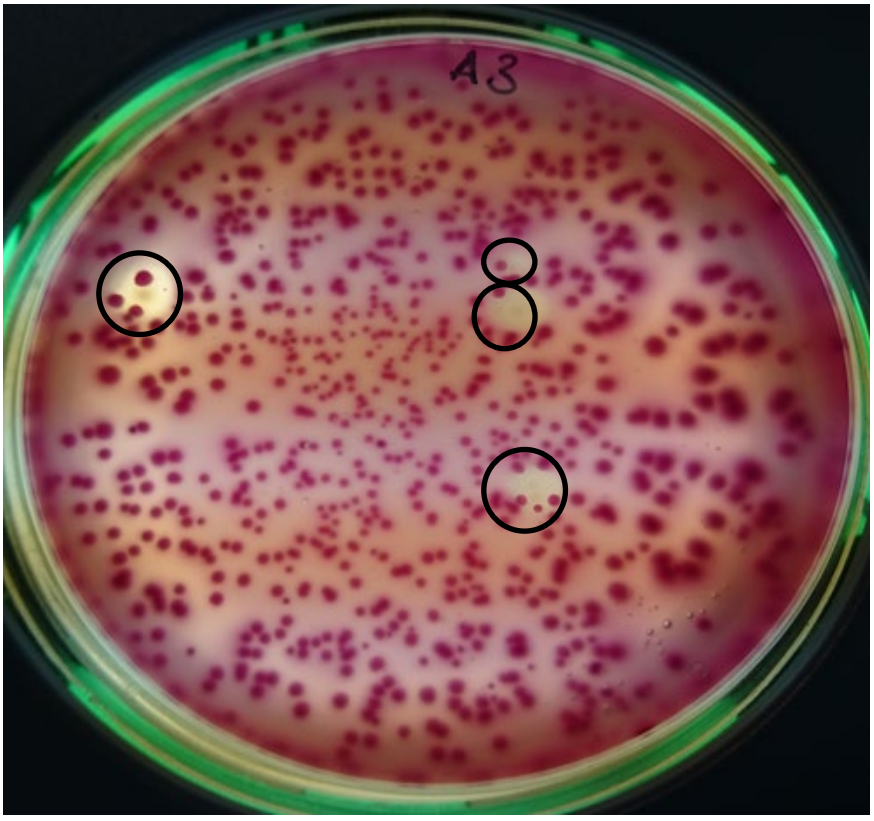


Figure 7.7 Plate of well A3 suspected to contain the spiked *lac* mutant.

The next step was finding unknown insertion sites, which meant there was no control reaction indicating whether the PCR reaction had worked or not in amplifying an unknown band size. The first to be targeted were mutants with insertions in the *nuo* operon, which was identified in both the TraDIS and RNAseq analysis as being potentially important for the acid response. A detailed description of these genes and the reason for their suspected role within the organic acid response can be found in Chapters Three and Four. Primers were designed for *nuoM* and *nuoF*. However, despite repeated attempts by both Dr Hadi Mohammed and and as part of this project, it did not appear possible to isolate these mutants. The methodology was changed to one of nested PCR where the flanking primers were first used to amplify the

region containing the gene (and hopefully the transposon) and then the transposon primer introduced to amplify any *nuo* mutants containing the transposon. This would hopefully amplify the specific transposon containing region in wells where there were particularly low frequency of target mutants. The two rounds of PCR were attempted as indicated in figure 7.8 below. This also proved unsuccessful and it was concluded that even if the transposon mutant was present within the culture, the smaller band of the non-interrupted *nuo* gene would be amplified in preference to the 1KB larger transposon mutant, meaning there would be almost no chance of isolating the much larger gene-plus-transposon product using this method. An alternative method for nested PCR was devised whereby two sets of flanking primers were designed, one external to the gene of interest and one internal, paired with the transposon primer, as shown in figure 7.9. Again this proved unsuccessful.

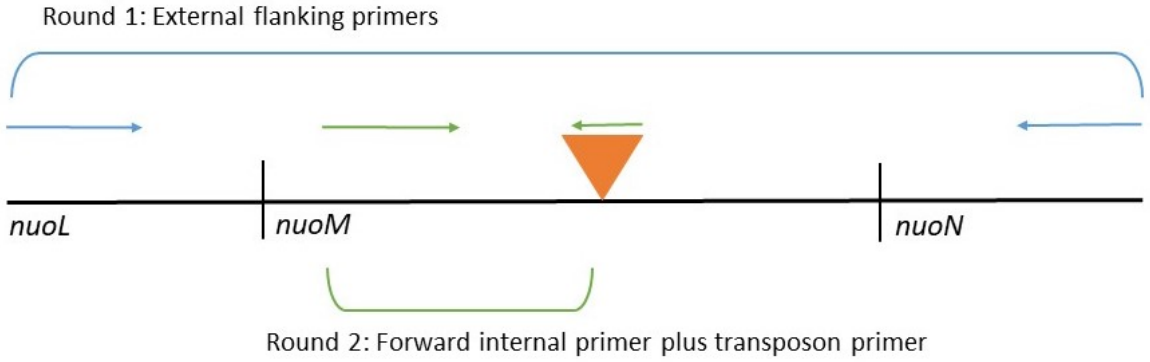


Figure 7.8 A representation of the first round of nested PCR, where the flanking primers were used for the first round and the internal forward primer and transposon primer were used for round two. The experiment was repeated using the reverse internal primer and transposon primer for round two.

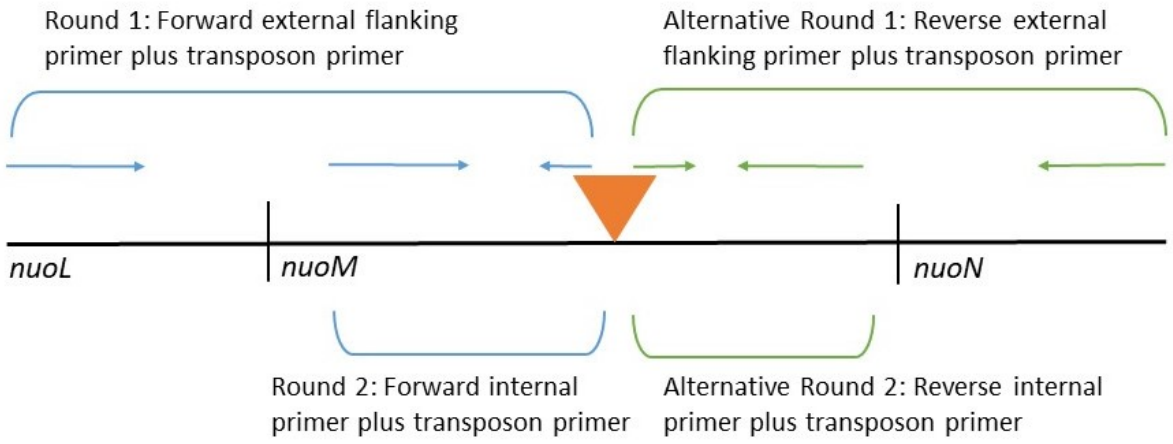


Figure 7.9 The second nested PCR attempt, using an external flanking primer and the tn primer for round one and an internal primer and tn primer for round two.

On further consideration of the saturation of transposons within the *nuo* operon, which is low overall compared to other regions of the genome, it was concluded that the number of transposon inserts within this region was too low to be able to find the mutants using this method. For example, the insertion index, calculated as the number of insertions found in a gene divided by the length of the gene, was 0.08 for *lacZ*. For *nuoM* it was 0.01, for *nuoG* it was 0.02 and for *nuoF* it was 0.01. On this basis, the decision was subsequently made to target genes with higher transposon saturation. The next gene selected for target on the basis of the sequencing analysis was *ytfP,* which was considered significant in both the TraDIS and RNAseq experiments. Investigation of the coverage of *ytfP* also showed high saturation of transposons within the gene (insertion index of 0.08, the same as *lacZ*) making it more likely that it would be possible to isolate individual mutants.

The transposon library was grown overnight and diluted into three plates so that there was a starting concentration of approximately 200 cells per well within the 96 well plate, as per the control experiments. This was confirmed via plating on LB agar. Three plates were used, to maximise chances of finding a

mutant. For one of the plates, after the first round of PCR, there were four potential column locations indicated by the gel in figure 7.10 below. Unfortunately, there were no corresponding row locations, with no bands appearing at all in A – G. Therefore, every well from the four columns indicated below: 3, 4, 7 and 8, were grown overnight, and those cultures used the following day for PCR using the forward flanking primer and the transposon primer.

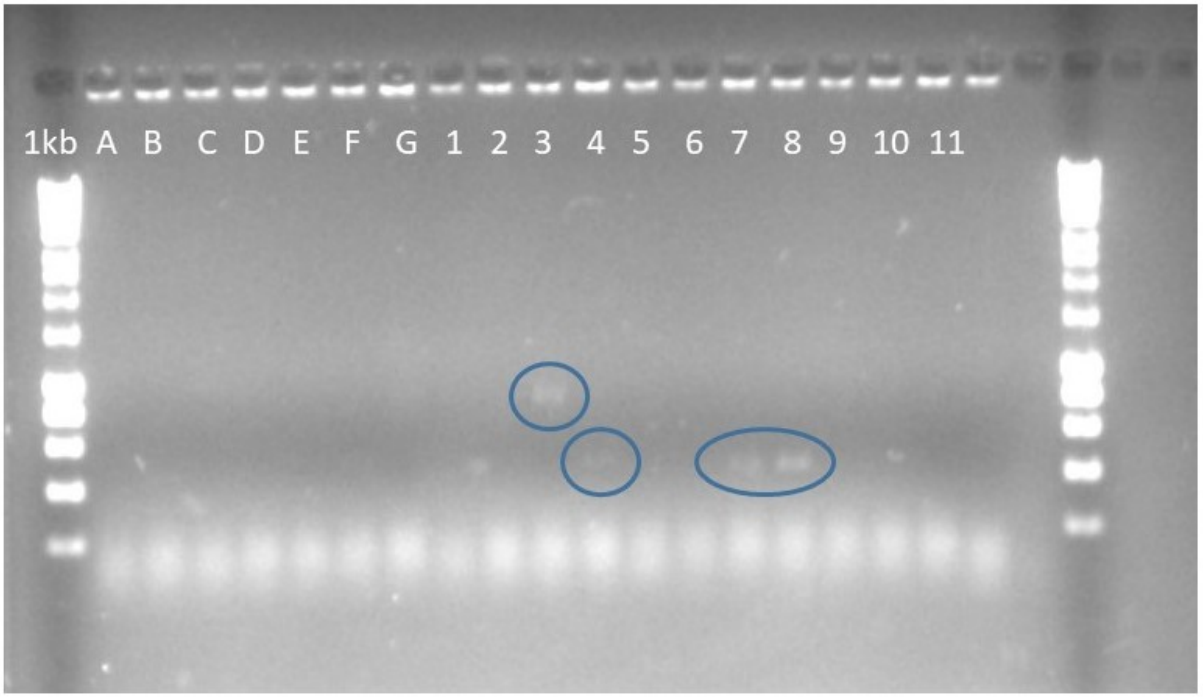


Figure 7.10 PCR result of the attempt to find *ytfP* transposon mutants using the 96 well plate method.

There were three bands identified following this next round of PCR of 28 wells (4 columns with 7 wells per column). Figure 7.11 show bands in wells 2 and 7 from column 3 and well 2 from column 8.

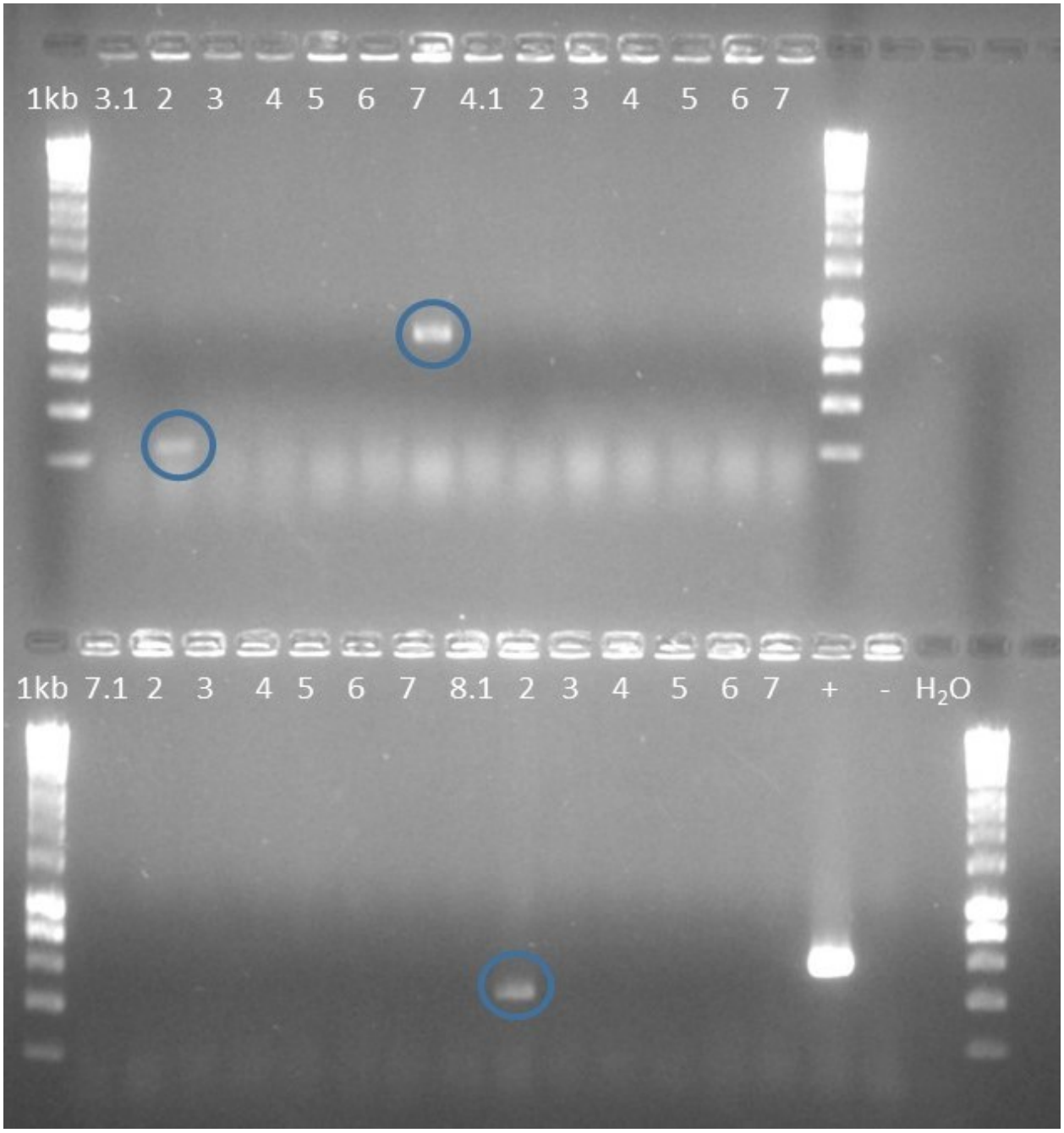


Figure 7.11 PCR result for the cultures grown from the columns indicated at figure… above. The gel shows three bands in wells 3.2, 3.7 and 8.2.

The next step in isolating individual mutants was individual colony PCR. As such, cultures from these wells was diluted and plated on LB agar. Three plates per well were used for subsequent colony PCR, and a total of 150 colonies were tested, 50 from each suspected location. Colonies grown from wells 3.7 and 8.2

showed no match with the *ytfP* forward flanking primer and transposon primer. Three colonies from 3.2 showed a band. These were confirmed by, first, using PCR. Figure 7.12 below shows a gel in which a culture grown from each of the three isolated colonies was grown overnight, and this culture tested in triplicate using the external flanking primers, which should give a band approximately 1kb larger than *ytfP* wildtype. The transposon was inserted approximately 100 bases into the *ytfP* gene in these particular mutants and so the band is much smaller than the other two (approximately 150 bases) and can be seen in the last nine lanes on the gel in figure 7.12. The middle band is WT EO 499 showing band size of WT *ytfP* (559 bases), and the first nine bands show the result from the flanking primers, indicating the presence of the 1kb transposon.

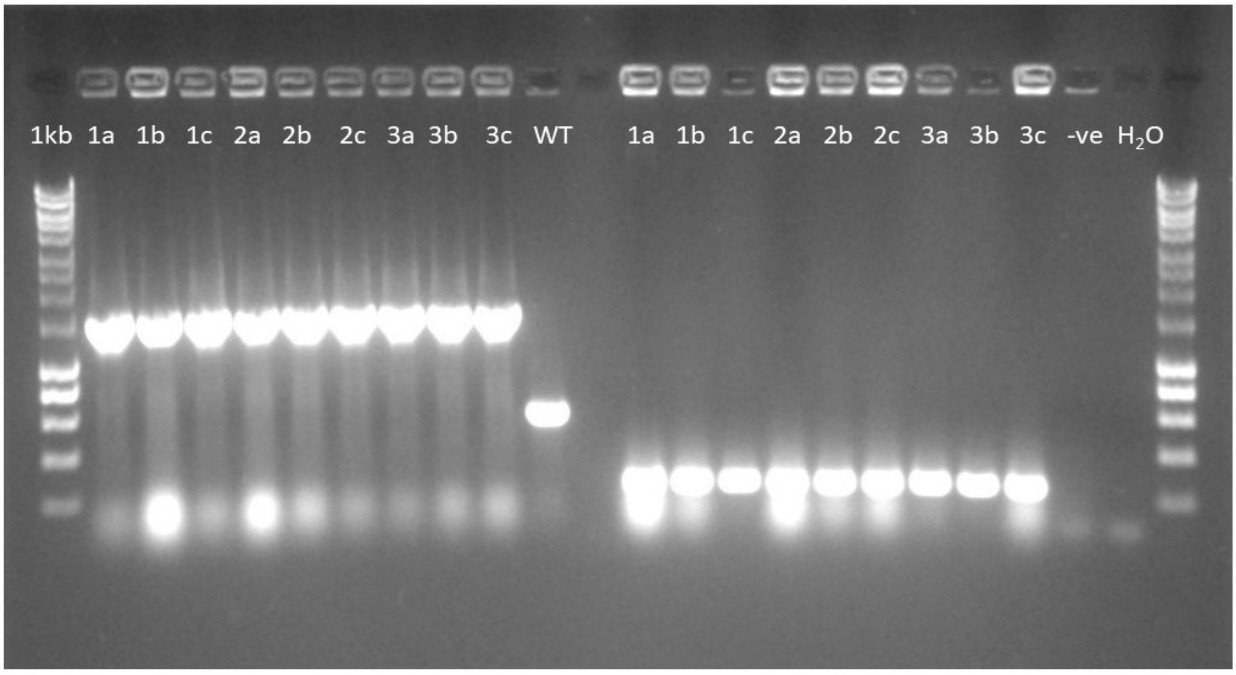


Figure 7.12 Confirmation gel showing the three mutants isolated. The first nine wells show results of just flanking primers, compared to flanking primer used with wild type. The mutants show a band 1kb bigger than the wild type cells. The second batch of nine bands is from the mutants using flanking primer paired with tn primer which shows a smaller band than wild type.

In order to confirm that these were in fact transposon inserts in the *ytfP* gene, a gel extraction was done and the DNA was sequenced, confirming the transposon insert was 116 bases downstream of the 5' end of the gene.

Following on from successfully isolating *ytfP* mutants, the next mutants selected for isolation were *rssB* and *apaH* for reason as outlined in Chapters Three and Four. These two genes had an insertion index of 0.04 for both, which was half of *ytfP* but four times higher than *nuoG.* Given their importance to this study it was considered appropriate to attempt to isolate these mutants. However, repeated attempts to isolating these particular mutants were unsuccessful.

7.2 Analysis of the *ytfP* phenotype using competition experiments

The overall intention in isolating these mutants was to test whether they were less able than wild type EO 499 to compete in organic acid stress in order to test the hypothesis that this gene is important under organic acid stress. The first step was to ensure there was no competitive advantage or disadvantage conferred by the transposon within *lacZ* and *lacY* as one of these mutants would be used as a proxy for wild type given it was not possible to construct a *lac*- mutant in EO 499. It was important to ensure that the *lac*- TraDIS mutant had the same phenotype as EO 499 wild type in order to ensure accurate results in the competition experiments. To test this, wild type EO 499 was competed against the three Lac^- mutants. The bacteria were competed at pH 7, pH 7 plus 40 mM acetic acid, pH 5.5 and pH 5.5 plus 4 mM acetic acid, reflecting the conditions used in the traDIS experiment. Below at figure 7.13 is the result of the competition experiment. Each mutant and condition was tested in triplicate. A competitive index of 1 indicates no competitive difference. The competition experiments and the competitive index were calculated as described in Chapter Two.

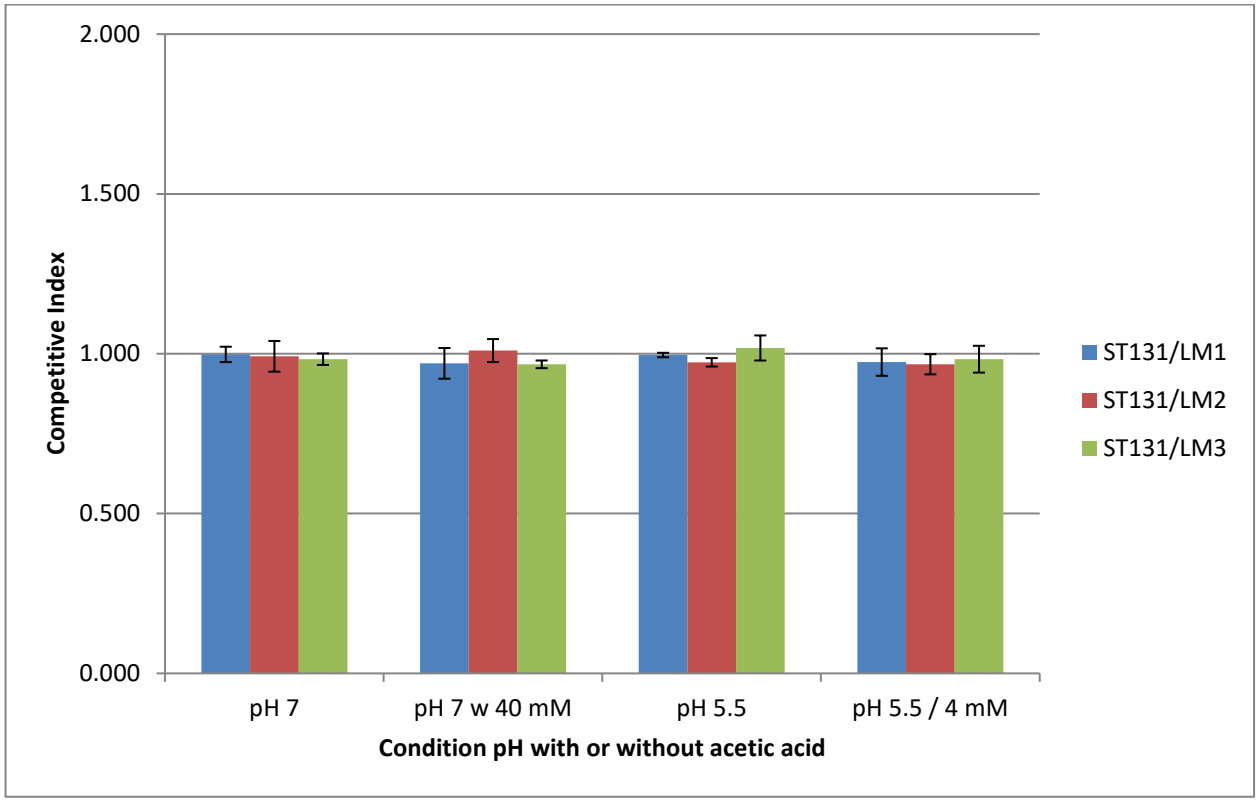


Figure 7.13 Competition experiment of wild type ST131 against the *lac* mutants (called LM1, 2 or 3) in all test conditions.

As the Lac$^-$ mutants and wild type showed no difference in the test conditions, the Lac$^-$ mutants were used for competition experiments with the *ytfP* mutant. The TraDIS experiment indicated that these *ytfP* mutants would display a competitive disadvantage in the test conditions. However, figure 7.14 below shows no difference between the Lac$^-$ mutant and the *ytfP* mutant.

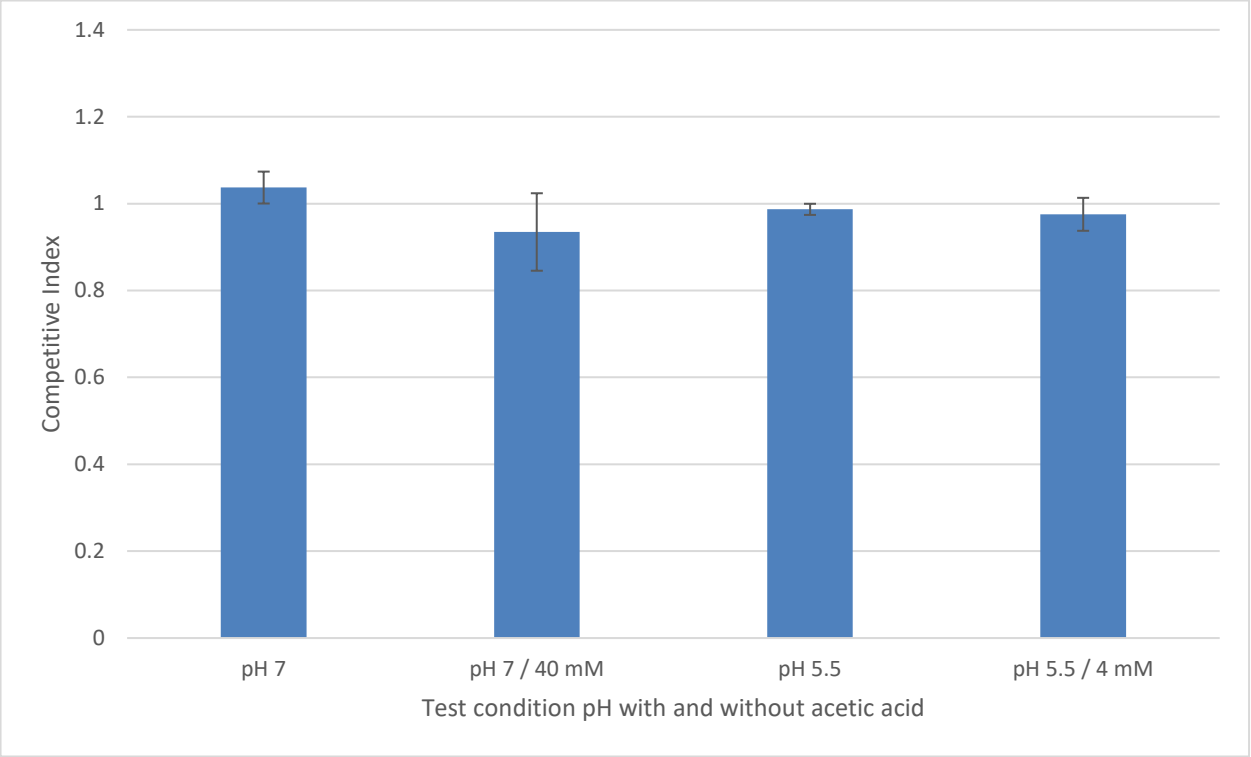


Figure 7.14 Competition experiment of transposon *lac* mutant versus the *ytfP* transposon mutant.

The result here indicates that the transposon insert into the *ytfP* gene did not affect its ability to compete in the test conditions. This result does not therefore validate the hypothesis from the TraDIS experiment. A PhD student in the group, Fatemah Alatar, did the same competition experiment using the Keio library knockout of *ytfP*, competed against a Lac^+ mutant of the background strain BW25113. For several reasons this may provide a different result to using EO 499. Primarily, EO 499 is a commensal organism and an opportunistic pathogen, whereas BW25113 is a non-pathogenic lab strain. Therefore, EO 499 has a higher tolerance to organic acids and low pH than BW25113 due to its normal environment of the intestinal tract. However, in order to make some headway with testing the TraDIS and RNAseq analysis, many of the genes of interest were tested by Fatemah Alatar including *ytfP*. The result also showed no competitive difference between the WT and mutant strain. Two additional conditions were tested: pH 7 with 100 mM acetic acid and pH 5.5 with 10 mM acetic acid. However, these conditions also showed no difference

between WT and *ytfP* mutant. Figure 7.15 below shows Ms Alatar's results from the competition of Lac^{+} BW25113 and the Keio *ytfP* knockout.

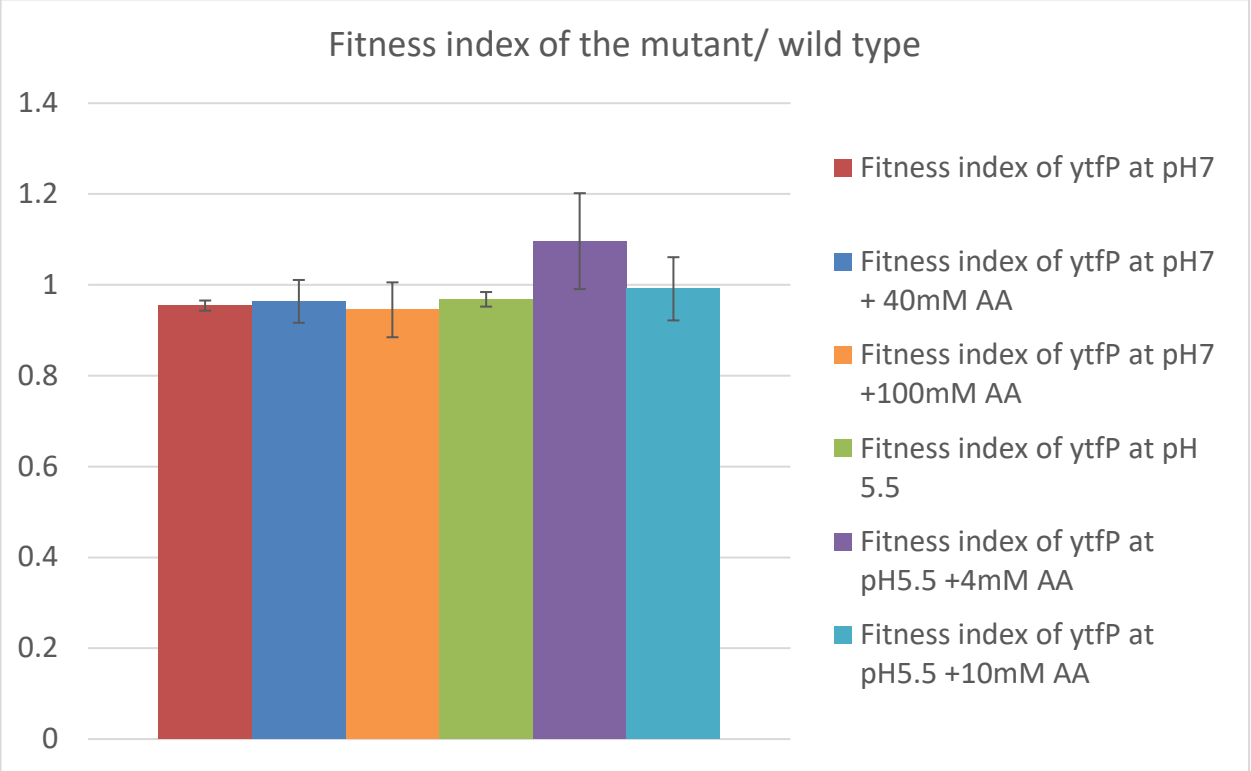


Figure 7.15 Data collated by Fatemah Alatar showing lac+ BW25113 versus Keio *ytfP* mutant in six conditions: those reflecting TraDIS and RNAseq, and two additional conditions reflecting harsher conditions.

7.3 Discussion

It was unexpectedly challenging to isolate the mutations, either by KO or by searching for transposon mutants, and this meant that it was not possible to fully validate the results of the TraDIS and RNAseq experiments in the manner which had been envisaged at the start of this PhD. However, due to the development of an alternative method to find TraDIS mutants, it was possible to test one of the genes which was indicated to play a role in acetic acid resistance. The biological function of *ytfP* is unknown, however it shares an operon with *tamA* and *tamB*, which are transmembrane secretion proteins (Selkrig et al., 2012). The results of the competition experiments with wild-type *ytfP* showed no competitive advantage compared to the *ytfP* TraDIS mutant.

The method of pooling in order to find TraDIS insertion mutants proved successful but highly inefficient, with large amounts of time and resources required to isolate a single mutant. A further complicating factor was transposon coverage within a gene of interest, making it difficult to isolate mutants with a low insertion index. This made it impossible to find and test these mutants. However, the isolation of *ytfP* was useful in the sense that it confirmed the results obtain from the competition experiments using the Keio library. Although these experiments were done in a different background strain of *E. coli* and, therefore, must be analysed as such, the fact that the EO 499 *ytfP* transposon mutant and the Keio *ytfP* mutant behaved in the same way adds weight to the other competition experiments done using the Keio library.

The competition experiments with *ytfP* with both the transposon mutant and the Keio mutant failed to validate the TraDIS and RNAseq data in this instance. It is not possible to know whether that would have been the case with other mutants, such as *apaH* or *rssB,* or with a true *ytfP* EO 499 knockout strain but it does imply that the TraDIS and RNAseq data should be treated with some caution when making inferences about the cellular response of bacteria to test conditions based on the transcriptional and/or transposon mutants' phenotypic responses. There was insufficient time to test the EO 499 *ytfP* transposon

mutant in more extreme conditions, to see the effect of higher concentrations of organic acid or lower pH, however this would make an interesting follow-up experiment. An option for future experiments would be to use either the Keio library, or an alternative UPEC strain in which knockouts can be made, in order to either validate the data presented in this thesis or generate new sets of data in this alternative strain so that directly comparable validation can be attempted.

Chapter Eight:

The effects of organic acid on growth and biofilm formation in *Pseudomonas aeruginosa*

Chapter 8 - The effects of organic acid on growth and biofilm formation in *Pseudomonas aeruginosa.*

8.1 Organic acids as agents to treat *P. aeruginosa* infections. A review of clinical and experimental data.

Weak organic acids are more able than strong inorganic acids to cross the cell membrane and lower the cytoplasmic pH, with a bacteriostatic effect at low concentrations, as described in detail in chapter one. The use of acetic acid to treat wounds, such as burns, is well established (Owen, 1946; Phillips et al., 1968). However, there has been limited research conducted into the effectiveness of acetic acid as a sole topical treatment for burns, especially against the multi-drug resistant (MDR) bacteria that typically colonise and infect wounds. There is also limited research into how acetic acid disrupts biofilm formation and whether bacteria are able to evolve resistance to the effects of weak organic acids over time. This last point seems particularly pertinent given the need to develop remedies against MDR pathogens which will be effective in the long-term. Literature on the effectiveness of acetic acid as a topical treatment is lacking (Halstead et al., 2015). Also, the number of investigations into alternative organic acids to acetic acid is surprisingly limited. Citric acid, one of the few other organic acids which has been considered by researchers and medics, has been identified as a useful antimicrobial agent against *P. aeruginosa*. (Nagoba et al., 2010; Yabanoglu et al., 2013). In medical studies, citric acid facilitated complete wound healing in 86.95% of cases, whereas conventional antibiotic therapy resulted in just 45% of patients demonstrating complete healing. However, it is also reported that citric acid was not an effective agent against the MDR strains tested (Yabanoglu et al., 2013). Most organic acids, including sorbic acid, benzoic acid and lactic acid do not appear to have been considered at all.

In historical studies, treatment using 5% acetic acid (approx. 800 mM) has been highly successful, eradicating all traces of infection. However, although there is a high level of effectiveness in treating *P. aeruginosa* infections, other bacterial species such as *S. aureus, S. pyogenes* and *E. coli,* which either appear after the elimination of the *P. aeruginosa* infection or are present throughout, prove harder to eliminate using the same methods (Phillips et al., 1968). Also, once treatment ceases *P. aeruginosa*

infections recur, with there being absolutely no effect on the level of infection at all in some cases (Phillips et al., 1968; Sloss et al., 1993). Previous studies demonstrate that although acetic acid treatment is successful against some more persistent gram-negative species such as *P. aeruginosa,* it has, overall, a narrow spectrum as an antimicrobial agent and was generally not effective against gram-positive and some *Proteus* species. This has been confirmed using *in vitro* tests, which show that acetic acid is inhibitory to *P. aeruginosa* growth at 2%. However, 8 – 10% acetic acid solution is required before the same effect is seen in *S. aureus* (with minimum inhibitory concentration of 4% for *E. coli* clinical isolates) (Sloss et al., 1993).

Studies with a wider scope have looked at a broader range of pathogens isolated from burn wounds and cross-compared their results from acetic acid treatment against commercially available antiseptic topical treatments being used in burn centres. It has been shown that acetic acid can be highly effective against gram-negative species such as *P. aeruginosa, Acinetobacter baumannii* and *P. vulgaris* at a concentration of 3% after only 5 minutes incubation at 37^{o}C, and after 30 minutes *E. coli* was also eliminated. It should be noted that this is in contradiction to the study cited above (Sloss et al., 1993), which demonstrated a MIC of 4%. A solution of 3% (~500 mM) was chosen as previous studies using 5% (~800 mM) solution produced complaints from patients of stinging and itching, although this study was carried out *in vitro* and not *in vivo* (Ryssel et al., 2009). A study examining *in vivo* application, post-operatively, showed a 3% solution of acetic acid to be successful against infection when combined with wound washing over a 10 day period. Importantly, it should also be noted that the application of acetic acid onto some wounds caused a large amount of pain, and in some cases, treatment had to be suspended. This seems significant when determining whether this sort of treatment is a realistic option for young children (Nagoba et al., 1997; Phillips et al., 1968; Ryssel et al., 2009).

As a result of the variation and contradictions within the literature, there is uncertainty about the most effective concentration solution to use when weighed against patient discomfort. A solution of 1% (~174 mM) acetic acid against clinical isolates of *P. aeruginosa* (2 strains)*, E. coli, S. aureus, S. haemolyticus* and *E. faecalis* was used in four *in vitro* burn models, the first where the solution was applied immediately to the burn, the second after 4 hours, the third after 6 hours and the fourth after 24 hours. The authors note that the solution was highly effective against *P. aeruginosa* after 0, 4 and 6 hours but lost efficacy when applied after 24 hours post-burn. This may indicate that once a biofilm had formed acetic acid lost its effectiveness (Hajská et al., 2014). It is also noted that the 1% solution had low effect against *E. coli* if immediately applied, and population counts showed no difference to the control group of *E. coli* if applied 4, 6 or 24 hours post-burn. This result is in general agreement with the studies above, which report an MIC for *E. coli* of between 3 – 4 %. An inhibitory effect against *P. aeruginosa* at concentrations as low as 0.166% and 0.312% and also against *A. baumannii* was demonstrated (Fraise et al., 2013). This study also showed that evaporation played no role in decreased efficacy of acetic acid treatment.

The laboratory-based (as opposed to clinical) study published by Halstead et al., (2015) focused on a wider range of bacterial strains and species, including eight strains of *P. aeruginosa,* seven strains of *A. baumannii*, four strains of *S. aureus* and two strains of *E. coli*, as well as others. There were three objectives of this study: to identify minimum inhibitory concentrations (MIC); minimum biofilm inhibitory concentration (MBIC); and minimum biofilm elimination concentration (MBEC) of acetic acid against these model strains and clinical isolates. Bacteria were grown in a range of acetic acid concentrations, from 0.026% to 1.6% at 37^{o}C. Growth was shown to be inhibited between a range of 0.16% to 0.31% (approximately 30 mM to 50 mM). Biofilm formation (measured in the 23 strains which reliably form biofilm) was shown to be inhibited across the same range. For biofilm elimination, however, a wider range of 0.1% to 2.5% (approximately 400 mM) was required to eliminate existing biofilm (grown at 33^{o}C to mimic the temperature of a surface wound). This study was able to support the use of acetic acid in

managing bacterial load in surface wounds. However, the synergistic effect of acetic acid combined with low pH was not tested (Halstead et al., 2015). As described in chapter one, this is important for a number of reasons and as such is a key point of focus within this chapter.

8.1.1 The role of biofilms in persistent infections

Biofilms are widely acknowledged to be a cause of sepsis in hospitalised patients, usually as a result of catheter use and central venous lines. However, it has been shown that bacteria are able to form biofilms within the burn wound itself. This is caused by a break in the skin integrity, and the most common source of bacteria into wounds comes from the patient's own skin flora, which colonise and infect the wound (Asati and Chaudhary, 2017). Biofilm formation has been noted in chronic wounds, such as those caused by diabetes, but are less common in acute wounds such as burns, given the standard medical practice of quickly excising the wound (surgical removal of the entire burned area, which is suitable for small burns) in preparation for a skin graft in order to close it (Kennedy et al., 2010). It has been shown that MIC and minimum biofilm eradication concentration (MBEC) of clinically relevant antimicrobials can be 100 – 1000 times higher in biofilm-bound sessile cells than in planktonic cells (Alav et al., 2018; Mulcahy et al., 2014).

In pig models, inducing full-thickness burn wounds that were then maintained over an eight week period and infected with a combination of *P. aeruginosa* and *A. baumannii* showed that even once the wound appeared healed to the eye, further more detailed analysis demonstrated that the newly formed epidermis was now 'leaky', with water loss at the site of infection demonstrating a weakness in the skin barrier that could leave the subject vulnerable to further infection (Roy et al., 2014). It was also reported that despite debridement (removal of foreign matter and dead tissue until healthy bleeding tissue is exposed) of the wounds by a plastic surgeon to remove infected tissue, micro-colonies of both bacterial species were discovered in deeper tissue, and the population of *P. aeruginosa* had returned to pre-debridement levels with 48 hours. This may indicate that rather than being a method of infection

management, debridement may be a contributory factor to encouraging infection in deeper tissue each time debridement is performed (Roy et al., 2014).

Relevant to this thesis, organic acids have been demonstrated to cause a breakdown of biofilms (Nostro et al., 2014). Acetic acid at a pH of 5 is more effective at breaking down the biofilm of *S. aureus* and *S. epidermidis* than lactic acid or hydrochloric acid. The effectiveness of acetic acid on biofilm formation is pH dependent, with a marked decrease in effectiveness at pH 6 compared to pH 5, with each acid looked at independently and pH determined by the amount of each respective acid added to the media. The authors of this particular study are not specific about the concentrations of acid used, rather the pH of the media, simply stating they added 'adequate amounts' of acid to media in order to achieve desired pH. Also, the focus of the paper was in relation to gram-positive bacteria in a food safety context (Nostro et al., 2014). Some success has been achieved in reducing biofilm formation in Shiga-toxin producing *E. coli* with a 2% acetic acid solution. However, this is not in the context of wounds but the food industry (Park and Chen, 2015). This is an important area of future study, given the known role biofilm formation plays in antibiotic resistance and the presence of biofilms in chronic wounds.

8.1.2 Limitations of the data available on organic acid as a clinical treatment.

In summary, there are few studies into the effectiveness of acetic acid as a treatment for burn wound infections. The studies that have been carried out both *in vivo* and *in vitro* are promising, but all the clinical studies outlined above showed that effectiveness was far from 100%, with some infections persisting after weeks of treatment. Also, although relatively low concentrations of acetic acid were required to treat *P. aeruginosa* infections, higher concentrations were required to treat infections caused by *E. coli* and *A. baumannii* and gram-positive bacteria. This throws some doubt on the effectiveness of acetic acid as a sole treatment given the need to weigh relative effectiveness at any given concentration against patient discomfort. No research was uncovered in the course of preparing this review that looked

into whether any pathogenic bacteria are able to evolve resistance to the effects of acetic acid over an extended period. As noted above, given the current situation with an ever changing landscape as far as drug resistance is concerned, it seems vital that this question is addressed. There is a significant lack of data demonstrating the effectiveness of acetic acid as a sole treatment in the control of burn wound infections. The study by Phillips et al., (1968) combined the use of acetic acid with antibiotics. This is an important area of further investigation and, despite this study having been conducted almost 50 years ago, does not seem to have been investigated any further. There has also been a consistent field of research into the use of other topical agents, such as silver sulfadiazine and dressings such as polyurethane foam (Imran et al., 2016; Nagoba et al., 2010; Yabanoglu et al., 2013). However, it was not possible to locate any studies which looked at the synergistic effect of using multiple antimicrobial agents either with or without conventional antibiotics, at once. This seems to be a vitally important area of further research, given that patients who suffer extensive burn injuries are much more likely to die from sepsis than from the burn injury itself (Church et al., 2006).

There are notable issues also with experimental design. As noted above at section 8.2.2 when discussing the study of Nostro et al. (2014), the lack of specifics as to concentrations used, but rather using pH only, is a considerable oversight, pointing to a lack of understanding of the relationship between concentration, environmental pH and the ionic state of each acid tested. In order to fully understand the results of that study, having information as to concentration is vital, given different medias respond differently in terms of altered pH to the addition of acid, depending on constituents. Throughout the literature there is little, if any, acknowledgement of the relationship between concentration, acid and pH. This important aspect will be considered in more depth within this chapter.

In order to look at this in more detail, together with biofilm inhibition, *Pseudomonas aeruginosa* will be grown in different organic acids at different pH and in increasing organic acid concentration, within

buffered media to control pH upon addition of organic acid. The resulting growth and biofilm data will be analysed in order to determine whether there are differences between organic acids, pH, concentration of acid and the strain of bacteria. To do that, it must first be determined how best to analyse growth data.

8.1.3 Bacterial Growth and Modelling Growth Data

Bacteria are required to continuously respond to changing levels of carbon availability, oxygen and any other environment stressor that may impact on growth, such as pH and nutrient starvation. Bacterial growth is separated into three distinct phases: lag, log (or exponential) and stationary phase. During lag phase, bacteria adapt to the environment and prepare themselves for division by synthesis of DNA, RNA and enzymes that will be needed to grow. If the bacteria have been transferred from one culture or environment where there was cell damage, due to high temperature or the presence of toxins, then the bacteria will require time to repair cellular damage before reproduction can begin. The length of the lag phase is usually determined by how well the bacteria can adapt to the growth conditions, and if the conditions are unfavourable with, for example, low pH or high temperature, the bacteria may take longer to adjust to or repair damage. This is because compensations will need to be made for the additional energy required by the cell to survive (and reproduce) in the hostile environment, which may come at the expense of cell division initially (Madigan et al., 2009). Lag phase is then followed by log, or exponential, phase. This phase is distinguished by cell doubling, which is proportional to the present population, and provided there are no obstacles to cell growth, such as unexpected nutrient limitation, this cell doubling happens at a continuous rate. The rate of doubling will depend on the conditions, with more favourable conditions resulting in a shorter doubling time compared with less favourable conditions. The exponential phase of growth is defined as when the doubling rate is constant. However, there are few (if any) environments in which nutrients, and other favourable environmental factors such as space, are unlimited and therefore a doubling population will eventually reach maximum population density. This is the stationary phase. Before the culture reaches stationary phase however there will first be a slowing of

growth, which can no longer be defined as the population doubling at a constant rate, and this is when the culture will be entering stationary phase, but while there is still bacterial growth. During stationary phase, many cell functions such as the biosynthesis of essential compounds continues, as does cellular reproduction, albeit at a much reduced rate. The overall population density of the culture remains constant as the rate of reproduction is equal to that of cell death. In the culture where nutrients are completely depleted, or there has been a detrimental build-up of toxic metabolic by-products, cell death may begin to exceed cell reproduction, and the overall population density will start to decrease. Ultimately cells will no longer reproduce at all, though biosynthesis may continue, and the cells will be in death phase. This is also an exponential function, though is usually much slower than exponential growth (Madigan et al., 2009). Analysing the specific rates and changes that are occurring within the culture during growth, particularly under different conditions that are to be compared such as different organic acids, or the same organic acid at different concentrations, is a challenge that has been tackled via routes including mathematical modelling, with varying degrees of success.

The mathematical modelling of bacterial growth is an established but unperfected method of predicting the way in which a particular bacterial species will respond to a given environment. This is particularly important in the fields of food preservation (and shelf-life) and fermentation, and in other industries such as agriculture, where the ability to predict microbiological behaviour would be highly advantageous. In the context of burn wounds, it could allow the modelling of bacterial growth in wound environments and has the potential to predict the microbiological response to a chosen method of treatment, such as topical application of organic acids, without the need to first have the results of many microbiological experiments. The modelling of bacterial growth data also allows prediction in the sense that a fitted curve can estimate when a population will reach stationary phase by using the data provided to estimate a growth rate and likely end point, or carrying capacity, even if that information is not available.

The logistic growth model is based on the sigmoidal shape of a bacterial growth curve, which is usually distinguished by three distinct phases: lag; log (or exponential); and stationary phase. These phases are determined by the growth conditions, with very favourable conditions resulting in a very short lag time, a quick doubling time during log phase, with the culture reaching its population density, and therefore stationary phase, within a relatively short space of time. If conditions are less favourable, for example if there is a limit on key nutrients, then the population density (the number of individuals the available resources are able to support) may be lower. If conditions are less favourable in terms of an environmental factor such as slightly acidic pH or moderately higher or lower temperature, then the overall population density may be the same. However, lag may be longer and doubling time slower, resulting in a different shaped curve. The aim of parametric models, such as logistic, is to provide an estimation of the three parameters described above to fit a curve to the data and predict where that curve may end up. The accuracy of this curve is then measured by calculating mean squared error (Mytilinaios et al., 2012; Tonner et al., 2017).

This study makes use of three methods of modelling growth data: two parametric models, logistic and Baranyi and one Bayesian based model, Gaussian process regression. Logistic modelling uses the presumption that with food, space and no threat, a population grows at a rate that is proportional to the population i.e. in each unit of time, a certain number of individuals produce new individuals. However, population growth is not unconstrained in the long-term and this is taken into account as the 'carrying capacity' i.e. the assumed population density. Baranyi modelling is based on the logistic model. However, this model contains an adjustment function which accommodates an extended lag phase. In the Baranyi model, all extracellular environmental factors, such as chemicals produced during metabolism, are considered to be growth-dependent and are therefore included as a variable. This model is also based on the assumption that lag phase is inversely proportional to maximum growth rate. This allows the modelling of growth in challenging environmental conditions which themselves may also be changing, and

which will be having a continuous effect on the lag and growth of the bacteria within the culture (Baranyi and Roberts, 1994; Lipkin and Smith, 2004).

The third method used in this study is that described by Tonner et al., (2017). The authors describe their Gaussian process (GP) regression model which is a Bayesian nonparametric model, which can adapt to the complexity of the data. This allows predictions to be made based on learning from previous data, and may be able to take into account unseen variables based on previous information (Gershman and Blei, 2011). Tonner et al., (2017) used their GP regression model *Baysian Growth Rate Effect Analysis and Test* (B-Great) in order to analyse the growth data of *Halobacterium salinarum* and note that the B-Great model outperforms the parametric models against which is was tested, determined by calculating the mean squared error (MSE) between the test data and the model output. The authors showed that under 'ideal' conditions, the parametric models and B-Great were able to fit the data with low MSE. However, when modelling data generated under oxidative stress conditions, GP regression MSE was significantly lower than that produced by the more commonly used parametric models.

In the current study, the effects of different organic acids on growth and biofilm formation were investigated. Two strains of *Pseudomonas aeruginosa* were used: PA01, the classical lab strain, and PA1054, a clinical isolate from a burns patient at the Queen Elizabeth Hospital, Birmingham. Cultures were grown in acetic acid, propionic acid, butyric acid, sorbic acid, citric acid, lactic acid, DL-malic acid and benzoic acid. The pH of the medium was also altered, in order to investigate the impact of changing pH, and therefore the charged states of the acids, on the growth rates and biofilm formation of the two strains. The aim of this investigation is to identify whether acetic acid is a suitable inhibitor of *P. aeruginosa* growth, whether any other organic acids are more effective at inhibiting growth and biofilm formation, what the effects are of lowering pH in each case and whether there is a significant difference across strains. Logistic, Baranyi and Phenom (GP regression) mathematical models were utilised to fit and

analyse the data generated from these experiments. The data obtained from this study will be used to inform a clinical trial within Queen Elizabeth Hospital Birmingham into the use of organic acids as a topical treatment for burn wounds.

8.2 Experimental design and refinement

Initially, *Pseudomonas aeruginosa* PA01 was grown in 96 well flat bottomed clear plastic plates containing supplemented M9 medium, covered with an oxygen permeable film. The FLUOstar Omega automated plate reader, produced by BMG Labtech was used, set to a temperature of 37^{o}C, with continuous shaking at 200rpm. It was decided that this high-throughput design was the most appropriate for the amount of data to be collected as using shake flasks would be laborious and very time consuming.

Problems were encountered with this method. The data proved to be highly variable and inconsistent. Figure 8.1 below shows PA01 growth in the control condition of pH 7 with no added organic acid for six independent biological replicates. The data was collected on different days but in identical conditions. Beneath at Figure 8.2 shows growth of PA1054 in pH 7 supplemented M9 media with no added organic acid. Again, the data are highly variable with completely different growth profile from one day to the next. Plates were seeded with bacteria grown from different colonies, but from the same agar plate, taken from a single colony.

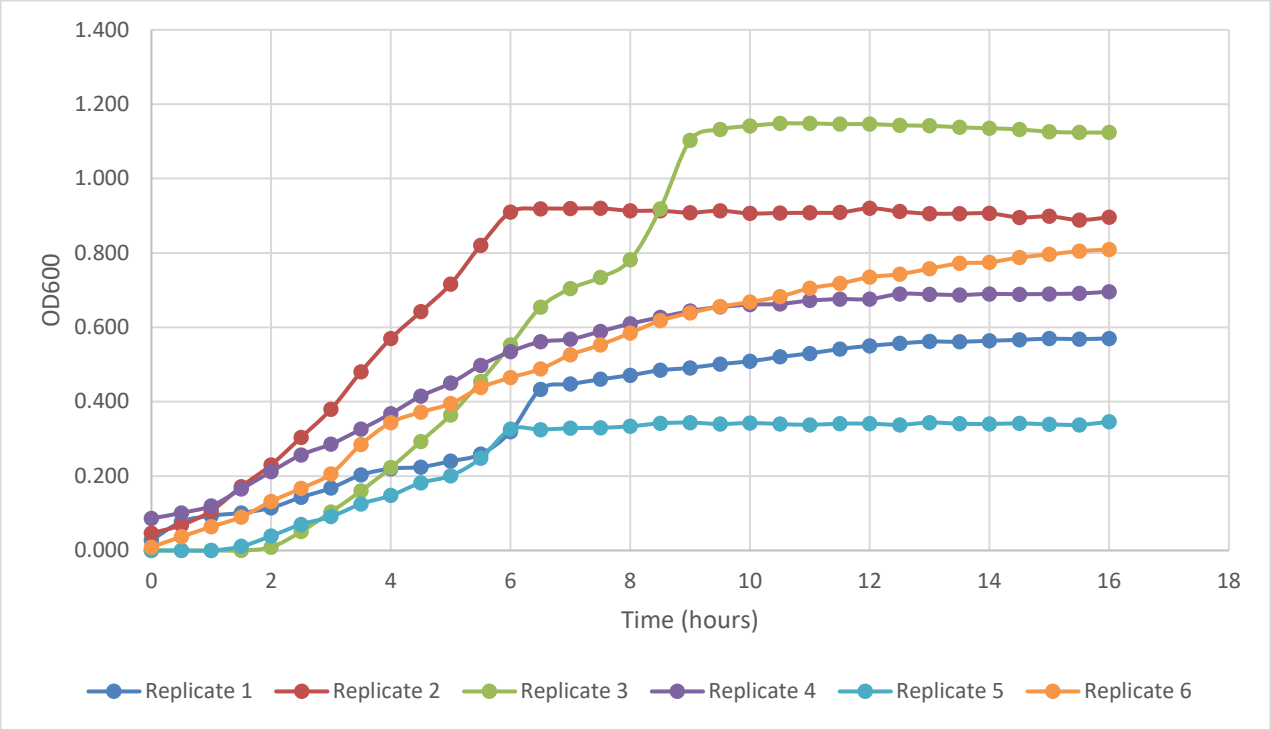


Figure 8.1 PA01 grown in supplemented M9 at pH 7 in a clear flat-bottomed microtitre plate covered with a gas-permeable film. Growth was measured in the FLUOstar Omega.

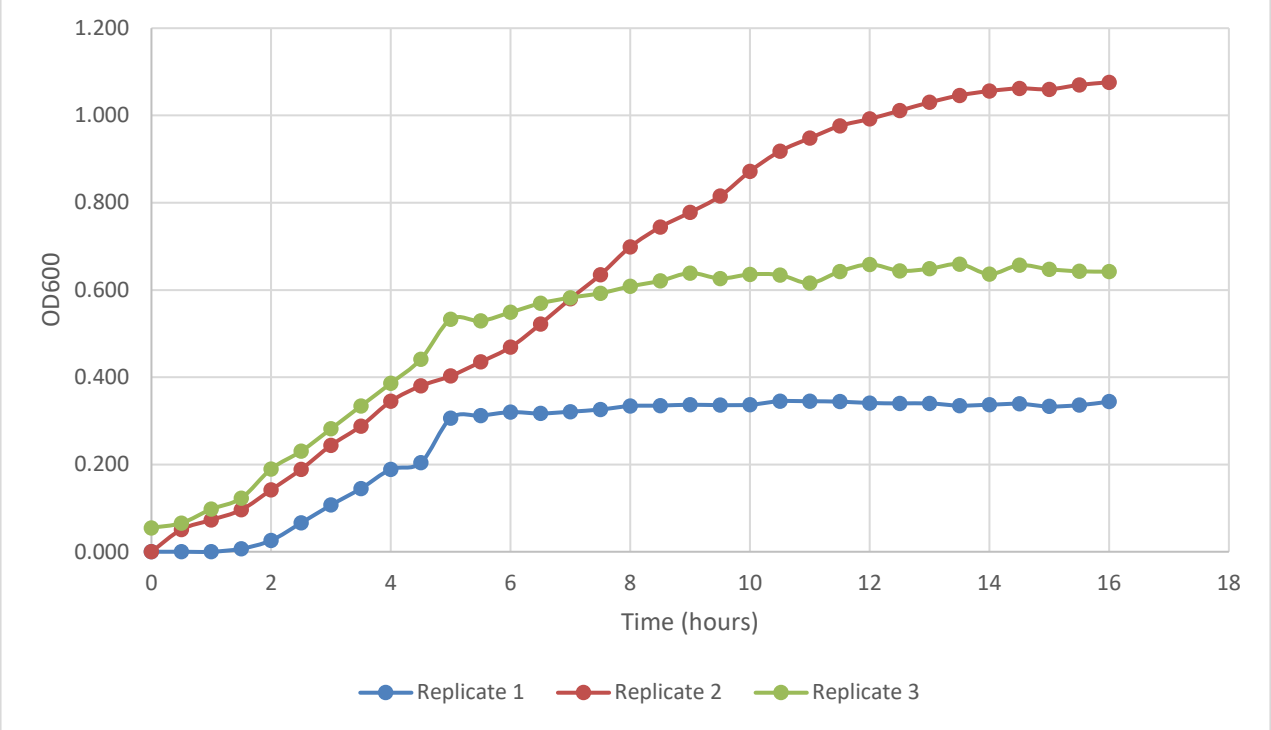


Figure 8.2 PA1054 grown in supplemented M9 at pH 6.5 in a clear flat-bottomed microtitre plate covered with a gas-permeable film. Growth was measured in the FLUOstar Omega.

A number of factors were considered as the possible source for the high variability, including inconsistency with the medium and supplements, a high rate of variation potentially inherent within the strains themselves, oxygen limitation due to the plastic film covering the plate and inconsistency in the set temperature within the growth chamber of the machine. The settings on the machine were also changed multiple times in order to determine whether there were optimal settings which reduced variability. In consultation with the company who produced the machine, BMG Labtech, the data was sent to their engineers in order for them to check settings, and they were unable to determine any limiting or unusual factors. Once investigation was completed into the issues listed above, the most limiting factor on reproducibility in the FLUOstar Omega machine appeared to be oxygen distribution to the growing culture. Figure 8.3 below shows PA01, in the same growth conditions as above at Figure 8.2, except with an evaporation resistant lid instead of a film covering the plate. This shows a clear improvement in both growth and reproducibility. However, variability is unlikely to be solely due to lower oxygen availability due to the use of a gas-permeable film rather than a lid, as this would simply reduce growth in all cases, rather than cause highly variable growth.

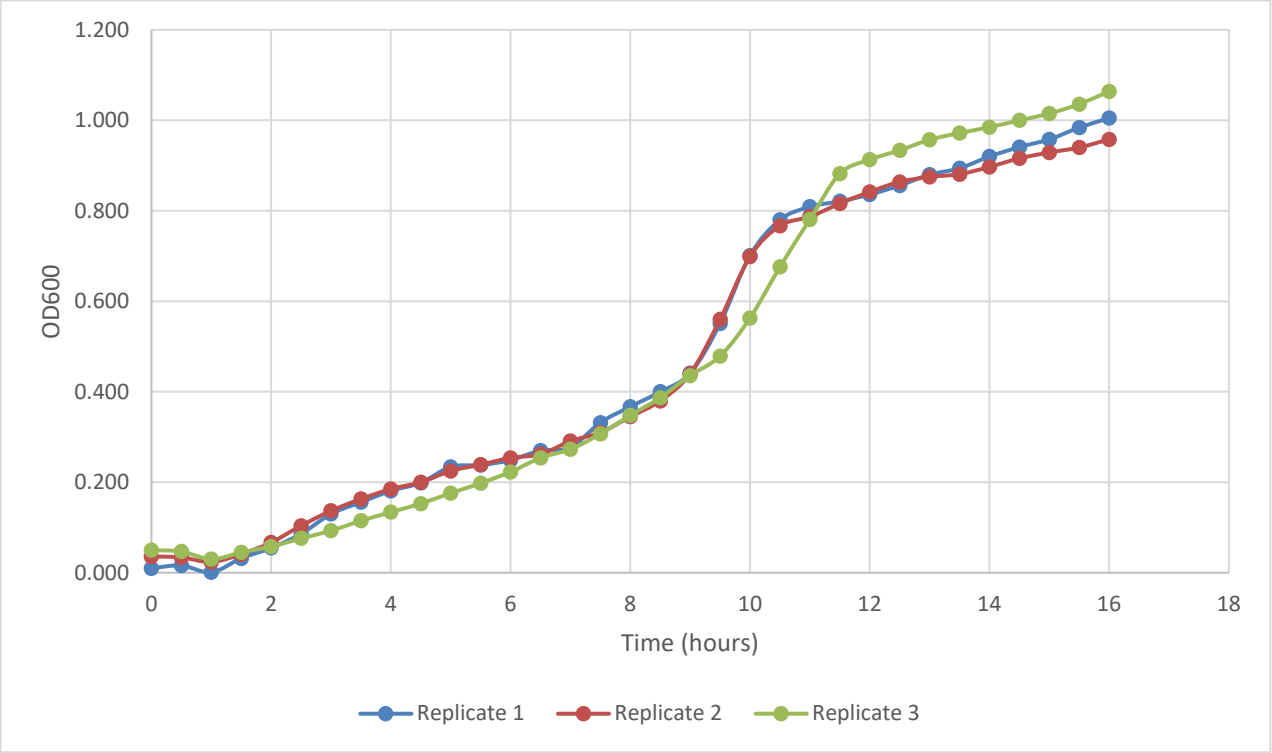


Figure 8.3 Growth of PA01 in supplemented M9 media at pH 7 with no added organic acid. Gas-permeable film has been replaced with evaporation resistant lid.

As mentioned above, we also considered whether there was a high level of variability within the strains themselves, with the possibility of random mutations occurring during the course of the growth experiment which led to increased or variable growth patterns. Though this explanation was considered highly unlikely, given the regular occurrence, to test it the cultures which showed much higher growth were grown again overnight and a plate seeded just with these cultures. However, the growth profiles returned to those of the other replicates, i.e. mostly depleted compared to using the lid instead of the film.

Despite the obvious difference in the overall growth and reproducibility between using the film and lid, a further difficulty arose in that the FLUOstar Omega scratched the top of the lid due to the internal space for the plate being too small to accommodate the lid, which itself distorted OD readings as the top of the plate was scratched and covered with a fine layer of plastic 'dust'.

Following these initial experiments, the experiments were therefore moved to using the CLARIOstar, an automated plate reader produced by the same company BMG Labtech, but with a larger internal compartment. Preliminary experiments demonstrated a growth bias and apparent edge effect on the plate as demonstrated in Figure 8.4 below. *E. coli* BW25113 was grown in LB media (produced by the School of Biosciences central services). Every well on the plate was seeded from the same overnight to a starting OD of 0.05 and allowed to grow for 24 hours in the CLARIOstar using the same settings as those for the subsequent experiments with *P. aeruginosa*. This was repeated twice and the heatmap below shows final OD (although the pattern represented below is apparent from earlier time points).

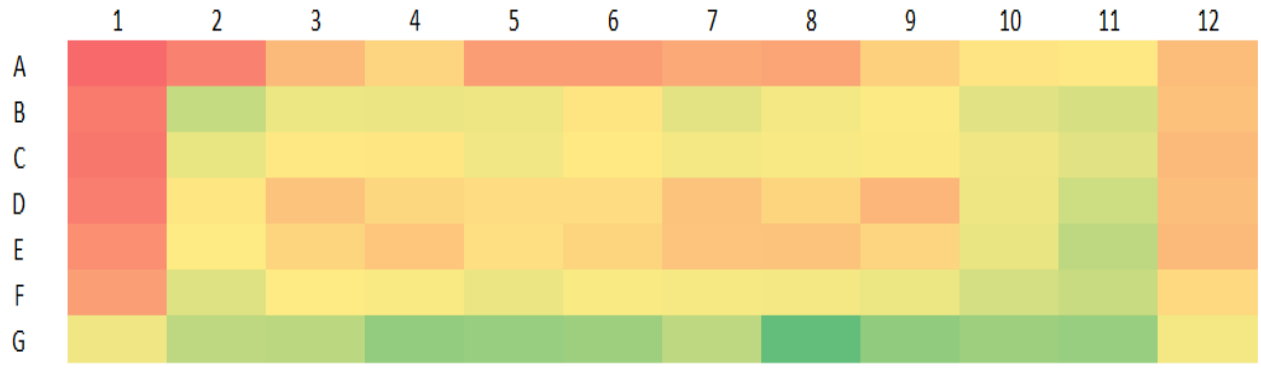


Figure 8.4 CLARIOstar BW25113 in LB edge effect heat map showing representation of final OD. The same culture was grown in every well and incubated for 16 hours. Row H (not shown) was used for blanks. Red indicates a lower final OD along a scale to dark green indicating the highest final OD.

As a consequence of this, the decision was made to use the external rows and columns as blanks only and to shift the experiments across the plate each day to reduce as far as possible any possible source of bias from the final data collection and analysis.

In order to test the reproducibility of the CLARIOstar machine, PA01 was grown in supplemented M9 media at 37^{o}C, with continuous shaking at 200 rpm and 10 seconds of vigorous shaking at 500 rpm before each read. A total of 60 wells were inoculated with PA01 from the same overnight to a starting OD of 0.05 and a reading taken every 10 minutes for 24 hours. Figure 8.5 below shows the mean plot for all 60 replicates (data shown with both a linear and logarithmic y axis). This graph also shows that PA01 growth was relatively slow in this experiment, demonstrated by the fact that stationary phase was not reached in 24 hours. This may have been due to oxygen limitation from use of a microtitre plate for growth. However, for practical reasons a decision was made for all subsequent growth experiments to be done using the same conditions and over the same time period. Data in this chapter will be presented using a linear y axis. This is because in the growth conditions used, there were often small difference in growth which are difficult to see on a logarithmic y axis, for example where growth is very slow but increasing.

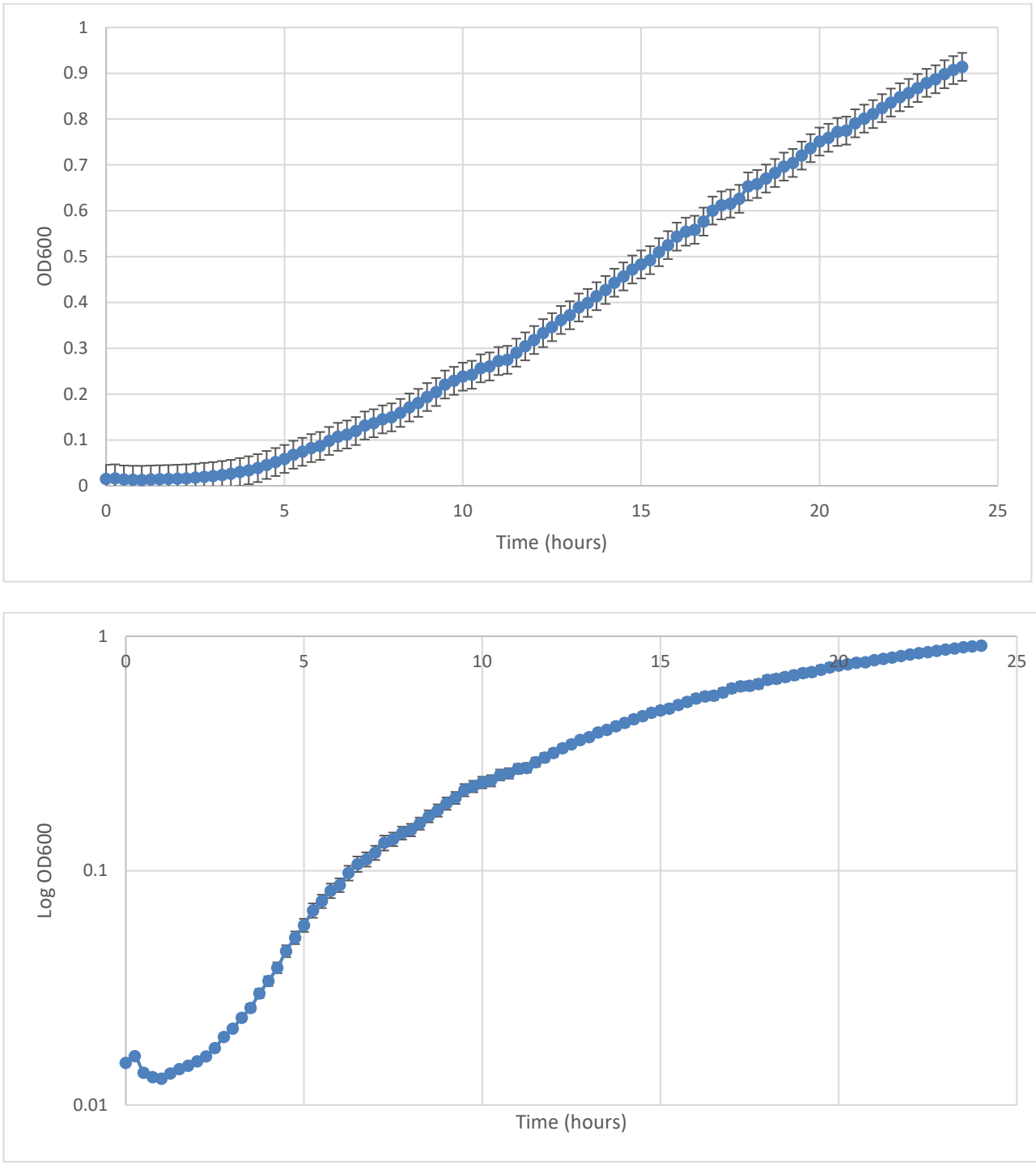


Figure 8.5 PA01 grown in CLARIOstar in supplemented M9 media, showing 60 technical replicates. Plate covered using an evaporation resistant lid. Error bars show standard error of the mean (SEM). Top: data shown with linear y axis. Below: data shown with logarithmic y axis.

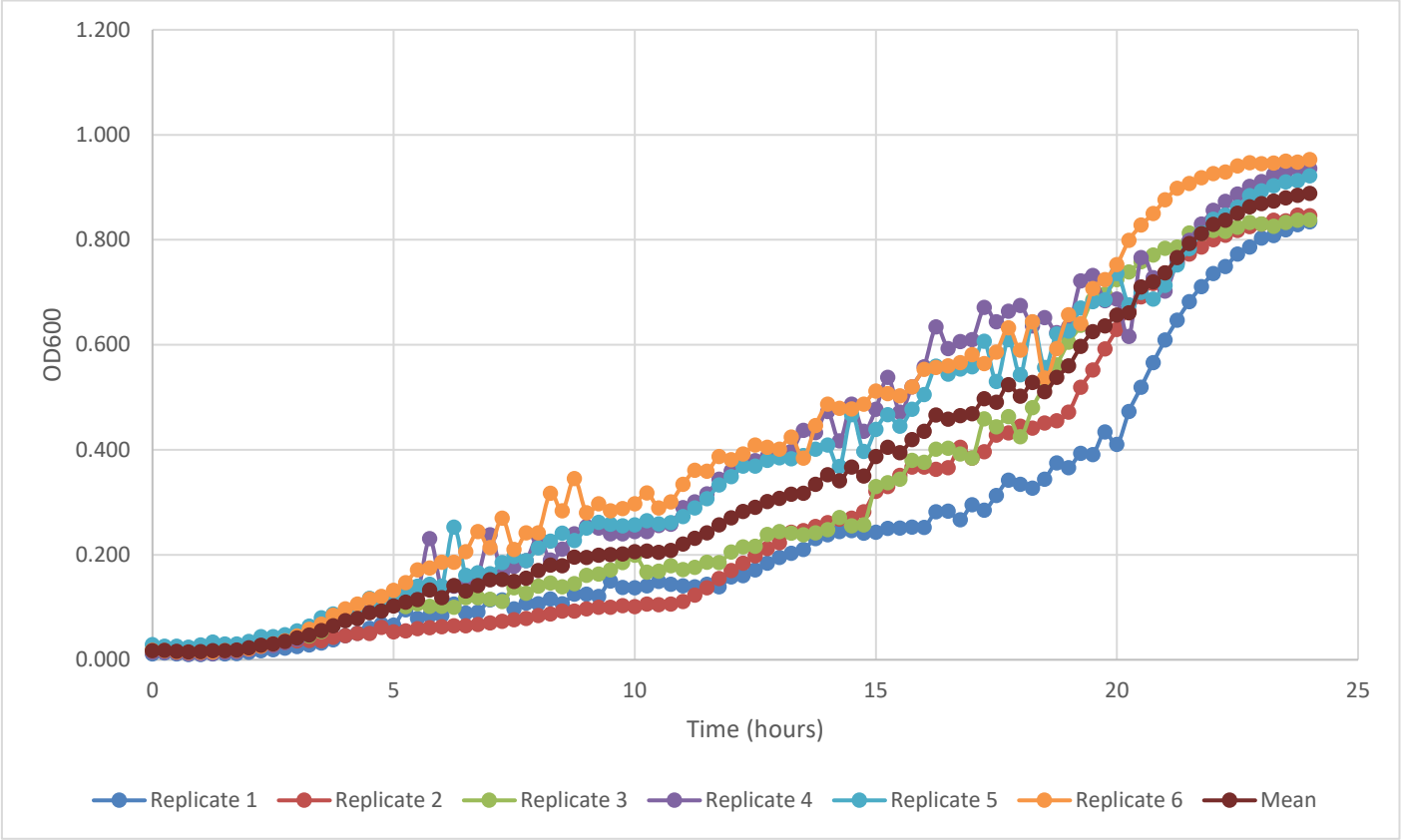


Figure 8.6 PA01 grown at pH 7 with 5 mM acetic acid**.** The first three data sets, replicates 1 – 3, falling towards the bottom of the six lines, were grown on a different day (a separate biological triplicate experiment) to the top three data lines, replicates 4 – 6.

One further issue that was encountered during data collection was day to day batch variation. The raw data above shown in Figure 8.6 clearly show the effect of batch differences from one day to the next. The six replicates have separated into two distinct groups with the mean represented by the brown line between the two. This is not something that could be overcome with refinement of the protocol. However, it was a source of variability to which we were alive and have considered via potentially normalising the data, discussed below. Ultimately, though day to day variation was apparent, by using six replicates for analysis instead of three and increasing the number of readings from every 30 minutes to every 15 minutes, enough data was generated to make reliable observations between data sets despite the inherent variability within that seemed apparent with this method. This will be discussed in more detail in Section 8.2.1 below.

Plates used for growth experiments were also stained for biofilm quantification. In order to determine the linear range of the CLARIOstar in this assay, a serial dilution of crystal violet was plated and absorbance measured at OD550nm. Figure 8.7 below shows that the machine reached maximum absorbance at approximately OD 3.4. Therefore, any biofilm quantification reading which registered as OD 2.5 or higher was diluted 10-fold to ensure the reading was well within the range of the machine.

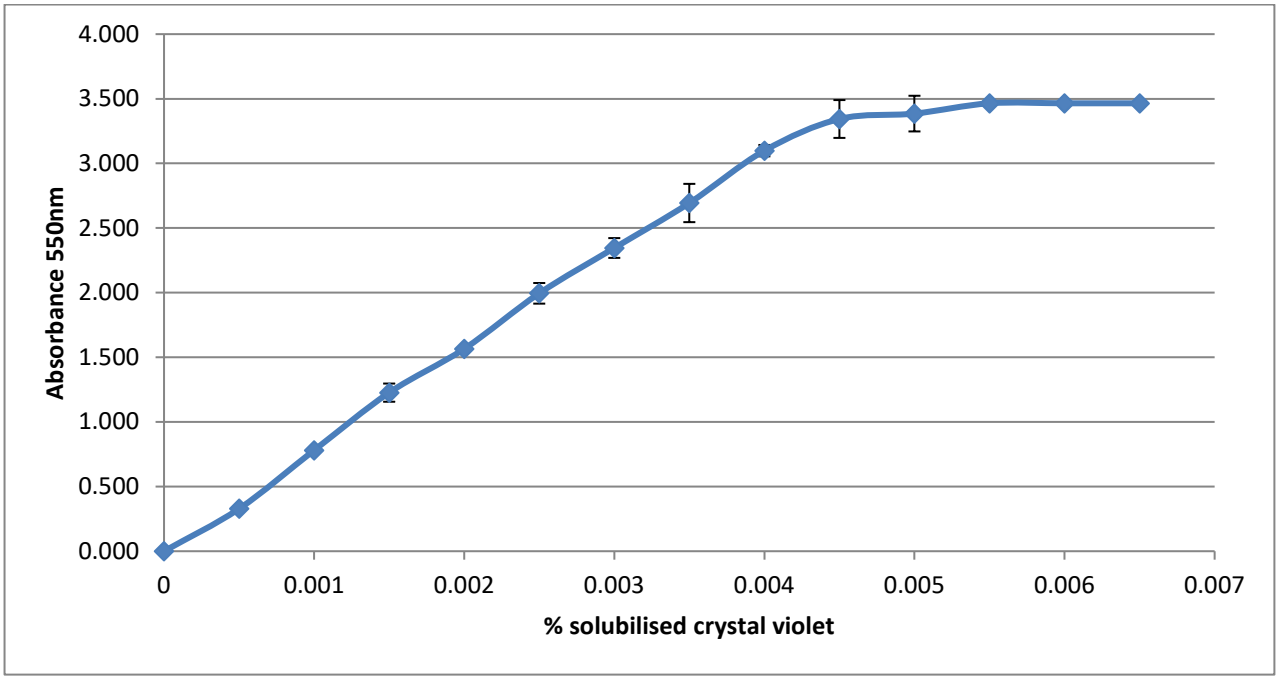


Figure 8.7 Determination of linear range of CLARIOstar using dilutions of crystal violet. Error bars show standard deviations. Dilutions were measured in triplicate.

8.2.1 Refined experimental data

We measured growth kinetics of two strains of *Pseudomonas aeruginosa* in eight different organic acids, at three concentrations across a range of 5 different pHs. A total of 1920 independent growth curves were generated (960 per *Pseudomonas* strain) and an overall total of 184,320 data points used in the analysis. The following conditions were tested with six replicates per condition:

	PA01				PA1054			
Acetic Acid	0 mM	5 mM	10 mM	20 mM	0 mM	5 mM	10 mM	20 mM
Butyric Acid	0 mM	5 mM	10 mM	20 mM	0 mM	5 mM	10 mM	20 mM
Propionic Acid	0 mM	5 mM	10 mM	20 mM	0 mM	5 mM	10 mM	20 mM
Citric Acid	0 mM	5 mM	10 mM	20 mM	0 mM	5 mM	10 mM	20 mM
Lactic Acid	0 mM	5 mM	10 mM	20 mM	0 mM	5 mM	10 mM	20 mM
Malic Acid	0 mM	5 mM	10 mM	20 mM	0 mM	5 mM	10 mM	20 mM
Sorbic Acid	0 mM	5 mM	10 mM	20 mM	0 mM	5 mM	10 mM	20 mM
Benzoic Acid	0 mM	5 mM	10 mM	20 mM	0 mM	5 mM	10 mM	20 mM

In order to analyse properly the large amount of data collected, three different curve fitting methods were used and a comparison of the results will be presented at section 8.3. However, it is possible to do a more general analysis from visual inspection of a selection of the raw data.

First, as expected, as the pH of the media decreased, the addition of a fixed concentration of organic acid had a more pronounced effect. This is demonstrated in figures 8.8, which shows PA01 at pH 7, pH 6.5, pH 6, pH 5.5 and pH 5 with the addition of 10 mM acetic acid. Second, increasing the concentration of organic acid at a fixed pH also had an impact on overall growth, demonstrated at figure 8.9 which shows growth of PA01 at pH 6 with the addition of 5 mM, 10 mM and 20 mM butyric acid. Third, there was an organic

acid dependent response, with some of the organic acids having a far greater impact on overall growth than others. This is demonstrated at figure 8.10, which shows PA01 growth data at pH 6 with 10 mM of each organic acid tested. Fourth, there was a strain dependent response. PA1054, the clinical isolate, was more tolerant to the addition of organic acids at all pH tested than the lab strain PA01. This is shown at figure 8.11 which shows PA1054 at pH 6 with 10 mM of each organic acid.

In some cases, the addition of organic acid, specifically citric acid and lactic acid, appeared to cause an increase in growth as concentrations increased. Figure 8.12 below demonstrates the apparent improved growth of PA01 in citric acid at pH 6 as concentration of citric acid increases. The microtitre plate was inspected visually following the experiment to ensure that the increased OD readings had not been caused by obvious cell clumping and there was no indication of this.

Given the very large amount of data which has been generated, together with the different interactions as outlined above, it is hard to make meaningful comparisons simply by looking at the growth data. Therefore, different mathematical models were used to extract growth parameters so that cross-condition comparisons could be made more easily using heat maps and, in the case of PHENOM, relative interactions (which will be discussed in detail below).

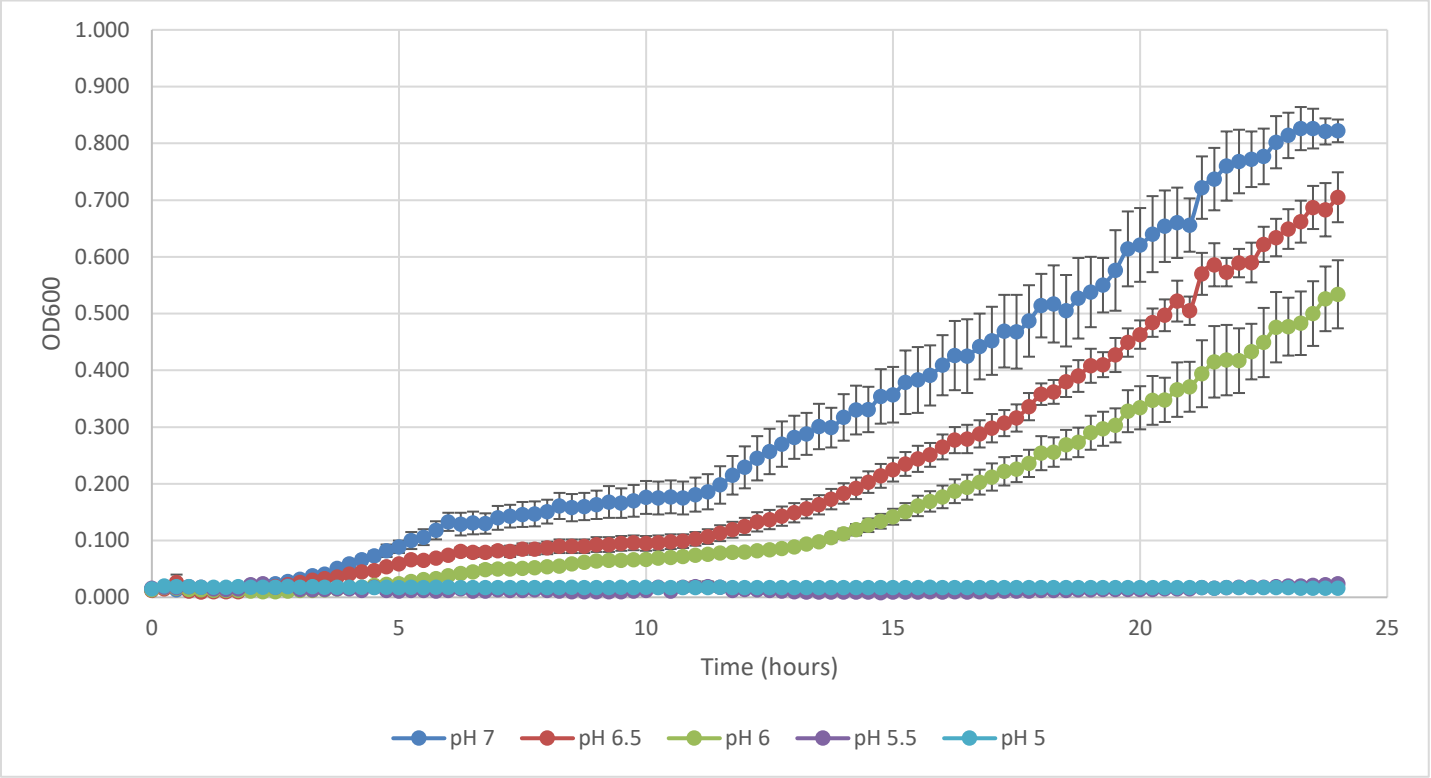


Figure 8.8 PA01 grown in supplemented M9 media across a range of 5 different pH with 10 mM added acetic acid. Error bars show SEM. N=6.

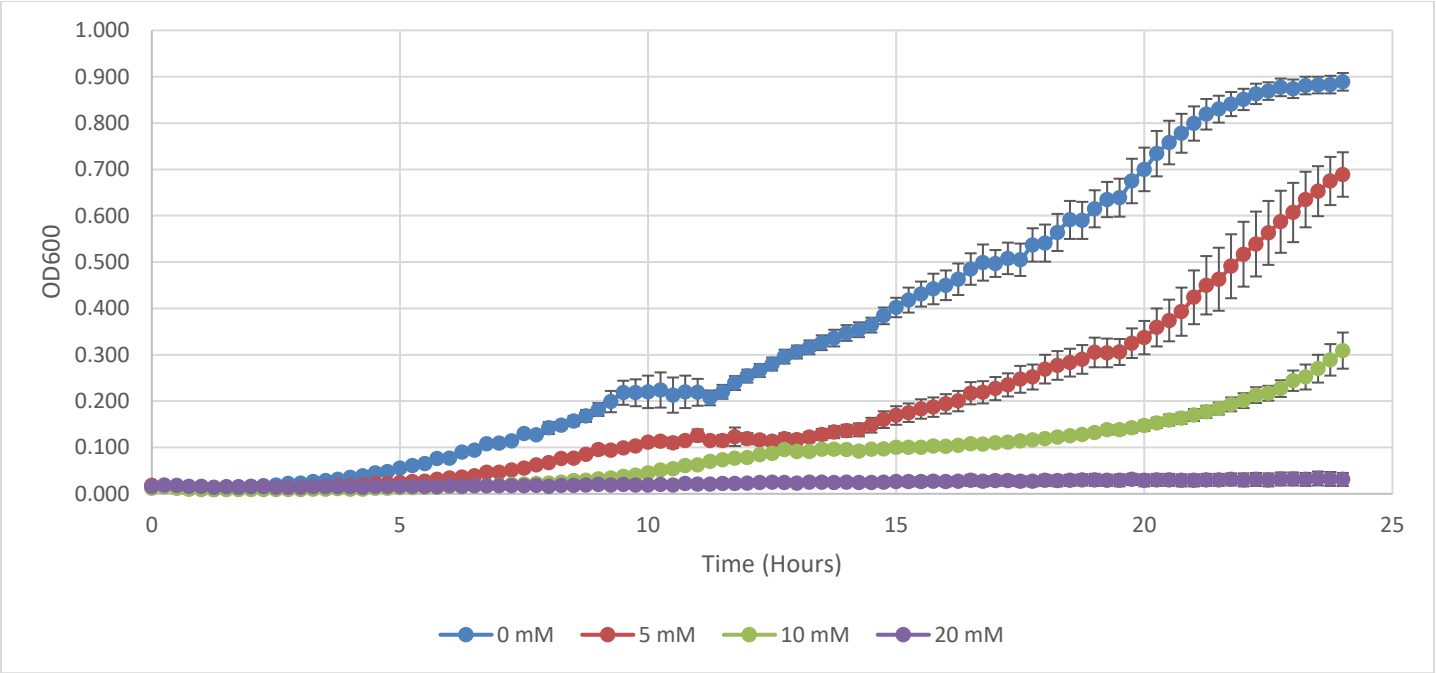


Figure 8.9 PA01 grown in supplemented M9 media at pH 6 with the addition of no organic acid, 5 mM butyric acid, 10 mM butyric acid and 20 mM butyric acid. Error bars show SEM. N=6

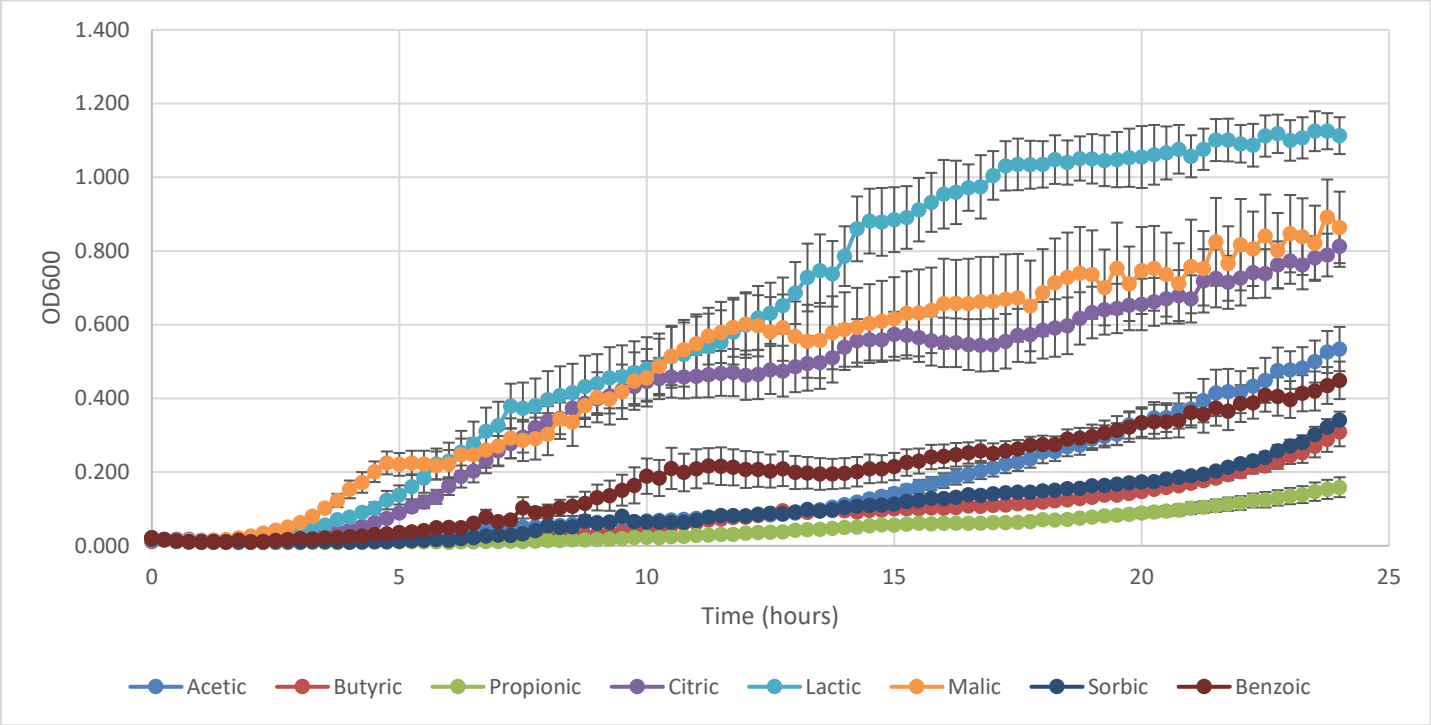


Figure 8.10 Growth of PA01 at pH 6 with 10 mM added organic acid. Error bars show SEM. N = 6.

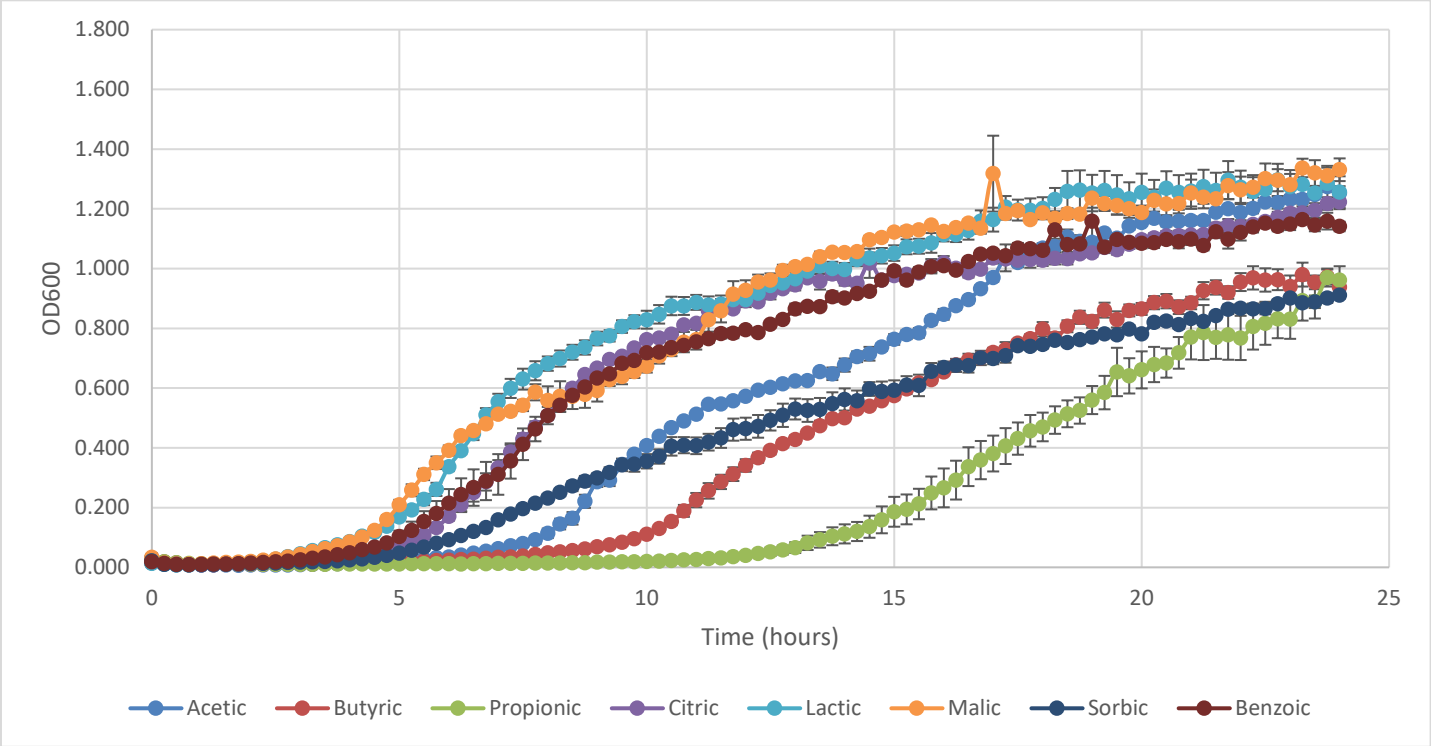


Figure 8.11 Growth of PA1054 at pH 6 with 10 mM added organic acid. Error bars show standard error of the mean (SEM). N = 6.

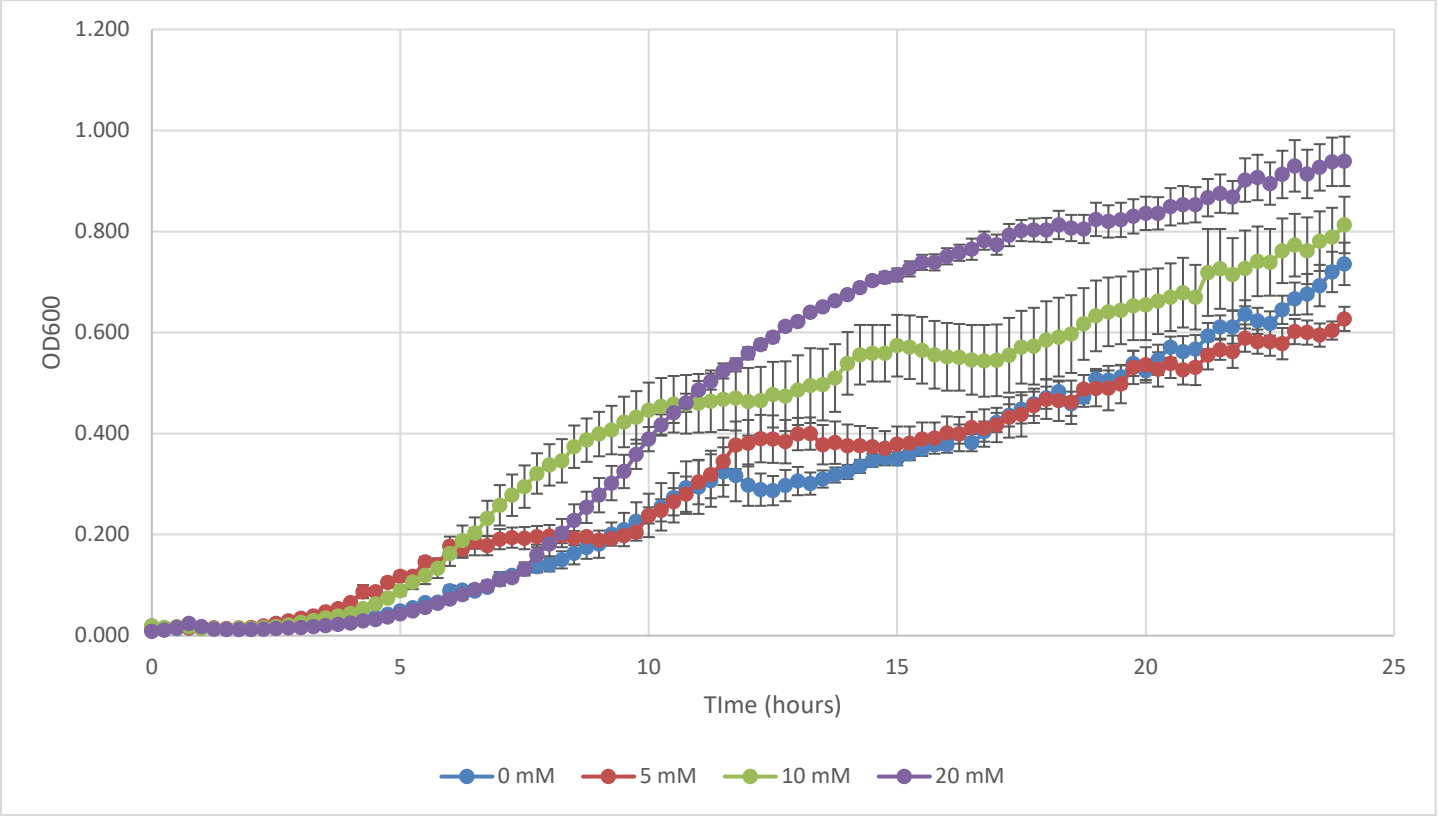


Figure 8.12 Growth of PA01 at pH 6 with added citric acid. Error bars show standard error of the mean (SEM). N=6

Data was assessed daily as it was generated to ensure consistency by comparing the control conditions of no added organic acid, which was included on every plate at each pH, as it was expected that these conditions should be highly reproducible day to day. There was some variation, as would be expected, however this was significantly smaller than compared to that previously noted with the FLUOstar Omega and earlier growth conditions. Consistency of the data being produced was tested by comparing the standard deviation between mean final OD of each condition, including the control conditions. The assumption was made that the standard deviation between all conditions with no added acetic acid would be low from one set of data to the next (on the basis that this condition is identical in all experiments). Conversely, it was expected that there would be greater variation between acids, given their differing impact on growth. For example, it was predicted that there would be greater variation between pH 7 with 10 mM acetic acid and pH 7 with 10 mM malic acid than between pH 7 with no added organic acid, which was the control condition in both cases. If the data was not consistent, then high levels of variation would be seen in the control conditions from day to day. This data collected from the control conditions proved to be reproducible, with the standard deviation of the mean final OD between control experiments lower than that between test conditions. This was true for all controls and test conditions except for pH 5 with 20 mM added organic acid in both PA01 and PA1054, which showed low standard deviation between the means of each acid because in every case there was little or no growth recorded whatsoever. Figure 8.13 below shows ranking from lowest to highest of standard deviation of the mean final OD from each experiment for both PA01 and PA1054.

	Acetic Acid	Propionic Acid	Butyric Acid	Sorbic Acid	Citric Acid	Lactic Acid	Benzoic Acid	Malic Acid	Standard Deviation
pH 6	0.877	0.804	0.889	0.884	0.692	0.879	0.892	0.879	0.070
pH 5 + 20 mM	0.022	0.017	0.017	0.029	0.064	0.066	0.301	0.019	0.097
pH 6.5	0.879	0.831	0.916	0.785	0.698	0.933	0.980	0.967	0.097
pH 5	1.217	1.208	1.248	1.296	0.995	1.309	1.285	1.315	0.105
pH 7	0.861	0.839	0.936	0.761	0.714	0.992	0.976	0.992	0.108
pH 5.5	0.962	0.663	0.729	0.850	0.454	0.796	0.568	0.690	0.160
pH 6.5 + 5 mM	0.831	0.621	0.888	0.664	0.549	1.024	0.736	0.979	0.172
pH 6 + 5 mM	0.679	0.473	0.689	0.557	0.546	1.003	0.715	0.942	0.188
pH 7 + 5 mM	0.888	0.519	0.858	0.552	0.726	1.063	0.804	1.028	0.199
pH 7 + 10 mM	0.822	0.379	0.643	0.516	0.593	1.099	0.741	1.098	0.261
pH 5 + 10 mM	0.016	0.014	0.018	0.027	0.069	0.082	0.772	0.008	0.263
pH 6.5 + 10 mM	0.705	0.273	0.514	0.567	0.471	1.230	0.691	1.049	0.314
pH 5.5 + 20 mM	0.031	0.012	0.017	0.083	0.017	0.029	0.570	0.992	0.365
pH 5.5 + 5 mM	0.326	0.132	0.191	0.326	0.866	1.182	0.371	0.707	0.368
pH 6 + 10 mM	0.534	0.159	0.309	0.341	0.77	1.17	0.558	1.165	0.382
pH 7 + 20 mM	0.671	0.22	0.525	0.433	0.58	1.406	0.489	1.224	0.408
pH 5.5 + 10 mM	0.024	0.006	0.041	0.343	0.730	1.110	0.649	0.698	0.409
pH 6.5 + 20 mM	0.536	0.266	0.303	0.401	0.825	1.415	0.438	1.113	0.417
pH 6 + 20 mM	0.029	0.038	0.031	0.229	0.869	1.285	0.388	1.208	0.534
pH 5 + 5 mM	0.058	0.010	0.016	0.479	0.010	1.057	1.172	1.079	0.535

	Acetic Acid	Propionic Acid	Butyric Acid	Sorbic Acid	Citric Acid	Lactic Acid	Benzoic Acid	Malic Acid	Standard Deviation
pH 5 + 20 mM	0.021	0.011	0.025	0.026	0.103	0.022	0.040	0.110	0.039
pH 5	1.094	1.089	1.082	1.019	0.957	1.095	1.039	1.152	0.059
pH 7 + 5 mM	1.259	1.262	1.242	1.323	1.234	1.362	1.126	1.311	0.071
pH 5.5	1.133	1.125	1.124	1.075	1.160	1.179	1.319	1.210	0.074
pH 6.5	1.249	1.320	1.363	1.123	1.213	1.169	1.237	1.247	0.077
pH 6	1.423	1.199	1.331	1.321	1.279	1.345	1.331	1.455	0.079
pH 7	1.254	1.241	1.206	1.088	1.139	1.348	1.163	1.333	0.091
pH 6.5 + 10 mM	1.491	1.258	1.380	1.191	1.354	1.315	1.192	1.347	0.101
pH 7 + 10 mM	1.333	1.296	1.311	1.258	1.076	1.404	1.192	1.380	0.106
pH 6 + 5 mM	1.390	1.183	1.244	1.185	1.264	1.376	1.135	1.439	0.112
pH 6.5 + 5 mM	1.487	1.381	1.414	1.207	1.421	1.326	1.131	1.180	0.130
pH 7 + 20 mM	1.345	1.291	1.164	1.133	1.175	1.399	1.138	1.565	0.154
pH 6.5 + 20 mM	1.328	0.956	1.063	1.107	1.139	1.229	1.122	1.464	0.160
pH 6 + 10 mM	1.253	0.962	0.938	0.911	1.223	1.256	1.141	1.331	0.166
pH 5 + 10 mM	0.016	0.006	0.019	0.020	0.042	0.020	0.558	0.113	0.188
pH 5.5 + 20 mM	0.014	0.003	0.016	0.500	0.037	0.037	0.787	0.129	0.293
pH 5 + 5 mM	0.149	0.002	0.013	0.048	0.063	0.018	0.940	0.356	0.322
pH 5.5 + 5 mM	1.102	0.391	0.298	0.822	1.131	1.269	1.014	1.170	0.367
pH 5.5 + 10 mM	0.029	0.002	0.008	0.689	0.754	0.018	0.882	1.084	0.462
pH 6 + 20 mM	0.328	0.001	0.011	0.740	1.205	1.006	1.004	1.335	0.527

Figure 8.13 Means of final OD for all test conditions, ranked by standard deviation between the means.

Top: PA01. Bottom: PA1054

In order to minimise the chance of drawing misleading conclusions based on high variability in the control conditions, this analysis was used to determine whether to normalise the growth data to the pH 7 control condition. However, given the high level of agreement between the controls from each experiment, the decision was made not to normalise the data for downstream analysis as this would potentially lead to a loss of meaningful comparison, particularly between strains.

8.3 Curve Fitting using Logistic, Baranyi and Phenom – a comparison

8.3.1 Comparison of parametric curve fitting versus Gaussian curve fitting.

Where growth was logistic, the parametric models were able to fit a curve to the data with low relative error. For each fit generated, the code calculated absolute error (the difference between the actual data value and the fit, using the square root of the square of the distance between the data and the curve summed over all data points) and the relative error (the absolute error divided by the distance between the data and the curve). These error estimates provide a means by which to determine how accurate the curve fit has been. If the error is very high, then the curve fit has been poorly estimated. By contrast, if the error has been calculated as being very low, it can be confidently said that the fit is accurate to the data. The relative error will be used in this analysis as it provides a value for error which can be more accurately compared across conditions as relative error gives the error relative to the size of what is being measured, therefore for data with very small values (for example where OD readings are very low) the relative error value can be meaningfully compared to that of the error where the data values are much higher. The code then estimated whether or not a curve fit was 'bad' by determining whether the relative error was more than 10% that of the error calculated for pH 7, the control condition which should, in theory, have a good curve fit with low error. The code also generated a growth rate, which is the maximum growth rate during logarithmic phase, and carrying capacity – the value at which the fitted curve plateaus and therefore no longer increases in value on the y axis.

Despite the shortcomings of the logistic models, even where data was not sigmoidal in shape, if there was a lag and log phase, the models were generally able to fit a curve, though not always with an acceptable level of error, as determined by the criteria as laid out above. However, the models were not able to produce a value in all cases for carrying capacity, as growth was not logistic and so the equation could not accurately estimate parameter fits. Some of the values generated by the parametric models also appear to be overestimated, with values for growth rate of cultures where there was very little, or no growth, being estimated at many times higher than for control conditions. This is most likely as a result of very low OD readings which doubled very quickly in the initial stages which was likely an artefact of the CLARIOstar machine providing inaccurate readings when OD is very low and growth is very poor (for example from OD 0.005 to OD 0.01 to OD 0.02) and therefore not reaching a point where the spectrophotometer readings are more accurate. This resulted in the models estimating a very high growth rate and subsequently enormous carrying capacity after 24 hours. For those sets of data for which the models were unable to fit a curve at all, consequently they were also unable to generate values for growth rate and carrying capacity. Figure 8.14 below shows curve fits generated by the logistic model. In this instance the model fit appears to be accurate. This is confirmed by the low relative error, depicted in the scale below.

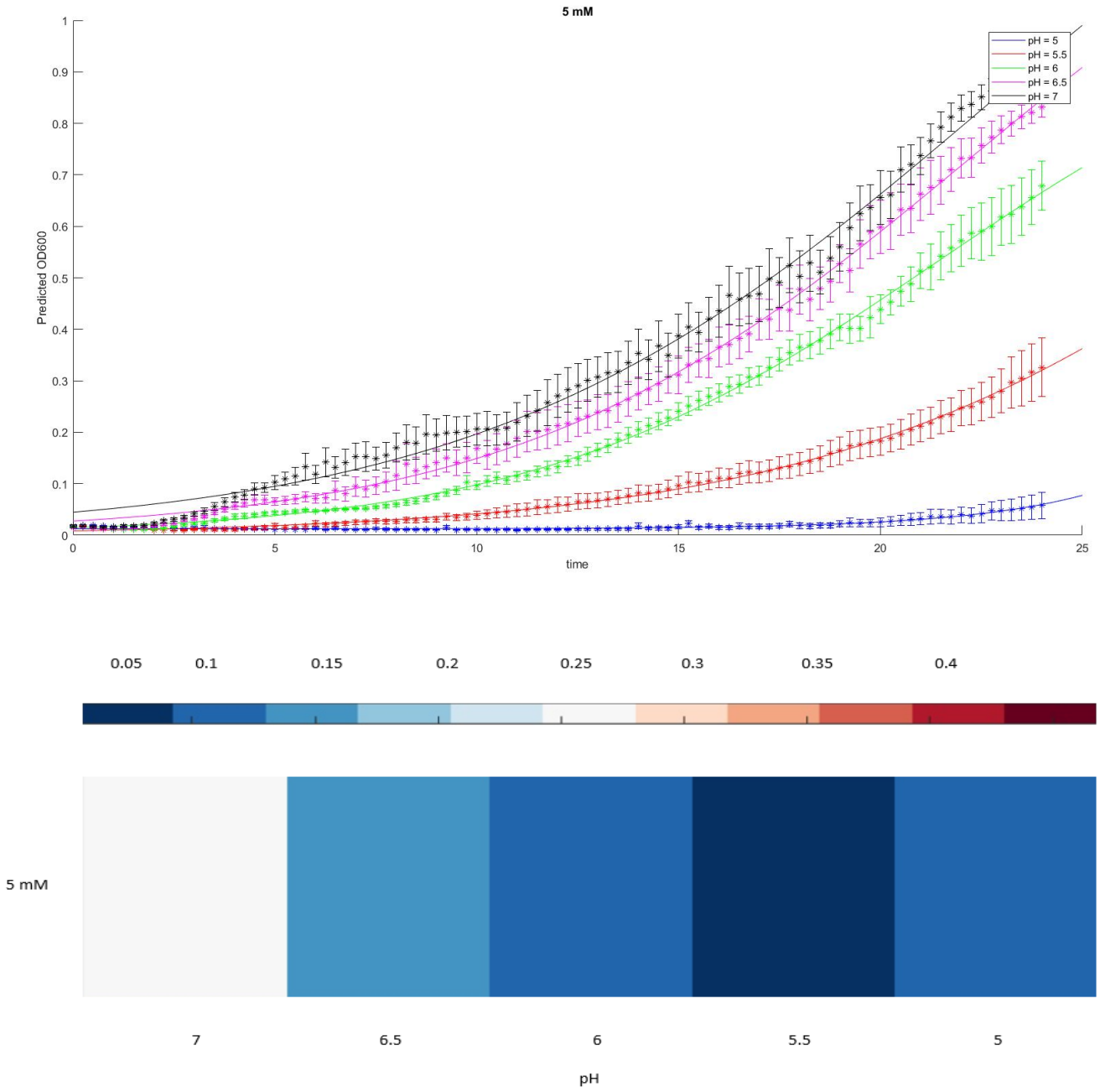


Figure 8.14 PA01 grown in 5 mM acetic acid. Curve fitting generated by Matlab using the logistic growth model. Points show raw data, lines show curve fit. Error bars are SEM. N = 6. Below: the relative error calculated by the model, as described above.

In contrast, due to the way the model works, the Phenom model successfully fitted curves to all data generated. A total of six replicates for each condition were generated for this model as the more data inputted the better the model is able to minimise error and analyse variation. The day to day difference (which has been previously discussed) was factored in as a variable in its own right within the Phenom model. Figure 8.15 below shows the curve fittings generated by Phenom to the same conditions as above (namely 5 mM acetic acid across the pH range). The Phenom model uses all of the data collected in order to establish the curve fits resulting in a less uniform curve, whereas the Baranyi and logistic models use only the mean of the data collected. A direct result of this approach is that the Phenom model is able to capture reproducible but unexpected changes in the growth of the bacteria which is lost when taking the mean of the data alone. The curves are therefore potentially more informative in that it is possible to see areas of the growth pattern which may be showing diauxic shift, for example, that would be lost in the generation of a simple logistic growth curve, which would not correctly model the raw data.

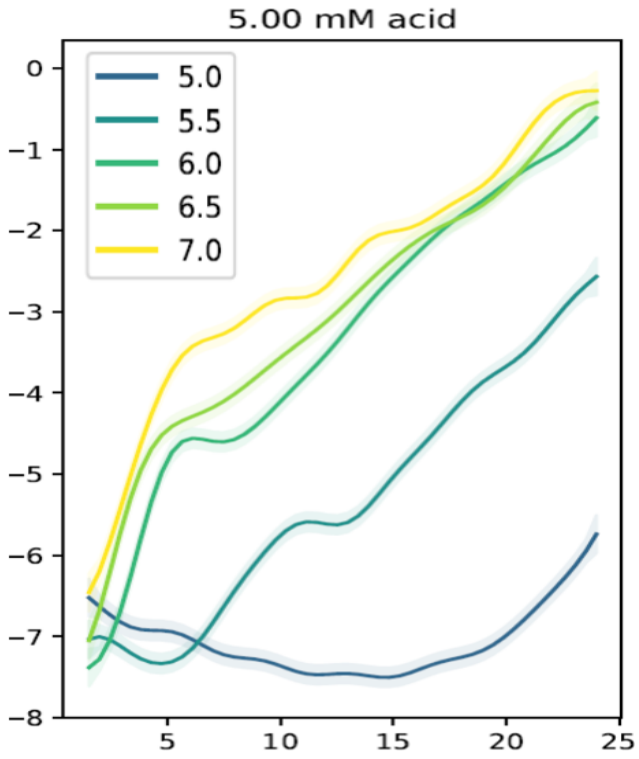


Figure 8.15 The same conditions as above, PA01 grown in 5 mM acetic acid across a range of 5 different pH. This curve was generated using the Phenom Gaussian process model. The x axis shows time in hours. The y axis represents the normalised growth function (as explained in Chapter Two).

As stated above, for many of the data sets, the logistic and Baranyi models were either unable to fit curves without generating high levels of relative error (deemed to be a relative error 10% higher than that calculated for pH 7 with no added organic acid) or rejected the data entirely (and subsequently generated no values or curves) and therefore the data could not be accurately fitted to a logistic growth curve, with or without a lag phase term. Figure 8.16 below shows the baranyi curve fits for 10 mM butyric acid across the pH range has resulted in a curve fit for only two sets of data. However, beneath that at Figure 8.17 are the corresponding curve fits generated by the Phenom model. This shows that the Phenom model is able to fit curves to data even when it is not falling into logistic growth, because this model does not use a logistic growth model in order to fit curves.

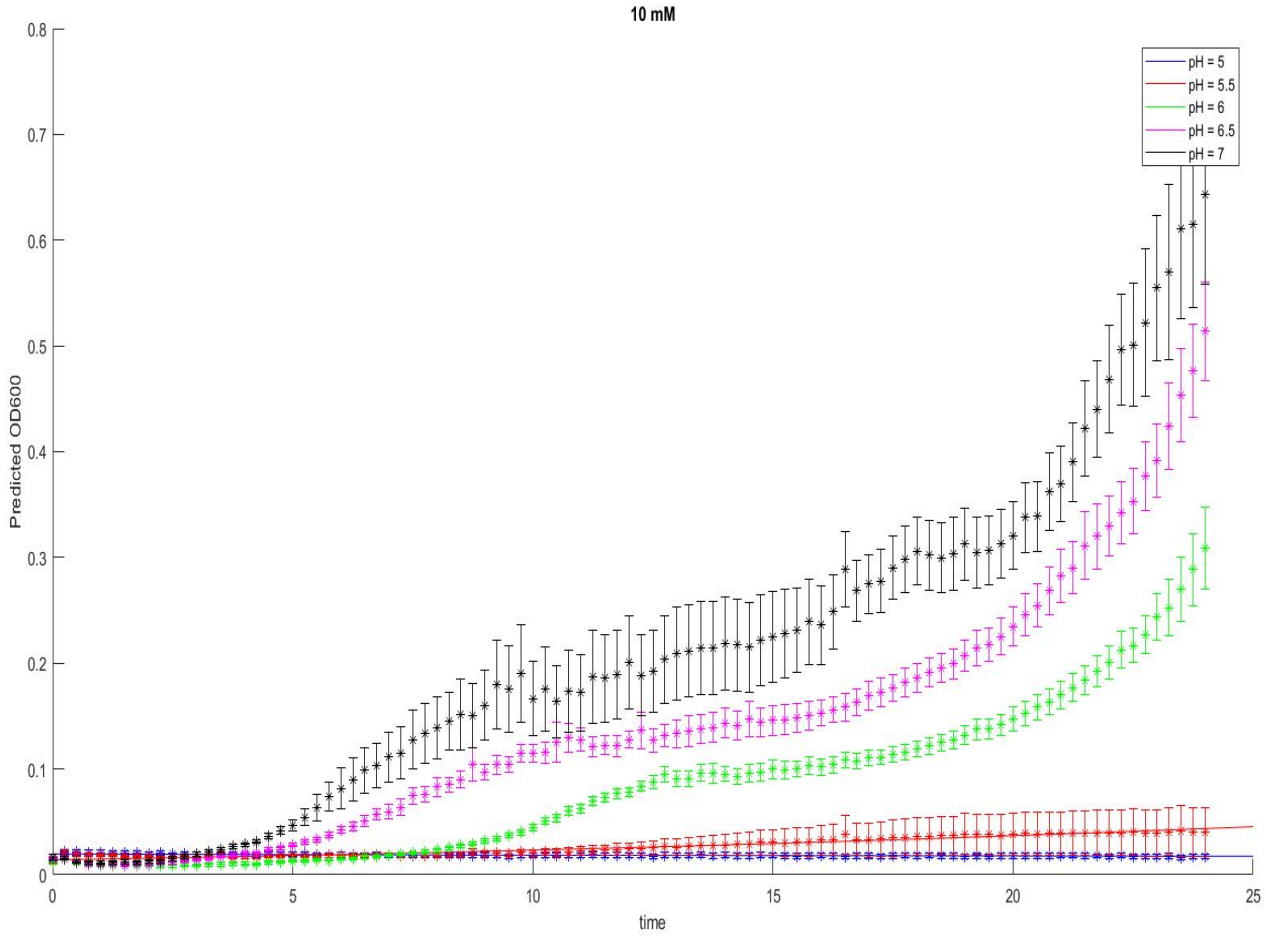


Figure 8.16 PA01 grown in 10 mM butyric acid across the pH range. Neither the Logistic or Baranyi model, based on the logistic equation, were able to fit curves to the top three datasets.

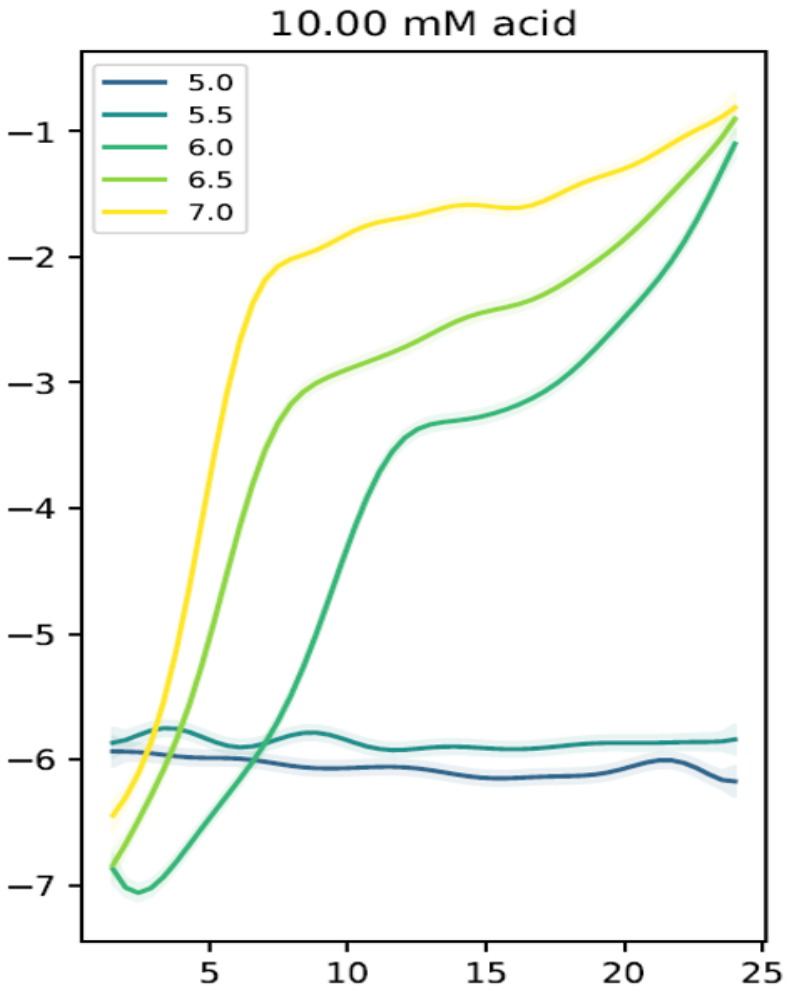


Figure 8.17 The same data set, PA01 grown in 10 mM butyric acid across a pH range. Curve fittings generated by the Phenom model. Y-axis shows normalised growth function. X-axis shows time in hours.

Below in Figure 8.18 the curve fits follow the general growth pattern of the data. However, it is possible to see from the fit against the raw data that there are fluctuations in growth which are not fully represented in the curve fits. Therefore, important and subtle changes in the way the bacteria are growing are lost in the smooth logistic curves that have been generated by the model. The model has been unable to fit a curve to pH 5 (blue line). In contrast the Phenom model offers fits which are a much more accurate representation of the raw data (Figure 8.19) and as a result may allow more insight into how the bacteria are behaving.

However, between the two parametric models, Baranyi was more successful than the logistic model at fitting curves to the data, due to the adapted function in Baranyi which allows for an extended lag phase.

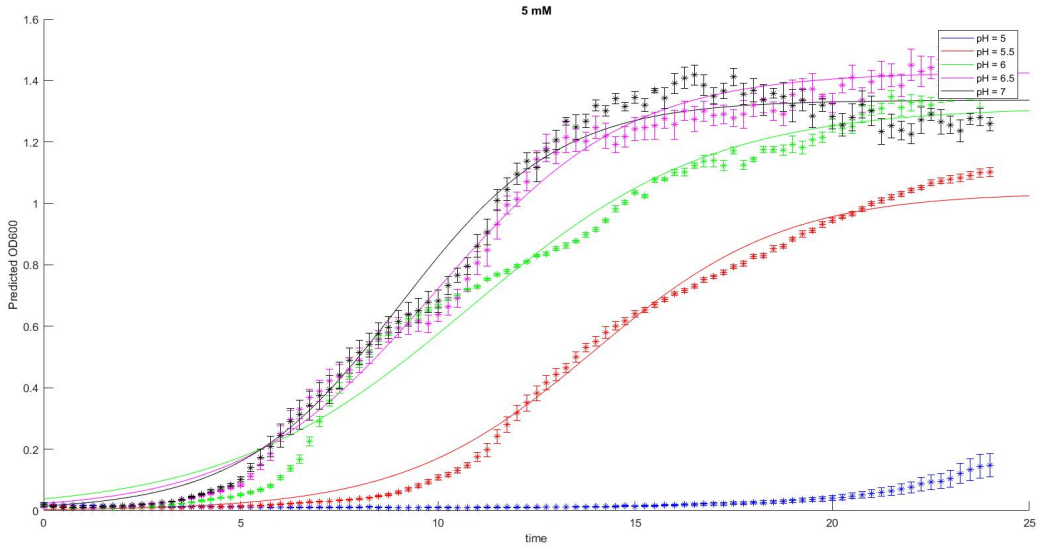


Figure 8.18 PA1054 grown in 5 mM acetic acid curve fits generated by the Baranyi model, generated curves plotted over raw data. Error bars represent SEM. N = 6.

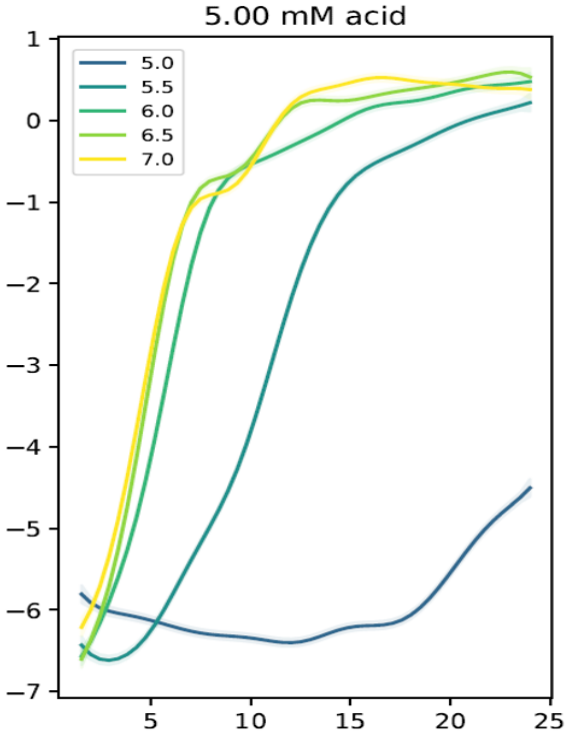


Figure 8.19 The same data fit by the Phenom model: PA1054 grown in 5 mM acetic acid.

There were incidences when bacterial growth deviated from parametric growth, but growth was still good (compared to the control condition of pH 7 with no added organic acid) and a high final OD was reached. In these instances, the Baranyi and logistic models were able to generate curves, but with high relative error. This is because the models were 'forcing' a logistic fit where growth was non-logistic. The Baranyi fits shown below at Figure 8.20 clearly do not fit well to the actual data and a great deal of information is lost if the fits are taken alone. The growth here shows a high level of fluctuation, which was repeatedly seen in general for PA01. The third graph in this figure shows growth of PA1054 in the same conditions. PA1054 growth showed much less fluctuation in general. The curve fittings here are of a better quality.

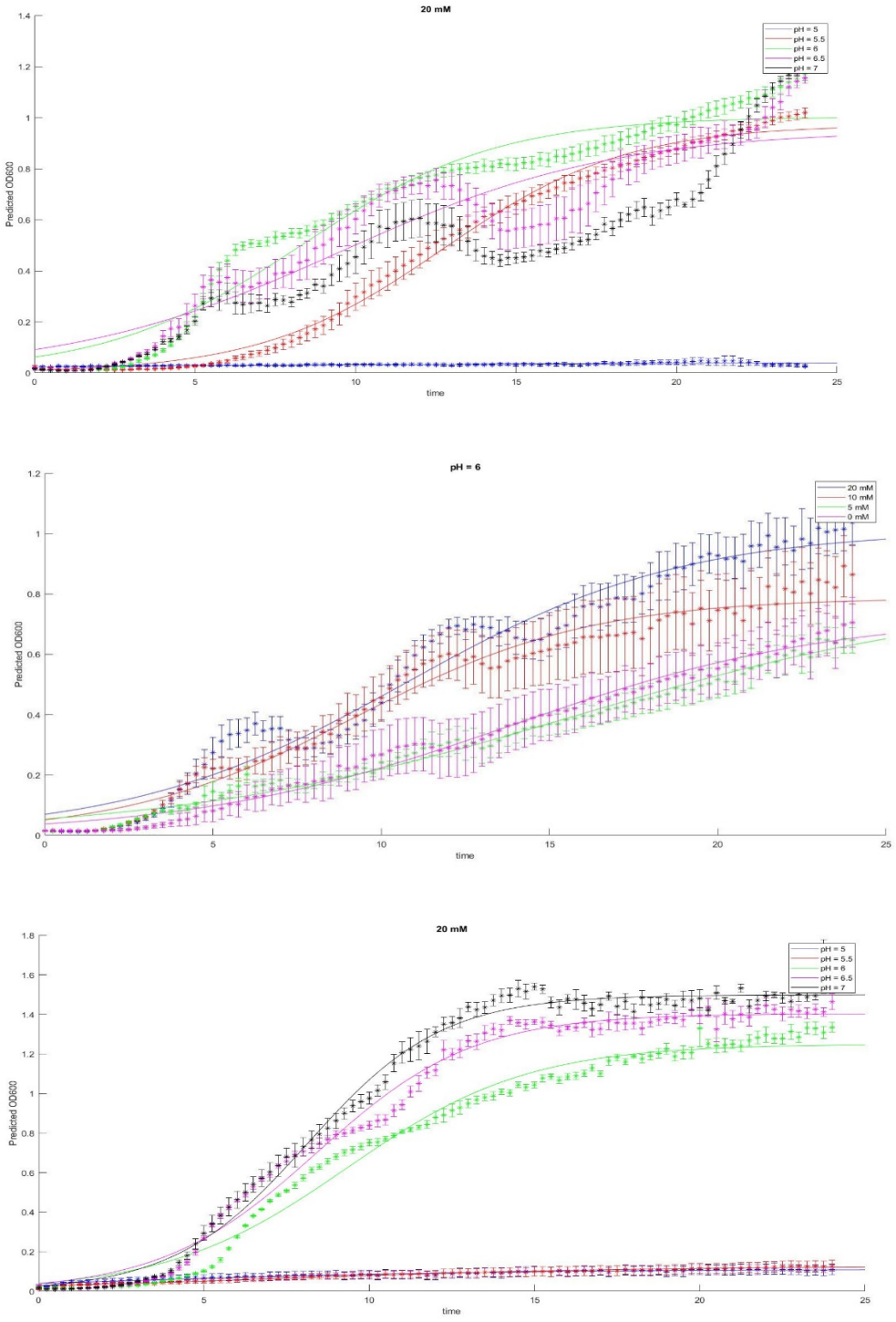


Figure 8.20 Top: PA01 grown in 20 mM Malic acid. Curves generated by Baranyi model. Error bars represent SEM. Middle: Baranyi generated curves and raw data for PA01 grown at pH 6 in 0, 5, 10 and 20 mM malic acid. Error bars represent SEM. Bottom: PA1054 growth in 20 mM malic acid. Curves generated by Baranyi model. Error bars represent SEM.

What is also visible here is the impact that 20 mM malic acid has on PA1054 at pH 5.5. This was not replicated in PA01, which does not show the severe growth impact in this condition as seen in PA1054. At Figure 8.21 below is the 'relative interaction' curve generated by the Phenom model. This is a method of displaying the data to compare differences between growth of each strain in a given condition, relative to the pH control (i.e. pH 6 with 10 mM organic acid relative to pH 6 with no added organic acid). So, any deviation from the pH control is deemed to be a result of the organic acid and concentration of that acid at any given pH. This allows the comparison of pH 7 plus 20 mM organic acid against pH 5.5 with 20 mM organic acid, which will have a different impact on the cell due to its charged state, but removes the growth impact caused by drop in pH itself. A good example of this is the difference between the two strains at pH 5.5 in 20 mM malic acid. The impact at pH 5.5 to growth of PA1054 is abrupt, which goes from steady growth at pH 6 to a flat-line (i.e. no growth) at pH 5.5. This method of comparison allows data to be displayed in a way that draws attention easily to obvious differences between strains. This will be discussed in more detail below when examining the comparisons more closely between strains, acids and pH. However, it does not provide a cross-acid comparison. Also shown are pH 6.5, 6 and 5. From the relative interaction plot, it appears that PA01 growth is improved at pH 6 with 20 mM malic acid. From Figure 8.20 (above) showing the Baranyi curve fits and raw data for this condition (top), from the raw data and curve (green line) it appears that growth is indeed higher overall than at pH 6 with no added acid (blue line compared to pink line, bottom).

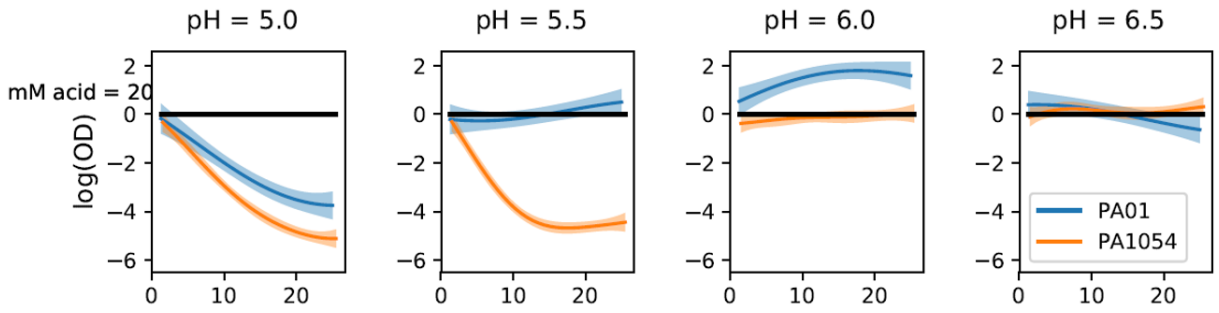


Figure 8.21 The relative interaction between pH and 20 mM malic acid compared to 0 mM controls at the same pH (black line) generated by Phenom. PA01 is indicated by the blue line, PA1054 is indicated by the orange line.

The most effective acids at inhibiting growth were propionic acid and butyric acid. Both acids completely inhibited growth in both the lab strain and the clinical isolate by pH 6 plus 20 mM of each respective acid. Both strains however still showed growth in this condition in the presence of acetic acid at the same concentration and pH. In all but one acid, benzoic acid, pH 5 plus 10 mM acid was sufficient to completely inhibit growth in both strains.

The next section will look closely at the estimates of carrying capacity and growth rates estimated by each of the models as an alternative way of analysing the data.

8.3.2 Comparison of growth rate & carrying capacity as generated by the three models.

Both parametric models, Baranyi and logistic, give an estimate for growth rate and for carrying capacity. Phenom is also able to estimate these parameters, although a clear distinction between these three models in that the parametric models attempt to estimate a final carrying capacity based on the fit of the curve once it has reached stationary phase (even if that point is never reached), whereas Phenom

provides a value for the final point reached in the fitted curve. Given that the bacteria often do not reach stationary phase in the 24 hour period of monitoring, the parametric models often struggle to accurately estimate this parameter, giving values which appear to be incorrectly estimated, or values for which is it not possible from the data to determine whether it is accurate or not. For example, figure 8.22 below shows the Baranyi curve fits for growth of PA01 in all pH with 5 mM added acetic acid. The curves are a good fit to the data. However, the accompanying carrying capacity estimate for pH 5.5 (the red line) is 1.255 (based on where the model has predicted the growth will settle once it has reached stationary phase). This is likely due to the fact that growth has appeared to increase rapidly towards the end of the 24 hour period, with a sharp upward trend in the curve. As growth never reaches stationary phase due to monitoring only lasting for 24 hours, it is impossible to know from this data whether that prediction is accurate. That would require further investigation. However, as the Phenom model estimates carrying capacity based on the final point in the curve fit (and does not attempt to estimate future 'unknown' growth), the estimate for carrying capacity is far lower, 0.315, and more representative of final OD reached in the growth period of 24 hours (OD 0.326 mean value across six replicates). This adds an element of difficulty in determining which model is better in terms of predicting unknown growth. Phenom, therefore, allows more accurate analysis of known (rather than predicted) parameters as the curve fits are truer to the data.

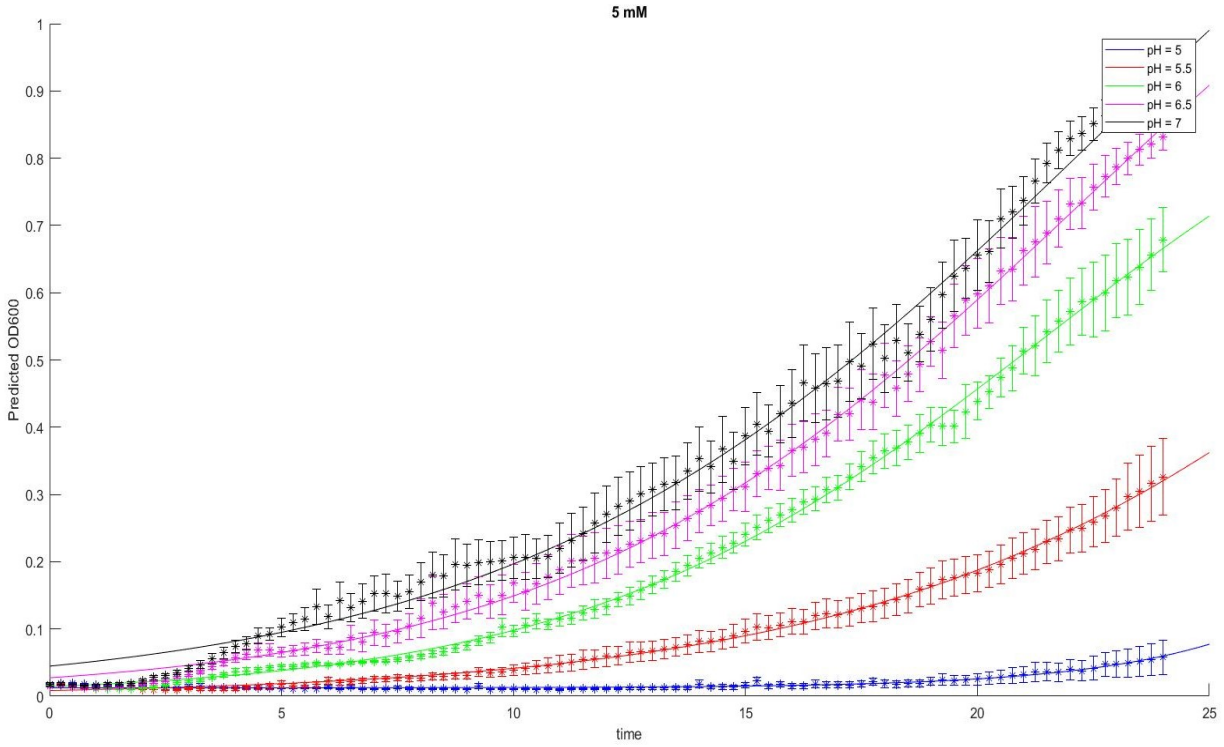


Figure 8.22 PA01 growth across 5 different pH with 5 mM added acetic acid. Baranyi curve fits plotted over raw data. Raw data shows average from six replicates. Error bars represented SEM.

The same problem was found in relation to growth rates. Exponential growth was not seen in many of the test conditions. The parametric models were therefore unable in many cases to estimate values for growth rate. The Phenom model was able to provide an estimate in all cases. In order to determine how accurately growth rates were predicted in all cases by each model, the carrying capacities were compared between the logistic model and Phenom with the hypothesis that the carrying capacities would follow a similar trend to one another if the fits to the data were accurate in both cases. The comparison was not done with Baranyi as the removal of the first five time points (the first one hour and 15 minutes) from the Phenom analysis means the data is not comparable between the two models as it removes the analysis of extended lag phase for which Baranyi has been specifically designed. On that basis it was determined that the output of the Phenom model was only comparable to the logistic model and only once the same initial five time points were removed. Below at Figure 8.23 are four pages of heat maps demonstrating carrying capacity for all PA01 and PA1054 conditions generated by Phenom and the logistic model. Where there is a blank space indicates where the logistic model was unable to generate a value for this parameter. This is an alternative way of showing the growth data in a more manageable, easy to view way, which more concisely demonstrates the difference across organic acid, *P. aeruginosa* strain and pH.

	Acetic PA01		Butyric PA01		Propionic PA01		Citric PA01	
	Phenom	Logistic	Phenom	Logistic	Phenom	Logistic	Phenom	Logistic
pH 7	5.890	5.974	5.795	5.951	5.738	5.508	5.665	6.059
pH 7 w 5 mM	5.575	5.741	6.511	6.341	5.210	5.149	5.441	5.056
pH 7 w 10 mM	5.807	5.711	5.290	6.045	5.052	4.893	5.371	4.971
pH 7 w 20 mM	5.433	5.513	5.534	5.673	4.582	4.363	5.229	5.438
pH 6.5	5.906	6.014	5.724	5.777	5.896	5.728	5.508	5.927
pH 6.5 w 5 mM	5.710	6.505	6.404	6.089	5.288	5.080	5.208	4.776
pH 6.5 w 10 mM	5.863	6.664	5.322	5.170	4.789	4.717	5.449	4.925
pH 6.5 w 20 mM	5.945	6.780	5.156	6.039	3.237	5.114	5.966	6.005
pH 6	5.774	5.711	5.721	5.862	5.584	5.605	5.743	5.515
pH 6 w 5 mM	5.735	6.127	5.869	6.766	4.983	5.903	5.223	4.947
pH 6 w 10 mM	5.658		4.982	7.233	3.555		5.987	5.717
pH 6 w 20 mM	0.232	-2.392	0.750	3.644	0.102	1.432	6.094	6.523
pH 5.5	5.924	6.052	5.364	5.447	5.362	5.559	5.478	5.644
pH 5.5 w 5 mM	4.755	7.117	4.410	5.924	3.081		6.464	6.423
pH 5.5 w 10 mM	0.514	-0.674	0.663	0.187	0.090	-2.477	3.603	
pH 5.5 w 20 mM	0.269	-0.617	0.060	-0.296	0.078	-0.856	0.399	0.937
pH 5	6.398	7.188	6.164	7.721	6.787	7.728	6.361	7.257
pH 5 w 5 mM	1.183	-0.651	0.310	0.227	0.043	-0.734	0.381	0.284
pH 5 w 10 mM	0.096	-0.157	0.061	-0.090	0.093	-1.121	0.493	0.510
pH 5 w 20 mM	0.235	0.391	0.147	-1.428	0.156	-0.560	0.301	0.261

Figure 8.23 (Four pages) Heat maps showing carrying capacity estimates generated by the Phenom model and the logistic model. Dark green indicates a high carrying capacity and red indicates a low carrying capacity. Colour coding was done for each data set separately i.e. carrying capacity for PA01 Logistic output is colour coded relative to the other values in that entire data set (including all acids), not to the Phenom data. First two heat maps show PA01 and the second two show PA1054.

	Malic PA01		Sorbic PA01		Benzoic PA01	
	Phenom	Logistic	Phenom	Logistic	Phenom	Logistic
pH 7	5.938	5.661	5.511	5.858	5.523	5.573
pH 7 w 5 mM	5.291	4.809	5.244	5.033	4.999	5.222
pH 7 w 10 mM	5.801	5.345	5.053	5.283	5.002	5.603
pH 7 w 20 mM	4.926	5.361	4.767	5.757	4.750	4.848
pH 6.5	6.252	6.169	6.039	6.903	5.710	5.697
pH 6.5 w 5 mM	5.446	5.392	5.619	5.752	5.274	5.205
pH 6.5 w 10 mM	4.428		5.638	6.020	4.973	5.407
pH 6.5 w 20 mM	4.627		4.890	5.005	4.640	4.655
pH 6	5.755	5.480	5.670	5.785	5.513	5.252
pH 6 w 5 mM	5.443	4.944	4.762	5.418	5.080	4.954
pH 6 w 10 mM	5.918	5.669	4.664	4.565	4.920	5.320
pH 6 w 20 mM	6.358	6.085	3.964		4.905	6.070
pH 5.5	5.524	5.493	6.186	6.116	5.021	5.025
pH 5.5 w 5 mM	5.266	5.089	5.007	8.351	4.836	4.833
pH 5.5 w 10 mM	5.884	5.736	4.151		5.275	5.532
pH 5.5 w 20 mM	5.012		1.428	-0.991	4.789	5.738
pH 5	6.530	7.110	6.462	7.233	6.226	7.388
pH 5 w 5 mM	4.654		4.819		5.776	7.043
pH 5 w 10 mM	1.377	3.692	0.095	-1.931	4.436	7.448
pH 5 w 20 mM	1.909		0.120	-0.874	1.663	-1.166

	Acetic PA1054		Butyric PA1054		Propionic PA1054		Citric PA1054	
	Phenom	Logistic	Phenom	Logistic	Phenom	Logistic	Phenom	Logistic
pH 7	6.611	6.615	6.556	6.563	6.451	6.475	6.297	6.343
pH 7 w 5 mM	6.565	6.554	6.554	6.624	6.680	6.723	6.409	6.384
pH 7 w 10 mM	6.960	6.889	6.649	6.761	7.015	7.153	6.398	6.492
pH 7 w 20 mM	6.852	6.976	7.026	7.082	6.911	7.989	5.921	6.081
pH 6.5	6.423	6.444	6.522	6.547	6.377	6.381	6.465	6.473
pH 6.5 w 5 mM	6.814	6.736	6.843	6.844	6.834	6.971	6.618	6.564
pH 6.5 w 10 mM	7.325	7.23	7.222	7.281	7.144	7.929	6.601	6.563
pH 6.5 w 20 mM	7.159	7.307	6.625	8.328	6.345		6.534	6.499
pH 6	6.580	6.506	6.603	6.623	6.365	6.390	6.611	6.592
pH 6 w 5 mM	7.020	6.921	6.586	6.754	6.779	7.356	6.252	6.145
pH 6 w 10 mM	7.453	7.555	6.742	7.823	6.339		6.892	6.797
pH 6 w 20 mM	3.700	-0.309	0.109	-0.903	0.068	-5.655	6.354	6.509
pH 5.5	5.859	5.997	6.321	6.537	6.299	6.4253	6.481	6.637
pH 5.5 w 5 mM	6.880	7.818	4.160		4.193		6.318	6.213
pH 5.5 w 10 mM	0.445	-0.141	0.046	-0.694	0.038	-12.2	5.926	7.824
pH 5.5 w 20 mM	0.076	-0.173	0.160	-0.433	0.132	-13.1	0.367	-0.07
pH 5	6.080	6.855	6.013	6.864	6.201	6.737	6.034	7.325
pH 5 w 5 mM	2.965	-0.094	0.047	-0.673	0.060	-13.80	0.917	1.587
pH 5 w 10 mM	0.133	-0.27	0.080	-0.288	0.097	-2.242	0.137	0.096
pH 5 w 20 mM	0.229	0.11	0.176	-0.274	0.296	-0.163	0.613	0.356

	Lactic PA1054		Malic PA1054		Sorbic PA1054		Benzoic PA1054	
	Phenom	Logistic	Phenom	Logistic	Phenom	Logistic	Phenom	Logistic
pH 7	6.145	6.175	6.105	6.174	6.025	6.0059	6.626	6.663
pH 7 w 5 mM	6.062	6.051	5.902	5.924	6.431	6.4371	6.480	6.485
pH 7 w 10 mM	6.152	6.119	6.008	5.976	6.496	6.4847	6.651	6.662
pH 7 w 20 mM	5.941	5.92	6.199	6.17	6.377	6.3462	6.651	6.650
pH 6.5	5.947	5.973	6.061	6.13	6.312	6.3806	6.583	6.610
pH 6.5 w 5 mM	6.127	6.06	5.861	5.87	6.503	6.3879	6.247	6.263
pH 6.5 w 10 mM	6.279	6.173	6.040	5.962	6.566	6.4622	6.951	6.930
pH 6.5 w 20 mM	6.400	6.245	6.341	6.253	6.347	6.274	6.700	6.680
pH 6	6.425	6.4	6.238	6.21	6.366	6.3833	6.317	6.302
pH 6 w 5 mM	6.197	6.068	5.915	5.787	6.490	6.4014	6.434	6.424
pH 6 w 10 mM	6.539	6.409	6.179	6.03	6.298	6.256	6.522	6.478
pH 6 w 20 mM	5.716	6.064	6.303	6.173	5.801	6.0609	6.359	6.280
pH 5.5	6.271	6.342	6.030	6.158	6.018	6.0934	6.398	6.331
pH 5.5 w 5 mM	6.072	6.003	5.976	5.952	5.780	5.8104	6.034	6.000
pH 5.5 w 10 mM	0.173	-0.558	6.107	6.361	5.794	6.0413	6.211	6.193
pH 5.5 w 20 mM	0.442	0.381	1.603		4.822		5.715	5.753
pH 5	6.179	6.96	6.079	6.829	6.184	7.0185	6.081	6.799
pH 5 w 5 mM	0.195	-0.457	4.475		0.886	0.3703	5.820	7.158
pH 5 w 10 mM	0.242	0.31	0.952	1.346	0.317	-0.016	5.122	
pH 5 w 20 mM	0.248	-0.404	1.027	1.247	0.428	0.1784	0.953	0.414

Figure 8.23 (Four pages) Heat maps showing carrying capacity estimates generated by the Phenom model and the logistic model. Dark green indicates a high carrying capacity and red indicates a low carrying capacity. Colour coding was done for each data set separately i.e. carrying capacity for PA01 Logistic output is colour coded relative to the other values in that entire data set (including all acids), not to the Phenom data. First two heat maps show PA01 and the second two show PA1054.

The comparison above shows a high level of similarity between the estimates produced by the two models. However, given the inability of the logistic and Baranyi models to successfully fit curves and estimate parameters in all cases due to the nature of the models, Phenom was selected in order to analyse the data in more detail. The PA01 data set is missing the values for lactic acid, this is because of an error in processing the data using the Phenom model that has yet to be rectified.

8.4 Comparison of strains, organic acids and pH

From the growth rate and carrying capacity data presented above, the organic acids with the biggest impact on growth are propionic acid and butyric acid. However, there is a clear difference across strains, with the clinical isolate PA1054 being more resistant to added organic acid than the lab strain PA01. Figure 8.24 below shows the GP regression curves for both PA01 and PA1054 across all pH with 5 mM added acetic acid, butyric acid or propionic acid. Although both strains grow well in neutral or very mildly acidic pH (pH 6), once the pH becomes slightly more acid at pH 5.5, both strains show an impact in growth, however this is more pronounced in PA01 than PA1054. In butyric and propionic acid, at pH 5 there is no growth in either strain, though in acetic acid, there appears to be some growth in PA1054, which would be of concern in a clinical setting.

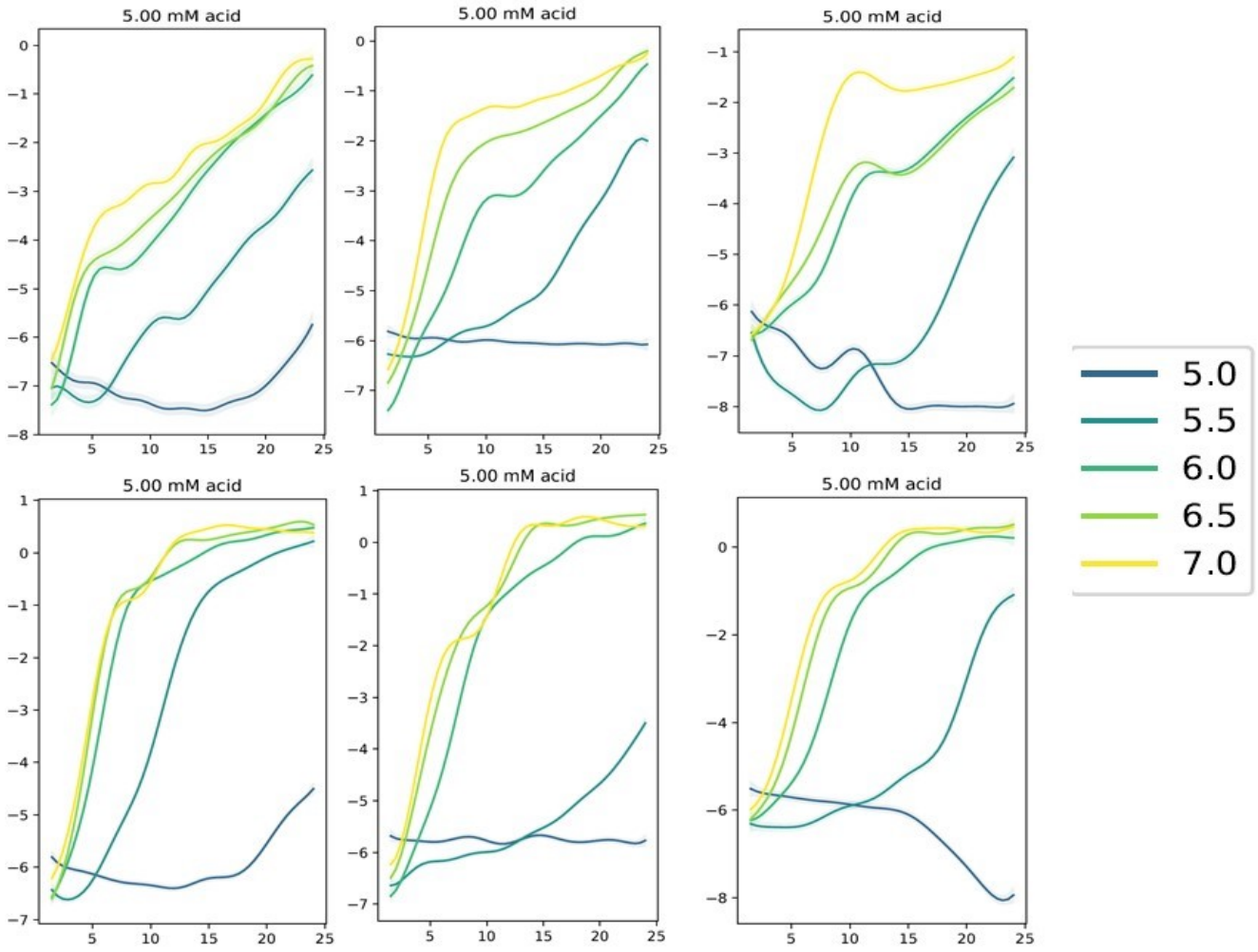


Figure 8.24 GP regression curve fits for PA01 (top) and PA1054 (bottom) with 5 mM of acetic acid (left), butyric acid (middle) and propionic acid (right) across 5 different pH. The key on the right shows colour code for each pH. X-axis shows time in hours. Y-axis shows normalised growth function.

The relative interaction plots discussed above provide a concise way of looking at these differences between strains. Figure 8.25 below shows the relative interaction plots for the addition of 20 mM acetic acid or citric acid in pH 6.5, 6, 5.5 and 5. These plots allow the analysis of the impact of the organic acid alone and removes the separate impact upon growth of the drop in pH. The comparison of acetic and citric acids is relevant because they are the two organic acids currently used in clinics as a topical treatment for burn wound infections. The plots below show that the two strains react in a similar way to the addition of organic acid. However, at pH 6 there is a big impact on growth in acetic acid that is not mirrored in citric acid, which appears to have no impact on growth whatsoever upon the addition of 20 mM citric acid, and growth appears somewhat improved compared to the control condition as discussed previously. This is also shown at pH 6.5. These plots are however limited, as they do not show whether there is any growth, only that it is reduced, the same, or improved compared to the control condition. It also does not show pH 7, which would allow comparison across the full range. However, it is useful in the way it gives a concise and clear way of looking at the impact of the added organic acid across both strains.

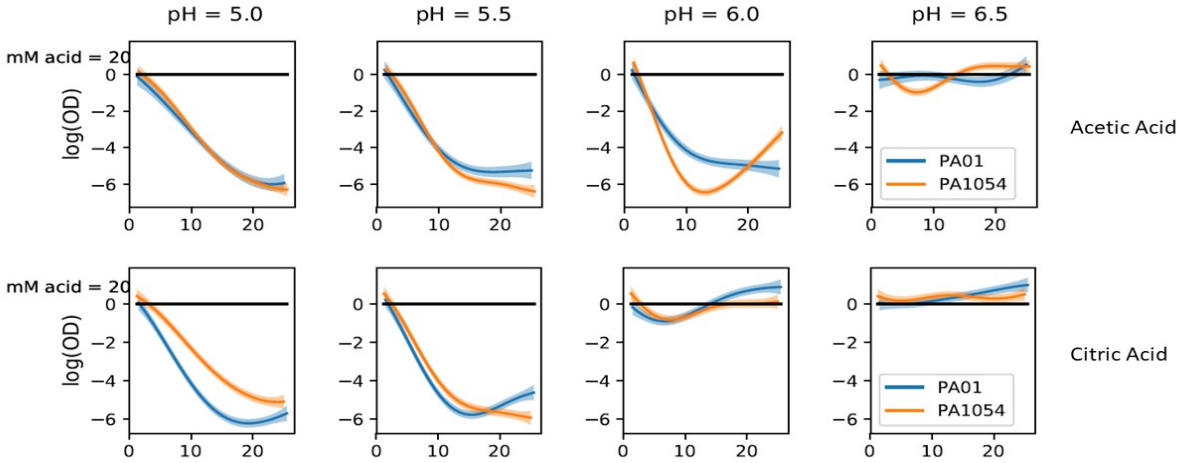


Figure 8.25 The relative interaction of each strain in acetic acid (top) and citric acid (bottom) compared to each pH control condition (no added organic acid) represented by the black line e.g. at pH 6.5 the black line represents pH 6.5 with no added acid.

An alternative way of looking at the impact of all the acids and conditions at once is by producing heat maps showing growth rates and carrying capacity and also clustering the growth rate and carrying capacity estimates generated by the Phenom model. Below at Figure 8.26 is a heat map showing carrying capacity estimates generated by the Phenom model for all conditions for PA01 and PA1054. The heat map shows a steady decrease in carrying capacity as pH falls with propionic, butyric and acetic acids being the most effective in both strains, and benzoic acid being the least effective in both strains. Beneath that in Figure 8.27 is a cluster heatmap generated by ClustVis, an open access online tool for clustering of multivariate data (https://biit.cs.ut.ee/clustvis/). The cluster map shows the organic acid, strain and pH / OA combinations which show the most similarity in carrying capacity estimates. The map shows that PA01 estimates and PA1054 estimates are clustered together (i.e. PA01 estimates are clustered separately to PA1054 estimates), indicating that the strains behaved more similarly regardless of condition. Lower pH conditions are clustered together, indicating that the carrying capacity was more similar in low pH regardless of addition of higher concentrations of organic acid at neutral pH, other than for pH 5 with no added organic acid.

	Acetic PA01	Butyric PA01	Propionic PA01	Citric PA01	Lactic PA01	Malic PA01	Sorbic PA01	Benzoic PA01
pH 7	5.890	5.795	5.738	5.665	2.973	5.938	5.511	5.523
pH 7 w 5 mM	5.575	6.511	5.210	5.441	1.619	5.291	5.244	4.999
pH 7 w 10 mM	5.807	5.290	5.052	5.371	1.939	5.801	5.053	5.002
pH 7 w 20 mM	5.433	5.534	4.582	5.229	2.813	4.926	4.767	4.750
pH 6.5	5.906	5.724	5.896	5.508	3.006	6.252	6.039	5.710
pH 6.5 w 5 mM	5.710	6.404	5.288	5.208	1.619	5.446	5.619	5.274
pH 6.5 w 10 mM	5.863	5.322	4.789	5.449	1.924	4.428	5.638	4.973
pH 6.5 w 20 mM	5.945	5.156	3.237	5.966	2.994	4.627	4.890	4.640
pH 6	5.774	5.721	5.584	5.743	3.310	5.755	5.670	5.513
pH 6 w 5 mM	5.735	5.869	4.983	5.223	2.017	5.443	4.762	5.080
pH 6 w 10 mM	5.658	4.982	3.555	5.987	2.448	5.918	4.664	4.920
pH 6 w 20 mM	0.232	0.750	0.102	6.094	3.281	6.358	3.964	4.905
pH 5.5	5.924	5.364	5.362	5.478	3.698	5.524	6.186	5.021
pH 5.5 w 5 mM	4.755	4.410	3.081	6.464	2.658	5.266	5.007	4.836
pH 5.5 w 10 mM	0.514	0.663	0.090	3.603	1.751	5.884	4.151	5.275
pH 5.5 w 20 mM	0.269	0.060	0.078	0.399	0.343	5.012	1.428	4.789
pH 5	6.398	6.164	6.787	6.361	3.453	6.530	6.462	6.226
pH 5 w 5 mM	1.183	0.310	0.043	0.381	1.314	4.654	4.819	5.776
pH 5 w 10 mM	0.096	0.061	0.093	0.493	0.573	1.377	0.095	4.436
pH 5 w 20 mM	0.235	0.147	0.156	0.301	0.476	1.909	0.120	1.663

	Acetic PA1054	Butyric PA1054	Propionic PA1054	Citric PA1054	Lactic PA1054	Malic PA1054	Sorbic PA1054	Benzoic PA1054
pH 7	6.611	6.556	6.451	6.297	6.145	6.105	6.025	6.626
pH 7 w 5 mM	6.565	6.554	6.680	6.409	6.062	5.902	6.431	6.480
pH 7 w 10 mM	6.960	6.649	7.015	6.398	6.152	6.008	6.496	6.651
pH 7 w 20 mM	6.852	7.026	6.911	5.921	5.941	6.199	6.377	6.651
pH 6.5	6.423	6.522	6.377	6.465	5.947	6.061	6.312	6.583
pH 6.5 w 5 mM	6.814	6.843	6.834	6.618	6.127	5.861	6.503	6.247
pH 6.5 w 10 mM	7.325	7.222	7.144	6.601	6.279	6.040	6.566	6.951
pH 6.5 w 20 mM	7.159	6.625	6.345	6.534	6.400	6.341	6.347	6.700
pH 6	6.580	6.603	6.365	6.611	6.425	6.238	6.366	6.317
pH 6 w 5 mM	7.020	6.586	6.779	6.252	6.197	5.915	6.490	6.434
pH 6 w 10 mM	7.453	6.742	6.339	6.892	6.539	6.179	6.298	6.522
pH 6 w 20 mM	3.700	0.109	0.068	6.354	5.716	6.303	5.801	6.359
pH 5.5	5.859	6.321	6.299	6.481	6.271	6.030	6.018	6.398
pH 5.5 w 5 mM	6.880	4.160	4.193	6.318	6.072	5.976	5.780	6.034
pH 5.5 w 10 mM	0.445	0.046	0.038	5.926	0.173	6.107	5.794	6.211
pH 5.5 w 20 mM	0.076	0.160	0.132	0.367	0.442	1.603	4.822	5.715
pH 5	6.080	6.013	6.201	6.034	6.179	6.079	6.184	6.081
pH 5 w 5 mM	2.965	0.047	0.060	0.917	0.195	4.475	0.886	5.820
pH 5 w 10 mM	0.133	0.080	0.097	0.137	0.242	0.952	0.317	5.122
pH 5 w 20 mM	0.229	0.176	0.296	0.613	0.248	1.027	0.428	0.953

Figure 8.26 Heat map showing carrying capacity estimates generated by the Phenom model. PA01 top and PA1054 bottom.

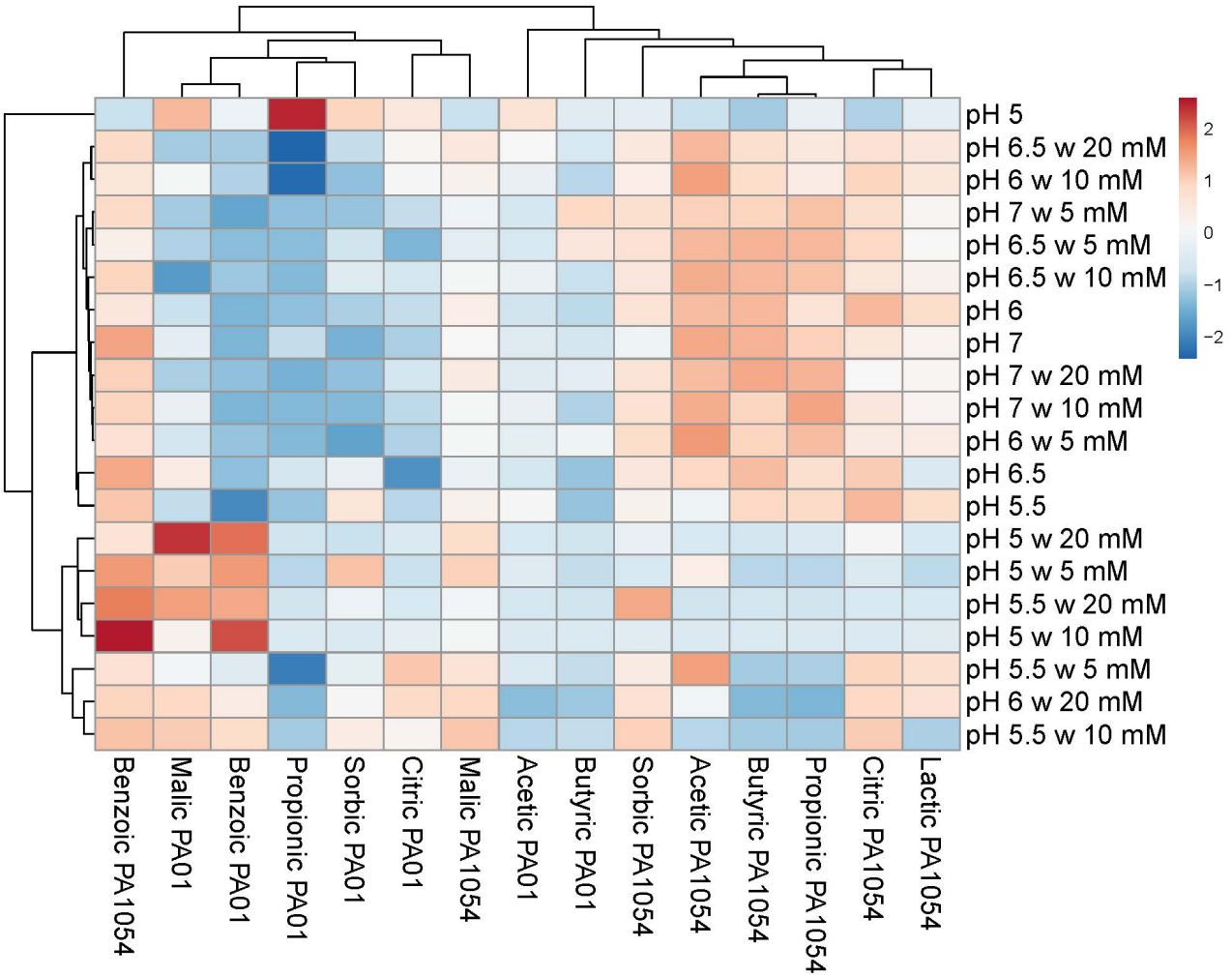


Figure 8.27 Clustvis (https://biit.cs.ut.ee/clustvis/) heat map of Phenom generated carrying capacities for PA01 and PA1054. Rows and columns are clustered using correlation distance and average linkage.

As with carrying capacity, the growth rates generated by the Phenom model show the same pattern. Below at Figure 8.28 is a heat map comparing all conditions for PA01 (top) and PA1054 (bottom) which indicates the strongest impact on growth rate is seen with acetic, butyric and propionic acid and the smallest impact in malic and benzoic acid.

	Acetic PA01	Butyric PA01	Propionic PA01	Citric PA01	Lactic PA01	Malic PA01	Sorbic PA01	Benzoic PA01
pH 7	0.688	0.688	0.574	0.583	0.593	0.866	0.474	0.677
pH 7 w 5 mM	0.676	0.676	0.576	0.735	0.480	0.855	0.470	0.657
pH 7 w 10 mM	0.682	0.682	0.573	0.806	0.483	0.927	0.469	0.653
pH 7 w 20 mM	0.663	0.663	0.570	0.739	0.526	0.852	0.466	0.639
pH 6.5	0.685	0.685	0.571	0.630	0.593	0.886	0.461	0.796
pH 6.5 w 5 mM	0.568	0.568	0.492	0.761	0.477	0.851	0.559	0.752
pH 6.5 w 10 mM	0.626	0.626	0.423	0.873	0.497	0.877	0.539	0.708
pH 6.5 w 20 mM	0.706	0.706	0.369	0.783	0.553	0.875	0.477	0.693
pH 6	0.628	0.628	0.482	0.657	0.594	1.004	0.520	0.756
pH 6 w 5 mM	0.454	0.454	0.405	0.784	0.507	0.972	0.522	0.711
pH 6 w 10 mM	0.465	0.465	0.266	0.932	0.533	1.113	0.467	0.721
pH 6 w 20 mM	0.111	0.111	0.116	0.706	0.563	1.132	0.324	0.670
pH 5.5	0.594	0.594	0.496	0.637	0.594	0.838	0.657	0.720
pH 5.5 w 5 mM	0.301	0.301	0.305	0.845	0.56	0.806	0.691	0.704
pH 5.5 w 10 mM	0.193	0.193	0.105	0.257	0.454	0.950	0.551	0.697
pH 5.5 w 20 mM	0.081	0.081	0.063	0.137	0.280	0.794	0.446	0.566
pH 5	0.466	0.466	0.518	0.506	0.594	0.656	0.516	0.469
pH 5 w 5 mM	0.251	0.251	0.073	0.150	0.394	0.554	0.541	0.456
pH 5 w 10 mM	0.089	0.089	0.071	0.156	0.322	0.441	0.269	0.354
pH 5 w 20 mM	0.092	0.092	0.065	0.129	0.305	0.425	0.167	0.230

	Acetic PA1054	Butyric PA1054	Propionic PA1054	Citric PA1054	Lactic PA1054	Malic PA1054	Sorbic PA1054	Benzoic PA1054
pH 7	1.111	0.998	0.971	0.912	1.080	0.904	1.110	1.068
pH 7 w 5 mM	1.102	0.887	0.913	0.909	1.110	0.964	1.079	1.056
pH 7 w 10 mM	1.105	0.876	0.816	0.914	1.126	1.051	1.086	1.078
pH 7 w 20 mM	1.129	0.837	0.619	0.912	1.092	1.083	1.051	1.053
pH 6.5	1.153	0.988	0.968	1.009	1.036	0.936	1.128	1.091
pH 6.5 w 5 mM	1.128	0.886	0.895	1.064	1.161	0.977	1.127	1.126
pH 6.5 w 10 mM	1.255	0.896	0.739	1.015	1.247	1.126	1.106	1.126
pH 6.5 w 20 mM	0.961	0.599	0.538	1.011	1.200	1.187	1.023	1.050
pH 6	0.976	0.904	0.910	0.976	0.989	0.912	1.060	0.966
pH 6 w 5 mM	1.025	0.790	0.715	0.981	1.113	0.874	0.963	1.052
pH 6 w 10 mM	0.991	0.656	0.728	1.059	1.126	1.047	0.947	1.040
pH 6 w 20 mM	0.471	0.113	0.028	0.841	0.638	1.079	0.764	0.905
pH 5.5	0.624	0.683	0.747	0.687	0.707	0.709	0.902	0.871
pH 5.5 w 5 mM	0.721	0.384	0.597	0.799	0.711	0.640	0.797	0.921
pH 5.5 w 10 mM	0.272	0.066	0.081	0.436	0.074	0.580	0.751	0.883
pH 5.5 w 20 mM	0.095	0.109	0.068	0.094	0.100	0.207	0.450	0.769
pH 5	0.471	0.480	0.520	0.427	0.503	0.521	0.511	0.466
pH 5 w 5 mM	0.474	0.081	0.048	0.125	0.086	0.393	0.176	0.507
pH 5 w 10 mM	0.107	0.082	0.066	0.119	0.113	0.169	0.153	0.460
pH 5 w 20 mM	0.119	0.111	0.136	0.113	0.103	0.174	0.178	0.255

Figure 8.28 Heat map of Phenom generated growth rates for PA01 (top) and PA1054 (bottom).

Shown in Figure 8.29 is a clustering heat map generated by ClustVis, again showing that the final growth profile is more dependent on the particular strain than the organic acid present, and that lower pH produced an impaired growth rate which was similar regardless of addition of organic acid. Again, this heat map shows that the biggest effect is seen in propionic and butyric acid.

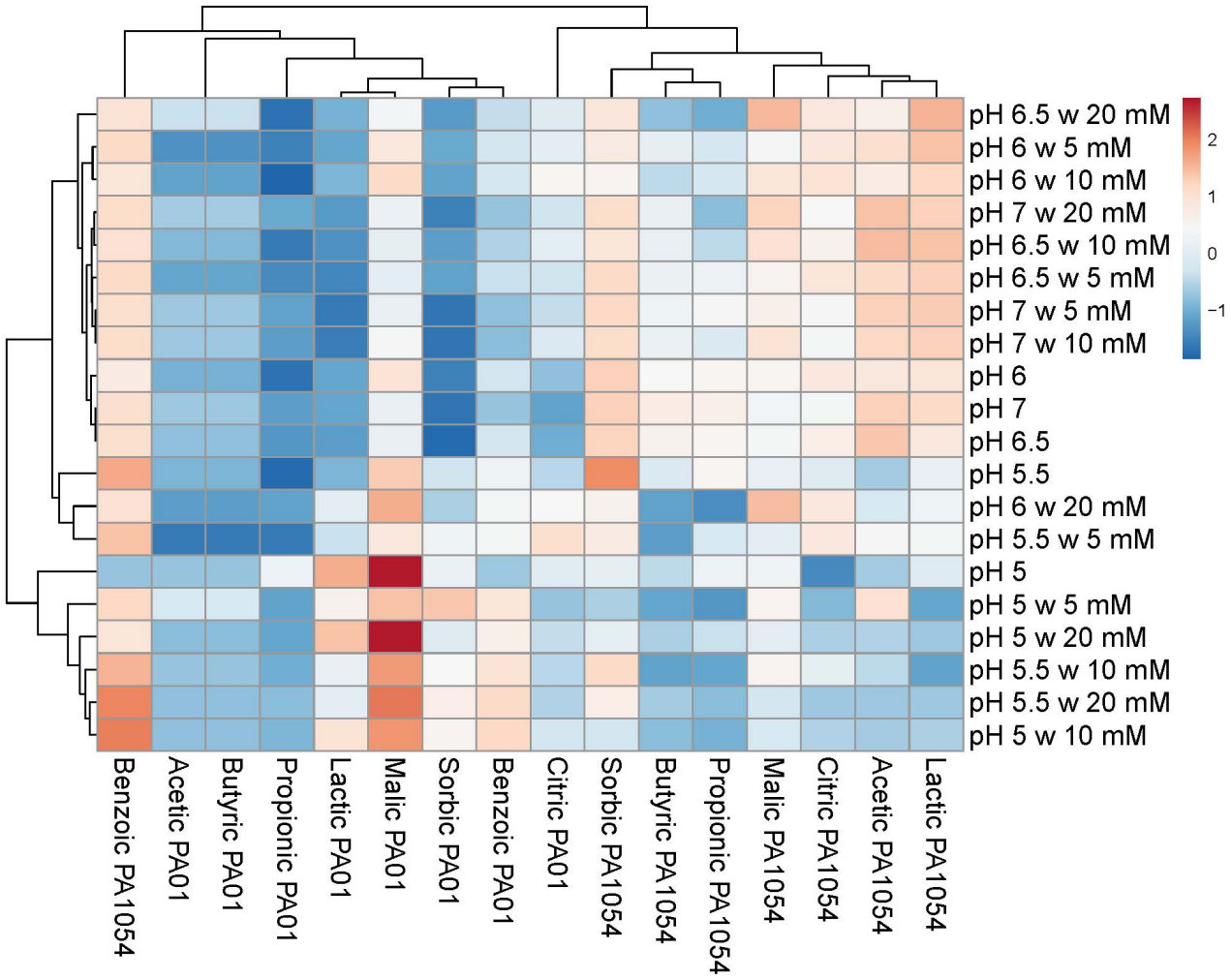


Figure 8.29 Clustvis (https://biit.cs.ut.ee/clustvis/) heat map of Phenom generated growth rates for PA01 and PA1054. Rows and columns are clustered using correlation distance and average linkage.

In conclusion, regardless of strain and pH, the biggest effect on both growth and carrying capacity was seen on the addition of propionic acid, with the next most effective being butyric acid. The least effective acids at inhibiting growth were benzoic acid and malic acid.

An additional complication seen in wounds is the formation of biofilms, which may also impact on growth. The next section of this chapter will focus on biofilm formation under these growth conditions in both strains.

8.5 The inhibition of biofilm formation by organic acids.

One of the biggest problems faced by clinicians when treating a *P. aeruginosa* wound infection is that of biofilm formation. A biofilm made up of extracellular substance secreted by the bacteria, forms around the colony making a physical barrier between antibiotics and the bacterial cells. This barrier is also effective again other forms of treatment, such as topical treatments, reaching the cells and diminishing infections. One of the key points to address when evaluating whether or not acetic acid (or an alternative organic acid) is effective as a topical treatment, is whether or not biofilm formation is inhibited by organic acids. If biofilm formation can be inhibited by organic acids, then synergistic treatments of infections with organic acids combined with antibiotics may be a more effective method of treating wound infections. In this study, biofilm formation under all conditions was measured by staining the same plate used for growth with 1% crystal violet. Plates were left overnight to dry and then the crystal violet re-solubilised in 30% acetic acid and measured at 550 nm (as described in Chapter Two).

PA01 produces much more biofilm than PA1054 in both stressful and non-stressful conditions. Figure 8.30 below shows biofilm quantification for both PA01 and PA1054 in acetic acid. It is clear from this data that the addition of acetic acid causes a large increase in biofilm formation indicating that biofilm production was stimulated even in conditions in which growth may have been simultaneously inhibited. For example;

at pH 6 with 10 mM acetic acid final OD in the growth experiment was 0.414 and biofilm production was measured at OD 5.152, whereas biofilm production was lower at pH 7 with no added acid (the control condition) where final OD was 0.861 and biofilm measured at OD 1.241. At pH 6.5 with 10 mM added acetic acid, biofilm production was measured at OD 6.047, compared to OD 1.157 at pH 6.5 with no added acetic acid.

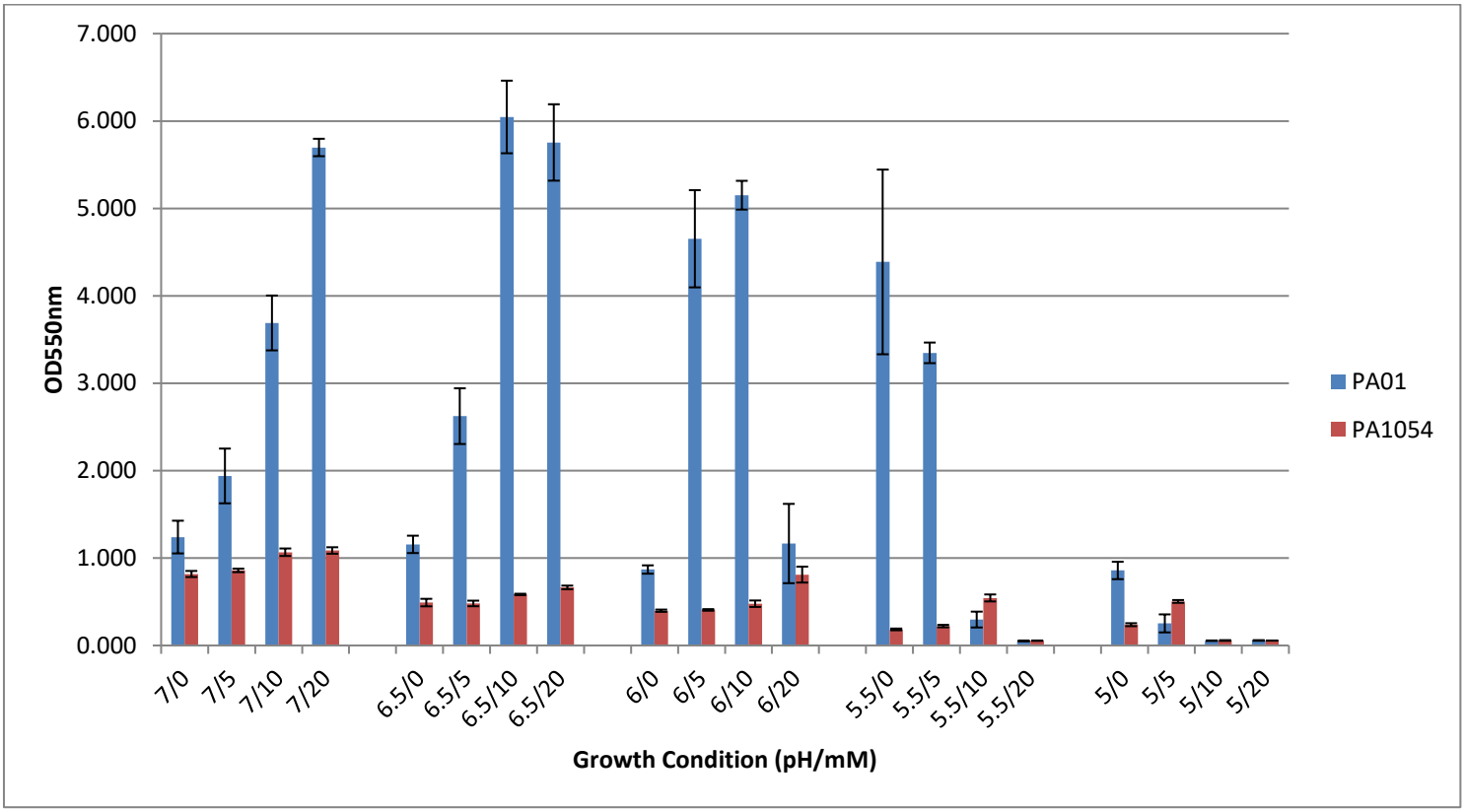


Figure 8.30 Biofilm quantification following growth in acetic acid across full pH range. Blue columns show PA01. Red columns show PA1054. Six replicates per condition. Error bars show SEM. Labels show pH/acid concentration i.e. 7/0 equals pH 7 with no added acetic acid and 7/5 equals pH 7 with 5 mM acetic acid.

Shown in Figure 8.31 is a heat map representing biofilm formation in PA01 and in Figure 8.32 for PA1054. The addition of propionic acid, even at neutral pH, had a greater overall impact on growth. However, at pH 7 with 10 mM and 20 mM propionic acid appears to stimulate large amounts of biofilm formation, causing six times more biofilm formation than pH 7 with no added acid. The same effect is seen again at pH 6.5 with 10 and 20 mM, and pH 6 with 10 mM acid. Where propionic acid has completely inhibited growth (at pH 6 with 20 mM) biofilm formation is also reduced, though it should be noted that there is still some biofilm being produced by these cells, which would otherwise be considered bacteriostatic. This may have an impact on the decision to use organic acids as a topical treatment, as there is an overall inability of organic acids to kill the cells at the concentrations tested and therefore the ability to produce biofilm is unaffected (though possibly reduced – this has not been quantified) and subsequently these cells could proliferate and cause infection once treatment has ceased.

As with growth, other organic acids had a lower impact on biofilm formation. The amount of biofilm formed generally followed a similar pattern to growth in terms of each acid, with lower growth resulting in less biofilm (other than above, where the stimulatory effect is noted). Overall benzoic acid had the weakest impact on biofilm formation, even in the most stringent conditions tested.

	Acetic	Propionic	Butyric	Sorbic	Citric	Lactic	Benzoic	Malic
pH 7 / 0 mM	1.241	1.558	2.131	1.261	4.835	1.670	1.902	1.964
pH 7 / 5 mM	1.940	4.892	4.383	2.247	4.468	2.179	4.192	2.231
pH 7 / 10 mM	3.690	9.742	4.925	3.367	4.673	3.067	6.808	3.017
pH 7 / 20 mM	5.698	9.362	5.605	3.680	6.312	5.250	6.690	4.358
pH 6.5 / 0 mM	1.157	2.564	1.705	1.342	4.142	2.303	1.712	1.726
pH 6.5 / 5 mM	2.624	4.182	2.442	1.931	5.030	1.911	3.838	1.936
pH 6.5 / 10 mM	6.047	9.202	5.418	3.693	4.490	3.502	4.555	3.527
pH 6.5 / 20 mM	5.755	9.543	4.675	3.513	3.250	2.942	5.240	5.145
pH 6 / 0 mM	0.869	1.831	1.036	1.051	2.950	1.886	1.500	2.078
pH 6 / 5 mM	4.653	6.273	2.208	2.496	4.597	1.847	3.550	2.627
pH 6 / 10 mM	5.152	7.482	4.727	3.061	2.988	3.287	4.980	4.747
pH 6 / 20 mM	1.167	0.256	0.247	0.295	1.824	3.063	4.797	6.175
pH 5.5 / 0 mM	4.388	6.310	2.410	2.577	2.995	2.446	4.073	4.043
pH 5.5 / 5 mM	3.348	3.712	4.755	2.077	1.876	4.163	5.987	5.407
pH 5.5 / 10 mM	0.297	0.146	0.067	0.097	1.098	3.031	6.327	5.405
pH 5.5 / 20 mM	0.052	0.056	0.064	0.039	0.067	0.111	5.017	2.100
pH 5 / 0 mM	0.858	0.995	0.827	1.059	0.703	0.560	0.712	0.596
pH 5 / 5 mM	0.253	0.090	0.063	0.082	0.071	0.805	1.661	1.041
pH 5 / 10 mM	0.055	0.058	0.060	0.037	0.063	0.055	4.582	0.059
pH 5 / 20 mM	0.058	0.056	0.062	0.039	0.063	0.097	1.994	0.061

	Acetic	Propionic	Butyric	Sorbic	Citric	Lactic	Benzoic	Malic
pH 7 / 0 mM	-0.564	-0.643	-0.124	-0.325	1.012	-0.384	-1.131	-0.511
pH 7 / 5 mM	-0.242	0.266	0.955	0.409	0.828	-0.021	0.100	-0.367
pH 7 / 10 mM	0.565	1.589	1.215	1.243	0.931	0.611	1.506	0.056
pH 7 / 20 mM	1.490	1.485	1.541	1.476	1.756	2.166	1.443	0.779
pH 6.5 / 0 mM	-0.603	-0.369	-0.329	-0.264	0.663	0.067	-1.233	-0.639
pH 6.5 / 5 mM	0.073	0.073	0.025	0.174	1.111	-0.212	-0.090	-0.526
pH 6.5 / 10 mM	1.651	1.442	1.451	1.485	0.839	0.921	0.295	0.331
pH 6.5 / 20 mM	1.516	1.535	1.095	1.351	0.214	0.522	0.663	1.203
pH 6 / 0 mM	-0.736	-0.568	-0.649	-0.481	0.063	-0.230	-1.347	-0.449
pH 6 / 5 mM	1.008	0.643	-0.087	0.594	0.893	-0.258	-0.245	-0.154
pH 6 / 10 mM	1.238	0.973	1.120	1.015	0.082	0.768	0.524	0.988
pH 6 / 20 mM	-0.598	-0.998	-1.027	-1.044	-0.504	0.608	0.425	1.757
pH 5.5 / 0 mM	0.886	0.653	0.009	0.655	0.086	0.169	0.036	0.609
pH 5.5 / 5 mM	0.407	-0.055	1.133	0.283	-0.478	1.392	1.065	1.344
pH 5.5 / 10 mM	-0.999	-1.028	-1.114	-1.191	-0.870	0.586	1.248	1.343
pH 5.5 / 20 mM	-1.112	-1.053	-1.115	-1.234	-1.389	-1.494	0.544	-0.437
pH 5 / 0 mM	-0.741	-0.796	-0.749	-0.475	-1.069	-1.174	-1.771	-1.248
pH 5 / 5 mM	-1.020	-1.043	-1.116	-1.202	-1.387	-1.000	-1.260	-1.008
pH 5 / 10 mM	-1.111	-1.052	-1.117	-1.236	-1.391	-1.534	0.310	-1.537
pH 5 / 20 mM	-1.109	-1.053	-1.116	-1.234	-1.391	-1.504	-1.081	-1.536

Figure 8.31 PA01 biofilm quantification in all conditions tested. Top: Heat map of raw data averages, bottom: heat map of standardised data. Data was standardised to the mean of pH 7 with no added organic acid (the control condition) and colour coded based on how many standard deviations the value

fell from the control. The colour scale runs from dark green to dark red with dark green indicating values that are higher (indicting more biofilm formation) and red indicating values that are lower (less biofilm formation).

The standardised and non-standardised heatmaps above at figure 8.31 are very similar in the trends that they show, with biofilm formation falling as pH decreases and organic acid concentration increases. On that basis, the data for PA1054 shown below at figure 8.32 shows only the mean values for biofilm production and not the standardised data.

	Acetic	Propionic	Butyric	Benzoic	Citric	Lactic	Malic	Sorbic
pH 7 / 0 mM	0.818	0.874	0.865	0.578	0.808	0.713	0.893	0.813
pH 7 / 5 mM	0.859	0.994	1.422	0.469	0.968	1.207	1.034	0.539
pH 7 / 10 mM	1.067	0.911	1.374	0.356	0.574	1.038	1.071	0.467
pH 7 / 20 mM	1.087	0.790	1.533	0.318	0.612	0.922	1.028	0.423
pH 6.5 / 0 mM	0.491	0.449	0.687	0.755	0.837	0.439	0.756	0.780
pH 6.5 / 5 mM	0.481	0.582	0.758	0.586	0.583	0.488	0.934	0.653
pH 6.5/10 mM	0.584	0.723	0.905	0.479	0.617	0.657	0.878	0.441
pH 6.5/20 mM	0.664	1.060	1.245	0.412	0.334	0.695	0.761	0.382
pH 6 / 0 mM	0.398	0.521	0.596	0.349	0.608	0.364	0.513	0.473
pH 6 / 5 mM	0.407	0.646	0.624	0.488	0.352	0.400	0.499	0.411
pH 6 / 10 mM	0.478	0.993	1.202	0.425	0.246	0.485	0.631	0.351
pH 6 / 20 mM	0.811	0.251	0.226	0.387	0.251	0.240	0.432	0.409
pH 5.5 / 0 mM	0.184	0.205	0.199	0.232	0.267	0.127	0.189	0.240
pH 5.5 / 5 mM	0.222	0.613	0.301	0.192	0.214	0.268	0.145	0.291
pH 5.5/10 mM	0.544	0.186	0.101	0.228	0.060	0.042	0.204	0.194
pH 5.5/20 mM	0.055	0.061	0.044	0.166	0.080	0.056	0.056	0.126
pH 5 / 0 mM	0.238	0.244	0.240	0.268	0.250	0.202	0.217	0.295
pH 5 / 5 mM	0.504	0.145	0.077	0.229	0.065	0.054	0.196	0.091
pH 5 / 10 mM	0.057	0.065	0.054	0.128	0.078	0.069	0.055	0.049
pH 5 / 20 mM	0.056	0.053	0.053	0.074	0.082	0.066	0.066	0.041

Figure 8.32 PA1054 biofilm quantification. Data shown is average OD550 nm of stained biofilm over six replicates. Colour scale runs from dark green for higher values to dark red for lower values.

At low pH PA1054 biofilm production is reduced compared to neutral pH. Sodium benzoate and potassium sorbate again seem to have less of an impact at low pH compared to the other organic acids, although both seem to have a greater impact than the other acids at neutral pH. As with PA01 there appears to be slight stimulation at neutral pH with the addition of organic acids, although this is less pronounced than in PA01 (this may be due to the effect noted with sodium benzoate of a lower internal pH at neutral external pH, and the reverse at lower pH). As demonstrated above, PA1054 produces less biofilm than PA01. However, this biofilm production is more resistant to the addition of organic acid than PA01, which reflects the pattern seen with growth.

One aspect not addressed in this chapter is that of biofilm eradication. Once an infection is established, it is likely that a biofilm will already have formed, in which case it would also be vital to understand which acids, and which concentrations or pH-concentration combinations, are effective at eradicating a biofilm that has already formed in order to then target the bacteria that live within it. This is something that is being addressed by another member of the laboratory.

8.5 Discussion

In this study we sought to analyse the impact of organic acids on growth and biofilm formation in two strains of *Pseudomonas aeruginosa,* a species which is commonly isolated from infected burn wounds. In order to properly analyse the large amounts of data collected, three mathematical models were employed to fit curves to the data and provide estimates for growth rate and carrying capacity. The Baranyi and logistic models proved to be inferior to the Bayesian approach. This was because growth of bacteria under stress conditions is often not logistic and therefore accurately estimating these parameters was either difficult or impossible. However, where growth was logistic, these models were able to fit curves and estimate associated parameters with low relative error. The Bayesian model however was able to fit a curve in all cases. This is because it makes no prior assumption about growth (unlike the

parametric models which assume logistic growth), therefore this method proved itself to be more accurate in all cases and therefore a more reliable tool.

Neither approach is perfect. In the case of the logistic and Baranyi models, if the culture has not reached stationary phase then the estimate is much more likely to be wrong (though not necessarily always wrong), especially when growth is not following a logistic growth pattern. There is a similar issue with Phenom, in that carrying capacity is given as the final point the curve has reached, and this may not give the whole story. If growth continues beyond the 24 hour time point, in the case of very slow growing conditions where the bacteria are highly stressed, this means that the 'carrying capacity' is likely to be higher after 48 hours than after 24 hours. However, the model cannot estimate this parameter based on data that it does not have. In both cases it is sensible to be cautious of relying too heavily on carrying capacity estimates.

Bacterial clumping was observed under one condition which may render some of the estimates for carrying capacity inaccurate, as the final ODs were unlikely to represent live cells. Every care was taken in the course of this study to eliminate this confounding factor including very vigorous shaking before a reading and reading multiple points in the well and taking an average. However, it was clear from visual inspection of the plates post-experiment that even very high levels of shaking were not sufficient to completely eliminate this problem. However, it was mostly likely not an issue and subsequent parameter estimates were not affected.

In terms of the actual effectiveness of each acid with regards to inhibiting growth and biolfim formation, the acids ranked as follows from most inhibitory to least: propionic; butyric; acetic; citric; lactic; sorbic; malic; and benzoic. It should be noted, however, that in some cases at neutral and near-neutral pH, citric and lactic acid improved growth, possibly due to being utilised as a carbon source, and propionic acid

appeared to stimulate biofilm formation. This is possibly an additional response by the cell to propionic acid stress, though this was not investigated further. Despite this, the organic acid with the most inhibitory effect on *P. aeruginosa* growth was propionic acid, and this was seen in both strains tested. The organic acid with the least impact was benzoic acid. A recent paper published by Chakraborty et al., (2017) showed an unusual effect seen with sodium benzoate. In this study, the strain of interest was *S. typhimurium*, used to investigate acid and osmotic stress. The authors used sodium benzoate as a pH clamp, as it has been reported to equilibrate internal pH (pH_i) and external pH (pH_e). However, this study showed that sodium benzoate is a poor choice of pH clamp, as the authors observed an effect whereby, in the presence of 30 mM sodium benzoate at pH_e of 7.2, pH_i dropped to 5.8. At pH_e of 5.6, pH_i was measured at 6.3, and at pH_e of 6, pH_i was 6.2. The effect was unexpected and has not been explained. However, in the context of the results presented in this report, exposure to 20 mM sodium benzoate had no impact upon the cell at pH 5.5, and in fact growth appeared improved at pH 5.5 in the presence of 10 mM and 20 mM compared with pH 7 (as demonstrated in the ClustVis map in Figure 8.27). Although the study outlined above was detailing *S. typhimurium*, there is a possibility that the same effect is being seen here in *P. aeruginosa.*

The molecular basis for the different impacts of each acid was not investigated. However, as well as the phenomenon described above, it is likely that there are myriad different cellular processes instigated and regulated upon the introduction of each organic acid. As has been discussed at length in this thesis, there is likely an impact on proton motive force and other cellular functions such as various metabolic pathways which may have an overall effect on how well the cells grow in these conditions. Given the unknown impact upon the cell of each organic acid anion upon dissociation within the cytoplasm, it is not possible to fully understand the reasons why some acids, and particularly propionic and butyric, have such a marked effect on the cell which is not replicated in other organic acids. For this reason, however, it is possible to say that this effect is not solely a result of H^+ ions accumulating in the cytoplasm, and there is

likely an effect here of different levels of toxicity of each anion and how efficiently it can be either metabolised or excreted from the cell (propionic acid, the most effective, has a pKa of 4.87 and benzoic acid, the least effective, has a pKa of 4.2).

In conclusion, despite the superior inhibitory effects of butyric and propionic acids, given the very strong and unpleasant smells of both acids (butyric acid in particular), these are not viable options for use within a hospital on patients. The next most effective acid at inhibiting both growth and biofilm formation was acetic acid. Given that this acid is currently used by clinicians to control surface wound infections and is generally well tolerated by patients, we conclude that the continued use of acetic acid is the most reasonable way forward, though it is likely that the current use of 5% acetic acid could be reduced and combined instead with mildly acidic solution to achieve the same results with less irritation. However, as stated above, there is an issue of biofilm stimulation at lower concentrations when combined with neutral and near-neutral pH. This needs to be considered when making recommendations for clinical use. The other major consideration when translating the data from this study to clinical use is that the methods used to generate this data do not recapitulate the conditions associated with a wound. There needs to be some caution therefore when applying these findings to a clinical setting. This study also did not look into synergistic treatment of topical organic acid combined with antibiotics. Given this would likely be a commonly used approach in clinic, it would seem prudent for more detailed research into this treatment to be devised and tested *in vivo*.

References

Al Zaid Siddiquee, K., Arauzo-Bravo, M.J., Shimizu, K., 2004. Metabolic flux analysis of pykF gene knockout *Escherichia coli* based on 13C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. Appl. Microbiol. Biotechnol. 63, 407–417.

Alav, I., Sutton, J.M., Rahman, K.M., 2018. Role of bacterial efflux pumps in biofilm formation. J. Antimicrob. Chemother.

Amarneh, B., Vik, S.B., 2005. Direct Transfer of NADH from Malate Dehydrogenase to Complex I in *Escherichia coli*. Cell Biochem. Biophys. 42, 251–262.

Anba-Mondoloni, J., Chaillou, S., Zagorec, M., Champomier-Vergès, M.-C., 2013. Catabolism of *N* -Acetylneuraminic Acid, a Fitness Function of the Food-Borne Lactic Acid Bacterium Lactobacillus sakei, Involves Two Newly Characterized Proteins. Appl. Environ. Microbiol. 79, 2012–2018.

Anderson, L.A., McNairn, E., Leubke, T., Pau, R.N., Boxer, D.H., 2000. ModE-Dependent Molybdate Regulation of the Molybdenum Cofactor Operon moa in *Escherichia coli*. J. Bacteriol. 182, 7035–7043.

Asati, S., Chaudhary, U., 2017. Prevalence of biofilm producing aerobic bacterial isolates in burn wound infections at a tertiary care hospital in northern India. Ann. Burns Fire Disasters 30, 39–42.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2: 2006 0008 10.1038/msb4100050

Baranyi, J., Roberts, T.A., 1994. A dynamic approach to predicting bacterial growth in food. Int. J. Food Microbiol. 23, 277–294.

Battesti, A., Majdalani, N., Gottesman, S., 2011. The RpoS-mediated general stress response in *Escherichia coli*. Annu. Rev. Microbiol. 65, 189–213.

Bekhit, A., Fukamachi, T., Saito, H., Kobayashi, H., 2011. The role of OmpC and OmpF in acidic resistance in *Escherichia coli*. Biol. Pharm. Bull. 34, 330–334.

Bernal, V., Castaño-Cerezo, S., Cánovas, M., 2016. Acetate metabolism regulation in *Escherichia coli*: carbon overflow, pathogenicity, and beyond. Appl. Microbiol. Biotechnol. 100, 8985–9001.

Blanchard, C.Z., Amspacher, D., Strongin, R., Waldrop, G.L., 1999. Inhibition of biotin carboxylase by a reaction intermediate analog: implications for the kinetic mechanism. Biochem. Biophys. Res. Commun. 266, 466–471.

Blattner, F.R., 1997. The Complete Genome Sequence of *Escherichia coli* K-12. Science 277, 1453–1462.

Bologna, F.P., Campos-Bermudez, V.A., Saavedra, D.D., Andreo, C.S., Drincovich, M.F., 2010. Characterization of *Escherichia coli* EutD: a phosphotransacetylase of the ethanolamine operon. J. Microbiol. 48, 629–636.

Bonnet, R., 2004. Growing Group of Extended-Spectrum -Lactamases: the CTX-M Enzymes. Antimicrob. Agents Chemother. 48, 1–14. 4

Booth, I.R., 1985. Regulation of cytoplasmic pH in bacteria. Microbiol. Rev. 49, 359–378.

Brégeon, D., Colot, V., Radman, M., Taddei, F., 2001. Translational misreading: a tRNA modification counteracts a +2 ribosomal frameshift. Genes Dev. 15, 2295–2306.

Breidt, F., Hayes, J.S., McFeeters, R.F., 2004. Independent effects of acetic acid and pH on survival of *Escherichia coli* in simulated acidified pickle products. J. Food Prot. 67, 12–18.

Brinkkötter, A., Klöss, H., Alpert, C., Lengeler, J.W., 2000. Pathways for the utilization of N-acetyl-galactosamine and galactosamine in *Escherichia coli*. Mol. Microbiol. 37, 125–135.

Brock, M., Maerker, C., Schütz, A., Völker, U., Buckel, W., 2002. Oxidation of propionate to pyruvate in *Escherichia coli*: Involvement of methylcitrate dehydratase and aconitase. Eur. J. Biochem. 269, 6184–6194.

Burton, N.A., Johnson, M.D., Antczak, P., Robinson, A., Lund, P.A., 2010. Novel Aspects of the Acid Response Network of *E. coli* K-12 Are Revealed by a Study of Transcriptional Dynamics. J. Mol. Biol. 401, 726–742.

Cabrera, R., Baez, M., Pereira, H.M., Caniuguir, A., Garratt, R.C., Babul, J., 2011. The Crystal Complex of Phosphofructokinase-2 of *Escherichia coli* with Fructose-6-phosphate: Kinetic and structural analysis of the allosteric ATP Inhibition. J. Biol. Chem. 286, 5774–5783.

Cantón, R., González-Alba, J.M., Galán, J.C., 2012. CTX-M Enzymes: Origin and Diffusion. Front. Microbiol. 3.

Casalino, M., Prosseda, G., Barbagallo, M., Iacobino, A., Ceccarini, P., Carmela Latella, M., Nicoletti, M., Colonna, B., 2010. Interference of the CadC regulator in the arginine-dependent acid resistance system of *Shigella* and enteroinvasive *E. coli*. Int. J. Med. Microbiol. 300, 289–295.

Castanie-Cornet, M.-P., Cam, K., Bastiat, B., Cros, A., Bordes, P., Gutierrez, C., 2010. Acid stress response in *Escherichia coli*: mechanism of regulation of gadA transcription by RcsB and GadE. Nucleic Acids Res. 38, 3546–3554.

Castanie-Cornet, M.P., Penfound, T.A., Smith, D., Elliott, J.F., Foster, J.W., 1999. Control of acid resistance in *Escherichia coli*. J. Bacteriol. 181, 3525–3535.

Chakraborty, S., Winardhi, R.S., Morgan, L.K., Yan, J., Kenney, L.J., 2017. Non-canonical activation of OmpR drives acid and osmotic stress responses in single bacterial cells. Nat. Commun. 8.

Charbit, A., Reizer, J., Saier, M.H., 1996. Function of the duplicated IIB domain and oligomeric structure of the fructose permease of *Escherichia coli*. J. Biol. Chem. 271, 9997–10003.

Chattopadhyay, M.K., Keembiyehetty, C.N., Chen, W., Tabor, H., 2015. Polyamines Stimulate the Level of the σ^{38} Subunit (RpoS) of *Escherichia coli* RNA Polymerase, Resulting in the Induction of the Glutamate Decarboxylase-dependent Acid Response System via the *gadE* Regulon. J. Biol. Chem. 290, 17809–17821.

Chevalier, S., Bouffartigues, E., Bodilis, J., Maillot, O., Lesouhaitier, O., Feuilloley, M.G.J., Orange, N., Dufour, A., Cornelis, P., 2017. Structure, function and regulation of *Pseudomonas aeruginosa* porins. FEMS Microbiol. Rev. 41, 698–722.

Chiang, R.C., Cavicchioli, R., Gunsalus, R.P., 1992. Identification and characterization of narQ, a second nitrate sensor for nitrate-dependent gene regulation in *Escherichia coli*. Mol. Microbiol. 6, 1913–1923.

Chilcott, G.S., Hughes, K.T., 2000. Coupling of Flagellar Gene Expression to Flagellar Assembly in *Salmonella enterica Serovar Typhimurium* and *Escherichia coli*. Microbiol. Mol. Biol. Rev. 64, 694–708.

Cholley, P., Ka, R., Guyeux, C., Thouverez, M., Guessennd, N., Ghebremedhin, B., Frank, T., Bertrand, X., Hocquet, D., 2014. Population structure of clinical *Pseudomonas aeruginosa* from West and Central African countries. PloS One 9, e107008.

Church, D., Elsayed, S., Reid, O., Winston, B., Lindsay, R., 2006. Burn Wound Infections. Clin. Microbiol. Rev. 19, 403–434.

Clermont, O., Johnson, J.R., Menard, M., Denamur, E., 2007. Determination of *Escherichia coli* O types by allele-specific polymerase chain reaction: application to the O types involved in human septicemia. Diagn. Microbiol. Infect. Dis. 57, 129–136.

Condemine, G., Berrier, C., Plumbridge, J., Ghazi, A., 2005. Function and Expression of an N-Acetylneuraminic Acid-Inducible Outer Membrane Channel in *Escherichia coli*. J. Bacteriol. 187, 1959–1965.

Cotter, P.A., Gunsalus, R.P., 1989. Oxygen, nitrate, and molybdenum regulation of dmsABC gene expression in *Escherichia coli*. J. Bacteriol. 171, 3817–3823.

Cummings, J.H., Pomare, E.W., Branch, W.J., Naylor, C.P., Macfarlane, G.T., 1987. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. Gut 28, 1221–1227.

Dahl, J.-U., Koldewey, P., Salmon, L., Horowitz, S., Bardwell, J.C.A., Jakob, U., 2015. HdeB functions as an acid-protective chaperone in bacteria. J. Biol. Chem. 290, 65–75.

De Biase, D., Pennacchietti, E., 2012. Glutamate decarboxylase-dependent acid resistance in orally acquired bacteria: function, distribution and biomedical implications of the *gadBC* operon: Role of *gadBC* operon in orally acquired bacteria. Mol. Microbiol. 86, 770–786.

De Lay, N.R., Cronan, J.E., 2008. Genetic Interaction Between the *Escherichia coli* AcpT Phosphopantetheinyl Transferase and the YejM Inner Membrane Protein. Genetics 178, 1327–1337.

Deatherage, D.E., Barrick, J.E., 2014. Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq, in: Sun, L., Shou, W. (Eds.), Engineering and Analyzing Multicellular Systems. Springer New York, New York, NY, pp. 165–188.

Diez-Gonzalez, F., Russell, J.B., 1997. The ability of *Escherichia coli* O157:H7 to decrease its intracellular pH and resist the toxicity of acetic acid. Microbiol. Read. Engl. 143 (Pt 4), 1175–1180.

Donovan, G.T., Norton, J.P., Bower, J.M., Mulvey, M.A., 2013. Adenylate cyclase and the cyclic AMP receptor protein modulate stress resistance and virulence capacity of uropathogenic *Escherichia coli*. Infect. Immun. 81, 249–258.

Dragosits, M., Mozhayskiy, V., Quinones-Soto, S., Park, J., Tagkopoulos, I., 2013. Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in Escherichia coli. Mol. Syst. Biol. 9, 643.

Eguchi, Y., Ishii, E., Hata, K., Utsumi, R., 2011. Regulation of Acid Resistance by Connectors of Two-Component Signal Transduction Systems in *Escherichia coli*. J. Bacteriol. 193, 1222–1228.

Eguchi, Y., Ishii, E., Yamane, M., Utsumi, R., 2012. The connector SafA interacts with the multi-sensing domain of PhoQ in *Escherichia coli*: SafA interacts with the sensor domain of PhoQ. Mol. Microbiol. 85, 299–313.

Eguchi, Y., Utsumi, R., 2014. Alkali metals in addition to acidic pH activate the EvgS histidine kinase sensor in *Escherichia coli*. J. Bacteriol. 196, 3140–3149.

El-Hajj, Z.W., Reyes-Lamothe, R., Newman, E.B., 2013. Cell division, one-carbon metabolism and methionine synthesis in a metK-deficient *Escherichia coli* mutant, and a role for MmuM. Microbiology 159, 2036–2048.

Enjalbert, B., Millard, P., Dinclaux, M., Portais, J.-C., Létisse, F., 2017. Acetate fluxes in *Escherichia coli* are determined by the thermodynamic control of the Pta-AckA pathway. Sci. Rep. 7, 42135.

Erhardt, H., Steimle, S., Muders, V., Pohl, T., Walter, J., Friedrich, T., 2012. Disruption of individual nuo-genes leads to the formation of partially assembled NADH:ubiquinone oxidoreductase (complex I) in *Escherichia coli*. Biochim. Biophys. Acta BBA - Bioenerg. 1817, 863–871.

Farr, S.B., Arnosti, D.N., Chamberlin, M.J., Ames, B.N., 1989. An apaH mutation causes AppppA to accumulate and affects motility and catabolite repression in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 86, 5010–5014.

Flores-Mireles, A.L., Walker, J.N., Caparon, M., Hultgren, S.J., 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat. Rev. Microbiol. 13, 269–284.

Fodor, S.P., 1997. DNA SEQUENCING: Massively Parallel Genomics. Science 277, 393–395.

Foster, J.W., 2004. *Escherichia coli* acid resistance: tales of an amateur acidophile. Nat. Rev. Microbiol. 2, 898–907.

Fraise, A.P., Wilkinson, M.A.C., Bradley, C.R., Oppenheim, B., Moiemen, N., 2013. The antibacterial activity and stability of acetic acid. J. Hosp. Infect. 84, 329–331.

Fricke, J., Neuhard, J., Kelln, R.A., Pedersen, S., 1995. The cmk gene encoding cytidine monophosphate kinase is located in the rpsA operon and is required for normal replication rate in *Escherichia coli*. J. Bacteriol. 177, 517–523.

Gale E F. 1946 The bacterial amino acid decarboxylases. Advan. Enzymol. 6:1-32.

Gershman, S., Blei, D., 2011. A Tutorial on Baysian Nonparametic Models.

Goodall, E.C.A., Robinson, A., Johnston, I.G., Jabbari, S., Turner, K.A., Cunningham, A.F., Lund, P.A., Cole, J.A., Henderson, I.R., 2018. The Essential Genome of *Escherichia coli* K-12. mBio 9.

Gunsalus, R., 2011. EcoCyc. Aerob. Respir.

Hajská, M., Slobodníková, L., Hupková, H., Koller, J., 2014. In vitro efficacy of various topical antimicrobial agents in different time periods from contamination to application against 6 multidrug-resistant bacterial strains isolated from burn patients. Burns 40, 713–718.

Halstead, F.D., Rauf, M., Moiemen, N.S., Bamford, A., Wearn, C.M., Fraise, A.P., Lund, P.A., Oppenheim, B.A., Webber, M.A., 2015. The Antibacterial Activity of Acetic Acid against Biofilm-Producing Pathogens of Relevance to Burns Patients. PLOS ONE 10, e0136190.

Harden, M.M., He, A., Creamer, K., Clark, M.W., Hamdallah, I., Martinez, K.A., Kresslein, R.L., Bush, S.P., Slonczewski, J.L., 2015. Acid-Adapted Strains of *Escherichia coli* K-12 Obtained by Experimental Evolution. Appl. Environ. Microbiol. 81, 1932–1941.

Harwani, D., Zangoui, P., Mahadevan, S., 2012. The -Glucoside (bgl) Operon of *Escherichia coli* Is Involved in the Regulation of *oppA*, Encoding an Oligopeptide Transporter. J. Bacteriol. 194, 90–99.

Hayashi, K., Nakashima, R., Sakurai, K., Kitagawa, K., Yamasaki, S., Nishino, K., Yamaguchi, A., 2016. AcrB-AcrA Fusion Proteins That Act as Multidrug Efflux Transporters. J. Bacteriol. 198, 332–342.

Helling, R.B., Vargas, C.N., Adams, J., 1987. Evolution of *Escherichia coli* during growth in a constant environment. Genetics 116, 349–358.

Henderson, J.C., Zimmerman, S.M., Crofts, A.A., Boll, J.M., Kuhns, L.G., Herrera, C.M., Trent, M.S., 2016. The Power of Asymmetry: Architecture and Assembly of the Gram-Negative Outer Membrane Lipid Bilayer. Annu. Rev. Microbiol. 70, 255–278.

Hengge-Aronis, R., 2002. Signal Transduction and Regulatory Mechanisms Involved in Control of the S (RpoS) Subunit of RNA Polymerase. Microbiol. Mol. Biol. Rev. 66, 373–395.

Hengge-Aronis, R., 1993. Survival of hunger and stress: the role of rpoS in early stationary phase gene regulation in *E. coli*. Cell 72, 165–168.

Hickey, G.I., Fletcher, H.L., Winter, P.C., 2007. Genetics, 3rd ed. ed, BIOS instant notes. Taylor & Francis Group, New York, NY.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Huang, Y.-J., Tsai, T.-Y., Pan, T.-M., 2007. Physiological Response and Protein Expression under Acid Stress of *Escherichia coli* O157:H7 TWC01 Isolated from Taiwan. J. Agric. Food Chem. 55, 7182–7191.

Imran, F.-H., Karim, R., Maat, N.H., 2016. Managing burn wounds with SMARTPORE Technology polyurethane foam: two case reports. J. Med. Case Reports 10.

Ishii, E., Eguchi, Y., Utsumi, R., 2013. Mechanism of Activation of PhoQ/PhoP Two-Component Signal Transduction by SafA, an Auxiliary Protein of PhoQ Histidine Kinase in *Escherichia coli*. Biosci. Biotechnol. Biochem. 77, 814–819.

Jenkins, L.S., Nunn, W.D., 1987. Regulation of the ato operon by the atoC gene in *Escherichia coli*. J. Bacteriol. 169, 2096–2102.

Jensen, P.A., Zhu, Z., van Opijnen, T., 2016. Network analysis links genome-wide phenotypic and transcriptional stress responses in a bacterial pathogen with a large pan-genome.

Johnson, M.D., Bell, J., Clarke, K., Chandler, R., Pathak, P., Xia, Y., Marshall, R.L., Weinstock, G.M., Loman, N.J., Winn, P.J., Lund, P.A., 2014. Characterization of mutations in the PAS domain of the EvgS sensor kinase selected by laboratory evolution for acid resistance in *E scherichia coli*: PAS domain mutants in the EvgS histidine kinase. Mol. Microbiol. 93, 911–927.

Johnson, M.D., Burton, N.A., Gutierrez, B., Painter, K., Lund, P.A., 2011. RcsB Is Required for Inducible Acid Resistance in *Escherichia coli* and Acts at gadE-Dependent and -Independent Promoters. J. Bacteriol. 193, 3653–3656.

Johnstone, D.B., Farr, S.B., 1991. AppppA binds to several proteins in *Escherichia coli*, including the heat shock and oxidative stress proteins DnaK, GroEL, E89, C45 and C40. EMBO J. 10, 3897–3904.

Journet, L., Cascales, E., 2016. The Type VI Secretion System in *Escherichia coli* and related species. EcoSal Plus 7.

Joyce, A.R., Reed, J.L., White, A., Edwards, R., Osterman, A., Baba, T., Mori, H., Lesely, S.A., Palsson, B.O., Agarwalla, S., 2006. Experimental and Computational Assessment of Conditionally Essential Genes in *Escherichia coli*. J. Bacteriol. 188, 8259–8271.

Kanjee, U., Houry, W.A., 2013a. Mechanisms of Acid Resistance in *Escherichia coli*. Annu. Rev. Microbiol. 67, 65–81. https://doi.org/10.1146/annurev-micro-092412-155708

Kennedy, P., Brammah, S., Wills, E., 2010. Burns, biofilm and a new appraisal of burn wound sepsis. Burns 36, 49–56.

King, T., Lucchini, S., Hinton, J.C.D., Gobius, K., 2010. Transcriptomic Analysis of *Escherichia coli* O157:H7 and K-12 Cultures Exposed to Inorganic and Organic Acids in Stationary Phase Reveals Acidulant- and Strain-Specific Acid Tolerance Responses. Appl. Environ. Microbiol. 76, 6514–6528.

Kitko, R.D., Wilks, J.C., Garduque, G.M., Slonczewski, J.L., 2010. Osmolytes contribute to pH homeostasis of *Escherichia coli*. PLoS ONE 5, e10078.

Kriek, M., Martins, F., Leonardi, R., Fairhurst, S.A., Lowe, D.J., Roach, P.L., 2007. Thiazole Synthase from *Escherichia coli*: AN investigation of the substrates and purified proteins required for activity *in vitro*. J. Biol. Chem. 282, 17413–17423.

Krisko, A., Copic, T., Gabaldón, T., Lehner, B., Supek, F., 2014. Inferring gene function from evolutionary change in signatures of translation efficiency. Genome Biol. 15, R44.

Kroll, R.G., Booth, I.R., 1981. The role of potassium transport in the generation of a pH gradient in *Escherichia coli*. Biochem. J. 198, 691–698.

Krulwich, T.A., 1983. Na+/H+ antiporters. Biochim. Biophys. Acta 726, 245–264.

Krulwich, T.A., Sachs, G., Padan, E., 2011. Molecular aspects of bacterial pH sensing and homeostasis. Nat. Rev. Microbiol. 9, 330–343.

Kwong, W.K., Zheng, H., Moran, N.A., 2017. Convergent evolution of a modified, acetate-driven TCA cycle in bacteria. Nat. Microbiol. 2, 17067.

Kyriakidis, D.A., Tiligada, E., 2009. Signal transduction and adaptive regulation through bacterial two-component systems: the *Escherichia coli* AtoSC paradigm. Amino Acids 37, 443–458.

Lackraj, T., Kim, J.I., Tran, S., Barnett Foster, D.E., 2016. Differential modulation of flagella expression in enterohaemorrhagic *Escherichia coli* O157: H7 by intestinal short-chain fatty acid mixes. Microbiology 162, 1761–1772.

Lambert, L.A., Abshire, K., Blankenhorn, D., Slonczewski, J.L., 1997. Proteins induced in *Escherichia coli* by benzoic acid. J. Bacteriol. 179, 7595–7599.

Langridge, G.C., Phan, M.-D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., Wain, J., Parkhill, J., Turner, A.K., 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. Genome Res. 19, 2308–2316.

Lavigne, J.-P., Vergunst, A.C., Goret, L., Sotto, A., Combescure, C., Blanco, J., O'Callaghan, D., Nicolas-Chanoine, M.-H., 2012. Virulence Potential and Genomic Mapping of the Worldwide Clone *Escherichia coli* ST131. PLoS ONE 7, e34294.

Lazar, S.W., Almirón, M., Tormo, A., Kolter, R., 1998. Role of the *Escherichia coli* SurA protein in stationary-phase survival. J. Bacteriol. 180, 5704–5711.

Lehninger, A.L., Nelson, D.L., Cox, M.M., 2008. Lehninger principles of biochemistry, 5th ed. ed. W.H. Freeman, New York.

Lei, R., Ye, K., Gu, Z., Sun, X., 2015. Diminishing returns in next-generation sequencing (NGS) transcriptome data. Gene 557, 82–87.

Leimkühler, S., 2014. The Biosynthesis of the Molybdenum Cofactor in *Escherichia coli* and Its Connection to FeS Cluster Assembly and the Thiolation of tRNA. Adv. Biol. 2014, 1–21.

Lenski, R.E., Rose, M.R., Simpson, S.C., Tadler, S.C., 1991a. Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. Am. Nat. 138, 1315–1341. https://doi.org/10.1086/285289

Lenski, R.E., Wiser, M.J., Ribeck, N., Blount, Z.D., Nahum, J.R., Morris, J.J., Zaman, L., Turner, C.B., Wade, B.D., Maddamsetti, R., Burmeister, A.R., Baird, E.J., Bundy, J., Grant, N.A., Card, K.J., Rowles, M., Weatherspoon, K., Papoulis, S.E., Sullivan, R., Clark, C., Mulka, J.S., Hajela, N., 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. Proc. Biol. Sci. 282, 20152292.

Lin, J., Lee, I.S., Frey, J., Slonczewski, J.L., Foster, J.W., 1995. Comparative analysis of extreme acid survival in *Salmonella typhimurium*, *Shigella flexneri*, and *Escherichia coli*. J. Bacteriol. 177, 4097–4104.

Lipkin, L., Smith, D., 2004. Logistic Growth Model - Background: Logistic Modelling. J. Online Math. Its Appl.

Lu, S., Zgurskaya, H.I., 2012. Role of ATP binding and hydrolysis in assembly of MacAB-TolC macrolide transporter: Role of ATP in assembly of MacAB-TolC. Mol. Microbiol. 86, 1132–1143.

Lund, P., Tramonti, A., De Biase, D., 2014a. Coping with low pH: molecular strategies in neutralophilic bacteria. FEMS Microbiol. Rev. 38, 1091–1125.

Ma, D., Lu, P., Shi, Y., 2013. Substrate Selectivity of the Acid-activated Glutamate/ -Aminobutyric acid (GABA) Antiporter GadC from *Escherichia coli*. J. Biol. Chem. 288, 15148–15153.

Macfarlane, S., Macfarlane, G.T., 2003. Regulation of short-chain fatty acid production. Proc. Nutr. Soc. 62, 67–72.

Madigan, M.T., Brock, T.D. (Eds.), 2009. Brock biology of microorganisms, 12. ed., internat. ed. ed. Pearson/Benjamin Cummings, San Francisco, Calif.

Marriott, A.S., Copeland, N.A., Cunningham, R., Wilkinson, M.C., McLennan, A.G., Jones, N.J., 2015. Diadenosine 5′, 5‴-P1,P4-tetraphosphate (Ap4A) is synthesized in response to DNA damage and inhibits the initiation of DNA replication. DNA Repair 33, 90–100.

Martorana, A.M., Sperandeo, P., Polissi, A., Dehò, G., 2011. Complex transcriptional organization regulates an *Escherichia coli* locus implicated in lipopolysaccharide biogenesis. Res. Microbiol. 162, 470–482.

Mates, A.K., Sayed, A.K., Foster, J.W., 2007. Products of the *Escherichia coli* Acid Fitness Island Attenuate Metabolite Stress at Extremely Low pH and Mediate a Cell Density-Dependent Acid Resistance. J. Bacteriol. 189, 2759–2768.

Matsumoto, Y., Ikeda, F., Kamimura, T., Yokota, Y., Mine, Y., 1988. Novel plasmid-mediated beta-lactamase from *Escherichia coli* that inactivates oxyimino-cephalosporins. Antimicrob. Agents Chemother. 32, 1243–1246.

Maurer, L.M., Yohannes, E., Bondurant, S.S., Radmacher, M., Slonczewski, J.L., 2005a. pH Regulates Genes for Flagellar Motility, Catabolism, and Oxidative Stress in *Escherichia coli* K-12. J. Bacteriol. 187, 304–319.

McLennan, A.G., Barnes, L.D., Blackburn, G.M., Brenner, C., Guranowski, A., Miller, A.D., Rovira, J.M., Rotllán, P., Soria, B., Tanner, J.A., Sillero, A., 2001. Recent progress in the study of the intracellular functions of diadenosine polyphosphates: Intracellular Diadenosine Polyphosphates. Drug Dev. Res. 52, 249–259.

McMurry, J., 1992. Organic chemistry. Brooks/Cole Pub., Pacific Grove, Calif.

McNicholas, P.M., Rech, S.A., Gunsalus, R.P., 1997. Characterization of the ModE DNA-binding sites in the control regions of *modABCD* and *moaABCDE* of *Escherichia coli*. Mol. Microbiol. 23, 515–524.

Mellies, J.L., Lorenzen, E., 2014. Enterohemorrhagic *Escherichia coli* Virulence Gene Regulation. Microbiol. Spectr. 2.

Miller, J.M., Lin, J., Li, T., Lucius, A.L., 2013. *E. coli* ClpA Catalyzed Polypeptide Translocation Is Allosterically Controlled by the Protease ClpP. J. Mol. Biol. 425, 2795–2812.

Mitchell, P. 1961. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature, 1961, Vol.191(4784), p.144

Mulcahy, L.R., Isabella, V.M., Lewis, K., 2014. *Pseudomonas aeruginos*a Biofilms in Disease. Microb. Ecol. 68, 1–12.

Mytilinaios, I., Salih, M., Schofield, H.K., Lambert, R.J.W., 2012. Growth curve prediction from optical density data. Int. J. Food Microbiol. 154, 169–176.

Nagoba, B.S., Deshmukh, S.R., Wadher, B.J., Patil, S.B., 1997. Acetic acid treatment of pseudomonal postoperative wound infection. J. Hosp. Infect. 36, 243–244.

Nagoba, B.S., Gandhi, R.C., Hartalkar, A.R., Wadher, B.J., Selkar, S.P., 2010. Simple, effective and affordable approach for the treatment of burns infections. Burns 36, 1242–1247.

Nishino, K., Inazumi, Y., Yamaguchi, A., 2003. Global Analysis of Genes Regulated by EvgA of the Two-Component Regulatory System in *Escherichia coli*. J. Bacteriol. 185, 2667–2672.

Nostro, A., Cellini, L., Ginestra, G., D'Arrigo, M., di Giulio, M., Marino, A., Blanco, A.R., Favaloro, A., Bisignano, G., 2014. Staphylococcal biofilm formation as affected by type acidulant. APMIS 122, 648–653.

Nurse, P., Marians, K.J., 2013. Purification and Characterization of *Escherichia coli MreB* Protein. J. Biol. Chem. 288, 3469–3475.

Olavarria, K., De Ingeniis, J., Zielinski, D.C., Fuentealba, M., Munoz, R., McCloskey, D., Feist, A.M., Cabrera, R., 2014. Metabolic impact of an NADH-producing glucose-6-phosphate dehydrogenase in *Escherichia coli*. Microbiology 160, 2780–2793.

O'Toole, G.A., 2011. Microtiter Dish Biofilm Formation Assay. J. Vis. Exp.

Owen, C.R., 1946. Acetic acid inhibition of gram-negative bacilli in culture media. J. Bacteriol. 52, 353–356.

Padan, E., 2014. Functional and structural dynamics of NhaA, a prototype for Na+ and H+ antiporters, which are responsible for Na+ and H+ homeostasis in cells. Biochim. Biophys. Acta BBA - Bioenerg. 1837, 1047–1062.

Pan, J.W., Macnab, R.M., 1990. Steady-state measurements of *Escherichia coli* sodium and proton potentials at alkaline pH support the hypothesis of electrogenic antiport. J. Biol. Chem. 265, 9247–9250.

Paquin, C.E., Adams, J., 1983. Relative fitness can decrease in evolving asexual populations of S. cerevisiae. Nature 306, 368–370.

Park, Y.J., Chen, J., 2015. Control of the biofilms formed by curli- and cellulose-expressing Shiga toxin–producing *Escherichia coli* using treatments with organic acids and commercial sanitizers. J. Food Prot. 78, 990–995.

Patrick, W.M., Quandt, E.M., Swartzlander, D.B., Matsumura, I., 2007. Multicopy suppression underpins metabolic evolvability. Mol. Biol. Evol. 24, 2716–2722.

Pauli, G., Overath, P., 1972. ato Operon: a highly inducible system for acetoacetate and butyrate degradation in *Escherichia coli*. Eur. J. Biochem. 29, 553–562.

Peng, L., Shimizu, K., 2004. Effect of ppc gene knockout on the metabolism of *Escherichia coli* in view of gene expressions, enzyme activities and intracellular metabolite concentrations. Appl. Microbiol. Biotechnol.

Phillips, I., Lobo, A.Z., Fernandes, R., Gundara, N.S., 1968. Acetic acid in the treatment of superficial wounds infected by *Pseudomonas aeruginosa*. Lancet Lond. Engl. 1, 11–14.

Phillips, J., Murray, P., Kirk, P. (Eds.), 2001. The biology of disease, 2nd ed. ed. Blackwell Science, Oxford ; Malden, MA.

Plague, G.R., Boodram, K.S., Dougherty, K.M., Bregg, S., Gilbert, D.P., Bakshi, H., Costa, D., 2017. Transposable Elements Mediate Adaptive Debilitation of Flagella in Experimental *Escherichia coli* Populations. J. Mol. Evol. 84, 279–284.

Plumbridge, J., 2001. DNA binding sites for the Mlc and NagC proteins: regulation of nagE, encoding the N-acetylglucosamine-specific transporter in *Escherichia coli*. Nucleic Acids Res. 29, 506–514.

Plumbridge, J., Vimr, E., 1999. Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by *Escherichia coli.* J. Bacteriol. 181, 47–54.

Potron, A., Poirel, L., Nordmann, P., 2015. Emerging broad-spectrum resistance in *Pseudomonas aeruginosa* and Acinetobacter baumannii: Mechanisms and epidemiology. Int. J. Antimicrob. Agents 45, 568–585.

Rahman, M., Hasan, M.R., Shimizu, K., 2008. Growth phase-dependent changes in the expression of global regulatory genes and associated metabolic pathways in *Escherichia coli*. Biotechnol. Lett. 30, 853–860.

Rajaraman, E., Agarwal, A., Crigler, J., Seipelt-Thiemann, R., Altman, E., Eiteman, M.A., 2016. Transcriptional analysis and adaptive evolution of *Escherichia coli* strains growing on acetate. Appl. Microbiol. Biotechnol. 100, 7777–7785.

Rao, A.K., Mendicino, J., 1978. Synthesis of UDN-N-[1-14C]acetyl d-glycosamine and UDP-N-[1-14C]acetyl-d-galactosamine from [1-14C]acetate. Anal. Biochem. 91, 490–495.

Rasamiravaka, T., Labtani, Q., Duez, P., El Jaziri, M., 2015. The Formation of Biofilms by *Pseudomonas aeruginosa* : A Review of the Natural and Synthetic Compounds Interfering with Control Mechanisms. BioMed Res. Int. 2015, 1–17.

Richard, H., Foster, J.W., 2007. Sodium regulates *Escherichia coli* acid resistance, and influences GadX- and GadW-dependent activation of *gadE*. Microbiology 153, 3154–3161.

Roe, A.J., McLaggan, D., Davidson, I., O'Byrne, C., Booth, I.R., 1998. Perturbation of anion balance during inhibition of growth of *Escherichia coli* by weak acids. J. Bacteriol. 180, 767–772.

Roy, S., Elgharably, H., Sinha, M., Ganesh, K., Chaney, S., Mann, E., Miller, C., Khanna, S., Bergdall, V.K., Powell, H.M., Cook, C.H., Gordillo, G.M., Wozniak, D.J., Sen, C.K., 2014. Mixed-species biofilm compromises wound healing by disrupting epidermal barrier function: Preclinical burn wound biofilm infection. J. Pathol. 233, 331–343.

Rudolph, B., Gebendorfer, K.M., Buchner, J., Winter, J., 2010. Evolution of *Escherichia coli* for Growth at High Temperatures. J. Biol. Chem. 285, 19029–19034.

Rybtke, M., Hultqvist, L.D., Givskov, M., Tolker-Nielsen, T., 2015. *Pseudomonas aeruginosa* Biofilm Infections: Community Structure, Antimicrobial Tolerance and Immune Response. J. Mol. Biol. 427, 3628–3645.

Ryssel, H., Kloeters, O., Germann, G., Schäfer, T., Wiedemann, G., Oehlbauer, M., 2009. The antimicrobial effect of acetic acid—An alternative to common local antiseptics? Burns 35, 695–700.

Sabo, D.L., Boeker, E.A., Byers, B., Waron, H., Fischer, E.H., 1974a. Purification and physical properties of inducible *Escherichia coli* lysine decarboxylase. Biochemistry 13, 662–670.

Saier Jr., M.H., 2015. The Bacterial Phosphotransferase System: New Frontiers 50 Years after Its Discovery. J. Mol. Microbiol. Biotechnol. 25, 73–78.

Sandoval, N.R., Mills, T.Y., Zhang, M., Gill, R.T., 2011. Elucidating acetate tolerance in *E. coli* using a genome-wide approach. Metab. Eng. 13, 214–224.

Sawers, R.G., Boxer, D.H., 1986. Purification and properties of membrane-bound hydrogenase isoenzyme 1 from anaerobically grown Escherichia coli K12. Eur. J. Biochem. 156, 265–275.

Schembri, M., Zakour, N., Phan, M.-D., Forde, B., Stanton-Cook, M., Beatson, S., 2015. Molecular Characterization of the Multidrug Resistant *Escherichia coli* ST131 Clone. Pathogens 4, 422–430.

Schmid, R., Bakker, E.P., Celik, C., Rahe, E., Harms, C., Domoto, Y., Stumpe, S., Nakamura, T., 2001. Identification of the ABC protein SapD as the subunit that confers ATP dependence to the K+-uptake systems TrkH and TrkG from *Escherichia coli* K-12. Microbiology 147, 2991–3003.

Schmidt, C., 2015. Mental Health: Thinking from the gut. Nature 518, S12–S15.

Selkrig, J., Mosbahi, K., Webb, C.T., Belousoff, M.J., Perry, A.J., Wells, T.J., Morris, F., Leyton, D.L., Totsika, M., Phan, M.-D., Celik, N., Kelly, M., Oates, C., Hartland, E.L., Robins-Browne, R.M., Ramarathinam, S.H., Purcell, A.W., Schembri, M.A., Strugnell, R.A., Henderson, I.R., Walker, D., Lithgow, T., 2012. Discovery of an archetypal protein transport system in bacterial outer membranes. Nat. Struct. Mol. Biol. 19, 506–510, S1.

Shi, R., Villarroya, M., Ruiz-Partida, R., Li, Y., Proteau, A., Prado, S., Moukadiri, I., Benitez-Paez, A., Lomas, R., Wagner, J., Matte, A., Velazquez-Campoy, A., Armengod, M.-E., Cygler, M., 2009. Structure-Function Analysis of *Escherichia coli* MnmG (GidA), a Highly Conserved tRNA-Modifying Enzyme. J. Bacteriol. 191, 7614–7619.

Shi, X., Bennett, G.N., 1994. Effects of rpoA and cysB mutations on acid induction of biodegradative arginine decarboxylase in *Escherichia coli*. J. Bacteriol. 176, 7017–7023.

Sim, M., Koirala, S., Picton, D., Strahl, H., Hoskisson, P.A., Rao, C.V., Gillespie, C.S., Aldridge, P.D., 2017. Growth rate control of flagellar assembly in *Escherichia coli* strain RP437. Sci. Rep. 7, 41189.

Sloss, J.M., Cumberland, N., Milner, S.M., 1993. Acetic acid used for the elimination of *Pseudomonas aeruginosa* from burn and soft tissue wounds. J. R. Army Med. Corps 139, 49–51.

Spehr, V., Schlitt, A., Scheide, D., Guénebaut, V., Friedrich, T., 1999. Overexpression of the *Escherichia coli nuo* -Operon and Isolation of the Overproduced NADH:Ubiquinone Oxidoreductase (Complex I) †. Biochemistry 38, 16261–16267.

Stincone, A., Daudi, N., Rahman, A.S., Antczak, P., Henderson, I., Cole, J., Johnson, M.D., Lund, P., Falciani, F., 2011a. A systems biology approach sheds new light on Escherichia coli acid resistance. Nucleic Acids Res. 39, 7512–7528. https://doi.org/10.1093/nar/gkr338

Sun, Y., Fukamachi, T., Saito, H., Kobayashi, H., 2011. ATP Requirement for Acidic Resistance in *Escherichia coli*. J. Bacteriol. 193, 3072–3077.

Tipney, H., Hunter, L., 2010. An introduction to effective use of enrichment analysis software. Hum. Genomics 4, 202.

Tonner, P.D., Darnell, C.L., Engelhardt, B.E., Schmid, A.K., 2017. Detecting differential growth of microbial populations with Gaussian process regression. Genome Res. 27, 320–333.

Tramonti, A., De Canio, M., De Biase, D., 2008a. 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. Mol. Microbiol.

Tsai, M.-F., McCarthy, P., Miller, C., 2013. Substrate selectivity in glutamate-dependent acid resistance in enteric bacteria. Proc. Natl. Acad. Sci. 110, 5898–5902.

Tucker, D.L., Tucker, N., Conway, T., 2002. Gene Expression Profiling of the pH Response in *Escherichia coli*. J. Bacteriol. 184, 6551–6558.

Unden, G., Bongaerts, J., 1997. Alternative respiratory pathways of Escherichia coli: energetics and transcriptional regulation in response to electron acceptors. Biochim. Biophys. Acta BBA - Bioenerg. 1320, 217–234.

Vera, A., Arís, A., Carrió, M., González-Montalbán, N., Villaverde, A., 2005. Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies. J. Biotechnol. 119, 163–171.

Vinolo, M.A.R., Rodrigues, H.G., Nachbar, R.T., Curi, R., 2011. Regulation of inflammation by short chain fatty acids. Nutrients 3, 858–876.

Vivijs, B., Aertsen, A., Michiels, C.W., 2016. Identification of genes required for growth of *Escherichia coli* MG1655 at moderately low pH. Front. Microbiol. 7.

Walker, N.J., Clark, E.A., Ford, D.C., Bullifent, H.L., McAlister, E.V., Duffield, M.L., Acharya, K.R., Oyston, P.C.F., 2012. Structure and function of cytidine monophosphate kinase from Yersinia pseudotuberculosis, essential for virulence but not for survival. Open Biol. 2, 120142.

Wang, T., Si, M., Song, Y., Zhu, W., Gao, F., Wang, Y., Zhang, L., Zhang, W., Wei, G., Luo, Z.-Q., Shen, X., 2015. Type VI Secretion System Transports Zn2+ to Combat Multiple Stresses and Host Immunity. PLoS Pathog. 11, e1005020.

West, I.C., Mitchell, P., 1974. Proton/sodium ion antiport in *Escherichia coli*. Biochem. J. 144, 87–90.

Yabanoglu, H., Basaran, O., Aydogan, C., Azap, O.K., Karakayali, F., Moray, G., 2013. Assessment of the Effectiveness of Silver-Coated Dressing, Chlorhexidine Acetate (0.5%), Citric Acid (3%), and Silver Sulfadiazine (1%) for Topical Antibacterial Effects Against the Multi-Drug Resistant *Pseudomonas Aeruginosa* Infecting Full-Skin Thickness Burn Wounds on Rats. Int. Surg. 98, 416–423.

Yamamoto, K., Matsumoto, F., Minagawa, S., Oshima, T., Fujita, N., Ogasawara, N., Ishihama, A., 2009. Characterization of CitA-CitB Signal Transduction Activating Genes Involved in Anaerobic Citrate Catabolism in *Escherichia coli*. Biosci. Biotechnol. Biochem. 73, 346–350.

Yamanaka, Y., Oshima, T., Ishihama, A., Yamamoto, K., 2014. Characterization of the YdeO regulon in *Escherichia coli*. PloS One 9, e111962.

Yang, B., Feng, L., Wang, F., Wang, L., 2015. Enterohemorrhagic *Escherichia coli* senses low biotin status in the large intestine for colonization and infection. Nat. Commun. 6.

Youenou, B., Brothier, E., Nazaret, S., 2014. Diversity among strains of *Pseudomonas aeruginosa* from manure and soil, evaluated by multiple locus variable number tandem repeat analysis and antibiotic resistance profiles. Res. Microbiol. 165, 2–13.

Zhao, J., 2004. Effect of zwf gene knockout on the metabolism of *Escherichia coli* grown on glucose or acetate. Metab. Eng. 6, 164–174.

Zhu, Z., Surujon, D., Pavao, A., Bento, J., van Opijnen, T., 2018. Forecasting bacterial survival-success and adaptive evolution through multi-omics stress response-mapping, network analyses and machine learning. BioRxiv, Aug 9. 2018

Zorraquino, V., Kim, M., Rai, N., Tagkopoulos, I., 2016. The genetic and transcriptional basis of short and long term adaptation across multiple stresses in *Escherichia coli*. Mol. Biol. Evol. msw269.